# Economies of Scale and Heterogeneity in the Provision of Public Goods: Evidence from School Consolidation in China

### Xiang Ma<sup>\*</sup>

### April 8, 2016

### Abstract

This paper examines the trade-offs faced by central decision-makers responsible for delivering local public goods. From 2000 to 2010, China closed more than 50% of its primary schools. While smaller and academically weaker schools were closed and students' test scores increased, students and their families paid additional costs due to a longer home-to-school commute post school closure. This paper builds and estimates a structural model of local governments' decision-making to understand these stylized facts and to recover the preferences of local authorities. I pay particular attention to cases when there may be heterogeneous productivity across schools, increasing returns in schooling production, and home-to-school commuting costs. The estimation shows that schools that were closed had lower mean productivity. Some very small schools with relatively high productivity were also closed. I use a moment-inequalities approach to estimating on the weight local governments placed on the cost to families attending a more distant school. The very low upper bound of this estimate implies that distance cost was under-weighted in local governments' decision.

**Keywords:** School Mergers; Education Production Function; Economies of Scale; Moment Inequalities.

### **JEL codes:** D61; H52; I21; O12.

<sup>\*</sup>Department of Economics, Yale University (e-mail: xiang.ma@yale.edu). The author is grateful to Daniel Keniston, Nancy Qian, Christopher Udry and Eric Weese for advice, and to Jose-Antonio Espin-Sanchez, Dean Karlan, Costas Meghir, Ahmed Mushfiq Mobarak, Nicholas Ryan, Laura Schechter and seminar participants at the Yale Development Lunch, the Yale Development Workshop, NEUDC 2015 at Brown University, the University of Hong Kong, the Chinese University of Hong Kong, the Hong Kong University of Science and Technology for helpful comments. Special thanks to Emily Hannum for providing the survey data, to the Gansu Education Bureau for providing the administrative data, to Yajing Chen, Shijun He, Shujuan Liu, Runyan Lu, Xuemei Peng, Shixia Qu, Wenjing Tao, Tian Xia, and Wei Xu for help in conducting the telephone survey. Financial support from the Ryoichi Sasakawa Young Leaders Fellowship is greatly acknowledged. All errors are my own.

## 1 Introduction

The provision of public goods, such as basic health care and public education, is an essential function of local governments and is especially important for developing countries seeking to end poverty and boost productivity. There are two aspects of the provision of public goods: physical access and quality. Physical access to public goods in developing countries has been improving over time. For example, according to the World Bank, the gross enrollment ratio of primary school worldwide reached 108% in 2013.<sup>1</sup> Recently policymakers and researchers have begun to focus on the quality of public goods, documenting the prevalence of public goods of poor quality and studying policies that could improve quality. For example, poor public education in some developing countries is due in part to teacher absenteeism (Duflo, Hanna, and Ryan 2012).

This paper examines the trade-offs faced by central decision-makers who are responsible for delivering local public goods, using a school consolidation program implemented in rural China as an example. From 2000 to 2010, China closed more than 50% of its primary schools. The main reason is that demographic change stemming from the one-child policy led to a sharp decline in the number of children of primary school age in rural China. In 2001, China's State Council ordered the school consolidation program to begin. Each county government managed the process of consolidation within their administrative areas and played the role of the local planner. Therefore the observed pattern of school consolidation reveals countylevel governments' objectives and incentives. I build and estimate a structural model of local county governments' decision-making to recover the preferences of the local authorities who choose the spatial allocation of these public goods.

The structural model incorporates two key factors: economies of scale arising from large school size and heterogeneity. I consider both heterogeneity in home-to-school commuting

<sup>&</sup>lt;sup>1</sup>http://data.worldbank.org/topic/education accessed on August 15th, 2015.

distance and heterogeneity in unobserved school productivity. The estimation of the model takes advantage of the exogenous demographic change resulting from the one-child policy and the exogenous policy shift initiated by China's State Council—in 2001, the State Council relaxed the binding constraint of one primary school per village and allowed its county governments to close primary schools in a large scale. The estimation results show that the key features of the school consolidation can be understood through an analysis of the trade-offs among economies of scale, heterogeneity in distance and heterogeneity in unobserved school productivity.<sup>2</sup> The estimation shows that schools that were closed had lower mean productivity. Some very small schools with relatively high productivity were also closed. I use a moment-inequalities approach to estimating on the weight local governments place on the cost to families attending a more distant school. The very low upper bound of this estimate implies that distance cost was under-weighted in local governments' decision.

Most empirical literature explaining the variation in allocation of public goods focuses on collective action by local communities, i.e., the demand side of public goods. By collective action, communities or groups compete in various ways to lay claim to limited public resources. The various characteristics of communities or groups, such as taste (Chattopadhyay and Duflo 2004), group size (Olson 2002), or ethnic diversity (Alesina, Baqir, and Easterly 1999), determine their ability to collectively invest in activities that bring them public goods. However, as Banerjee et al. summarize the literature on collective action (Banerjee, Iyer, and Somanathan 2008):

These studies...account for a small part of the observed variation in provision. Access to public goods is often better explained by 'top-down' interventions rather than the 'bottom-up' processes highlighted in the collective action literature... If public good access were determined primarily by local population characteristics, we could rarely see rapid changes in such access, since many of these characteristics (religion, caste, ethnicity) change very slowly over time.

<sup>&</sup>lt;sup>2</sup>Weese (2015) similarly shows that a simple model of a central planner incorporating the trade-off between economies of scale and geographic distance cost in provision of public goods can well explain the observed pattern of municipal mergers in Meiji Japan.

By focusing on the demand side, these studies ignore the supply of public goods.<sup>3</sup> In practice, central planners, especially those of developing countries, determine where and how much to invest in public goods. This central decision-making is much less studied in current empirical literature. Regarding central decision-making, the first question is what the preferences of central decision makers are and what incentives or trade-offs are involved when central planners make decisions. A second question is how to measure the welfare impacts of policies made by central planners. One strand of literature in development economics claims that poor countries are poor because their governments make policies without regard to people's welfare (Baland, Moene, and J. A. Robinson 2010). This paper recovers the preferences of central decision-makers, examines the trade-offs involved in central decision-making, estimates the welfare impacts of the school consolidation program and thus contributes to the literature.

Although almost absent from the empirical literature, central decision-making is a focus of the theoretical literature, especially the literature on optimal partition of political jurisdiction.<sup>4</sup> For example, Alesina and Spolaore (1997) argue that the trade-off between the benefits of economies of scale resulting from large jurisdictions and the costs of preference heterogeneity of large and diverse populations is an important explanation of the number and size of nations; Bolton and Roland (1997) focuses instead on the trade-off between economies of scale and redistribution conflicts arising from differences in income distribution across regions. Alesina, Baqir, and C. Hoxby (2004) test these ideas using American school districts and school attendance areas and find the trade-off between economies of scale and racial heterogeneity tends to be larger in magnitude and more robust empirically than the

<sup>&</sup>lt;sup>3</sup>Demand side underlines willingness to pay or ability to pay, and supply side underlines efficiency. Suppose there is a public school in village A rather than village B. The explanation according to the demand side literature would be that it is because it is easier for village A to act collectively, while the explanation according to the supply side literature would be that given limited budget, local planners think it is more efficient to place a public school in village A rather than in village B.

<sup>&</sup>lt;sup>4</sup>Here, the public good is the "government", which provides a bundle of administrative, judicial, economic services, and public policies.

trade-off between economies of scale and income heterogeneity. This paper echoes the point emphasized in the aforementioned theoretical literature: I find that the trade-offs between economies of scale and heterogeneity in preferences played an important role in shaping the observed school consolidation pattern. Moreover, I extend the literature by modeling and estimating the heterogeneity in productivity (in translating inputs into public good services) across local providers.

This paper is organized as follows. Section 2.1 and Section 2.2 introduce the school consolidation program and its institutional background respectively. Section 3 describes the data sets: the Gansu Survey of Children and Families (GSCF), the Gansu Administrative School Data and a survey I conducted. In Section 4.1, I use descriptive regressions to establish three stylized facts. First, smaller and academically weaker schools were more likely to be closed during school consolidation. Second, post school consolidation, the test scores of students whose schools were closed increased, while the test scores of students of receiving schools did not significantly change. Third, students whose schools were closed paid additional costs due to a longer commute between home and school after school consolidation. In Section 4.2, I build a structural model of local governments' decision-making to rationalize these stylized facts. The structural model consists of a production function of per capita school quality in Section 4.2.1, and a payoff function for local governments in Section 4.2.2. The production function of school quality incorporates two factors: economies of scale and unobserved heterogeneity in school productivity. The payoff function for local governments is modeled as a sum of students' test scores minus students' home-to-school distance cost and schools' running cost.

In Section 5.1, I discuss the challenges in estimating the structural model: the heterogeneity in production function is partially unobserved and the number of choices facing local governments as they consolidated schools is unusually large. In Section 5.2, I provide the estimation strategies to deal with these econometric challenges. I show how the insight by Olley and Pakes (1996) can be used to solve the problem of unobserved heterogeneity in school production and how a moment-inequalities approach can be taken to circumvent the problem caused by the unusual number of choices facing local governments. I discuss the estimation results in Section 5.3. In Section 6, I further examine the functional form assumptions and alternative mechanisms. I conclude in Section 7.

## 2 Background

### 2.1 The Rural Primary School Consolidation Program

In 2001, China's State Council announced the Rural Primary School Consolidation Program. County governments were allowed to consolidate primary schools in rural areas with the condition that the ability of students to attend school after consolidation would not be compromised—elementary education is compulsory by law in China and all students have to be able to attend schools even after the school consolidation (The China State Council 2001). The foremost reason for the policy change was that demographic change stemming from the one-child policy led to a sharp decline in the number of primary school age children in rural China. Figure 1 shows that the number of primary schools across China dropped from about 550 thousand in 2000 to about 250 thousand in 2010, while the number of students in primary school dropped from about 130 million to about 100 million. Within ten years, one out of two primary schools was closed in China. Moreover, primary schools were closed faster than students decreased, which suggest there might be a failure to close schools that should have been closed previously.

Because of the *hukou* system, almost all rural students could only attend primary schools in or nearby their villages before the school consolidation.<sup>5</sup> During the school consolidation,

 $<sup>{}^{5}</sup>$ In 1958, the Chinese government officially promulgated the family register system to control the movement of people between urban and rural areas. Under the *hukou*, each individual had an official place of residence, and the documents verifying residence are similar to a passport. Individuals were broadly catego-

each county-level government was responsible for school mergers within its administrative areas. Students whose schools were closed were reallocated to another school designated by local governments. Either one whole school was merged into another school, or certain grades were merged into another school. In many cases, multiple schools merged into one simultaneously. Teachers whose schools were closed could be fired or reallocated by local county governments.<sup>6</sup> Teachers and students whose schools were closed were not necessarily reallocated into the same school. Each county government was also responsible for the allocation of teachers within its district. There was great heterogeneity and little specific guidance by the State Council. Local governments claimed that the massive consolidation could help make use of scarce educational resources more economically and efficiently and improve the quality of education for all rural students. The most commonly cited reason for school consolidation by local governments was economies of scale (Personal Interviews).<sup>7</sup> If the inputs of the production of education are teachers and students and the output is the total amount of test scores, economies of scale implies that when the number of teachers and the number of students are both doubled, average test scores increase. Economies of scale could result from division of labor and specialization on the part of teachers. Instructors in a small school with a limited number of teachers often taught multiple subjects at the same time, with the extreme case that one teacher had to teach all subjects if he or she

rized as a "rural" or "urban" resident. Urban residents received state-allocated jobs and access to an array of social services, including food rations, grain subsidies, employer-provided housing, free education, medical care and old-age pensions, while rural residents were expected to work on the collective farms. People were allowed to work legally, and to receive social security benefits only in their place of residence. A change in official place of residence can be granted only by permission, similar to a local authority granting a visa. Rural-to-urban migration should be allowed only if compatible with economic development and was controlled strictly. Moreover, parents pass their *hukou* status to their children, solidifying these administrative categories into inheritable social identities.

