12-29-2017

School district consolidation policies: endogenous cost inefficiency and saving reversals

Mustafa U. Karakaplan
Governors State University

Levent Kutlu
The University of Texas Rio Grande Valley, levent.kutlu@utrgv.edu

Recommended Citation

Karakaplan, Mustafa U., and Levent Kutlu. "School District Consolidation Policies: Endogenous Cost Inefficiency and Saving Reversals." Empirical Economics 56, no. 5 (December 29, 2017): 1729–68. https://doi.org/10.1007/s00181-017-1398-z.

School District Consolidation Policies:

Endogenous Cost Inefficiency and Saving Reversals

Mustafa U. Karakaplan[a] and Levent Kutlu[b]

[a] Corresponding Author. Georgetown University, Department of Economics, 37th and O Streets, NW, Washington DC, 20057, USA

email: mukarakaplan@yahoo.com, phone: +1-979-422-0606

[b] School of Economics, Georgia Institute of Technology, Atlanta, GA, USA

email: levent.kutlu@econ.gatech.edu, phone: +1-404-385-1363

About the authors: Mustafa U. Karakaplan has a PhD in Economics from Texas A&M University. Levent Kutlu has a PhD in Economics from Rice University.

Version: June 20, 2017

School District Consolidation Policies: Endogenous Cost Inefficiency and Saving Reversals

Abstract

Some education policy studies suggest that consolidation of public school districts saves resources. However, endogeneity in cost models would result in incorrect estimates of the effects of consolidation. We use a new stochastic frontier methodology to examine district expenditures while handling endogeneity. Using the data from California, we find that the effects of student achievement and education market concentration on expenditure per pupil are substantially larger when endogeneity is handled. Our findings are robust to concerns such as instrumental variable adequacy and spatial interactions. Our consolidation simulations indicate that failure to address endogeneity can result in unrealistic expectations of savings.

Keywords: Educational Finance; Expenditures; Costs; Consolidation; Economies of Scale; Efficiency; Endogeneity

JEL Classification: I22, I28, I21, H72, H75, L38

1. Introduction

The Supplemental Report of the 2010-11 Budget Package compiled by California's Legislative Analyst's Office initiated an analysis about the options for consolidating small public school districts in California. Although the results did not indicate substantial support for district consolidation, the analysis was presented to the fiscal committees of the legislature with recommendations such as increasing the minimum threshold for district size to achieve greater cost efficiencies. While the analysis assessed the cost reductions due to size, it mostly failed to notice the effects of increased market concentration as a secondary outcome of consolidation and the potential endogeneity of student achievement and education market competition.

It is not uncommon for such education finance analyses to propose school district consolidation to exploit size economies and save resources. What is often overlooked, however, is how consolidation would change the education market structure, and more importantly, how endogeneity in the cost model would result in inaccurate estimates of the effects of consolidation. To have a better understanding of the effects of consolidation on school district cost, we use a recently-developed stochastic cost frontier methodology to estimate the impacts of student achievement and market concentration on expenditure while handling endogeneity. We analyze the same cross-sectional data from California in 2010-11 school year. Our main results show that the effects of student achievement and market concentration are substantially larger when their endogeneity is controlled for. Also, our counterfactual analysis of the district consolidation shows that when endogeneity is handled, a saving reversal happens to an extent that the state actually loses resources in overall. These results indicate that the actual effects of consolidation on school district expenditure may be much different than what simple cost analyses would suggest, and hence, policy implications and recommendations based on simple analyses may not be reliable.

2. An Overview

2.1. California's Public Education Funding System

California's public education system supports more than 6 million K-12 students. The system's resources add up to 60 billion dollars. While this amount is sizable, California is behind other states in terms of per pupil funding. Federal reports indicate that California's average per pupil expenditure has been well below national averages.[1]

In terms of resources, California has one of the most centralized systems in the United States with local school districts having little discretion over revenues. According to Proposition 13, property taxes are essentially state taxes in California with no local government control. Some other methods to raise revenues, such as parcel tax and voluntary contributions, are available to the local governments, but these methods are limited in terms of generating funds. As a result, the state provides most of the total K-12 funding in California.

Most of the resources provided to districts in California are unrestricted general-purpose funding based on enrollments. Once the resources are received by the school districts, they have considerable control over how to use these resources.[2] The lack of a strict auditing mechanism for how funds are spent by districts allows across-district variation in per pupil expenditures.[3] Based on outcomes, some districts can be more cost inefficient compared to others. Hence, this setting calls for an examination of the determinants of district cost efficiencies in California.

---

[1] For details, see the expenditure reports of the Institute of Education Sciences (IES).

[2] Even for the use of restricted funds, there is little or no formal accountability, and as explained by Timar (2006), some districts even acknowledge that they spend the restricted categorical funds for general purpose use.

[3] Reduced regulation may result in efficiency variation depending on the type of intervention and market structure. See Grosskopf et al. (2009), Gronberg et al. (2012), and Gronberg et al. (2015) for examples.

2.2. Effects of Size and Competition on Education Cost

Several researchers examine the issue of economies of size and cost efficiency in education. Andrews et al. (2002) survey this literature and focus on economies of size. They document that considerable savings are expected for districts with enrollment levels between 2,000 and 4,000 students relative to a very small district with 500 or fewer pupils. However, substantial diseconomies may be experienced by districts with enrollment above 15,000 students. Following the average enrollment in schools, production function studies at the school level indicate that decreasing returns may be seen in high schools with enrollment above 1,000 students and in elementary schools with more than 600 students.

Many researchers specifically analyze the size effects of district consolidation. For example, Duncombe et al. (1995) find that consolidation of districts with less than 500 pupils would potentially result in substantial savings. Simulating savings from consolidating Arkansas school districts, Dodson and Garrett (2004) conclude that district consolidation brings about 34% cost savings per student in spite of other potential implicit costs to students and local communities. Duncombe and Yinger (2007) report that district consolidation makes fiscal sense predominantly for very small districts. Zimmer et al. (2009) use Indiana school district data and determine 1,942 students to be the optimal enrollment for cost reduction through district consolidation.

The effects of education market concentration on outcomes are surveyed by Belfield and Levin (2002). Hoxby (2003) and others present that introducing more competition to the market improves public school performance. Gronberg et al. (2015) evaluate the degree of the tradeoff between scale and cost efficiency in a model with exogeneity assumption and provide evidence that increased concentration in an education market would increase cost inefficiency, which would reduce potential benefits of consolidation.

3. Empirical and Econometric Models

3.1. Public School District Cost Efficiency

In this paper, we use a stochastic frontier model to analyze districts cost efficiencies. Stochastic frontier analysis is introduced by Meeusen and van den Broeck (1977) and Aigner et al. (1977), and a way to estimate efficiency is provided by Jondrow et al. (1982). In the education setting, efforts to cost-out adequate education is deemed to be futile by Hanushek (2005, 2006). He considers methods to identify cost and inefficiencies in education as alchemy, i.e., not scientific.[4] Costrell et al. (2008) have strong reservations about using cost functions in education settings since minimum cost is not equivalent to the estimated expenditure. They recommend avoiding education cost functions. On the other hand, Gronberg et al. (2011a) provide support for the appropriateness of stochastic frontier models for the analysis of public school district cost efficiencies. As they express, the reason for the existence of stochastic cost frontier methodology is to identify the minimum (efficient) cost of producers and their additional expenditures due to their cost inefficiencies. Researchers have been successfully applying these models in various fields. Kumbhakar and Lovell (2003) provide an excellent coverage of this econometric methodology. Fried et al. (2008) present its scientific applicability in different industries and settings including education.

In a standard stochastic cost frontier model, school district expenditure can be specified as a function of outcomes, input prices, quasi-fixed inputs, the cost inefficiency component, and an error term. In the education literature, it is common to estimate per pupil expenditure. So, instead

[4] This statement rules out the entire literature of stochastic frontier analysis, which is acknowledged by strands of studies. A detailed defense of this literature is beyond the scope of our study.

of estimating total expenditures, we estimate the natural logarithm of the expenditures per pupil. Right-hand-side variables in our model are outcome variables, which include enrollment and its square[5], quality measures, input prices, and other factors. All right-hand-side variables are in natural logarithm or ratios.

There is evidence in the literature that the level of competition in the school district's education market may be a crucial feature of the expenditure equation. Competition may be a significant force that corrects district cost inefficiencies through various channels. Therefore, we model the distribution of the one-sided error term as a function of (so-called) environmental variables including a measure of competition in the education market that potentially explain the school district cost inefficiency.

3.2. Traditional Stochastic Frontier Models

A traditional stochastic cost frontier model can be written as:[6]

$$y_i = x_{1i}'\beta + v_i + u_i \tag{1}$$

where y_i is the natural logarithm of the expenditure per pupil of district i; v_i is the usual two-sided error term; x_{1i} is a vector of exogenous variables in the sense that x_{1i} is uncorrelated with v_i; and $u_i \geq 0$ is a one-sided error term capturing inefficiency and is uncorrelated with v_i. A variety of distributions is proposed for the one-sided error term. For example, Aigner et al. (1977) used the half normal; Meeusen and van den Broeck (1977) used the exponential; Stevenson (1980)

---

[5] We modify the Cobb-Douglas form by adding the square of enrollment to allow testing the economies of size and the shape of the cost function.

[6] Kumbhakar and Lovell (2003) provide an extensive survey of traditional stochastic frontier models.

used the truncated normal; Greene (1980a, 1980b, 2003) used the gamma distributions. One way to analyze the effect of exogenous variables on efficiency is modeling u_i as follows:

$$u_i = h(x_{2i}'\varphi)u_i^* \tag{2}$$
$$u_i^* \sim \mathbf{N}^+(\mu, \sigma_u^2)$$

where $h > 0$ is a function and x_{2i} is a vector of environmental variables that affect inefficiency.[7] Common choices for μ and h are:

$$\mu = 0 \tag{3}$$
$$h(x_{2i}'\varphi)^2 = \exp(x_{2i}'\varphi).$$

When x_{1i} is endogenous or u_i is correlated with v_i, the parameter and efficiency estimates of this model are inconsistent. Hence, if there is an endogeneity problem in our education cost model, we would need to handle them for consistent results.

3.3. Endogeneity Issues

3.3.1. Endogeneity in Education Models

Endogeneity in education models is discussed by some researchers. To give examples, Hoxby (2000) presents reasons why education market concentration would cause endogeneity in education production models. Izadi et al. (2002) mention that their education cost function might be suffering from endogeneity. Duncombe and Yinger (2011) point out that the education output quality in their cost equation is endogenous, and Gronberg et al. (2011a) discuss the endogeneity of output quality as well. The literature indicates that depending on a model that is used to evaluate an education policy, endogeneity can cause problems that need to be addressed.

[7] See Wang and Schmidt (2002).

Millimet and Collier (2008) analyze the endogeneity due to simultaneity of technical inefficiencies of the neighboring school districts in their spatial production spillovers model. They investigate the endogeneity of the education market structure with their economic modelling. They use a two-stage approach to examine the spillover effects of neighboring district efficiencies. Their first-stage is a distribution-free stochastic frontier model in the style of Schmidt and Sickles (1984).[8] In their second stage, they model the efficiency of a public school district with a spatial reaction function where the efficiency is assumed to be a linear function of weighted average of neighboring school districts, exogenous district characteristics, and an error term. The coefficient of the weighted average of efficiencies capture the spillover effect, which turns out to be positive. This study concludes that school districts become more efficient as neighboring school districts become more efficient. Note that their first-stage requires a panel data and it is not applicable to a cross-sectional data set. If one rather uses a standard maximum likelihood based stochastic frontier model in the first-stage, the parameter estimates would be inconsistent. Therefore, the type of application by Millimet and Collier (2008) is not suitable to address the endogeneity in our study properly.

Endogeneity problems in our stochastic cost frontier model can arise due to a couple of major reasons: First, the determinants of the cost frontier and the two-sided error term can be correlated. For instance, a random event such as an unexpected temporary school closing due to adverse weather conditions or natural disasters can reduce instructional time and that would influence the performance of the school district in statewide performance tests. Marcotte and Hemelt (2008) find that the effects of unscheduled closings on school performance on state

[8] See Duygun et al. (2016) and Kutlu (2017) for recent variations and extensions of this model.

assessments is substantially negative. Fitzpatrick et al. (2011) show that gains over the school-year-period are mostly an output of days spent in classrooms and that additional days at school matter for better performance. Since unexpected school closures are common in the United States and millions of students get affected every year (Wong et al. (2014)), such a correlation between education cost frontier and the two-sided error term would raise endogeneity issues.

Secondly, the inefficiency term and two-sided error term can be correlated, or in particular, the determinants of the inefficiency can cause this correlation. To give an example, consolidation of school districts, education market concentration, and the district resources to be spent are all determined simultaneously. Gronberg et al. (2015) provide evidence that districts in more concentrated education markets are more wasteful since they are not under competitive pressure to be more efficient with their spending. In the meanwhile, based on district expenditure structures, the government decides whether or not to consolidate districts, and as a consequence, change market concentration. This decision simultaneously determines the distribution of resources to the districts. The existence of such simultaneity would result in endogeneity.

3.3.2. Handling Endogeneity in Stochastic Frontier Analysis

Endogeneity in a stochastic frontier model would lead to inconsistent parameter estimates, and hence, it would need to be addressed properly. In the empirical literature, there is a growing concern about the endogeneity issues in the stochastic frontier models. Compared to the standard regression models, dealing with the endogeneity issue is more complicated in the stochastic frontier analysis.

