Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

BACHELOR THESIS

Efficiency Analysis of Grammar Schools in the Czech Republic

Author: **Kateřina Červenková**

Supervisor: **PhDr. Lenka Šťastná, Ph.D.**

Academic Year: **2012/2013**

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature. The author also declares that she has not used this thesis to acquire another academic degree.

The author grants to Charles University the permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 16, 2013

Signature

Acknowledgments

I would like to express my deepest gratitude to my thesis supervisor, PhDr. Lenka Šťastná, Ph.D., for her valuable comments and suggestions. I would also like to thank PhDr. Ladislav Krištoufek for answering my queries relating to the econometric part.

Bibliographic entry

Červenková, K. (2013): "Efficiency Analysis of Grammar Schools in the Czech Republic." *(Unpublished bachelor thesis).* Charles University in Prague. Supervisor: PhDr. Lenka Šťastná, Ph.D.

Length: 63 719 characters

Abstract

This thesis aims to assess the technical efficiency of grammar schools in the Czech Republic. We analyze a sample of 263 grammar schools (out of 362) and use data of cohort that graduated in 2012. We adopt a two-stage approach to efficiency analysis. In the first stage, efficiency scores are obtained through the data envelopment analysis under different returns to scale assumptions. The average efficiency of Czech grammar schools is 92% if variable returns to scale are considered.

In the second stage, the efficiency is related to school and environmental characteristics using a Tobit regression. Our results suggest that the percentage of female students has a positive effect on efficiency as well as the school size and the share of students attending the 6 or 8 year study program. On the contrary, schools offering a vocational program along with a grammar school program are found to be less efficient. We have not found evidence that the share of not qualified teachers, yearly salary per teacher, state/private status, size of the town in which a school is located or any other environmental variable affect efficiency.

JEL Classification	F12, C14, H52, I21
Keywords	Efficiency, grammar schools, DEA

Author's e-mail	`cervenkova.katerina@gmail.com`
Supervisor's e-mail	`stastna@fsv.cuni.cz`

Abstrakt

Tato práce si klade za cíl zhodnotit technickou efektivitu gymnázií v České republice. Analyzujeme vzorek 263 gymnázií (z celkového počtu 362) a používáme data o studentech, kteří odmaturovali v roce 2012. Volíme dvoufázový přístup k analýze efektivity. Nejprve získáme pomocí obálkové metody dat skóre efektivity za různých předpokladů o výnosech z rozsahu. Zjišťujeme, že gymnázia v průměru dosahují 92% efektivity za předpokladu proměnlivých výnosů z rozsahu.

V druhé fázi se snažíme pomocí Tobitova regresního modelu vysvětlit efektivitu pomocí dalších charakteristik školy a oblasti, ve kterém se škola nachází. Došli jsme k závěru, že procento studentek, velikost školy a procento studentů víceletých gymnaziálních oborů má kladný vliv na efektivitu. Naopak, školy, které kromě gymnaziálních studijních oborů nabízejí navíc odborně zaměřené obory zakončené maturitní zkouškou, dosahují nižší efektivity. Nepotvrdilo se, že by procento kvalifikovaných učitelů, plat učitelů, státní/soukromý charakter školy či velikost města, ve kterém se škola nachází, měly vliv na efektivitu školy.

Klasifikace JEL F12, C14, H52, I21
Klíčová slova Efektivita, gymnázia, DEA

E-mail autora `cervenkova.katerina@gmail.com`
E-mail vedoucího práce `stastna@fsv.cuni.cz`

Contents

List of Tables

List of Figures

Acronyms

CRS Constant Returns to Scale

DEA Data Envelopment Analysis

DMU Decision Making Unit

FDH Free-Disposal Hull

GDP Gross Domestic Product

MF Ministry of Finance

MLE Maximum Likelihood Estimator

MSMT Ministry of Education

OLS Ordinary Least Squares

OECD Organisation for Economic Co-operation and Development

SFA Stochastic Frontier Analysis

VRS Variable Returns to Scale

CSU Czech Statistical Office

Chapter 1

Introduction

> "Our progress as a nation can be no swifter than our progress in education. The human mind is our fundamental resource."
>
> John F. Kennedy

It is essential for every country to have a steady supply of skilled labor force to move its economy forward and beyond any doubts, education plays a crucial role in forming such labor force. Consequently, education system and educational attainment of a country's population is under scrutiny by researches as well as subject to international comparison.

Expenditure as a percentage of GDP on educational institutions in the Czech Republic grew since 2000 by 1 percentage point reaching approximately 5% in 2009 (OECD 2012a). However, the educational attainment of Czech 15-year-old pupils declined over the same period according to Program for International Student Assessment (OECD 2010). Reasons for decline in educational attainment while the funds are pumped into the education system may be numerous, one of which may be inefficient educational process of individual schools. In other words, schools may not employ their resources efficiently and do not maximize their potential. Even though schools are in many aspects different from production firms, some methods from operational research, that mainly focuses on production processes of profit making entities, may be applied with caution to providers of education.

The objective of this thesis is to measure and analyze efficiency of grammar schools in the Czech Republic. We focus on this particular type of secondary schools as grammar schools offering comprehensive education are the gateway to tertiary education and the share of students entering a grammar school program has risen steadily since 2002 (Chamoutová & Vojtěch 2013). Furthermore, grammar schools are more or less similar in curriculum which makes the comparison of schools' achievement as well as the amounts of their resources plausible.

We perform efficiency analysis of Czech grammar schools based on a two-stage approach. Firstly, we obtain efficiency scores of individual schools through the data envelopment analysis (DEA). We benefit from the recent introduction of compulsory standardized final school-leaving exams, results of which we include in DEA as an output. To the best of our knowledge no efficiency analysis of Czech grammar schools taking into consideration results of these exams has been published. In the second stage, efficiency scores are regressed on several school and environmental characteristics not appearing in DEA. Tobit regression model is adopted for the second stage due to the nature and distribution of efficiency scores.

The thesis is structured as follows: Chapter 2 presents an overview of methods employed in efficiency assessment and their development. Theoretical underpinnings of DEA and a Tobit regression are described in Chapter 3. Chapter 4 is devoted to description of the data set used for empirical analysis. We present the results of our empirical analysis in Chapter 5. Chapter 6 summarizes our findings.

Chapter 2

Literature Review

Ways how the resources are transformed into the final products usually differ significantly across companies and institutions. Therefore, efforts have been made to develop a method for assessment of these production processes that serve as a basis for comparison of performance of similar companies or institutions or, more generally, similar decision making units (DMUs). Obviously, a production unit should aim for high outputs while consuming as little inputs as possible. The question then arises as how a production process of a unit can be assessed and compared to a production process of similar units? There does not exist a trivial answer especially in case the units under observation employ different mix of inputs to produce a different mix of outputs.

2.1 Efficiency Analysis Methodology Development

Substantive contribution in the field of productivity analysis is attributed to Farrell (1957) for introduction of a more satisfactory measurement of productive efficiency than it was available at that time. The work of Farrell (1957) considers mainly the case of multiple inputs and one single output, which was previously dealt with by creating an index of inputs through assigning the predetermined weights to inputs. Subsequently, the unit with the highest output to input index ratio would be considered to be the most efficient. Farrell's intention is to overcome the problem of indexing as setting of uniform weights may not be fair and would lead to the situation that the input index of some units may decrease if the arbitrarily set weights are changed which, in turn, causes the output to input index ratio to increase. Farrell's approach consists of determination of an efficiency frontier, formed by the best performing observed units, being the benchmark against which other observations are compared. The distance of a unit of observation from the efficiency frontier represents inefficiency. The output-oriented efficiency frontier represents the highest output relative to given level of input. Similarly, input-oriented efficiency frontier represents the lowest input relative to given output. Therefore, efficiency can be measured from input or output point of view. A unit that is assessed as inefficient from the output perspective should produce more output given its input. The unit inefficient from the input point of view should reduce its input while maintaining the level of its output.

As the efficiency frontier is based on empirical data rather than on a theoretical ideal production process, relative efficiency is assessed instead of absolute efficiency. Cooper *et al.* (2004) suggests the definition of relative efficiency as follows:

> A DMU is to be rated as fully (100%) efficient on the basis of available evidence if and only if the performances of other DMUs does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs. (Cooper *et al.* 2004, p. 3)

Therefore, if we use the term efficiency throughout this thesis we refer to the relative efficiency if not stated otherwise.

The concept of efficiency measurement of similar decision making units was further developed by Charnes *et al.* (1979), who introduced efficiency measure as a ratio of weighted inputs to weighted outputs, known as CCR ratio, where the weights specific for each observation are set so that the ratio is maximized. Furthermore, they referred to the method of technical efficiency analysis as the data envelopment analysis (DEA), an expression under which this method is known until now.

Banker *et al.* (1984) relaxed the assumption of constant returns to scale and provided a more suitable method to calculate the efficiency scores, while allowing for the variable returns to scale, than Farrell (1957). For a comprehensive review of reactions and further developments of the method presented by Farrel see Foersund & Sarafoglou (2000).

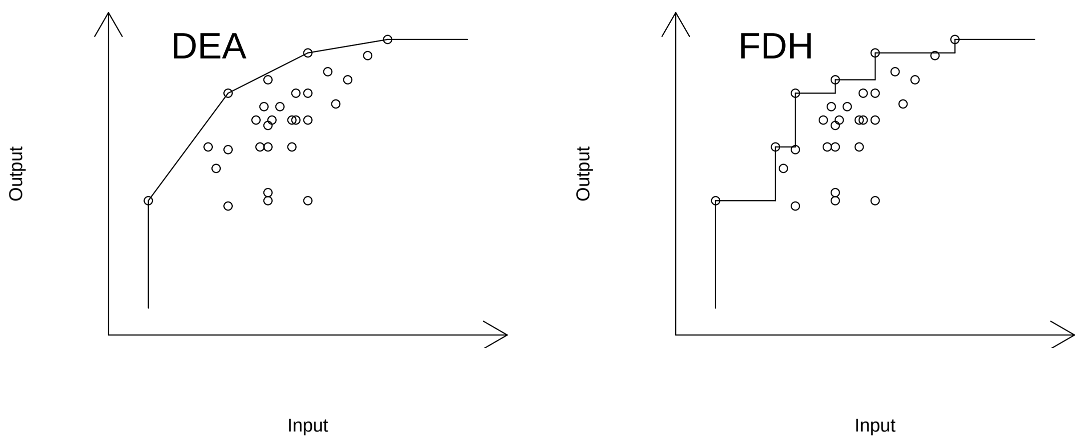
Farrell (1957) also introduced the concept of overall efficiency consisting of technical and price (allocative) efficiency. Technical efficiency relates to the distance of a unit from the efficiency frontier, while the allocative efficiency takes into consideration the value of inputs and outputs. Units located on the efficiency frontier, therefore being technically fully efficient, may use a different mix of inputs to produce various combinations of outputs in their production.