<sup>&</sup>lt;sup>6</sup>In China, teachers are either substitute or regular. Regular teachers are guaranteed to be employed and substitute teachers could be fired. Many substitute teachers were fired during the school consolidation. For example, it was reported that by 2004, Gansu Province of China had closed 2585 primary and secondary schools and fired 9876 substitute teachers (Peng 2005, page 277).

<sup>&</sup>lt;sup>7</sup>In 2013, I visited more than ten counties of Gansu Province and interviewed government officials in local education bureaus about the school consolidation. The official documents provided to me during the interviews also confirmed these points.

was the only teacher in a school, which damaged the quality of education.<sup>8</sup> After school consolidation, given the increase in the total number of teachers, a teacher in a larger school could specialize in one subject he or she was good at. Economies of scale could also result from the savings in fixed costs, such as gas and coal expenditure in the winter.

The massive rural primary school merger program generated major controversy in China. Media reports focused on the negative side associated with the increased distance to schools after school consolidation. Nonetheless, little research was done to systematically evaluate the impacts of the program, with two exceptions. Liu et al. (2010) find that overall the primary school mergers did not harm the academic performance of students. Mo et al. (2012) further confirm that there is a large positive resource effect associated with the transfers of students from less centralized schools to more centralized schools. There papers provide important descriptive evidence regarding the pattern of school consolidation. However, given limited data, neither can they address the endogeneity of school closure, nor can they evaluate the overall impact.

### 2.2 Institutional Background

The geographical distribution of rural primary schools was largely determined prior to the 1980s. After the founding of the People's Republic of China in 1949, China swung between a "pro-elite education policy" and a "pro-mass education policy" before the beginning of the Cultural Revolution in 1966. The pro-elite education policy focused on cultivating skilled graduates for heavy industries and national defense. The pro-mass education policy, advocated by Chairman Mao, empowered the proletariat and the peasant masses, which was the mission of the Communist Party. In the Cultural Revolution, Mao's mass education policy was implemented in practice. The official standard then was a primary school in

<sup>&</sup>lt;sup>8</sup>In contrast to primary schools in the U.S., teachers are classified by subjects and teach different subjects in the primary schools of China.

every production brigade, known today as a village and a middle school in every commune, known today as a town (Pepper 1996, page 418).<sup>9</sup> As a result, primary-school education had been mostly expanded before the beginning of reform in 1978. In 1976, when the Cultural Revolution ended, the net enrollment rate of school-age children was 96% (Department of Planning 1984, page 226). The high literacy rate is generally regarded as one important factor in China's success in its reform starting from the earlier 1980s. As Sen puts it (1999, page 42)

While pre-reform China was deeply skeptical of markets, it was not skeptical of basic education and widely shared health care. When China turned to marketization in 1979, it already had a highly literate people, especially the young, with good schooling facilities across the bulk of the country.

After the Cultural Revolution, economic-oriented incentives replaced politically-oriented incentives for local governments.<sup>10</sup> In 2001, the central government announced the Rural Primary School Consolidation Program and relaxed the binding constraint of one primary school per village and implicitly encouraged school consolidation.

During the period of school consolidation, county governments were responsible for local compulsory education. They managed principals, teachers, their wages, the school consolidation, and the education budget (The China State Council 2001). The education budget of each school was formula-based and depended on its number of teachers and its number of students. The budget of each school had two main components: the salary of its teachers and the expenses of teaching (Yang 2001, page 585–586).<sup>11</sup> The budget for teachers' salary, of course, depended on the number of teachers. The budget for teaching expenses

<sup>&</sup>lt;sup>9</sup>The tiers of government in China, in the descending order, are the central government, province, prefecture, county, town and village. By law, villages are not local administrative governments. Nonetheless, in practice, they function as the lowest-level governments.

<sup>&</sup>lt;sup>10</sup>The total number of rural high schools decreased by more than 80% after the Cultural Revolution between 1977 and 1983. Exploiting the sharp decrease, Zhang (2014) studies the effect of the negative shock to maternal education on infant health, without any significant effect found.

<sup>&</sup>lt;sup>11</sup>The teaching expenses covered expenses of teachers' training, lab equipment, sports equipment, water, electricity, heating, library books, school maintenance, etc.

depended on the number of students. The per-student budget for teaching expenses was adjusted frequently and increased over time.<sup>12</sup> This budget structure implies that the amount of transfer that local county-level governments could get had little to do with the school consolidation within their districts and thus was not the driving force underlying the school consolidation.<sup>13</sup>

## 3 Data

### 3.1 Data Sources

I employ three main data sources. The first data set, the Gansu Survey of Children and Families (GSCF), is a longitudinal, multi-level study of rural children's welfare outcomes, including education, health, and psycho-social development.<sup>14</sup> Data was collected in 2000, 2004, 2007, and 2009 from Gansu Province, China. Gansu has 87 counties in total. On average, each county has 19 towns and each town has 11 villages. Due to limitations in accessing the data, this paper only uses data from 2000, 2004 and 2007. In 2000, the GSCF used a four-stage stratified sampling procedure to draw a sample of 2000 children aged 9 to 12 living in 20 counties, 42 townships and 100 villages. Figure 2 shows the sampling area. In 2004, the same 2000 children were resurveyed. Their eldest younger brother or sister was included in the survey. In 2007, the same 2000 children were surveyed again. In addition,

<sup>&</sup>lt;sup>12</sup>Taking Gansu Province as an example, the budget for teaching expenses in primary schools was 150 Yuan per student in 2007 (The Ministry of Finance and The Ministry of Education 2007), 300 Yuan per student in 2009 (The Ministry of Finance 2010) and 400 Yuan per student in 2011 (The National People's Congress 2011).

<sup>&</sup>lt;sup>13</sup>The budgetary source for teaching expenses saw a significant change during this period. Taking Gansu Province as an example, in 2005, every student had to pay miscellaneous fee 94-184 Yuan for teaching expenses, and the remaining part 20-90 Yuan per student was covered by public finance (The Gansu Finance Bureau and The Gansu Education Bureau 2005). In 2006, the miscellaneous fee was exempted and was covered by public finance: the finance from the central government covered 80% of the original miscellaneous fee per student, and the finance from the provincial government covered the remaining 20% part. Moreover, in 2006, the textbook fee was also exempted and was all covered by the finance from the central government (The Gansu Education Bureau 2006).

<sup>&</sup>lt;sup>14</sup>Interested readers can refer to the website http://china.pop.upenn.edu for details.

a new sample of 1400 children aged 9 to 15 from the same villages were also surveyed. In each wave, the children's parents, families, villages, schools and homeroom teachers were surveyed; standardized Chinese and math tests were administered.<sup>15</sup> Therefore, there are cross-cohort difference embedded in the sampling design of the GSCF, since in each later wave, there was a new cohort of children included in the survey. I will take advantage of this cross-cohort difference in my following difference-in-difference identification strategy.

The second data set is the Gansu Administrative School Data, kindly provided by the Gansu Education Bureau. Every primary school in Gansu Province reports its basic information to the Education Bureau at the beginning of each academic year (September), including its name and address, class size by grade, number of students by grade and age, number of teachers by age, education, professional rank and teaching subject, students and teachers' turnover, and various school facilities.<sup>16</sup> This data is available for 2001, 2004, 2006, 2007, 2008 and 2009. This administrative data will allow me to check which schools were closed during the school consolidation. Moreover, the GSCF and the administrative data can be linked through the village names in the GSCF and the school addresses in the administrative data.<sup>17</sup>

The final data set records the student flow after the school closure, that is, which schools students went to attend after their original schools were closed. I conducted a telephone survey specially inquiring the student flow of every village in those 20 counties sampled by the GSCF.<sup>18</sup> The survey was town-based, since almost all school consolidation happened within the scope of a town. The telephone survey finds that cross-town school consolidation was rare. I will use this third data set to calculate the home-to-school commuting distance

<sup>&</sup>lt;sup>15</sup>Glewwe, Huang, and Park (2013) contains detailed introduction about the GSCF.

<sup>&</sup>lt;sup>16</sup>There are six distinct professional ranks in the administrative school data: middle school high level, primary school high level, primary school level one, primary school level two, primary school level three, and no rank.

<sup>&</sup>lt;sup>17</sup>Almost all villages have one school at most.

<sup>&</sup>lt;sup>18</sup>Special thanks to the Gansu Education Bureau for providing the latest telephone number of each school.

after the school consolidation.

In the following section I examine the basic pattern of school consolidation using these data sets.

### 3.2 Data Description

Table 1, based on the 2000 GSCF, examines the effect of *Hukou* system on students' school choices. Among the 4115 students surveyed, less than 3% reported that they did not attend the nearest primary school. Among those who did not attend the nearest primary school, less than half cited school quality as the reason for not attending the nearest one. Table 2 shows that in 2001, when the school consolidation began, almost every village in Gansu had a primary school. Therefore, Table 1 and Table 2 are consistent with the claim that the *Hukou* system severely limited students' choices of school. In almost all cases, children had to attend the primary school in their own village.

Figure 3 shows the number of primary schools, the number of students in primary schools, and the number of full-time teachers in Gansu from 1990 to 2010. From 1990 to 2000, the change in the number of primary schools was minimal, with only about 1% decrease per year. Only starting from 2001 did the number of primary schools in Gansu begin to decrease dramatically. Though the one-child policy began in the early 1980s, not until 2000 did the number of students in primary schools begin to decrease, as the middle graph of Figure 3 shows. The upper graph of Figure 3 also shows that though the number of students and the number of primary schools decreased after 2000, the total number of full-time teachers increased from about 122,000 in 2001 to about 140,400 in 2010, about 15% increase within ten years.

Figure 4, based on the Gansu Administrative Education Data, shows that the number of schools with the first grade declined by about 29% from 2001 to 2009, while the number of schools with the fifth grade declined by about 18% over the same period. Comparison with the trend across the rest of China reveals that the number of schools elsewhere declined more. Therefore, the effects of school consolidation that we detect in Gansu are likely lower bounds for the effects elsewhere in China. Table 3 compares schools whose third grades remained open after 2001 with schools whose third grades were closed at some point after 2001. In general, schools whose third grades were closed were much smaller, with about 70 students on average in 2001, than schools whose third grades remained open, with about 200 students on average in 2001. Moreover, pupil-teacher ratio of schools whose third grades were closed after 2001 were larger, with average educational level of teachers lower, and value of per-student fixed capital much smaller. All indicators clearly show that schools whose third grades were closed after 2001 were smaller and academically weaker than schools whose third grades remained open after 2001. Comparison between schools whose first grades (or fifth grades) were closed after 2001 with schools whose first grades (or fifth grades) remained open reveals the same pattern.

Before examining the determinants of school consolidation and checking its impact on students' education, I use descriptive regressions to establish some stylized facts which later more formal econometric model can build on.