Guan et al. (2009) follow a two-step estimation methodology to handle the endogenous frontier regressors. In the first step of their methodology, they get the consistent estimates of the frontier parameters using GMM; and in the second step, they use the residuals from the first-stage as the dependent variable and estimate a standard (maximum likelihood based) stochastic frontier model. Their efficiency estimates would not be consistent when the two-sided and one-sided error terms are correlated.

In the maximum likelihood context, Kutlu (2010) proposes a model that solves the endogeneity problem due to the correlation between the regressors included in the frontier and two-sided error term. This model does not have environmental variables and does not discuss endogeneity problem due to correlation between one-sided and two-sided error terms. However, the importance of this model is that it introduces a novel way to handle endogeneity in a single stage for stochastic frontier models through decomposing the two-sided error term by a Cholesky decomposition, which became a standard trick to handle endogeneity in almost all stochastic frontier models. Hence, it may be considered the cornerstone of the current endogeneity papers in the literature.

Tran and Tsionas (2013) propose a GMM variation of Kutlu (2010). Mutter et al. (2013) explain of why omitting the variable causing the endogeneity is not a viable solution. Shee and Stefanou (2014) extends the methodological approach in Levinsohn and Petrin (2003) to overcome the problem of endogenous input choice due to production shocks that are predictable by the productive unit but unknown to the econometrician. Gronberg et al. (2015) try to solve the problem through pseudo-IV methodologies. However, most of their empirical analyses rely on exogeneity assumptions.

Amsler et al. (2016) use a copula approach to solve endogeneity problem due to not only the correlation between regressors included in frontier and two-sided error term but also due to correlation between one-sided error term and two-sided error term. Hence, they allow a more general correlation structures compared to Kutlu (2010) and its variations such as Tran and Tsionas (2013). Their method is computationally intensive and considerably harder to estimate compared to standard maximum likelihood models. Tran and Tsionas (2015) present another copula based approach that allows estimation of efficiency when there are no external instruments. However, none of these models allow for endogenous environmental variables.

Karakaplan and Kutlu (2017) extend the work of Kutlu (2010) to allow environmental variables in the cross-sectional data context. They carry out Monte Carlo simulations to analyze the small sample performance of our estimator in a variety of endogeneity scenarios; and find that when there is endogeneity in the model, their estimator outperforms the model, which assumes exogeneity.[9]

In this paper, we use the econometric methodology presented by Kutlu (2010) and Karakaplan and Kutlu (2017) as that enables us to control for the endogeneity of district achievement and education market concentration. This approach provides consistent estimates of the effects of student achievement on cost and market concentration on cost inefficiency in our model.

3.4. Econometric Model

Consider the following stochastic frontier methodology of Karakaplan and Kutlu (2017)

[9] Historically, Karakaplan and Kutlu (2017) is the first paper (2013) that introduced endogenous environmental variables and endogeneity test in the maximum likelihood estimation setting to our knowledge. Amsler et al. (2017) is a recent paper that allows endogenous environmental variables too.

with endogenous explanatory variables:[10]

$$y_i = x_{1i}'\beta + v_i + u_i \tag{4}$$

$$x_i = Z_i\delta + \varepsilon_i$$

$$\begin{bmatrix} \tilde{\varepsilon}_i \\ v_i \end{bmatrix} \equiv \begin{bmatrix} \Omega^{-1/2}\varepsilon_i \\ v_i \end{bmatrix} \sim \mathbf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_p & \sigma_v\rho \\ \sigma_v\rho' & \sigma_v^2 \end{bmatrix}\right)$$

where y_i is the natural logarithm of the expenditure per pupil of district i ; x_{1i} is a vector of exogenous and endogenous variables; x_i is a $p \times 1$ vector of all endogenous variables (excluding y_i), $Z_i = I_p \otimes z_i'$ where z_i is a $q \times 1$ vector of all exogenous variables, v_i and ε_i are two-sided error terms, and $u_i \geq 0$ is a one-sided error term capturing the inefficiency. In our framework, a variable is endogenous if it is not independent from v_i. Ω is the variance-covariance matrix of ε_i, σ_v^2 is the variance of v_i, and ρ is the vector representing the correlation between $\tilde{\varepsilon}_i$ and v_i.

The identification requires number of excluded instruments to be greater or equal to number of endogenous variables. Moreover, the ratio of variances for u_i and v_i, should not be too close to zero, which is necessary for empirical identification. More details about identification can be found in Amsler et al. (2016) who present the moment conditions for GMM counterparts of our model.[11]

---

[10] We programmed the analysis in this section using Stata 13 and MATLAB 2014a. The programs and the Stata ado files are available upon request. The Stata command is also available on the Statistical Software Components (SSC) at Boston College, and on the Stata repository (package st0466). For more information about this Stata module see Karakaplan (2017) or type the following command in Stata: findit sfkk.

[11] See their Section 4.2 for the moment conditions and Section 4.3 for specific comments on Kutlu (2010) and Karakaplan and Kutlu (2017).

The log-likelihood of the model is given by:

$$\ln L(\theta) = \ln L_{y|x}(\theta) + \ln L_x(\theta) \tag{5}$$

where

$$\ln L_{y|x}(\theta) = \sum_{i=1}^{n} \left(\frac{\ln(2/\pi) - \ln\sigma_i^2 - (e_i^2/\sigma_i^2)}{2} + \ln\Phi\left(\frac{\lambda_i e_i}{\sigma_i}\right) \right)$$

$$\ln L_x(\theta) = \sum_{i=1}^{n} \left(\frac{-p \cdot \ln 2\pi - \ln(|\Omega|) - \varepsilon_i' \, \Omega^{-1} \varepsilon_i}{2} \right)$$

$$e_i = y_i - x_{1i}'\beta - \eta'(x_i - Z_i\delta)$$

$$\varepsilon_i = x_i - Z_i\delta$$

$$\sigma_i^2 = \sigma_w^2 + \sigma_{ui}^2$$

$$\lambda_i = \frac{\sigma_{ui}}{\sigma_w}$$

where θ is the vector of coefficients; $y = (y_1, y_2, \ldots, y_n)'$ is the vector of dependent variables; $x = (x_1', x_2', \ldots, x_n')'$ is the matrix of endogenous variables in the model; ϕ and Φ denote the standard normal probability density function and the cumulative distribution function; $u_i = \sigma_u(x_{2i}; \varphi_u)u_i^*$; x_{2i} is a vector of exogenous and endogenous variables; $u_i^* \sim \mathbf{N}^+(0,1)$ is a producer specific random component; $\sigma_{ui}^2 = \exp(x_{2i}'\varphi_u)$; $w_i = \sigma_v\sqrt{1 - \rho'\rho}\widetilde{w}_i = \sigma_w\widetilde{w}_i$; $\sigma_w^2 = \exp(\varphi_w)$; $\widetilde{w}_i \sim \mathbf{N}(0,1)$; Ω is the variance-covariance matrix of ε_i; σ_v^2 is the variance of v_i; and ρ is the vector representing the correlation between $\tilde{\varepsilon}_i$ and v_i.[12]

The following formula is used to predict the efficiency, $EFF_i = \exp(-u_i)$:[13]

---

[12] The details about the assumptions and how the estimator is derived are available in Kutlu (2010) and Karakaplan and Kutlu (2017).

[13] See Kutlu (2010) and Karakaplan and Kutlu (2017) for further details.

$$E[\exp(u_i)|e_i]^{-1} = \left(\frac{\Phi(\sigma_i^* + \mu_i^*/\sigma_i^*)}{\Phi(\mu_i^*/\sigma_i^*)} \exp\left(\mu_i^* + \frac{1}{2}\sigma_i^{*2}\right)\right)^{-1} \tag{6}$$

where

$$\mu_i^* = \frac{e_i \sigma_{ui}^2}{\sigma_i^2}$$

$$\sigma_i^{*2} = \frac{\sigma_{wi}^2 \sigma_{ui}^2}{\sigma_i^2}.$$

Note that there is not any practical difference between the two types of endogeneity (the correlation between the determinants of the cost frontier and the two-sided error term; and the correlation between the determinants of the inefficiency and the two-sided error term) If not handled, the endogenous covariates may have a larger bias effect on the frontier parameters and the endogenous variables in the one-sided error term may have a larger bias effect on the efficiency parameters.

The reason why we choose to use a cross-sectional model is to demonstrate the importance of solving endogeneity problem in a comparable real-world cross-sectional policy-making scenario outlined in the introduction. However, all cross-sectional stochastic frontier models have the disadvantage that they may not differentiate heterogeneity and inefficiency well. Hence, in these models, the variables that are included in the frontier model are assumed to control for all heterogeneity.[14]

---

[14] The violation of this assumption may bias efficiency estimates. A generalization of true fixed effects model of Greene (2005a, 2005b) to the endogeneity case as done by Kutlu et al. (2017) may overcome such difficulties in the panel data context.

4. Data and Variables

We get the data from the National Center for Education Statistics (NCES) and California Department of Education (CDE). Our analysis is at the school district level and contains 913 traditional public school districts that provide K-12 education in the 2010-11 school year. We exclude charter school districts from this study because they may not have the same education production technology that traditional public school districts have. With a similar reasoning, we exclude some other structurally different educational entities that have unique education production technologies such as County Offices of Education (COE), Regional Occupational Centers/Programs (ROC, ROP, and ROCP), Division of Juvenile Justice, state schools for disabilities, and community administrations.[15] Table 1 presents the summary statistics of the variables that we use.

---

[15] For statistical and financial purposes, California Department of Education began classifying County Offices of Education, California Education Authority School District, state schools for disabilities, and state charter schools as school districts in 2004. Structurally, however, these institutions are considerably different than traditional elementary, high, and unified public school districts. For example, County Offices of Education provide complementary and supplementary services, staff development, technical assistance, legal and financial advice, curriculum and instructional support to the public school districts. To give another example, Regional Occupational Centers and Programs provide career education, development, and workforce preparation by operating through Joint Powers Agreements (JPA) with traditional school districts, or through County Boards of Education. Without taking such structural differences into account, many print and online sources count these institutions as public school districts due to the change in the classification, and report that the number of public school districts in California is more than 1,000. Our data set, on the other hand, shows that the number of traditional public school districts in 2010-11 school year is 935. Our sample size drops from 935 to 913 due to the unavailability of some of the observations mostly in the student achievement data.

[Table 1: Somewhere here.]

The dependent variable in our study is the natural logarithm of actual current operating expenditures per pupil. We do not include the food expenditures in the dependent variable because the value of food and other in-kind transfers for lunch programs cannot be measured well, and that would cause additional noise. To avoid such noise, we also leave the transportation expenditures out because they are not expected to be identified by the factors that identify the student performance. Furthermore, following Gronberg et al. (2011b), we do not include facility acquisition and construction, intergovernmental payments, debt service, and community service.

Education production technology is a multiple output process. The explanatory variables in our study include variables that capture quantity or quality aspects of outcomes in our cost model. For the quantity aspect, we use the total number of students in a public school district in fall enrollment. The mean and median of this quantity measure are 6,544 and 1,932, respectively, with a minimum of 10 and a maximum of 670,746.

There are many quality aspects of outcomes that need to be taken into account in our cost model. For simplicity, we use the Academic Performance Index (API) of the districts, which is an aggregated measure of the test scores created by CDE to assess overall district performance. API is based on annual statewide assessment results in math, reading, writing, science, and history at grades 2 through 12. These statewide assessments had been the Standardized Testing and Reporting (STAR) and the California High School Exit Examination (CAHSEE) between 1999 and 2014. API can range from a low of 200 to a high of 1000, with higher scores indicating better aggregate district performance. In our sample, the mean of this measure is 787.27 with a low of 545 and a high of 967.

In addition to the quality and quantity indicators of the outcome, our explanatory variables also include input prices. In order to control for the primary educational input prices, that is the teacher salaries, we use a comparable wage index (CWI) originally created by Taylor and Fowler Jr (2006).[16] As indicated in Gronberg et al. (2015) the salaries of other education personnel such as principals are highly correlated with teacher salaries so we do not include them separately, and treat the comparable wage index as an aggregate personnel wage index.

Moreover, in order to take into account the cost of classroom material, and instructional equipment and services, we use a measure of geographic isolation, or to be exact, the linear remoteness of a district from the nearest major metropolitan area. We employ this proxy variable because the local price data of such inputs are not directly available and the variation in the local prices of these inputs is largely due to the transportation costs.[17]

Along with the outcome and input price indicators, our model also contains several environmental factors, which are not purchased but likely to influence the school district cost. In order to assess the changes in costs due to the diverse needs of students, we use the percentages of students in each public school district who are classified as limited English proficient, special

---

[16] The CWI data is available for all states at http://bush.tamu.edu/research/faculty/Taylor_CWI/

[17] Even in the age of the Internet, there is Internet costs and shipping costs, which are the counterpart of transportation cost. Since the warehouses of online shops and the hubs of package delivery companies are in central locations, shipping costs for transporting the items to remote locations would be more than transporting them to central locations. In addition, a policy brief of the Rural School and Community Trust by Hobbs (2004) explains how newer education technologies are much more expensive in rural areas. Also, the report indicates that maintenance and technical assistance for these technologies are costlier in remote locations.

education, high school, or economically disadvantaged.[18]

We introduce a measure of market concentration based on the Herfindahl-Hirschman Index (HHI) to analyze the effects of competition in the education market on school cost. While HHI is commonly used in many studies in the economics of education literature, researchers have different definitions of the education market: Hoxby (2000), Grosskopf et al. (2001), Gronberg et al. (2015), and others define education markets as the Core Based Statistical Areas, whereas Borland and Howsen (1992), Zanzig (1997), Millimet and Collier (2008) and others define education markets as the counties. The guidelines that are used by the U.S. governmental offices to delineate the statistical and jurisdictional areas involve many components including politics, which may not be necessarily relevant to the education markets. Therefore, these predetermined borders may not be completely suitable to measure education market concentration. In this paper, instead of relying on a predetermined delineation of education markets, we define the school district-specific markets based on the spatial distribution of the competitors. We assume that the relevant geographic market for each school district includes all districts with at least one school within an average driving distance around each of the district's own schools. Figure 1 illustrates our approach to construct dynamic education markets.

