I N S T I T U T D E S T A T I S T I Q U E

B I O S T A T I S T I Q U E E T

S C I E N C E S A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2013/13

Schooling inputs, property tax caps and effciency scores in public schools

HENDERSON, D., SIMAR, L. and L. WANG

Schooling inputs, property tax caps and efficiency scores in public schools*

Daniel J. Henderson†
Department of Economics, Finance and Legal Studies
University of Alabama

Léopold Simar‡
Institute of Statistics
Université catholique de Louvain

Le Wang§
Department of Economics
University of New Hampshire

February 22, 2013

*The authors would like to thank Subal Kumbhakar, Christopher Parmeter, Daniel Millimet, Sebastien Van Bellegem and Arnold Zellner for useful comments on the subject matter of this paper as well as participants of the 9th European Workshop on Efficiency and Productivity Analysis, Brussels, Belgium (June 2005) and the first annual New York Camp Econometrics, Saratoga Springs, NY (February 2006). The research on this project was partially completed during two separate visits by Henderson to the Institute of Statistics (June 2004, 2005) and he thanks both the Université catholique de Louvain and Binghamton University for financial support. Simar wishes to acknowledge research support from the IAP iResearch Network P7/06 of the Belgian State (Belgian Science Policy). This paper previously circulated under the title "A Fully Nonparametric Stochastic Frontier Model for Panel Data".

†Daniel J. Henderson, Department of Economics, Finance and Legal Studies, Box 870224, University of Alabama, Tuscaloosa, AL 35487-0224, U.S.A., 1-205-348-8991, Fax: 1-205-348-0590, e-mail: djhender@cba.ua.edu.

‡Léopold Simar, Institute of Statistics, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium, +32 10 47.43.08, Fax: +32 10 47.30.32, e-mail: leopold.simar@uclouvain.be.

§Le Wang, Department of Economics, University of New Hampshire, Durham, NH 03824, U.S.A., 1-603-826-0818, Fax: 1-603-862-3383, e-mail: le.wang@unh.edu.

Schooling inputs, property tax caps and efficiency scores in public schools

Abstract

Using a data set comprised of public school districts in the state of Illinois, we examine the role of traditional "schooling inputs" and "non-schooling inputs" in the determination of test scores. Using recently developed nonparametric kernel methods, we are able to determine the relevance (or irrelevance) of many commonly used inputs in an educational production function. We find that most commonly used 'schooling inputs' (e.g., capital per pupil expenditure) are irrelevant in determining student test scores in our sample, even in a very general setting. For the relevant inputs, our findings show that heterogeneous returns (which are allowed by our approach) are able to reconcile some of the insignificant or modest impacts of (non-schooling) inputs typically found in the literature. We further extend the existing analysis by developing an efficiency estimator which allows for discrete inputs. Previous attempts in the literature have either used partially linear models, treated discrete inputs as continuous, ran separate regressions for each value of the discrete regressor or excluded them altogether. Our estimators (which use discrete kernels to incorporate discrete inputs) of the educational production function and technical efficiency scores are consistent under standard assumptions and regularity conditions. In our application, we find that the standard parametric estimators drastically underestimate the level of efficiency. This is not only important from an econometric standpoint, but it may also be important for policy. If we believe that schools are underperforming, we may unnecessarily penalize schools, teachers and/or students. We fail to find any evidence that school districts facing property tax caps operate more efficiently so as to offset the negative effects of lack of resources on students' performance.

Keywords: Technical Efficiency, Nonparametric Kernel, Panel Data, School Inputs, Student Achievement

JEL Codes: C14, C23, I21

1 Introduction

The literature on the estimation of educational efficiency is typically based on the educational production approach, which entails measuring educational outputs and exploration of the underlying determinants/"inputs" of these outputs. The vast majority of papers in this area have focused on spending related "inputs" such as class size and teachers' credentials (e.g., Angrist and Lavy, 1999; Hoxby, 1999; Clofelter et al., 2007). Although the previous literature provides careful and important evidence on the effects of educational inputs, there are at least two gaps remaining. First, there may be heterogeneity in the returns to educational inputs. There is no reason to believe that the return to each input is constant across schools or within each state. Second, the existing educational production function literature relies mostly on parametric regression models (e.g., Millimet and Collier, 2008). Although popular, parametric models require stringent assumptions. In particular, the errors are generally assumed to come from a specified distribution and the functional form of the educational production function is given *a priori*. Since there is little reason to believe that the relationship between each educational input and test scores are linear and separable, a parametric specification which fully captures the true relation may be difficult to find. Further, if the functional form assumption does not hold, the parametric model will most likely lead to biased and inconsistent estimates of the return to each input and of efficiency scores themselves.

In light of these some of these potential shortcomings, a nonparametric approach could be useful (e.g., Duncombe et al., 1997; Grosskopf et al., 2001). Nonparametric estimation procedures relax the functional form assumptions associated with traditional parametric regression models and create a tighter fitting regression curve through the data. They also allow for potentially complicated interactions among the inputs of the model. These procedures do not require restrictive assumptions on the distribution of the error; nor do they require specific assumptions on the form of the underlying production function. Furthermore, the procedures generate unique estimates for each observation for each variable. This attribute enables us to make inference regarding heterogeneity in the returns.

While nonparametric methods are useful, nonparametric estimators typically assume that all variables are continuous in nature (e.g., Fan et al., 1996; Park et al., 1998, 2003). Educational production functions are filled with variables which are discrete in nature (e.g., the number of administrators). There are cases where this may be relatively harmless (e.g., including a time trend), but even in such cases, the curse of dimensionality is present. Recent research (e.g., Racine and Li, 2004) has shown that only the number of continuous regressors impacts the rate of convergence of nonparametric estimators. In other words, additional discrete variables do not add to the curse of dimensionality. Our suggestion is not simply about semantics, it leads to asymptotic improvements.

Here we propose using discrete kernels to incorporate discrete inputs into educational production functions. Our approach will both allow for "appropriate" classification of variables (without having to resort to ad hoc specifications), while at the same time lessens the curse of dimensionality problem so commonly referred to in the nonparametric realm. Our estimators are consistent under standard assumptions and regularity conditions.

By using automated bandwidth selection criteria that could effectively evaluate the relevance of variables in nonparametric settings, the results of our exercise show that many of the commonly used "inputs" in educational production functions are irrelevant in our sample. In fact, 9 of the 16 inputs we include, are shown to have no effect on the prediction of test scores. Further, this includes *all* of the commonly used "schooling inputs" such as number of schools in the district, current per pupil expenditure and capital per pupil expenditure. Because we do not impose any functional form restrictions and allow for heterogeneity in returns in assessing the relevance of these variables, these results are even stronger than that of the growing literature that suggests that schooling inputs have limited success in improving test scores (e.g., Eherenberg and Brewer, 1994; Hanushek, 2003).