## 4 Econometric Model

### 4.1 Descriptive Regressions

Table 4, based on the Gansu Administrative School Data, checks the determinants of school consolidation. This is a linear probability model that regresses the dummy variable denoting whether a school's first grade or third grade was closed at some point after 2001 on its schooling conditions, which includes log number of students, log pupil-teacher ratio, average education of teachers and log amount of capital per capita. All schooling conditions, except the amount of per-capita capital, are significant at 1% level. A 1% decrease in the number

of students increases the probability of its being closed in the future by about 0.25%. The coefficients of the other regressors also confirm the suggestion from Table 3 that academically weaker schools tend to be closed: schools with higher pupil-teacher ratio and lower average educational level of teachers have a higher chance of being closed in the future. Moreover, the adjusted  $R^2$  shows that the explanatory power of the number of students dominates all other explanatory variables. It alone explains about 25% of the total variation. When other variables are added into the regressor list, the explanatory power of the whole regression barely changes.

After checking the determinants of school consolidation, I now turn to examine the impact of school consolidation on students' education. Both Table 3 and Table 4 show that academically weaker schools were closed. Therefore, it is plausible that students might gain from the school consolidation. Table 5, based on the GSCF, shows that this is indeed the case. Table 5 employs a difference-in-difference (DID) strategy. The first difference is across villages at the same point in time: schools in some villages were closed while in other villages schools remained open. The second is the difference across cohorts in the same village: earlier cohorts enjoyed schools open in their villages while later cohorts might have to travel to schools in other villages. The DID shows that post school consolidation, students whose schools were closed benefited from the consolidation: their Chinese-language test score increased by about 19%, significant at 5% level. The point estimate is that their math test score increased by about 2%, but this is insignificant. Table 6 examines the impacts of school consolidation on the test scores of students-stayers whose schools received students transferred from other schools due to the school consolidation. The regressions in Table 6 employ a similar DID strategy—the first is the difference across village at the same point in time: schools in some villages received students transferred from other schools due to the school consolidation while in other villages schools there did not; the second is the difference across cohorts in the same village: earlier cohorts did not experience student inflow from other villages while later cohorts studied together with students from other villages. These regressions provide no evidence that the school consolidation affected the students-stayers negatively. On the contrary, there were slightly positive effects for them, though insignificant.<sup>19</sup> Overall, post school consolidation, the test scores of students whose schools were closed increased, while the test scores of students-stayers whose schools received students from other places due to the school consolidation showed small and insignificant change.

However, the school consolidation was not without costs. Table 7, based on the same specification as Table 5 and employing the same DID strategy, shows that post school consolidation, family expenditure on bus, board, and food in school more than doubled. Due to the longer distances that students travel to school, students' leisure time, measured as the number of hours students spent watching TV per week, decreased by about 25%. The reduction in leisure time was not because students studied harder or longer, as the last column in Table 7 shows that the numbers of hours students spent on homework per week barely changed after the school consolidation. Thus, though the test scores of students' whose schools were closed increased, they paid additional costs given reduced leisure time and increased family expenditure on transportation, food and board due to a longer commute. This is clearly a trade-off.

However, the regressions in Table 4, Table 5, Table 6 and Table 7 are only descriptive. The regressions in Table 4 treat school closure decisions as independent from each other. In practice, school closure decisions were interdependent. Whether a school was closed depended on not only its own academic strength, but also whether or not there was another school nearby and if there was, whether that school was academically sound. Indeed, it is the interdependent nature of school consolidation that makes its study challenging. The regressions in Table 5, Table 6 and Table 7 do not take the endogeneity of school closure

<sup>&</sup>lt;sup>19</sup>Using evidence from Boston's Metco Program that sends mostly black students out of the Boston district into schools in the surrounding, mostly white suburban districts, Angrist and Lang (2004) also find no impact of Metco on the scores of white non-Metco students.

into account. School closures were decided by the local governments. This choice depended on the objectives of the local government as well as on the interdependent nature of school consolidation. Nor can these descriptive regressions enable me to conduct welfare analysis. Table 5 shows that the test score of students' whose schools were closed increased while Table 7 shows that these students paid additional costs. Thus, it is unclear whether there was a net gain for these students. If so, how much was the net gain? Nor is it clear that the marginal gain for students is positive.<sup>20</sup> In order to deal with the endogeneity of school closure and to calculate the net welfare effects and the marginal effects of school consolidation, it is necessary to learn the objective function of the local county-level governments. Their objective function generally will not be revealed by the effect of treatment on the treated (TT) estimated in Table 5, Table 6 and Table 7, and needs to to learned through the observed school consolidation pattern and the marginal choices that they made.

These considerations call for a more formal structural model. The formal structural model should be explicit about the local governments' decision about school closure. It should also be consistent with the stylized fact that smaller and academically weaker schools were closed and that the test scores of students whose schools were closed increased after the school consolidation but that they paid additional costs due to a longer home-to-school commuting distance. Furthermore, the estimation of the structural model allows me to conduct some counter-factual and welfare analysis.

<sup>&</sup>lt;sup>20</sup>Consider calculating the average additional distance travelled by students from closed schools, and comparing that to the average increase in test scores from closing a school. This sort of calculation might show that on average students were sent only a small additional distance in exchange for a substantial increase in test scores, suggesting that the county governments were placing substantial weight on rural students' commuting distance. This average, however, includes many closed schools that were close to other schools: it is obvious that these schools should have been closed, and including them in the average potentially obscures the fact that the government may have been closing other schools that resulted in students travelling long distances in exchange for relatively insubstantial gain in test scores. That is, even if the county governments has very strange weights in its objective function, I will probably not learn this by looking at the "average effect" shown in Table 5, Table 6 and Table 7.

### 4.2 Structural Model

### 4.2.1 Production Function of School Quality

Assume the production function of school i's quality at time t is

$$\tilde{q}_{it} = \eta_{it}\omega_{it}L^{\alpha}_{it}N^{-\beta}_{it} \tag{1}$$

 $\tilde{q}_{it}$  is per capita school quality with measurement error. In the empirical analysis, it is measured by students' residual test scores after partialling out the effect of personal and family attributes.  $\eta_{it}$  is measurement error.  $\omega_{it}$  denotes area(or school) specific productivity which is observed by local bureaucrats but not by the econometricians. This specific productivity term could result from schools' management style, comfortableness of teaching buildings for students, and whether the location of the school is convenient and attractive to teachers. In practice, it is difficult for schools in remote villages to attract good teachers, and the productivity term  $\omega_{it}$  there is low as a consequence.  $L_{it}$  is the number of quality-adjusted teachers allocated by local governments.<sup>21</sup>  $N_{it}$  is the number of students.  $\alpha$  and  $\beta$  are parameters of the production function to be estimated.  $\alpha > \beta$  implies economies of scale in school quality provision.

The assumed production function of per capita school quality (1) is consistent with findings in the literature.<sup>22</sup> First, a positive  $\beta$  (negative  $-\beta$ ) is consistent with the findings in the literature of the negative effect of large class size on students' test scores. (See, for ex-

<sup>&</sup>lt;sup>21</sup>In empirical work, the quality is adjusted according to teachers' professional ranks, since teachers' wage are set according to their professional ranks. These professional ranks are conferred by local county educational bureaus based on individual's educational level, teaching skill, teaching experience and teaching achievement. Lai, Sadoulet, and Janvry (2011) show that these professional ranks have about the same predictive power for student test scores as do school fixed effects.

 $<sup>^{22}</sup>$ Here I consider a "value-added" production function in the sense that the intermediate input does not enter the production function to be estimated. For the identification problem with the gross output production functions where the intermediate input does enter the production function to be estimated, see Gandhi, Navarro, and Rivers (2014).

ample, the papers by Krueger (1999) and Angrist and Lavy (1999)).<sup>23</sup> Second, a positive  $\alpha$  accords well with the findings of the significant positive impacts of teacher quality on student achievement and adult earnings (for example, Card and Krueger (1992), Behrman, Rosenzweig, and Taubman (1996), Chetty, Friedman, and Rockoff (2014a) and Chetty, Friedman, and Rockoff (2014b)). Third, positive productivity term  $\omega$  captures school quality, which has been studied extensively (for example, Altonji and Dunn (1996), Dearden, Ferri, and Meghir (2002), and Rivkin, Hanushek, and Kain (2005)). Rather than focusing on and trying to estimate the marginal benefit of one particular component of the production function of education quality, this paper contributes to the literature by putting all three factors into a single unified framework and estimating the whole production function. Policymakers not only need to know the magnitude of the marginal benefit of any particular factor, but also require the knowledge of the whole production function for the school consolidation program.

The production function (1) incorporates two important factors. One is the productivity term  $\omega$ . If schools are heterogeneous and  $\omega_{it}$  are different from each other, students reallocated from a school with low  $\omega_{it}$  to a school with high  $\omega_{it}$  in the school consolidation will benefit. The other is the economies of scale term  $\alpha - \beta$ . If this term is positive, school consolidation itself increases school size and benefits students as a result. That is, there are potentially two sources for welfare improvement due to school consolidation, one being students reallocation from a worse school with lower  $\omega_{it}$  to a better school with a higher  $\omega_{it}$ , the other being economies of scale realized in the consolidation. With a formal structural analysis, we can distinguish these two effects in a clear way.

The Cobb-Douglas functional from in (1) is consistent with the descriptive regressions in

 $<sup>^{23}</sup>$ C. M. Hoxby (2000) is an exception. Using variation in class size driven by idiosyncratic variation in the population, she does not find a statistically significant effect of class size on student achievement. Moreover, Deaton (2010) mentions that there might be sorting across the boundary by the parents of students who are potentially affected by Maimonides' rule so that the children in the smaller classes have richer and more educated parents than the children in the larger class, which thus might compromise the generality of the results by Angrist and Lavy (1999).

Table 8. The regressions in Table 8 are valid only when there is no unobserved heterogeneity  $\omega_{it}$ . If this is true, the difference between the coefficient of the number of students and the coefficient of the number of teachers would measure the economies of scale term  $\alpha - \beta$ , which is about 12% in the production of Chinese test score and about 4% in the production of math test score. However, when unobserved heterogeneity  $\omega_{it}$  is present, these coefficients are biased—the productivity term  $\omega_{it}$  can be positively correlated with the number of teachers, either because local governments allocate more teachers to schools with higher productivity, or schools with a higher productivity could attract good teachers; the productivity term  $\omega_{it}$  can also be positively correlated with the number of students, either because local governments directly closed schools with lower productivity and sent students to schools with higher productivity during the school consolidation, or schools with higher productivity were traditionally located in larger villages where more people who gave birth to more children. Biased estimates of  $\alpha$  and  $\beta$  generally also lead to biased estimates of economies of scale term  $\alpha - \beta$ .<sup>24</sup> If  $\alpha$ ,  $\beta$  and  $\alpha - \beta$  change significantly after the unobserved heterogeneity  $\omega_{it}$  is taken into account in the following empirical analysis, it is an indication that the unobserved heterogeneity  $\omega_{it}$  differs across locations and that it affects the allocation of teachers and school consolidation.

### 4.2.2 Payoff Function for Local Governments

Assume the utility for a student attending school in village i but living in village j at time t is

$$q_{it} = \omega_{it} L^{\alpha}_{it} N^{-\beta}_{it} \tag{2}$$

$$U_{jt} = q_{i(j)t} - \gamma d_{i(j)j} \tag{3}$$

<sup>&</sup>lt;sup>24</sup>In appendix A, I derive the biases for  $\alpha$  and  $\beta$ , and show that the direction and the magnitude of these biases generally depend on the pairwise correlation among the log number of students, the log number of teachers, and the log school productivity.

where *i* is a school and *j* is a village.  $q_{it}$  in equation (2) denotes per capita education quality without measurement error. i(j) in equation (3) represents a student who attends school in village *i* but lives in village *j*.  $d_{i(j)j}$  represents the distance from school *i* which student i(j)attends to village *j* where student i(j) lives. *i* and *j* are not necessarily different from each other. If *i* and *j* are the same,  $d_{i(j)j}$  is zero.  $\gamma$  denotes unit distance cost. Per capita utility  $U_{jt}$  for student i(j) is the benefit from attending school *i* minus distance cost to school *i* from village *j*.