[Figure 1: Somewhere here.]

This approach captures the commuting distance argument better by allowing the education markets for each school district to vary with respect to the spatial locations of the schools. We

---

[18] We do not control for the percentage middle school students since variation in this variable is low.

must note that ideally, the distance from one school to another should be measured as the shortest driving distance between the two schools. We realize that our 2-dimensional circular measure may not be fully accurate to capture commuting distances because of certain 3-dimentional topographical features such as hills and lakes. However, such driving distance data are not available to us, and hence, we use circles with a fixed radius around each school. Also, we assume that 10 miles would be a reasonable average of daily commuting distance each way, which may seem arbitrary to determine the education market commuting distances. Nevertheless, according to the American Community Survey of the U.S. Census Bureau, mean travel time to work of the regular education market commuters is 23 minutes, which is less than that of other workers. Also, average travel distance to work for workers of all types is reported as 15 miles by the U.S. Census Bureau. So, considering that regular education market commuters have a preference for shorter commuting, our 10-mile radius selection is appropriate. Yet, we analyze the sensitivity of our findings to using 15, 20, and 30-mile radii and other definitions of education markets in the alternative specifications section.

In this paper, we control for the endogeneity of the student achievement measure and the education market concentration in our stochastic cost frontier model. For the endogenous student achievement variable, a common practice in the literature is to employ determinants of local demand for education as instrumental variables. For instance, Imazeki and Reschovsky (2004) present a set of possible indicators of local education demand. Such a variable set includes the unemployment rate of the population aged 16 years and over, aggregated at the school district level.[19] This variable can be effectively used as an instrumental variable for output quality. We

[19] The source of this data is demographics from Census 2010.

employ the unemployment rate to control for the endogeneity of the ln(district API) variable in our cost frontier model.

Hoxby (2000) argues and illustrates that topographical data can be used to create valid instrumental variables for education market concentration. Rothstein (2007) uses alternative measures of topography and finds that while the effect of choice on student performance is weaker than that presented by Hoxby (2000), topographical measures are essentially applicable as instruments for education market concentration. With a similar approach, we use the number of springs in a county as an exogenous source of local education supply.[20] We employ this measure as an instrumental variable for the education market concentration variable in our model. [21]

As a robustness check, in addition to the variables described above, we also try using different sets of instrumental variables explained in the alternative specifications section.

5. Estimation Results

5.1. Baseline Findings

Table 2 presents our baseline estimation results. Model EX specification assumes that all variables explaining per pupil expenditures are exogenous.[22] With that assumption, we find that

---

[20] The source of data is the U.S. Geological Survey (USGS) Geographic Names Information System (GNIS).

[21] The correlation between the proxy for material costs and the number of springs is -0.0879. This outcome would indicate that our material cost/geographic isolation proxy variable is not necessarily correlated with the number of springs topography measure. While the fact that material costs measure is uncorrelated with the springs variable does not necessarily indicate that the measurement error is not correlated with this IV, such a correlation is highly unlikely.

[22] Note that for Model EX, $\sigma_w^2 = \sigma_v^2$.

the estimated attributes of the expenditure function are mostly consistent with the theoretical expectations. There is evidence of a U-shaped cost function with respect to enrollment with minimum cost at 5,868 students in California. The indicator of district achievement has a significant and positive coefficient. That is, holding other variables constant, a higher level of district achievement would require a higher level of expenditure per pupil.[23] Except for the low-income students, student bodies that necessitate more resource intensive educational technologies also increase the costs of the school districts. For instance, one percentage point increase in the share of special education students would result in a 2.28 percent increase in cost. Additionally, districts that pay higher personnel salaries or are in remote geographical locations have higher per pupil costs. Our findings also indicate that districts in more competitive education markets (smaller HHI) are more cost efficient. Finally, we find that the average district cost efficiency is 0.9296 with minimum efficiency at 0.4474 and maximum at 0.9763.

[Table 2: Somewhere here.]

Model EN specification in Table 2 presents the estimation results using our estimator with the two instrumental variables described above for potentially endogenous variables in our model. The results are striking. First, looking at the endogeneity test results, the F-test of joint significance of the components of η indicates that the correction for the endogeneity of ln(district API) and HHI is needed. The details of the η endogeneity test are available in Karakaplan and Kutlu

[23] Expenditure variable in our paper aggregates what needs to be spent for variables that other researchers find to be effective to get more quality.

(2017). The prediction equations[24] are available in the appendix. Looking at the prediction equations, all excluded instruments are statistically significant at 0.1% level. Their z-values are reasonable. In particular, unemployment rate's z-values are −4.51 in ln(District API)'s prediction equation, and −4.88 in HHI's prediction equation. The number of springs variable's z-values are −4.28 in ln(District API)'s prediction equation and 5.78 in HHI's prediction equation. For a single endogenous variable, a commonly used rule-of-thumb to justify the strength of an instrument is to have its z-value greater than $\sqrt{10} \cong 3.16$ (or F-value > 10). In our case, all relevant z-values are in line with this rule-of-thumb.

The estimated effect of ln(district API) on per pupil cost is positive and considerably larger in Model EN than in Model EX.[25] In order to evaluate the effects of this difference on per pupil expenditures, we first assume that an imaginary school district's independent variables are equal to the state averages. We can predict this hypothetical district's expenditure per pupil by using the corresponding parameter estimates from our models. Since the expenditure information of this hypothetical district is not observed, an estimate of this district's error term is not available. Instead, we assume that the estimate of the inefficiency term u is:

$$\hat{u} = E(u) = \sqrt{\exp(x_2'\hat{\varphi}_u)} \cdot \frac{\sqrt{2}}{\sqrt{\pi}} \tag{7}$$

where $\hat{\varphi}$ is the coefficient vector from the inefficiency function of either Model EX or Model EN. Using Equation (7), when we evaluate the coefficient estimates of Model EX, we find that the

[24] As explained by Karakaplan and Kutlu (2017), estimations with this model are done in a single-stage. So, to avoid confusion, we do not use the name "first-stage statistics" and instead, call them prediction equations.

[25] Our simulation results confirm the possibility of such large biases.

predicted expenditure per pupil of the hypothetical district is $8050.48, and an increase in the average API by 10% (equivalent to a 78.37 points increase) would require an increase in the per pupil expenditure by $330.40 (4.1% increase), ceteris paribus.[26] Considering that mean district enrollment in California is 6,544 students, an average-sized school district would have a predicted total expenditure of $52.68 million and would need to incur an additional $2.16 million to attain 10% increase in the average API. If we examine the same scenario with the coefficient estimates of Model EN, we find that the predicted expenditure per pupil of the imaginary district is $8285.31, and for an increase in average API by 10%, the needed increase in per pupil expenditure would be about $4489.54 (54.19% increase). The difference between what Model EX and Model EN predict for the increase in per pupil expenditures is about $4159.14 (about 13-fold). At the district level, total predicted expenditure would be $54.19 million with Model EN and the increase in the total predicted expenditure is $29.38 million when the average API is increased by 10%, ceteris paribus. The difference between the predictions is substantial especially when the overall enrollment in California is taken into account. If the state was composed of 913 of these hypothetical districts, to increase in the average API by 10% in every district, the state would potentially spend $1.97 billion more in addition to the $48.10 billion predicted state expenditures according to Model EX, and $26.82 billion more on top of the $49.50 billion predicted state expenditures according to Model EN. The huge difference of $24.85 billion additional expenditures is a quantitative indication of the underestimated effect of student achievement on per pupil expenditure under the exogeneity assumption.

[26] We calculate ln(District API) corresponding to 10% above the average API as $\ln(1.1 \times \exp(6.664)) = 6.760$.

Moreover, the effect of HHI on expenditure per pupil is higher in Model EN than that in Model EX. For the imaginary district outlined above, when the coefficient estimates from Model EX are evaluated, we find that an increase in the average HHI by 0.1 would result in an increase in the per pupil expenditure by $125.58 (1.56% increase).[27] Such an increase in per pupil expenditure due to an increase in HHI would result in a total increase in the expenditure by about $0.82 million for an average-sized school district. On the other hand, when the same scenario is evaluated with the coefficient estimates from Model EN, the increase in the average HHI by 0.1 would result in an increase in the per pupil expenditure by $188.78 (2.28% increase). This is $63.19 (50.32%) more than what Model EX predicts. The total increase in the expenditure of an average-sized district due to the change in HHI would be $1.24 million in the endogenous model, which is $0.42 million more than the total district expenditure predicted with the exogenous model. These findings show that while it is essential to include a market concentration measure in the cost model to capture the extra expenditures due to increased cost inefficiencies, it is also critical to handle the endogeneity of that measure carefully to comprehend the potential increases in the expenditures better.

The rest of the coefficient estimates in Model EN are also different than those in Model EX. The cost function is still U-shaped in Model EN with respect to enrollment, but the minimum cost is at 6,704 students. The coefficients of inputs and input prices are higher in Model EN. Also, the coefficient of the percent low income students is positive and significant in Model EN, which

[27] Such a change in a district's HHI is not necessarily an outcome of a change in the enrollment of the district. The chance in a district's HHI can solely be a result of changes in the enrollments of other districts in the district's education market.

suits the theoretical expectations better.

Finally, Figure 2 presents the histograms of district cost efficiencies in Model EX and Model EN. This visual illustration of the differences between these two models provides a better understanding of the effect of endogeneity correction on cost efficiencies. We find that the average district cost efficiency in Model EN is 0.9061 with minimum efficiency at 0.3242 and maximum at 0.9765. As Figure 2 shows, the cost efficiencies of the school districts in Model EN are somewhat less than their cost efficiencies reported in Model EX. We find that this difference is 0.0235 on average and its median is 0.0142 with a minimum of 0.0002 and a maximum of 0.1989. That is, in the exogenous model, school districts appear slightly more cost efficient than they actually are. Moreover, we find that this average difference is larger for districts in highly concentrated education markets. For instance, the average cost efficiency difference between Model EX and Model EN is 0.0694 for the 132 districts that are in education markets with concentrations larger than 0.5. The same average difference is 0.0923 for the 53 districts that are in education markets with concentrations larger than 0.8. The intuition of this finding is that even though education market concentration has a negative effect on cost efficiency, that effect is underestimated with the exogeneity assumption. Since concentration would possibly decrease the district performance and a decrease in district performance decreases the expenditures, exogeneity assumption would result in an upward-biased cost efficiency. In Model EN, we find that concentration is actually endogenous and therefore treating the endogeneity removes the upward bias in cost efficiencies. Finally, a Kolmogorov-Smirnov test of equality of distributions indicates that the distributions of cost efficiencies in these two models are significantly different at 0.01% level ($p = 0.000$), and the school districts are significantly less cost efficient in Model EN than in Model EX.

[Figure 2: Somewhere here.]

5.2. Sensitivity Analyses

5.2.1. IV Adequacy

In this section, we offer alternative strategies to investigate if our findings rely on the set of instruments we chose to use or if the results stay essentially the same with different approaches. In particular, we try using only exogenous sources of local education supply as IVs. As explained in the previous section, topographical measures provide valid IVs for endogenous educational variables. With the same reasoning, we explore employing different topographical measures as instrumental variables for both student achievement and education market concentration variables. This approach would eliminate the concerns based on the validity of the IVs that are based on demand side equations such as the unemployment rate.

[Table 3: Somewhere here.]

In the first column of Table 3, we use the number of summits and the number of swamps in a county as the only two IVs to handle the endogeneity in the model. In the second column of Table 3, we use the number of lakes and the number of plains in a county as the only two IVs. The results in Table 2 and Table 3 are quite similar, especially the coefficients and significance of the endogenous variables and the results from the endogeneity tests. Looking at Table 3's prediction equations, excluded instruments are statistically significant and their z-values reasonably satisfy the rule-of-thumb mentioned earlier to justify the strength of them. We also find that using other

IV sets with different combinations of topographical measures based on valleys, basins, streams, etc. generate results that consistently resemble the results in Table 2. We omit the details of our finding here but they are available upon request. These findings indicate that the results in Table 2 do not heavily depend on the selection of IVs. Furthermore, these findings indirectly support that the set of IVs in Table 2, particularly the demand side variable, is usable. We also tried weighted averages of instruments proposed by Lewbel (2012) so that for each endogenous variable we have one instrument. Even though these generated instruments had very low correlations with our original instruments, the qualitative and quantitative results were very similar. We believe that all these outcomes outlined in this subsection provide strong support for the exclusion constraints.

