Spring 5-22-2021

The Conditional Mode Estimator Of Technical Efficiency: Theory And Application To Nyc Public Schools

Yi Yang

Syracuse University, yychcsg@gmail.com

Abstract

This dissertation studies the properties of the conditional mode estimator of stochastic frontier models and applies it to measure technical inefficiency. It consists of two chapters. The first chapter analyzes the conditional mode estimator's closed-form expressions, convergence, near-minimax optimality when interpreted using Lasso, and selection rules. The second chapter applies the true fixed effect stochastic frontier model (Greene, 2005a,b) to analyze the persistent and transient technical inefficiencies of 425 NYC public middle schools for cohorts of students that graduated between 2014 to 2016.

The Conditional Mode Estimator of Technical Efficiency: Theory and Application to NYC Public Schools

Yi Yang

B.S., Nanyang Technological University, 2016

Dissertation

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics.

Syracuse University

May 2021

Acknowledgements

I would like to express my deepest gratitude to my advisor William C. Horrace, for the countless hours he spent helping me grow into a better researcher. This dissertation would not have been possible without him. I want to thank my co-authors Hyunseok Jung and Michah W. Rothbart, for their invaluable inputs. I am also thankful to the rest of my dissertation defense committee: Maria Zhu, Jeffrey D. Kubik, Hugo Jales, and the chair Wei Li for their efforts and insightful comments.

I am grateful to my friends Samuel Saltmarsh and Yimin Yi, who are the greatest companions one can find when embarking on this journey.

I want to thank my mom, Chunrong Wang, for being the strongest pillar in my life. I want to also give a special thanks to my father, Dongwei Yang, who would have been so proud.

Contents

List of Figures

List of Tables

Introduction

This dissertation studies the estimation of technical inefficiency under stochastic frontier models. In particular, it revisits the conditional mode estimator, which was proposed by Jondrow et al. (1982) but has been by and large overlooked. This dissertation fills in the vacuum of research on the conditional mode estimator by taking a closer look at its various properties in the first chapter and applying it empirically to measure education inefficiency in the second chapter.

The first chapter is a taxonomy of the various properties of the conditional mode estimator. It also advocates its use for three reasons. First, similar to the conditional mean estimator, the conditional mode estimator converges to the real inefficiency when its signal-to-noise ratio converges to zero. Second, the conditional mode estimator can take the value zero, allowing policymakers to find efficient units. We also introduce Horrace (2005) probabilities to assess the selection rule inferred by the zero conditional mode estimates. Third, we show that the conditional mode estimator achieves near-minimax optimality as it can be re-interpreted as a LASSO estimator. This chapter also includes several Monte Carlo experiments and an application to U.S. electric plants to illustrate how to apply the conditional mode estimator.

The second chapter applies the conditional mode estimator to analyze school efficiency. Specifically, we estimate the persistent and transient technical inefficiencies in Mathematics (Math) and English Language Arts (ELA) test score gains in NYC public middle schools from 2014 to 2016, using panel data and a "true" fixed effect stochastic frontier model (Greene, 2005a,b). It compares several measures of transient technical inefficiency and show that around 58% of NYC middle schools are efficient in Math gains, while 16% are efficient in ELA gains. Multivariate inference techniques are used to determine subsets of efficient schools, providing actionable decision rules to help policymakers target resources and incentives.

Chapter 1

The Conditional Mode Estimator in Parametric Frontier Models

1.1 Introduction

This chapter studies the estimation of technical inefficiency in stochastic frontier models. Stochastic frontier models (SFM) were introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977). They are statistical techniques used to estimate production (or cost) functions. Here, we primarily consider the canonical cross-sectional stochastic frontier model, written as:

$$\begin{aligned} Y_i &= f(X_i, \beta) + \varepsilon_i, \\ &= f(X_i, \beta) + v_i - u_i, \end{aligned} \tag{1.1}$$

as the basis of our analysis. Y_i and X_i are the production output and vector of inputs of firm i, where $i = 1, ..., N$. SFM differs from traditional production functions in its error term. The composite error ε is made up of two components, $\varepsilon = v - u$, where v represents statistical noise and u captures the inefficiency that prevents production from reaching its frontier. Hence, it is non-negative[1]. We refer to u as technical inefficiency. Here we also assume that u and v are independent and identically distributed as well as independent of each other and X.

One of the main applications of stochastic frontier models is to estimate the technical inefficiency, u_i, for each firm i. The challenge lies in separating technical inefficiency u_i from statistical noise v_i, as both are random variables and hence can't be separately identified. To address this problem, Jondrow et al. (1982) proposes using the mean and mode of the conditional distribution of $u_i|\varepsilon_i$, namely, $E(u_i|\varepsilon_i)$ and $M(u_i|\varepsilon_i)$, to measure inefficiency. We refer to them as the conditional mean estimator and the conditional mode estimator hereafter. Jondrow et al. (1982) also derives the closed-form expressions of both the conditional mean and the conditional mode estimator under normal-half normal (NHN) distributions, which means that $v_i \sim N(0, \sigma_v^2)$ and $u_i \sim |N(0, \sigma_u^2)|$, and normal-exponential (NE) distributions, which we write as $v_i \sim N(0, \sigma_v^2)$ and $u_i \sim \text{Exp}(1/\sigma_u)$, namely, $f(u_i) = (1/\sigma_u)\exp(-u_i/\sigma_u)$.

Since Jondrow et al. (1982), a plethora of papers have been written either studying the properties of the conditional mean estimator or applying the conditional mean estimator in an empirical setting to estimate inefficiency. In terms of its closed-form expressions, the conditional mean estimator is derived in a wider range of distributional assumptions. For example, the normal-truncated normal distributional assumption is introduced by Stevenson (1980) and the conditional mean estimator's expression in this setting is derived in Kumbhakar and Lovell (2003); Normal-gamma distributions are considered by Greene (1980a), Greene (1980b), and Stevenson (1980) and the parametric form of the conditional mean estimator in this setting is derived in Greene (1990); Horrace and Parmeter (2018) considers the Laplace distribution in stochastic frontier models. Nguyen (2010) derives the closed-form expressions under normal-uniform distributions, Laplace-exponential distributions and

[1] It would be non-positive if it is a cost frontier.

Cauthy-half Cauthy distributions. In terms of properties, Wang and Schmidt (2009) derives the distribution of the conditional mean estimator and studies how it converges when the signal-to-noise ratio converges to zero or infinity; Simar and Wilson (2010) proposes a bootstrap method to construct the confidence interval of the conditional mean estimator. The conditional mean estimator has also been applied to analyze the inefficiency of schools, banks, and others. Chakraborty et al. (2001) uses the conditional mean estimator to measure the technical inefficiencies of school districts in Utah; Mokhtar et al. (2006) uses it to compare the efficiencies between Islamic banking and conventional banking; Flores-Lagunes et al. (2007) applies it to measure fishing vessels' technical efficiencies, just to name a few.

In sheer contrast to the popularity of the conditional mean estimator, the conditional mode estimator has been rarely considered. Its closed-form expressions were only extended to normal-gamma and normal-truncated normal distribution assumptions by Kumbhakar and Lovell (2003), and we have also yet to find studies that analyze its properties or apply it to any empirical settings. This chapter tries to fill in this gap by deriving the closed-form expressions of the conditional mode estimator under various distributional assumptions and studying their properties.

Based on the findings, we advocate using the conditional mode estimator, in addition to the conditional mean estimator, when measuring technical inefficiency for several reasons. First, when the conditional mean estimator is solely used, the estimates are strictly positive, which implies that all the firms are not operating efficiently. One way to interpret the results is to consider the firm with the lowest inefficiency estimate as the efficient firm, which is a rather restrictive assumption[2]. Also, the value of the conditional mean estimator is strictly monotonic with the value of the residual, which precludes the possibility of firms being

[2] Kumbhakar et al. (2013) attempts to relax this assumption by proposing a 'zero inefficiency stochastic frontier' that allows for the existence of both efficient and inefficient units with another set of assumptions.

tied in inefficiency, as it is highly unlikely that different firms' residuals are identical. The conditional mode estimator is able to address the aforementioned problems. As one can easily tell from its closed-form expressions under common distributional assumptions, the conditional mode estimator allows firms' estimated technical inefficiencies to be tied at zero when their residuals do not surpass a certain threshold.

Second, the conditional mode estimator has similar properties with the conditional mean estimator in terms of convergence. Specifically, following Wang and Schmidt (2009), we also show that when σ_v converges to zero, the conditional mode estimator converges to the real technical inefficiency, and when σ_v converges to infinity, the conditional mode estimator converges to the mode of the distribution of real technical inefficiency[3].

Third, the conditional mode estimator achieves near-minimax optimality. It is well known that LASSO (Least Absolute Shrinkage and Selection Operator) minimizes the risk of estimation of inefficiency in the panel data setting when there are many efficient firms (or firms with zero inefficiency). In this chapter, we argue that the conditional mode estimator for inefficiency can be interpreted as a LASSO type estimator for cross-sectional SFM. It also possesses near-minimax properties like the LASSO estimator. In other words, using the conditional mode estimator to estimate technical inefficiency minimizes the maximum risk.

In addition, this chapter also assesses the selection rule based on the results of conditional mode estimates. As Horrace and Schmidt (1996) point out, both the conditional mean and the conditional mode estimator are point estimators of the technical inefficiency, which will not sufficiently capture the entire distribution of inefficiency. To address this problem, several approaches have been proposed. For example, Horrace and Schmidt (2000) introduce multiple comparisons with the best (MCB) approach and apply it to assess time-invariant

[3]Both half normal and exponential distributions' modes are at 0.

technical inefficiency in the fixed-effect stochastic frontier model. Bera and Sharma (1999) calculate the conditional variance of the estimate. Here we apply the probability statement method proposed by Horrace (2005) as our inference method. It was used previously to calculate the probability of particular fishing vessels being efficient (Flores-Lagunes et al., 2007). Here, combined with the results of technical inefficiency estimates, we use it to measure the probability of selection rules inferred by the conditional mode estimates containing the most efficient firm.

The rest of the chapter is organized as follows. In section 1.2, we compile the closed-form expressions of the conditional mode estimator under the most commonly used distributional assumptions and extend the expressions to cases under the Laplace distribution (Horrace and Parmeter, 2018). The conditional mode estimator's density function is also derived. We show that it converges in a similar way with the conditional mean estimator when the signal-to-noise ratio changes. We proceed to use conditional mean and conditional mode estimates to construct selection rules and assess their reliability using the probability statement approach in section 1.3. Re-interpretation of the conditional mode estimator as a LASSO estimator is explained in detail in section 1.4. In section 1.5, we conduct several Monte Carlo experiments to compare the performances of the conditional mean and the conditional mode estimator under a range of signal-to-noise ratios. An empirical application analyzing the inefficiency of U.S. electric utility firms is provided in section 1.6. A summary of the chapter is included in section 1.7.

1.2 Closed-Form Expressions and Distributions of the Conditional Mode Estimator

1.2.1 Closed-Form Expressions of The Conditional Mode Estimator

Most of the closed-form expressions of conditional mode are derived under the assumption that v_i is normally distributed. Conveniently, under such assumption, $u_i|\varepsilon_i$ often also follows a normal distribution truncated at 0. Specifically, Jondrow et al. (1982) finds that when u_i is distributed as a half normal distribution, $u_i \sim |N(0, \sigma_u^2)|$, $u_i|\varepsilon_i$ is distributed as $N^+(-\frac{\sigma_u^2 \varepsilon_i}{\sigma_u^2+\sigma_v^2}, \frac{\sigma_u^2 \sigma_v^2}{\sigma_u^2+\sigma_v^2})$. If u_i follows exponential distribution, $f(u) = \frac{1}{\sigma_u} e^{-\frac{u}{\sigma_u}}$, $u_i|\varepsilon_i$ is distributed as $N^+(-\varepsilon_i + \frac{\sigma_v^2}{\sigma_u}, \sigma_v^2)$. When u_i is distributed as truncated normal, namely, $f(u_i) = \frac{1}{\sqrt{2\pi}\Phi(\mu/\sigma_u)} \exp\left(-\frac{(u_i-\mu)^2}{2\sigma_u^2}\right)$, Kumbhakar and Lovell (2003) has shown that $u_i|\varepsilon_i$ would be distributed as $N^+(\frac{-\sigma_u^2 \varepsilon_i + \mu\sigma_v^2}{\sigma_v^2+\sigma_u^2}, \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2+\sigma_u^2})$. In addition, Nguyen (2010) has proven that when u_i is distributed uniformly on $[0, A]$, where $A \in R^+$, the conditional distribution is simply $N(0, \sigma_v^2)$ truncated at 0 and A. Therefore, we conclude that:

Remark 1 *When v is distributed as $N(0, \sigma_v^2)$, if $u \in R_{\geq 0}$ follows uniform, exponential or truncated normal distributions, the conditional distribution, $f(u_i|\varepsilon_i)$, is a normal distribution truncated at 0. The mode of such distribution is either the pre-truncated mean or zero, whichever is larger.*

We complied the closed-form expressions of the conditional mode estimator under the assumption that v_i follows a normal distribution in table 1.1. Comparing them to those of the conditional mean estimator, we can find that: First, the conditional mode estimator is monotonic with the value of ε, whereas the conditional mean estimator is strictly monotonic with ε, which precludes ties. Second, unlike the conditional mean estimator, which only generates positive estimates, the conditional mode estimator allows us to have zero estimates. This feature is particularly relevant for policymakers as it can be used as a selection rule of firms that are potentially efficient. Observing how the percentage of firms that are estimated to have zero conditional mode estimates change also helps us gauge the overall efficiency of the market.

Table 1.1: Closed-form Formulae for the Conditional Mode Estimator

$f(u)$	$M(u\|\varepsilon)$	
When v follows a Normal distribution, $v \sim N(0, {\sigma_v}^2)$:		
Uniform on $[0, A]$: $f(u) = \frac{1}{A}$	0	$\varepsilon \geq 0$
	$-\varepsilon$	$-A \leq \varepsilon < 0$
	A	$\varepsilon < -A$
Doubly Truncated Normal on $[0, A]$:	A	$\frac{-\sigma_u^2\varepsilon+\mu\sigma_v^2}{\sigma_v^2+\sigma_u^2} \geq A$
$f(u) = \frac{1}{\sqrt{2\pi}\sigma_u \Phi(\mu/\sigma_u)} \exp\left(-\frac{(u-\mu)^2}{2\sigma_u^2}\right)$	$\frac{-\sigma_u^2\varepsilon+\mu\sigma_v^2}{\sigma_v^2+\sigma_u^2}$	$A > \frac{-\sigma_u^2\varepsilon+\mu\sigma_v^2}{\sigma_v^2+\sigma_u^2} \geq 0$
	0	$\frac{-\sigma_u^2\varepsilon+\mu\sigma_v^2}{\sigma_v^2+\sigma_u^2} < 0$
Exponential: $f(u) = \exp(-u/\kappa)/\kappa$	$-\varepsilon - \sigma_v^2/\kappa$	$\varepsilon \leq -\sigma_v^2/\kappa$
	0	$\varepsilon > -\sigma_v^2/\kappa$
When v follows a Laplace distribution, $f(v) = \exp(-\|v\|/\gamma)/(2\gamma)$:		
Uniform on $[0, A]$: $f(u) = \frac{1}{A}$	0	$\varepsilon > 0$
	$-\varepsilon$	$-A < \varepsilon \leq 0$
	A	$\varepsilon \leq -A$
Half Normal: $u \sim N(0, {\sigma_u}^2)$	0	$\varepsilon \geq 0$
	$-\varepsilon$	$-\sigma_u^2/\gamma \leq \varepsilon < 0$
	σ_u^2/γ	$\varepsilon < -\sigma_u^2/\gamma$
Exponential: $f(u) = \exp(-u/\kappa)/\kappa$	0	$\varepsilon \geq 0$
	0	$\varepsilon < 0$ and $\kappa < \gamma$
	$-\varepsilon$	$\varepsilon < 0$ and $\kappa > \gamma$
	$[0, -\varepsilon]$	$\varepsilon < 0$ and $\kappa = \gamma$
Truncated Laplace:	0	$\varepsilon \geq 0$ and $\theta > \gamma$
$f(u) = [(1 - 0.5\exp(-\mu/\theta))/(2\theta)]\exp^{-\|u-\mu\|/\theta}$, where $\mu > 0$	μ	$\varepsilon \geq 0$ and $\theta < \gamma$
	$[0, \mu]$	$\varepsilon \geq 0$ and $\theta = \gamma$
	0	$\varepsilon < 0$, $\theta = \gamma$ and $\mu > -\varepsilon$
	μ	$\varepsilon < 0$, $\theta < \gamma$
	$-\varepsilon$	$\varepsilon < 0$ $\theta > \gamma$ and $\mu < -\varepsilon$

Truncated Laplace distributions here refer to Laplace distributions truncated at 0 with positive pre-truncated mean. Truncated Laplace distributions of non-positive pre-truncated mean are exponential distributions.

Recent literature also considered Laplace distributions. Specifically, Horrace and Parmeter (2018)[4] analyzes SFM under both Laplace-exponential and Laplace-truncated Laplace distributions, and shows that when Laplace distribution is assumed, the estimators have faster polynomial convergence rates and smaller MSEs for the majority of the time than their normal-exponential counterparts in situations of misspecification. To allow for the conditional mode estimation under Laplace distributions, we also derive the closed-form expressions under Laplace-uniform, Laplace-half normal, Laplace-exponential, and Laplace-truncated Laplace distributions. The results are reported in table 1.1.

The parametric formulae show that the value of the conditional mode estimator under Laplace distributions is determined jointly by ε_i and the relative scale of distributional parameters. This complication comes from the absolute value sign in the Laplace distributions. Another distinct feature is that in certain cases, conditional mode estimates are not points but intervals. For instance, under Laplace-truncated Laplace distributions, when $\theta = \gamma$, the conditional mode estimator can be any point from 0 to μ, as long as ε is non-negative. This feature could be potentially applied to model more complicated situations.

1.2.2 Distributions of The Conditional Mode Estimator

Wang and Schmidt (2009) analyzes the properties of the conditional mean estimator and shows that it converges to u when the signal-to-noise ratio converges to infinity and to $E(u)$ when the signal-to-noise ratio converges to zero. Here, we investigate if the conditional mode estimator has similar properties. Specifically, we analyze how the value of the conditional mode estimator changes when its signal-to-noise ratio, σ_u/σ_v, converges to extreme values. It should be pointed out that, following Wang and Schmidt (2009), we ignore sampling errors and use the term regression residual ($\hat{\varepsilon}_i$ or e_i) and regression error (ε_i) interchangeably.

[4] Corrections to some typos in the math formula of the paper are included in the Appendix.

We first derive the probability density function of the conditional mode estimator. Recall that Jondrow et al. (1982) has previously shown that, under the normal-half normal assumption, the conditional mode estimator (denoted as u^m) can be written as:

$$u_i^m = h(\varepsilon_i) = \begin{cases} -\varepsilon_i(\frac{\sigma_u^2}{\sigma^2}), & \text{if } \varepsilon_i \leq 0 \\ 0, & \text{if } \varepsilon_i > 0 \end{cases} \tag{1.2}$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$. The fact that when ε is non-positive, the conditional mode value decreases with ε allows us to perform a change of variable as following:

$$\varepsilon_i = h^{-1}(u_i^m) = g(u_i^m), \tag{1.3}$$

$$f_{u^m}(u_i^m) = f_\varepsilon(g(u_i^m))|\frac{\partial g(u_i^m)}{\partial u_i^m}|. \tag{1.4}$$

When v_i and u_i are distributed as normal and half normal (NHN), Aigner et al. (1977) proves that $f_\varepsilon(\varepsilon_i) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon_i}{\sigma}\right)[1 - \Phi(\varepsilon_i\lambda\sigma^{-1})]$, where $\lambda = \sigma_u/\sigma_v$. Therefore, as $g(u_i^m) = -\frac{\sigma^2}{\sigma_u^2}u_i^m$, we show that the probability density function of conditional mode as:

$$f_{u^m}^{NHN}(u_i^m) = F_{u^m}^{NHN}(0)\delta(u_i^m) + \frac{2\sigma}{\sigma_u^2}\phi\left(\frac{\sigma}{\sigma_u^2}u_i^m\right)\Phi\left(\frac{\sigma}{\sigma_u\sigma_v}u_i^m\right)\mathbf{I}\{u_i^m > 0\}, \tag{1.5}$$

where $\delta(.)$ is a Dirac's delta function and $F_{u^m}^{NHN}(u_i^m) = 1 - F_\varepsilon^{NHN}(-\frac{\sigma^2}{\sigma_u^2}u_i^m)$. Here u^m is a mixed random variable that is discontinuous at 0 and continuous in $(0, +\infty)$.