Our examination of the partial effects of the educational production function find, as expected, heterogeneity across the sample. For example, we find that the unemployment rate typically has a negative and significant impact on test scores, but the degree to which varies. We also find that property tax caps have negative impacts on test scores on average, but find that the impact is often insignificant and sometimes is even positive.

Even though we find some evidence of negative impacts on test scores, we are unable to find any differences in efficiency scores between schools which are financially constrained by property tax cap laws and those who are not. In technical terms, what this is saying is that even though the educational production function shifts downwards, the school districts are no closer to the frontier than they were previously. This disputes some claims which assume that schools can or will become more efficient with these constraints. By contrast, our findings may suggest some type of strategic behavior amongst schools that try to argue for more funding by letting students academic performances drop, instead of operating more efficiently (Figlio and O'Sullivan, 2001).

With respect to the efficiency scores in general, we find that the parametric specifications drastically underestimate the efficiency scores of schools. The nonparametric estimates paint a brighter picture with respect to how the schools are performing given their "inputs". It also shows the harm in misspecification. If we solely relied on the parametric model(s), we would falsely conclude that schools are drastically underperforming. This could lead to potentially unnecessary and costly reforms to schools which may not lead to eventual gains in test scores.

The remainder of the paper is organized as follows: Section 2 discusses standard parametric estimation of an educational production function and determination of efficiency scores while Section 3 discusses an existing nonparametric stochastic frontier estimator. The fourth section outlines our extension of that estimator which allows for discrete inputs and gives its asymptotic properties. Section 5 presents the empirical results from our study and the final section concludes.

2 Parametric estimation

In this section we describe parametric methods used for estimating educational production functions and technical efficiency in a panel data setting.[1] In these models the production functions are estimated, and time invariant estimates of output oriented technical efficiency for each firm are obtained as a by-product of the exercise. This basic framework assumes that we observe a cross-section of data on N firms over T time periods. Quantities of d inputs are used to produce a scalar output through a production function. More specifically, the production frontier model can be written as

$$y_{it} = f(x_{it}, \beta) + \varepsilon_{it} - u_i,$$

where y_{it} represents the level of output for firm i at time period t ($i = 1, 2, \dots, N, t = 1, 2, \dots, T$), f is the production function, β is a vector of unknown parameters, ε_{it} represents the two sided noise component, and u_i is the non-negative technical inefficiency component of the error term.

Although many methods exist for estimating a parametric production frontier model with panel data, here we choose one of the most popular methods and one that is comparable to our approach. Fixed Effects (FE) estimation of the production frontier, introduced by Schmidt and Sickles (1984), can be obtained, for example, from the log-linear Cobb-Douglas one-way error component model

$$y_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it}, \tag{1}$$

[1]In our applicaiton the unit of observation is a school district for which we have (average) test scores (our repeated measure) for different subjects in different grades in two time periods. However, we will refer to firms as our unit of observation and time as our repeated measure in our theoretical discussion to stay consistent with the literature and to emphasize the broad applicability of the method discussed here.

where α_i $(= \alpha - u_i)$ is the firm fixed effect. In other words, we assume that each firm shares the same parametric technology in each time period, but that differences between them are captured by a location (firm) effect α_i. Estimation of β can be obtained, for example, by means of the within estimator. Firm specific estimates of α are then obtained by

$$\widehat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - x_{it}\widehat{\beta} \right).$$

Here we note that the estimate of β is consistent as $NT \to \infty$, but $\widehat{\alpha}_i$ is consistent only if $T \to \infty$. Further, we estimate the individual u_i by means of the normalization

$$\widehat{u}_i = \max_i \widehat{\alpha}_i - \widehat{\alpha}_i,$$

and if the left-hand-side variable is measured in logs, firm specific estimates of technical efficiency are given by

$$\widehat{TE}_i = \exp\left(-\widehat{u}_i\right).$$

3 Nonparametric estimation

The (linear) parametric assumption in (1) may not be suitable for all panel data sets. If a researcher assumes a linear specification and the Data Generating Process (DGP) is non-linear, then the estimates will most likely be biased and inconsistent. To counter situations such as this, Kneip and Simar (1996) suggest a more general form for the production function:

$$y_{it} = g(x_{it}) + \alpha_i + \varepsilon_{it},$$

where $g(\cdot)$ is an unknown smooth production function that each firm shares, but differences between them are captured by the location effect α_i. Identification of the individual effect requires the "average" of α_i to be equal to 0.

Estimation of $g(\cdot)$ can be obtained several ways. Kneip and Simar (1996) propose using a local-constant (Nadaraya-Watson) type estimator, but here we suggest using the Local-Linear Least-Squares (LLLS) estimator. Not only does the local-linear estimator give more reliable estimates of $g(\cdot)$, it also allows for estimation of both the production function and its gradient in one step. Estimation of $\delta(x) \equiv (g(x), \beta'(x))'$, where $\beta(x) \equiv \bigtriangledown g(x)$, is obtained as

$$\begin{aligned}\widehat{\delta}(x) &\equiv \left(\widehat{g}(x), \widehat{\beta}'(x)\right)' \\ &= \left(X'K(x)X\right)^{-1} X'K(x)y,\end{aligned}$$

where y is a $NT \times 1$ vector with the itth row being y_{it}, X is a $NT \times (d+1)$ matrix with itth row being $(1, (x_{it} - x))$ and $K(x)$ is a $NT \times NT$ diagonal matrix of (Gaussian) kernel weighting functions with bandwidth vector $h = (h_1, h_2, \ldots, h_d)$.

Estimation of the bandwidths can be obtained by using the Least-Squares Cross-Validation (LSCV) procedure. In short, the procedure chooses h such that it minimizes the LSCV function given by

$$CV(h) = \sum_{t=1}^{T} \sum_{i=1}^{N} [y_{it} - \widehat{g}_{-i}(x_{it})]^2,$$

where $\widehat{g}_{-i}(x_{it})$ is computed by leaving out all the T observations (y_{it}, x_{it}) of the ith firm (leave-one firm-out estimator).