5.2.2. Restricted Sample and Additional Endogeneity

We also report some other robustness analyses to check the sensitivity of our results presented in the previous section. We begin with examining the effects of district enrollment density on the coefficients. Imazeki and Reschovsky (2004) exclude two outlier districts (Houston and Dallas school districts) from their analysis because of the large variation in public school district size in Texas. Similarly, Gronberg et al. (2015) exclude low density school districts as a robustness check. In order to control if enrollment density is a significant factor in our study, we drop the districts below the 5^{th} percentile and above the 95^{th} percentile in Table 4. When we exclude potentially outlying districts from the sample, we find that the coefficients change fractionally, but their signs and significance do not change and the general findings presented in Table 2 still hold.

[Table 4: Somewhere here.]

In Table 5, we consider the possibility of endogeneity of the enrollment variable and its squared term.[28] While the literature treats enrollment as an exogenous variable, there are theoretical reasons, similar to that of the endogenous output quality, to believe that the output quantity and the two-sided error term may be correlated. For example, in our case, districts with low enrollments have fiscal incentives through state programs and other channels that would motivate them to try to remain small and use resources to accomplish that goal. In fact, consolidation recommendations such as the report of California's Legislative Analyst's Office focus not only on the savings due to the economies of size, but also on stopping funding advantages given to small districts and their expenses made towards acquiring these additional funds. Hence, district enrollment in our study may be endogenous. We instrument the enrollment variables (as well as the concentration and achievement variables) with the instrumental variables we used earlier and another topographical supply side variable, namely the number of basins in a county. It turns out that in this case, some of our previous findings change. Under the endogenous enrollment assumption, the coefficients of the concentration and enrollment terms are not appreciably different than the corresponding coefficients in Model EN in Table 2 in terms of size

[28] We can include only a single bias correction term for enrollment variable and its squared term. Intuitively, this is because the squared enrollment term is a function of enrollment, and conditional on enrollment, these two variables are exogenous. This may be seen from the decomposition of the log-likelihood $\ln L(\theta) = \ln L_{y|x}(\theta) + \ln L_x(\theta)$. In the standard instrumental variables approaches such as 2SLS, a prediction equation for each function of an endogenous variable would be needed. In our setting, however, we need a prediction equation only for the original endogenous variable but not for its functions. For more details regarding this issue, see Wooldridge (2010) and Amsler et al. (2016).

and significance, but the coefficient of ln(district API) seems to be smaller than that in Table 2 by about 1.7 points. It is also important to note that in Table 5, the significance of the wage index, percent low income students, and the constant term goes away. This is not surprising given the fact that in this particular setting, we have more endogenous variables. Even though some of the results in Table 5 are different, we find that the qualitative outcomes that we specifically concentrate on in this paper are valid.

[Table 5: Somewhere here.]

5.2.3. Spatial Interactions and Variation in Cost Efficiency

Spatial interaction effect may be considerable in education markets, and if so, ignoring the spatial interaction can lead to biased parameter estimates and misleading policy advice. It could be nice to evaluate the possibility of interaction and spillover effect using a spatial lag model. The literature on spatial stochastic frontier models is sparse. Druska and Horrace (2004) extend the spatial cross-sectional data model of Kelejian and Prucha (1999) to the panel data setting where the time-invariant efficiencies are calculated using the fixed effects as in Schmidt and Sickles (1984). Glass et al. (2013a) use a similar approach following Cornwell et al. (1990), which allows the efficiency to depend on time through second degree time polynomials. These models are not suitable for our purpose as the interplay between the education market concentration and inefficiency is ignored. Basically, these two models are subject to similar criticism that we discussed for Millimet and Collier (2008). In contrast to these studies, Adetutu et al. (2014) and Glass et al. (2013b) introduce spatial stochastic frontier models where distributional assumptions about the inefficiency are made. However, these models do not handle endogeneity. Therefore, these models are not suitable to analyze our policy questions here. Extending these models to address endogeneity is possible using a control function approach similar to ours but that is beyond the scope of this study.

To investigate the interaction and spillover effects, Millimet and Collier (2008) propose that the efficiency of a school district is a function of the average values of these variables in the neighboring districts. They define the neighboring districts of a district as all other districts in the county of that district. We modify their approach and define the neighboring value as the average of a variable in all districts that share a physical border with the observed district. While our approach requires a more complicated algorithm to find the neighbors, it is more refined as we relax the implicit assumption that districts that share a border but not in the same county do not directly affect each other's efficiency.

Moreover, in terms of model specification, the list of variables that we used in Table 2 is rather conservative. Specifically speaking, even though using a restricted set of input and control variables in education models is not uncommon in the literature[29], comments can be made about some variables that are possibly omitted in our model, and how that may impact the cost inefficiency. Especially since we specify the market concentration as the sole source of variation in cost efficiency, we think that there is a need to justify our approach by testing if our findings would change with the inclusion of more variation. For this test, we use the extensive list of education input and control variables provided by Millimet and Collier (2008).

The first column in Table 6 presents an extended version of the Model EN in Table 2, with a complete set of neighboring values as determinants of the district cost inefficiency. The second

---

[29] See Grosskopf et al. (2001), Imazeki and Reschovsky (2004), Gronberg et al. (2015), and others.

column in Table 6, extends the model further by including possibly omitted variables in the cost frontier. These variables are summarized in Table 2. We find that most spillover variables are not significantly effective. Only neighboring values of white population and a few others are significant at the 5% level. Evaluating the spillover variables as endogenous does not yield different findings.[30] Hence, the results in Table 6 would raise questions about the importance of spillover effects. Cost efficiencies in Table 6, however, are less than that in our baseline model EN, which points out their sensitivity to the inclusion of the neighboring values and other additional variables. Extending the baseline model improves the log likelihood substantially, and a likelihood ratio test would indicate that Extended Model EN-2 fits significantly better than Extended Model EN-1 or Model EN in Table 2.[31] Looking at the variables of our interest, district achievement and market concentration variables are endogenous, their effects are positive and significant as in Model EN of Table 2, and their effect sizes are not substantially different.

[Table 6: Somewhere here.]

5.2.4. Alternative Measures of Achievement and Concentration

A commonly raised criticism is about the credibility of the output quality measures used in education models. As we mentioned earlier, education production technology can be considered

---

[30] These findings are available upon request.

[31] It is important to note that Table 6 includes only two extended versions of Model EN in Table 3. Model EXs are excluded from Table 6 but available upon request. Assuming that Model EN in Table 3 is the null model, Extended Model EN-1 as an alternative model has a significantly better fit, and Extended Model EN-2 has an even better fit than Extended Model EN-1 according to the likelihood ratio test.

as a multiple output process, but some of the education outputs may be hard to measure. While the school district API variable that we utilized is an aggregated measure of the overall district performance, there may be some quality criteria that the district API misses. To control for this possibility, we use a different aggregate measure called Adequate Yearly Progress (AYP) which is California's statewide accountability system for districts mandated by the No Child Left Behind Act of 2001. AYP measure captures a wider range of output quality involving several district-specific performance criteria based on participation rate, proficiency rate, and graduation rate.[32] In Table 7, instead of using ln(district API) as the aggregate output quality measure, we use the percentage of AYP criteria that are met by each district. The variable has a mean of 0.796 and a standard deviation of 0.166, with its minimum and maximum at 0.115 and 1. Larger AYP rates indicate better overall district performance. Table 7 shows that when the endogeneity of the district AYP is handled properly, AYP's positive and significant effect increases more than fourfold, a rather similar result to that in Table 2. The rest of the qualitative findings are essentially the same as our baseline estimation results.

[Table 7: Somewhere here.]

Table 8 presents regression results using the concentration measure based on 15-mile radius circular approach (HHI-R15). The mean of HHI-R15 is 0.197 with a standard deviation of 0.177, minimum of 0.013, and maximum of 1.[33] HHI-R15 has a larger effect on cost inefficiency

---

[32] School districts have different numbers of criteria to be met depending on the applicability of the criteria.

[33] While the education market areas are larger for each district with the HHI-R15 measure, HHI-R15 of a

compared to HHI, and when we control for its endogeneity, the effect size gets even larger.[34]

[Table 8: Somewhere here.]

In Table 9, instead of using our radial approach to calculate the market concentration, we follow the literature mentioned above and construct a metropolitan area based concentration measure (HHI-MA), which is the sum of squared enrollment shares of all public and private school districts in Core Based Statistical Areas (CBSA).[35] [36] We find that when the concentration measure is calculated with this approach, there are 48 education markets in our sample. The district average of HHI-MA is 0.120, and its standard deviation is 0.106 with the minimum at 0.030 and

---

district can be larger or smaller than, or equal to the HHI measure of that district depending on how much the enrollment and the number of districts in the district's education market change with the increase in the education market area. In our sample, HHI of 723 districts is larger than their HHI-R15, and 166 districts have smaller HHI than their HHI-R15. HHI of 24 districts is equal to their HHI-R15.

[34] The regression results with 20-mile and 30-mile radii approaches are in line with the results in Table 8, so we omitted them.

[35] CBSAs are identified by the Office of Management and Budget. A metropolitan area contains a core urban population of 50,000 or more, and a micropolitan area contains an urban core population of more than 10,000 but less than 50,000. Each metropolitan or micropolitan area contains counties with the core urban areas and any neighboring counties with a high degree of social and economic integration with the urban core. If a county does not belong to a metropolitan or micropolitan area, then we take that county as a separate educational market.

[36] CDE assigns each cross-county school district to a single county. To determine the locations of public school districts, we use that designation. We get the location and enrollment data of private schools from the National Center for Education Statistics (NCES) Private School Universe Survey (PSS).

the maximum at 1. This measure has even a larger impact on cost inefficiency than HHI-R15, and handling its endogeneity increases this impact further. The two analyses in Table 8 and 9 would indicate that while the effect size of concentration variable is sensitive to how concentration is measured, its endogeneity, coefficient sign and significance, and the change in its effect size after its endogeneity is addressed remain the same in essence.

[Table 9: Somewhere here.]

Table 10 summarizes the estimation results from all the cases above including the baseline estimation. First, we find that in all our estimations, the coefficients of the concentration and achievement measures are significant. The top panel of Table 10 compares the Model EX and Model EN results in each table. When the endogeneity of concentration measure is controlled for, its coefficient increases by 0.44 on average with a 0.23 standard deviation. Percent increase in its coefficient is 10.19 on average. The average increase in the coefficient of achievement measure is 3.66 when its endogeneity is handled. The effect of the achievement measure on cost increases by 9.40-fold on average in the Model ENs.

[Table 10: Somewhere here.]

The second panel of Table 10 summarizes the predicted expenditures of a hypothetical district with its independent variables assumed to be at state averages. Using Equation (7), according to Model EXs, the predicted per pupil expenditure of a hypothetical district is $8,124.82 on average, and according to Model ENs, the predicted per pupil expenditure is $8,256.61 on average. The average-sized district's total expenditure is predicted to be $53.17 million on average based on Model EXs, and $54.03 million on average based on Model ENs. If the state had 913 of these imaginary districts, total predicted expenditure of the state would be $48.54 billion on average according to Model EXs, and $49.33 billion on average according to Model ENs.

The third panel of Table 10 presents the required additional expenditure for increasing the average student achievement measure by 10% in the imaginary district while keeping every other independent variable at the state averages. We use Equation (7), and find that this change requires an increase in the expenditure per pupil by $284.92 on average in Model EXs. In Model ENs, however, the same effect requires an increase by $3,598.87 on average. When the endogeneity in Model EXs is treated, the effect on required expenditure per pupil goes up by 11.67-fold on average. The difference between the average increase in the total expenditure of the average-sized district in Model EXs and Model ENs is more than $20 million. Furthermore, as a result of 10% increase in the average student achievement measure, the average required increase in the total expenditure of the state of 913 identical hypothetical districts is $19.8 billion more in Model ENs than that in Model EXs.

The bottom panel of Table 10 displays the findings when concentration measure is increased by 0.1 while holding all other right-hand-side variables of the hypothetical district at the state averages. Using Equation (7), we find that this change increases the expenditure per pupil by $129.13 on average in Model EXs with a minimum of $59.15 and maximum of $197.07. When the endogeneity of concentration and achievement measures is mitigated, the increase in the expenditure per pupil goes up by 63.79% on average with a 46.35% standard deviation. The increase in the expenditure per pupil is $68.72 more on average in Model ENs compared to the corresponding Model EXs. Change in the average concentration measure increases the total expenditure of an average-sized district by about $0.85 million on average in the exogenous models compared to $1.29 million on average in the endogenous models.

These findings in Table 10 indicate that the baseline results presented in the previous section are robust. That is, the effects of concentration and achievement measures on cost are significant and these effects are considerably larger when their endogeneity is properly addressed.

6. Economic Simulation of a Consolidation Policy

One of the recommendations in the Supplemental Report of the 2010-11 Budget Package of California's Legislative Analyst's Office is to increase the minimum threshold for district size to at least 100 students to save money. About 10% of California's school districts have less than 100 students. If the state implements the recommended threshold, these "very small" districts would be possibly required to consolidate with nearby districts to form larger school districts.[37]

In this section, we simulate the proposed consolidation plan outlined above to analyze the effects of market concentration and its endogeneity on total state expenditures. It is important to clarify that our objective here is not an analysis of a realized consolidation. Instead, our aim in this section is to evaluate how handling the endogeneity in our education cost model changes the predicted outcomes of the consolidation policy. To answer that question, we use Equation (7) and calculate the predicted costs of the districts and their cost inefficiencies before and after a consolidation scenario, and compare the potential overall savings or losses using the estimates from Model EX and Model EN in Table 2. As an alternative, estimates from Table 5 with

---

[37] California State is known to be carrying out various consolidation projects in the past that the number of traditional public school districts decreased from more than 2,000 to less than 1,000 in fifty years.

endogenous enrollment, Table 6 with extended models, or one of our other specifications could be used as well. Our decision to use the results from Table 2 in this section is based on the fact that results from Table 2 closely represent the average results from all of our alternative specifications in Table 10. When we examine the simulation exercise outlined in this section with Model EX and Model EN estimates from the alternative models, we find that the quantitative results presented in this section do not appreciably change. Moreover, qualitatively, our conclusion presented in this section that ignoring endogeneity leads to overestimated saving expectations from a consolidation plan holds regardless of the alternative model specification we use.