Assume the payoff that a local government gets from running schools  $\{i\}$  is

$$N_{it} = \sum_{j \in J(i)} N_{j(i)t} \tag{4}$$

$$C_{it} = VC(L_{it}) + F \tag{5}$$

$$W_t = \sum_j U_{jt} N_{jt} - \sum_i C_{it} + \epsilon_t \tag{6}$$

J(i) in equation (4) denotes the collection of villages whose children attend school *i*. Equation (4) denotes that the number of students attending school *i* is the sum of the number of students in each village *j* whose children attend school *i*.  $VC(L_{it})$  is the variable cost of running school *i*, as a function of the number of teachers  $L_{it}$  in school *i*. *F* is the fixed cost for running each school. Equation (5) denotes that the total cost for running each school consists of the total variable cost and the fixed cost. The existence of the fixed cost represents another form of economies of scale that differs from the economies of scale in the production function  $\alpha - \beta$ . Therefore, school consolidation can bring in economies of scale through economies of scale  $\alpha - \beta$  in the production of educational quality directly, or through saving in fixed cost *F* for running each school. The total payoff to the local government  $W_t$ is the total utility to the students within its district  $\sum_j U_{jt}N_{jt}$ , minus the total cost for running schools  $\{i\}$ , plus  $\epsilon_t$ .  $\epsilon_t$  is a payoff-irrelevant measurement error or a payoff-relevant shock, which is to be specified more accurately in the estimation Section 5.2.1. The local government observes per capita quality  $q_{it}$  without measurement error, while we only observe  $\tilde{q}_{it}$ , per capita quality measured with error.

Several key assumptions are embedded in the model. The first assumption is that in the production function of school quality, unobserved state heterogeneity is captured by the onedimension productivity term  $\omega_{it}$  in equation (1), and that the unobserved productivity term is place specific. The place specificity of the productivity term further implies that when students from village j transfer to school i during the school consolidation, they will enjoy the same productivity as students from village i. Another assumption is that students pay additional distance costs if they attend schools outside their own villages. Local governments' objective is assumed to maximize the total payoff in equation (6), composed of total quality provision to students, total distance cost incurred by students and total running cost of schools. Every student is treated equally by local governments and they enjoy the same weight in the payoff function (6) for local governments.

In the model, the marginal benefit of quality q for the local governments is normalized to be one. The magnitude of the unit distance cost  $\gamma$  measures the importance of distance in the decision of school consolidation by the local governments. If the unit distance cost  $\gamma$ is large, it means that the local governments care a lot about the distance cost incurred by students whose schools get closed. If, on the other hand, the unit distance cost  $\gamma$  is zero, it means that the local governments do not care about the cost incurred by students whose schools get closed, and that decisions about school closures by the local government are mostly driven by economies of scale in provision of school quality and savings in the fixed cost of running schools.

Having specified the structural model, I now turn to its estimation. First, I discuss the econometric challenges of estimating the structural model, followed by my estimation strategy to deal with the challenges. Then I detail the implementation of the estimation strategy and discuss the results.

## 5 Econometric Approach

### 5.1 Econometric Challenges

The econometric challenge associated with the production function is that the productivity term  $\omega_{it}$  in the production of school quality is unobserved. However, the productivity term affects school consolidation as well as teacher allocation. Estimates of the production function (1) ignoring the unobserved  $\omega_{it}$ , such as that in Table 8, are biased. As a result, the economies of scale in the production of school quality, measured by  $\alpha - \beta$  in equation (1), will also be biased. In the case that the economies of scale term  $\alpha - \beta$  is overestimated, most welfare gains, if any, are likely to be attributed to the economies of scale in the production of school quality when in reality it is modest. That is, ignoring unobserved productivity  $\omega_{it}$ , not only leads to inconsistent estimates of the production function, but also further leads to incorrect welfare analysis.

At the same time, there is a data constraint. There are no standardized test scores for all schools in the administrative data, which are the output measures. Only several hundred schools in the GSCF administered standardized tests as part of the survey. Without output measures for all schools in the administrative data, it is challenging to estimate the production function. In Section 5.2.1, I will kill two birds with one stone—I will develop an approach to deal with the challenge arising from the unobserved school productivity and the data constraint.

The econometric challenge of estimating the payoff function for local governments is the unusual number of choices facing local governments as they consolidated schools. With N schools, the number of choices is  $N^{N}$ ,<sup>25</sup> which quickly exceeds the limit of modern computing power as N increases.<sup>26</sup> As a result, traditional Methods of Simulated Likelihood (MSL) or

<sup>&</sup>lt;sup>25</sup>Each school has N choices: either remain open (send students to itself) or send its students to one of the other N - 1 schools.

<sup>&</sup>lt;sup>26</sup>Facility location problem is known to be NP hard in computer science.

Simulated Methods of Moments (SMM) are unlikely to work.<sup>27</sup>

A possible solution to the challenge arising from the unusual number of discrete choices is to assume that the error term  $\epsilon_t$  in the payoff function (6) is payoff relevant and is an independently, identically distributed extreme value. Then estimation can be performed on only a random subset of alternatives without inducing inconsistency (McFadden 1978). However, one problem associated with this approach is that the payoff relevant shocks  $\epsilon_t$  are unlikely to be independent from each other.<sup>28</sup> The other problem is that the logit model exhibits the property of independence from irrelevant alternatives (IIA), which is unlikely to hold in this combinatorial case.<sup>29</sup>

A second possible solution is to explore whether there is some special structure that can reduce the number of potential optimal choices. For example, Jia (2008) makes use of the supermodularity in the games between Wal-mart and Kmart. However, in the school consolidation, there are vertical differentiation resulting from heterogenous school productivity  $\omega_{it}$  as well as horizontal differentiation resulting from heterogeneous home-to-school distance  $d_{it}$  and heterogeneous school size  $N_{it}$  that affects economies of scale. As a result, no special structure can be taken advantage of because of the interactions among heterogeneous productivity across schools, heterogeneous home-to-school commuting costs, and heterogeneous

<sup>&</sup>lt;sup>27</sup>Setting the limit of modern computing power aside, another problem with MSL is that it is simply too difficult to write down the likelihood function. A common problem with both MSL and SMM is that simulation error would be large in practice when there are too many discrete choices. Simulation methods generally require a computer to draw a random shock conditional on that the observed choice is the best. With too many choices, the random shock to be drawn has to be extremely large at non-true parameters so that the observed choice is the best and will be chosen by economic agents. It is difficult for a computer to draw an extremely large random shock repeatedly, resulting a large simulation error.

<sup>&</sup>lt;sup>28</sup>Suppose now there are three schools A, B, and C. Denote the payoff relevant shock associated with the choice of closing A and merging it with B as  $\epsilon_{AB,C}$  and the payoff relevant shock associated with the choice of closing A and merging it with C as  $\epsilon_{AC,B}$ . It is unlikely that  $\epsilon_{AB,C}$  and  $\epsilon_{AC,B}$  are independent with each other since both of these two choices involve closing A.

<sup>&</sup>lt;sup>29</sup>Suppose now there are four schools A, B, C and D. The relative probability of the choice of closing A and merging it with D and the choice of closing B and merging it with D might be irrelevant with the alternative that closing C and merging it with D. It is unlikely that this relative probability is irrelevant with the alternative of closing A and C at the same time and merging the two with D. If the latter alternative is of higher probability, it is likely that the relative probability would also be larger.

economies of scale resulting from heterogeneous school size.

A third possible solution is to adopt a moment-inequalities approach and estimate a bound for the parameters. Indeed, the difficulty in the motivating example in Pakes (2010) also arises from the unusual large size of choice set. Moment inequalities neither restrict choice sets nor require a parametric form for the disturbance distributions. However, this approach will often result in partial rather than point identification of the parameters of interest (Pakes 2010). Pakes et al. (2015) provide several application examples of moment inequalities.

In the following section I will discuss how to address the challenge arising from the unobserved school productivity  $\omega_{it}$  and the data constraint and how to use moment inequalities to bound the parameters in the payoff function.

### 5.2 Estimation Strategy

#### 5.2.1 Estimation of the Production Function

The solution to the econometric challenge above relies on the following insight: unobserved heterogeneity  $\omega_{it}$  affects school consolidation and teacher allocation at the same time. If conditional on the number of students  $N_{it}$ , there is a monotonic relationship between the unobserved heterogeneity  $\omega_{it}$  and the number of teachers  $L_{it}$ , I can infer the unobserved heterogeneity  $\omega_{it}$  for all schools in the administrative data using the information on the number of teachers in each school. With the unobserved heterogeneity  $\omega_{it}$  inferred, I can correctly estimate the parameters in the payoff function for local governments and correct welfare analysis is possible. It is the same idea pioneered by Olley and Pakes (1996).

Olley and Pakes (1996) use a control function derived from the monotonic relationship between firms' amount of investment and unobserved productivity conditional on firms' labor choice to control for the unobserved productivity when estimating production functions. Levinsohn and Petrin (2003) extend the approach and use intermediate inputs to proxy for the unobserved productivity. Ackerberg, Caves, and Frazer (2006) unify these two approaches in GMM framework.

Like Olley and Pakes (1996), take logs on both sides of equation (1) and get

$$\tilde{q}_{it} = \omega_{it} + \alpha l_{it} - \beta n_{it} + \eta_{it} \tag{7}$$

With minor abuse of notation here,  $\tilde{q}_{it}$  denotes  $\log(\tilde{q}_{it})$  in equation (1),  $\eta_{it}$  denotes  $\log(\exp(\eta_{it}))$ ,  $l_{it}$  denotes  $\log(L_{it})$  and  $n_{it}$  denotes  $\log(N_{it})$ . Optimal labor allocation implies that the number of teachers depends on unobserved productivity  $\omega_{it}$  and the number of students  $n_{it}$  as

$$l_{it} = f_t(\omega_{it}; n_{it}) \tag{8}$$

Suppose that conditional on the number of students  $n_{it}$ , the relationship  $f_t$  in equation (8) is monotonic.<sup>30</sup> That is, conditional on the number of students  $n_{it}$  the number of teachers allocated  $l_{it}$  is a monotonic function of  $\omega_{it}$ . Invert the function  $f_t$  and get the productivity  $\omega_{it}$  as

$$\omega_{it} = f_t^{-1}(l_{it}; n_{it}) \tag{9}$$

Substitute  $\omega_{it}$  in (9) into the log production function (7) and get

$$\tilde{q}_{it} = f_t^{-1}(l_{it}, n_{it}) + \alpha l_{it} - \beta n_{it} + \eta_{it}$$
(10)

I implement the econometric strategy as follows. First, with the GSCF, I nonparametrically

<sup>&</sup>lt;sup>30</sup>This monotonic relationship between the school productivity and the the number quality-adjusted teachers conditional on the number of students is not directly implied by the assumed Cobb-Douglas functional form. Whether the monotonic relationship holds depends on the cost structure in equation (5), which is not specified here. For example, when the marginal cost of teachers is constant, this monotonic relationship holds.

regress  $\tilde{q}_{it}$  on  $l_{it}$  and  $n_{it}$ , and obtain estimates  $\hat{\Phi}_{it}$ 

$$\Phi_{it} = f_t^{-1}(l_{it}, n_{it}) + \alpha l_{it} - \beta n_{it}$$
(11)

 $\hat{\Phi}_{it}$  is the nonparametric prediction of log test score  $q_{it}$  given the log number of teachers  $l_{it}$ and the log number of students  $n_{it}$ . Use this estimated function  $\hat{\Phi}_{it}(l_{it}, n_{it})$  from the GSCF to predict log test scores on all schools in the administrative data. Even though there is no data on standardized test scores for schools in the administrative data, this prediction step allows me to infer their test scores consistently using information from the GSCF and thus solves the problem associated with this data constraint.