Having determined an estimator $\widehat{g}(\cdot)$ of $g(\cdot)$, estimators of α for each i are also obtained by the method of least squares as

$$\widehat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - \widehat{g}(x_{it})\right).$$

Under the regularity assumptions of Kneip and Simar (1996), as $N \to \infty$

$$|\widehat{g}_i(x) - g_i(x)| = O_p\left((NT)^{-2/(d+4)} + N^{-(1/2)}\right),$$

where $g_i(x) = g(x) + \alpha_i$. Further, under the same regularity conditions, when N and $T \to \infty$

$$|\widehat{\alpha}_i - \alpha_i| = O_p\left((NT)^{-2/(d+4)} + T^{-(1/2)} + N^{-(1/2)}\right).$$

Again, we estimate the individual u_i by means of the normalization

$$\widehat{u}_i = \max_i \widehat{\alpha}_i - \widehat{\alpha}_i,$$

and if the dependent variable is measured in logs, firm specific estimates of technical efficiency are given by

$$\widehat{TE}_i = \exp\left(-\widehat{u}_i\right).$$

4 Discrete inputs

Although the nonparametric approach above was a step in the right direction, it assumes that all inputs are continuous in nature (also see Park and Simar, 1994; Sickles, 2005). It is common in economic research for inputs to be discrete (e.g., number of schools in a district). Previous attempts in the literature have either used partially linear models, treated the discrete inputs as continuous, ran separate regressions for each value of the discrete regressor or excluded them altogether. Often times this is harmless. For example, assuming that the number of hospital beds is a continuous variable would not worry most researchers. However, sometimes the nature of the categories are too important to ignore. For example, assume we are studying the efficiency of power plants. Given that there are different types of ways to generate power (coal, nuclear, etc.), when using nonparametric methods, authors are often forced to resort to approaches which are less than ideal. Here we circumvent these problems by introducing an estimator of the production function which allows for inputs that are discrete.

Here we propose using discrete kernels to incorporate discrete inputs correctly. A further benefit of this approach is that the rate of convergence of the production function will depend only on the number of continuous inputs and hence, properly treating the discrete inputs as discrete will improve the asymptotic results. This is especially important in a nonparametric model.

4.1 Estimation

In order to introduce the estimator, recall the estimator of Kneip and Simar (1996)

$$y_{it} = g(x_{it}) + \alpha_i + \varepsilon_{it}, \tag{2}$$

where $g(\cdot)$ is an unknown smooth production function that each firm shares, but differences between them are captured by the location effect α_i. The major difference between our estimator here and theirs is that we allow the vector x_{it} to be composed of three different types of regressors: continuous, unordered discrete and ordered discrete variables. Formally, $g(\cdot)$ is the unknown smooth function (now) with argument $x_{it} = [x_{it}^c, x_{it}^u, x_{it}^o]$. x_{it}^c is a vector of continuous regressors, x_{it}^u is a vector of regressors that assume unordered discrete values, and x_{it}^o is a vector of regressors that assume ordered discrete values. We now define d as the number of continuous inputs. The number of unordered and ordered discrete inputs are defined as d_u and d_o, respectively.

Taking a first-order Taylor expansion of (2) with respect to x^c yields

$$y_{it} \approx g(x) + (x_{it}^c - x^c)\beta(x) + \alpha_i + \varepsilon_{it}$$

where $\beta(x)$ is defined as the partial derivative of $g(x)$ with respect to x^c. Note that the Taylor expansion can only be taken with respect to the continuous inputs.

The LLLS estimator of $\delta(x) \equiv [g(x)\ \beta'(x)]'$ is given by

$$\begin{aligned}\widehat{\delta}(x) &= \left(\widehat{g}(x), \widehat{\beta}'(x)\right)' \\ &= \left(X'K\left(x\right)X\right)^{-1} X'K\left(x\right)y\end{aligned}$$

where X is a $NT \times (d+1)$ matrix with itth row being $(1, (x_{it}^c - x^c))$ and $K(x)$ is a $NT \times NT$ diagonal matrix of kernel weighting functions for mixed continuous and discrete data with bandwidth vector $(h,\lambda^u, \lambda^o) = (h_1, h_2, \ldots, h_d, \lambda_1^u, \ldots, \lambda_{d_u}^u, \lambda_1^o, \ldots, \lambda_{d_o}^o)$ for the continuous variables and discrete variables, respectively.

According to the properties established for nonparametric estimation of regression with both categorical and continuous regressors, our nonparametric estimator of $\widehat{g}(x_{it})$ achieves the standard nonparametric rate of convergence for mixed data types $\left(N^{2/(d+4)}\right)$ when $N \rightarrow \infty$ (remember that T is fixed), under the appropriate regularity conditions and technical assumptions on the bandwidths, as described in (the local-linear case) Li and Racine (2004) and (the local-constant case) Racine and Li (2004). Under these conditions, we can write, as $N \rightarrow \infty$

$$(NT)^{2/(d+4)}\left(\widehat{g}(x) - g(x) - b(x)\right) \sim \mathcal{N}(0, \sigma^2(x)),$$

where $b(x) = O((NT)^{-2/(d+4)})$ is the bias term. The order of the bias and the rate of convergence correspond to optimal choice of the bandwidths. As shown in the same references, the same result holds for bandwidths determined by cross-validation, of course, then $b(x)$ is replaced by $\widehat{b}(x)$ with $\widehat{b}(x) = O_p((NT)^{-2/(d+4)})$. From this result, we obtain:

$$|\widehat{g}(x) - g(x)| = O_p\left((NT)^{-2/(d+4)}\right), \tag{3}$$

where of course the last term is in fact $O_p(N^{-2/(d+4)})$, since T is fixed, but the writing in (3) exposes the fact that the NT observations are used in the estimation.

4.2 Bandwidths and relevance of variables

Estimation of the bandwidths can be obtained by using the LSCV procedure. In short, the procedure chooses bandwidths (h,λ^u, λ^o) such that they minimizes the LSCV function given by

$$CV(h,\lambda^u,\lambda^o) = \sum_{t=1}^{T}\sum_{i=1}^{N}[y_{it} - \widehat{g}_{-i}(x_{it})]^2,$$

where $\widehat{g}_{-i}(x_{it})$ is computed by leaving out, as in Kneip and Simar (1996), all the T observations (y_{it}, x_{it}) of the ith firm (leave-one firm-out estimator).

A useful feature of the LSCV procedure is its ability to determine the relevance of the variables of interest (Hall, Li and Racine, 2007). For continuous regressors, in the local-constant case, a bandwidth equal to the upper bound implies that the variable is irrelevant. In the local-linear case, a bandwidth equal to the upper bound determines that the (continuous) variable enters in linearly. The upper bound for the bandwidth on a continuous regressor in either case is infinity. This is impossible to observe in practice. However, when using a Gaussian kernel function, any bandwidth in excess of a few standard deviations of the regressor gives essentially equal weight to all observations. In other words, in the local-constant setting, the local average with respect to that variable is actually a global average of the left-hand-side variable and hence the regressor (essentially) has no impact on the conditional mean. In the local-linear setting, all observations are given equal weight and hence the regressor enters the model (essentially) in a linear fashion. Thus, we follow the suggestion of Hall, Li and Racine (2007) and use two standard deviations of the regressor as the bound for relevance/linearity. Thus, if any bandwidth on a continuous regressor exceeds two standard deviations of its associated variable, we conclude that it enters in an irrelevant fashion (in the local-constant setting) or linearly (in the local-linear setting).