Following similar procedure, we find that the probability density function of conditional mode under normal-exponential (NE) is:

$$f_{u^m}^{NE}(u_i^m) = F_{u_m}^{NE}(0)\,\delta(u_i^m) + \frac{1}{\sigma_u}\exp\left\{-\frac{u_i^m}{\sigma_u} - \frac{\sigma_v^2}{2\sigma_u^2}\right\}\Phi\left(\frac{u_i^m}{\sigma_v}\right)\mathbf{I}\{u_i^m > 0\}, \tag{1.6}$$

where $F^{NE}_{u^m}(u^m_i) = 1 - F^{NE}_{\varepsilon}(-u^m_i - \frac{\sigma_v^2}{\sigma_u})$. Here u^m_i is also a mixed variable that is discontinous at 0.

Based on the probability density functions, we proved that:

Theorem 1

Under normal-half normal or normal-exponential distributions,

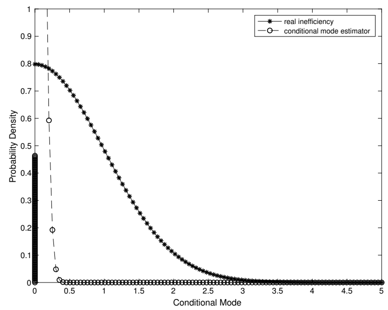
(1) When $\sigma_v^2 \to 0$, $(u^m - u) \xrightarrow{p} 0$ and $f_{u^m} \xrightarrow{d} f_u$.

(2) When $\sigma_v^2 \to \infty$, $u^m \xrightarrow{p} Mode(u)$ and $f_{u^m} \xrightarrow{d} \delta(u^m)$.

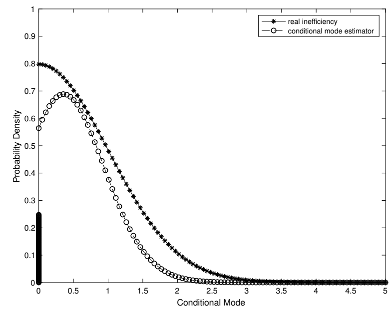
with proof included in the Appendix.

Under the assumption that v is distributed as $N(0, \sigma_v^2)$, when σ_v^2 converges to 0, theorem 1 shows that not only does u^m converge to the true underlying u, the distribution of u^m also converges to u's true distribution. As σ_v^2 converges to infinity, the statistical noise dominates the error term, u^m is converging to the mode of their true distribution, which is 0.

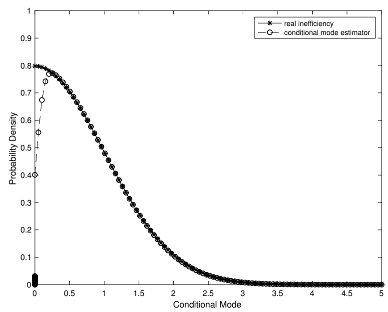
We also demonstrate the difference between u^m and u by plotting their distributions in figures 1.1 and 1.2 under signal-to-noise ratios ranging from 0.1 to 100. The value of σ_u is set to be 1 so that the curves are comparable in scale. The graphs essentially corroborate with the findings in theorem 1. One can see that the probability density function of the conditional mode estimator is clearly different than the distribution of u_i when $\sigma_v = 10$, but converges to the distribution of u_i when σ_v unilaterally decreases. This is true for both normal-half normal distributions and normal-exponential distributions.



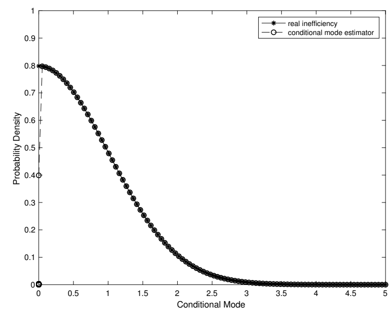
(a) $\sigma_v = 10$, $\sigma_u = 1$.



(b) $\sigma_v = 1$, $\sigma_u = 1$.

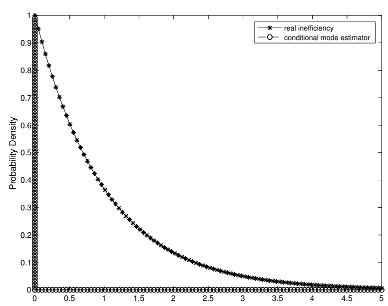


(c) $\sigma_v = 0.1$, $\sigma_u = 1$.

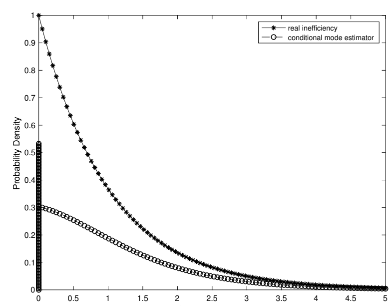


(d) $\sigma_v = 0.01$, $\sigma_u = 1$.

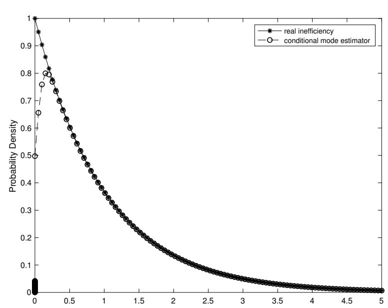
Figure 1.1: Distribution of u^m And u Under Normal-Half Normal Distributions



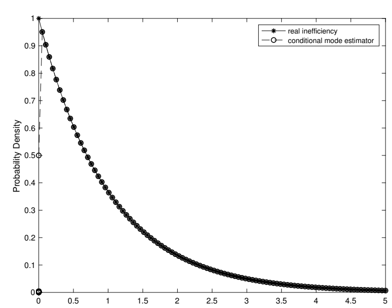
(a) $\sigma_v = 10$, $\sigma_u = 1$.



(b) $\sigma_v = 1$, $\sigma_u = 1$.



(c) $\sigma_v = 0.1$, $\sigma_u = 1$.



(d) $\sigma_v = 0.01$, $\sigma_u = 1$.

Figure 1.2: Distribution of u^m And u Under Normal-Exponential Distributions

1.3 Assessing Selection Rules using Probability Statement Approach

In this section, we apply the probability statement approach (Horrace, 2005) to assess the selection rules inferred by the conditional mode estimates.

1.3.1 Probability Statement Approach

The probability statement approach allows us to calculate the probabilities of a variety of probability statements to be true. For example, it allows us to assess the probability that firm j has the smallest real technical inefficiency, which can be written as:

$$\mathrm{P}_j = \Pr\left\{u_j \le u_i \ \forall \ i \neq j|\varepsilon_1, \dots, \varepsilon_n\right\} = \int_0^\infty f_{u_j|\varepsilon_j}(u) \prod_{i\neq j}^{n} \left[1 - F_{u_i|\varepsilon_i}(u)\right] du, \tag{1.7}$$

where $f_{u|\varepsilon}(u)$ and $F_{u|\varepsilon}(u)$ are the probability density function and cumulative distribution function of u conditional on ε. Similarly, the probability of firm j being the least efficient, thus having the largest technical inefficiency, can be written as:

$$\mathrm{P}_j^* = \Pr\left\{u_j \ge u_i \ \forall \ i \neq j|\varepsilon_1, \dots, \varepsilon_n\right\} = \int_0^\infty f_{u_j|\varepsilon_j}(u) \prod_{i\neq j}^{n} F_{u_i|\varepsilon_i}(u) du.$$

In the following, we will use P_j to denote the probability of firm j being the most efficient and P_j^* to denote the probability of firm j being the least efficient.

In the context of stochastic frontier models, this approach allows us to perform inference on the ranking of firm-level technical inefficiencies. Specifically, if we rank firms based on their conditional mean estimates (denoted as u^e), from the smallest to the largest, as:

$$\hat{u}^e_{[1]} < \hat{u}^e_{[2]} < \ldots\ldots < \hat{u}^e_{[N]}, \tag{1.8}$$

then $\mathrm{P}_{[1]} = \Pr\left\{u_{[1]} < u_{[i]} \ \forall\, i \neq 1|\varepsilon\right\}$ measures the probability that all the firms with $\hat{u}^e$ larger than $\hat{u}^e_{[1]}$ indeed have larger u than $u_{[1]}$. $u_{[1]}$ is the real inefficiency of the firm with the smallest estimated inefficiency. In other words, it measures the probability that firm with the smallest $\hat{u}^e$ also has the smallest u. Similarly, $\mathrm{P}_{[2]} = \Pr\left\{u_{[2]} < u_{[i]} \ \forall\, i \neq 2|\varepsilon\right\}$ measures the probability that firm with the second smallest $\hat{u}^e$ is the most efficient firm. Therefore, $\mathrm{P}_{[1]} + \mathrm{P}_{[2]}$ is the probability that the most efficient firm is the firm with either the smallest $\hat{u}^e$ or the second smallest $\hat{u}^e$.

1.3.2 Selection Rules based on Conditional Mode Estimates

One of the main applications of estimating technical inefficiencies is to find relatively efficient firms. When conditional mean estimates are used to find the most technically efficient firm, it is typically assumed that firm with the smallest conditional mean estimate is the most efficient. We can formalize it as a selection rule, which can be written as:

$$R^{mean}: \quad \text{Select firm } \ j \ \text{ if } \hat{u}^e_j = \min \ \hat{u}^e_i \ \text{ for all } i.$$

and the corresponding subset is $S^{mean} \equiv \left\{j : \hat{u}^e_j = \min \ \hat{u}^e_i \ \text{ for all } i\right\}$, which contains only one element.

Here we propose another selection rule, but this time based on the conditional mode estimates. Since the conditional mode estimator can take value zero, we propose only selecting those with zero conditional mode estimates into a set, which can be written as:

$$R^{mode}: \text{ Select firm } j \text{ if } \hat{u}_j^m = 0,$$

then the corresponding subset S^{mode} is $\left\{j : \hat{u}_j^m = 0\right\}$. Unlike S^{mean}, whose cardinality is fixed at one, the cardinality of S^{mode} can be larger than 1 as there can be multiple firms with zero conditional mode estimates. It is also possible that none of the firms are estimated with zero conditional mode, in which case the cardinality of the subset is 0.

Using probability statement approach, we can assess the likelihood that subsets based on the aforementioned selection rules capture the most efficient firm. In Bayesian statistics, a credible interval is defined as the interval that an unobserved parameter value falls with a particular probability. It allows us to make probability statements for a particular sample. Here, we can loosely think of those subsets based on selection rules as "credible" subsets in the sense that there is a certain probability that they contain the most efficient firm. The credible subset based on the conditional mean results would be a singleton, and the probability of it capturing the most efficient firm can be written as $\sum_{j \in S^{mean}} P_j = P_{[1]}$. The credible subset based on conditional mode being zero is $\sum_{j \in S^{mode}} P_j$. In other words, the probability of a credible subset containing the most efficient firm is the sum of the probabilities of each firm in that subset being the most efficient firm.

We can also construct a credible subset with a particular probability of containing the most efficient firm. Specifically,

Remark 2 *One can construct a subset of firms that has ω probability of containing the most* *efficient firm, where* $\omega \in [0, 1]$*. This can be done by setting a selection rule* R^{ω} *as:*

$$R^{\omega}: \quad \textit{Select all firm ranked from } [1] \textit{ to } [J] \textit{ such that } \sum_{[j]=1}^{J} P_{[j]} \geq \omega \textit{ and } \sum_{[j]=1}^{J-1} P_{[j]} < \omega$$

The corresponding subset of this selection S^{ω} *contains all firms that are ranked from [1] to [J].*

Note that while we can construct a credible subset of any ω by including firms as long as the sum of their probability statements is or exceeds ω, in order to have the credible subset with the smallest cardinality, we add into firms sequentially based on their ranking of conditional mean estimates from the smallest to the largest[5]. We analyze the performances of selections rules using Monte Carlo experiments in section 1.5,

[5]This is similar to the highest posterior density interval concept in Bayesian statistics in that it takes the interval with the highest density. Here we sequentially include firms from the largest probability of being the most efficient to the smallest.

1.4 Re-Interpreting the Conditional Mode Estimator using Lasso

Our re-interpretation of the conditional mode estimators is based on the Bayesian interpretation of the LASSO and Jondrow et al. (1982). It is well known that the LASSO can be viewed as the posterior mode estimator in the Bayesian context when imposing independent Laplace priors on the coefficients (Tibshirani, 1996; Park and Casella, 2008). Another popular shrinkage estimator Ridge can be viewed similarly as the posterior mode estimator when imposing independent normal priors on the coefficients. The mode of the posterior distribution is the maximum likelihood estimator when it use the posterior distribution as its likelihood. Noting that the essence of the Bayesian inference is to update the priors on the parameters of interest using data, the distributions imposed on inefficiency can be seen as priors in the Bayesian language and the conditional distributions we are using to reconstruct inefficiency can be viewed as the posterior conditional distributions updated by the composed error estimate. Materov (1981) shows that conditional mode is the MLE when we use the joint density of u_i and v_i given $\epsilon_i = v_i - u_i$, which is directly related to this Bayesian mode estimators.

From this context, we can see the exponential conditional mode distribution as a Lasso type estimator and the normal conditional mode as a Ridge type estimator where $\frac{\sigma_v^2}{\sigma_u}$ and $\frac{\sigma_u^2}{\sigma^2}$ are the tuning parameters driven by the distributional assumption on the inefficiency. The only difference is the shrinkage effect from the LASSO and the Ridge is symmetric around the zero, but is asymmetric in conditional mode due to the positiveness of the inefficiency. In the following section, we will show that the conditional mode possesses the minimax optimality by a theorem and a brief simulation exercise. The minimax optimality implies that conditional mode is one of the estimators that minimize the worst estimation error for the inefficiency. More importantly, we will also show it is more robust to the misspecification of the distribution of inefficiency and sparsity of it.[6]

1.4.1 Near-Minimax Optimality

To show the near-minimax optimality of the exponential conditional mode estimator, we basically use the setting used in Donoho et al. (1992) and Donoho and Johnstone (1994). The multiple estimation problem considered in the papers is as follows: We are given observations $\varepsilon = (\varepsilon_i)_{i=1}^n$ where $\varepsilon_i = \theta_i + \sigma_v v_i, v_i \sim N(0,1)$, and the objective is to estimate θ_i by some estimator, $\hat{\theta}_i$. The quality of the estimator is measured by l_2 risk, $R(\theta) = E[\sum_{i=1}^n (\hat{\theta}_i - \theta_i)]$. In our problem, $-u_i$ corresponds to θ_i with a constraint $u_i \geq 0$. The problem of the positively constrained parameters are considered in Donoho et al. (1992) so the reader can find more relevant pieces about the problem setting from the chapter.

For the mini-max optimality, we assume $\frac{\sigma_v}{\sigma_u} = (2\log n)^{1/2}$, that is the signal to noise ratio is decreasing as we have more firms, which is quite consistent with the competitive market theory as we will have less inefficient firms as we have more firms to compete. By denoting $\lambda = \frac{\sigma_v^2}{\sigma_u}$, conditional mode becomes

$$\hat{u}_i^m(\lambda) = [-\hat{\varepsilon}_i - \lambda]_+. \tag{1.9}$$

Following Donoho and Johnstone (1994) and Zou (2006), we can derive the oracle inequality

[6]The properties of the conditional mean may be discussed similarly. It can be viewed as the mean of the posterior distribution in the Bayesian context and the mean of the posterior distribution is known to minimize the Bayes Risk when we use l_2 loss function, where Bayes Risk is the expected risk of the parameters of interest evaluated by the posterior distribution of the parameters. This is related to the superiority of the conditional mode in terms of mean squared error under correctly specified distribution for inefficiency, which has been shown in Waldman (1984). Also, conditional mean and conditional mode in general can be viewed as shrinkage estimators as conditional mode shrinks $\hat{\varepsilon}_i$ toward $E(u_i)$ (Wang and Schmidt, 2009) and conditional mode shrinks $\hat{\varepsilon}_i$ toward zero.

for equation 1.9 such that:

Theorem 2

$$R(\hat{u}^m(\lambda)) \leq \left(2\log n + \frac{3}{2}\right)\left(\sum_{i=1}^{n}\min(u_i^2, \sigma_v^2) + \frac{\sigma_v^2}{\kappa}\right) \tag{1.10}$$

where $\kappa = (\log n)^{2/3} \cdot 2 \cdot \sqrt{\pi}$. The proof is provided in the appendix. This theorem says that the exponential conditional mode estimator attains the performance measured by the mean squared error which is differs from the ideal performance ($\sum_{i=1}^{n}\min(u_i^2, \sigma_v^2)$) with the oracle (Donoho and Johnstone, 1994) by at most a factor of approximately $2\log n$. Combined with the result from Donoho and Johnstone (1994) that

$$\inf_{\hat{u}}\sup_{u}\frac{R(\hat{u}^m)}{\sigma_v^2 + \sum_{i=1}^{n}\min(u_i^2, \sigma_v^2)} \sim 2\log n. \tag{1.11}$$

(1.11) indicates that the exponential conditional mode estimator achieve the near-minimax risk. In other words, it is one of the estimators which minimize the maximum risk.[7] If we do not have a prior on the distribution of inefficiency and are willing to minimize the worst estimation error, the exponential conditional mode estimator could be one of the options we can take. A couple of remarks follow.

Remark 3 *As Donoho et al. (1992) explain, the risk saving is at a price of bias induced by the shrinkage. This improvement would be more pronounced when there are many efficient firms. Therefore the exponential u^m estimator is basically for the competitive markets where a majority of firms are efficient.*

Remark 4 *Note that the MSE we are considering is for all the firms,* $R(u) = E[\sum_{i=1}^{n}(\hat{u}_i^m -$

[7] Donoho et al. (1992) derive a little bit different risk bound (due to the different oracle they are considering of), but the implication of the result is the same. "... even without knowing a priori ... we can obtain a mean-squared error which is worse only by logarithmic terms." (Donoho and Johnstone, 1992, p49)

$u_i)]$. *This implies we are more interested in the market level inefficiency or aggregate level inefficiency. The message from the oracle inequality is that the shrinkage leads to bias on some of the firms' inefficiencies but we obtain a larger variance saving from other firms. So this method would be preferred if we want to predict the overall market inefficiency (or production).*[8]

Remark 5 *Ondrich and Ruggiero (2001) show that the rank correlation between the estimated composed errors, $\hat{\varepsilon}_i$, from any SF models with log-concave noise (e.g. normal) and the associated conditional mode estimates is unity. This implies that, if we are primarily interested in the efficiency ranking of the firms, qualitatively conditional mode and $\hat{\varepsilon}_i$ are the same in most of the cases as we conventionally assume $v_i \sim N(0, \sigma_v^2)$).*[9] *However, this is not true for conditional mode estimators as they produce many ties on the top. Only conditional mode is qualitatively different from $\hat{\varepsilon}_i$, giving us an informational gain about the group of efficient firms.*

These points are borne in our simulations next section.[10]

[8] The estimation of the density of inefficiency may be one of the examples.

[9] They also show that the inefficiency we are identifying from the SF models inherently a relative measure of inefficiency, which may imply that only ranking statistics is informative among the results of SF models.

[10] Note that the exercise in Waldman (1984) is only for the conditional mean and without the sparsity in inefficiency.

1.5 Monte Carlo Experiments

In the following, we conduct three Monte Carlo experiments to examine how the conditional mean and the conditional mode estimators perform with varying sample sizes, signal-to-noise ratios, and time periods and how they perform under the sparsity assumption.