Second, given a candidate value of  $(\alpha, \beta)$ , I compute the implied  $\hat{\omega}_{it}(\alpha, \beta), \forall t$  using the formula:

$$\hat{\omega}_{it}(\alpha,\beta) = \hat{\Phi}_{it} - \alpha l_{it} + \beta n_{it} \tag{12}$$

Further assume that  $\omega_{it}$  evolves as a first-order Markov process:

$$\omega_{it} = g(\omega_{it-1}) + \zeta_{it} \tag{13}$$

With  $\hat{\omega}_{it}(\alpha,\beta)$ ,  $\forall t$  computed in equation (12) given  $(\alpha,\beta)$ , the residuals from nonparametric regression  $\hat{\omega}_{it}(\alpha,\beta)$  on  $\hat{\omega}_{it-1}(\alpha,\beta)$  are the estimated  $\hat{\zeta}_{it}$ .

$$\hat{\zeta}_{it}(\alpha,\beta) = \hat{\omega}_{it}(\alpha,\beta) - \hat{g}(\hat{\omega}_{it-1}(\alpha,\beta))$$
(14)

Since  $\zeta_{it}$  is the mean-zero innovation in the productivity evolution, it is uncorrelated with the number of teachers last period  $l_{it-1}$  and the number of students last period  $n_{it-1}$ . Therefore, the following moment conditions hold

$$\mathbb{E}[\hat{\zeta}_{it}(\alpha,\beta)\cdot n_{it-1}] = 0 \tag{15}$$

$$\mathbb{E}[\hat{\zeta}_{it}(\alpha,\beta)\cdot l_{it-1}] = 0 \tag{16}$$

These two moment conditions do not involve data on test scores and therefore should be satisfied by the administrative data. The moment conditions (15) and (16) identify the two parameters  $\alpha$  and  $\beta$  in the production function.

Notice that in the framework above, school consolidation can increase the number of students  $n_{it}$  in receiving schools and thus affect the number of teachers through the teacher allocation rule in equation (8). However, school consolidation is assumed not to affect the evolution of productivity  $\omega_{it}$  of receiving schools, and is independent of the productivity innovation term  $\zeta_{it}$  in equation (13).

In practice, instead of non-parametrically estimating  $\hat{\Phi}_{it}(l_{it}, n_{it})$  directly, I use the P. Robinson (1988) estimator to estimate the following semi-parametric relationship

$$\tilde{q}_{it} = \rho X + \Phi(l_{it}, n_{it}) + \eta_{it} \tag{17}$$

where X includes sex, age, parents' age, age square and their education level to control for the influence of personal and family characteristics on test scores. I use a fifth-order polynomial to approximate the Markov process, the function g, in equation (13).

The approach in this paper differs from that taken by Olley and Pakes (1996). Olley and Pakes (1996) use investment of the previous period to proxy for the unobserved productivity of the previous period, and the transition from last period to current period introduces the endogenous exit problem that Olley and Pakes (1996) have to control for. However, this present paper directly uses number of teachers of current period to control for unobserved productivity of current period. Thus there is no need to control for the endogenous exit.<sup>31</sup>

### 5.2.2 Estimation of the Payoff Function

The aforementioned strategy of estimating the production function only assumes a Cobb-Douglas functional form for the production function and a monotonic relationship between the number of teachers  $L_{it}$  and the unobserved school productivity  $\omega_{it}$ , and no assumption is imposed on the payoff function. To further estimate the remaining cost parameters, I assume that the marginal cost of quality-adjusted teachers is constant w. The Cobb-Douglas production function and the constant marginal cost imply that the variable cost in (5) is a share of total output, per-capital quality multiplied by the total number of students.

$$VC(N_{it}) = wL_{it} = \alpha q_{it} N_{it} \tag{18}$$

Thus

$$w = \frac{\alpha q_{it} N_{it}}{L_{it}} \tag{19}$$

Equation (19) provides the link between test scores and their monetary value. The fixed cost F in (5) will be calibrated using data from the GSCF.<sup>32</sup> With wage w from (19) and the fixed cost F taken from the GSCF, only the unit distance cost  $\gamma$  in the payoff function (6) is remaining to be estimated.

I adopt a moment-inequalities approach and estimate a bound of the unit distance cost  $\gamma$ . To gain intuition of the procedure, suppose now there are two schools A and B. I use

<sup>&</sup>lt;sup>31</sup>Olley and Pakes (1996) cannot use investment of current period to control for productivity of current period because that would introduce collinearity in its second stage estimation. This paper can do this because the number of students  $n_{it}$  is given. Moreover, because of data limitations, Olley and Pakes (1996) use log sales to proxy for physical output in equation (7) rather than physical output itself. Log sales is the sum of log output and log price. As long as the market is not competitive, the right hand side in equation (7) contains an unobserved price error. However, this paper uses test scores to control for "physical" output directly, and thus no further error needs to be corrected as that in De Loecker (2011).

<sup>&</sup>lt;sup>32</sup>The GCSF contains a school questionnaire that includes information on school cost.

 $W_A$  and  $W_B$  to denote the payoffs for local governments for running school A and school B separately:

$$W_A = q_A N_A - VC(L_A) - F = (1 - \alpha)\omega_A L_A^{\alpha} N_A^{1-\beta} - F$$
(20)

$$W_B = q_B N_B - VC(L_B) - F = (1 - \alpha)\omega_B L_B^{\alpha} N_B^{1-\beta} - F$$
(21)

Equation (11) implies that

$$W_A = (1 - \alpha) \exp(\Phi(l_A, n_A)) - F$$
(22)

$$W_B = (1 - \alpha) \exp(\Phi(l_B, n_B)) - F$$
(23)

Suppose A is not merged into B. The total payoff for local governments is the sum of  $W_A$  and  $W_B$ . Now consider the counterfactual case that A is merged into B. The counterfactual total payoff for local governments is

$$W_{AB} = q_B (N_A + N_B) - VC(L_{AB}) - \gamma d_{AB} N_A - F$$
  
=  $(1 - \alpha) \omega_B L^{\alpha}_{AB} (N_A + N_B)^{1-\beta} - \gamma d_{AB} N_A - F$  (24)

Having estimated parameters  $\alpha$  and  $\beta$  in the production function using the strategy in Section 5.2.1, I can infer  $\omega$  according to equation (12), repeated here as

$$\hat{\omega}_{it}(\alpha,\beta) = \hat{\Phi}_{it} - \alpha l_{it} + \beta n_{it} \tag{12}$$

In particular, I can infer  $\omega_B$  in equation (24). Nonetheless, I still do not observe  $L_{AB}$ , the number of teachers in the counterfactual case. But equation (8) shows that  $L_{AB}$  is a function of school productivity  $\omega_B$  and the number of students  $N_A + N_B$  in counterfactual case. More

specifically, equation (8) implies that

$$\Phi(l_{it}, n_{it}) = \Phi(f(\omega_{it}; n_{it}), n_{it}) \equiv \Psi(\omega_{it}, n_{it})$$
(25)

After estimating  $\Phi(l_{it}, n_{it})$  according to (11) and inferring  $\omega_{it}$  according to equation (12), I non-parametrically regress  $\hat{\Phi}(l_{it}, n_{it})$  on  $\omega_{it}$  and  $n_{it}$  and get  $\hat{\Psi}(\omega_{it}, n_{it})$ . Due to the monotonic relationship between school productivity and the number of teachers conditional on the number of students, test scores can be expressed as a function of the number of students and the number of teachers, and can also be expressed as a function of school productivity and the number of students. These two functions,  $\Phi$  and  $\Psi$  in equation (25) are equivalent to each other.

Therefore,  $W_{AB}$  in equation (24) can be written as

$$W_{AB} = (1 - \alpha) \exp(\Psi(\omega_B, \log(N_A + N_B))) - \gamma d_{AB}N_A - F$$
(26)

The comparison between the observed case when A is not merged into B and the counterfactual case when A is merged into B reveals that

$$\mathbb{E}[W_A + W_B - W_{AB}] \ge 0 \tag{27}$$

Substitute  $W_A$  in equation (22),  $W_B$  in equation (23), and  $W_{AB}$  in equation (26) into equation (27), leading to

$$(1 - \alpha)\mathbb{E}[\exp(\Psi(\omega_B, \log(N_A + N_B))) - [\exp(\Phi(l_A, n_A)) + \exp(\Phi(l_B, n_B))]] + F \le \gamma \mathbb{E}[d_{AB}N_A]$$
(28)

which provides a lower bound of the unit distance cost  $\gamma$ . Regarding the upper bound of the

unit distance cost, consider the opposite case that school A is merged into B in practice and the counterfactual case that school A is not merged into school B. By similar reasoning, I get

$$(1 - \alpha)\mathbb{E}[\exp(\Phi(L_{AB}, \log(N_A + N_B))) - [\exp(\Psi(\omega_A, n_A)) + \exp(\Psi(\omega_B, n_B))]] + F \ge \gamma \mathbb{E}[d_{AB}N_A]$$
(29)

which provides an upper bound of the unit distance cost  $\gamma$ . In contrast with the case when I estimate the lower bound of the unit distance cost, in this case, I observe  $L_{AB}$  but not  $L_A$  and  $L_B$ . Taking advantage of the equivalence between  $\Phi$  and  $\Psi$  established in equation (25), I express the payoff difference between the observed case and the counterfactual case as equation (29), an equation with  $L_{AB}$  but without  $L_A$  and  $L_B$ .

The problem with equation (29) is that  $\omega_A$  cannot be inferred as that in equation (12) since A is closed and neither the number of teachers nor its test scores can be observed. However, I have assumed that  $\omega_A$  evolves as a first-order Markov process, repeated here as

$$\omega_t^A = g(\omega_{t-1}^A) + \zeta_t^A \tag{13}$$

After inferring  $\omega$  of each period according to equation (12), I can estimate the markov process g in equation (13). Though  $\omega_t^A$  cannot be observed any more when school A is closed,  $\omega_{t-1}^A$  can be inferred according to equation (12). With  $\omega_{t-1}^A$  inferred, I can predict  $\omega_t^A$  using estimated  $\hat{g}$  when innovation shock  $\zeta_t^A$  is ignored.

I assume that when local governments made school consolidation decisions, they did not observe the innovation shock in school productivity to the schools that were closed. Local governments' expectation is based on last period's productivity, that is, on  $g(\omega_{t-1})$ . In this case, I can substitute  $\hat{g}(\omega_{t-1}^A)$  into equation (29), and all terms on the left-hand side now can be observed or inferred. As a result, I can obtain an upper bound of the unit distance cost  $\gamma$ .

### 5.3 Estimation Results

Table 9 shows the estimation results of the production function. The measured economies of scale, measured by  $\alpha - \beta$  in the production function, are around 8% in Chinese-language and around 5% in math. Even after taking account the observed school productivity, I still find economies of scale in education production, and of course the estimate of the fixed cost F provides a further source of increasing returns to scale.

