1 Introduction

The United States' constitution does not directly protect the right to education, nor does it imply any such right[1]. This was the decision reached in San Antonio Independent School District v. Rodriguez (1973), a case which centred on the inequalities which natural result from a school financing system based on local property taxation[2]. In particular it centred around the vastly different tax bases in two neighbouring Texan public school districts and the effect this had on the resources available to students in each district. The decision has stood ever since.

Almost twenty years would pass until Texans would successfully sue the Texan Commisioner of Education for discrimination against students in poor districts[3] in Edgewood Independent School District v. Kirby (Tex 1989). During the trial it was found that the "[the] wealthiest district [have] over $14,000,000 of property wealth per student, while the poorest [have] approximately$ 20,000; this disparity reflects a 700 to 1 ratio"[4]. While as Americans the students in Edgewood ISD had no right to education, as Texans it was decided they did. The solution forged by the wrestle between parents, lawmakers and judges which followed was finance-equalisation legislation passed in 1993 and taken into effect in the 1993/94 school year. Though nicknamed 'Robin Hood' by the press, this belied the widespread ire the legislation would draw from economists[5], lawmakers[6] and parents[7] (rich and poor alike), and the two further rounds of lawsuits it would face. In January 2019 Republican Governor Greg Abbott tweeted: "We must put Robin Hood school funding on a path of extinction", later congratulating himself (more mutedly) for "reform" on the issue as the 86th Legislative Session ended[8].

Far from being a unique, arcane example, interesting only in its novelty, Texas' school finance system mirrors dozens[9] of other states which rely on similar local funding models; as do its problems. Moreover, the "byzantine"[10] nature of the system presents two intruiging settings to the endogeneity-minded econometrician, which to my knowledge are previously unexplored. Arbitrary restrictions on tax rates due to the unconstitutionality of statewide taxes can induce cash-starved districts to max out their tax rates, presenting windows in which their revenue is completely dependent on property values and state aid. Further to this, random flucuations in the oil price can provide positive and negative revenue shocks which may constitute a viable instrument for district revenue, allowing for a more plausible estimate of the causal impact of spending on school performance. These combined with the time-series, campus-level nature of the data allows for estimates which plausibly address some endogeneity concerns which plague other educational production function estimates. Many remain, and are addressed in due course.

This paper begins by describing the institutional structure of school funding and performance measurement in Texas, then proceeds to review the relevant theoretical and empirical literature, describe the original data and the sample selection, estimate production functions using both identification strategies in several different settings, presents and interprets the findings and then provides opportunities to extend the analysis.

2 Model

The ultimate aim

Our ultimate aim is to estimate a production function of the form:

$$Score = f(ability, resources, teacher, family, peers, community) + \epsilon$$

Once we have done this policy questions can be answered by investigating the parameters obtained from this estimate. For example how does the effect of class size differ in different communities, what proportion of the variation in expenditure estimates can be attributed to district-level effects?

[Obviously I'll need to put the general background, data and theory here too. (Importantly I need to talk about data selection and assumptions about sampling process. That is, that the data I have dropped were dropped randomly).]

3 Model

Consider a simple linear model for academic performance over campuses $i = 1, \ldots, n$; school districts $j = 1, \ldots, 957$; and periods $t = 2003, \ldots, 2011$ (where 2003 represents the academic year '02-'03):

$$\begin{aligned}
\text{Gr5.Avg}_{ijt} = \beta_0 &+ \beta_1 \text{Gr4.Avg}_{ijt-1} + \beta_2 \text{Gr3.Avg}_{ijt-2} \\
&+ \beta_3 \text{Per.Pupil.Exp}_{ijt} + \beta_4 \text{Econ.Disadv.Per}_{ijt} \\
&+ \beta_5 \text{T.Avg.Sal}_{ijt} + \beta_6 \text{T.Avg.Exp}_{ijt} \\
&+ \beta_7 \text{Gr5.Class.Size}_{ijt} + \epsilon_{ijt}
\end{aligned}$$

Before estimating this model it is instructive to restrict it to more specialised cases and clearly step through the assumptions required for OLS to be an efficient, unbiased estimation strategy. This makes suggesting alternative estimation strategies much more intuitive.