1.5.1 Comparing the Conditional Mean Estimator and the Conditional Mode Estimator

We first compare how the conditional mean and the conditional mode estimators change when their signal-to-noise ratio changes. The data generating process is based on the canonical linear stochastic frontier function, $Y = X\beta + v - u$. Specifically, $X = [X_1, X_2, X_3]$, with parameters $(\beta_1, \beta_2, \beta_3) = (2, 0.5, 2)$. X_1 is a vector of ones, whereas X_2 and X_3 are random draws from normal distributions with variance σ_v^2 and mean at 1 and 0.5, respectively. We assume that noise v is generated from a normal distribution, $N(0, \sigma_v^2)$, and signal u is generated by either a half normal or an exponential distribution with a scale parameter σ_u so that the variance parameters satisfy $\sigma_v^2 + \sigma_u^2 = 1$. The signal-to-noise ratios considered are $\lambda = \{10^{-\frac{1}{2}}, 10^{-\frac{1}{4}}, 1, 10^{\frac{1}{4}}, 10^{\frac{1}{2}}\}$. The sample size N is 1,000. The experiment is iterated for $M = 100$ times using Maximum Likelihood Estimation. The Mean Squared Errors (MSE) of the conditional mode and mean estimates against their true inefficiencies are reported in table 1.2.[11]

There are several takeaways from the results in table 1.2. First, when λ is bigger than 1, MSE decreases with increasing signal-to-noise ratios. For instance, as λ increases from 1 to $10^{\frac{1}{2}}$, MSE(u^{mean}) drops from 0.252 to 0.070 and MSE(u^{mode}) drops from 0.284 to 0.078. When λ

[11] $\text{MSE} = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{N}\sum_{i=1}^{N}(\hat{u}_{mi} - u_{mi})^2$. This is different from conventional sense of MSE where $u_{mi} = u_m$.

Table 1.2: MSEs of the Conditional Mean and Conditional Mode Estimator

	Normal-Half Normal					Normal-Exponential				
	$10^{-\frac{1}{2}}$	$10^{-\frac{1}{4}}$	10^{0}	$10^{\frac{1}{4}}$	$10^{\frac{1}{2}}$	$10^{-\frac{1}{2}}$	$10^{-\frac{1}{4}}$	10^{0}	$10^{\frac{1}{4}}$	$10^{\frac{1}{2}}$
Conditional Mean										
MSE(u^{mean})	0.092	0.170	0.197	0.142	0.072	0.309	0.306	0.252	0.153	0.070
MSE($u^{mean}\|u^{mode}$>0)	0.114	0.198	0.224	0.153	0.076	0.700	0.546	0.360	0.186	0.078
MSE($u^{mean}\|u^{mode}$=0)	0.075	0.239	0.192	0.138	0.070	0.163	0.138	0.096	0.052	0.022
Conditional Mode										
MSE(u^{mode})	0.117	0.217	0.285	0.195	0.093	0.236	0.301	0.284	0.177	0.078
MSE($u^{mean}\|u^{mode}$>0)	0.114	0.209	0.248	0.161	0.077	0.541	0.487	0.360	0.197	0.082
MSE($u^{mean}\|u^{mode}$=0)	0.091	0.242	0.474	0.752	0.863	0.122	0.170	0.172	0.117	0.054

is less than 1, however, MSE is no longer monotonic with λ. Second, MSE(u^{mean}) is smaller than MSE(u^{mode}) in most situations. This is reasonable because of the conditional expectation function prediction property of the conditional mean (Angrist and Pischke, 2008). One exception, however, is when $\lambda < 1$, under normal-exponential distributions, MSE(u^{mode}) is smaller than MSE(u^{mean}). In addition, we break down the MSEs of conditional mean and conditional mode estimates by firms with $\hat{u}^m > 0$ and those with $\hat{u}^m = 0$. The results show that the differences of MSEs are not driven by this breakdown. In other words, when the overall MSE(u^{mode}) is larger (or smaller) than MSE(u^{mean}) , both MSE(u^{mode}) for firms with positive mode and MSE(u^{mode}) for firms with zero mode are larger (or smaller) than their mean counterpart.

1.5.2 Comparing Selection Rules

We also compare the performance of selection rules by conditional mean estimates, R^{mean}, and by conditional mode estimates, R^{mode}. Without loss of generality, we assume $\sigma_v^2+\sigma_u^2 = 1$ and a pre-specified signal-to-noise ratio $\lambda = \sigma_u/\sigma_v$. Under either normal-half normal or normal-exponential distributions, we obtain a sample of size N by taking random draws of v and u from their respective distribution with variance parameter σ_v and σ_u. Note that in the event of $T > 1$, we still take N draws of u as time-invariant technical inefficiency but take NT draws for v. To prevent the impact of potential sampling error of estimated coefficients, we do not include any independent variable, equivalently, $X = \mathbf{0}$. The regression function can be written as $Y = X\beta + \varepsilon = v - u$. We can then obtain the conditional mean estimates and conditional mode estimates of each firm and the results of the selection rules. The process is iterated for $M = 1,000$ times.

First, we calculate the percentages of iterations where subsets of selection rules by conditional mean estimates and conditional mode estimates actually contain the firm with minimum inefficiency, written as:

$$\begin{aligned}
\hat{P}^{mean} &\equiv \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}\left\{j \in S_m^{mean} | u_{mj} = \min\, u_{mi} \text{ for all } i\right\}, \\
\hat{P}^{mode} &\equiv \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}\left\{j \in S_m^{mode} | u_{mj} = \min\, u_{mi} \text{ for all } i\right\},
\end{aligned}$$

where $\mathbb{1}$ is an indicator function that is only satisfied when the subset based on the selection rule contains firm with smallest u.

Second, we calculate the probabilities of the subsets containing the most efficient firm, $\hat{P}_m$, using the probability statement approach in each iteration m. We also report the average $\hat{P}_m$ over all the iterations, calculated as:

$$\begin{aligned}
\bar{\hat{P}}^{mean} &\equiv \frac{1}{M}\sum_{m=1}^{M} \hat{P}_m^{mean}. \\
\bar{\hat{P}}^{mode} &\equiv \frac{1}{M}\sum_{m=1}^{M} \hat{P}_m^{mode}.
\end{aligned}$$

Note that subsets based on different selection rules often have different cardinalities. It is also intuitive that the larger a subset's cardinality is, the more likely it contains the least inefficient firm. To adjust for the cardinality, we also calculate the average probability per firm in the subset being the most efficient, namely, $P_m^{mode}/|\bar{S}_m^{mode}|$ and $\hat{\bar{P}}_m^{mode}/|\bar{S}_m^{mode}|$.[12] There is no adjustment needed for the selection rule by conditional mean estimates as its subset only contains one firm.

The results are reported in table 1.3. The first panel reports the results when sample size N increases from 5 to 10, 20, 50, and 100.[13] Second panel shows the results when we adjust the squared signal-to-noise ratio λ^2 from 0.1 to 10. We also investigate how the probability varies with varying T and show the results when $T = 1$, 5, 10, 50, and 100 in the last panel. Column 2 - 6 report the results under the normal-half normal assumptions, and column 7 - 11 report the results under the normal-exponential distributions.

We can see that, as sample size increases, P^{mean} and $\hat{P}^{mean}$ decrease. This can be explained by the fact that S^{mean} only contains one firm. When the sample gets large, the possibility of it capturing the correct unit decreases. In contrast, P^{mode} and $\hat{P}^{mode}$ fluctuate around the same range as the sample size increases. This is because the cardinality of S^{mode} increases proportionally to sample size.

Also, as λ^2 increases, the probability of each selected firm being the most efficient firm increases, which is reflected in the increases of P^{mean}, $\hat{\bar{P}}^{mean}$ for R^{mean} and $P^{mode}/|\bar{S}^{mode}|$, $\hat{\bar{P}}^{mode}/|\bar{S}^{mode}|$ for R^{mode}. The impacts to P^{mode} and $\hat{P}^{mode}$ are less straightforward - Both of them are subject to both λ^2 and the cardinality of S^{mode}. In the meantime, changing λ^2 also affect the cardinality of S^{mode}. As a result, we find that P^{mode} and $\hat{P}^{mode}$ remain by and large constant under normal-half normal distributions but decreases under normal-

[12] The more rigorous way of calculating the probability of each unit being selected is $\hat{P}_m^{mode}/|S_m^{mode}|$ and average it across number of iterations. We use $\hat{\bar{P}}^{mode}/|\bar{S}^{mode}|$ instead because $|S_m^{mode}|$ can take value 0 in some iterations.

[13] It is worth noting that as the sample size increases, the computing power required to calculate $\hat{P}^M$ also increases significantly, as it is an integral of the product of $(N-1)$ cumulative density functions and 1 probability density function.

Table 1.3: Monte Carlo Experiment Results

Statistic	Normal-Half Normal					Normal-Exponential				
	N									
	5	10	20	50	100	5	10	20	50	100
P^{mode}	0.393	0.488	0.483	0.462	0.504	0.819	0.820	0.845	0.839	0.845
$\bar{\hat{P}}^{mode}$	0.405	0.475	0.488	0.496	0.497	0.882	0.955	0.964	0.966	0.968
P^{mean}	0.362	0.116	0.061	0.026	0.020	0.403	0.106	0.045	0.039	0.022
$\bar{\hat{P}}^{mean}$	0.363	0.107	0.061	0.033	0.018	0.520	0.208	0.127	0.076	0.043
$\lvert\bar{\hat{S}}^{mode}\rvert$	1.221	6.168	12.378	25.063	50.018	2.720	13.350	26.992	53.715	107.509
$P^{mode}/\lvert\bar{\hat{S}}^{mode}\rvert$	0.322	0.079	0.039	0.018	0.010	0.301	0.061	0.031	0.016	0.008
$\bar{\hat{P}}^{mode}/\lvert\bar{\hat{S}}^{mode}\rvert$	0.332	0.077	0.039	0.020	0.010	0.324	0.072	0.036	0.018	0.009
	λ^2									
	0.1	0.5	1	5	10	0.1	0.5	1	5	10
P^{mode}	0.491	0.485	0.483	0.472	0.463	1.000	0.923	0.845	0.644	0.591
$\bar{\hat{P}}^{mode}$	0.493	0.491	0.488	0.476	0.471	1.000	0.996	0.964	0.769	0.685
P^{mean}	0.032	0.046	0.061	0.098	0.128	0.023	0.034	0.045	0.082	0.107
$\bar{\hat{P}}^{mean}$	0.032	0.048	0.061	0.109	0.141	0.232	0.137	0.127	0.142	0.161
$\lvert\bar{\hat{S}}^{mode}\rvert$	19.955	15.080	12.378	6.590	4.844	49.788	37.019	26.992	11.122	7.571
$P^{mode}/\lvert\bar{\hat{S}}^{mode}\rvert$	0.025	0.032	0.039	0.072	0.096	0.020	0.025	0.031	0.058	0.078
$\bar{\hat{P}}^{mode}/\lvert\bar{\hat{S}}^{mode}\rvert$	0.025	0.033	0.039	0.072	0.097	0.020	0.027	0.036	0.069	0.091
	T									
	1	5	10	50	100	1	5	10	50	100
P^{mode}	0.483	0.476	0.496	0.438	0.411	0.845	0.664	0.588	0.492	0.469
$\bar{\hat{P}}^{mode}$	0.488	0.476	0.467	0.436	0.400	0.964	0.761	0.678	0.536	0.476
P^{mean}	0.061	0.101	0.149	0.246	0.329	0.045	0.099	0.147	0.231	0.284
$\bar{\hat{P}}^{mean}$	0.061	0.108	0.136	0.247	0.317	0.127	0.142	0.163	0.249	0.300
$\lvert\bar{\hat{S}}^{mode}\rvert$	12.378	6.625	4.887	2.287	1.530	26.992	10.979	7.406	3.054	2.096
$P^{mode}/\lvert\bar{\hat{S}}^{mode}\rvert$	0.039	0.072	0.101	0.192	0.269	0.031	0.060	0.079	0.161	0.224
$\bar{\hat{P}}^{mode}/\lvert\bar{\hat{S}}^{mode}\rvert$	0.039	0.072	0.095	0.191	0.262	0.036	0.069	0.092	0.176	0.227

The default setting is that M=1000; $\lambda^2 = 1$; N=50; T=1;

exponential distributions.

In addition, increasing T has a similar effect with increasing λ. This is because, under the time-invariant inefficiency setting, observing more periods of v lowers the variance of $\bar{v}$, which makes the signal-to-noise ratio increase.

1.5.3 Comparing the Conditional Mean Estimator and the Conditional Mode Estimator with Sparsity

We construct an experiemnt that relaxs the no sparsity assumption. For simplicity, we consider a data generating process such that $y = \varepsilon = v - u$, where $v \sim N(0,1)$ and $u \sim F = p \cdot \tau_0 + (1-p) \cdot \chi^2(k)$. τ_0 is Dirac mass at 0, p is a sparsity parameter taking a positive scalar in $[0,1]$ and k is a randomly selected integer from 1 to 10. This setting is intended to impose different degrees of sparsity in inefficiency and also non-standard specifications for the distribution of inefficiency. We consider applying four different estimation methods to the data. First, we estimate σ_u and σ_v using the normal-half normal stochastic frontier model of Aigner et al. (1977) and apply conditional mean and conditional mode to estimate u. Then, we do the same thing but under normal - exponential. We set $N = 1,000$ and simulate $1,000$ times for each case with $p \in \{0.1, 0.5, 0.9\}$. We report two statistics: average RMSE ($\sqrt{\sum_{i=1}^{n} (\hat{u}_i - u_i)^2}$), and average rank correlation between $\hat{u}$ and u using Spearman's ρ in table 1.4. [14]

We can see that the conditional mode estimator, under normal-exponential distribution, produces the worst results when there is little sparsity in inefficiency in both RMSE and the rank correlation, but the differences between the results of them are negligible. On the other hand, when there is little inefficiency, it produces remarkably better results than the other

[14] We exclude the cases conditional mode estimate inefficiency all zero in the calculation of the spearman's ρ

Table 1.4: Simulation Results

	RMSE				Rank Correlation			
Sparcity	1	2	3	4	1	2	3	**4**
p = 0.1	30.76	29.94	30.98	31.71	0.90	0.90	0.90	0.90
p = 0.5	48.27	37.06	34.08	28.27	0.83	0.84	0.83	0.86
p = 0.9	54.95	52.05	24.49	17.56	0.45	0.47	0.45	0.69

1: Conditional Mean under NHN; 2: Conditional Mode under NHN; 3: Conditional Mean under NE; 4: Conditional Mode under NE.

estimators. This simulation results roughly confirm the minimax property of the exponential conditional mode estimator.

1.6 Empirical Application

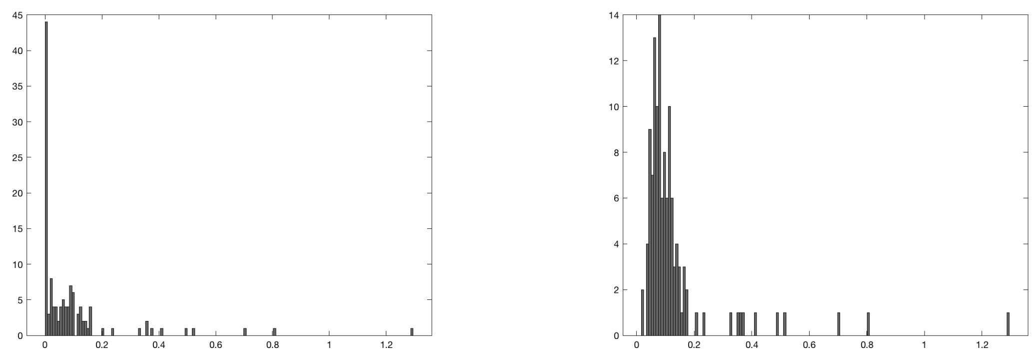
In this section, we apply a stochastic frontier model to analyze U.S. electric utility industry. This cross-sectional data set has been previously considered in Greene (1990) and Nguyen (2010). It contains observations of output (Q), labor (L), capital (K), and fuel (F) from 123 electric utility firms in the U.S, written as:

$$\ln Q_i = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + \beta_3 \ln F_i + v_i - u_i. \tag{1.12}$$

Here we estimate equation (1.12) using MLE under the normal-exponential distributional assumption and find that $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (8.497, -0.127, 0.089, 1.103)$. The variance parameters $(\sigma_v,\ \sigma_u)$ are estimated to be $(0.088, 0.129)$, respectively. Hence, the estimated signal-to-noise ratio $\hat{\sigma}_u/\hat{\sigma}_v$ in this case is 1.4659.

We proceed to estimate the technical inefficiencies and probability statements of each firm being the most efficient. The results are reported in tables 1.5, 1.6, 1.7 and 1.8 in the appendix. We rank the firms based on their conditional mean estimates reported in column 3. Note that ranking by conditional mode estimates would be the same with ranking by conditional mode estimates for firms with positive conditional mode estimates. The estimated conditional mode values are reported in column 4. Column 5 reports the probability of each of them being the most efficient firm. Column 6 reports the probability that the subset of that firm and all the higher-ranked firms containing the firm with the smallest real technical inefficiency. For instance, row 3 of table 1.5 shows the probability that the most efficient firm is among S'westernP.S, NortheastUtil, Orange&Rockln is 0.104. Column 6 suggests that there are 40 firms with zero conditional modes. If we use zero conditional mode as a selection rule, then the corresponding subset of firms has a 59.4% probability of containing the most efficient firm. We can also construct a subset that has a certain probability. If one is interested in constructing a subset with, for instance, a 95% probability of containing the most efficient firm, then the results suggest that such subset would require at least 96 firms.

The distributions of conditional mean and conditional mean estimates are plotted in figure 1.3, where we can see that some firms' conditional mode estimates are stacked at 0. But the same is not observed in conditional mean estimates, as it only allows for positive estimates.



(a) Conditional Mode

(b) Conditional Mean

Figure 1.3: Histogram of Electric Utility Firms' Estimated Technical Inefficiencies

1.7 Conclusion

One of the main applications of stochastic frontier models is to measure firm-specific technical inefficiency. To do so, Jondrow et al. (1982) proposes the conditional mean estimator and the conditional mode estimator. While there are numerous papers studying the properties of the conditional mean estimator and applying it to various empirical settings, the conditional mode estimator has been by and large overlooked.

This paper fills in the gap of research on the conditional mode estimator. We derive its parametric expressions under various distributional assumptions. Similar to the conditional mean estimator, we find that the conditional mode estimator converges to the true underlying inefficiencies when σ_v converges to zero. To assess the ranking of firms based on estimated technical inefficiencies, we also propose using the probability statement approach. It measures the probability of any firm being the most efficient or least efficient. We also use it to measure the probability that selection rules inferred by conditional mean and conditional mode estimates capture the most efficient firm. We then re-interpret the conditional mode estimator as a LASSO-type estimator and find that under the exponential distribution, it achieves the near-minimax risk. Several Monte Carlo experiments are included to compare the differences between the conditional mean estimator and conditional mode estimator. Lastly, we demonstrate how to use the conditional mode estimator by analyzing the technical inefficiencies of 123 U.S. electric utility firms.

1.8 Appendix

1.8.1 Proofs

Relevant Results under Laplace-Half Normal Distributional Assumption

When v is distributed as Laplace distribution at 0, $f_v(v) = \frac{1}{2\gamma}\exp(-\frac{|v|}{\gamma})$, and u is distributed as half-normal distribution, $f_u(u) = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\exp(-\frac{u^2}{2\sigma^2})$, we show that $f_{v,u}(\varepsilon+u,u) = \frac{1}{\sqrt{2\pi}\sigma\gamma}\exp(-\frac{|\varepsilon+u|}{\gamma} - \frac{u^2}{2\sigma^2})$. After integration with respect to u, we can find that

$$f(\varepsilon) = \begin{cases} \frac{1}{\gamma}\exp\left(-\frac{\varepsilon}{\gamma}+\frac{\sigma^2}{2\gamma^2}\right)\left[1-\Phi(\frac{\sigma}{\gamma})\right] & \varepsilon \geq 0 \\ \frac{1}{\gamma}\exp(\frac{\sigma^2}{2\gamma^2})\left[\exp(\frac{\varepsilon}{\gamma})\left\{\Phi(-\frac{\varepsilon}{\sigma}-\frac{\sigma}{\gamma})-\Phi(\frac{-\sigma}{\gamma})\right\}+\exp(-\frac{\varepsilon}{\gamma})\left\{1-\Phi(-\frac{\varepsilon}{\sigma}+\frac{\sigma}{\gamma})\right\}\right] & \varepsilon < 0 \end{cases}$$

which allow us to conduct MLE estimation.