We calculate the pre-consolidation predicted expenditure per pupil of a district by using the observed district-level values of the variables in Table 2, and their coefficients. We include the school districts with missing observations to our simulation analysis by assigning the state averages for the corresponding missing values. This enables us to evaluate the expenditures of all 935 traditional public school districts instead of examining only the regression subsample of 913 districts.[38]

We make a set of assumptions to calculate the post-consolidation predicted expenditures per pupil. First of all, we assume that families do not move as an immediate reaction to the consolidation decision of districts.[39] With that assumption, we perform the following

[38] Using only the regression subsample is an alternative economic simulation approach. However, our findings indicate that excluding the 22 school districts with missing observations from the analysis does not change the results qualitatively.

[39] There is not any strong factual evidence against this assumption, and families do not have many incentives to move solely due to a consolidation decision of a district, at least not within the first year of consolidation during

consolidation algorithm: We find the smallest district in the state with less than 100 students, and then we consolidate that district with the smallest neighboring district in the same county to form a single larger district. To give an example, in 2010-11 school year, Big Creek Elementary School District has 38 students and is the smallest district in Fresno County. Pine Ridge Elementary School District in Fresno County has 92 students, and is the smallest neighboring district of Big Creek Elementary. Our consolidation algorithm merges these two districts and form a single district with 130 students. To give another example, Santa Clara Elementary School District has 56 students and is the smallest district in Ventura County. Mupu Elementary School District in Ventura County has 132 students and is the smallest neighboring district of Santa Clara Elementary. Our algorithm consolidates these two districts and make a single district with 188 students. We repeat this consolidation algorithm until the smallest district in California has at least 100 students. As a result of this process, the total number of districts in California decreases by 78.[40]

---

when the overall consequences of consolidation would be ambiguous to the public. If we relax this assumption, and if families expect that the consolidated district will be worse than their current district, they would move to other districts, or if they are already in other districts, they would not move to the consolidated district. That would decrease the education market share of the consolidated district compared to sum of pre-consolidation market shares of the consolidated districts and decrease the resources that were available to them to spend. In this case, education market concentration would not increase as much as it would if parents do not move. So, the cost inefficiency of the consolidated district would not be as big as it would be if parents do not move.

[40] In 2010, California has 90 "very small" districts which has less than 100 students. The decrease in the total number of districts is less than 90 because some very small districts have another very small district in the neighborhood so that they merge and form a district with more than 100 students by eliminating two very small

Because of consolidations, district boundaries change, and so do the education market boundaries. Hence, we recalculate the HHI measure of all 857 post-consolidation districts, while assuming that charter and private school locations and enrollments would not change as a direct response to traditional public school district consolidations.[41] Before consolidation, the average HHI of 935 school districts is 0.275. However, the average HHI of districts with less than 100 students is 0.603, which is largely due to the fact that most very small districts in California are isolated such as Death Valley Unified School District in Inyo County, remote such as Desert Center Unified School District in Riverside County, or peripheral to the competitive core such as Lincoln Elementary School District in Marin County. Through consolidation, these very small districts with high concentration merge with other districts and become larger districts with larger education markets. After consolidation, the average HHI of 857 school districts is 0.249.[42]

Finally, for the price index variable of a new district, we assign the linear average of the price indices from the two consolidating districts[43], and for all the other variables in Table 2, we assign the enrollment weighted average of each variable from the two consolidating districts to the new district they form. For instance, in 2010-11 school year, Meridian Elementary School District

---

districts at the same time.

[41] This assumption is in line with the assumption that parents do not move as an immediate reaction to the consolidation decision of districts. As explained in Footnote 39, we do not have any substantial real-life evidence for students' changing schools as a response to accommodate the consolidation of traditional school districts.

[42] Median HHI before consolidation is 0.194 and the average HHI of the districts more than or equal to 100 students is 0.240 before consolidation. These numbers are smaller than 0.249 as expected.

[43] Since price index is a proxy based on linear remoteness of a district from the nearest major metropolitan area, the linear average of the price indices from the two consolidation districts is a suitable proxy for the new district.

in Sutter County has 85 students and 19% of them have limited English proficiency. The smallest neighboring district of Meridian Elementary is Winship-Robbins School District in Sutter County which has 183 students and 36% of them have limited English proficiency. We assume that when these two districts consolidate, 30.6% (the enrollment weighted average of 19% and 36%) of the students of the new district would have limited English proficiency. Pre- and post-consolidation descriptive statistics are presented in Table 11.

[Table 11: Somewhere here.]

The state-level predicted outcomes of our consolidation simulation have remarkable implications. The first and second column of Table 12 is based on Model EX and Model EN estimates in Table 2, respectively. According to Model EX predictions, California's total efficient cost[44] before consolidation is $47.10 billion, total inefficiency is $2.17 billion, and total predicted expenditure is their sum, which is $49.27 billion. After consolidation, California's total efficient cost decreases to $47.06 billion, total inefficiency increases to $2.19 billion, and total predicted expenditure decreases by about $14.28 million. To compare, we find that a stochastic cost model that excludes the concentration measure predicts that our consolidation simulation would decrease state's total expenditure by $24.87 million. Such decreases in total expenditure may be viewed by some audiences as too small to be accounted for, especially when these amounts are compared to the state's actual total expenditures which is around $50 billion. On the other hand, actual total expenditure of the districts with less than 100 students is $57.87 million, so even a $14-$24 million

[44] By efficient cost, we mean the cost when there is no inefficiency.

decrease in total state expenditures is important when the amount is compared to the total budget of very small districts. Hence, such seemingly small decreases in total expenditure may be viewed by others such as consolidation policy advocates as desirable, and they may recommend consolidation based on these outcomes.

[Table 12: Somewhere here.]

It is important to emphasize again that our intention with this exercise is to investigate how handling the endogeneity in the education cost model changes the predicted findings of a proposed consolidation scenario. Predictions based on Model EX indicate that the proposed consolidation would result in some savings as explained above. Model EN, however, predicts that the change in state's total expenditure can be in the opposite direction. According to Model EN, California's total predicted efficient cost is $49.96 billion, total predicted inefficiency is $2.98 billion, and total predicted expenditure is $52.94 billion before consolidation. Post-consolidation, state's total efficient cost decreases to $49.93 billion, and total predicted inefficiency increases to $3.02 billion. As a result, total predicted expenditure of the state increases by $10.53 million, which indicates that consolidation of school districts can lead saving reversals to an extent that the state actually loses resources in overall. This finding is substantially different than that based on Model EX. Consolidation policies cannot be recommended based on such losses. Therefore, our consolidation simulation exercise indicates that it is crucial to address the endogeneity in the education cost model to have a better idea about what to expect out of a consolidation policy and its suitability.

On a final note, with both Model EX and Model EN estimates, consolidation decreases the predicted efficient cost and increases the predicted inefficiency. The difference in the direction of the change in state's predicted expenditure is due to the fact that with Model EX, savings in efficient cost dominates the additional inefficiency, while with Model EN, the additional inefficiency outweighs the savings in efficient cost. In other words, handling the endogeneity in our model modifies the set of coefficients in a fashion that boosts the effects of consolidation on inefficiency, and dwarfs the effects of consolidation on savings from the efficient cost to a point that flips the sign of their overall difference. Our conclusion highlights that ignoring endogeneity in the model leads to overestimated savings as in the exogenous model of Gronberg et al. (2015), and mitigating endogeneity results in a dramatic reduction in those savings. Treating the endogeneity problem in our education cost model with our econometric methodology reveals the possibility of saving reversals and aggregate losses due to a consolidation policy which may appear like a proper cost-saving strategy under exogeneity assumptions.

7. Concluding Remarks

Education finance policies often include a clause that recommends school district consolidation to take advantage of size economies and save resources. Consolidation, however, would transform the education market structure, and policy makers generally do not pay much attention to this important detail. In this study, we use a stochastic education cost frontier model to estimate the determinants of school district expenditures. This analysis is challenging because of the endogeneity issues in the stochastic cost frontier model.

We present a methodology that would estimate the degrees of public school district cost inefficiency in California while handling the endogeneity in the model. We show that the estimated effect of student achievement on expenditure per pupil is substantially larger in the model which remedies the endogeneity of achievement measure. We also show that the effect of education market concentration on cost inefficiency is larger in the model that considers the endogeneity of concentration. These differences provide a quantitative evidence that the effects of achievement and concentration on per pupil expenditure are underestimated under the exogeneity assumption.

We provide many alternative specifications including a spatial interactions model that present the robustness of our findings. Our economic simulation of a consolidation scenario indicates that handling endogeneity can diminish the predicted savings even to an extent in which a consolidation policy results in overall losses while the exogenous model predicts savings. Due to the possibility of such saving reversals, we conclude that the policies or studies that ignore the effects of important cost and inefficiency determinants, or mishandle their endogeneity may not be reliable. Consequently, we recommend being cautious with policies to consolidate school districts.

References

Adetutu M, Glass AJ, Kenjegalieva K, Sickles RC (2014) The effects of efficiency and TFP growth on pollution in Europe: A multistage spatial analysis. Journal of Productivity Analysis 43:307-326

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6:21-37

Amsler C, Prokhorov A, Schmidt P (2016) Endogenous stochastic frontier models. Journal of Econometrics 190:280-288

Amsler C, Prokhorov A, Schmidt P (2017) Endogenous environmental variables in stochastic frontier models. Journal of Econometrics:Forthcoming

Andrews M, Duncombe W, Yinger J (2002) Revisiting economies of size in American education: Are we any closer to a consensus? Economics of Education Review 21:245-262

Belfield CR, Levin HM (2002) The effects of competition between schools on educational outcomes: A review for the United States. Review of Educational Research 72:279-341

Borland MV, Howsen RM (1992) Student academic achievement and the degree of market concentration in education. Economics of Education Review 11:31-39

Cornwell C, Schmidt P, Sickles RC (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics 46:185-200

Costrell R, Hanushek E, Loeb S (2008) What do cost functions tell us about the cost of an adequate education? Peabody Journal of Education 83:198-223

Dodson ME, III, Garrett TA (2004) Inefficient education spending in public school districts: A case for consolidation? Contemporary Economic Policy 22:270-280

Druska V, Horrace WC (2004) Generalized moments estimation for spatial panel data: Indonesian rice farming. American Journal of Agricultural Economics 86:185-198

Duncombe W, Miner J, Ruggiero J (1995) Potential cost savings from school district consolidation: A case study of New York. Economics of Education Review 14:265-284

Duncombe W, Yinger J (2007) Does school district consolidation cut costs? Education Finance and Policy 2:341-375

Duncombe W, Yinger J (2011) Making do: State constraints and local responses in California's education finance system. International Tax and Public Finance 18:337-368

Duygun M, Kutlu L, Sickles RC (2016) Measuring productivity and efficiency: A Kalman filter approach. Journal of Productivity Analysis 46:155-167

Fitzpatrick MD, Grissmer D, Hastedt S (2011) What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. Economics of Education Review 30:269-279

Fried HO, Lovell CAK, Schmidt SS (2008) The measurement of productive efficiency and productivity growth. Oxford University Press, Oxford

Glass A, Kenjegalieva K, Paez-Farrell J (2013a) Productivity growth decomposition using a spatial autoregressive frontier model. Economics Letters 119:291-295

Glass A, Kenjegalieva K, Sickles RC (2013b) A spatial autoregressive production frontier model for panel data: With an application to european countries. Available at SSRN 2227720

Greene W (2005a) Fixed and random effects in stochastic frontier models. Journal of productivity analysis 23:7-32

Greene W (2005b) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126:269-303

Greene WH (1980a) Maximum likelihood estimation of econometric frontier functions. Journal of

Econometrics 3:27-56

Greene WH (1980b) On the estimation of a flexible frontier production model. Journal of Econometrics 13:101-115

Greene WH (2003) Simulated likelihood estimation of the normal-gamma stochastic frontier function. Journal of Productivity Analysis 19:179-190

Gronberg TJ, Jansen DW, Karakaplan MU, Taylor LL (2015) School district consolidation: Market concentration and the scale-efficiency tradeoff. Southern Economic Journal 82:580-597

Gronberg TJ, Jansen DW, Taylor LL (2011a) The adequacy of educational cost functions: Lessons from Texas. Peabody Journal of Education 86:3-27

Gronberg TJ, Jansen DW, Taylor LL (2011b) The impact of facilities on the cost of education. National Tax Journal 64:193-218

Gronberg TJ, Jansen DW, Taylor LL (2012) The relative efficiency of charter schools: A cost frontier approach. Economics of Education Review 31:302-317

Grosskopf S, Hayes KJ, Taylor LL (2009) The relative efficiency of charter schools. Annals of Public and Cooperative Economics 80:67-87 doi:10.1111/j.1467-8292.2008.00381.x

Grosskopf S, Hayes KJ, Taylor LL, Weber WL (2001) On the determinants of school district efficiency: Competition and monitoring. Journal of Urban Economics 49:453-478 doi:10.1006/juec.2000.2201

Guan Z, Kumbhakar SC, Myers RJ, Lansink AO (2009) Measuring excess capital capacity in agricultural production. American Journal of Agricultural Economics 91:765-776

Hanushek EA (2005) The alchemy of ‘costing out’ an adequate education. In: Conference on Adequacy Lawsuits, 2005.