The allocative efficiency reflects the costs of inputs and the price of outputs. In case of efficiency analysis of firms, that purchase all their inputs entering the production or sell all their final products on the market, the usage of overall efficiency considering both the technical and price efficiency is a plausible tool for assessment of production processes as the prices of inputs and outputs are observed. In case of public services, the allocative efficiency analysis is no longer appropriate as the values of at least some inputs or outputs are not observed.

In this thesis we analyze the efficiency of grammar schools in the Czech Republic, therefore, we consider only the technical efficiency as the nature of inputs and outputs included in our analysis does not allow us to perform the allocative efficiency analysis.

Another non-parametric method based on envelopment of the observed data developed for efficiency analysis is called Free-Disposal Hull (FDH) which was introduced by De Prins *et al.* (1984). The difference between DEA and FDH is that the former assumes the convexity of the efficiency frontier whereas the latter does not. The difference between DEA and FDH efficiency frontier determination in a single-output and single-input case is shown in Figure 2.1. FDH is therefore suitable for analysis of production processes with inputs and outputs, that are not perfectly divisible.

Figure 2.1: Comparison of DEA and FDH efficiency frontier



Source: author's layout.

The methods introduced so far are non-parametric (efficiency frontier is constructed from the data) and deterministic, whole distance from the frontier is inefficiency.

Farrell (1957) had an influence on a group of researchers, who developed a parametric approach to efficiency analysis. Aigner *et al.* (1977) were among the first researchers, that introduced the Stochastic Frontier Analysis (SFA), the most famous parametric method. This approach, unlike the previously mentioned non-parametric methods, requires a specification of a functional form between the input and output variables. The next step is running a regression with a composite error term consisting of a symmetrically distributed stochastic component and one-sided distribution component that represents inefficiency. Due to SFA's parametric nature, this method enables straightforward testing for significance using the whole battery of statistical tools. On the other hand, SFA or any other parametric method has one significant drawback, it can be employed on models with multiple inputs but only one output (output-oriented model) or with multiple outputs but only one input (input-oriented model). A multiple outputs or inputs problem may be solved by introducing a composite index which assigns the weight to every output or input. Such approach requires determination of the output or input weights which are the same for all units involved in the analysis. However, uniformly set output or input weights lead to the fact that some units may improve its output or input index only by modification of the weights (Coelli *et al.* 2005).

2.2 Efficiency Analysis Applications

During the last decades efficiency has been analyzed widely by researchers in academia as well as those in private sector partly due to the progression of computer technology. Usage of efficiency analysis may be spotted in papers assessing efficiency of hospitals, educational institutions, airports, police departments, banks and many others.

The reasons for conducting an efficiency analysis are numerous. Firstly, decision making units may be ranked according to efficiency achieved, the operation of the least efficient units may be terminated and, on the other hand, the most efficient units may be rewarded. Alternatively, the drivers of inefficiency may be identified and a course of action may be taken so that the least efficient units improve their operation to achieve a more desirable level of efficiency. In addition, efficiency analysis may prove useful in assessing the impact of a new policy by comparison of the efficiency attained by the units at time before the new policy was introduced to the efficiency attained at a time after the new policy took effect.

Efficiency Analysis of Educational Institutions

In this thesis, we analyze the efficiency of Czech grammar schools, secondary educational institutions providing a comprehensive education. Let us therefore briefly comment on the efficiency analyses conducted in the field of education provision so far. A summary of variables and methods utilized in previous studies relating to the efficiency analysis of educational institutions is provided in Table A.1 in the appendix. We are particularly interested in the selection of inputs and outputs employed in the efficiency analyses as including an irrelevant variable or omitting an important one may lead to different results (Bradley *et al.* 2001).

We observe the tendency to opt for results of standardized nation-wide school-leaving examination as the output in previous studies. Some, for example Kirjavainen & Loikkanen (1998) or Stupnytskyy (2004), included more outputs, such as number of graduates, attendance rates or the success at admission to an institution of a higher education, as the DEA unlike the SFA allows for a multiple input and multiple output model. Johnes *et al.* (2010) was, due to data limitation, forced to include the number of admitted students. He argued in his paper that the number of students admitted to an educational institution is highly correlated with the number of graduates which is obviously a more satisfactory quantitative measure. Unfortunately, he did not include any qualitative output in his analysis.

Although the choice of outputs seems to be rather uniform across the papers, this is not the case for the inputs selection. We see a wide variation in inputs used, among which most common are student-teacher ratio, size of class, expenditure etc. Several papers, Stupnytskyy (2004) or Kirjavainen & Loikkanen (1998), controlled for students' abilities prior admission to a secondary school. Kirjavainen & Loikkanen (1998) actually tested several DEA model specifications containing various inputs and outputs. Firstly, they estimated a simple model with quantitative variables only. Then they continued by adding qualitative inputs and outputs, observed how increasing number of variables shifted the efficiency score distribution and studied the stability of the models' results.

Estimating efficiency scores for decision making units is not usually the ultimate goal of the papers. It is furthermore possible to explain inefficiency and search for drivers of efficiency. One of the most common approaches to determine and quantify variables that have significant effect on efficiency is to use a Tobit regression as non-trivial fraction of observations achieves the highest possible score of efficiency. The variables entering a Tobit regression may be non-discretionary, meaning the DMU has no authority to influence them, or discretionary. Explanatory variables frequently used in a Tobit regression include socio-economic background of students, status of the school (private, public), school's location, male-female ratio of students etc.

Certain papers examined the effect of one specific variable on the efficiency of DMUs and offered policy recommendations. Davutyan *et al.* (2010) related scale inefficiencies of Turkish schools to the highly centralized educational structure concluding that a higher extent of decentralization would be fruitful in terms of efficiency. Bradley *et al.* (2001) argued that the competition among non-selective schools (measured as the number of non-selective schools within different radii) led to higher efficiency scores and this effect strengthens over time. Therefore, the implication for the policy makers is that closing a school may reduce not only public expenditure but also the efficiency of the schools in the neighborhood as the competition among schools decreases.

The efficiency analyses of education institutions have been conducted in numerous countries, the Czech Republic not being an exception. Stupnytskyy (2004) estimated the efficiency of 270 grammar schools in the Czech Republic by means of DEA using data for academic year 1997/1998. He considered the following inputs: students' skills prior admission to grammar school measured as the average grade at completion of elementary school, classroom per student ratio, physical facility index and following outputs: scores in mathematics and Czech language of graduates and also the rate of students admitted to university.

Furthermore Stupnytskyy (2004) discovered that a school with a student career advice center, higher percentage of male students in class, a cooperation with foreign schools and sorting their students into class according to their abilities achieved higher levels of efficiency. Interestingly, the teacher-student ratio negatively influenced the performance of a school in terms of efficiency. Author's suggested explanation is that the higher teacher-student ratio, the smaller the class is. As schools are primarily funded based on the number of students enrolled, smaller classes may lead to a situation when the salary of teacher is not high enough, not incentivizing the teacher to deliver the best quality of teaching.

Franta & Konečný (2009) approached the efficiency measurement differently than Stupnytskyy (2004). They opted for Stochastic Frontier Analysis (SFA), a parametric approach. The output considered was a probability of admission to a university program (share of students successfully admitted taking into consideration the selectivity of the program they are applying for). The direct factors influencing the output were students-teacher ratio, students-class ratio, proxy for management size (turns out to be insignificant) and number of students. Inefficiency was detected therefore further analysis of inefficiency drivers was performed using specific other characteristics of schools, local economic and other conditions. Their results suggest that the student-teacher ratio, unemployment rate of district in which a school is located decrease efficiency. Average class size, female teachers and positively influence efficiency. Grammar schools from relatively small towns (with less than 80,000 inhabitants) outperform on average other schools.

This thesis is going to perform efficiency analysis using the most recent data that is information relating to the cohort having graduated in academic year 2011/2012. We are going to include results achieved in the standardized final school-leaving exams, that were not available when Stupnytskyy (2004) and Franta & Konečný (2009) conducted their analyses. Furthermore, we are going to consider an additional output of the educational institutions. Therefore, we are going to apply DEA, a parametric method that can cope with multiple inputs and multiple outputs. At the end, we will compare our conclusions about the determinants of efficiency to the results of previous studies conducted in the same field.

Chapter 3

Methodology

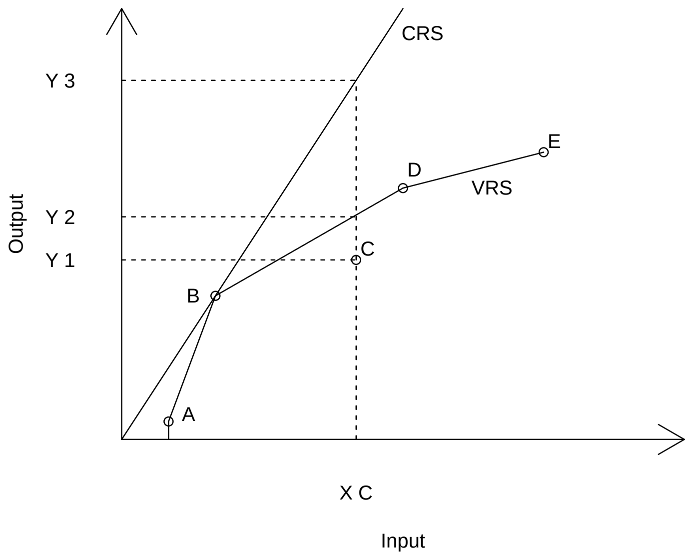
In this chapter, we described the methods used in our analysis. To measure and analyze school-level efficiency we adopt a two-stage approach. Firstly, we estimate the efficiency scores through a non-parametric method DEA. We opt for DEA (instead of more general FDH method) as our input and output variables (described inChapter 3) are perfectly divisible. The second stage consists of a Tobit regression to identify determinants of efficiency. This chapter is devoted to theoretical foundations of DEA and a Tobit regression.

3.1 Data Envelopment Analysis

Let us introduce a simple example illustrating the logic behind DEA. We assume a simplified scenario in which schools have the same single input and there is only one output of their educational process. For illustration let us consider five schools — A, B, C, D, E — that differ in terms of input and output. These schools are depicted in the Figure 3.1 as points where the x and y coordinate represents respectively inputs and outputs of an individual school.