For the purpose of finding conditional mode under this circumstance, because $f_u(u|\varepsilon) \propto f_u(u)f_v(\varepsilon+u)$, when ε is given, finding the the conditional mode is equivalent with finding the mode of $f_u(u)f_v(\varepsilon+u)$.

When $u+\varepsilon > 0$, we can write

$$f_u(u)f_v(\varepsilon+u) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(u+\frac{\sigma^2}{\gamma})^2}{2\sigma^2}\right)\times\left[\frac{1}{\gamma}\exp(-\frac{\varepsilon}{\gamma}+\frac{\sigma^2}{2\gamma^2})\right].$$

When $u+\varepsilon < 0$, $0 < u < -\varepsilon$,

$$f_u(u)f_v(\varepsilon+u) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(u-\frac{\sigma^2}{\gamma})^2}{2\sigma^2}\right)\times\left[\frac{1}{\gamma}\exp(\frac{\varepsilon}{\gamma}+\frac{\sigma^2}{2\gamma^2})\right].$$

Proof of Theorem 1

Case I: $\sigma_v^2 \to 0$ under normal-half normal

As v is assumed to be distributed as $N(0, \sigma_v^2)$, when $\sigma_v^2 \to 0$, the distribution of v collapses into a Dirac Delta function (An application of Dirac Delta function in stochastic frontier model can be referred to Horrace and Wright (2016)) at 0, which makes $\varepsilon \equiv v - u \xrightarrow{p} -u$. In addition, $\frac{\sigma_u^2}{\sigma_v^2+\sigma_u^2} \xrightarrow{p} 1$. Therefore, $u^m \equiv -\frac{\sigma_u^2}{\sigma_v^2+\sigma_u^2}\varepsilon \xrightarrow{p} u$ when $\varepsilon \leq 0$. The case when $\varepsilon > 0$ is trivial, because $\varepsilon \to -u$, and $u \in R_{\geq 0}$. In other words, we will not observe positive valued ε when $\sigma_v^2 \to 0$.

Also, when $\sigma_v^2 \to 0$, $\sigma \xrightarrow{p} \sigma_u$, which means that $\Phi(u^m\sigma/\sigma_v\sigma_u) = \Phi(u^m/\sigma_v) \xrightarrow{p} 1$. Thus, we can write $f_{u^m}(u_i^m, \varepsilon_i \leq 0)$ as $\frac{2\sigma}{\sigma_u^2}\phi\left(\frac{\sigma}{\sigma_u^2}u_i^m\right) \xrightarrow{d} 2\frac{1}{\sigma_u}\phi(\frac{u_i^m}{\sigma_u})$, which is a half normal distribution, identical with u_i.

Case II: $\sigma_v^2 \to \infty$ under normal-half normal:

As u is half normally distributed, the mode of u is at 0. When $\sigma_v^2 \xrightarrow{p} \infty$, $\frac{\sigma_u^2}{\sigma^2} \xrightarrow{p} 0$. Therefore, $u^m = -\varepsilon(\frac{\sigma_u^2}{\sigma^2}) \xrightarrow{p} 0 = \text{Mode}(u)$.

Intuitively, because all u^m converge to 0, the probability density function of u^m is a dirac delta function at 0.

Case III: $\sigma_v^2 \to 0$ under normal-exponential:

When $\sigma_v^2 \to 0$, $\varepsilon \xrightarrow{p} -u$ and $\frac{\sigma_v^2}{\sigma_u} \xrightarrow{p} 0$. Therefore, $u^m = -\varepsilon - \sigma_v^2/\sigma_u \xrightarrow{p} u$ when $u \geq 0$ and 0 when $u < 0$. It is the same with u.

As u_m is distributed as $\frac{1}{\sigma_u}\exp\{-u_m/\sigma_u - \sigma_v^2/2\sigma_u^2\}\,\Phi\left(u_m/\sigma_v\right)$, where $\Phi\left(u_m/\sigma_v\right) \xrightarrow{p} 1$ and

$\exp\{-u_m/\sigma_u - \sigma_v^2/2\sigma_u^2\}$ converges to $\exp\{-u_m/\sigma_u\}$, the probability density functions eventually converges to $\frac{1}{\sigma_u}e^{-\frac{u^m}{\sigma_u}}$, which is identical with the distribution of u.

Case IV: $\sigma_v^2 \to \infty$ under normal-exponential:

When $\sigma_v^2 \to \infty$, $-\sigma_v^2/\sigma_u$ converges to negative infinite, which makes $\varepsilon \leq -\sigma_v^2/\sigma_u$ holds for every finite ε value. Thus, u^m is always 0.

The distribution of u^m is $\frac{1}{\sigma_u}\exp\{-u^m/\sigma_u - \sigma_v^2/2\sigma_u^2\}[1-\Phi(-u^m/\sigma_v)]$. When $\sigma_v^2 \to \infty$, $\exp\{-u^m/\sigma_u - \sigma_v^2/2\sigma_u^2\}$ converges to 0. $\Phi(-u^m/\sigma_v)$ converges to 1/2. Therefore, the probability density function converges to $\delta(u^m)$.

Proof of Theorem 2

We are mostly following the proof for the theorem 3 of Zou (2006) but account for the positive constraint on inefficiency in our proof. We first prove the univariate case. Let $\varepsilon \sim N(-u, \sigma_v^2)$ where $u \geq 0$, and $\hat{u}^m(\lambda) = [-\varepsilon - \lambda]_+$ with $\lambda = \sigma_v \cdot \sqrt{2\log n}$.

We first expand the mean squared error of $\hat{u}^m(\lambda)$ such that

$$
\begin{aligned}
E[(\hat{u}^m(\lambda) - u)^2] &= E[(\hat{u}^m(\lambda) + \varepsilon)^2] + E[(-\varepsilon - u)^2] - 2E[\hat{u}^m(\lambda)(u+\varepsilon)] - 2E[\varepsilon(u+\varepsilon)] \\
&= E[(\hat{u}^m(\lambda) + \varepsilon)^2] + \sigma_v^2 - 2E[\hat{u}^m(\lambda)(u+\varepsilon)] - 2\sigma_v^2 \\
&= E[(\hat{u}^m(\lambda) + \varepsilon)^2] - 2\sigma_v^2 E\left[\frac{\partial \hat{u}^m(\lambda)}{\partial \varepsilon}\right] - \sigma_v^2
\end{aligned}
\tag{1.13}
$$

where we have used the Stein's lemma (Stein, 1981) that $E[\hat{u}^m(\lambda)\frac{u+\varepsilon}{\sigma_v^2}] = E\left[\frac{\partial \hat{u}^m(\lambda)}{\partial \varepsilon}\right]$. We consider a partition such that

$$\begin{cases} (\hat{u}^m(\lambda)+\varepsilon)^2 = \lambda^2 \ \text{ and } \ \frac{\partial \hat{u}(\lambda)}{\partial \varepsilon} = -1, \ \text{ if} -\varepsilon - \lambda > 0 \\ (\hat{u}^m(\lambda)+\varepsilon)^2 = \varepsilon^2 \ \text{ and } \ \frac{\partial \hat{u}^m(\lambda)}{\partial \varepsilon} = 0, \ \text{ if} -\varepsilon - \lambda \leq 0 \end{cases} \tag{1.14}$$

Thus, we have

$$E[(\hat{u}^m(\lambda)-u)^2] = E[\varepsilon^2 \cdot I(-\varepsilon-\lambda \leq 0)] + E[(\lambda^2+2\sigma_v^2)\cdot I(-\varepsilon-\lambda>0)] - \sigma_v^2 \tag{1.15}$$

For the first term of the right hand side of equation (1.15), we can show

$$\begin{aligned} E[\varepsilon^2 \cdot I(-\varepsilon-\lambda \leq 0)] &= E[\varepsilon^2 \cdot I(-\lambda \leq \varepsilon \leq 0)] + E[\varepsilon^2 \cdot I(0 \leq \varepsilon)] \\ &\leq \lambda^2 \cdot P(-\lambda \leq \varepsilon \leq 0) + \frac{1}{2}\sigma_v^2 \end{aligned} \tag{1.16}$$

due to $\sup E[\varepsilon^2 \cdot I(0 \leq \varepsilon)] = \frac{1}{2}\sigma_v^2$. Thus we have

$$\begin{aligned} E[(\hat{u}^m(\lambda)-u)^2] &\leq \lambda^2 \cdot P(-\lambda \leq \varepsilon \leq 0) + (\lambda^2+2\sigma_v^2)\cdot P(\varepsilon < -\lambda) - \frac{1}{2}\sigma_v^2 \\ &= \lambda^2 \cdot P(\varepsilon \leq 0) + 2\sigma_v^2 \cdot P(\varepsilon < -\lambda) - \frac{1}{2}\sigma_v^2 \\ &\leq \lambda^2 + \frac{3}{2}\sigma_v^2 \leq \left(\frac{\lambda^2}{\sigma_v^2} + \frac{3}{2}\right)\cdot\left(\sigma_v^2 + \frac{\delta}{n}\right) \end{aligned} \tag{1.17}$$

where δ is an arbitrary positive value. Next, we will use (1.15), again to derive another upper bound such that

$$\begin{aligned} E[(\hat{u}^m(\lambda)-u)^2] &= E[\varepsilon^2] + E[(\lambda^2+2\sigma_v^2-\varepsilon^2)\cdot I(\varepsilon<-\lambda)] - \sigma_v^2 \\ &\leq u^2 + 2\sigma_v^2 \cdot P(\varepsilon < -\lambda) \end{aligned} \tag{1.18}$$

Note that $P(\varepsilon < -\lambda)$ is a function of u such that

$$P(\varepsilon < -\lambda) = \int_{-\infty}^{-\lambda} \frac{1}{\sigma_v} \phi\left(\frac{\varepsilon + u}{\sigma_v}\right) d\varepsilon = \Phi\left(\frac{u-\lambda}{\sigma_v}\right) \tag{1.19}$$

Let $g(u) = P(\varepsilon < -\lambda)$, then $g(u)$ is a monotonic increasing function in u with $g^{'}(u) = \frac{1}{\sigma_v}\phi\left(\frac{u-\lambda}{\sigma_v}\right)$ and has the largest derivative at $u = \lambda$. We consider a quadratic function, $z(u) = a + \frac{1}{2}bu^2$, which satisfies $g(u) \leq z(u)$ for all u. Due to the geometric properties of $g(u)$ and $z(u)$,[15] one of the sufficient conditions for $z(u)$ is that $z(0) \geq g(0)$ and $z(\lambda) \geq g(\lambda)$ while $z^{'}(k) \geq \frac{1}{\sigma_v}\phi(0)$ for $k \geq \lambda$. The reader can verify the condition is satisfied when $a^* = \Phi\left(-\sqrt{2\log 2}\right)^{-1} \cdot \Phi\left(-\sqrt{2\log n}\right)$ and $b^* = \frac{\phi(0)}{\lambda\sigma_v}$. Furthermore, $a^* + b^* u^2 \leq a^* + \frac{\lambda^2}{2\sigma_v^4}u^2$ for all $u \geq 0$ as $b^* \leq \frac{\lambda^2}{2\sigma_v^4}$ for $n \geq 2$. So (1.18) becomes

$$\begin{aligned}
E[(\hat{u}^m(\lambda) - u)^2] &\leq u^2 + 2\sigma_v^2 \cdot \left(a^* + \frac{\lambda^2}{2\sigma_v^4}u^2\right) \\
&\leq \left(\frac{\lambda^2}{\sigma_v^2} + 1\right)\left(u^2 + \frac{2\sigma_v^4}{\lambda^2} \cdot a^*\right) \\
&\leq \left(\frac{\lambda^2}{\sigma_v^2} + 1\right)\left(u^2 + \frac{\sigma_v^2}{n \cdot (\log n)^{2/3} \cdot 2 \cdot \sqrt{\pi}}\right)
\end{aligned} \tag{1.20}$$

The last equality is due to $\Phi\left(-\sqrt{2\log n}\right) \leq \int_{-\infty}^{-\sqrt{2\log n}} \frac{x}{-\sqrt{2\log n}}\phi(x)dx = \frac{1}{\sqrt{2\log n}} \int_{-\infty}^{-\sqrt{2\log n}} \phi^{'}(x)dx = \frac{1}{\sqrt{2\log n}}\phi(-\sqrt{2\log n}) = \frac{1}{n \cdot 2 \cdot \sqrt{\pi \log n}}$. Finally, equation (1.17) and (1.20) imply

$$E[(\hat{u}^m(\lambda) - u)^2] \leq \left(\frac{\lambda^2}{\sigma_v^2} + \frac{3}{2}\right)\left(\min(u^2, \sigma_v^2) + \frac{\sigma_v^2}{n \cdot (\log n)^{2/3} \cdot 2 \cdot \sqrt{\pi}}\right) \tag{1.21}$$

The multivariate case follows by summation.

[15]The derivative of $g(u)$ is increasing until $u = \lambda$ and then decreasing whereas the derivative of $z(u)$ keep increasing. Also, the minimum value of $g(u)$ is $\Phi\left(\frac{-\lambda}{\sigma_v}\right) = \Phi\left(-\sqrt{2\log n}\right)$, which is not varying by σ_v.

Math Correction in Horrace and Parmeter (2018)

- In equation(5), under $\varepsilon < 0$, $\theta \neq \gamma$, $f_\varepsilon(\varepsilon, \mu \leq 0) = \frac{1}{2\gamma\theta}[(\lambda_+ - \lambda_-)e^{\varepsilon/\theta} + \lambda_- e^{\varepsilon/\gamma}]$. Under $\varepsilon < 0$ and $\theta = \gamma$, $f_\varepsilon(\varepsilon, \mu \leq 0) = \frac{1}{2\gamma\theta}(\lambda_+ - \varepsilon)e^{\varepsilon/\theta}$.

- In the equation below equation (8), $f_u(u|\varepsilon) = \frac{c(\mu_*)}{4\gamma\theta f_\varepsilon(0)} e^{-\frac{|u-\mu_*|}{\theta} - \frac{u}{\gamma}}$.

- The log-likelihood function under section 3.3 is $\ln L(\varepsilon_i|\beta, \gamma, \theta) = const. - \sum_{i:\varepsilon_i \geq 0}[\ln(\gamma + \theta) + \varepsilon_i/\gamma] + \sum_{i:\varepsilon_i < 0} \ln[\frac{1}{\gamma - \theta}(e^{\varepsilon_i/\gamma} - e^{\varepsilon_i/\theta}) + \frac{1}{\gamma + \theta} e^{\varepsilon_i/\theta}]$.

1.8.2 Table

Table 1.5: Inefficiencies of Electric Utility Firms Ranked 1st - 33rd.

Rank	Name	$E(u\|\varepsilon)$	$M(u\|\varepsilon)$	Probability	Cumulative Probability
1	S'westernP.S.	0.017	0.000	0.048	0.048
2	NortheastUtil.	0.023	0.000	0.035	0.083
3	Orange&Rockln.	0.036	0.000	0.021	0.104
4	DaytonPwr.&Lt.	0.037	0.000	0.021	0.125
5	BostonEdison	0.038	0.000	0.020	0.144
6	NewMex.Elec.Ser.	0.041	0.000	0.018	0.162
7	MontanaPower	0.042	0.000	0.018	0.180
8	WestTunasUtil.	0.043	0.000	0.017	0.198
9	SierraPac.Pwr.	0.043	0.000	0.017	0.215
10	ToledoEdison	0.043	0.000	0.017	0.232
11	Ctrl.HudsonG.&E.	0.044	0.000	0.017	0.249
12	PacicP&L	0.045	0.000	0.017	0.266
13	HawaiianElec.	0.045	0.000	0.017	0.282
14	LouisvilleG.&E.	0.048	0.000	0.015	0.298
15	Ctrl.III.Pub.Ser.	0.049	0.000	0.015	0.312
16	BaiigorHydro.	0.050	0.000	0.014	0.327
17	Wisc.Pub.Ser.	0.052	0.000	0.014	0.340
18	Wisc.Pwr.&Light	0.056	0.000	0.013	0.353
19	NevadaPower	0.057	0.000	0.012	0.366
20	Indy.Power&L.	0.057	0.000	0.012	0.378
21	NewEnglandEl.	0.058	0.000	0.012	0.390
22	Ctrl.Tel.&Util.	0.058	0.000	0.012	0.402
23	So.Car.EI.&Gas	0.059	0.000	0.012	0.414
24	ElPasoElec.	0.060	0.000	0.012	0.426
25	Atl.CityElec.	0.061	0.000	0.011	0.437
26	KentuckyUtils.	0.061	0.000	0.011	0.448
27	UtahPower&Lt.	0.061	0.000	0.011	0.460
28	DelmarvaP.&L.	0.062	0.000	0.011	0.471
29	MauiElectric	0.062	0.000	0.011	0.482
30	P.S.Co.ofN.H.	0.063	0.000	0.011	0.493
31	Ark.Mo.Power	0.063	0.000	0.011	0.504
32	FloridaPower	0.065	0.000	0.010	0.514
33	CommunityP.S.	0.066	0.000	0.010	0.525

Table 1.6: Inefficiencies of Electric Utility Firms Ranked 34th - 66th.

Rank	Name	$E(u\|\varepsilon)$	$M(u\|\varepsilon)$	Probability	Cumulative Probability
34	CentralLa.Pwr.	0.066	0.000	0.010	0.535
35	ClevelandEl.I.	0.066	0.000	0.010	0.545
36	CentralKansas	0.067	0.000	0.010	0.555
37	TucsonGas&E.	0.069	0.000	0.010	0.565
38	SavannahE.&P.	0.070	0.000	0.010	0.575
39	NiagaraMohawk	0.071	0.000	0.009	0.584
40	DukePowerCo.	0.071	0.000	0.009	0.594
41	KansasGas&El.	0.072	0.003	0.009	0.603
42	IowaPub.Ser.	0.072	0.005	0.009	0.612
43	SanDiegoG.&E.	0.073	0.005	0.009	0.621
44	Balt.Gas&El.	0.073	0.006	0.009	0.630
45	IowaSouthern	0.075	0.012	0.009	0.639
46	CarolinaP.&L.	0.077	0.016	0.008	0.647
47	BlackHillsP&L	0.077	0.016	0.008	0.656
48	Cinci.Gas&El.	0.078	0.018	0.008	0.664
49	LongIs.Light	0.078	0.019	0.008	0.673
50	Ariz.Pub.Ser.	0.078	0.020	0.008	0.681
51	EmpireDist.El.	0.078	0.020	0.008	0.689
52	Cent.MainePwr.	0.079	0.020	0.008	0.697
53	So.Ind.G.&E.	0.079	0.022	0.008	0.705
54	Ark.Power&Lt.	0.080	0.025	0.008	0.713
55	No.Ind.Pub.Ser.	0.081	0.026	0.008	0.721
56	MinnesotaP.&L.	0.081	0.027	0.008	0.729
57	InterstatePwr.	0.083	0.030	0.008	0.737
58	Pub.Ser.Colo.	0.083	0.030	0.008	0.745
59	NewOrleansP.S.	0.084	0.033	0.007	0.752
60	OhioEdisonCo.	0.085	0.035	0.007	0.759
61	Mo.PublicSer.	0.087	0.040	0.007	0.767
62	MadisonGas&E.	0.088	0.041	0.007	0.774
63	S'westernEl.Pr.	0.089	0.043	0.007	0.781
64	CentralPwr.&L.	0.090	0.046	0.007	0.787
65	LouisianaP.&L.	0.093	0.051	0.007	0.794
66	TampaElectric	0.093	0.052	0.006	0.800

Table 1.7: Inefficiencies of Electric Utility Firms Ranked 57th - 99th.