Figure 5 compares the cumulative density function of productivity of schools which did not experience any closure with that of schools which did experience closure from year 2001 to year 2008. On average, the productivity of schools which experienced any closure was lower. Therefore, students whose schools were closed were transferred to schools with higher productivity. This accounts for one reason that students' test scores increased post school closure. Notice that the two cumulative density functions cross with each other towards the end, which implies that some schools with relatively high productivity also experienced school closure during the school consolidation. This point is further confirmed in Figure 6.

Figure 6 compares the relationship between school size and estimated school productivity before the school consolidation with that after the school consolidation. It confirms that some small schools with relatively high productivity were closed during the school consolidation, which explains the crossing in Figure 5. It also shows that small schools with lower productivity were heavily selected out during the school consolidation. Post school consolidation, small schools tended to be better than large schools in terms of productivity. Therefore, smaller schools do not necessarily exhibit higher productivity; rather, they are selected to be of higher productivity. Small schools and large schools coexist with each other in the world: small schools are better in school productivity but worse in terms of economies of scale; large schools are worse in school productivity but better in terms of economies of scale.

The selection shown in Figure 6 is striking when one realizes that small schools with lower productivity were those relatively cheap ones rather than those relatively expensive ones—lower productivity means that conditional on the number of students, those schools closed during the school consolidation employed less teachers than other schools remaining open; that is, on a per-student basis, the former were cheaper—therefore, a model with local governments only minimizing costs would not explain the observed pattern of school consolidation.

After estimating the parameters in the production function and inferring school productivity for all schools in the administrative school, I go on to estimate bounds on the unit distance cost.

For robustness, I impose the assumption that there is no fixed cost. That is, F is zero. I estimate the lower bound of the unit distance cost using the administrative data in 2009. Within a town, I select out two schools with about 50 students and another two largest schools.<sup>33</sup> The schools with about 50 students were not merged with the largest school (or the second largest school) in 2009. The counterfactual case I use is that the schools with about 50 students were merged with either of the two largest schools in 2009. The estimated lower bound of the distance cost per student per 5km is from 0\$(in Math) to 1.5\$(in Chinese) per year.

I estimate the upper bound of the distance cost using the mergers happened in 2007, 2008 and 2009. The counterfactual I use is that these mergers did not happen. The estimated upper bound of the distance cost per student per 5km is from 31\$(in Chinese) to 47\$ (in Math) per year. I compare this estimated upper bound of the distance cost with the actual

<sup>&</sup>lt;sup>33</sup>As long as the selection criteria of counterfactual case are not correlated with payoff-relevant shock or payoff-irrelevant measurement error  $\epsilon_t$  in equation (6), the estimation is valid. 50 is on the 30th percentile of school-size distribution in 2009.

distance cost in Table 12 reported by families in the 2012 China Family Panel Studies.<sup>34</sup> The estimated upper bound is at most one half of the actual monetary cost incurred by families.<sup>35</sup> Therefore, distance cost was under-weighted in local governments' decision.

Before I reach the conclusion of this paper, I discuss the assumptions of the functional form, and examine whether alternatives stories like political influence and peer effects could drive the data pattern I see.

## 6 Discussion

### 6.1 Assumption on Functional Forms

A common functional form assumption is to assume that the total cost of running each school with students  $N_{it}$  is

$$C_{it} = F + \nu N_{it}^{\gamma} \tag{30}$$

where F is fixed cost and  $\gamma$  measures economies of scale.

This subsection shows that under certain restrictions, the cost function in equation (30) and the production function in equation(2) are equivalent. The production function for school quality is assumed to be of the Cobb-Douglas form in equation(2):

$$q_{it} = \omega_{it} L^{\alpha}_{it} N^{-\beta}_{it}$$

Thus, given  $N_{it}$ , production of  $q_{it}$  requires

$$L_{it} = \left(\frac{q_{it}}{\omega_{it}}\right)^{\frac{1}{\alpha}} N_{it}^{\frac{\beta}{\alpha}} \tag{31}$$

<sup>&</sup>lt;sup>34</sup>Readers can refer to the website http://www.isss.edu.cn/cfps/EN/ for details on this data.

 $<sup>^{35}</sup>$ This does not take into account the subsidies paid by governments on transportation, food in school and boarding, etc.

If I further assume that the marginal cost of  $L_{it}$  is constant c, and the fixed cost of running each school is F, the total cost of running school i ends up being

$$C_{it} = F + c \left(\frac{q_{it}}{\omega_{it}}\right)^{\frac{1}{\alpha}} N_{it}^{\frac{\beta}{\alpha}}$$
(32)

If  $q_{it}$  is q, the same for every school, it further simplifies to

$$C_{it} = F + cq\omega_{it}^{-\frac{1}{\alpha}} N_{it}^{\frac{\beta}{\alpha}}$$
(33)

which is exactly of the same form as that in equation (30) when there is no unobserved heterogeneity in  $\omega_{it}$ , with the measure of economies of scale in equation (30) corresponding to  $\frac{\beta}{\alpha}$  in equation (33).

Therefore, the production function (2) assumed in this paper is more general than the common assumption on the cost function (30).

### 6.2 Shift in Political Power

The current empirical literature explaining provision of public goods focuses on collective action. An alternative explanation for the observed pattern could be that political power shifted from small villages to large villages during the studying period. It might be the shift in political power from small villages to large villages rather than the economies of scale from consolidating small schools into large schools that drove the school consolidation. This alternative explanation would imply not only the change in allocation of public schools over time, but also changes in other policy dimensions over time, such as subsidies and grants from upper level governments, and possible public projects supported by upper level governments.

However, Table 10, based on a panel data covering over 200 villages from year 2000 to year

2008,<sup>36</sup> does not show any significant change in these other policy dimensions: neither per capita subsidies nor per capita grants from upper level government decreased (increased) as the number of children aged 7 to 13 became smaller (larger) over time; per capita investments and subsidies in various public projects, potentially supported by upper level government, did not deteriorate as the number of children aged 7 to 13 became smaller of primary school within each village,<sup>37</sup> I use a proxy of school closure—whether the total number of children aged 7 to 13 exceeded 100 or not—to measure the effect of school closure on the per capita subsidies, grants from upper level governments. These results show no evidence that there was a change in other policy dimensions as villages to large villages drove the school consolidation during the study period.

### 6.3 Peer Effects

The school consolidation not only changed the school productivity, the number of teachers and the number of students—students whose schools were closed were transferred to schools with different productivity, different number of teachers and different number of students but also changed many other aspects of studying environment, most importantly, the peer structure. To the extent that the model above in equation (2) does not fully capture the change in other aspects of studying environment, such as peer structure, one might be concerned that it is the change in peer structure rather than the change in school productivity and economies of scale that drives the observed increase in test scores in Table 5.

Exact modeling and estimation of peer structure is beyond the scope of this paper.

<sup>&</sup>lt;sup>36</sup>Thanks Nancy Qian for sharing the data.

<sup>&</sup>lt;sup>37</sup>Probably because initially there was a primary school in every village, the questionnaire did not even bother to include such a question.

However, to the extent that peer structure affects students through the total number of students in school, it has been captured by the negative term  $-\beta n_{it}$  in the production function for school quality in equation (7). Moreover, the change in peer structure should also change students' studying behavior accordingly. Nonetheless, Table 7 shows that school closure had no effect on students' time spent on their homework, which casts doubt on the possibility that a change in peer structure drives the observed increase in test scores. Finally, I explicitly test the null hypothesis that it is not the change in peer structure (or any other aspects that one can think of) that drives the observed results. I modify the log production function for school quality in equation (7) to be

$$\tilde{q}_{it} = \omega_{it} + \alpha l_{it} - \beta n_{it} + p_{it} + \eta_{it} \tag{34}$$

where the additional term  $p_{it}$  stands for peer effect. Within the framework set up in Section 5.2, the sum of the first three terms in equation (34) is a function of  $l_{it}$  and  $n_{it}$ , as shown in equation (11). Therefore, I have

$$\tilde{q}_{it} = \Phi(l_{it}, n_{it}) + p_{it} + \eta_{it} \tag{35}$$

and

$$\Delta \tilde{q}_{it} = \Delta \Phi(l_{it}, n_{it}) + \Delta p_{it} + \Delta \eta_{it} \tag{36}$$

where  $\Delta$  denotes change over time, or the change after school closure. Therefore, once the change in  $\Phi(l_{it}, n_{it})$  is controlled for, any significant (insignificant) remaining effect of the school closure on students' test scores will imply significant (insignificant) change in the peer structure brought about by the school consolidation.

Table 11 shows that there is no further significant change in test scores once I have controlled for the change in  $\Phi(l_{it}, n_{it})$ , both for students in sending schools and for students in receiving schools. Therefore, the change in the peer structure did not drive the change in students' test scores. Rather, it is mainly the change in  $\Phi(l_{it}, n_{it})$ , i.e., the change in school productivity and economies of scale, that causes the increase in the test scores of students from sending schools. The school consolidation is mainly about the consolidation in terms of school productivity and economies of scale. Any other aspect is insignificant once I have controlled for these changes.

## 7 Conclusion

This paper shows that the school consolidation process can be rationalized by considerations of economies of scale, and trade-offs between heterogeneity in school productivity and heterogeneity in home-to-school commuting distance. The general lesson is that factors from the supply side cannot be ignored in explaining variation in the provision of public goods, especially in developing countries, where the budget for public goods is more limited and the trade-offs involved are more significant. These factors, such as economies of scale, heterogeneity in the productivity of public goods at different places and heterogeneity in people's preferences towards public goods, shape but also constrain local governments' policies in delivering public goods.

The estimation shows that there is great heterogeneity in school productivity. This heterogeneity affects how local governments allocate teachers to schools and also has implications for the estimation of economies of scale in school production. More importantly, this heterogeneity affects local governments' decisions in choosing which schools to close or keep open during the school consolidation.

I find that smaller and academically weaker schools with lower productivity were heavily selected out during the school consolidation. Ultimately, small schools were selected to be better than large schools in terms of school productivity. This selection story can help shed light on the current debate on small schools versus large schools. It is often claimed that small schools are better than large schools. However, small schools and large schools coexist in this world in practice. If the claim is true, competition might force large schools out of market over time, which is not the case. The findings of the paper show that heterogeneity is often multidimensional: small schools are better in productivity but worse in terms of economies of scale; large schools are worse in productivity but better in terms of economies of scale. The balance among the different dimensions of heterogeneity assures the coexistence of small schools and large schools.

The estimation using moment-inequalities shows that the value of the estimated distance cost was not enough to cover families' expenditure on home-to-school commute, indicating that local governments closed more schools than needed from families' perspective. Though compared with traditional estimation methods, moment-inequalities can only partially identify parameters in the model, this partial identification is informative and reveals whether the massive school consolidation program examined in this paper was marginally valuable.

## References

- Ackerberg Daniel A., Caves Kevin, and Frazer Garth (2006). "Structural identification of production functions".
- Alesina Alberto, Baqir Reza, and Easterly William (1999). "Public goods and ethnic divisions". The Quarterly Journal of Economics 114eth.4, pp. 1243–1284.
- Alesina Alberto, Baqir Reza, and Hoxby Caroline (2004). "Political Jurisdictions in Heterogeneous Communities". Journal of Political Economy 112.2, pp. 348–396.
- Alesina Alberto and Spolaore Enrico (1997). "On the Number and Size of Nations". Quarterly Journal of Economics 112.4, pp. 1027–1056.