It is obvious that overall the unit with the highest productivity (measured by the output to input ratio) is school B as its output to input ratio is the greatest (the ratio is actually the slope of the line called "CRS"). In case the unit B was capable of maintaining the same input-output ratio regardless the amount of inputs used, its production function would be represented by line "CRS" connecting the origin and point B. This line therefore serves as an efficiency frontier against which other schools are compared. All units except for school B lie below the "CRS" line therefore in the setting of constant returns to scale they exhibit inefficiency. We may also be interested in quantifying the

Figure 3.1: Efficiency frontier determination



Source: author's layout.

degree of inefficiency of individual units. We observe that school C achieves output Y_1 while the output of fully efficient unit (virtual unit constructed based on production process of school B) using the same amount of inputs X_C should equal Y_3. The gap between Y_3 and Y_1 is caused by inefficient production process of unit C. The degree of inefficiency is represented by ratio $\frac{\mathrm{Y}_3-\mathrm{Y}_1}{\mathrm{Y}_3}(\cdot 100\%)$. Alternatively, we may say that the school C achieves $\frac{\mathrm{Y}_1}{\mathrm{Y}_3}(\cdot 100\%)$ efficiency when transforming its input to its output.

In case we assume the variable returns to scale, the efficiency production frontier is represented by a convex piecewise linear curve connecting the units with the highest productivity at certain levels of inputs. Efficiency frontier under variable returns to scale based on our set of five schools is depicted in Figure 3.1 as curve "VRS". We notice that only school C is now inefficient. The degree of achieved efficiency is the ratio $\frac{\mathrm{Y}_2}{\mathrm{Y}_1}(\cdot 100\%)$ that is the actual output of school C divided by the output of a virtual producer that lies on the efficiency frontier and uses the same amount of input as unit C.

Determination of the efficiency score in multiple inputs and multiple outputs scenario requires solving the linear optimization problem (Kirjavainen & Loikkanen 1998). Let us assume that a decision making unit i, $i \in \{1, 2, \ldots, N\}$ using a vector of inputs $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{ni})$ produces a vector of outputs $\mathbf{y}_i = (y_{1i}, y_{2i}, \ldots, y_{mi})$

Efficiency score is obtained as the maximum ratio of weighted outputs to weighted inputs for each decision making unit i, while the most efficient unit(s) achieves the ratio of 1 (or 100%), which can be written as follows:

$$
\begin{aligned}
\max_{\omega,\nu} z_i(\omega,\nu) &= \frac{\sum_{j=1}^{m}\omega_j y_{ji}}{\sum_{k=1}^{n}\nu_k x_{ki}}, \\
\text{s.t.} \quad & \frac{\sum_{j=1}^{m}\omega_j y_{jr}}{\sum_{k=1}^{n}\nu_k x_{kr}} \leq 1, \quad r = 1,\dots,N \\
& \omega_j, \nu_k \geq 0, \quad j = 1,\dots,m \quad k = 1,\dots,n
\end{aligned}
\tag{3.1}
$$

where ω_j, ν_k are weights specific for DMU i of j-th output and k-th input respectively.

The above mentioned programming problem has infinitely many solutions (Coelli *et al.* 2005). If $\boldsymbol{\nu^*} = (\nu_1^*,...,\nu_n^*)$ and $\boldsymbol{\omega^*} = (\omega_1^*,...,\omega_m^*)$ are solutions to the problem then $a\boldsymbol{\nu}^*$ and $a\boldsymbol{\omega}^*$, $a \in \mathbb{R}$ are solutions as well. Therefore a modification consisting of setting the weighted inputs equal to one is necessary. We finally obtain:

$$
\begin{aligned}
\max_{w,v} z_i(w,v) &= \sum_{j=1}^{m} w_j y_{ji}, \\
\text{s.t.} \quad & \sum_{k=1}^{n} v_k x_{ki} = 1, \\
& \sum_{j=1}^{m} w_j y_{jr} - \sum_{k=1}^{n} v_k x_{kr} \leq 0, \quad r = 1,\dots,N \\
& w_j, v_k \geq 0, \quad j = 1,\dots,m \quad k = 1,\dots,n
\end{aligned}
\tag{3.2}
$$

where w_j, v_k are weights specific for DMU i of j-th output and k-th input respectively.

The DEA model represented by equation (3.2) relates to situations when constant returns to scale (CRS) are assumed.

The CRS assumption should be relaxed as decision making units may not be operating at an optimal scale due to for example financial constraints (Coelli *et al.* 2005). The efficient production frontier under the variable returns to scale (VRS) may contain segments of increasing, constant and decreasing returns to scale. The linear optimization problem requires a modification reflecting the relaxed assumption. Banker *et al.* (1984) suggested inclusion of a constant term

u_i in the objective function and in one of the constraints in the following way:

$$\begin{aligned} \max_{w,v} z_i(w,v) &= \sum_{j=1}^{m} w_j y_{ji} - u_i, \\ \text{s.t.} \quad & \sum_{k=1}^{n} v_k x_{ki} = 1, \\ & \sum_{j=1}^{m} w_j y_{jr} - \sum_{k=1}^{n} v_k x_{kr} - u_i \leq 0, \quad r = 1, \ldots, N \\ & w_j, v_k \geq 0, \quad j = 1, \ldots, m \quad k = 1, \ldots, n \\ & u_i \text{ unrestricted} \end{aligned} \tag{3.3}$$

Simar & Wilson (2007) points out that the efficiency scores are serially correlated in an unknown and complicated way.[1] Hence, we can not assume within sample independence (Ramalho *et al.* 2010), which is required for subsequent regression analysis. Simar & Wilson (2007) suggests a solution to this issue that is however complex and beyond the scope of this study.

3.2 Tobit Regression

Once efficiency score for every unit of observation are obtained, determinants of efficiency may be further investigated. In this thesis the efficiency scores are regressed on the school or environmental characteristics not included in the DEA stage to reveal significant drivers of efficiency. As the efficiency scores are by their construction bounded by 0 and 1 (or 0% and 100%) we shall consider application of the limited-dependent variable model. It is tempting to apply the following two-limit Tobit model discussed by for example Maddala (1993):

$$y_i^* = \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2) \tag{3.4}$$

where y_i^* denotes the latent normally distributed variable that satisfies the classical linear model assumptions (Wooldridge 2008). $\boldsymbol{\beta}$ is a vector of parameters to be estimated and $\boldsymbol{x}_i$ represents a vector of explanatory variables. If y_i is the observed dependent variable and L_1 and L_2 are upper (1 or 100%) and lower (0) bounds respectively, we obtain:

[1] Efficiency scores are dependent on each other as the efficiency scores will be different if the efficiency frontier, that is created by the fully efficient units, changes.

$$y_i = \begin{cases} L_1 & \text{if } y_i^* \geq L_1, \\ y_i^* & \text{if } L_1 > y_i^* > L_2, \\ L_2 & \text{if } y_i^* \leq L_2. \end{cases} \tag{3.5}$$

As a matter of fact, the last line in 3.5 is redundant in our analysis. It is impossible for a school to obtain negative efficiency score as it can not generate negative outputs because the lowest possible score in the standardized national exam is 0. Achieving the efficiency score of 0 would arise from a situation when the unit of observation generates only a very small output while consuming enormous amounts of inputs, being completely relatively inefficient. It is highly unlikely that such units appear in our analysis so we abandon the lower limit and employ the following right-censored Tobit model to estimate the vector of parameters $\boldsymbol{\beta}$:

$$\begin{aligned} y_i^* &= \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \qquad (3.6) \\ y_i &= \begin{cases} L_1 & \text{if } y_i^* \geq L_1, \\ y_i^* & \text{if } y_i^* < L_1. \end{cases} \end{aligned}$$

The final model is subsequently estimated by the means of maximum likelihood method, where the likelihood function is stated as:[2]

$$L = \prod_{y_i=1} \left[1 - \Phi\left(\frac{L_1 - \boldsymbol{\beta}^T \boldsymbol{x}_i}{\sigma} \right) \right] \prod_{y_i=y^*} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_i - \boldsymbol{\beta}^T \boldsymbol{x}_i)^2}{2\sigma^2} \right] \tag{3.7}$$

where the first part is a product over the observations achieving the full efficiency ($y_i = 1$), the second second part is a product over the units that exhibits inefficiency ($y_i \leq 1$). $\Phi(a)$ represents the standard normal distribution function evaluated at a:

$$\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2} \right) dt$$

When conducting a regression analysis one should be aware of its assumptions. It is argued that the maximum likelihood estimators (MLEs) are very sensitive to violation of the underlying assumptions. According to e.g. Arabmazar & Schmidt (1982) and Brown & Moffitt (1983) non-normality or het-

[2] Equation 3.7 is based on the logic presented in Maddala (1993).

eroskedasticity of errors lead to MLEs being inconsistent. Therefore, these issues need to be addressed properly.

A Tobit regression in the second-stage is widely used. This approach is found in papers by Kirjavainen & Loikkanen (1998), Bradley *et al.* (2001), Johnes *et al.* (2010), Stupnytskyy (2004) etc. However, McDonald (2009) claims that the usage of Tobit model is inappropriate as he thinks that the DEA efficiency scores are not generated by a censoring data generating process. Instead, he recommends OLS with White's heteroskedastic-consistent standard errors to obtain consistent estimates. On the other hand, he also notes that Tobit estimates are usually similar to OLS estimates. Davutyan *et al.* (2010) uses both OLS and Tobit regression in his study and concludes that the results from both regressions are similar. We will also check the similarity of results obtained though Tobit model with the OLS results.

Chapter 4

Data Description

This chapter is intended to introduce the data employed in our two-stage efficiency analysis. The data set consists of 263 cross-sectional observations of Czech grammar schools. Only schools offering a comprehensive education are considered. Even though 362 such institutions[1] exist in the Czech Republic, we were unable to gather the complete data for some of these schools. However, the sample of 263 schools is representative enough. The data collected relate to the cohort that graduated in academic year 2011/2012.

Variables used in the first stage — the data envelopment analysis — represent inputs and outputs of educational process. Then the effect of several school and environmental characteristics on efficiency is tested through a Tobit regression to identify possible determinants of efficiency.

4.1 Inputs

There is hardly a consensus on a complete list of inputs entering the educational process. Researchers are therefore left to experiment with various variables. Furthermore, they face limitations in terms of data availability. In the present paper the following variables are considered as inputs:

- Number of teachers per 100 students (*teach*)
- Expenditures per student in thousands CZK excluding salary expenditures (*expend*)
- Percentage of rejected students at admission (*reject*)

[1]We classify an school as an institution offering comprehensive education if at least one student of a grammar school program graduated in 2012.

Number of Teachers per 100 Students

The number of teachers per 100 students is calculated as the number of teachers (full-time equivalents) employed by a school as at 30th September 2011 divided by the number of students enrolled at a school at the same date, times one hundred.

It can be reasonably expected that the higher is the teacher per 100 student ratio, the more individual care a student receives which, in turn, probably leads to better educational outcomes such as higher scores in the final examination.