Rank	Name	$E(u\|\varepsilon)$	$M(u\|\varepsilon)$	Probability	Cumulative Probability
67	IllinoisPower	0.094	0.054	0.006	0.807
68	Pub.Scr.NewMex.	0.096	0.057	0.006	0.813
69	Penn.Pwr.&Lt.	0.096	0.058	0.006	0.819
70	NYStateEl.&Gas	0.099	0.064	0.006	0.825
71	Vir.Elec&Pwr.	0.100	0.064	0.006	0.831
72	Nrth.Sts.Pwr.	0.100	0.064	0.006	0.837
73	TexasPower&L.	0.101	0.066	0.006	0.843
74	East.Utl.Ass.	0.101	0.067	0.006	0.848
75	KansasPwr.&L.	0.104	0.072	0.006	0.854
76	Okla.Gas&Elec.	0.105	0.074	0.005	0.859
77	Miss.Power&L.	0.107	0.077	0.005	0.865
78	ConsumersPwr.	0.107	0.078	0.005	0.870
79	IowaPwr.&Light	0.108	0.079	0.005	0.875
80	GeneralPub.U.	0.110	0.082	0.005	0.880
81	PotomacEl.Pr.	0.111	0.084	0.005	0.885
82	SouthernCo.	0.112	0.086	0.005	0.890
83	DuquesneLight	0.113	0.087	0.005	0.895
84	Pub.Ser.El.&G.	0.113	0.087	0.005	0.900
85	DallasPwr.&L.	0.115	0.090	0.005	0.904
86	UnionElec.Co.	0.115	0.090	0.005	0.909
87	Pac.Gas&Elec.	0.117	0.094	0.005	0.914
88	HoustonLt.&Pr.	0.117	0.094	0.005	0.918
89	MontDak.Utils.	0.118	0.094	0.005	0.923
90	RochesterG.&E.	0.119	0.096	0.004	0.927
91	Pub.Ser.Okla.	0.119	0.096	0.004	0.932
92	Kan.CityP.&L.	0.119	0.097	0.004	0.936
93	UpperPen.Pwr.	0.121	0.099	0.004	0.940
94	OtterTailPwr.	0.121	0.100	0.004	0.945
95	AlleghenyPr.	0.122	0.101	0.004	0.949
96	Amer.Elec.Pr.	0.131	0.113	0.004	0.953
97	Pub.Ser.OfInd.	0.131	0.114	0.004	0.957
98	Phila.Elect.	0.132	0.115	0.004	0.960
99	LakeSup.Dist.Pr.	0.136	0.121	0.004	0.964

Table 1.8: Inefficiencies of Electric Utility Firms Ranked 100th - 123rd.

Rank	Name	$E(u\|\varepsilon)$	$M(u\|\varepsilon)$	Probability	Cumulative Probability
100	UnitedIII.Co.	0.139	0.124	0.003	0.967
101	Colms&So.Ohio	0.140	0.127	0.003	0.971
102	GulfStatesUtl.	0.141	0.128	0.003	0.974
103	NewEng.G.&E.Ass.	0.145	0.132	0.003	0.977
104	TexasElec.Ser.	0.148	0.137	0.003	0.980
105	Common.Edison	0.152	0.141	0.003	0.983
106	FloridaPwr.&L.	0.154	0.144	0.003	0.986
107	Wisc.Elec.Pwr.	0.162	0.154	0.002	0.988
108	St.JosephL&P	0.168	0.161	0.002	0.990
109	So.Cal.Edison	0.168	0.161	0.002	0.993
110	IowaElec.L.&Pwr.	0.170	0.163	0.002	0.995
111	DetroitEdison	0.170	0.163	0.002	0.997
112	HiloElec.Light	0.205	0.202	0.001	0.999
113	Consol.Edison	0.234	0.233	0.001	0.999
114	NewportElec.	0.330	0.330	0.000	1.000
115	MainePub.Ser.	0.355	0.355	0.000	1.000
116	CitizensUtils.	0.362	0.362	0.000	1.000
117	FitchburgG.&E.	0.373	0.373	0.000	1.000
118	UnitedGas.I.	0.414	0.414	0.000	1.000
119	Mt.CarmelPub.	0.493	0.493	0.000	1.000
120	IowaIII.G.&E.	0.519	0.519	0.000	1.000
121	N'westernP.S.	0.702	0.702	0.000	1.000
122	Cal.Pac.Util	0.806	0.806	0.000	1.000
123	Ctrl.Ver.Pub.Ser.	1.294	1.294	0.000	1.000

Chapter 2

Technical Efficiency of Public Middle Schools in New York City

2.1 Introduction

While improving public school education has been an empirical concern of parents, teachers, researchers, and policymakers for decades, a challenge has been the debate over the balance between increasing financial resources or pressing schools to improve efficiency. This has led to a multi-pronged policy approach in the United States (US), including both increased public-school spending - real per-pupil expenditures in public education increased from $7,000 in 1980 to $14,000 in 2015 (Baron, 2019) - and increased public school accountability - for example, the No Child Left Behind Act of 2001 (NCLB; Public Law 107-110). Nonetheless, student academic performance in the US continues to lag other Organization for Economic Co-operation and Development (OECD) countries despite spending more per pupil (Grosskopf et al., 2014). This suggests inefficiency in US public schools, where a lack of competitive market forces may allow it to persist. Consequently, econometrics production models that account for the existence of inefficiency are required, and this paper leverages the stochastic frontier literature (due to Aigner et al. (1977) and Meeusen and van den Broeck (1977)) to estimate and perform inference on inefficiency measures for public middle schools (serving grades 6-8) in New York City from 2014 to 2016. The nearest neighbors to our research are three stochastic frontier analysis of US public schools: Previous research that also conducted stochastic frontier analyses of US public schools: Chakraborty et al. (2001), Kang and Greene (2002) and Grosskopf et al. (2014). Our research adds to this literature by estimating a more flexible production specification Greene (2005a,b) and modern inference techniques (Horrace, 2005; Flores-Lagunes et al., 2007), applied to data from the largest and one of the most diverse public-school systems in the country.

Public schools in New York City (NYC) enroll over 1.1 million students in more than 1,700 schools, of which over 200,000 are in middle school grades (grades 6 through 8) in more than 500 schools. The city's size and diversity provide a unique backdrop for a school efficiency study, because it has many schools (the primary unit of observation) that operate under a common set of regulations, funding mechanisms, and procedures, reducing the potential for heterogeneity bias due to differences in the economic and policy environment. Moreover, understanding school inefficiency in this environment is of great importance as 72.8% of students in NYC public schools are from economically disadvantaged backgrounds, a characteristic often negatively associated with educational attainment (Hanushek and Luque, 2003; Kirjavainen, 2012). To this end, we construct a balanced panel of 425 public middle schools that operate from 2012 to 2016 to estimate each school's technical inefficiency for the cohorts of students in grade 8 between the 2014 and 2016 academic years (AY). We begin with a school-level educational production function that measures output during middle school as the gains in mean students' test scores in Math and English Language Arts (ELA) between grade 5 (in the spring semester before students enter middle school grades) and grade 8 (in the last spring semester of middle school). We use gains in testing outcomes to address concerns that produced outputs (e.g., proficiency rates or mean test scores) are a result of student quality (selection into middle schools) rather than school efficiency. Our production function, then, also includes inputs that broadly fit into three groups - student characteristics, teacher characteristics, and school characteristics - in order to provide estimates of and to control for the marginal effects of other features of the middle school environment.

Aside from being the first stochastic frontier analysis of NYC public schools, to the best of our knowledge this paper is the first to apply the "true fixed effect stochastic frontier model" of Greene (2005a,b) to US school production[1]. This model is highly flexible, because it accounts for both persistent (time-invariant) and transient (time-varying) inefficiency shocks. For example, Chakraborty et al. (2001) estimate only persistent inefficiency in a cross-section of Utah public schools. Kang and Greene (2002) estimate only transient inefficiency in an upstate NY public school district.Grosskopf et al. (2014) estimate only persistent inefficiency in public districts in Texas. We find that both persistent and transient inefficiency are present in NYC middle school production and ignoring either component is an empirical mistake.

In addition to improved flexibility of our specification relative to others, our paper considers different measures of transient inefficiency and uses inferential techniques that offer policy-makers a methodology to determine groups of schools that are on the efficient frontier. In particular, under common distribution assumptions, parametric stochastic frontier models only yield a truncated (below zero) normal distribution of inefficiency conditional on the production function residual for each school[2]. The most common approach to attain point estimates of school-level inefficiency is then to calculate the means of these conditional distributions (Jondrow et al., 1982) and rank them. However, the mean of a positive and

[1] Kirjavainen (2012) is the only other education paper that applies Greene's model but to Finnish secondary schools

[2] An exception in the stochastic frontier literature is the Laplace model of Horrace and Parmeter (2018), which yields conditional distributions with a probability mass at zero inefficiency.

continuous random variable can never be zero, which means these point estimates can never identify efficient (inefficiency equal to zero) schools. Therefore, in addition to calculating the means of these truncated normal distributions for each school, we calculate their modes as a point estimate of school-level efficiency (Jondrow et al., 1982). Since the truncated normal distribution for each school has a mode at zero inefficiency with positive probability, the mode measure allows for efficiency ties, potentially producing a group of firms that are on the efficient frontier. We also "salvage" the conditional mean point estimate using the inferential techniques in Horrace (2005) and Flores-Lagunes et al. (2007), which may be used to select a subset of schools that are efficient at the 95% level. We compare the cardinality of the set of mode-zero schools to the cardinality of the selected subset based on Horrace (2005)[3].

In the absence of frontier-based analyses, many studies estimate school (and teacher) effectiveness using value-added models (Ladd and Walsh, 2002; Meyer, 1997). We note these techniques are different in both purpose and form from the models we use here. Beginning with purpose, value-added models typically aim to identify the benefits of educational inputs (for example, if value-added increases when a policy is implemented) or the underlying quality of an education-producing unit (i.e., school or teacher), thus largely ignoring transient technical inefficiency. In fact, one of the major controversies of using value-added models for high-stakes public policy decisions stems from the assumption that deviations from each school's (or teacher's) fixed effect[4] may provide evidence that estimates are unstable (Koedel et al., 2015; Schochet and Chiang, 2013)[5]. The true fixed effect stochastic frontier model allows for a portion of annual deviations to reflect transient inefficiencies in education pro-

[3]Mizala et al. (2002) proposed an approach for salvaging the conditional mean point-estimate. The divide production units into four quadrants using an efficiency-achievement matrix and treating those in the first quadrant as efficient. However, the approach is ad hoc, and is no substitute for a proper inference procedure.

[4]Some use random effects to estimate value-added, but this is relatively rare in the value-added literature.

[5]Another major controversy stems from bias that results from non-random student selection into schools (Angrist et al., 2017; Ladd and Walsh, 2002)

duction (perhaps, for example, related to effort or changes in curriculum) and to estimate the size of transient inefficiency for each unit. Then, in terms of difference in form, traditional value-added models estimate the value-added of a unit as deviations from the conditional mean, while in our model we use the regression equation to develop an efficiency frontier. Using our probability statement technique, then, we can estimate the likelihood that individual units or groups of units operate on this efficiency frontier in a given observation year (or not). Conversely, value-added methods require decisionmakers to designate ad hoc cut-offs to assign policy levers, perhaps flagging high-value-added units for rewards or low-value-added units for penalty. Taken together, we believe the true fixed effect stochastic frontier model can address some of the major controversies that surround the use of value-added models or previous stochastic frontier techniques used for education policymaking, in part because the model is intended to identify inefficiency rather than quality, and in part because it separates persistent from transient inefficiencies, which allows for better targeting of policy levers towards each form of inefficiency.

In short, we find that student composition of a school is more predictive of production in ELA, while the teacher composition of a school is more predictive of Math production, which is consistent with conventional wisdom that ELA achievement is more reflective of home and individual characteristics, and Math achievement is more reflective of classroom characteristics (Bryk and Raudenbush, 1989). Second, by separating persistent technical inefficiency from transient technical inefficiency, we are able to show that both sources of inefficiency harm the productivity of middle schools in NYC (the conditonal means of both sources range from about one-half to a whole standard deviation, depending on subject considered and estimator used). Third, we offer evidence that both efficient and inefficient schools operate in all five boroughs of NYC, suggesting school inefficiency is geographically dispersed and dispersed across schools serving high and low performing students. Fourth, by separating inefficiency from the error term (under our set of distributional assumptions), decisionmakers are better able to assess the extent to which declining exam performance during middle school is due to inefficiency as opposed to statistical noise. Finally, we offer policymakers a pair of actionable decision rules that are methodologically rigorous and reflect true performance of schools, both derived from the true fixed effects model, including application of the conditional mode estimator to identify when schools operate efficiently and the more rigorous Horrace (2005) probabilities to identify a subset of the best.

The rest of the paper is organized as follows. The next section presents the econometric model and reviews the stochastic frontier literature as it relates to research in educational inefficiency. Section 2.3 discusses the data. Section 2.4 presents the empirical results. Section 2.5 concludes.

2.2 Stochastic Frontier Models in Education Efficiency

Stochastic frontier analysis (SFA) is an econometric technique to estimate a production function while accounting for statistical noise and inefficiency. A highly flexible specification for panel data is due to Greene (2005a,b), who considers the linear production function:

$$y_{it} = \alpha + x_{it}'\beta + v_{it} - u_{it} - w_i, \tag{2.1}$$

where $i = 1, \ldots, n$ and $t = 1, \ldots T$. Here u_{it} represents transient (time-varying) inefficiency of the i-th school in period t, w_i is a fixed- (or random-) effect, and v_{it} is the usual mean-zero random error term (or regression noise). The variable y_{it} is productive output (e.g., student proficiency rates, average test scores, or gains in test scores). The x_{it} is a vector of production inputs (e.g., student characteristics, teacher quality and experience, principal quality, and others), β is an unknown vector of marginal products, and α is an unknown constant. Assuming w_i is fixed, let unobserved heterogeneity be $\alpha_i = \alpha - w_i$, leading to the Greene (2005a,b) true fixed effect stocahstic fronteir model[6]. In general, w_i captures all forms of time-invariant unobserved heterogeneity. Nonetheless, the SFA literature refers to w_i as "persistent technical inefficiency" and we will follow the same practice in what follows. Our empirical focus is characterizing and making inferences on u_{it}.

Identification of the model requires mutual independence of the random error components and the ipnuts over i and t. Since the mean of u_{it} (conditional on inputs) is non-zero, identification also requires parametric distributional assumptions on the random error components. Typically it is assumed $v_{it} \sim N(0, \sigma_v^2)$ with $u_{it} \sim |N(0, \sigma_u^2)|$ (half normal) or u_{it} distributed exponentially with variance σ_u^2.[7] Then the marginal maximum simulated likelihood estima-

[6]Assuming fixed w_i allows identification of the model even when w_i is correlated with x, the usual panel date resutls

[7]Other distributions for u have been proposed, such as truncated normal (Stevenson, 1980), gamma (Greene,

tion (MMSLE), proposed by Belotti and Ilardi (2018), leads to consistent estimates of α, β, σ_u^2, σ_v^2 (as T or $n \to \infty$), and the residuals can be used to consistently estimate α_i (as $T \to \infty$). A consistent estimate of α is the maximum of the estimated α_i, and a consistent estimate of persistent inefficiency (w_i) is the difference between the estimated α and each estimated α_i. The parametric assumptions (whether u is half normal or exponential) imply that the distribution of transient inefficiency (u) conditional on $\varepsilon_{it} = v_{it} - u_{it}$ is a truncated (at zero) normal distribution parameterized in terms of the estimates of σ_u^2, σ_v^2, and T with the regression residuals (e_{it}, say), substituted for errors ε_{it} (Aigner et al., 1977).

Point estimation of school-level (transient) technical inefficiency proceeds by calculating moments of the truncated normal distribution of u conditional on $\varepsilon_{it} = e_{it}$. Jondrow et al. (1982) provide formulae for the conditional expectation, $E(u|\varepsilon_{it})$, and the conditional mode, $M(u|\varepsilon_{it})$, which are reproduced in the Appendix. The conditional mean is more commonly employed in empirical exercises as a point estimate for inefficiency but has the shortcomings that it is always positive and that the probability of ties across i is zero[8]. That is, no firm is on the efficient frontier and there are never ties in the efficiency scores. On the other hand, the conditional mode allows for ties at zero[9]. We calculate both point estimates of transient inefficiency in our application, but suggest that the oft-ignored conditional mode may be a more useful point estimate for policymakers. That is, the mode determines a group of schools to be on the efficient frontier, so policy prescriptions can be made for the group of schools that are under-preforming or to reward schools operating efficiently. This phenomenon is illustrated in 2.1, which plots the conditional mean and mode for the Normal-Half Normal (NHN) specification and for the Normal-Exponential (NE) specification for continuous values

1980b,a), uniform and half Cauchy distribution (Nguyen, 2010) and truncated Laplace (Horrace and Parmeter, 2018). Kumbhakar and Lovell (2003) show that the choice of distribution most likely does not affect the relative ranking of estimated firm-level inefficiency.

[8]This is an empirical fact to anyone familiar with the empirical literature. It is likely due to economist's preferences for conditional expectations

[9]To see this, consider a $N(\mu, \sigma^2)$ density truncated at zero. If $\mu > 0$, the mode is positive, otherwise it is zero.

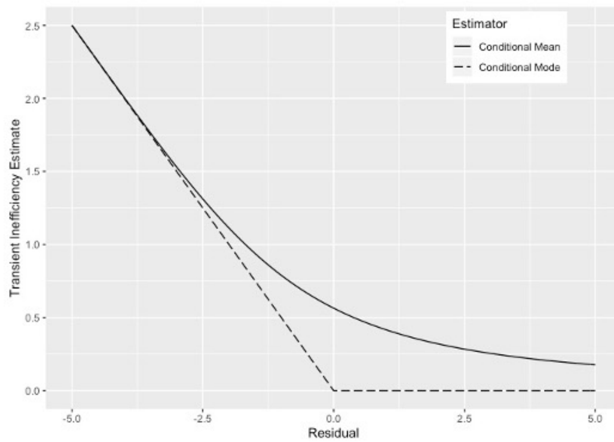


Figure 2.1: Relationship between Transient Technical Inefficiency and Residual under NHN

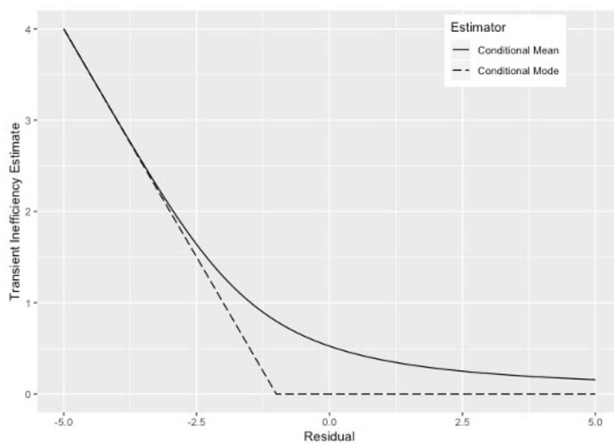


Figure 2.2: Relationship between Transient Technical Inefficiency and Residual under NE

of ε_{it} with $\sigma_u^2 = \sigma_v^2 = 1$ and $\alpha = \beta = 0$.