For the discrete variables, the bandwidths, either for local-constant or local-linear, indicate which variables are relevant, as well as the extent of smoothing in the estimation. From the definitions for the ordered and unordered kernels, it follows that if the bandwidth for a particular unordered or ordered discrete variable equals zero, then the kernel reduces to an indicator function and no weight is given to observations for which $x_{it}^o \neq x^o$ or $x_{it}^u \neq x^u$. On the other hand, if the bandwidth for a particular unordered or ordered discrete variable reaches its upper bound, then equal weight is given to observations with $x_{it}^o = x^o$ and $x_{it}^o \neq x^o$. In this case, the variable is completely smoothed out (and thus does not impact the estimation results). For both unordered discrete variables, the upper bound is $(c-1)/c$, where c is the number of distinct values the discrete regressor takes. For ordered discrete variables, the upper bound is unity. See Hall, Li and Racine (2007) for further details.

In our application we will first use the local-constant estimator with LSCV bandwidths to determine relevance of our inputs. We will then remove irrelevant inputs and use a local-linear regression to examine the partial effects as well as determine our efficiency scores in light of the theoretical findings on local-constant versus local-linear estimation.

4.3 Efficiency scores

As we saw before, having determined an estimator $\widehat{g}(\cdot)$ of $g(\cdot)$, estimators of α for each i are again obtained by the method of least squares as

$$\widehat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \widehat{g}(x_{it})\right).$$

As we noted previously, the rate of convergence of the unknown function depends only upon the number of continuous regressors. Therefore, all the asymptotics follow from Kneip and Simar (1996), but we note that now d will be smaller if any of the inputs are treated as discrete. Again, we estimate the individual u_i by means of the normalization

$$\widehat{u}_i = \max_i \widehat{\alpha}_i - \widehat{\alpha}_i,$$

and if the dependent variable is measured in logs, firm specific estimates of technical efficiency are given by

$$\widehat{TE}_i = \exp\left(-\widehat{u}_i\right).$$

5 An educational production function

The lack of substantial gains in test scores which have been accompanied by huge increases in expenditures over the last four decades have caused many to wonder how well our schools are using these resources. This has caused many authors to examine the efficiency of schools (e.g., Collier and Millimet, 2009; Ruggiero et al., 2002). While there is a large body of research, many of these papers use simplistic models which fail to capture the complexity of an educational production function. Given that there is no consensus on the functional form of the educational production function, an assumption of a linearly separable production function may be a bit naive. Here we consider a nonparametric educational production function which incorporates discrete inputs and allows them to interact amongst themselves and with the continuous inputs in an unknown way. Not only do we expect to uncover heterogeneity in the partial effects, but we also expect to obtain more accurate efficiency scores.

5.1 Data

The data are obtained from Millimet and Collier (2008), and we provide limited details here. The data include public school districts in the state of Illinois. As noted in Hanushek (2007), policies are usually set at the state level, and focusing on schools within a state would allow us to compare schools that "operate within the same basic policy environment" and thus to eliminate differences in unmeasured policies (see,e.g., Ruggiero, 1996). The outcome of interest is test scores in reading and mathematics administered by the Illinois State Board of Education annually in grades three, six, eight and ten. These eight test scores are available for the 1996-1997 and 1997-1998 academic years. Following Millimet and Collier (2008), the districts with missing information in either year are excluded, and the resulting data are a balanced-panel sample with 850 school districts.

We utilize the following inputs at the district level in the estimation of the educational production function, and these inputs could be loosely categorized into two groups: school v.s. non-school inputs. School inputs include (1) Resources (current per pupil expenditure, capital per pupil expenditure, full-time equivalent (FTE) teachers); (2) District Organization (the number of schools, Local Education Agency (LEA) administrators). Non-school inputs include district-level characteristics and demographic information (total population, median household income, persons aged 20+ without a high school diploma, persons aged 20+ with a college education, unemployment rate, percentages of children aged 4-19 whose primary language is something other than English, percentages of occupied housing, and percentages of owner-occupied housing. The number of FTE teachers, schools, and LEA administrators are categorical variables in nature.

We also include a measure of fiscal constraints - whether the Property Tax Extension Limitation Law (PTELL), commonly called "tax caps", is implemented; the law places a cap on the *growth* of property tax revenue, which regulates that the growth of tax revenue cannot exceed the growth of the inflation rate. Unlike other levels of local government, school districts do not have any other sources of revenue, and hence they are more dependent on property tax revenue. Schools under PTELL counties are thus more financially constrained. Millimet and Rangaprasad (2007) and Millimet and Collier (2008) show that this financial constraint could play an important role in analysis of an educational production function and efficiency. The measure used is a binary variable equal to 1 if PTELL is implemented and zero otherwise. In addition, we also include test ID to indicate the type of test scores (an unordered discrete variable taking values 1-8 corresponding to a particular subject and grade level) and a variable which indicates the year the test was taken (an ordered discrete variable).

5.2 Estimation results

We first examine the bandwidths from the cross-validation procedure. As mentioned above, recent advances in nonparametric regression have shown that data driven bandwidth selection procedures can shed light onto the role of the regressors in the prediction of the dependent variable. Second, we will focus on the partial effects of the educational production function. Our estimators allow for the effects to vary across the population and we find ample evidence that they do. We then turn our attention to the efficiency scores.

5.2.1 Bandwidths and relevance of variables

Table (1) gives the bandwidth for each of the regressors in our study.[2] First, we find that nearly all non-school inputs (e.g., median household income and unemployment rate) are important determinants of students' academic achievements. This result is not surprising. Non-school inputs such as families and peer effects have been found to be important since the publication of the Coleman (1966) report (Harris, 2010), and the importance of these inputs, especially family background, is evident in more recent studies (Rothstein, 2004).

Our results with regard to the (percentage) educated population and income are consistent with prior findings that socioeconomic status is strongly associated with educational outcome (e.g., Vigdor and Ludwig, 2010). However, these results seem suggestive that among the components of socioeconomic status, education may matter more than income. Specifically, we find that median household income in the school district is irrelevant, while education, measured by the percentage of the population with bachelors degree, matters for student achievement. We also find PTELL to be relevant. This policy variable is highly debated and its impact on test scores will be examined more later.