Expenditures per Student

Expenditures per student excluding salary expenditures are obtained as the total costs of a school less the salary expenses. Data for each school are extracted from their statement of income for the year end 31st December 2011. We have excluded the salary expenses (including social and health security expenses) as it closely relates to the number of teachers per student, another input variable. The higher the number of teachers per hundred students, the higher is the salary expenditure per student. The variable *expend* should therefore contain costs of running the school such as utilities costs, costs of materials consumed, maintenance etc. As the total costs are extracted from the statement of income, the depreciation of long term assets is also included in the variable *expend*. Through the concept of depreciation, we avoid the situation when a substantial expense for, for example major refurbishment, incurred in one year leads to very high level of the *expense* variable. Omitting the depreciation completely would not be appropriate as the investments in long term assets are also costs borne by a school.

We expect that the increase in expenditures per student should have positive effect on educational achievements. The more funding is available, the more state-of-the-art equipment a school can afford leading to more interactive and compelling teaching. As a consequence, it may stimulate students' interest in schooling resulting in higher educational achievements.

Percentage of Rejected Students at Admission

So far, we have not considered the quality of students entering the educational process. Naturally, if the students admitted at one school are more skilled than the students admitted at another school, the outcome of educational process of the former school would very likely be higher than the one of the latter even if both schools had exactly the same other inputs. Therefore, additional input variable is necessary. The percentage of rejected students at admission may serve as a proxy for students' ability prior entering the school. The percentage is calculated for each study program finished with final school-leaving examination[2] as the number of rejected students divided by the number of applicants. Given the cohort that graduated in 2012 is the subject of our interest, the number of rejected students and number of applicants relates to admission process taking place in 2008, 2006, 2004 for four, six and eight year programs respectively. If a school offers several study programs the overall rejection rate is obtained as weighted average of rejection rates for each program weighted by the number of students accepted to each program.

The last input (percentage of rejected students) could be considered as a questionable one. One can argue that if the percentage of rejected students is high, only the best students are accepted. On the other hand, we feel that the selected proxy may not be a completely accurate measurement of student's ability at admission. To some extent, a study program with demanding curriculum (focused for example on mathematics) may attract lower number of prospective students. However, the students interested in this difficult program may possess a sound knowledge and learning ability prior enrollment to a secondary school.

The mobility of students at this stage of education is rather limited, meaning that they do not usually leave the household of their parents because of secondary education. Therefore, the institution at which a student is schooled is located not far from the address of their residence depending on the transportation options. In addition, children of university educated parents are expected to achieve better educational results than others. Consequently, it may be argued that the students' quality at admission may be closely related to the percentage of university educated people in school's vicinity[3]. We thus perform DEA using rejection rate as a proxy of students' ability at admission and then test the effect on the percentage of university educated people in the area on the efficiency score. If the percentage of university educated people turned out to be significant, new efficiency scores would be obtained through DEA when input rejection rate would be replaced by percentage of university educated people in school's vicinity.

[2]A grammar school in the Czech Republic may offer several learning programs finished with final school-leaving examination. These programs may differ in curriculum (focus on natural sciences, foreign languages etc.) and length (four, six, eight years).

[3]The description of this variable is provided in Section 4.3.

Furthermore, limited mobility of students also leads to the fact that grammar school is small towns may receive relatively fewer applications than schools in larger towns because the number of prospective students living in small or villages nearby is lower than in the larger towns.

Considering all above mentioned issues, we conclude that a more suitable proxy for student's ability at admission would be results obtained in standardized entrance exams. Several dozen secondary schools outsource the preparation of entrance exams to a private firm Scio.[4] The number of schools to which Scio provides this service is growing so future research in this field may benefit from it in case Scio is willing to provide information aggregated on a school-level about the entrance exams results. Alternatively, if Ministry of Education launches next year nation-wide testing in mathematics, Czech and English language, of students in their final year of elementary school, these results may be used in DEA as good proxy of students' skills prior admission to grammar school.

4.2 Outputs

In contrast with an ordinary production company manufacturing goods, educational institutions are service providers associated with many externalities. Therefore, defining the outputs of an educational process constitutes a challenge for researchers.

It seems rational to include the results obtained at the standardized final exams. The same set of problems for all students sitting the exams together with uniform guidance on marking the answers should ensure objectivity. Concerning other outputs, one may even consider a salary after securing a job to be also an outcome of educational process. Therefore, the selection of outputs depends again on the researcher and on data availability. We include the following two variables as outputs of educational process in our DEA:

- Average score achieved at standardized final school-leaving examination (*avgscore*)
- Financial award per student obtained in "Program Excelence" (*pe*)

[4] See www.scio.cz.

Average Score at Final Examination

Prior academic year 2010/2011 the final school-leaving examinations were prepared and evaluated by individual schools, therefore the results of different schools were incomparable. The final standardized exams were firstly introduced for academic year 2010/2011. Since then, in order to graduate, every student is required to pass two standardized exams, that are the same for all schools, along with additional exams set by the given school. Czech language is the compulsory standardized exam for everybody and students choose the second exam to be either mathematics or foreign language.

Furthermore, the student could decide to take the higher level of the exams which was more difficult or the lower level that was easier. In academic year 2011/2012 students did not have many incentives to take the higher level as tertiary educational institutions in the Czech Republic did not require prospective students to pass the higher level in order to be admitted. Nevertheless, we feel that the students who took the higher level of exams deserve a premium as they would have probably scored more if they had taken the lower level. CERMAT, the institution responsible for standardized final examination did not provide us with guidance as to how a student sitting a higher level exam would have scored if he or she had sat the less difficult exam. Therefore, we decided to increase the scores achieved in the higher level exams by arbitrary 5% while maintaining the highest score to be 100%. As the standardized final exams have been introduced very recently, the Ministry of Education keeps changing the format of these exams. As of academic year 2012/2013 only one level of difficulty will be available, therefore the future research in this area should not face the dilemma whether or not to reward the students for taking the higher level and how.

The overall average score for individual school is calculated as a weighted average of average scores achieved in all the exams where the weights are set to be the number of students who took the exam in spring or fall 2012. The weighted average is more appropriate than the simple average as it prevents the data to distorted in case a single student achieves 100% in for example higher level of Spanish while the remaining students reach much lower scores.

Financial Awards in "Program Excelence"

The second output variable is represented by the money obtained in the "Program Excelence", a program introduced by Ministry of Education in 2011 for the first time. The aim of this program is to financially support the teachers of the students who achieved the best results in competitions focused mainly on natural sciences in order to incentivize the care of talented students. We feel that this variable represents a proxy for an important output of educational process, that is highly successful students in a certain field. Therefore, the variable financial award obtained in"Program Excelence" per students is included in DEA as an output. In 2012 the amount of financial support provided by Ministry of Education reached 20 million CZK which was divided between 385 schools across the country. 189 out of 263 schools in our sample benefited from this program in 2012. For better comparison across schools it is required that the amount of funding a school received is divided by the number of students enrolled at this school.

Table 4.1 summarizes the descriptive statistics of all variables (inputs and outputs) used in DEA.

Table 4.1: Inputs and outputs — descriptive statistics

Variable	Mean	St. Dev.	Min	Max
Inputs				
Teachers per 100 students	8.296	1.378	4.890	19.930
Expenditures per student (TCZK)	15.087	11.351	4.240	165.550
% rejected students	30.537	16.957	0	72.510
Outputs				
Average score in the stand. final exams	80.274	4.079	66.950	92.600
Financial award per student	105.547	131.185	0	966.590

Source: MSMT, CERMAT, MF; author's computations.

4.3 Determinants

In the second stage of our analysis we focus on the search for possible drivers of efficiency. Several school and environmental characteristics have been gathered in order to test their influence on the efficiency score of the schools.

School Characteristics

The following school characteristics have been considered:

- Percentage of female students (pf)

- Percentage of not qualified teachers (*pnq*)
- Yearly salary expense per teacher (*sal*)
- School founded by the state (*state*)
- Percentage of 2012 graduates that attended six year or eight year program (*lng*)
- School offers also vocational program(s) finished with standardized final school-leaving exams (*nongr*)
- School size measured as total number of students attending the school (*studnum*)

The variable percentage of female students is calculated as the share of female students that graduated in 2012. It is commonly believed that female students tend to be more diligent in their approach to education, therefore we feel that this variable should influence positively the efficiency of individual schools.

The percentage of not qualified teacher is a share of teachers that have not obtained a professional qualification[5] as at 30$^{\text{th}}$ September 2011. Essentially, a master degree in pedagogy is a sufficient qualification for teachers employed at grammar schools. It is expected that the percentage of not qualified teacher has a negative effect on efficiency.

The variable yearly salary per teacher is obtained by dividing the total expenditures on salaries incurred in year 2011 (excluding social and health insurance) by the total number of teachers (full-time equivalents) as at 30$^{\text{th}}$ September 2011. Unfortunately, the provided salary expenditures do not include teachers' salaries only but they also contain the salaries of administrative employees, principals etc. However, we still feel that the variable approximately reflects the level of teachers' salaries. It is expected that the level of salary has a positive influence on the efficiency as better paid teachers may be more motivated and deliver better standard of teaching.

The next variable classifies the schools according to the founder. It equals one when the school is public (founded and maintained by state). Private legal entities or a church are other founders of schools. We have not distinguished

[5] Act No. 563/2004 Coll. as amended, on Pedagogical Staff states the conditions for teachers to obtain the professional qualification.

between private and religious schools as the number of religious school in our sample is very low.

Grammar schools in the Czech Republic also offer 6 or 8 year programs, meaning that the students enroll at the school after completion of 7th or 5th grade of elementary school. These programs are usually more demanding and attract hard working students. Consequently, we expect that the more students are enrolled in 6 or 8 year programs the higher should be the efficiency. Therefore, we include variable percentage of 2012 graduates in six or eight year program. The percentage was calculated based on number of admitted students but as the drop-out rate at this stage of education is negligible we believe that the calculated percentage represents the true composition of the graduates.

The next variable is a dummy variable that indicates whether a school also offers a vocational program finished with standardized final school-leaving exam. The results of students attending such vocational program(s) are also included in the variable average score achieved in final exams, so the fact that the school offers such program may negatively influence the efficiency of that school. Essentially, vocational programs at secondary school level are still perceived as inferior to comprehensive education.

Finally, we would like to test whether bigger schools measured in terms of students enrolled achieve higher or lower efficiency than smaller school or if the school size does not affect efficiency at all.

Environmental Characteristics

Two environmental factors relating to the area where the school is located are also included in our model:

- Percentage of people with university degree (*pud*)
- Size of the city/town based on its population (*s1*), (*s2*), (*s3*), (*prg*)

The share of university educated people is calculated as the number of local people who obtained any university degree divided by the total number of people living in the area. The students at the secondary level of education tend to visit school near to their household due to their limited mobility. Children of people with university degree are more likely to perform better in their studies as they are usually encouraged by their parents. It is expected that the percentage of university educated people should have positive effect on the efficiency. The data were collected in 2011 during the national census, taking place every 10 years. Therefore, the considered data are up-to-date.