Selecting the schools with conditional mode equal to zero is a useful policy tool, but it is not a decision rule grounded in statistical theory, so we also appeal to the selection rule in Flores-Lagunes et al. (2007) based on the efficiency probabilities of Horrace (2005), which we briefly describe here and for which we provide more details in the Appendix. Given the n truncated normal conditional (transient) inefficiency distributions of u and given a specific time period t, we follow Horrace (2005) to characterize transient technical inefficiency as the probability that school i's draw of u is the smallest in any period t as

$$\pi_{it} = \Pr\left(u_{it} < u_{jt}, \forall j \neq i \mid \varepsilon_{1t}, \ldots, \varepsilon_{nt}\right) \tag{2.2}$$

These are within-sample, relative "efficiency probabilities". Then one may estimate the probabilities by substituting $\varepsilon_{it} = e_{it}$ above and use the estimated efficiency probabilities to select a subset of schools that contains the unknown efficient school at a prespecified confidence level (e.g., 95%), following Flores-Lagunes et al. (2007).[10] Let the population rankings of the unknown efficiency probabilities be,

$$\pi_{[n]t} > \pi_{[n-1]t} > \cdots > \pi_{[1]t},$$

and let the sample rankings of the estimated probabilities, $\hat{\pi}_{it}$ be

$$\hat{\pi}_{(n)t} > \hat{\pi}_{(n-1)t} > \cdots > \hat{\pi}_{(1)t}$$

where $[i] \neq (i)$ in general. Then, the Flores-Lagunes et al. (2007) procedure is to sum the estimated probabilities, $\hat{\pi}_{it}$, from largest to smallest until the sum is at least 0.95. Then the school indices in the sum represent a "subset of the best schools", containing the unknown best school, $i = [n]$, with probability at least 95%. Equivalently, the school indices in the subset of the best cannot be distinguished and are all on the within-sample efficient frontier (in a statistical sense). If the subset of the best is a singleton, then there is only one efficient school, $[n] = (n)$. The subset could contain all n schools, so all schools are on the frontier. The lower the cardinality of the subset, the sharper the statistical inference on $[n]$.

Education researchers have adopted SFA to estimate production frontiers and to analyze

[10]We do not show how to do this, so the reader is referred to Horrace (2005) and Flores-Lagunes et al. (2007).

school inefficiency, including: universities (Dolton et al., 2003; Gronberg et al., 2012; Stevens, 2005; Zoghbi et al., 2013); school districts (Chakraborty et al., 2001; Grosskopf et al., 2014; Kang and Greene, 2002); and primary and middle schools (Garcia-Diaz et al., 2016; Kirjavainen, 2012; Pereira and Moreira, 2007; Salas-Velasco, 2020).[11] Only a few of these studies focus on inefficiency in US public school education. Chakraborty et al. (2001) set $T = 1$ and $w = 0$ in (1) to measure the inefficiency of public education in Utah. Kang and Greene (2002) set $w = 0$ in (1) to analyze technical inefficiency in an upstate NY public school district from 1989 to 1993. Grosskopf et al. (2014) set $T = 1$ and $w = 0$ in (1) to analyze data from 965 public school districts in Texas. In all these papers, the only estimator of US school-level inefficiency considered is the conditional mean, $E(u|\varepsilon_{it})$, and none of these papers consider inference over the identification of efficient and inefficient schools in any meaningful way.

Compared to the other, earlier models, the true fixed effect model relaxes the assumption that technical inefficiency must be time invariant and allows for unobserved school heterogeneity. Unlike Greene (2005a,b), however, we estimate the model using Marginal Maximum Simulated Likelihood Estimation (MMSLE), proposed by Belotti and Ilardi (2018)[12]. The maximum likelihood dummy variable estimation originally proposed by Greene (2005a,b) suffers from an incidental parameter problem, resulting in inconsistent estimates of σ_u^2 and σ_v^2.[13] MMSLE addresses the incidental parameter problem by treating the marginal likelihood function as an expectation with respect to the change of residuals and estimates variances through simulation. MMSLE also allows for consideration of both normal-half normal and normal-exponential distribution assumptions for the technical inefficiency parameter, u_{it}.[14]

[11]Surveys of SFA in education are Worthington (2001), Johnes (2004) and De Witte and López-Torres (2017).

[12]This estimation is available on Stata in command $sftfe$.

[13]More detailed explanation of the incidental parameter problem can be found in Neyman and Scott (1948) and Lancaster (2000).

[14]Chen et al. (2014) proposes an alternative using marginal maximum likelihood estimation (MMLE), which utilizes

2.3 Data

We use data from the New York State Education Department (NYSED) and New York City Departments of Education (NYC DOE) to construct a balanced panel of education outputs (test score gains) and education inputs (student, teacher, and school characteristics) for cohorts of NYC public school students that completed middle school between AY 2014 and AY 2016. Specifically, we use school-level data from the NYS School Report Cards (SRC), which contains information on school enrollments by grade, student demographics, and teacher characteristics in every NYS public school. We merge SRC data to aggregated student data that summarizes the mean gains in Math and English Language Arts (ELA) test scores between grades 5 and 8 for each cohort in every school as well as mean characteristics of those test takers[15]. The resulting panel contains 425 public middle schools in NYC, excluding charter schools and schools that open, close, or otherwise are missing data during our sample period. The schools are scattered across all five NYC boroughs, including 133 in Brooklyn, 115 in the Bronx, 84 in Manhattan, 80 in Queens, and 13 in Staten Island.

2.3.1 Educational Outcomes

We construct cohort-level measures of normalized test score gains to measure schools' education production. We use test scores on annual standardized exams implemented by the New York State Testing Program (NYSTP), which administrates state-wide mathematics (Math) and English language arts (ELA) tests to students from grade 3 to grade 8 in compliance with the standards of the NCLB Act and, later, the "Every Student Succeeds Act (ESSA)

closed skew normal distributions properties González-Farias et al. (2004) to derive closed-form expressions of the marginal likelihood function to address the incidental parameter problem.

[15] In the following, unless specified, we use test-takers and students interchangeably.

of 2015" (Public Law 114-95, 2015)[16].

Following common practice in education economics research, we normalize student test performance across grades and years as standardized z-scores with a mean of zero and a standard deviation of one for each grade and year, thus pegging performance to the citywide mean for each cohort. The standardized exams are administered in the second half of each academic year (usually in April or May), so we calculate z-score changes ("gains") between grade 5 and grade 8 to reflect education production during the middle school period (which spans grade 6 to grade 8)[17]. Thus, for example, if a student is at precisely the citywide mean for students in grade 5 in AY 2012 and one standard deviation above average in grade 8 in AY 2015, their gain score takes a value of one (1). This has implications for interpretations of the marginal products in equation 2.1. For example, if β equals 0.5 for a variable in x_{it}, such as the share of students with limited English proficiency, then increasing this share of students from 0 to 1 increases average gains in test scores by one-half of a standard deviation. For our main sample, we restrict each cohort to those students who take both the Math and the ELA standardized exams in both grade 5 and grade 8 to limit the extent to which the composition of a cohort changes by students transferring into and out of NYC schools and the bias that results from nonrandom selection into the testing population by exam (such as students taking one exam but not the other due to expected performance). By including only students with complete exam data in each cohort, we ensure that the mean cohort-level gain scores reflect true changes in performance over time for the same students, rather than changes in the composition of test takers[18].

[16]More information can be found on `https://www.schools.nyc.gov/learning/in-our-classrooms/testing`.

[17]We also use specifications that treat grade 8 z-scores as the output, either with baseline performance in grade 5 included as a student characteristic or without that additional variable. The first of these models are akin to value-added models and produce similar results to those presented in this paper. The second do not control for baseline performance (an all-too-common practice in previous SFA research), so some estimates differ because they reflect both marginal effects and uncontrolled student quality.

[18]To test the sensitivity of our results to cohort restrictions, we relax the sample constraints to keep students with either complete (grade 5 and 8) Math or ELA exams (rather than both subjects). Results are substantively similar

2.3.2 Educational Inputs

Following Grosskopf et al. (2014), we include school, teacher, and test-taker characteristics among our educational inputs. Column one of Table 2.1 lists input variables included in this study. Test-taker characteristics include sociodemographic information, such as share of the cohort by race/ethnicity (white, black, Hispanic, Asian, or multiracial), gender, with limited English proficiency, with disabilities, and from economically disadvantaged households. Teacher characteristics include the number of teachers per one hundred students, and teacher quality measures, such as the share of teachers with a master's degree or greater, teaching without valid certification, out of certification, and who have more than three years of experiences. School characteristics include the share of classes taught by teachers without certification, the average number of classes per one hundred students, the number of staff (excluding teachers) per one hundred pupils, and the number of principal and assistant principals per one hundred students.

The second column of Table 2.1 reports citywide summary statistics of the educational inputs. Hispanic students are the largest racial/ethnic group in NYC, accounting for nearly half of students in the average middle school, followed by black students at 34.7%. More than three-fourths of students in the average NYC public middle school are economically disadvantaged, and roughly 17% are students with disabilities. We also report summary statistics by borough in columns 2-6 of Table 2.1. The share of white students accounts for only 3.88% in middle schools in the Bronx, but nearly half for the schools in Staten Island. Compared with other boroughs, schools in the Bronx also have the largest share of students from economically disadvantaged backgrounds (83.94%) and with limited English proficiency (8.83%). In terms of teacher and school inputs, middle schools in the Bronx have the highest share of teachers out of certificate (20.62%) and without valid certification (1.58%). Schools

(in magnitude and direction) to the main results reported and are available from the authors upon request.

Table 2.1: Summary Statistics for NYC Public Middle Schools

Variable	NYC	Manhattan	The Bronx	Brooklyn	Queens	Staten Island
Test-Taker Characteristics						
Share Male	49.60%	48.98%	49.47%	49.05%	50.88%	51.00%
Share Female	50.40%	51.02%	50.53%	50.95%	49.12%	49.00%
Share White	10.40%	8.87%	3.88%	10.20%	15.45%	49.27%
Share Black	34.70%	26.55%	28.03%	49.73%	31.12%	14.14%
Share Hispanic	45.00%	56.26%	64.54%	30.90%	31.50%	27.47%
Share Asian	9.07%	7.18%	3.03%	8.61%	20.60%	8.29%
Share Multiracial	0.81%	1.14%	0.50%	0.56%	1.33%	0.84%
Share Limited English	6.09%	7.48%	8.82%	4.56%	4.02%	1.54%
Share Disadvantaged	77.00%	75.03%	83.94%	78.72%	69.31%	58.52%
Share Disabled	16.90%	20.74%	17.06%	16.14%	13.54%	18.52%
Number of Test-Takers	93.83	56.86	76.64	90.48	143.61	212.51
Teacher Characteristics						
No. Teachers / 100 Students	7.42	8.01	7.47	7.84	6.49	6.71
Share Master Deg. or Higher	40.50%	35.62%	32.67%	44.02%	47.02%	66.58%
Share More 3yrs Experience	86.00%	84.40%	80.80%	89.40%	88.34%	94.12%
Share Out of Certificate	15.90%	16.36%	20.62%	14.38%	11.78%	11.31%
Share Without Certificate	1.12%	1.14%	1.58%	1.13%	0.59%	0.17%
School Characteristics						
No. of Classes /100 Students	26.76	27.26	27.01	28.6	24.05	25.5
Share Classes Uncertified	15.10%	15.44%	19.41%	13.79%	11.37%	11.12%
No. Staff / 100 Students	1	1.13	1.06	1.03	0.81	0.95
No. Principals / 100 Students	3.02	2.44	3	2.99	3.49	4.23
Mean z-score						
Grade 5 Math	-0.11	-0.13	-0.24	-0.11	0.04	0.11
Grade 8 Math	0.01	0.02	-0.16	0.01	0.21	0.25
Grade 5 ELA	-0.09	-0.12	-0.21	-0.08	0.07	0.13
Grade 8 ELA	-0.03	0	-0.2	-0.03	0.13	0.21
n	425	84	115	133	80	13

in Staten Island is at the other end of the spectrum, having the lowest mean shares of students from economically disadvantaged background (58.52%) or with limited English proficiency (1.52%). The share of teachers with master or higher degrees (66.58%) and with three or more years of experience (94.12%) are also the highest in Staten Island. We note, as well, that performance varies across districts, with the mean grade 8 Math and ELA z-scores 16% and 20% of a standard deviation below average for schools in the Bronx, but 25% and 21% of a standard deviation above average for schools in Staten Island. Average middle school gains in test performance also vary by district, but not to the same degree; the borough with the smallest gains is the Bronx with 7% and 1% of a standard deviation gains in Math and ELA, respectively, and the borough with the greatest gains is Manhattan with 16% and 11% of a standard deviation gains in those two subjects[19].

[19] All gain scores calculated as the difference between grades 8 and 5 mean performance. At first blush, it is counterintuitive that gain scores are above 0 for all boroughs, but we note that this reflects that students entering the district in middle school are lower performing than those enrolled and who take the exams in both grades.

2.4 Results

Estimates from the "true" fixed-effect stochastic frontier model in equation 2.1 are shown in Table 2.2. We only present estimates for Math (column 2) and ELA (columns 3) scores using normal-exponential and normal-half normal specifications of the model, respectively. The normal-half model for Math and the normal-exponential model for ELA did not converge, so estimates are not presented.

2.4.1 Marginal Effect of Education Inputs

Columns 2 and 3 of Table 2.2 contain the marginal effects for improvements in Math and ELA scores, respectively. Generally speaking, we find that improvements in Math scores are largely uncorrelated with test-taker and school characteristics, while teacher characteristics are important. Improvement in ELA scores are largely due to student characteristics.

Beginning with the marginal effects of test-taker characteristics, we find none of the student characteristics are correlated with middle school Math gains at the 95% significance level (though "share multiracial" is positively and "share limited English proficient" is negatively correlated with Math gains with p-values less than 0.1). Conversely, share female, Asian, and limited English proficiency are all positively correlated with ELA gains (while other test taker characteristics are not). For example, an increase in the share of a cohort who is female from none to all (0 to 1) is associated with greater gains during middle school of nearly one-fifth of a standard deviation (0.190). Put differently, a 10 percentage-point increase in the female share of students is correlated with 1.90 percent of a standard deviation greater increases in gain scores. Similarly, 10 percentage-point increases in share of a cohort who are Asian or with limited English proficiency increase ELA gains by 4.53% and 3.85% of a

Table 2.2: Results of the "True" Fixed Effect Model Estimated by MMSLE

	Math	ELA
Test-Taker Characteristics		
Share Female	-0.147	0.190***
Share Black	-0.0731	0.00227
Share Hispanic	0.174	0.134
Share Asian	0.301	0.453***
Share Multiracial	0.832*	0.284
Share Limited English	-0.262*	0.385***
Share Disadvantaged	-0.165	-0.0826
Share Disabled	-0.0932	0.0971
Teacher Characteristics		
No. Teachers / 100 Students	0.0349***	0.0146
Share Master Deg. or Higher	-0.662***	0.00811
Share More 3yrs Experience	0.635***	0.09
Share Out of Certificate	-0.13	-0.026
Share Without Certificate	1.364**	0.00143
School Characteristics		
No. of Classes /100 Students	-0.00289	0.00284*
Share Classes Uncertified	0.351	0.323
No. Staff / 100 Students	0.0818*	0.0012
No. Principals / 100 Students	-0.0122	-0.0108
σ_u	0.124	0.132
σ_v	0.137	0.111
λ	0.9051	1.1892
Observations	1,275	1,275
n	425	425
Distribution Assumed	NE	NHN

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

standard deviation, respectively.

Unlike test taker characteristics, we find teacher characteristics are more strongly correlated with Math performance gains than ELA. As the number of teachers per 100 pupils increases by 1, Math gains increase by 3.49% of a standard deviation. As the share of teachers with at least three years of experience increases by 10 percentage-points, Math gains increase by 6.62% of a standard deviation. Perhaps surprisingly, share of teachers with master's or doctorate degrees is negatively associated with gains in Math (a 10 percentage-point increase is linked with 6.62% of a standard deviation decrease in Math gains) and share of teachers without certification are positively associated (a 10 percentage-point increase is linked with 13.64% of a standard deviation greater gains). None of these teacher characteristics are correlated with ELA gains.

School characteristics appear to matter little for education production in both subjects, because none of the school characteristics are significantly correlated with gains in middle school Math or ELA performance at the 95% level (though the number of professional staff per 100 pupils is positively correlated with Math gains at the 90% level and the number of classes per 100 pupils is positively correlated with ELA gains at the 90% level).

2.4.2 Persistent Technical Inefficiency Estimates

After controlling for production inputs, Figure 2.3 summarizes the distribution of our estimates of Persistent Technical Inefficiency (PTI) by borough and by test subject (Math or ELA). That is, the figure plots the empirical distribution of our estimates of $w_i = \alpha - \alpha_i$. The rectangular boxes show the medians, 25th, and 75th percentiles of Persistent Technical Inefficiency (PTI) for each subject and borough. The lower and upper whiskers below and above each box are the percentiles that are 1.5 times the interquartile range below and above the 25th and 75th percentiles, respectively, for each subject and borough. The dots are individual estimates of PTI for schools outside the whisker percentiles: the most and least persistently efficient schools in the sample. For example, there are two dots at PTI = 0, indicating that the persistently efficient ELA school is in Brooklyn and the persistently efficient Math school in the Bronx. It also appears that there is a second Bronx school that is very close to the efficient frontier in the Math test. In general, we find that the interquartile ranges of PTI are largely higher (and, perhaps, wider) in the Bronx, Brooklyn, and Manhattan than in Queens and Staten Island. Differences in estimated PTI are less stark for ELA, but it does seem they are slightly higher in the Bronx than elsewhere. Of greater note, perhaps, is that the distributions of inefficiency across the NYC's boroughs are not so large as to reflect a "tale of two cities" - there are schools in the Bronx that are estimated to have low PTI as well as schools in Staten Island with moderate to moderately high estimated PTI. We note that direct comparisons across the two subjects should be avoided, because the educational production functions for Math and ELA are estimated separately with different distributional assumptions on the transient inefficiency component, u.

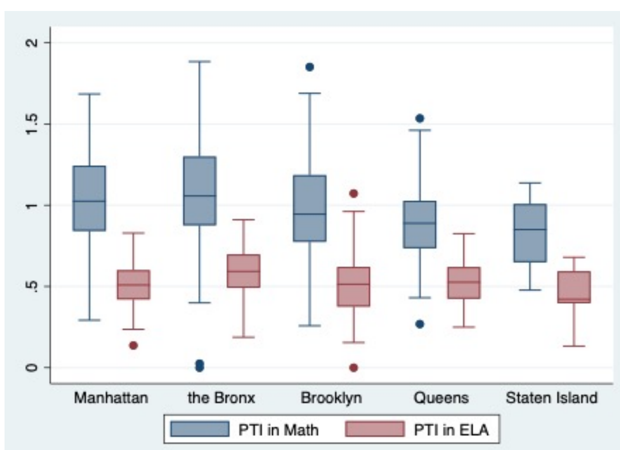


Figure 2.3: Persistent Technical Efficiency (PTI) by Borough

We report the mean and standard error of Persistent Technical Inefficiency (PTI) in Table 2.3. Consistent with Figure 2.3, the Bronx has the highest mean PTI: 1.08 for Math and

0.58 for ELA, both of which are significantly higher than the average citywide PTI. In other words, over the period the Bronx is persistently about one standard deviation below the efficient frontier of normalized test score improvements in Math and about a half a standard deviation below the frontier in ELA. Conversely, Staten Island has the smallest PTI for Math and ELA (0.82 and 0.45 for Math and ELA, respectively), and differences from the citywide mean are statistically significant. Under our modelling assumptions, this implies that schools in the Bronx persistently operate less efficiently on average than those in Staten Island (or Queens, for the matter). Given that these schools also serve the lowest performing students, as shown in Table 2.1, the results suggest that PTI increased the student achievement gap across boroughs during this period.

Table 2.3: Mean of Persistent Technical Efficiency by Subject and Borough

Subject	NYC	Manhattan	The Bronx	Brooklyn	Queens	Staten Island
Math	0.99	1.04	1.08***	0.97	0.90***	0.82**
	(0.29)	(0.29)	(0.31)	(0.3)	(0.23)	(0.24)
ELA	0.53	0.51	0.59***	0.51	0.53	0.45*
	(0.16)	(0.14)	(0.15)	(0.18)	(0.13)	(0.15)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Do schools with large Persistent Technical Inefficiency (PTI) in Math also have large PTI in ELA? Figure 2.4 presents a scatterplot of PTI in Math against PTI in ELA in all years with a linear fit line superimposed (the slope of the line is 1.13, with a t-statistic of 16.35). A Spearman test, comparing school ranks in Math PTI and ELA PTI, finds a positive (0.6169) and significant statistic (p-value = 0.0000), suggesting a strong monotonic relationship between PTI in Math and PTI in ELA.