Hanushek EA (2006) Science violated: Spending projections and the ‘costing out’of an adequate education. Courting failure:257-308

Hobbs V (2004) The promise and the power of distance learning in rural education. Policy brief. Rural School and Community Trust

Hoxby CM (2000) Does competition among public schools benefit students and taxpayers? American Economic Review 90:1209-1238

Hoxby CM (2003) School choice and school productivity: Could school choice be a tide that lifts all boats? Chicago and London: University of Chicago Press,

Imazeki J, Reschovsky A (2004) Is no child left behind an un (or under) funded federal mandate? Evidence from Texas. National Tax Journal 57:571-588

Izadi H, Johnes G, Oskrochi R, Crouchley R (2002) Stochastic frontier estimation of a CES cost function: The case of higher education in Britain. Economics of Education Review 21:63-71

Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19:233-233

Karakaplan MU (2017) Fitting endogenous stochastic frontier models in Stata. The Stata Journal 17:39-55

Karakaplan MU, Kutlu L (2017) Handling endogeneity in stochastic frontier analysis. Economics Bulletin 37:889-891

Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40:509-533

Kumbhakar SC, Lovell CK (2003) Stochastic frontier analysis. Cambridge University Press,

Kutlu L (2010) Battese-Coelli estimator with endogenous regressors. Economics Letters 109:79-

81

Kutlu L (2017) A constrained state space approach for estimating firm efficiency. Economics Letters 152:54-56

Kutlu L, Tran KC, Tsionas MG (2017) A time-varying true individual effects model with endogenous regressors. Unpublished Manuscript

Levinsohn J, Petrin A (2003) Estimating production functions using inputs to control for unobservables. The Review of Economic Studies 70:317-341

Lewbel A (2012) Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. Journal of Business & Economic Statistics 30:67-80

Marcotte DE, Hemelt SW (2008) Unscheduled school closings and student performance. Education Finance and Policy 3:316-338 doi:10.1162/edfp.2008.3.3.316

Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production function with composed errors. International Economic Review 18:435-444

Millimet DL, Collier T (2008) Efficiency in public schools: Does competition matter? Journal of Econometrics 145:134-157 doi:10.1016/j.jeconom.2008.05.001

Mutter RL, Greene WH, Spector W, Rosko MD, Mukamel DB (2013) Investigating the impact of endogeneity on inefficiency estimates in the application of stochastic frontier analysis to nursing homes. Journal of Productivity Analysis 39:101-110

O'Donnell CJ, Coelli TJ (2005) A bayesian approach to imposing curvature on distance functions. Journal of Econometrics 126:493-523

Rothstein J (2007) Does competition among public schools benefit students and taxpayers? Comment. American Economic Review 97:2026-2037

Schmidt P, Sickles RC (1984) Production frontiers and panel data. Journal of Business and

Economic Statistics 2:367-374

Shee A, Stefanou SE (2014) Endogeneity corrected stochastic production frontier and technical efficiency. American Journal of Agricultural Economics:aau083

Stevenson R (1980) Likelihood functions for generalized stochastic frontier functions. Journal of Econometrics 13:57-66

Taylor LL, Fowler Jr WJ (2006) A comparable wage approach to geographic cost adjustment. Research and development report. NCES-2006-321. National Center for Education Statistics

Timar T (2006) Financing k–12 education in California. A system overview. Unpublished Manuscript, University of California, Davis, March

Tran KC, Tsionas EG (2013) GMM estimation of stochastic frontier model with endogenous regressors. Economics Letters 118:233-236

Tran KC, Tsionas EG (2015) Endogeneity in stochastic frontier models: Copula approach without external instruments. Economics Letters 133:85-88

Wang H-J, Schmidt P (2002) One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. Journal of Productivity Analysis 18:129-144

Wong KK et al. (2014) Why is school closed today? Unplanned k-12 school closures in the United States, 2011–2013. PloS one 9:e113755

Zanzig BR (1997) Measuring the impact of competition in local government education markets on the cognitive achievement of students. Economics of Education Review 16:431-444

Zimmer T, DeBoer L, Hirth M (2009) Examining economies of scale in school consolidation: Assessment of indiana school districts. Journal of Education Finance 35:103-127

Table 1: Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
ln(Expenditure per pupil)	9.03	0.335	6.268	10.497
ln(Enrollment)	7.344	1.879	2.303	13.416
Percent limited English proficiency	0.186	0.173	0	0.8
Percent Special Education students	0.088	0.043	0	0.299
Percent high school students	0.207	0.28	0	1
Percent low income students	0.514	0.262	0	0.995
Price Index	0.47	0.06	0.026	0.64
Wage Index	0.389	0.05	0.314	0.742
ln(District API)	6.664	0.092	6.301	6.874
HHI	0.267	0.236	0.018	1
Unemployment rate	5.679	2.722	0	30.4
Number of springs	68.805	72.389	0	270
ln(Median income)	10.953	0.37	9.964	12.26
Percent occupied houses	87.405	12.619	11.2	100
Percent owner occupied houses	64.997	13.686	6.7	97.4
ln(Rental rate)	6.908	0.319	5.451	8.008
ln(Population)	9.361	1.91	4.407	15.323
Percent population no high school	19.669	14.783	0	81
Percent population college or higher	25.101	16.741	0	83.3
Percent white population	77.010	16.621	17	100
Percent receiving assistance	9.018	9.425	0.286	85.714
Percent with at least one child	33.649	11.004	0	70

Number of observations = 913
(except for the ln(Rental rate) variable which has 897 observations)
Abbreviations: API is Academic Performance Index. HHI is Herfindahl-Hirschman Index.

Table 2: Baseline Estimation Results

Dependent variable: ln(Expenditure per pupil)	Model EX		Model EN	
Constant	6.957 ***	(1.156)	-21.227 **	(8.016)
ln(Enrollment)	-0.406 ***	(0.035)	-0.471 ***	(0.053)
ln(Enrollment)^2	0.023 ***	(0.002)	0.027 ***	(0.003)
Percent limited English proficiency	0.289 ***	(0.073)	0.495 ***	(0.111)
Percent Special Education students	2.283 ***	(0.227)	2.698 ***	(0.314)
Percent high school students	0.178 ***	(0.039)	0.604 ***	(0.134)
Percent low income students	-0.049	(0.065)	0.851 **	(0.272)
Price index	0.550 ***	(0.153)	0.730 ***	(0.205)
Wage index	0.768 **	(0.242)	1.454 ***	(0.372)
ln(District API)	0.422 *	(0.166)	4.543 ***	(1.179)
Dependent variable: $\ln(\sigma_u^2)$				
Constant	-6.328 ***	(1.132)	-5.881 ***	(0.663)
HHI	4.497 ***	(1.084)	4.905 ***	(0.698)
Dependent variable: $\ln(\sigma_v^2)$				
Constant	-2.686 ***	(0.056)		
Dependent variable: $\ln(\sigma_w^2)$				
Constant			-2.767 ***	(0.058)
η_1 (ln(District API))			-4.283 ***	(1.190)
η_2 (HHI)			-0.239 ***	(0.072)
η endogeneity test (F-Stat = 11.22399)			P > F = 0.00002	
Observations	913		913	
Log Likelihood	-101.17		1534.59	
Mean Cost Efficiency	0.9296		0.9061	
Median Cost Efficiency	0.9498		0.9353	

Notes: Standard errors are in parentheses. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels.
Abbreviations: API is the Academic Performance Index. HHI is the Herfindahl-Hirschman Index. Model EX is the exogenous model. Model EN is the endogenous model.
Endogenous Variables: ln(District API), HHI.
Instrumental Variables: Unemployment rate, Number of springs.

Table 3: Estimation Results with Different IV Sets

Dependent variable: ln(Expenditure per pupil)	Model EN IV Set 1		Model EN IV Set 2	
Constant	-14.108	(9.269)	-24.929	(16.25)
ln(Enrollment)	-0.448***	(0.051)	-0.488***	(0.073)
ln(Enrollment)^2	0.026***	(0.003)	0.028***	(0.004)
Percent limited English proficiency	0.448***	(0.109)	0.521***	(0.152)
Percent Special Education students	2.599***	(0.295)	2.725***	(0.375)
Percent high school students	0.493***	(0.149)	0.663*	(0.258)
Percent low income students	0.619*	(0.308)	0.974	(0.537)
Price index	0.686***	(0.188)	0.730**	(0.225)
Wage index	1.257***	(0.363)	1.538**	(0.516)
ln(District API)	3.499*	(1.362)	5.093*	(2.391)
Dependent variable: $\ln(\sigma_u^2)$				
Constant	-5.998***	(0.706)	-5.989***	(0.713)
HHI	4.974***	(0.735)	4.930***	(0.740)
Dependent variable: $\ln(\sigma_v^2)$				
Constant				
Dependent variable: $\ln(\sigma_w^2)$				
Constant	-2.737***	(0.058)	-2.736***	(0.058)
η_1 (ln(District API))	-3.138*	(1.370)	-4.720*	(2.396)
η_2 (HHI)	-0.193**	(0.073)	-0.192**	(0.073)
η endogeneity test (X^2-Stat = 10.06)	P > X^2 = 0.007			
η endogeneity test (X^2-Stat = 9.39)			P > X^2 = 0.009	
Observations	913		913	
Log Likelihood	1512.20		1507.26	
Mean Cost Efficiency	0.9096		0.9100	
Median Cost Efficiency	0.9388		0.9386	

Notes: Standard errors are in parentheses. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels.
Abbreviations: API is the Academic Performance Index. HHI is the Herfindahl-Hirschman Index. Model EX is the exogenous model. Model EN is the endogenous model.
Endogenous Variables: ln(District API), HHI.
Instrumental Variables: IV Set 1: Number of summits, Number of swamps
Instrumental Variables: IV Set 2: Number of lakes, Number of flats

Table 4: Estimation Results with Low (<%5) and High (>%95) Density Districts Excluded

Dependent variable: ln(Expenditure per pupil)	Model EX		Model EN	
Constant	5.636 ***	(1.342)	-20.927 **	(7.980)
ln(Enrollment)	-0.414 ***	(0.059)	-0.405 ***	(0.074)
ln(Enrollment)^2	0.024 ***	(0.004)	0.023 ***	(0.005)
Percent limited English proficiency	0.324 ***	(0.079)	0.498 ***	(0.107)
Percent Special Education students	2.522 ***	(0.251)	2.999 ***	(0.329)
Percent high school students	0.200 ***	(0.041)	0.587 ***	(0.128)
Percent low income students	-0.063	(0.074)	0.803 **	(0.278)
Price index	0.618 ***	(0.164)	0.642 **	(0.202)
Wage index	1.063 ***	(0.277)	1.758 ***	(0.409)
ln(District API)	0.600 **	(0.191)	4.449 ***	(1.163)
Dependent variable: $\ln(\sigma_u^2)$				
Constant	-6.720 ***	(1.660)	-5.939 ***	(0.767)
HHI	4.239 **	(1.519)	4.633 ***	(0.793)
Dependent variable: $\ln(\sigma_v^2)$				
Constant	-2.645 ***	(0.057)		
Dependent variable: $\ln(\sigma_w^2)$				
Constant			-2.722 ***	(0.061)
η_1 (ln(District API))			-4.009 ***	(1.176)
η_2 (HHI)			-0.222 **	(0.082)
η endogeneity test (F-Stat = 8.78161)			P > F = 0.00017	
Observations	823		823	
Log Likelihood	-94.72		1471.93	
Mean Cost Efficiency	0.9457		0.9152	
Median Cost Efficiency	0.9594		0.9388	

Notes: Standard errors are in parentheses. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels.
Abbreviations: API is the Academic Performance Index. HHI is the Herfindahl-Hirschman Index. Model EX is the exogenous model. Model EN is the endogenous model.
Endogenous Variables: ln(District API), HHI.
Instrumental Variables: Unemployment rate, Number of springs.

Table 5: Estimation Results with Endogenous Enrollment

Dependent variable: ln(Expenditure per pupil)	Model EX		Model EN	
Constant	6.957 ***	(1.156)	-9.288	(10.15)
ln(Enrollment)	-0.406 ***	(0.035)	-0.470***	(0.049)
ln(Enrollment)^2	0.023 ***	(0.002)	0.022***	(0.002)
Percent limited English proficiency	0.289 ***	(0.073)	0.750***	(0.173)
Percent Special Education students	2.283 ***	(0.227)	3.416***	(0.473)
Percent high school students	0.178 ***	(0.039)	0.608***	(0.119)
Percent low income students	-0.049	(0.065)	0.395	(0.365)
Price index	0.550 ***	(0.153)	0.725***	(0.191)
Wage index	0.768 **	(0.242)	0.755	(0.520)
ln(District API)	0.422 *	(0.166)	2.851*	(1.448)
Dependent variable: $\ln(\sigma_u^2)$				
Constant	-6.328 ***	(1.132)	-5.944***	(0.669)
HHI	4.497 ***	(1.084)	4.949***	(0.701)
Dependent variable: $\ln(\sigma_v^2)$				
Constant	-2.686 ***	(0.056)		
Dependent variable: $\ln(\sigma_w^2)$				
Constant			-2.776***	(0.058)
η_1 (ln(District API))			-2.618*	(1.330)
η_2 (HHI)			-0.246***	(0.071)
η_3 (ln(Enrollment))			0.084*	(0.037)
η endogeneity test (X^2-Stat = 32.7)			P > X^2 = 0.000	
Observations	913		913	
Log Likelihood	-101.17		-130.19	
Mean Cost Efficiency	0.9296		0.9080	
Median Cost Efficiency	0.9498		0.9374	

Notes: Standard errors are in parentheses. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels.
Abbreviations: API is the Academic Performance Index. HHI is the Herfindahl-Hirschman Index. Model EX is the exogenous model. Model EN is the endogenous model.
Endogenous Variables: ln(District API), HHI, ln(Enrollment), ln(Enrollment)^2.