The last set of variables indicates the size of the town or city where a school is located measured by its population. If the town has less than 20,000 inhabitants it is classified as small (s1); more than 20,000 but less than 50,000 inhabitants, it is considered as medium sized town (s2); and finally all towns with population above 50,000 excluding Prague are regarded as large (s3). There is a separate dummy variable for schools located in the capital city, Prague. The population is stated as at 1st January 2012.

Table 4.2 provides summary statistics of variables used in the second stage of the efficiency analysis. In addition Table A.2 in the appendix shows correlation between determinants. We notice there is very high correlation between the dummy variable for Prague and percentage of university educated people.

Table 4.2: Determinants of efficiency — descriptive statistics

Variable	Mean	St. Dev.	Min	Max
School characteristics				
% female students	60.897	9.451	25.000	86.540
% not qualified teachers	4.203	5.480	0	34.620
Yearly salary per teacher	405.190	57.527	171.930	648.620
State school	0.863	0.344	0	1
Vocational program	0.057	0.232	0	1
% graduates from 6 or 8 y. program	49.814	28.405	0	100.000
Total number of students	434.589	184.011	79	1,095
Enviromental variables				
% university educated population	13.203	6.512	4.140	30.410
Population $<$ 20,000	0.226	0.419	0	1
Population 20,000 - 50,000	0.491	0.501	0	1
Population $>$ 50,000	0.283	0.451	0	1
Prague	0.175	0.381	0	1

Source: MSMT, CSU, MF; author's computations.

4.4 Data Sources

The data employed in this study were obtained from several sources. School characteristics such as the number of students, the number of full-time equivalent teachers, study programs, the percentage of admitted students etc. were provided by the Ministry of Education, Youth and Sports (MSMT).

The total expenditures and salary expenditures of public schools are disclosed by the Ministry of Finance as a part of the statement of income of individual schools. Private school expenditures can be extracted from their statement of income that is available in Commercial Register at www.justice.cz. As far as the schools established by church are concerned, the expenditures are often found in their annual reports that include the statement of income. To obtain the sample of schools as extensive as possible, we have requested the income statement from those schools we were unable to acquire their expenditures by the means described earlier. Unfortunately, only several schools provided us with their statement of income in this case.

Average score achieved in the final school-leaving examination is calculated based on data provided by CERMAT, the entity responsible for preparation, execution and evaluation of the standardized final school-leaving examination in the Czech Republic.

"Program Excelence", the program recently introduced by the MSMT, rewards teachers of students who participated and succeed in competitions in mathematics, physics, foreign languages etc. MSMT discloses the amount of financial rewards assigned to each school on their website.

Information about the area in which a school is located (population of the city/town, percentage of people with university degree) are extracted from the data released by the Czech Statistical Office.

Chapter 5

Empirical Analysis

In this chapter, results of two-stage efficiency analysis are presented. Cross-sectional data on grammar schools in the Czech Republic relating to the cohort that graduated in academic year 2011/2012 were used. Firstly, DEA was performed to obtain an efficiency score for each unit of observation (school). The second stage of efficiency analysis consisted of a Tobit regression through which the impact of various school characteristics as well as environmental variables was tested. We performed all our calculations in software R.

5.1 DEA Results

We perform DEA[1] under the assumption of constant returns to scale and an output-oriented DEA allowing for variable returns to scale. Under variable returns to scale, unlike under constant returns to scale, input and output oriented DEA yield different results. The reason for the output orientation under VRS is that each school should aim for the most efficient utilization of its inputs.

Table 5.1 shows basic information about the distribution of efficiency scores obtained under different return to scale assumptions. It is not surprising that the number of fully efficient schools drops from 21 to 11 when constant returns to scale are assumed instead of variable returns to scale. The disparity stems from the difference in determination of the efficiency frontier. Mean and minimum of the efficiency scores in VRS model are also higher than in the CRS case. Minimal efficiency under VRS is 76.77 meaning that the least performing unit achieves only 76.77% of the output that a fully efficient school would attain given the same amount of inputs. In other words, the most inefficient

[1] Command `dea` (package `Benchmarking` required)

Table 5.1: Basic summary statistics of efficiency scores obtained under VRS and CSR assumptions

	VRS	CRS
Mean	92.46	79.12
Minimum	76.77	35.35
Maximum	100.00	100.00
No. of efficient schools	21	11

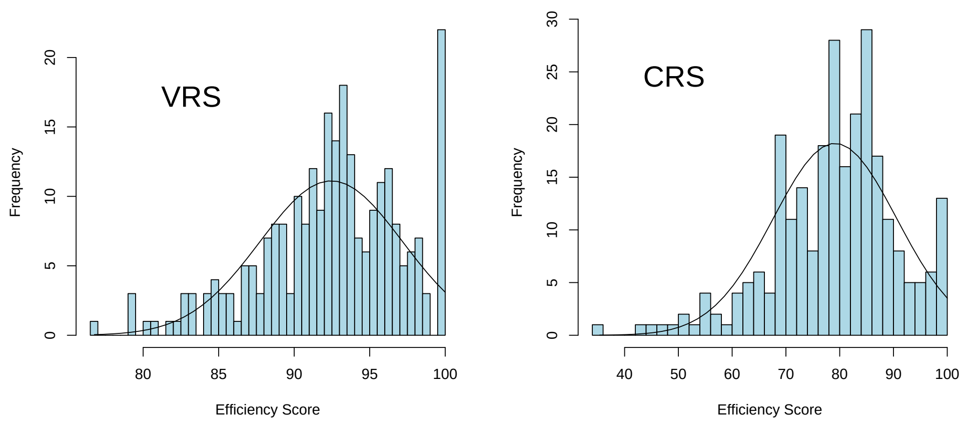
Source: author's computations.

unit under VRS needs to increase its outputs by 30.3% while keeping the level of inputs constant in order to achieve full efficiency.

The minimum efficiency score under CRS turns out to be substantially lower, 35.35. Apart from the above described output-oriented interpretation, efficiency estimates under CRS may indicate how inputs can be reduced to attain full efficiency. The unit with the minimum efficiency score may attain full efficiency if it reduces the inputs employed in its production process by 64.65% and keeps outputs at the current level.

More information about the distribution of efficiency scores can be obtained from a histogram. We provide a histogram for both models in Figure 5.1.

Figure 5.1: Histogram of efficiency scores obtained under VRS and CSR assumptions



Source: author's computations.

We conclude that the distribution is roughly normal but the efficiency scores are limited from above by value 100 and a non-trivial number of observations takes this upper value, which has an effect on the choice of a regression model in the second stage.

Another overview of efficiency scores is contained in Table 5.2 where we can observe ten most efficient schools, median school, ten least efficient schools together with their mix of inputs and outputs. The ranking is based on efficiency score obtained under VRS.

Table 5.2: Ranking of schools according to their efficiency

			DEA inputs			DEA outputs	
Efficiency rank	Efficiency estimate		% Rejected students	Teachers	Expend.	Financial award	Exam results
	VRS	CRS					
School 1	100.00	100.00	26.47	4.97	23.90	0.00	76.50
School 2	100.00	100.00	17.26	4.89	13.48	3.73	74.97
School 3	100.00	100.00	21.43	7.19	6.11	77.62	81.07
School 4	100.00	100.00	11.65	7.86	7.43	322.12	85.67
School 5	100.00	89.35	40.18	6.45	20.45	20.71	87.87
School 6	100.00	100.00	43.28	7.50	8.68	966.59	83.96
School 7	100.00	100.00	0.00	7.75	11.43	9.58	86.58
School 8	100.00	100.00	6.19	8.03	6.75	175.42	78.26
School 9	100.00	100.00	0.00	8.83	11.22	186.39	79.34
School 10	100.00	100.00	1.27	7.77	13.62	108.35	80.23
Median school	92.82	68.54	64.27	8.10	29.74	85.65	82.59
School 254	82.85	60.48	24.52	8.19	27.93	0.00	72.81
School 255	82.71	64.74	22.88	7.60	23.18	0.00	72.42
School 256	82.43	62.97	33.49	9.44	12.23	90.76	72.86
School 257	81.86	65.41	15.15	8.20	51.52	0.00	71.60
School 258	80.84	76.67	14.69	7.49	10.95	126.72	69.55
School 259	80.48	53.69	25.00	9.71	20.61	0.00	71.23
School 260	79.43	50.45	20.74	10.29	25.36	0.00	70.33
School 261	79.22	46.47	37.04	9.90	28.92	0.00	70.56
School 262	79.14	61.42	35.03	8.23	16.69	0.00	69.92
School 263	76.77	63.01	7.78	9.44	12.32	0.00	66.95

Source: MSMT, MF, CERMAT; author's computations.

Table 5.2 essentially shows that there are several possible ways how to achieve full efficiency. Both average score in final examination and financial award obtained in "Program Excelence" of school 2 are almost the lowest compared to the group of fully efficient schools. High efficiency of this particular school can be achieved due to the quite low level of inputs. On the contrary, the students of school 5 attained on average the best results in final exams but the school 5 required substantially more inputs than school 2.

As far as ten most inefficient schools are considered, their students tend to attain very low final exam results and most of the schools do not obtain any funds from the "Program Excelence". The levels of inputs vary within this group but we notice that they are on average higher than the levels of inputs of ten most efficient units.

Furthermore, we see that the median efficiency is 92.82 under VRS while the median efficiency score under CRS is only 68.54.

State vs Private schools

One may be interested in the comparison of the performance of state schools with the private[2] schools' performance. Table 5.3 summarizes the mean and median of efficiency scores by school type. In addition, Figure 5.2 provides a cumulative distributions of efficiency scores of state and private schools.