In stark contrast to non-school inputs, *none* of the school inputs are relevant. While these results are qualitatively in agreement with the literature, our results are much stronger. Recall that in parametric models, relevance of a variable is usually tested by statistical significance of the single coefficient on that variable (obtained, say via OLS). This coefficient represents the average effect of this variable. In the presence of heterogeneous effects where both negative and positive effects exist across districts, it is possible that these opposing effects could offset each other and the average effect becomes zero. This could result in a statistically insignificant coefficient and in turn lead one to falsely conclude that this variable is irrelevant. However, this is not the case here since we do not impose any functional form restrictions and allow for heterogeneity in returns in assessing the relevance of these variables.

[2] Out of curiosity, we also considered the case where we incorrectly treated all variables as continuous. As expected, these created some unexpected results which are available from the authors upon request.

Thus, it is a rather strong result that *all* school inputs are found to be irrelevant even in such a general setting.

5.2.2 Partial effects

Once we determined the irrelevant variables, we removed them from the sample and ran a new (this time local-linear) regression model (with a separate set of bandwidth estimated via LSCV). Table (2) presents the results for the gradient estimates from that local-linear regression. Nonparametric estimation provides observation-specific partial effects and standard errors. To conserve space, we present the nonparametric estimates corresponding to the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of the estimated gradient distributions, along with each corresponding wild bootstrapped standard error in parentheses.

Table (2) reveals an important message: precision set aside, there exists a large extent of heterogeneity in partial effects for each input. Not only do these partial effects vary in magnitudes, but also in signs. This heterogeneity likely comes from either nonlinearities and/or interactions between variables. We note that the types of nonlinearities and interactions are generally unknown which makes this difficult to model in a parametric setting. As we shall see below, even a relatively flexible trans-log production cannot capture these nonlinearities and interactions and leads to downward biased estimates of efficiencies.

Several findings are noteworthy. First, we find that for the majority of school districts, financial constraints (PTELL) typically exert a negative impact on students' educational outcomes. Roughly 63 percent of the school districts have a negative coefficient on PTELL. That being said, there are a large percentage which are insignificant. This result is consistent with the growing evidence that imposition of tax limits results in reductions in student achievement (e.g. Downes and Figlio, 1999). Even though the literature generally finds that the levels of school inputs do not affect student performance, as well evident above, Figlio and Kim (2001) have shown that tax limits are associated with reduced quality of these inputs, which could in turn reduce student performance. However, unlike prior studies, we also recover a proportion of the districts where these constraints play a positive role in student's academic achievements. This result could be masked by simple OLS estimates which are weighted average of these positive and negative effects. In light of the large literature on strategic competition among schools (e.g. Millimet and Collier, 2008), such positive effects may reflect that because of the competitive pressure imposed by districts without limits, fiscally constrained schools may be forced to be more efficient by reallocating resources productively. We will examine this matter later. The magnitudes of the effects also vary across districts. For example, the estimated effect at the 90^{th} percentile is roughly 4 times larger than at the 75^{th} percentile.

Second, students tend to perform better in districts with better educated populations. Specifically, we find that the effect of percentage of population with Bachelors on students' outcomes are generally positive. However, we do find a small fraction of the districts where better educated population does not necessarily benefit students. This is consistent with theoretical ambiguities regarding the effects of having neighbors of higher socioeconomic status. As noted in Kling et al. (2007), it could be beneficial because of mechanisms such as positive role models and reduced exposure to violence; or it could be adverse because children may face competition with advantaged peers and discrimination. Moreover, even for those districts with positive effects, the magnitudes of the effects again vary. For more than 5% of the school districts the effect, is at least twice larger than half of the school districts (0.062 at the (unreported) 95^{th} vs. 0.029 at the 75^{th}).

Third, we also find that the percentage of kids whose primary home language is not English has a large positive effect on educational outcomes in a majority of school districts. This result is broadly consistent with the existing literature, which fails to find harmful effects of immigration on natives' educational outcomes (e.g. Liu, 2000), and even finds slightly positive effects (e.g. Neymotin, 2009). Chin et al. (2012) similarly find that the number of limited English proficient (LEP) children does not affect ever-LEP students but has significant positive spillover effects on non-LEP students. Such a result could be explained by the fact that resources provided for LEP students are also beneficial to non-LEP students. Nevertheless, we find that the percentage of kids whose home language is not English does harm student achievements in some districts (marginal significance). Such a result is masked by ordinary regressions that examine only the average effects.

Finally, we find that for a majority of school districts, the unemployment rate predominantly has a negative effect on student achievement, while housing generally has a positive effect. These results may be explained by the fact that high mobility due to lack of affordable housing and unemployment could depress achievement for both students who move and stable children (Rothstein, 2004, p.46). Our results suggest that housing-mobility interventions such as moving-to-opportunity experiment that generate changes in neighborhood characteristics could be quite effective in improving student outcomes for some districts, although not for all. Despite the negative effects in most school districts, we also find that the unemployment rate is not necessarily harmful to student achievement. In particular, we find that there exists a fraction of districts where the unemployment rate has actually a positive effect on test scores. This result may seem surprising at first, but could be explained, e.g. by the fact that parental unemployment could increase the time spent with children on education-related activities (Levine, 2009); this positive effect could be potentially dominant for families with more resources. When splitting the sample by SES groups, Levine (2009) indeed finds such a pattern that the coefficients on the unemployment rate for children of mothers who attend college are close to zero and even positive in some specifications, while the coefficients for the children of mothers who dropped out of high school are negative. Our result confirms this pattern. Relative to the coefficients reported in Levine (2009), the absolute magnitudes of our estimates are larger. This may be because the functional form may not be linear and because there could even exist a large heterogeneity in impacts within the groups used in Levine (2009). Our nonparametric results relaxes the functional form assumption and could recover within-group heterogeneity, rather than focus solely on across-group heterogeneity as in Levine (2009).

In sum, we find that there exists great extent of heterogeneity in the effects of various factors on test scores. The findings of large heterogeneity in impacts helps to reconcile the fact that the literature consistently find at most modest associations between neighborhood characteristics and student achievement (Vigdor and Ludwig, 2010). These findings also highlight the importance of employing nonparametric techniques to complement the existing approaches to investigate the factors contributing to student achievement.

5.2.3 Technical efficiencies

Although there is a lot to be learned in terms of the partial effects of the model, another main purpose of the paper is to develop a new technique to estimate technical efficiency. Here, we compare our measures to ones obtained using Cobb-Douglas and Translog specifications. The summary statistics of the normalized efficiency scores for each approach given in Table (3). We present the 25^{th}, 50^{th}, 75^{th}, percentiles of the estimated efficiency scores as well as the minimum, mean, and maximum values.