3.1 Within district, one time period

Let the first restriction be j = 1 (WLOG let this represent DALLAS ISD) and t = 2006. These are arbitrary parameter values. The model then can be simplified to:

$$
\begin{aligned}
\text{Gr5.Avg}_i = \beta_0 &+ \beta_1\text{Gr4.Avg.Lag1}_i + \beta_2\text{Gr3.Avg.Lag2}_i \\
&+ \beta_3\text{Per.Pupil.Exp}_i + \beta_4\text{Econ.Disadv.Per}_i \\
&+ \beta_5\text{T.Avg.Sal}_i + \beta_6\text{T.Avg.Exp}_i \\
&+ \beta_7\text{Gr5.Class.Size}_i + \epsilon_i
\end{aligned}
$$

In making these restrictions we can automatically rule out concerns due to time-trends and some spatial autocorrelation, however we must still be confident that the GM assumptions are satisfied.

3.1.1 Multicollinearity

We require that the matrix $\boldsymbol{X}'\boldsymbol{X}$ where

$$
\boldsymbol{X} = \begin{pmatrix} 1 & x_1' \\ 1 & x_2' \\ \vdots & \vdots \\ 1 & x_n' \end{pmatrix}
$$

,

$$x_i = \begin{pmatrix} \text{Gr4.Avg.Lag1}_i \\ \text{Gr3.Avg.Lag2}_i \\ \vdots \\ \text{Gr5.Class.Size}_i \end{pmatrix}$$

be invertible.

3.1.2 Exogeneity

Measurement error, simultaneity.

3.1.3 Homoskedasticity

Downward bias, local spatial autocorrelation.

3.1.4 Normality

??

3.2 Within district, panel

A simple linear model for performance in Dallas ISD:

$$\begin{aligned} \text{Gr5.Avg}_{it} = \beta_0 &+ \beta_1 \text{Gr4.Avg.Lag1}_{it} + \beta_2 \text{Gr3.Avg.Lag2}_{it} \\ &+ \beta_3 \text{Per.Pupil.Exp}_{it} + \beta_4 \text{Econ.Disadv.Per}_{it} \\ &+ \beta_5 \text{T.Avg.Sal}_{it} + \beta_6 \text{T.Avg.Exp}_{it} \\ &+ \beta_7 \text{Gr5.Class.Size}_{it} + \epsilon_{it} \end{aligned}$$

3.3 Across districts, one time period

A simple linear model for performance in 2006:

$$\begin{aligned}\text{Gr5.Avg}_{ij} = \beta_0 &+ \beta_1\text{Gr4.Avg.Lag1}_{ij} + \beta_2\text{Gr3.Avg.Lag2}_{ij} \\ &+ \beta_3\text{Per.Pupil.Exp}_{ij} + \beta_4\text{Econ.Disadv.Per}_{ij} \\ &+ \beta_5\text{T.Avg.Sal}_{ij} + \beta_6\text{T.Avg.Exp}_{ij} \\ &+ \beta_7\text{Gr5.Class.Size}_{ij} + \epsilon_{ij}\end{aligned}$$

3.4 Across districts, panel

This linear model utilises all our data.

$$\begin{aligned}\text{Gr5.Avg}_{ijt} = \beta_0 &+ \beta_1\text{Gr4.Avg}_{ijt-1} + \beta_2\text{Gr3.Avg}_{ijt-2} \\ &+ \beta_3\text{Per.Pupil.Exp}_{ijt} + \beta_4\text{Econ.Disadv.Per}_{ijt} \\ &+ \beta_5\text{T.Avg.Sal}_{ijt} + \beta_6\text{T.Avg.Exp}_{ijt} \\ &+ \beta_7\text{Gr5.Class.Size}_{ijt} + \epsilon_{ijt}\end{aligned}$$

where $\epsilon_{ijt} = \alpha_i + \mu_j + \upsilon_{ijt}$ and

$$\upsilon_{ijt} \sim \mathcal{N}(0, \sigma^2) \quad \text{for all } i, j, t.$$

That is our errors are composed of time-invariant, campus-level and district-level heterogeneity and an idiosyncratic error.

3.5 Other Concerns

3.5.1 Hierarchical Error Structure

3.5.2 Parameter Heterogeneity

4 Estimation Strategy

Fixed effects is likely not to be appropriate in the presence of time trends.

4.1 Non-Linear Models

GMM estimation.

4.2 Dynamic Panel Model

Small T (9), large N (957).