Table 5.3: Mean and median of state and private schools

	Mean	Median
VRS		
State	92.64	92.81
Private	91.31	93.28
CRS		
State	81.23	81.34
Private	65.84	64.67

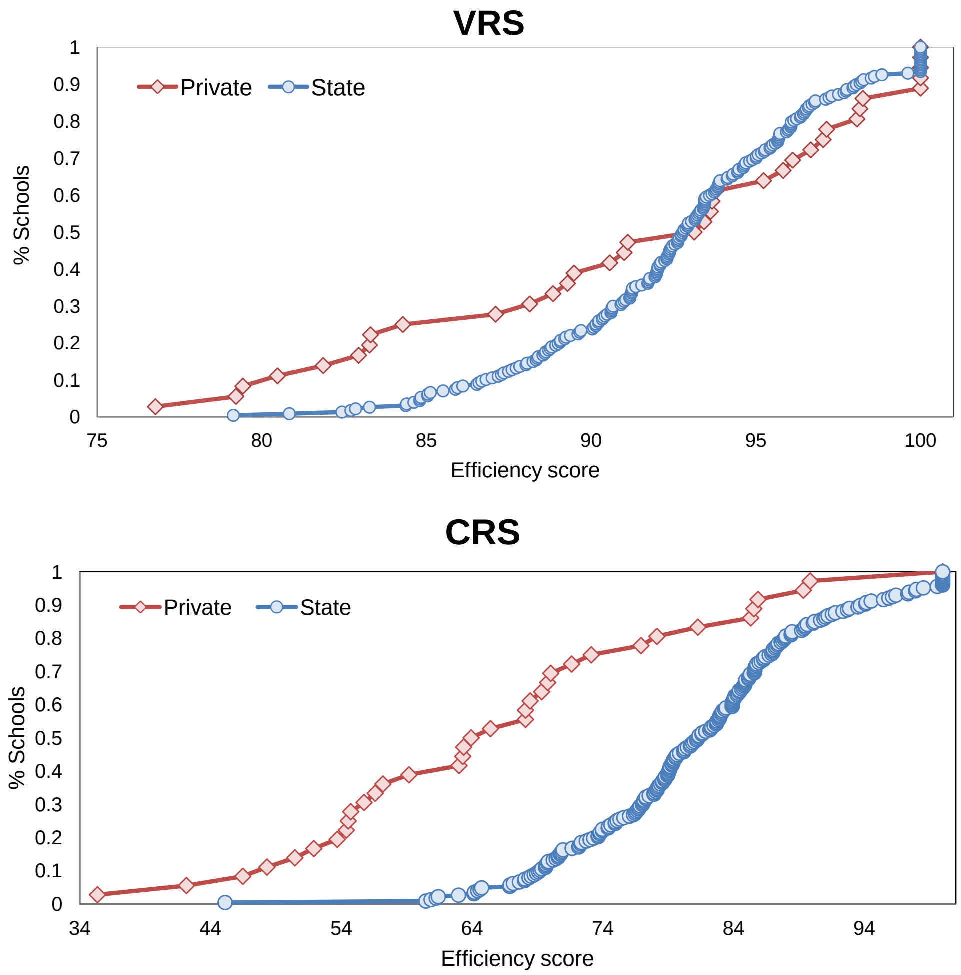
Source: author's computations.

State schools are on average more efficient than the private ones under both variable and constant return to scale assumptions. Under VRS the difference in mean efficiency is rather marginal, 1.33 whereas the mean efficiency of state school is greater by 15.39 under the CRS.

The cumulative distribution functions of state and private schools under VRS are more similar to each other than the cumulative distribution functions under CRS. The consequence of not allowing for variable returns to scale is that the schools which employ high amounts of inputs but achieve excellent educational outcomes are severely disadvantaged. Several private schools employs high amounts of inputs but their graduates attain one of the best results in the final exams. We have even observed that several fully efficient units under

[2]Schools founded and maintained by a church are treated as private as only a small number of type of school is included in our analysis.

Figure 5.2: Cumulative distribution of efficiency scores of private and state schools

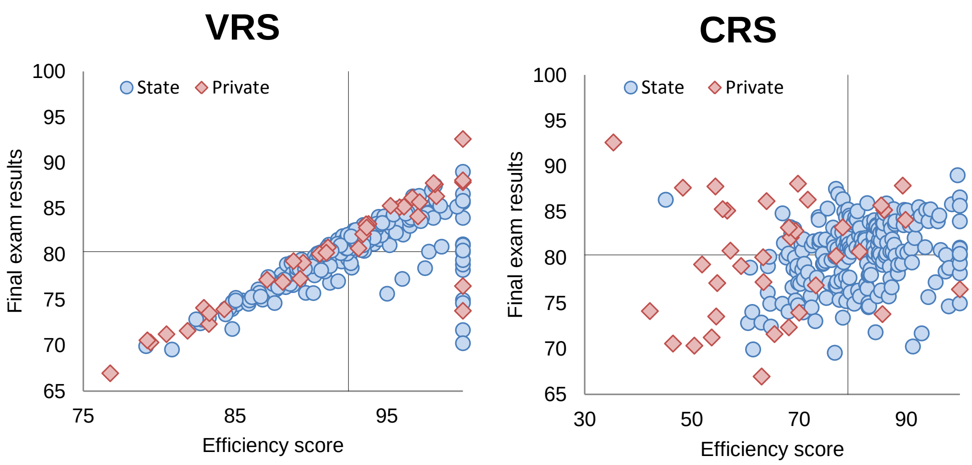


Source: author's computations.

VRS achieve very low efficiency scores under CRS. It is important to mention that it may be even impossible to achieve full efficiency under the CRS for high levels of inputs as one of the educational outcomes — the results of final school-leaving examination — is capped at 100.

Speaking of final examination scores, Figure 5.3 shows the relationship between the average result in final school-leaving exams and efficiency scores obtained under VRS and CRS. Scatter plots are divided into four quadrants. Schools in quadrant I achieves above average final exam results as well as above average efficiency scores. Quadrant II contains schools that achieve above average exam results but could perform even better with the same amounts of inputs. Academic achievement of schools falling into quadrant III is poor as well as their efficiency scores. Finally, schools that tend to maximize, to a large extent, their exam results given their resources but their exam results are still below the average, perhaps due to lack of resources, are depicted in quadrant IV. State and private schools are marked differently in the scatter plot so that further analysis of relationship between exam results and efficiency scores by school type is possible.

Figure 5.3: Efficiency score vs average results obtained in final shool-leaving examinations



Source: author's computations.

Table 5.4 summarizes the information contained in Figure 5.3, e.i. it shows how many schools of each type together with the percentage belongs to each quadrant.

We observe that in every quadrant the percentage of state schools is more or less similar to the percentage of private school under VRS. On the other hand, the percentage of private schools in quadrant I decreases significantly — from 47% to 14% — and we see the same increase, in terms of percentage points, in the percentage of private schools in quadrant II when the efficiency scores are calculated under CRS instead of VRS. The migration between these two quadrants in caused by the fact that private schools tend to have more resources at their disposal and some of them actually achieve stellar academic results. However, the additional increase in academic achievement is outweighed by the increase in additional resources when the CRS are assumed in the calculation of efficiency scores.

Table 5.4: Number of schools in each quadrant

	VRS	CRS
High efficiency, high exam results (QI)		
State	107 (47%)	87 (38%)
Private	17 (47%)	5 (14%)
Low efficiency, high exam results (QII)		
State	20 (9%)	40 (18%)
Private	1 (3%)	13 (36%)
Low efficiency, low exam results (QIII)		
State	84 (37%)	52 (23%)
Private	16 (44%)	16 (44%)
High efficiency, low exam results (QIV)		
State	16 (7%)	48 (21%)
Private	2 (6%)	2 (6%)

Source: author's computations.

Efficiency Analysis by Region

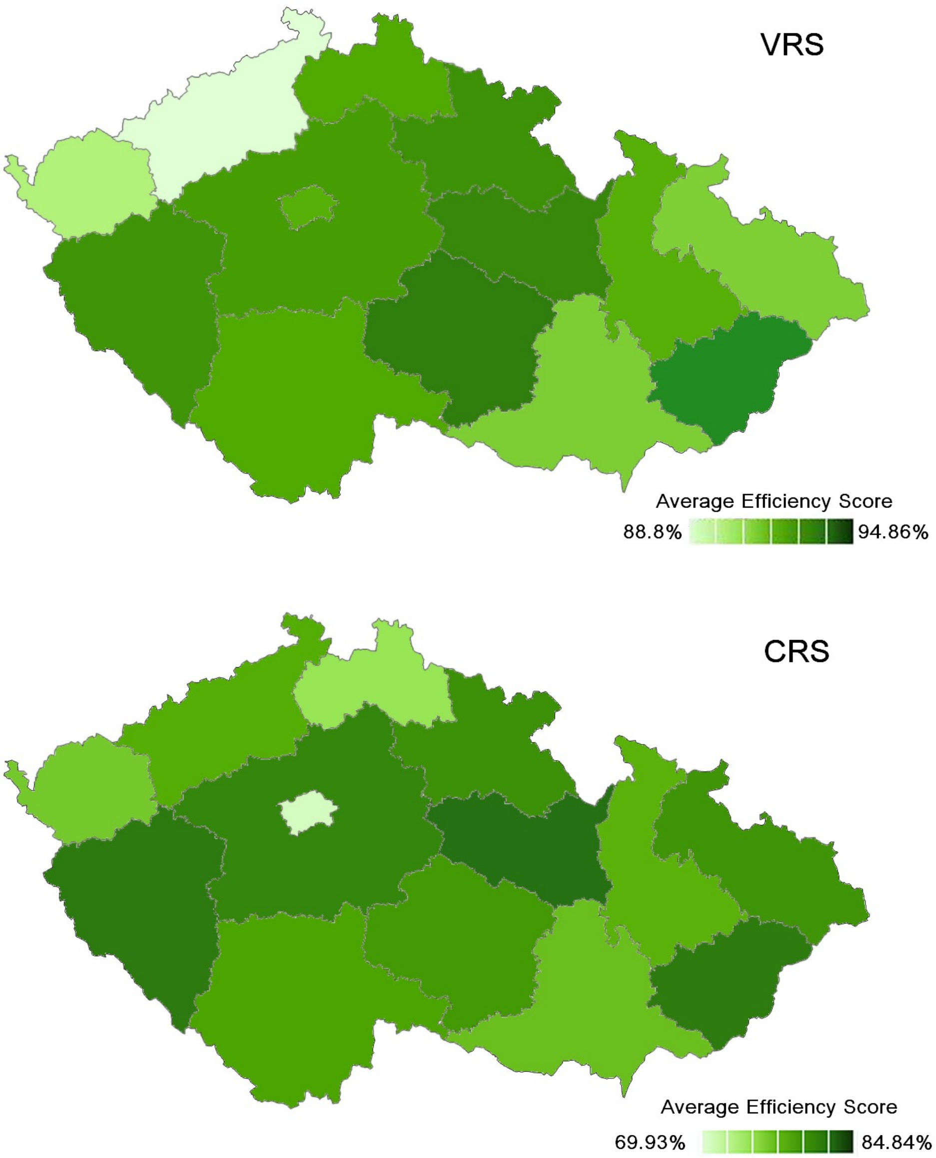
It is also possible to geographically analyze efficiency. Figure 5.4 depicts the average efficiency of schools located in one of 14 regions of the Czech Republic[3] through the use of a choropleth map. Numerical values of average efficiencies for respective regions are presented in Table A.3 in the appendix. The interpretation of the choropleth maps is as follows: the darker the shade the more efficient schools on average are located in this region.