The results are quite interesting. First, we notice that our estimates of efficiencies are larger than their parametric counterparts. The magnitudes of the differences are also sizable. For example, the minimum efficiency score for the nonparametric approach is 0.265, while the minimum scores are 0.038 and 0.003 for the Cobb-Douglas and Translog functions, respectively. The mean efficiency score for the nonparametric approach is twice as large as either parametric result. Second, the spread of the efficiency scores are much smaller for the nonparametric approach than for the parametric ones. Specifically, the coefficient of variation for efficiency scores using the nonparametric approach is only 5.100, while that for the Cobb-Douglas and Translog functions are 38.617 and 44.719, respectively. Note that the coefficient of variation for the level of test scores is about 16.300.

Finally, we notice that the parametric approaches produce rather similar efficiency scores. To further examine the relationship between different approaches, we examined the correlation rankings. Here we consider three different correlation coefficients: Kendall, Pearson and

Spearman. All three measures consistently show a very high correlation between the rank estimates of the parametric approaches. By contrast, the correlation between our nonparametric approach and parametric ones is relatively small. These results are even more evident when examining the densities of technical efficiencies (Figure 1). We find that the distribution of technical efficiencies from the parametric approaches are surprisingly close to one another, while that from the nonparametric approach is very distant from these parametric results. Stronger visual evidence is seen in Figure 2 where we look at the empirical cumulative distribution functions. These are complemented by Kolmogorov-Smirnov tests (Table 5) where we reject equality between each of the parametric and nonparametric distribution estimates of technical efficiency and at the same time fail to reject first-order uniform dominance of the nonparametric efficiency estimates over each of the corresponding parametric estimates.

We find that parametric approaches drastically underestimate the technical efficiency scores of schools. Given the findings in Gong and Sickles (1992), this suggests that the parametric estimates are downward biased. This is not only important from an econometric standpoint, but it may also be important for policy. If we believe that schools are underperforming, we may unnecessarily penalize schools, teachers and students. These findings again highlight the importance of use of the nonparametric approach.

As mentioned above, property tax caps may play a particularly important role in school efficiency. The proponents of PTELL argue that financially constrained (PTELL) schools may be forced to run schools more efficiently. Here, we test this claim by examining the efficiency scores by PTELL. In Figure 3 we look at separate distributions of the efficiency scores based on whether PTELL was equal to zero or one. The distributions of efficiency scores are essentially identical. In other words, the negative shift in the production frontier brought about by property tax caps is not met with a movement toward the frontier. This suggests that the fiscal constraints do not make the schools any more efficient. In fact, a simple check of correlation shows very little or even a negative relationship (albeit insignificant) relationship between PTELL and efficiency scores. This result is in stark contrast to the beliefs held by some researchers (Brennan and Buchanan, 1979) and many voters (e.g. Citrin, 1979) that tax and expenditure limits would make school districts eliminate waste. Remember that the majority (about 63%) of the schools constrained by PTELL experience reductions in test scores to varying extents. This indicates that lack of growth in tax revenues leads to less resources devoted to education, thereby negatively affecting students' school performances. In our analysis, if we run a simple regression of whether a school has a positive coefficient on PTELL on technical efficiency scores, the coefficient is positive and statistically significant, suggesting that efficient use of resources could potentially improve student achievement. The question is: why would administrators in these districts *not* have an incentive to be more efficient to offset the negative impact? One possible explanation is put forth in the literature (Downes et al. 1998): on one hand, inefficiency could remain unchanged because school administrators have better information about "how resources can be used productively"; on the other hand, the budget-maximizing administrators could use the decline in student performances to argue for more resources. Consistent with this explanation, Figlio and O'Sullivan (2001) provide evidence that cities subject to tax limits strategically manipulate the provisions of inputs (particularly by cutting largely instruction-related inputs but not administrative inputs) to encourage local voters to override the limit.

6 Conclusion

This paper examines the roll of schooling and non-schooling inputs in the determination of test scores in public schools in Illinois. We exploit recent advances in nonparametric kernel methods which allows us to smooth discrete variables. We show that these estimators are useful in the estimation of both production functions and in calculating technical efficiency scores as it avoids the need to 'continuize' categorical inputs.

Our results are able to uncover the relevance/irrelevance of many variables commonly used in the literature. In fact, we find that *none* of the traditional schooling inputs are relevant in the prediction of test scores. Further, we were able to discover heterogeneity in the partial effects across school districts. For example, we found that property tax caps negatively impacted test scores on average, but that some schools did not have any significant impacts. We also found no differences in efficiency scores between districts which were constrained by property tax caps and those who were not. One political argument for these caps are that they will force schools to become more efficient and we find no evidence of this. Finally, the efficiency analysis showed that parametric approaches may severely underestimate the efficiencies and overestimate the differences across school districts. Given the drastic differences between our nonparametric approach and parametric ones, it is our hope that some of the important studies of school efficiencies will be re-visited using our approach to provide information required for sound policies.

References

[1] Chin, A. N.M. Daysal and S.A. Imberman (2012). "Impact of Bilingual Education Programs on Limited English Proficient Students and Their Peers: Regression Discontinuity Evidence from Texas, " *NBER Working Papers # 18197.*

[2] Collier, T. and D.L. Millimet (2009). "Institutional Arrangements in Educational Systems and Student Achievement: A Cross-National Analysis," *Empirical Economics*, 37, 329-381.

[3] Duncombe, W., J. Miner and J. Ruggiero (1997). "Empirical Evaluation of Bureaucratic Models of Inefficiency," *Public Choice*, 93, 1-18.

[4] Eherenberg, R.G. and D.J. Brewer (1994). "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond," *Economics of Education Review*, 13, 1-17.

[5] Fan, Y., Q. Li and A. Weersink (1996). "Semi-parametric Estimation of Stochastic Production Frontier Models," *Journal of Business and Economic Statistics*, 14, 460-8.

[6] Figlio, D.N. and A. O'Sullivan (2001). "The Local Response to Tax Limitations Measures: Do Local Governments Manipulate Voters to Increase Revenues?" *Journal of Law and Economics*, 44, 233-257.

[7] Gong, B.-H. and R.C. Sickles (1992). "A Comparison Between Stochastic Frontier and Data Envelopment Methods Using Panel Data," *Journal of Econometrics,* 51, 259-84.

[8] Grosskopf, S., K.J. Hayes, L.T. Taylor and W.L. Weber (2001). "On the Determinants of School District Efficiency: Competition and Monitoring," *Journal of Urban Economics,* 49, 453-478.