Instrumental Variables: Unemployment rate, Number of springs, Number of basins.
See Footnote 28 about including only a single bias correction term for ln(Enrollment), and $\ln(\text{Enrollment})^2$.

Table 6: Estimation Results with Spatial Interactions and Variation in Cost Efficiency

Dependent variable: ln(Expenditure per pupil)	Extended Model EN-1		Extended Model EN-2	
Constant	-21.236 **	(7.790)	-16.709 *	(8.413)
ln(Enrollment)	-0.461 ***	(0.054)	-0.666 ***	(0.054)
$\ln(\text{Enrollment})^2$	0.027 ***	(0.003)	0.026 ***	(0.004)
Percent limited English proficiency	0.295 **	(0.113)	0.345 **	(0.134)
Percent Special Education students	2.466 ***	(0.328)	1.400 ***	(0.312)
Percent high school students	0.605 ***	(0.128)	0.376 **	(0.127)
Percent low income students	0.997 ***	(0.254)	0.834 ***	(0.210)
Price index	1.014 ***	(0.209)	0.922 ***	(0.171)
Wage index	1.514 ***	(0.373)	0.947 **	(0.355)
ln(Median income)			-0.055	(0.075)
Percent occupied houses			-0.006 ***	(0.001)
Percent owner occupied houses			0.002	(0.001)
ln(Rental rate)			0.070	(0.061)
ln(Population)			0.237 ***	(0.024)
Percent population no high school			0.003	(0.002)
Percent population college or higher			0.003	(0.002)
Percent white population			0.001	(0.001)
Percent receiving assistance			-0.006	(0.003)
Percent with at least one child			0.004 *	(0.002)
ln(District API)	4.500 ***	(1.147)	3.836 **	(1.266)
Dependent variable: $\ln(\sigma_u^2)$				
Constant	-26.482	(14.65)	-8.213	(11.06)
μ(ln(Median income)$_j$)	3.763 *	(1.883)	2.108	(1.423)
μ(Percent occupied houses$_j$)	-0.045	(0.028)	-0.027	(0.023)
μ(Percent owner occupied houses$_j$)	-0.060	(0.047)	-0.090 *	(0.035)
μ(ln(Rental rate)$_j$)	-4.090	(2.346)	-3.081	(1.810)
μ(ln(Population)$_j$)	0.429	(0.621)	-0.056	(0.456)
μ(Percent population no high school$_j$)	0.052	(0.073)	-0.047	(0.044)
μ(Percent population college or higher$_j$)	0.119 *	(0.048)	0.062	(0.033)
μ(Percent white population$_j$)	0.075 *	(0.030)	0.067 **	(0.024)
μ(Percent receiving assistance$_j$)	0.083	(0.119)	0.050	(0.088)
μ(Percent with at least one child$_j$)	0.038	(0.074)	0.120 *	(0.050)
HHI	5.141 ***	(0.615)	4.217 ***	(0.448)

Table 6 continues on the following page.

Dependent variable: $\ln(\sigma_w^2)$	Extended Model EN-1		Extended Model EN-2	
Constant	-2.976 ***	(0.060)	-3.704 ***	(0.077)
η_1 (ln(District API))	-4.415 ***	(1.162)	-3.970 **	(1.278)
η_2 (HHI)	-0.260 ***	(0.073)	-0.188 ***	(0.056)
η endogeneity test (F-Stat = 12.30331)	P > F = 0.00001			
η endogeneity test (F-Stat = 10.46985)			P > F = 0.00003	
Observations	895		895	
Log Likelihood	1575.847		1879.18	
Mean Cost Efficiency	0.8840		0.8663	
Median Cost Efficiency	0.9148		0.9022	

Notes: Standard errors are in parentheses. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels. μ(Variable $_j$) denotes the neighboring value which is the average of a variable in all districts that share a physical border with the observed district.
Abbreviations: API is the Academic Performance Index. HHI is the Herfindahl-Hirschman Index. Model EX is the exogenous model. Model EN is the endogenous model.
Endogenous Variables: ln(District API), HHI.
Instrumental Variables: Unemployment rate, Number of springs.

Table 7: Estimation Results with District Adequate Yearly Progress

Dependent variable: ln(Expenditure per pupil)	Model EX		Model EN	
Constant	9.617 ***	(0.186)	8.412 ***	(0.435)
ln(Enrollment)	-0.388 ***	(0.034)	-0.300 ***	(0.050)
ln(Enrollment)^2	0.023 ***	(0.002)	0.019 ***	(0.003)
Percent limited English proficiency	0.362 ***	(0.073)	0.801 ***	(0.177)
Percent Special Education students	2.229 ***	(0.222)	2.255 ***	(0.300)
Percent high school students	0.139 ***	(0.035)	0.143 **	(0.047)
Percent low income students	-0.093	(0.054)	0.083	(0.098)
Price index	0.521 ***	(0.150)	0.510 *	(0.205)
Wage index	0.697 **	(0.235)	0.682 *	(0.318)
District AYP	0.632 ***	(0.101)	3.358 ***	(0.945)
Dependent variable: $\ln(\sigma_u^2)$				
Constant	-6.614 ***	(1.350)	-6.063 ***	(0.741)
HHI	4.620 ***	(1.287)	4.890 ***	(0.755)
Dependent variable: $\ln(\sigma_v^2)$				
Constant	-2.713 ***	(0.056)		
Dependent variable: $\ln(\sigma_w^2)$				
Constant			-2.784 ***	(0.058)
η_1 (District AYP)			-2.784 **	(0.950)
η_2 (HHI)			-0.222 **	(0.070)
η endogeneity test (F-Stat = 8.8552)			P > F = 0.00016	
Observations	913		913	
Log Likelihood	-85.31		1128.60	
Mean Cost Efficiency	0.9370		0.9139	
Median Cost Efficiency	0.9559		0.9410	

Notes: Standard errors are in parentheses. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels.
Abbreviations: AYP is the Adequate Yearly Progress. HHI is the Herfindahl-Hirschman Index. Model EX is the exogenous model. Model EN is the endogenous model.
Endogenous Variables: ln(District AYP), HHI.
Instrumental Variables: Unemployment rate, Number of springs.

Table 8: Estimation Results with HHI based on 15-Mile Radius Circular Approach

Dependent variable: ln(Expenditure per pupil)	Model EX		Model EN	
Constant	7.516 ***	(1.114)	-19.254 *	(7.837)
ln(Enrollment)	-0.426 ***	(0.034)	-0.496 ***	(0.052)
ln(Enrollment)^2	0.024 ***	(0.002)	0.028 ***	(0.003)
Percent limited English proficiency	0.286 ***	(0.073)	0.478 ***	(0.108)
Percent Special Education students	2.240 ***	(0.225)	2.604 ***	(0.307)
Percent high school students	0.173 ***	(0.039)	0.580 ***	(0.131)
Percent low income students	-0.050	(0.064)	0.815 **	(0.266)
Price index	0.535 ***	(0.153)	0.705 ***	(0.203)
Wage index	0.701 **	(0.240)	1.311 ***	(0.361)
ln(District API)	0.357 *	(0.161)	4.279 ***	(1.153)
Dependent variable: $\ln(\sigma_u^2)$				
Constant	-6.249 ***	(0.887)	-5.770 ***	(0.579)
HHI-R15	5.276 ***	(1.020)	5.806 ***	(0.811)
Dependent variable: $\ln(\sigma_v^2)$				
Constant	-2.678 ***	(0.053)		
Dependent variable: $\ln(\sigma_w^2)$				
Constant			-2.750 ***	(0.056)
η_1 (ln(District API))			-4.089 ***	(1.166)
η_2 (HHI-R15)			-0.264 *	(0.106)
η endogeneity test (F-Stat = 8.62874)			P > F = 0.00019	
Observations	913		913	
Log Likelihood	-98.97		1730.784	
Mean Cost Efficiency	0.9341		0.9112	
Median Cost Efficiency	0.9494		0.9338	

Notes: Standard errors are in parentheses. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels.
Abbreviations: API is the Academic Performance Index. HHI-R15 is the Herfindahl-Hirschman Index based on 15-mile radius circular approach. Model EX is the exogenous model. Model EN is the endogenous model.
Endogenous Variables: ln(District API), HHI-R15.
Instrumental Variables: Unemployment rate, Number of springs.

Table 9: Estimation Results with HHI based on Metropolitan Areas as Education Markets

Dependent variable: ln(Expenditure per pupil)	Model EX		Model EN	
Constant	8.218 ***	(1.107)	-20.603 *	(10.48)
ln(Enrollment)	-0.474 ***	(0.033)	-0.580 ***	(0.057)
ln(Enrollment)^2	0.027 ***	(0.002)	0.033 ***	(0.004)
Percent limited English proficiency	0.266 ***	(0.074)	0.481 ***	(0.122)
Percent Special Education students	2.131 ***	(0.229)	2.457 ***	(0.328)
Percent high school students	0.184 ***	(0.040)	0.628 ***	(0.171)
Percent low income students	-0.044	(0.065)	0.894 *	(0.352)
Price index	0.475 **	(0.154)	0.592 **	(0.209)
Wage index	0.783 ***	(0.237)	1.477 ***	(0.404)
ln(District API)	0.290	(0.161)	4.533 **	(1.543)
Dependent variable: $\ln(\sigma_u^2)$				
Constant	-8.753 *	(3.745)	-6.807 ***	(1.204)
HHI-MA	7.094	(4.248)	7.147 ***	(1.709)
Dependent variable: $\ln(\sigma_v^2)$				
Constant	-2.591 ***	(0.048)		
Dependent variable: $\ln(\sigma_w^2)$				
Constant			-2.637 ***	(0.051)
η_1 (ln(District API))			-4.336 **	(1.551)
η_2 (HHI-MA)			-0.336 *	(0.152)
η endogeneity test (F-Stat = 6.36716)			P > F = 0.00180	
Observations	913		913	
Log Likelihood	-115.10		1960.53	
Mean Cost Efficiency	0.9808		0.9503	
Median Cost Efficiency	0.9851		0.9606	

Notes: Standard errors are in parentheses. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels.
Abbreviations: API is the Academic Performance Index. HHI-MA is the Herfindahl-Hirschman Index based on metropolitan areas as education markets. Model EX is the exogenous model. Model EN is the endogenous model.
Endogenous Variables: ln(District API), HHI-MA.
Instrumental Variables: Unemployment rate, Number of springs.

Table 10: Summary of the Differences between Model EXs and Model ENs

	Mean	Standard Deviation	Minimum	Maximum
Controlling for the endogeneity of concentration and achievement measures results in a(n):				
Increase in the coefficient of the concentration measure	0.44	0.23	0.05	0.79
Percent increase in the coefficient of the concentration measure	10.19	6.06	0.76	19.26
Increase in the coefficient of the achievement measure	3.66	0.69	2.44	4.24
Fold increase in the coefficient of the achievement measure	9.40	4.05	4.32	14.63
Predicted expenditures based on Model EXs and Model ENs:				
Predicted expenditure per pupil in Model EXs	$8,124.82	$165.17	$7,903.61	$8,376.53
Predicted expenditure per pupil in Model ENs	$8,256.61	$224.29	$8,025.81	$8,721.27
Predicted total expenditure of the average-sized district in Model EXs	$53.17 M	$1.08 M	$51.72 M	$54.82 M
Predicted total expenditure of the average-sized district in Model ENs	$54.03 M	$1.47 M	$52.52 M	$57.07 M
Predicted total expenditure of the state in Model EXs	$48.54 B	$0.99 B	$47.22 B	$50.05 B
Predicted total expenditure of the state in Model ENs	$49.33 B	$1.34 B	$47.95 B	$52.11 B

Table 10 continues on the following page.

	Mean	Standard Deviation	Minimum	Maximum
Increasing the average student achievement measure by 10% requires a(n):				
Increase in the expenditure per pupil in Model EXs	$284.92	$115.92	$109.80	$484.76
Increase in the expenditure per pupil in Model ENs	$3,598.87	$1,379.70	$574.97	$4,534.22
Fold increase in the increase in the expenditure per pupil when the endogeneity in Model EXs is handled	11.67	5.52	4.24	18.62
Increase in the total expenditure of the average-sized district in Model EXs	$1.86 M	$0.76 M	$0.72 M	$3.17 M
Increase in the total expenditure of the average-sized district in Model ENs	$23.55 M	$9.03 M	$3.76 M	$29.67 M
Increase in the total expenditure of the state in Model EXs	$1.70 B	$0.69 B	$0.66 B	$2.90 B
Increase in the total expenditure of the state in Model ENs	$21.50 B	$8.24 B	$3.44 B	$27.09 B
Increasing the average concentration measure by 0.1 results in a(n):				
Increase in the expenditure per pupil in Model EXs	$129.13	$42.18	$59.15	$196.07
Increase in the expenditure per pupil in Model ENs	$197.85	$36.58	$162.26	$256.07
Percent increase in the increase in the expenditure per pupil when the endogeneity in Model EXs is handled	63.79	46.35	30.60	174.32
Increase in the total expenditure of the average-sized district in Model EXs	$0.85 M	$0.28 M	$0.39 M	$1.28 M
Increase in the total expenditure of the average-sized district in Model ENs	$1.29 M	$0.24 M	$1.06 M	$1.68 M

Abbreviations: Model EX is the exogenous model. Model EN is the endogenous model.