We observe that the choropleth maps show different results under VRS and CRS assumptions. We notice that the most significant disparity in average efficiencies occurs in the capital city, Prague. As a matter of fact, Prague schools are on average assigned the lowest efficiency scores under CRS. Upon a closer look it is detected that 16 private school schools, which is half of the private schools in our sample, are situated in the capital city. As established earlier, the private schools are attributed low efficiency scores under CRS. Regions with on average relatively high efficient schools under both assumptions are Zlínský, Pardubický and Západočeský kraj.

At this point, we feel that DEA under VRS more fairly and realistically determine the efficiency frontier as: (i) minimum score achieved under CRS

[3]Figure A.1 in the appendix presents an overview and geographical position of the regions in the Czech republic.

Figure 5.4: Average efficiency scores by region



Source: author's computations.

assumption is much less realistic as the least efficient unit under CRS that by the way achieved very good results in the final school-leaving examination, should increase outputs by more than 200% (corresponds to the efficiency score of 35.5), (ii) one of the output - results of the school leaving exams are limited by 100, therefore observations with large inputs would be disadvantaged even if their exam results were excellent, (iii) positive relationship between results of final exams and VRS scores as depicted in Figure 5.3 is much more intuitive. Therefore, the second stage — search for determinant of efficiency — will take into consideration only the efficiency scores calculated under VRS.

5.2 Tobit Regression Outcomes

In order to identify the variables that have significant influence on the efficiency achieved by individual schools, regression analysis is performed. Tobit regression model has been selected due to the fact that the upper values of the dependent variable, efficiency score, are limited by 100.

Several model specifications are estimated using the Tobit regression[4], results of which are displayed in Table 5.5. The purpose of running several regressions is to find out whether the estimates change substantially if a variable or a set of variables is excluded or added to the model.

First, we run a regression that includes all determinants that, according to our opinion, could influence the efficiency scores. The determinants are described in detail in Section 4.3. The results of the first regression suggest that percentage of female graduates as well as the percentage of graduates attending the six or eight year program increase efficiency. School size also positively influences efficiency. On the other hand, schools offering vocational program tend to achieve lower level of efficiency. Furthermore, it seems that the schools in cities with population above 50,000 (including Prague) achieve lower efficiency scores. Coefficient of percentage of university educated people appears to positively influence the efficiency and is significant at 10% level of significance. However, we know that there is a strong positive correlation between the percentage of educated people and the dummy variable for Prague (which may result in multicollinearity), so it is unlikely that the estimates should differ in sign.

The second model excludes the variable percentage of university educated

[4]Command `tobit` (package `AER` required)

Table 5.5: Tobit regression estimates

	Model (1)	Model (2)	Model (3)	Model (4)
% female students	0.090***	0.086***	0.086***	0.095***
	(0.033)	(0.033)	(0.033)	(0.032)
% not qualified teachers	0.054	0.048		
	(0.055)	(0.055)		
Yearly salary per teacher	0.001			
	(0.005)			
State	−0.751	−1.047	−0.886	
	(1.153)	(1.145)	(1.134)	
Vocational program	−5.383***	−5.609***	−5.372***	−5.056***
	(1.348)	(1.343)	(1.327)	(1.310)
% graduates attending a 6 or 8 year program	0.030***	0.028**	0.028**	0.030***
	(0.011)	(0.011)	(0.011)	(0.011)
School size	0.005**	0.005**	0.005**	0.003**
	(0.002)	(0.002)	(0.002)	(0.002)
Small town			0.990	
			(0.832)	
Middle sized town	−0.686	−0.706		
	(0.940)	(0.946)		
Large town	−2.150**	−1.242		
	(1.089)	(0.963)		
Prague	−4.305**	−1.894*	−0.770	
	(1.682)	(0.973)	(0.857)	
% university educated	0.165*			
	(0.095)			
Constant	83.102***	85.349***	84.541***	84.224***
	(3.426)	(2.330)	(2.230)	(2.162)
Observations	263	263	263	263
Log likelihood	−741.245	−742.773	−743.332	−745.004
Wald Test	44.605***	40.994***	39.670***	35.873***
	(df = 11)	(df = 9)	(df = 7)	(df = 4)

Notes: *p<0.1; **p<0.05; ***p<0.01
standard errors are reported in brackets

Source: author's computations.

people for the reason stated above[5] and one insignificant school characteristic — the yearly salary per teacher. Under this model specification, the dummy variable for large cities proves to be insignificant at all standard levels of significance. The dummy variable for Prague is now significant only at 10% as opposed to 5% under the model 1 specification.

We further exclude variable percentage of not qualified teachers, dummy variables of medium and large city and include the dummy variable small city in Model 3. Dummy variables for Prague as well as for small city turn out to be insignificant. Variables that are significant at 5% or 10% under previous model specifications prove to be significant under model 3 specification as well. Furthermore, there is no change in signs of the estimates so the direction of influence on the efficiency is consistent across the models.

Model 4 includes only the variables that are significant in all previous model to check the stability of the results.

As we mention in the Methodology chapter, Tobit regression hinges on the assumption of homoskedasticity and normality of errors. If these assumptions are violated the obtained estimates are not consistent. We comment on homoskedasticity and normality of errors in Appendix B. Furthermore, we verify that our results are not different, in terms of significance and the signs of coefficient, from a OLS estimates when White heteroskedasticity-robust standard errors are calculated. The OLS results are shown in Table A.4 in appendix. We see that the OLS results are comparable to those obtained using Tobit model.

5.3 Discussion of Results

Our results indicate that the percentage of female students increases efficiency which is in line with Bradley *et al.* (2001) who concludes that girls-only schools perform better than other institutions. On the contrary, Stupnytskyy (2004) identifies percentage of male students as a driver of efficiency. According to OECD (2012b) 15-year old female students substantially outperformed their male counterparts in 2009 PISA reading assessment in all OECD countries while the male students achieved slightly better results in mathematics assessment. The final school-leaving examination, scores of which are considered as one of two outputs in our study, consists of two exams, one of which the Czech language is compulsory for all students. Mathematics is only an optional exam

[5] If we exclude the dummy variable for Prague, the percentage of university educated people is insignificant.

therefore, not all students were tested in mathematics. Stupnytskyy includes scores in mathematics that was compulsory for the students in 1998 as an output in his DEA and states that the male students achieved on average 10% higher scores in mathematics and only 1% lower scores of Czech language when compared to female students. We conclude that our results are in line with the facts presented in the up-to-date OECD report and we are not surprised that our conclusion about this variable contradicts the results obtained by Stupnytskyy.

The percentage of graduates from 6 or 8 year program is according to our analysis positively related to efficiency. Stupnytskyy (2004) who includes dummy variable for schools that offer 6 or 8 year program, arrives at the same result that a school offering such program is more efficient.

Stupnystky claims that the school size measured as the number of students attending the school does not have an effect on efficiency. Kirjavainen & Loikkanen (1998) reach the same result. Contrarily, evidence in this thesis suggests that a school size positively influence the efficiency. The idea of a larger school higher school achieving higher efficiency may be actually plausible. A large school may offer extra-curricular activities, such as experiments in the biology lab, singing in the choir, theater club and attract enough of its students to participate in them so that these activities are economically viable. Students participating in such activities may become more interested and eager to learn. It would be interesting to study the link between the efficiency scores and the extent to which a school offer these extra-curricular activities. However, we are unable to test the relationship between the two variables due to data limitation.

Our hypothesis about the schools offering a vocational study program is supported by the data as it was established that schools offering such program typically attain lower efficiency.

Percentage of not qualified teachers, yearly salary per teacher and ownership of the school proves to insignificant school characteristics. Stupnytskyy (2004) discovers that private schools achieve higher efficiency. Our results suggest that the founder of a school does not have an effect on efficiency. It is apparent from Figure 5.2 that some private schools achieve very low efficiency scores while there are private schools that attain very high efficiency under VRS.

We have not been successful in finding a link between efficiency and environmental factors.

Chapter 6

Conclusion

In this thesis efficiency of 263 grammar schools in the Czech Republic was studied by DEA and Tobit regression using an up-to-date dataset. The inputs considered in the first stage included number of teachers per 100 students, expenditure per student and rejection rate prior admission to grammar school. Two variables were considered as outputs of the education process. The average score achieved in the standardized final school-leaving examination and a financial award given to teachers of students who successfully participated in competitions.

Efficiency estimates under variable and constant returns to scale were calculated. The efficiency scores obtained under VRS ranged from 76.77 to 100 while the least efficient units achieved efficiency score of only 35.35 under CRS. We assessed that DEA under CRS disadvantaged the schools that use higher level of inputs but their students achieved top results in the final school-leaving examination. As we believe that it is harder, therefore requiring more resources, to score additional points when a student achieves 90% in a final exam that when he or she achieves 60%. Thus, it seems more sensible to consider only the efficiency scores obtained under VRS assumption in the second stage. Based on VRS efficiency scores we concluded that the most efficient schools on average were located in the region of Vysočina and Zlínský kraj. The least efficient schools turned out to be situated in Ústecký kraj.

In the second-stage we performed Tobit regression which revealed that efficiency is positively influenced by school size, share of female students and share of students that attended six or eight year study program. The results furthermore suggested that a grammar school offering a vocational study program typically achieves lower efficiency. The obtained results were in line with our expectation. These finding were also supported by OLS regression with White heteroskedasticity-consistent standard errors.

We are aware that one may be skeptical about one of our inputs — the rejection rate — being a good proxy for students' ability prior admission. Results of standardized entrance exams or nation-wide standardized tests of pupils upon completion of elementary school would be a much more accurate and objective measure of students' ability prior enrollment to a grammar school. As a matter of fact, starting next year, a nation-wide testing in mathematics, Czech and English language of students in their final year of elementary school is going to be carried out. Results of these tests will definitely be valuable for the future research in this field.

Bibliography

ADKINS, L. C. & R. L. MOOMAW (2007): "Analyzing the Technical Efficiency of School Districts in Oklahoma." *The Journal of Economics* **33(2)**: pp. 41–61.

AIGNER, D., C. LOVELL, & P. SCHMIDT (1977): "Formulation and estimation of stochastic frontier production function models." *Journal of Econometrics* **6(1)**: pp. 21 – 37.

ALEXANDER, W. R., A. A. HAUG, & M. JAFORULLAH (2010): "A two-stage double-bootstrap data envelopment analysis of efficiency differences of New Zealand secondary schools." *Journal of Productivity Analysis* **34(2)**: pp. 99–110.

ARABMAZAR, A. & P. SCHMIDT (1982): "An investigation of the robustness of the tobit estimator to non-normality." *Econometrica* **50(4)**: pp. pp. 1055–1063.