[9] Hall, P., Q. Li and J.S. Racine (2007). "Nonparametric Estimation of Regression Functions in the Presence of Irrelevant Regressors," *Review of Economics and Statistics*, 89, 784-789.

[10] Hanushek, E.A. (2003). "The Failure of Input-Based Schooling Policies," *Economic Journal*, 113, F64-F98.

[11] Harris, D.N. (2010). "Education Production Functions: Concepts. " in Economics of Education, edited by D.J. Brewer and P.J. McEwan

[12] Hoxby, C.M. (1999). "The Productivity of Schools and Other Local Public Good Producers," *Journal of Public Economics*, 74, 1-30.

[13] Kling, J.R. J.B. Liebman and L.F.Katz (2007). "Experimental Analysis of Neighborhood Effects. " *Econometrica* 75(1), 83-119

[14] Kneip, A. and L. Simar (1996). "A General Framework for Frontier Estimation with Panel Data," *Journal of Productivity Analysis,* 7, 187-212.

[15] Li, Q. and J. Racine (2004). "Cross-Validated Local Linear Nonparametric Regression," *Statistica Sinica*, 14, 485-512.

[16] Liu, S. T. (2000). *Essays on the Effects of Immigration on Education and Crime.* Ph.D. thesis, MIT.

[17] Millimet, D.L. and T. Collier (2008). "Efficiency in Public Schools: Does Competition Matter?" *Journal of Econometrics*, 145, 134-157.

[18] Neymotin, F. "Immigration and its Effect on the College-Going Outcomes of Natives, " *Economics of Education Review* 28, 2009: 538-550.

[19] Park, B., R.C. Sickles and L. Simar (1998). "Stochastic Frontiers: A Semiparametric Approach," *Journal of Econometrics*, 84, 273-301.

[20] Park, B., R.C. Sickles and L. Simar (2003). "Semiparametric Efficient Estimation of AR(1) Panel Data Models," *Journal of Econometrics*, 117, 279-309.

[21] Park, B. and L. Simar (1994). "Efficient Semiparametric Estimation in a Stochastic Frontier Model," *Journal of the American Statistical Association,* 89, 929-36.

[22] Racine, J. and Q. Li (2004), "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data," *Journal of Econometrics,* 119, 99-130.

[23] Rosthstein, R. (2004). *Class and Schools: Using Social, Economic and Educational Reform to Close the Black-White Achievement Gap.* New York: Teachers College.

[24] Ruggiero, J. (1996). "Efficiency of Educational Production: An Analysis of New York School Districts," *Review of Economics and Statistics*, 78, 499-509.

[25] Ruggiero, J., J. Miner and L. Blanchard (2002). "Measureing Equity of Educational Outcomes in the Presence of Inefficiency," *European Journal of Operational Research*, 142, 642-652.

[26] Schmidt, P., and R.C. Sickles (1984). "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics,* 2, 367-74.

[27] Sickles, R.C. (2005). "Panel Estimators and the Identification of Firm-Specific Efficiency Levels in Parametric, Semiparametric and Nonparametric Settings," *Journal of Econometrics*, 126, 305-34.

[28] Todd, P. and K. Wolpin (2003). "On the Specification and Estimation of the Production Function for Cognitive Achievement, " *Economic Journal*, 113(485), F3-F33.

[29] Vigdor, J. and J. Ludwig (2010). Neighborhoods and Peers in the Production of Schooling. " in Economics of Education, edited by D.J. Brewer and P.J. McEwan

Table 1: Bandwidths (based on Least Squared Cross Validation) and Variable Relevance

Variable	Mixed Data		
	Bandwidth (1)	Bound (2)	Relevant (3)
1 = PTELL in place in county	0.308	0.5	Yes
# schools in district	0.995	1	No
FTE teachers in district	0.993	1	No
LEA administrators in district	1	1	No
current per pupil expenditure	394264.4	0.413	No
capital per pupil exp.	10891594	2.053	No
total population	667149.3	2.338	No
median hh income	271521.4	0.623	No
persons age 20+ w/o HS diploma	633337.4	2.337	No
persona 20+ w/ bachelors degree	0.594	3.059	Yes
Unemployment Rate	0.005	0.052	Yes
% kids, 4-19, speak language other than English	0.033	0.203	Yes
% occupied housing	0.071	0.086	Yes
% owner-occupied housing	0.01	0.194	Yes
Test ID	0.038	0.875	Yes
Year	1	1	No

[1] Notes: For continuous variables, bounds are calculated by 2 times the standard deviation of the variable. For unordered discrete variables, the upper bound is given by $(c-1)/c$, where c is the number of unique values the unordered categorical variable may take. For ordered discrete variables, the upper bound is unity.

Table 2: Gradient Estimates

Variable	**Gradient Estimates**				
	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
	(1)	(2)	(3)	(4)	(5)
PTELL	-0.047	-0.024	-0.004	0.005	0.022
	(0.010)	(0.013)	(0.003)	(0.004)	(0.011)
% Population with Bachelors Degree	-0.019	0.000	0.016	0.029	0.044
	(0.012)	(0.018)	(0.008)	(0.007)	(0.019)
Unemployment Rate	-6.102	-3.805	-2.062	-0.728	0.688
	(1.297)	(1.707)	(0.660)	(0.910)	(14.077)
% Non-English Speaking Kids	-0.383	-0.136	0.043	0.203	0.405
	(0.195)	(0.158)	(0.132)	(0.079)	(0.181)
% Occupied Housing	-1.511	-0.521	0.067	0.521	1.012
	(0.278)	(0.287)	(0.210)	(0.790)	(0.420)
% Owner-Occupied Housing	-0.431	-0.120	0.172	0.463	0.809
	(0.255)	(0.146)	(0.140)	(0.222)	(0.214)

[1] Notes: Bootstrapped standard errors are reported in parentheses.

Table 3: Technical Efficiencies Descriptives

	Minimum (1)	Q1 (2)	Median (3)	Mean (4)	Q3 (5)	Maximum (6)
Nonparametric Production	0.265	0.815	0.835	0.836	0.853	1.000
Cobb-Douglas Production	0.038	0.292	0.413	0.416	0.524	1.000
Translog Production	0.003	0.278	0.414	0.418	0.557	1.000

[1] Notes: $Q1$ and $Q3$ represent the 25^{th} and 75^{th} percentiles of the distribution of technical efficiencies.