The positive correlation of inefficiency across subjects to a degree echoes the view of Helmstadter and Walton (1985) and Luyten (1994), where schools are viewed as "classic bureaucracies" and are expected to have consistent performance across subjects. On the other hand, the fact that PTIs across subjects are not perfectly correlated also implies that schools are differentially effective in their education in different subject areas (Matthews et al., 1981; Mandeville and Anderson, 1987).

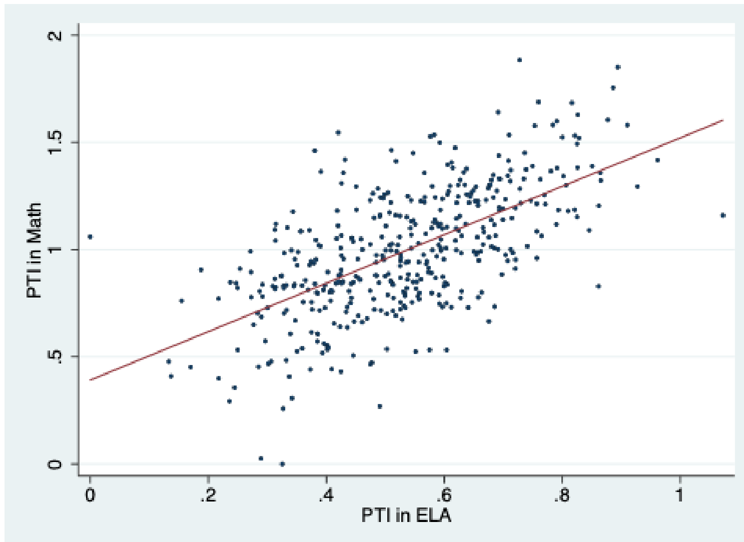


Figure 2.4: Correlation of Persistent Technical Inefficiency in Math and ELA.

2.4.3 Transient Technical Inefficiency

Table 2.4 shows summary statistics of each school's Transient Technical Inefficiency (TTI) with plotted distributions for Math and ELA shown in Figures 2.5 and 2.6, respectively. Remember, all that these models admit is the truncated (at zero) normal distribution of TTI conditional on the residual values of 425 school in each of 3 years. Here we point estimate (summarize) these conditional distributions for each of the $425 * 3 = 1,275$ school-years using their conditional means and modes (and later the conditional probability that each school is efficient) as described in section 2 and the Appendix. The first row of Table 2.4 contains summary statistics for the conditional mean of the Math TTI distributions for all schools in all years. For example, the mean of the conditional mean point estimates of TTI for Math is 0.115. That is, conditional on the residuals, we expect that Math TTI is 0.115 (3rd column) for all schools and years. Thus, on average TTI reduces improvements in Math scores by 0.115 standard deviations in the sample, which is comparable in magnitude to the mean gains in Math scores during this period (0.12 standard deviations citywide as reported in Table 2.1). Put differently, the grade 8 student achievement gap in Math for schools in the Bronx and Staten Island is approximately 0.41 standard deviations (as indicated in Table 2.1); the mean citywide TTI is 28% the size of that gap.

Table 2.4: Summary Statistics of TTI - All Schools in All Years

Subject	Estimator	Mean	S.D.	Min	25%ile	50%ile	75%ile	Max
Math	Conditional Mean	0.115	0.063	0.022	0.076	0.1	0.132	0.715
	Conditional Mode	0.045	0.081	0	0	0	0.056	0.715
ELA	Conditional Mean	0.102	0.038	0.024	0.075	0.096	0.121	0.339
	Conditional Mode	0.067	0.056	0	0.018	0.062	0.103	0.339

S.D. = Standard Deviation

The first row of Table 2.4 contains other statistics for the conditional mean estimates of Math TTI as well. For example, the observation with the minimal conditional mean point estimate for Math TTI has a value of 0.022, implying that it is 0.022 standard deviations below the efficient frontier. That is, based on the conditional mean estimates, the most efficient school-year in the sample for Math TTI is inefficient in expectation. Therefore, the conditional mean point estimate of TTI is made relative to an out-of-sample standard (a theoretical best school whose TTI distribution can be described as a Dirac delta at $u = 0$). The first row of Table 2.4 also reports the 25th, 50th and 75th percentiles of the conditional means of Math TTI distributions, as well as the maximal point estimate, which implies that we expect the least efficient school-year in the sample to be 0.715 standard deviations below the (theoretical) efficiency frontier.

The second row in Table 2.4 summarizes the conditional mode point estimates of the Math

TTI distributions. Compared to the conditional mean point estimates (first row), which are expectations, the conditional modes provide estimates of the most common (or likely) value of TTI for each observation. While the conditional mean is a measure of central tendency that can never equal zero for a non-negative u, the conditional mode can occur anywhere in the non-negative support of the truncated normal distributions that characterizes TTI. In particular we see in the second row of Table 2.4 that the average of the conditional mode point estimates is 0.045, which is considerably lower than the average of the conditional mean estimates (0.115) in the first row. We also see that the minimal estimate of the conditional mode is exactly zero (5th column). That is, for this school-year the most likely draw from its conditional distribution of TTI is $u = 0$, an efficient draw. Looking across the second row in Table 2.4, this is also true of the school at the 25th percentile (6th column) and the median school (7th column), meaning that at least half the schools in the sample are likely to be efficient (their conditional mode is on the frontier) even though they are appear inefficient in expectation (their conditional mean is not). While the conditional mean and the conditional mode of TTI summarize the truncated normal distributions in different ways, the mode has the added benefit of providing an ad hoc decision rule for selecting efficient schools: those with conditional modes equal to zero[20]. For example, in the last row of the table we see that the minimal value for the conditional mode of the ELA TTI is zero (5th column), as expected. However, the 25th percentile is positive 0.018 (6th column), implying that at least 75% of the observed school-years are unlikely to be efficient.

Finally, we note that in Table 2.4 the maximum conditional mean and the mode estimates appear to be the same for Math TTI, 0.715 (last column, first and second rows) and for ELA TTI, 0.339 (last column, third and fourth rows), but this equivalence is rounding error. Due to the nature of a normal distribution truncated at zero, the distribution's mean is always

[20] A statistically rigorous decision rule is based on the Horrace (2005) efficiency probabilities, and is considered in the sequel.

larger than its mode. For the maximal school-year observation in the last column of Table 2.4, however, the amount of truncation is so small that at three significant digits (0.715 for Math and 0.339 for ELA) its effect is negligible[21].

Figures 2.5 and 2.6 plot the empirical distributions of our TTI estimates in Table 2.4(note the axes scales for the two subjects differ). The conditional mean (red) and mode (blue) distributions in Figure 2.5 correspond to the summary statistics in rows 1 and 2 (respectively) of Table 2.4, while the distributions in Figure 2.6 correspond to the statistics in rows 3 and 4 of the table. The usefulness of the conditional mode as a decision rule for selecting efficient schools each year is clear. In Figure 2.5, the blue spike at zero indicates that more than 60% of the school-year observations in the sample are likely to be efficient in terms of their conditional distributions of Math TTI. In Figure 2.6 the blue spike indicates that about 19% of the school-year observations in the sample are likely to be efficient in terms of their conditional distributions of ELA TTI. Again, this is an ad hoc decision rule, but one that is easily understood by policymakers. What could a policymaker make of the red distributions of the conditional means in Figures 2.5 and 2.6? Not much compared to the blue distributions of the conditional modes in these figures.

In Table 2.5 we compare TTI by borough and academic year, reporting the percentage of schools with zero estimated TTI based on the conditional mode point estimate and our ad hoc decision rule. The table is self-explanatory. Staten Island has the highest percentage of schools operating efficiently (5th column) in Math (62%), followed by Queens (60%), the Bronx (59%), Manhattan (58%) and Brooklyn (55%) in order. Conversely, Manhattan and Brooklyn have the highest percentage of transiently efficient schools (last column) in ELA

[21] As with any truncated normal distribution with a very large mean equal to its mode (due to symmetry), the distribution is no longer symmetric after truncation at zero. That is, its new, post-truncation mean is necessarily larger than its mode, which is unchanged when the pre-truncated mean is positive. Moreover, the mean shifts further to the right as the amount of truncation increases.

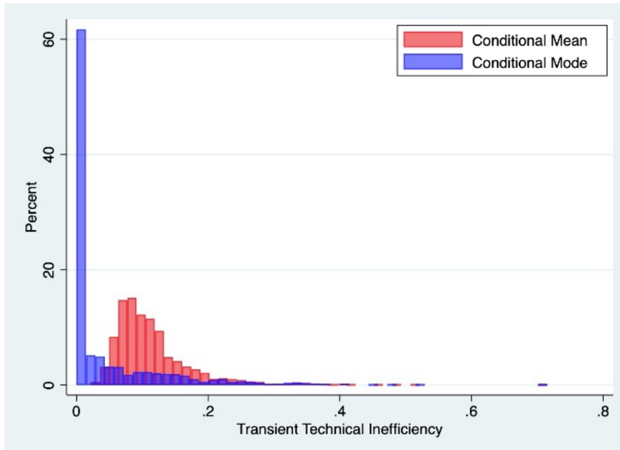


Figure 2.5: Distribution of Transient Technical Inefficiency in Math

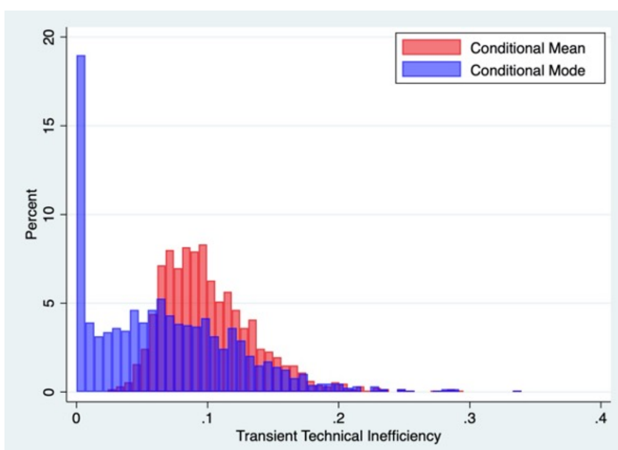


Figure 2.6: Distribution of Transient Technical Inefficiency in ELA

(18%), followed by the Bronx (17%), Queens (13%), and then Staten Island (10%). Looking at the trend over time, we find middle schools in the Bronx show consistent improvements in the percentage of schools operating efficiently, while other boroughs do not have consistent improvements in efficiency over the period. The share of middle schools in the Bronx with zero TTI increases from 49% to 60% to 67% in Math and 11% to 16% to 24% in ELA. Schools in Queens, on the other hand, are less likely to operate with zero TTI in Math over time (71% to 65% to 45%) with no consistent trends in ELA (20% to 6% to 11%). All other boroughs also do not display consistent positive or negative trends in TTI[22].

Table 2.5: Percentage of Zero-Mode Transiently Efficient Schools by Subject, Borough and Year.

	Math				ELA			
	AY2014	AY2015	AY2016	Average	AY2014	AY2015	AY2016	Average
Manhattan	50%	70%	55%	58%	11%	30%	13%	18%
The Bronx	49%	60%	67%	59%	11%	16%	24%	17%
Brooklyn	62%	52%	50%	55%	17%	19%	17%	18%
Queens	71%	65%	45%	60%	20%	6%	11%	13%
Staten Island	62%	77%	46%	62%	8%	15%	8%	10%
Total	58%	61%	54%	58%	14%	18%	17%	16%

Figure 2.7 shows a weak but positive relationship between Math TTI and ELA TTI during our sample period (the slope of the line is 0.56 with a t-statistic of 12.62). The Spearman test statistics each year range between 0.2834 and 0.3312 and are statistically significant.

2.4.4 Efficiency Probabilities

As suggested previously, the above-described ad hoc rule to identify efficient school-year observations lacks statistical rigor. Therefore, we calculate school-level efficiency probabilities Horrace (2005) to identify the subset of schools that operate efficiently each year in terms of

[22]While it is tempting to compare the magnitudes of TTI in Table 2.4 to the PTI in Table 2.3, the reader is reminded that PTI may also contain other sources of time-invariant unobserved heterogeneity, so comping TTI to PTI is ill-advised in general.

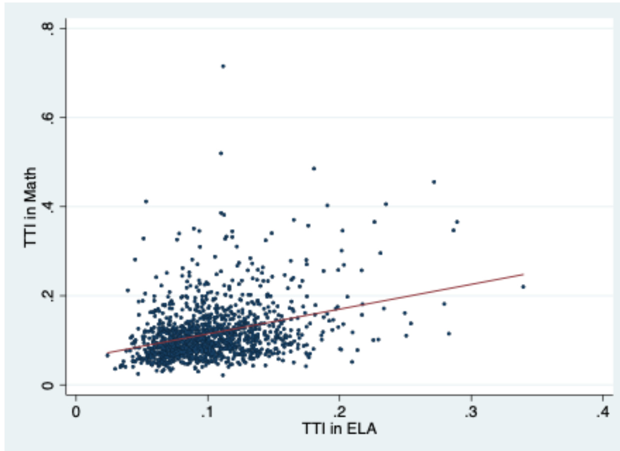


Figure 2.7: Correlation of Transient Technical Inefficiency between Math and ELA

TTI (Flores-Lagunes et al., 2007). A more thorough discussion of the technique is contained in the Appendix, but as stated earlier, it uses the conditional truncated normal distribution of TTI for each school to calculate the probability that each school is efficient in each year (has the smallest u conditional on the data), then selects a minimum cardinality subset of schools that contain the efficient school with at least 95% probability. This is a rigorous statistical decision rule, and we have two goals in the analyses that follow. The first is to calculate this minimal cardinality "subset of the best" schools in each year at the 95% level, and the second is to compare the cardinality of the subset of the best in each year to the cardinality of the subset of zero-mode schools in each year.

Table 2.6 contains the results. The first row of the table is for Math TTI in AY 2014. In 2014 there were 246 schools (3rd column) with conditional modes equal to zero. Since we have a balanced panel of 425 schools, this means that 58% percent of schools are efficient in 2014 based on our ad hoc decision rule, and this number corresponds to the 58% in the last row, 2nd column of Table 2.5. Call these 246 schools the "zero-mode subset" of efficient schools. The 4th column of Table 2.6 is the sum of the efficiency probabilities for the schools in the zero-mode subset. That is, the probability that the most efficient firm in the sample is contained in the zero-mode subset is 75.22% in 2014 for Math TTI. Put another way, the schools in the zero-mode subset are statistically indistinguishable from the unknown efficient school with probability 75.22%. Thus, the efficiency probabilities allow us to assign a confidence level to our ad hoc decision rule, however it is well below typical confidence levels like 90% or 95%. Nonetheless there is policy value in knowing the zero-mode subset. The second and third rows of Table 2.6 show that the cardinalities of the zero-mode subsets (3rd column) in AY 2015 and 2016 (259 and 231 schools respectively) are similar to AY 2014, as are the zero-mode probabilities (4th column), which are 76.10% and 73.09%, respectively. In sum, we are about 75% confident that about 60% of NYC schools are likely operating efficiently in terms of Math TTI.

Table 2.6: Subsets of the Best Schools in terms of TTI

Test	Cohort	Zero-Mode School Count	Zero-Mode Probabilities	Schools in 95% Best Subset
Math	AY2014	246	75.22%	372
	AY2015	259	76.10%	374
	AY2016	231	73.09%	369
ELA	AY2014	61	32.30%	334
	AY2015	75	37.97%	339
	AY2016	72	40.77%	326

Zero-mode probabilites are probabilites that the zero-mode subset contains the transient technical efficient school in the sample.

Moving to the ELA test results in Table 2.6 (rows 4-6), we see that the cardinalities of the zero-mode subsets (3rd column) are 61, 75, 72 with probabilities (4th column) 32.30%, 37.97% and 40.77% in AY 2014, 2015 and 2016, respectively. These are much lower cardinalities and probabilities than the Math TTI results, and this is reflected in the smaller blue spike at zero in compared to Figure 2.5 and 2.6 (which have different scales). Why might this be the case? Is there something inherent in ELA education that lends itself to lower efficiencies relative to Math education? It is, in fact, a commonly found empirical phenomenon that ELA achievement is more reflective of home environment and individual characteristics, while Math performance is more responsive to classroom characteristics - a finding that is also consistent with theory about the pedagogy of ELA and Math instruction (Bryk and Raudenbush, 1989). It is noteworthy too, that our results in Table 2.2 also show test-taker characteristics are more predictive of greater ELA gains, while teacher characteristics are correlated with greater Math gains. Keep in mind, however, that we are assuming that the distribution of Math TTI is half-normal and that of ELA TTI is exponential, and this was driven by nonconvergence of the likelihood maximization in alternative specifications. Aside from this technical detail, it may simply be that mathematical standards for "correctness" are objective and those for the language arts are more subjective, so identifying "best practices" in ELA may be more difficult than in Math. There is a branch of SFA that attempts to explore the determinants of technical inefficiency (Cho and Schmidt, 2020). Perhaps such an analysis may be helpful here, but this is left for future research.

The 5th column of Table 2.6 contains the cardinality of the 95% minimal cardinality subset of the best (Flores-Lagunes et al., 2007), and in the first row we see that for Math TTI in 2014, 372 of our 425 middle schools were indistinguishable from the best middle school in the sample with 95% probability. Here, and in the other rows of the table, the zero-mode subset is always a proper subset of the subset of the best. (Whether this is a coincidence or not remains to be seen and is left for future research.) The implication is that even if the zero-mode decision rule is ad hoc and does not achieve usual confidence levels, at least it identifies a subset of schools that are contained in the subset of the best, as based on a rigorous statistical decision rule.

Looking down the 5th column across academic years for Math TTI, 369 to 374 (depending on year) of the 425 schools are statistically indistinguishable from efficient at the 95% level. For ELA, 326 to 339 of 425 schools are statistically indistinguishable from efficient. These are useful statistics that policymakers may use to determine which and how many schools to target when designing interventions that are intended to improve performance.

2.5 Conclusion

This study provides summaries of persistent and transient technical efficiency estimates for each of 425 NYC middle schools using recent advancements in stochastic frontier modeling. Using the "true" fixed effect stochastic frontier model to estimate gains in mathematics (Math) and English language arts (ELA), we find substantial variation in persistent technical inefficiency across the city and between boroughs. We note that while some boroughs have higher shares of persistently inefficient schools than others, there is a wide and overlapping distribution across each of the five boroughs in the city, suggesting school efficiency in NYC is not a "tale of two cities". Thus, while the mean Math and ELA persistent technical inefficiency in the city are 0.99 and 0.53 standard deviations, respectively - both larger than the student achievement gap between schools in the borough that enrolls the highest performing students (Staten Island) and schools in the borough that teaches the lowest performing students (the Bronx) - school inefficiency itself is widely distributed across the NYC's boroughs and schools. Still, to give a sense of magnitude of the results, if the city could find a way to remove persistent technical inefficiency in schools in the Bronx, for example, it would eliminate the achievement gap across boroughs (and, in fact, even overshoot the target).

We next produce estimates of transient technical inefficiency, using both a conditional mean and a conditional mode estimator. Under the conditional mode estimator and an ad hoc decision rule, we find around 58% of schools are transiently technically efficient in Math and 16% in ELA. We then apply a probability statement approach to offer rigorous inferential insights on which school-years are statistically on the efficient frontier, and which are very likely not. Based on the results of our "zero-mode subset" and the minimal cardinality subset of the best, the model can be used for both subjects to provide substantial information to decisionmakers on which schools likely did and did not operate efficiently each year. These are important distinctions for policymakers to be able to make; for example, the difference in the mean achievement gains for students attending a school-year observation in the zero-mode subset in ELA using the conditional mode is estimated to make 6.2 percent of a standard deviation greater gains than if that school were operating at the median level of inefficiency for ELA in that year (equivalent to about 15% of the gap between mean grade 8 achievement in Staten Island and the Bronx).

Another innovation of this study is the use of student-level academic performance data to estimate gains over time, which are then aggregated to the cohort-school-level to more accurately measure the education produced during the middle school years. These sorts of "gains models" are common in other education research but have not yet been used in stochastic frontier modelling. This innovation allows for improved estimates of the marginal effects of student, teacher, and school inputs on education production as well as a more compelling methodology for determining which schools are persistently efficient in each year.