- Altonji Joseph G. and Dunn Thomas A. (1996). "Using Siblings to Estimate the Effect of School Quality on Wages". *Review of Economics and Statistics* 78.4, pp. 665–671.
- Angrist Joshua D. and Lang Kevin (2004). "Does school integration generate peer effects? Evidence from Boston's metco program". American Economic Review 94.5, pp. 1613– 1634.
- Angrist Joshua D. and Lavy Victor (1999). "Using Maimonides' Rule To Estimate The Effect Of Class Size On Scholastic Achievement". The Quarterly Journal of Economics May, pp. 533–575.
- Baland Jean-Marie, Moene Karl Ove, and Robinson James A. (2010). "Governance and Development". Handbook of Development Economics. Handbook of Development Economics 5, pp. 4597–4656.
- Banerjee Abhijit, Iyer Lakshmi, and Somanathan Rohini (2008). "Public action for public goods". Handbook of Development Economics 4.07, pp. 3117–3154.
- Behrman Jere R., Rosenzweig Mark R., and Taubman Paul (1996). "College Choice and Wages: Estimates Using Data on Female Twins". The Review of Economics and Statistics 78.4, pp. 672–685.
- Bolton Patrick and Roland Gerard (1997). "The Breakup of Nations: A Political Economy Analysis". *The Quarterly Journal of Economics* 112.4, pp. 1057–1090.
- Card David and Krueger Alan B. (1992). "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States". Journal of Political Economy 100.1, p. 1.
- Chattopadhyay Raghabendra and Duflo Esther (2004). "Women as policy makers: Evidence from a randomized policy experiment in India". *Econometrica* 72.5, pp. 1409–1443.
- Chetty Raj, Friedman John N., and Rockoff Jonah E. (2014a). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates". American Economic Review 104.9, pp. 2593–2632.

- Chetty Raj, Friedman John N., and Rockoff Jonah E. (2014b). "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood". American Economic Review 104.9, pp. 2633–2679.
- De Loecker Jan (2011). "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity". *Econometrica* 79.5, pp. 1407–1451.
- Dearden Lorraine, Ferri Javier, and Meghir Costas (2002). "The Effect of School Quality on Educational Attainment and Wages". *Review of Economics and Statistics* 84.1, pp. 1–20.
- Deaton Angus (2010). "Instruments, Randomization, and Learning about Development". Journal of Economic Literature 48.2, pp. 424–455.
- Department of Planning Ministry of Education (1984). Achievement of Education in China
  : Statistics 1949-1983. People's Education Press, p. 492.
- Duflo Esther, Hanna Rema, and Ryan Stephen P. (2012). "Incentives work: Getting teachers to come to school". American Economic Review 102.4, pp. 1241–1278.
- Gandhi Amit, Navarro Salvador, and Rivers David (2014). "On the Identification of Production Functions: How Heterogeneous is Productivity?"
- Glewwe Paul, Huang Qiuqiong, and Park Albert (2013). "Cognitive Skills, Noncognitive Skills, and the Employment and Wages of Young Adults in Rural China".
- Hoxby Caroline M. (2000). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation". The Quarterly Journal of Economics 115.4, pp. 1239–1285.
- Jia Panle (2008). "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry". *Econometrica* 76.6, pp. 1263–1316.
- Krueger Alan B. (1999). "Experimental estimates of education production functions". The Quarterly Journal of Economics 114.2, pp. 497–532.
- Lai Fang, Sadoulet Elisabeth, and Janvry Alain de (2011). "The Contributions of School Quality and Teacher Qualifications to Student Performance: Evidence from a Natural Experiment in Beijing Middle Schools". Journal of Human Resources 46.1, pp. 123–153.

- Levinsohn James and Petrin Amil (2003). "Estimating production functions using inputs to control for unobservables". *Review of Economic Studies* 70.2, pp. 317–341.
- Liu Chengfang et al. (2010). "The effect of primary school mergers on academic performance of students in rural China". International Journal of Educational Development 30.6, pp. 570–585.
- McFadden Daniel (1978). "Modelling the choice of residential location". Spatial Interaction Theory and Planning Models. Vol. 673. 477, pp. 75–96.
- Mo Di et al. (2012). "Transfer paths and academic performance: The primary school merger program in china". *International Journal of Educational Development* 32.3, pp. 423–431.
- Olley Steven and Pakes Ariel (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica* 64.6, pp. 1263–1297.
- Olson Mancur (2002). The Logic of Collective Action: Public Goods and the Theory of Groups, p. 186.
- Pakes Ariel (2010). "Alternative Models for Moment Inequalities". Econometrica 78.6, pp. 1783– 1822.
- Pakes Ariel et al. (2015). "Moment Inequalities and Their Application". Econometrica 83.1, pp. 315–334.
- Peng Sen, ed. (2005). 2004 China Economic Systems Reform Yearbook. Beijing: China Financial and Economic Publishing House.
- Pepper Suzanne (1996). Radicalism and education reform in 20th-century China: The search for an ideal development model. Cambridge University Press.
- Rivkin Steven G., Hanushek Eric A., and Kain John F. (2005). "Teachers, Schools, and Academic Achievement". *Econometrica* 73.2, pp. 417–458.
- Robinson Peter (1988). "Root-N-consistent semiparametric regression". *Econometrica* 56.4, pp. 931–954.
- Sen Amartya (1999). Development as Freedom. Anchor Books.

- The China State Council (2001). The State Council's Decision on the Reform and Development of Basic Education.
- The Gansu Education Bureau (2006). The Plan to Ensure Funding for Compulsory Education.
- The Gansu Finance Bureau and The Gansu Education Bureau (2005). On the Amount of Teaching Budget Per Student.
- The Ministry of Finance (2010). The Response to the Proposal on the Improvement of the Mechanism to Ensure Funding for Rural Compulsory Education.
- The Ministry of Finance and The Ministry of Education (2007). On the Adjustment and Improvement of the Mechanism to Ensure Funding for Rural Compulsory Education.
- The National People's Congress (2011). The Approval of 2011 Central Budget by the Standing Committee of the National People's Congress.
- Weese Eric (2015). "Inefficiency and Self-Determination: Simulation-Based Evidence From Meiji Japan".
- Yang Zhoufu (2001). China Educational Finance Statistical Yearbook 2001. China Statistics Press.
- Zhang Shuang (2014). "Mother's Education and Infant Health: Evidence from High School Closures in China".

## A Appendix

In this appendix, I derive the biases in the estimation of the production function of per capita school quality and the implication for the measure of economies of scale in schooling production when the school productivity is ignored. Repeat Equation (7) here as

$$\tilde{q} = \omega + \alpha l - \beta n + \eta \tag{7}$$

OLS estimates ignoring the unobserved school productivity lead to

$$\hat{\alpha} = \alpha + \frac{\operatorname{Var}(n)\operatorname{Cov}(\omega, l) - \operatorname{Cov}(l, n)\operatorname{Cov}(\omega, n)}{\operatorname{Var}(l)\operatorname{Var}(n) - \operatorname{Cov}(l, n)^2}$$

$$= \alpha + \frac{1}{1 - \rho_{l,n}^2} \frac{\sigma_\omega}{\sigma_l} (\rho_{\omega,l} - \rho_{l,n}\rho_{\omega,n}) \qquad (A-1)$$

$$\hat{\beta} = \beta - \frac{\operatorname{Var}(l)\operatorname{Cov}(\omega, n) - \operatorname{Cov}(l, n)\operatorname{Cov}(\omega, l)}{\operatorname{Var}(l)\operatorname{Var}(n) - \operatorname{Cov}(l, n)^2}$$

$$= \beta - \frac{1}{1 - \rho_{l,n}^2} \frac{\sigma_\omega}{\sigma_n} (\rho_{\omega,n} - \rho_{l,n}\rho_{\omega,l}) \qquad (A-2)$$

where  $\rho_{l,n}$  is the correlation coefficient between l and n,  $\rho_{\omega,n}$  is the correlation coefficient between  $\omega$  and n, and  $\rho_{\omega,l}$  is the correlation coefficient between  $\omega$  and l. Therefore,

$$\rho_{\omega,n} = 0, \ \rho_{\omega,l} = 0 \Longrightarrow \hat{\alpha} = \alpha, \ \hat{\beta} = \beta$$

$$\rho_{\omega,n} = 0, \ \rho_{\omega,l} > 0 \Longrightarrow \hat{\alpha} > \alpha, \ \hat{\beta} > \beta$$

$$\rho_{\omega,n} > 0, \ \rho_{\omega,l} = 0 \Longrightarrow \hat{\alpha} < \alpha, \ \hat{\beta} < \beta$$

$$\frac{\rho_{\omega,l}}{\rho_{\omega,n}} < \rho_{l,n} < \frac{\rho_{\omega,n}}{\rho_{\omega,l}} \Longrightarrow \hat{\alpha} < \alpha, \ \hat{\beta} < \beta$$
(A-3)

Therefore, when the log number of students n is sufficiently tightly related to the log productivity  $\omega$  relative to the tightness how much the log number of teachers l is related to the log productivity  $\omega$ , OLS underestimate both  $\alpha$  and  $\beta$ .

Equation (A-1) and (A-2) imply that the measure of economies of scale in schooling production is

$$\hat{\alpha} - \hat{\beta} = \alpha - \beta + \frac{[\operatorname{Var}(n) - \operatorname{Cov}(l, n)]\operatorname{Cov}(\omega, l) + [\operatorname{Var}(l) - \operatorname{Cov}(l, n)]\operatorname{Cov}(\omega, n)}{\operatorname{Var}(l)\operatorname{Var}(n) - \operatorname{Cov}(l, n)^2} = \alpha - \beta + \frac{1}{1 - \rho_{l,n}^2} [\frac{\sigma_{\omega}}{\sigma_n} \rho_{\omega,l} (\frac{\sigma_n}{\sigma_l} - \rho_{l,n}) + \frac{\sigma_{\omega}}{\sigma_l} \rho_{\omega,n} (\frac{\sigma_l}{\sigma_n} - \rho_{l,n})]$$
(A-4)

Therefore, when  $\rho_{l,n} < \frac{\sigma_n}{\sigma_l}$  and  $\rho_{l,n} < \frac{\sigma_l}{\sigma_n}$ , the measure of economies of scale measured by

OLS is overestimated when  $\rho_{\omega,l}$  and  $\rho_{\omega,n}$  are both positive.

	$\log(\text{students})$	$\log(\text{teachers})$
$\log(\text{students})$	1.42	
$\log(\text{teachers})$	1.22	1.48

Table A1: Variance and Covariance

Source: calculated from the Gansu Administrative School Data 2001, 2004, 2006, 2007, 2008 and 2009. Teachers are quality-rated according to their professional ranks.

Since  $\rho_{l,n}$  is 0.84, and both  $\frac{\sigma_n}{\sigma_l}$  and  $\frac{\sigma_l}{\sigma_n}$  are close to 1, it is very likely that the estimated economies of scale by OLS will be biased up.



Figure 1: Number of Schools and Students

Source: China Statistical Yearbook 2011



Figure 2: Sample Counties in the GSCF



Figure 3: Number of Primary Schools, Students and Full-Times Teachers in Gansu 1990-2010

Source: http://data.stats.gov.cn/ accessed on October 19th, 2015. The units for the number of students and the number of teachers are both 10,000 persons.