Table 4: Correlation Coefficients between Different Measures of Technical Efficiencies

Panel A: Spearman Correlation

	Nonparametric	Cobb-Douglas	Translog
Nonparametric	1.000		
Cobb-Douglas	0.136	1.000	
Translog	0.114	0.981	1.000

Panel B: Kendall Correlation

	Nonparametric	Cobb-Douglas	Translog
Nonparametric	1.000		
Cobb-Douglas	0.088	1.000	
Translog	0.073	0.893	1.000

Panel C: Pearson Correlation

	Nonparametric	Cobb-Douglas	Translog
Nonparametric	1.000		
Cobb-Douglas	0.095	1.000	
Translog	0.094	0.971	1.000

Table 5: Kolmogorov-Smirnov Tests of the Efficiency Distributions

	Equality of Distributions p-values	**Stochastic Dominance** p-values
	(1)	(2)
Nonparametric v.s. Cobb-Douglas	0.0000	0.9999
Nonparametric v.s. Translog	0.0000	0.9999

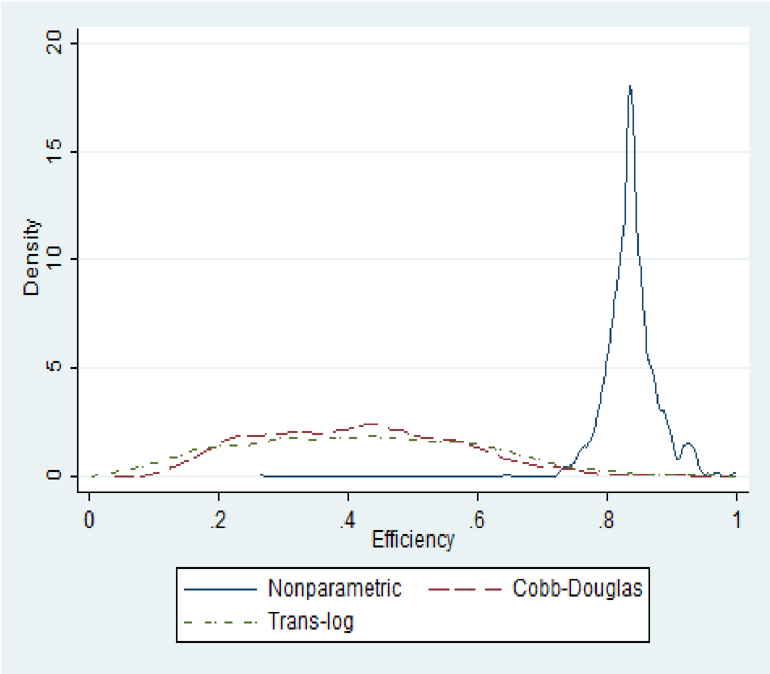
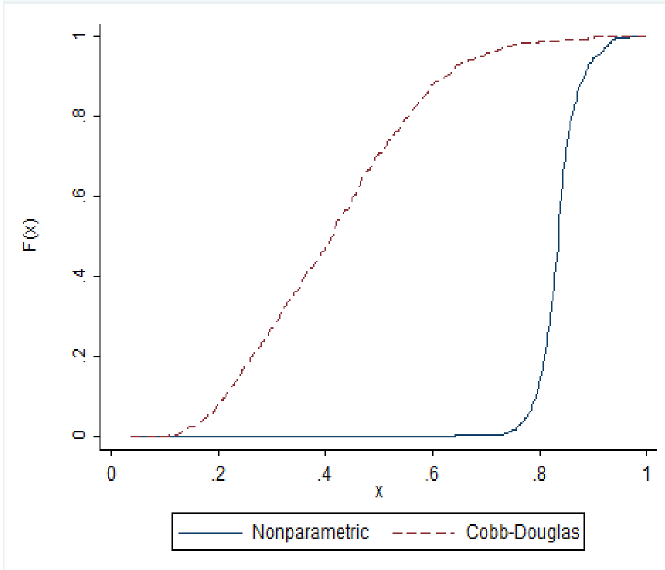
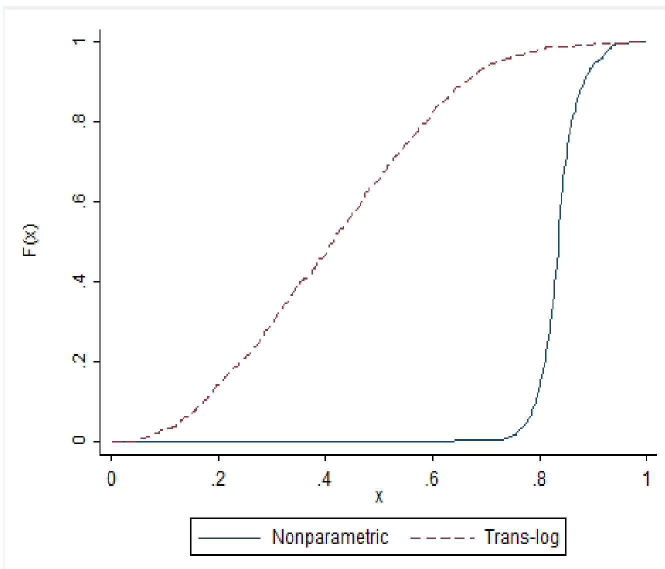


Figure 1: Density of Technical Efficiencies



(a) Nonparametric v.s. Cobb-Douglas



(b) Nonparametric v.s. Translog

Figure 2: Comparisons of Cumulative Distribution Functions

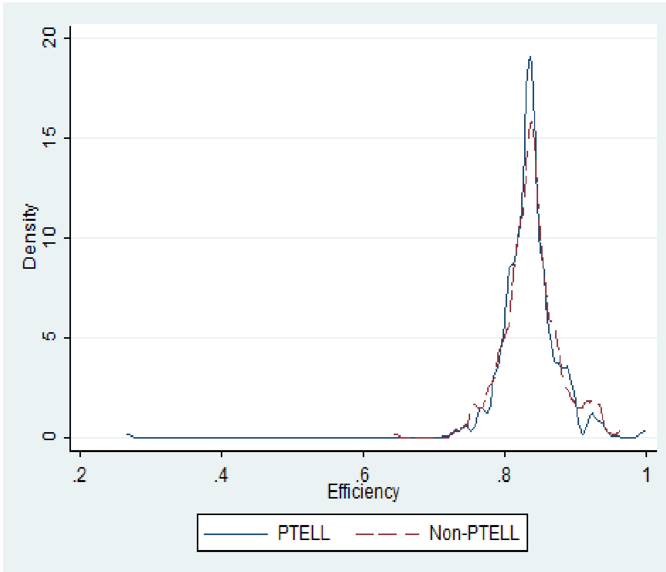


Figure 3: Density of Technical Efficiencies: PTELL v.s. Non-PTELL