Table 11: Descriptive Statistics Before and After Consolidation

Before Consolidation

Variable	Mean	Standard Deviation	Minimum	Maximum
ln(Enrollment)	7.269	1.959	2.079	13.416
Percent limited English proficiency	0.185	0.175	0	1
Percent Special Education students	0.087	0.045	0	0.375
Percent high school students	0.205	0.278	0	1
Percent low income students	0.514	0.262	0	0.995
Price Index	0.47	0.061	0.026	0.727
Wage Index	0.389	0.051	0.314	0.742
ln(District API)	6.664	0.091	6.301	6.874
HHI	0.275	0.244	0.018	1

Number of observations = 935
Abbreviations: API is the Academic Performance Index. HHI is the Herfindahl-Hirschman Index.

After Consolidation

Variable	Mean	Standard Deviation	Minimum	Maximum
ln(Enrollment)	7.630	1.620	4.606	13.416
Percent limited English proficiency	0.193	0.171	0	0.8
Percent Special Education students	0.089	0.04	0	0.299
Percent high school students	0.222	0.283	0	1
Percent low income students	0.511	0.26	0	0.995
Price Index	0.469	0.056	0.13	0.623
Wage Index	0.388	0.046	0.314	0.591
ln(District API)	6.666	0.087	6.332	6.874
HHI	0.249	0.214	0.019	1

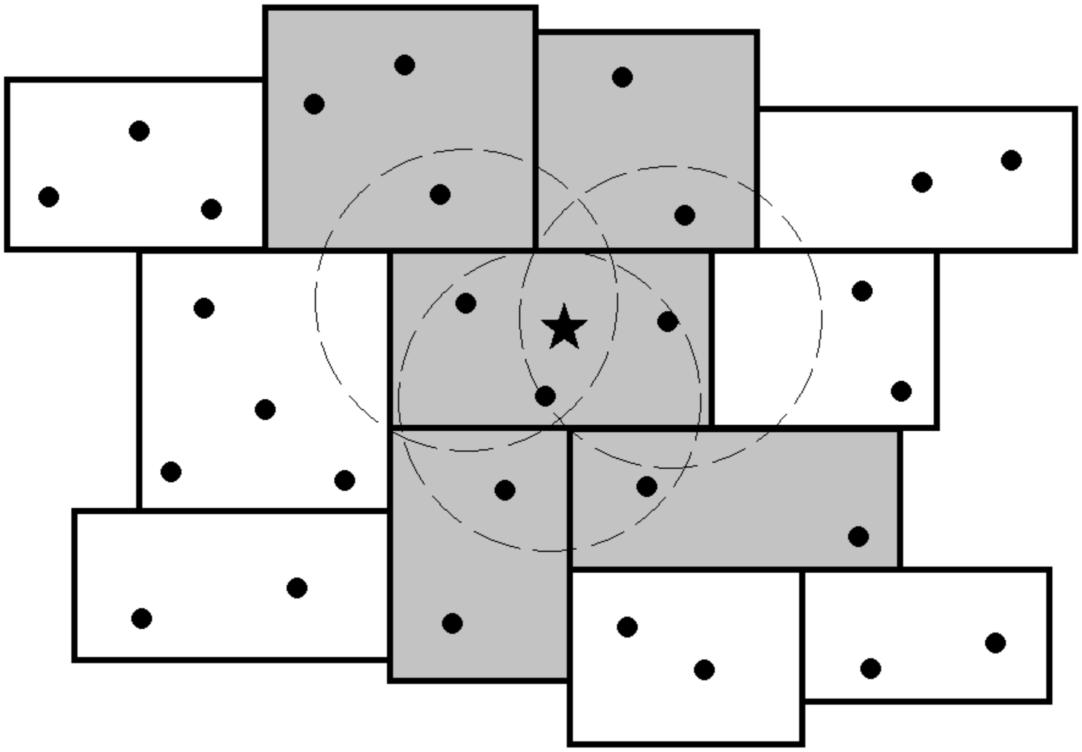
Number of observations = 857
Abbreviations: API is the Academic Performance Index. HHI is the Herfindahl-Hirschman Index.

Table 12: Predicted Outcomes of the Consolidation Simulation

	Model EX	Model EN
State's Predicted Expenditure Pre-Consolidation	$49,268,256,768	$52,939,563,008
State's Predicted Expenditure Post-Consolidation	$49,253,982,208	$52,950,093,824
Change in State's Predicted Expenditure	–$14,274,560	$10,530,816
State's Predicted Efficient Cost Pre-Consolidation	$47,095,889,920	$49,955,708,928
State's Predicted Efficient Cost Post-Consolidation	$47,061,114,880	$49,931,218,944
State's Predicted Inefficiency Pre-Consolidation	$2,172,366,848	$2,983,854,080
State's Predicted Inefficiency Post-Consolidation	$2,192,867,328	$3,018,874,880

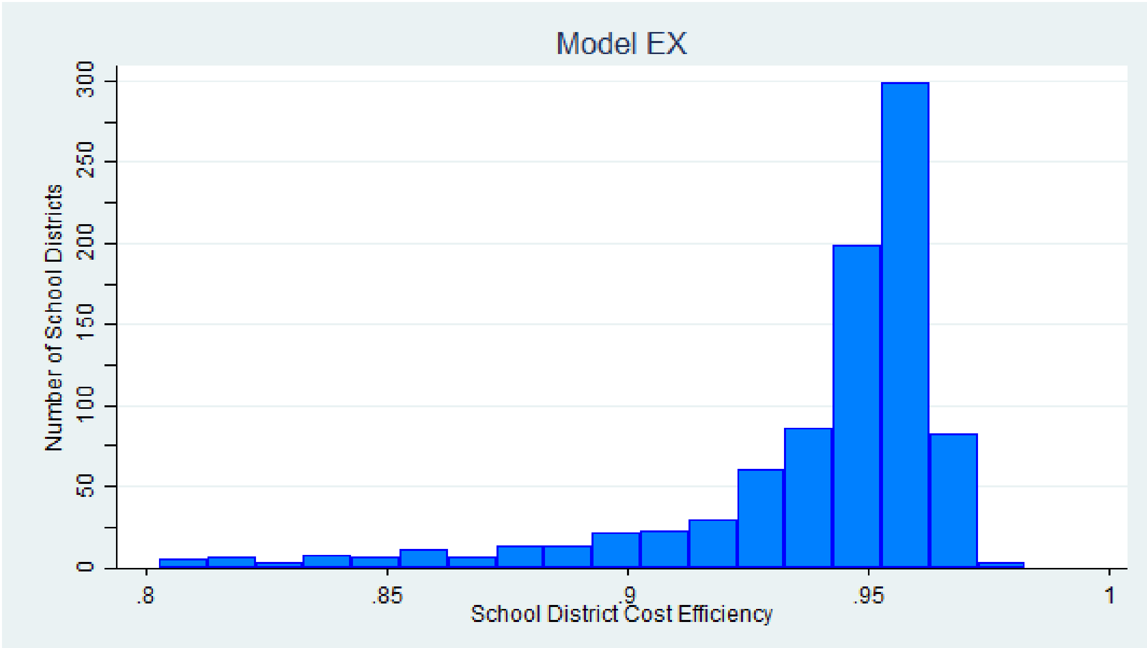
Abbreviations: Model EX is the exogenous model. Model EN is the endogenous model.

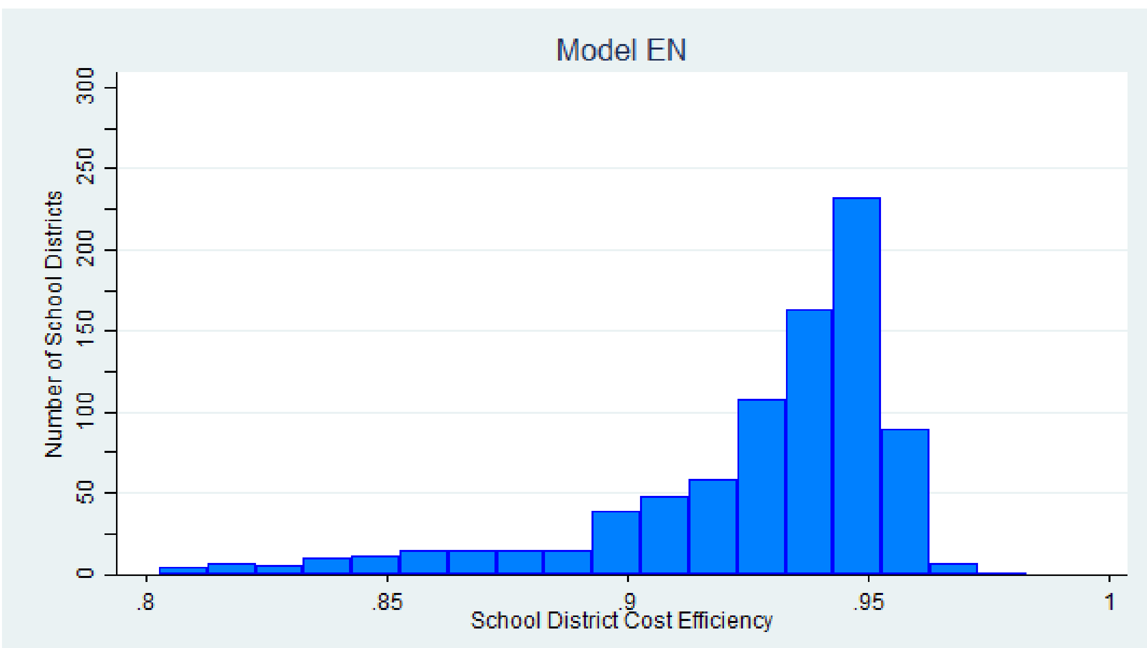
Figure 1: Determination of the Education Markets using the Radial Method



Notes: The rectangles represent the school districts, and the black dots represent the schools. There are three schools in the district with the star. When circles of a fixed radius are drawn around those schools, we see that four other districts surrounding the district with the star have schools in those circles. Those districts along with the district with the star are shaded in gray which represents the education market of the district with the star. It is important to notice that this approach allows for education market variation, that is, each school district would have a differently delineated education market. For example, the education market of the gray shaded district at the top right would not include the gray shaded districts at the bottom.

Figure 2: Cost Efficiencies of the School Districts





Notes: Each figure uses baseline estimation results in Table 2 and displays more than 92% of all school districts. The rest of the district cost efficiencies are less than 0.8 and the long and thin left tails extend to 0.32. These tails are omitted from the figures above for a better visual comparison. Abbreviations: Model EX is the exogenous model. Model EN is the endogenous model.

Appendix: Baseline Prediction Equation Estimates for Endogenous Variables

Dependent variable: ln(District API)			
Constant	6.7887 ***	(0.0341)	[198.82]
ln(Enrollment)	0.0253 ***	(0.0066)	[3.83]
ln(Enrollment)^2	-0.0013 **	(0.0004)	[-2.80]
Percent limited English proficiency	-0.0545 ***	(0.0150)	[-3.62]
Percent Special Education students	-0.1226 **	(0.0463)	[-2.65]
Percent high school students	-0.1070 ***	(0.0073)	[-14.67]
Percent low income students	-0.2068 ***	(0.0114)	[-18.14]
Price index	-0.0412	(0.0313)	[-1.32]
Wage index	-0.1137 *	(0.0486)	[-2.34]
Unemployment rate	-0.0031 ***	(0.0007)	[-4.51]
Number of springs	-0.0001 ***	(0.0000)	[-4.28]
Observations		913	

Notes: Standard errors are in parentheses. z-values are in brackets. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels. t-values are in brackets.
Abbreviations: API is the Academic Performance Index.

Dependent variable: HHI			
Constant	0.7466 ***	(0.1113)	[6.71]
ln(Enrollment)	-0.1194 ***	(0.0215)	[-5.56]
ln(Enrollment)^2	0.0035 *	(0.0015)	[2.42]
Percent limited English proficiency	-0.0061	(0.0493)	[-0.12]
Percent Special Education students	-0.2189	(0.1507)	[-1.45]
Percent high school students	0.0915 ***	(0.0238)	[3.85]
Percent low income students	0.1427 ***	(0.0372)	[3.84]
Price index	-0.0252	(0.1020)	[-0.25]
Wage index	0.4270 **	(0.1587)	[2.69]
Unemployment rate	-0.0120 ***	(0.0025)	[-4.88]
Number of springs	0.0005 ***	(0.0001)	[5.78]
Observations		913	

Notes: Standard errors are in parentheses. z-values are in brackets. Asterisks indicate significance at the 0.1% (***), 1% (**) and 5% (*) levels. t-values are in brackets.
Abbreviations: HHI is the Herfindahl-Hirschman Index.