Figure 4: Number of Grades in Gansu



Source: The Gansu Administrative School Data



Figure 5: Difference in Productivity: Any Closure vs. No Closure



(a) Difference in Productivity in Chinese (b) Difference in Productivity in Math

Figure 6: Productivity vs. Economies of Scale

Attended the Nearest	Frequency	Percentage	
Primary School?	1 0		
Yes	4,000	97.21	
No	115	2.79	
If No:			
Because School Quality	54	1.31	

### Table 1: Hukou System on School Choice

Source: the 2000 GSCF. The 2000 GSCF surveyed 2000 households. Each household reported the basic school information of all children in the family. Thus, there were over 4000 children in total. For each household, the enumerator also surveyed one child specially, and the same child was resurveyed in the following two rounds.

### Table 2: Villages and Schools of Gansu in 2001

Number of Villages in 2001	17,834
Number of Primary Schools in 2001	$17,\!477$
Courses National Durgen of Statistics	

Source: National Bureau of Statistics

	Grade 3 is Closed after 2001			
	No	No Yes		
	N=12616	N = 5083	Difference	
Number of students	211.88	72.71	139.17	
	(1.86)	(1.80)	$(2.59)^{***}$	
Pupil-teacher ratio	26.28	29.26	-2.98	
	(0.17)	(0.40)	$(0.43)^{***}$	
Average education of teachers	10.84	10.11	0.73	
	(0.01)	(0.02)	$(0.02)^{***}$	
Per-capita fixed capital(\$)	619.28	176.69	442.60	
	(189.21)	(21.74)	$(190.45)^{**}$	

Table 3: Closed vs. Not Closed

Source: the Gansu Administrative School Data. All data was reported to the Education Bureau in 2001. Teachers are wage-rated according to their professional rank. Estimated standard errors are in parenthesis. \*\*\* denotes significance at 1% level, \*\* at 5% level, and \* at 10% level.

	Grade 1	l Closed	Grade 3	8 Closed
ln(students)	-0.266	-0.209	-0.265	-0.229
	$(0.011)^{***}$	$(0.011)^{***}$	(0.010)***	$(0.009)^{***}$
$\ln(\text{pupil/teacher})$		-0.464		-0.562
		$(0.075)^{***}$		$(0.101)^{***}$
$\ln(\text{pupil/teacher})$ square		0.064		0.082
		$(0.012)^{***}$		$(0.016)^{***}$
average edu. of teachers		-0.056		-0.035
		$(0.006)^{***}$		$(0.005)^{***}$
$\ln(\text{capital/pupil})(\$)$		-0.006		0.018
		(0.013)		(0.011)
$\ln(\text{capital/pupil})$ square		0.001		-0.001
		(0.001)		(0.001)
Adjusted $\mathbb{R}^2$	0.27	0.28	0.26	0.27
N	21,087	19,446	$17,\!699$	$16,\!605$

Table 4: Determinants of Closure after 2001: Linear Probability

Source: the Gansu Administrative School Data. ln(students) refers to the number of students in the relative grades. ln(pupil-teacher ratio) refers to the pupil-teacher ratio of the whole school, with teachers being wage-rated according to their professional rank. Robust standard errors clustered on county level. \*\*\* denotes significance at 1% level, \*\* at 5% level, and \* at 10% level.

	$\log(\text{Chinese})$	$\log(Math)$
school closure*post	0.188	0.017
	$(0.075)^{**}$	(0.067)
Village FE	Yes	Yes
Year FE	Yes	Yes
Adjusted $\mathbb{R}^2$	0.32	0.18
N	2,854	2,820

Table 5: School Closure on Test Scores

Source: the GSCF. In 2000, half of the 2000 children surveyed were randomly assigned to take the Chinese test, and the other half took the math test. In 2004 and 2007, all children in the relevant age took both the Chinese-language and math tests. Scores are scaled from 0 to 100. *school closure\*post* is 1 if a student was living in village where either there was no school or there was a school but without the grade that the student was attending. This information is inferred from the GSCF: the GSCF recorded the village that any student was living and the school information of that village. Robust standard errors are clustered on village-year level. Other control variables include children's age and sex, parents' age, their age square, and their education level. \*\*\* denotes significance at 1% level, \*\* at 5% level, and \* at 10% level.

	$\log(\text{Chinese})$	$\log(\mathrm{Math})$
receiving*post	0.069	0.049
	(0.072)	(0.086)
Village FE	Yes	Yes
Year FE	Yes	Yes
Adjusted $\mathbb{R}^2$	0.32	0.18
N	2,854	2,820

Table 6: Receiving School on Test Scores of Stayers

Source: the GSCF. In 2000, half of the 2000 children surveyed were randomly assigned to take the Chinese test, and the other half took the math test. In 2004 and 2007, all children in the relevant age took both the Chinese-language and math tests. Scores are scaled from 0 to 100. *receiving\*post* is 1 if a student was attending a school that had received students transferred from other schools due to the school consolidation. Whether a school, here denoted as school A, had received students transferred from other schools due to the school consolidation. Whether a school hat edenoted as school A, had received students transferred from the telephone survey data and the administrative school data: first I use the telephone survey data to check whether there is any school that reported to be (partially) merged into school A. If so, I further use the administrative data to check whether the school merger happened by the time in question (for example, by 2007). Robust standard errors are clustered on village-year level. Other control variables include children's age and sex, parents' age, their age square, and their education level. \*\*\* denotes significance at 1% level, \*\* at 5% level, and \* at 10% level.

	$\ln(expenditure on$	ln(hours on	ln(hours on
	on bus, board	watching TV	homework
	and food in school)	per week)	per week)
school closure*post	1.273	-0.254	0.016
	$(0.271)^{***}$	$(0.073)^{***}$	(0.077)
Village FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Adjusted $\mathbb{R}^2$	0.28	0.19	0.10
N	3,837	3,837	3,837
Mean in 2001	9.67	5.96	10.28

Table 7: School Closure on Expenditure and Children's Time Allocation

Source: the GSCF. school closure\*post is 1 if a student was living in village where either there was no school or there was a school but without the grade that the student was attending. This information is inferred from the GSCF: the GSCF recorded the village that any student was living and the school information of that village. Robust standard errors are clustered on village-year level. Other control variables include children's age and sex, parents' age, their age square, and their education level. *Mean in 2001* refers to the absolute level. \*\*\* denotes significance at 1% level, \*\* at 5% level, and \* at 10% level.

Table 8: Estimation: A First Pass with OLS

	$\log(\text{Chinese})$	$\log(Math)$
ln(students)	-0.119	-0.137
	$(0.061)^{**}$	$(0.066)^{**}$
$\ln(\text{teachers})$	0.239	0.171
	$(0.065)^{***}$	$(0.062)^{***}$
Adjusted $\mathbb{R}^2$	0.11	0.05
N	2,854	2,820

Source: the GSCF. In 2000, half of the 2000 children surveyed were randomly assigned to take the Chineselanguage test, and the other half took the math test. In 2004 and 2007, all children in the relevant age took both the Chinese language and math tests. Scores are scaled from 0 to 100. Robust standard errors are clustered on village-year level. Other control variables include children's age and sex, parents' age, their age square, and their education level. Teachers are quality adjusted according to their professional rank. \*\*\* denotes significance at 1% level, \*\* at 5% level, and \* at 10% level.

	ln(Ch	inese)	$\ln(Math)$	
	ACF	OLS	ACF	OLS
α	0.50	0.24	0.31	0.17
$\beta$	0.42	0.12	0.26	0.14
scale: $\alpha - \beta$	0.08	0.12	0.05	0.03

Table 9: Estimate of Economies of Scale

	ln(per capita		$\ln(\text{per capita})$		log(per	· capita
	subsid	y from	grant	from	investment	& subsidy
	upper	govt.) upper govt.)		govt.)	by vi	llage)
$\ln(\text{per capita})$	0.055	0.052	0.088	0.088	0.112	0.112
net income)	(0.035)	(0.034)	$(0.039)^{**}$	$(0.039)^{**}$	$(0.039)^{***}$	$(0.041)^{***}$
$\ln(no. of$	-0.045		-0.009		0.016	
children 7-13)	(0.030)		(0.011)		(0.066)	
no. of children		0.042		0.015		0.000
$7-13 \le 100$		(0.049)		(0.027)		(0.104)
Village FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted $\mathbb{R}^2$	0.03	0.02	0.06	0.06	0.00	0.00
N	$1,\!264$	$1,\!264$	1,898	1,898	1,260	1,260

Table 10: Village Size on Transfer and Investment by Upper Government

Source: Rural Fixed Point Village Data 2000-2008. per capita subsidy from upper government is defined as total subsidy from upper level governments, only reported from 2003 on, divided by total number of permanent residents at the beginning of each year; per capita grant from upper government is defined as total amount of funds granted from upper level governments divided by total number of permanent residents at the beginning of each year; per capital investment and subsidy by village is defined as total productive investment and subsidy by village, only reported from 2003 on, divided by total number of laborers at the ending of each year. Productive investment and subsidy by village include 13 categories: investment in farmland infrastructure, investment in constructions and reconstructions of electronic nets, investment in country roads constructions, investment in planting trees to improve ecological environment, investment in small water conservancy and watering project, investment in township enterprises, investment in agricultural technology extension, subsidy for production of grain crop, subsidy for farming irrigation, subsidy for farming inputs, expense in hiring workers, and others. All subsidies, grants and investments are in unit of 100 RMB, and deflated accordingly. Observations with negative per capita inc income are excluded from the regressions. Robust standard errors are clustered at village level. \*\*\* denotes significance at 1% level, \*\* at 5% level, and \* at 10% level.

	Sending S	Schools	Receiving	Schools
	$\ln(\text{Chinese})$	$\ln(Math)$	$\ln(\text{Chinese})$	$\ln(Math)$
school closure*post	0.128	-0.037	0.038	0.015
	(0.095)	(0.087)	(0.077)	(0.087)
$\Phi(l_{it}, n_{it})$	Yes	Yes	Yes	Yes
Village FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
p value for $\Phi(l_{it}, n_{it})$	0.05	0.00	0.00	0.00
Adjusted $\mathbb{R}^2$	0.33	0.19	0.33	0.19
N	2,854	2,820	2,854	2,820

Table 11: Test for Peer Effects

Source: the GSCF. In 2000, half of the 2000 children surveyed were randomly assigned to take the Chinese test, and the other half took the math test. In 2004 and 2007, all children in the relevant age took both the Chinese-language and math tests. Scores are scaled from 0 to 100. For sending schools, school closure\*post is 1 if a student was living in village where either there was no school or there was a school but without the grade that the student was attending. This information is inferred from the GSCF: the GSCF recorded the village that any student was living and the school information of that village. For receiving schools, school closure\*post is 1 if a student was attending a school that had received students transferred from other schools due to the school consolidation. Whether a school, here denoted as school A, had received students transferred from other schools due to the school consolidation by a certain time (for example, by 2007) is inferred from the telephone survey data and the administrative school data: first I use the telephone survey data to check whether there is any school that reported to be (partially) merged into school A. If so, I further use the administrative data to check whether the school merger happened by the time in question (for example, by 2007). Robust standard errors are clustered on village-year level. Other control variables include children's age and sex, parents' age, their age square, and their education level. I use a fourth-order polynomial to approximate  $\Phi(l_{it}, n_{it})$ . \*\*\* denotes significance at 1% level, \*\* at 5% level, and \* at 10% level.

	Travel Distance	
	2-5km	$> 5 \mathrm{km}$
	N = 502	N=317
Transportation	174.14	408.91
	(424.30)	(803.13)
Food in School	415.84	886.32
	(764.04)	(1024.27)
Boarding	40.34	120.46
	(191.05)	(546.62)
Total	619.13	1370.75
	(988.11)	(1571.08)

Table 12: Actual Distance Cost

*Source*: the 2012 China Family Panel Studies. Unit: RMB. Statistics are for children studying in rural primary schools.