Our results suggest that policymakers should more rigorously consider the role inefficiency plays in reducing education production in public schools. First, we identify which features (and types of features) in the school environment are beneficial or harmful to education production in middle schools, interestingly finding that student composition of a school is more important for the production of ELA gains, while teacher composition of a school is more important for the production of Math gains. These results are consistent with the conventional wisdom that ELA achievement is more responsive to home and individual characteristics and Math achievement is more responsive to classroom characteristics (Bryk and Raudenbush, 1989).

Second, by separating persistent technical inefficiency from transient technical inefficiency, we offer a methodology for school district administrators to separate the systemic features of a school that harm efficiency (such as, perhaps, building or principal quality) from those that change perennially (such as, perhaps, teacher effort or curriculum design). Third, by separating inefficiency from the error term (under the above-described distributional assumptions), decisionmakers are better able to assess the extent to which declining exam performance is due to inefficiency as opposed to statistical noise.

Arming policymakers with actionable decision rules that are methodologically rigorous and reflect the true performance of schools is a tall ask of any statistical model. By applying the true fixed effect model with both the conditional mode estimator and the more rigorous Horrace (2005) probabilities, this paper expands the tool kit of policymakers and illustrates how to apply those tools to measure inefficiency in education. To avoid making an overstatement of the implications of this study, it is important to recognize that the methodologies in this paper, while can address some of the shortcomings of previous work such as separating persistent from transient inefficiencies, are no exceptions to having limitations: on top of the distribution assumptions made when estimating the model, the intuition-driven linear functional form, the limited data availability, and regression to the mean all could potentially influence the validity of our findings. Addressing those concerns requires future efforts such as cross-validating with alternative methodologies including value-added models (Ladd and Walsh, 2002; Meyer, 1997) and DEA (Farrell, 1957; Charnes et al., 1978), exploring how regression to the mean affects stochastic frontier models, conducting field experiments and qualitative research like focus groups among the educators and students, and others. These efforts may also allow for better targeting of policy levers towards disincentivizing each form of inefficiency.

2.6 Appendix

2.6.1 The Conditional Mean Estimator and the Conditional Mode Estimator

When v is normal and u is half-normal, the model is Normal-Half Normal (NHN). When u is exponential, the model is Normal-Exponential (NE). Per Jondrow et al. (1982), the closed-form expressions of the conditional mean under normal-half normal and normal-exponential assumptions are:

$$E\left(u_{it} \mid e_{it}, NHN\right) = \hat{\sigma}_* \left[\frac{\phi\left(\frac{e_{it}\hat{\lambda}}{\hat{\sigma}}\right)}{1 - \Phi\left(\frac{e_{it}\hat{\lambda}}{\hat{\sigma}}\right)} - \left(\frac{e_{it}\hat{\lambda}}{\hat{\sigma}}\right) \right], \tag{2.3}$$

$$E\left(u_{it} \mid e_{it}, NE\right) = \hat{\sigma}_v \left[\frac{\phi(\hat{A})}{1 - \Phi(\hat{A})} - \hat{A} \right] \tag{2.4}$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\sigma_*^2 = \sigma_u^2\sigma_v^2 / \left(\sigma_u^2 + \sigma_v^2\right)$, $\lambda = \frac{\sigma_u}{\sigma_v}$ and $A = \frac{\varepsilon_{it}}{\sigma_v} + \frac{\sigma_v}{\sigma_u}$. ϕ and Φ are the probability density function and cumulative distribution function of standard normal distribution. Estimates are formed by substituting the MMSLE estimates for their population parameters into these formulae while setting $\varepsilon_{it} = e_{it}$.

A less commonly employed estimator proposed by Jondrow et al. (1982) is the mode of the conditional distribution of $u_{it}|\varepsilon_{it}$, denoted as $M(u_{it}|\varepsilon_{it})$, to measure transient technical inefficiency. Under normal-half normal and normal-exponential distribution assumptions, the conditional mode estimator can be written as:

$$M\left(u_{it} \mid e_{it}, NHN\right) = \begin{cases} -e_{it}\left(\frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}\right), & \text{if } e_{it} \leq 0 \\ 0, & \text{if } e_{it} > 0 \end{cases} \tag{2.5}$$

$$M\left(u_{it} \mid e_{it}, NE\right) = \begin{cases} -e_{it} - \frac{\hat{\sigma}_v^2}{\hat{\sigma}_u}, & \text{if } e_{it} \leq -\frac{\hat{\sigma}_v^2}{\hat{\sigma}_u} \\ 0, & \text{if } e_{it} > -\frac{\hat{\sigma}_v^2}{\hat{\sigma}_u} \end{cases} \tag{2.6}$$

The parametric forms of both conditional mean and conditional mode estimators under NHN and NE are functions of e_{it}. To better understand the differences between the conditional mean and the conditional mode estimators, we standardize the standard errors σ_v and σ_u to one and plot their relationships with e_{it} under NHN in Figure 2.1 and under NE in Figure 2.2. The figures show that both conditional mean and conditional mode estimators are monotonically decreasing with the regression residual. The conditional mode estimator, however, is always below the conditional mean estimate given the same residual. Moreover, when the residual surpasses a threshold (0 under NHN or $-\sigma_v^2/\sigma_u$ under NE), the conditional mode estimator takes a value of zero whereas the conditional mean estimator is positive and monotonically decreasing. This is intuitive - the more negative the regression residual, the farther the school is below that frontier and the more likely it is to be operating with large inefficiency. When the regression residual is large and positive, the school's estimated productivity is above the production frontier, suggesting the inefficiency is likely to be small. The difference between the estimators, then, is that, when above the threshold, the estimated TTI using the conditional mean estimator is small but still positive, whereas using the conditional mode estimator is zero. We use this conditional mode property to identify "zero-mode" schools that are likely to be operating efficiently.

Similar to the conditional mean estimator, the conditional mode estimator can be used to rank schools. However, unlike the conditional mean, the ranking allows for ties if more than one school is estimated to have zero TTI. Among schools with positive conditional mode estimates (non-zero estimated inefficiency), however, the order of the rankings is the same as from the conditional mean.

2.6.2 Conditional Efficiency Probabilities and the Subset of the Best Schools

While conditional mean estimates can be used to rank schools and conditional mode estimates can be used to find zero-mode efficient schools, neither estimate can produce joint probability statements on the relative ranking of the schools. To assess the reliability of the results and to draw inference on the efficiency rankings, we turn to the probability statement approach (Horrace, 2005; Flores-Lagunes et al., 2007; Horrace and Richards-Shubik, 2012; Horrace et al., 2015). Assuming independence of u over i and t, the probability of the event "school i is efficient at time t" is:

$$\pi_{it} = P\left\{u_{it} \leq u_{jt}\ \forall i \neq j \mid \varepsilon_{1t}, \ldots, \varepsilon_{nt}\right\} = \int_0^{\infty} f_{u_{it}|\varepsilon_{it}}(u) \prod_{j \neq i}^{n} \left[1 - F_{u_{jt}|\varepsilon_{jt}}(u)\right] du \qquad (2.7)$$

where $f_{u_{it}|\varepsilon_{it}}(u)$ and $Ff_{u_{it}|\varepsilon_{it}}(u)$ are the probability density function and cumulative distribution function of $u_{it}|\varepsilon_{it}$, respectively. If u is half-normal with variance σ_u^2, then $u_{it}|\varepsilon_{it}$ is $N^+\left(-\frac{\varepsilon_{it}\sigma_u^2}{\sigma_u^2+\sigma_u^2}, \frac{\sigma_u^2\sigma_v^2}{\sigma_u^2+\sigma_n^2}\right)$. If u is exponential, then $u_{it}|\varepsilon_{it}$ is $N^+\left(-\varepsilon_{it} + \sigma_v^2/\sigma_u, \sigma_v^2\right)$. To estimate the probabilities π_{it}, the regression residuals, e_{it}, are substituted into the above formulas for errors, ε_{it}. Then, given any subset of the n schools (like our zero-mode subset), we can assign a confidence level to the set containing the efficient school by summing the probabilities π_{it} for the schools in the set. Alternatively, let the population rankings of the unknown efficiency probabilities be,

$$\pi_{[n]t} > \pi_{[n-1]t} > \cdots > \pi_{[1]t} \qquad (2.8)$$

and let the sample rankings of the estimated probabilities, $\hat{\pi}_{it}$, be

$$\hat{\pi}_{(n)t} > \hat{\pi}_{(n-1)t} > \cdots > \hat{\pi}_{(1)t} \qquad (2.9)$$

where $[i] \neq (i)$ in general. We can determine a 95% minimal cardinality subset of the best school by summing the probabilities from the largest (n) to the smallest (1) until the sum is at least 0.95. Then, the school indices in the sum are "in contention for the best school" with probability at least 95% at time t. In other words, these schools cannot be statistically distinguished from the (unknown) best school in the population, $[n]$. For example, if $\hat{\pi}_{(n)t} > 0.95$, then the minimal cardinality subset is a singleton containing only the index (n), and the inference is very sharp. If $\hat{\pi}_{(n)t} < 0.95$, but $\hat{\pi}_{(n)t} + \hat{\pi}_{(n-1)t} > 0.95$ (say), then the minimal cardinality subset is $\{(n), (n-1)\}$. It is possible that the subset contains all n schools, $\{(n), (n-1), \ldots, (1)\}$. This occurs when $\sum_{i=1}^{n-1} \hat{\pi}_{(i)t} < 0.95$ or equivalently when $\hat{\pi}_{(1)t} > 1 - 0.95$.

Bibliography

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6(1):21 – 37.

Angrist, J. D., Hull, P. D., Pathak, P. A., and Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. The Quarterly Journal of Economics, 132(2):871–919.

Angrist, J. D. and Pischke, J.-S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Baron, E. J. (2019). School spending and student outcomes: Evidence from revenue limit elections in wisconsin. Available at SSRN 3430766.

Belotti, F. and Ilardi, G. (2018). Consistent inference in fixed-effects stochastic frontier models. Journal of Econometrics, 202(2):161–177.

Bera, A. K. and Sharma, S. C. (1999). Estimating Production Uncertainty in Stochastic Frontier Production Function Models. Journal of Productivity Analysis, 12(3):187–210.

Bryk, A. S. and Raudenbush, S. W. (1989). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In Multilevel analysis of educational data, pages 159–204. Elsevier.

Chakraborty, K., Biswas, B., and Lewis, W. C. (2001). Measurement of Technical Efficiency in Public Education: A Stochastic and Nonstochastic Production Function Approach. Southern Economic Journal.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. European journal of operational research, 2(6):429–444.

Chen, Y.-Y., Schmidt, P., and Wang, H.-J. (2014). Consistent estimation of the fixed effects stochastic frontier model. Journal of Econometrics, 181(2):65–76.

Cho, C.-K. and Schmidt, P. (2020). The wrong skew problem in stochastic frontier models when inefficiency depends on environmental variables. Empirical Economics, 58(5):2031–2047.

De Witte, K. and López-Torres, L. (2017). Efficiency in education: a review of literature and a way forward. Journal of the Operational Research Society, 68(4):339–363.

Dolton, P., Marcenaro, O. D., and Navarro, L. (2003). The effective use of student time: a stochastic frontier production function case study. Economics of Education review, 22(6):547–560.

Donoho, D. L., Johnstone, I. M., Hoch, J. C., and Stern, A. S. (1992). Maximum entropy and the nearly black object. Journal of the Royal Statistical Society. Series B (Methodological), 54(1):41–81.

Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. biometrika, 81(3):425–455.

Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society. Series A (General), 120(3):253–290.

Flores-Lagunes, A., Horrace, W. C., and Schnier, K. E. (2007). Identifying technically efficient fishing vessels: a non-empty, minimal subset approach. Journal of Applied Econometrics, 22(4):729–745.

Garcia-Diaz, R., del Castillo, E., and Cabral, R. (2016). School competition and efficiency in elementary schools in mexico. International Journal of Educational Development, 46:23–34.

González-Farias, G., DomıÌĄnguez-Molina, A., and Gupta, A. K. (2004). Additive properties of skew normal random vectors. Journal of statistical planning and inference, 126(2):521–534.

Greene, W. (2005a). Fixed and random effects in stochastic frontier models. Journal of productivity analysis, 23(1):7–32.

Greene, W. (2005b). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of econometrics, 126(2):269–303.

Greene, W. H. (1980a). Maximum likelihood estimation of econometric frontier functions. Journal of econometrics, 13(1):27–56.

Greene, W. H. (1980b). On the estimation of a flexible frontier production model. Journal of Econometrics, 13(1):101–115.

Greene, W. H. (1990). A gamma-distributed stochastic frontier model. Journal of Econometrics, 46(1):141 – 163.

Gronberg, T. J., Jansen, D. W., and Taylor, L. L. (2012). The relative efficiency of charter schools: A cost frontier approach. Economics of Education Review, 31(2):302–317.

Grosskopf, S., Hayes, K. J., and Taylor, L. L. (2014). Efficiency in education: Research and implications. Applied Economic Perspectives and Policy, 36(2):175–210.

Hanushek, E. A. and Luque, J. A. (2003). Efficiency and equity in schools around the world. Economics of education Review, 22(5):481–502.

Helmstadter, G. C. and Walton, M. (1985). The generalizability of residual indexes of effective schooling. In annual meeting of the American Educational Research Association, Chicago.

Horrace, W. C. (2005). On ranking and selection from independent truncated normal distributions. Journal of Econometrics, 126(2):335 – 354. Current developments in productivity and efficiency measurement.

Horrace, W. C. and Parmeter, C. F. (2018). A laplace stochastic frontier model. Econometric Reviews, 37(3):260–280.

Horrace, W. C. and Richards-Shubik, S. (2012). A monte carlo study of ranked efficiency estimates from frontier models. Journal of Productivity Analysis, 38(2):155–165.

Horrace, W. C., Richards-Shubik, S., and Wright, I. A. (2015). Expected efficiency ranks from parametric stochastic frontier models. Empirical Economics, 48(2):829–848.

Horrace, W. C. and Schmidt, P. (1996). Confidence statements for efficiency estimates from stochastic frontier models. Journal of Productivity Analysis, 7(2-3):257–282.

Horrace, W. C. and Schmidt, P. (2000). Multiple comparisons with the best, with economic applications. Journal of Applied Econometrics, 15(1):1–26.

Horrace, W. C. and Wright, I. A. (2016). Stationary points for parametric stochastic frontier models.

Johnes, J. (2004). 16 efficiency measurement. International handbook on the economics of education, page 613.

Jondrow, J., Lovell, C. K., Materov, I. S., and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, 19(2):233 – 238.

Kang, B.-G. and Greene, K. V. (2002). The effects of monitoring and competition on public education outputs: A stochastic frontier approach. Public Finance Review, 30(1):3–26.

Kirjavainen, T. (2012). Efficiency of finnish general upper secondary schools: An application of stochastic frontier analysis with panel data. Education Economics, 20(4):343–364.

Koedel, C., Mihaly, K., and Rockoff, J. E. (2015). Value-added modeling: A review. Economics of Education Review, 47:180–195.

Kumbhakar, S. C. and Lovell, C. K. (2003). Stochastic frontier analysis. Cambridge university press.

Kumbhakar, S. C., Parmeter, C. F., and Tsionas, E. G. (2013). A zero inefficiency stochastic frontier model. Journal of Econometrics, 172(1):66 – 76.

Ladd, H. F. and Walsh, R. P. (2002). Implementing value-added measures of school effectiveness: getting the incentives right. Economics of Education review, 21(1):1–17.

Lancaster, T. (2000). The incidental parameter problem since 1948. Journal of econometrics, 95(2):391–413.

Luyten, H. (1994). Stability of school effects in dutch secondary education: The impact of variance across subjects and years. International journal of educational research, 21(2):197–216.

Mandeville, G. K. and Anderson, L. W. (1987). The stability of school effectiveness indices across grade levels and subject areas. Journal of educational measurement, 24(3):203–216.

Materov, I. (1981). On full identification of the stochastic production frontier model. Ekonomika i matematicheskie metody, 17:784–788.

Matthews, T. A. et al. (1981). Use of districtwide test scores to compare the academic effectiveness of schools.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. International Economic Review, 18(2):435–44.

Meyer, R. H. (1997). Value-added indicators of school performance: A primer. Economics of Education Review, 16(3):283–301.

Mizala, A., Romaguera, P., and Farren, D. (2002). The technical efficiency of schools in chile. Applied Economics, 34(12):1533–1552.

Mokhtar, H. S. A., Abdullah, N., and Al-Habshi, S. M. (2006). Efficiency of islamic banking in malaysia: A stochastic frontier approach. Journal of Economic Cooperation, 27(2):37–70.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. Econometrica: Journal of the Econometric Society, pages 1–32.

Nguyen, N. B. (2010). Estimation of technical efficiency in stochastic frontier analysis. PhD thesis, Bowling Green State University.

Ondrich, J. and Ruggiero, J. (2001). Efficiency measurement in the stochastic frontier model. European Journal of Operational Research, 129(2):434–442.

Park, T. and Casella, G. (2008). The bayesian lasso. Journal of the American Statistical Association, 103(482):681–686.

Pereira, M. C. and Moreira, S. (2007). A stochastic frontier analysis of secondary education output in portugal. Available at SSRN 1398692.

Salas-Velasco, M. (2020). Assessing the performance of spanish secondary education institutions: Distinguishing between transient and persistent inefficiency, separated from heterogeneity. The Manchester School, 88(4):531–555.

Schochet, P. Z. and Chiang, H. S. (2013). What are error rates for classifying teacher and school performance using value-added models? Journal of Educational and Behavioral Statistics, 38(2):142–171.

Simar, L. and Wilson, P. (2010). Inferences from cross-sectional, stochastic frontier models. Econometric Reviews, 29(1):62–98.

Stevens, P. A. (2005). A stochastic frontier analysis of english and welsh universities. Education Economics, 13(4):355–374.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics, 13(1):57 – 66.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1):267–288.

Waldman, D. M. (1984). Properties of technical efficiency estimators in the stochastic frontier model. Journal of econometrics, 25(3):353–364.

Wang, W. S. and Schmidt, P. (2009). On the distribution of estimated technical efficiency in stochastic frontier models. Journal of Econometrics, 148(1):36–45.

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. Education economics, 9(3):245–268.

Zoghbi, A. C., Rocha, F., and Mattos, E. (2013). Education production efficiency: Evidence from brazilian universities. Economic Modelling, 31:94–103.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476):1418–1429.

Yi Yang

Department of Economics
426 Eggers Hall
Syracuse, NY, 13203

(315) 876-5960
yyang64@syr.edu

EDUCATION

2016 - 2021 Syracuse University, Ph.D. in Economics.
2012 - 2016 Nanyang Technological University, B.S. in Math & Economics.

PROFESSIONAL EXPERIENCE

June - Aug 2020 Amazon, Economist Intern
July - Sep 2019 Amazon, Economist Intern

RESEARCH

Working Paper

2021 W.C. Horrace, M.W. Rothbart, Y. Yang, *Technical Efficiency of Public Middle Schools in New York City*
2019 W.C. Horrace, H. Jung, Y. Yang, *Estimation of Technical Inefficiency using Mode in Parametric Stochastic Frontier Model.*
2016 Z. Hu, Z, Long, Y. Yang, *An Economic Model of Shadow Banking: Double Tranches and Heterogeneous Risk Tolerance*

GRANTS AND AWARDS

2018 - 2020 Research Excellence Doctoral Fellowship
2016 - 2017 Syracuse University Economics Teaching Fellowship
2016 Best Thesis of the Year Award from Ministry of Trade and Industry
2015 Case Competition Championship - Bain & Company
2012 NTU Science and Engineering Undergraduate Scholarship.

CONFERENCE TALKS

2018 Syracuse University Econometrics Luncheon.
2016 MTI Economic Dialogue.

TEACHING

2019 Teaching Assistant of Introductory Macroeconomics.
2017-2018 Teaching Assistant of Economic Issues and Ideas.
2016 Teaching Assistant of Intermediate Macroeconomics.