BANKER, R. D., A. CHARNES, & W. W. COOPER (1984): "Some models for estimating technical and scale inefficiencies in data envelopment analysis." *Management Science (pre-1986)* **30(9)**: pp. 1078–1078.

BRADLEY, S., G. JOHNES, & J. MILLINGTON (2001): "The effect of competition on the efficiency of secondary schools in England." *European Journal of Operational Research* **135(3)**: pp. 545–568.

BROWN, C. & R. MOFFITT (1983): "The effect of ignoring heteroscedasticity on estimates of the tobit model." *Working Paper 27*, National Bureau of Economic Research.

CHAMOUTOVÁ, D. & J. VOJTĚCH (2013): "Vývoj vzdělanostní a oborové struktury žáků a studentů ve středním a vyšším odborném vzdělávání v Čr a v krajích Čr a postavení mladých lidí na trhu práce ve srovnání se stavem v evropské unii." *Technical report*, Národní ústav pro vzdělání.

Charnes, A., W. Cooper, & E. Rhodes (1979): "Measuring the efficiency of decision-making units." *European Journal of Operational Research* **3(4)**.

Coelli, T., D. P. Rao, C. O'Donnell, & G. Battese (2005): *An introduction to efficiency and productivity analysis.* New York, NY: Springer, 2. ed. edition.

Cooper, W. W., L. M. Seiford, & J. Zhu (2004): "Data Envelopment Analysis - History, Models and Interpretation." In W. W. Cooper, L. M. Seiford, & J. Zhu (editors), "Handbook on Data Envelopment Analysis (International Series in Operations Research & Management Science)," chapter 1. Kluwer Academic Publishers.

Davutyan, N., M. Demir, & S. Polat (2010): "Assessing the efficiency of Turkish secondary education: Heterogeneity, centralization, and scale diseconomies." *Socio-Economic Planning Sciences* **44(1)**: pp. 35 – 44.

De Prins, D., L. Simar, & H. Tulkens (1984): "Measuring labour efficiency in post offices." In P. P. M. Marchand & H. Tulkens (editors), "The Performance of Public Enterprises: Concepts and Measurement," pp. 243–267.

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

Foersund, F. R. & N. Sarafoglou (2000): "On the origins of data envelopment analysis." *Memorandum 24/2000*, Oslo University, Department of Economics.

Franta, M. & T. Konečný (2009): "Stochastic Frontier Analysis of the Efficiency of Czech Grammar Schools." *Czech Sociological Review* **45**: pp. 1265–1282.

Johnes, J., N. Burney, M. Al-Enezi, & M. Al-Musallam (2010): "An assessment of efficiency of public schools in Kuwait using data envelopemnt approach (DEA) and Tobit regression." *Journal of Development and Economic Policies* **12(2)**: pp. 5–33.

Kirjavainen, T. & H. A. Loikkanen (1998): "Efficiency differences of finnish senior secondary schools: An application of DEA and Tobit analysis." *Economics of Education Review* **17(4)**: pp. 377 – 394.

Maddala, G. (1993): *Limited-dependent and Qualitative Variables in Econometrics*. Number 3 in Econometric Society monographs. Cambridge [u.a.]: Cambridge University Press, repr. edition.

McDonald, J. (2009): "Using least squares and tobit in second stage DEA efficiency analyses." *European Journal of Operational Research* **197(2)**: pp. 792 – 798.

OECD (2010): "Pisa 2009 results: Executive summary."

OECD (2012a): "Education at a Glance 2012: OECD Indicators." OECD Publishing.

OECD (2012b): "How are girls doing in school – and women doing in employment – around the world?" Education Indicators in Focus.

Ramalho, E. A., J. J. Ramalho, & P. D. Henriques (2010): "Fractional regression models for second stage DEA efficiency analyses." *Technical report*, University of Evora.

Simar, L. & P. W. Wilson (2007): "Estimation and inference in two-stage, semi-parametric models of production processes." *Journal of Econometrics* **136(1)**: pp. 31 – 64.

Stupnytskyy, O. (2004): "Secondary schools efficiency in the Czech Republic." Center for Economic Research and Graduate Education and the Economics Institute, Prague.

Wooldridge, J. M. (2008): *Introductory Econometrics - A modern approach*. Cengage Learning Emea, 4 edition.

Appendix A

Additional Tables and Figures

Table A.1: Overview of previous studies focusing on efficiency analysis of educational institutions

Author	Sample	Method	Inputs	Outputs	Notes
Bradley *et al.* (2001)	Secondary schools in England	DEA	• Students ineligible for free school meals • Qualified teachers	• Attendance rate • GSCE examination results (% of A*- C grades)	Use Tobit to relate efficiency to: • Type of school (independent, public etc.) • Degree of competition between non-selective schools and selective ones • All girls, All boys or co-educated • Unemployment rate • Population density • % professional and managerial workers • Expenditures on teachers and books • School size • Student-teacher ratio
Adkins & Moomaw (2007)	School districts in Oklahoma	SFA	• Instructional and other expenditures per student • % students eligible for federally funded lunch • % nonwhite students • % students with limited English proficiency	• Test scores – average percentiles of state examination for grades 3, 7, 9 and 11	Explain inefficiency using the following variables: • Average teacher salary • Number of full time teacher equivalent • Teachers' years of experience • % staff with an advanced degree • Students-teacher ratio • Total enrollment

Continued on next page

Table A.1 – continued from previous page

Author	Sample	Method	Inputs	Outputs	Notes
Alexander *et al.* (2010)	Secondary schools in New Zealand	DEA	• Administration expense • Expenditure on learning resources • Depreciation expense • Expenditure for raising local funds • Property management expenses • full time equivalent teachers • students at each of years level • teacher aides	• Results from national examinations (several years)	Use truncated regression to relate efficiency to: • State owned • Years 9-13, 7-13 or 1-13 • Main, secondary, minor urban or rural area • All girls, all boys or co-educated • Teacher experience • Teacher qualification
Kirjavainen & Loikkanen (1998)	Senior secondary schools in Finnland	DEA	• Teaching hours per week • Non-teaching hours per week • Experience of teachers • Admission level • Educational level of students' parent	• Number of students who passed their grade (moved up) • Number of graduates • Score in compulsory and elective subjects in national school-leaving examination	Use Tobit to relate efficiency to: • School size • Class size • Private school • Urban, Densely populated area • Parents' education • Heterogenity in grades • Actual/acceptable expenditure • State grants Use Jackknife procedure to test robustness of the efficiency estimates

Continued on next page

Table A.1 – continued from previous page

Author	Sample	Method	Inputs	Outputs	Notes
Stupnytskyy (2004)	Grammar schools in the Czech Republic	DEA	• Students skills at admission to grammar school • Classrooms per student ratio • Physical facility index	• Score in Mathematics and Czech language of graduates • Admission rate to university	Use Tobit to relate efficiency to: • Teacher-student ratio • Teachers age • % internal teachers • % female teachers • Fluctuation of teachers • Years director is in function • % male students • Students career advice center • School council • Public relations and meeting with parents • School age • Cooperation with foreign schools • Sorting of students by their skills • Number of students in Use Jackknife procedure to test robustness of the efficiency estimates
Johnes *et al.* (2010)	Public schools in Kuwait	DEA	• Number of teachers • Number of classrooms	• Number of enrolled students (according to the article, highly correlated with number of graduates – more satisfactory measure, but data unavailable incomplete)	Use Tobit to relate efficiency to: • Geographical location of school • % Kuwait nationals teaching staff • Average teaching staff salary • All boys/girls school

Continued on next page

Table A.1 – continued from previous page

Author	Sample	Method	Inputs	Outputs	Notes
Davutyan *et al.* (2010)	Turkish secondary schools by provinces	DEA	• number of teachers • number of classrooms • Average score in a nationwide entrance exam • Standard deviation of quantitative scores in the nationwide entrance exam	• Number of high school students in each provinces • Average quantitative score in the nationwide university entrance exam.	Use OLS and Tobit to relate scale inefficiencies to number of school types in a province and number of secondary schools in a province controlled by a ministry other that Ministry of National Education
Franta & Konečný (2009)	Grammar schools in the Czech Republic	SFA	• Students/teacher • Students/class • Student/school management • Number of students	• Probability of admission index Or Normalized probability of admission index	Explain inefficiency using the following variables: • Share of women in teaching staff • more than 2 foreign languages • Unemployed/population • Unemployed white collars/blue collars • Vacancies/1000 inhabitants • Hourly wage tertiary educated/secondary educated • Foreign direct investment/ inhabitant • District > 80 000 pop. • District > 170 000 pop. • District > 500 000 pop.

Source: author's survey.

Table A.2: Correlation matrix of determinants

	A	B	C	D	E	F	G	H	I	J	K
% female students (A)											
% not qualified teachers (B)	-0.07										
Yearly salary per teacher (C)	-0.10	-0.05									
State(D)	0.22	-0.02	0.01								
Vocational program (E)	0.21	0.14	0.06	-0.05							
% graduates attending a 6 or 8 year program (F)	-0.15	-0.01	-0.13	-0.14	-0.22						
School size (G)	0.07	-0.12	0.12	0.44	-0.04	-0.11					
Small town (H)	0.15	0.07	-0.16	0.21	0.12	0.11	-0.44				
Middle sized town (I)	0.11	-0.07	-0.03	0.07	0.02	-0.10	0.23	-0.40			
Large town (J)	-0.09	-0.13	0.05	-0.10	-0.09	-0.12	0.28	-0.46	-0.25		
Prague (K)	-0.21	0.12	0.19	-0.22	-0.07	0.09	0.04	-0.40	-0.21	-0.25	
% university educated (L)	-0.25	0.02	0.22	-0.30	-0.14	0.01	0.15	-0.56	-0.26	0.19	0.79

Source: MSMT, MF, CSU; authors' computations.

Table A.3: Average efficiency scores by region

Region	Average Efficiency Score	
	VRS	CRS
Hlavní město Praha	91.68	69.93
Jihočeský kraj	92.83	80.25
Jihomoravský kraj	91.77	78.35
Karlovarský kraj	90.38	77.63
Kraj Vysočina	94.35	81.27
Královéhradecký kraj	93.66	82.02
Liberecký kraj	92.83	75.61
Moravskoslezský kraj	91.71	81.73
Olomoucký kraj	92.66	79.29
Pardubický kraj	93.98	84.86
Plzeňský kraj	93.6	83.93
Středočeský kraj	93.25	83.02
Ústecký kraj	88.8	79.56
Zlínský kraj	94.86	83.85

Source: author's computations.

Figure A.1: Regions of the Czech Republic

Source: `http://www.eu2009.cz/cz/czech-republic/regions/regiony-cr-328/`

Table A.4: OLS regression estimates

	Model (1)	Model (2)	Model (3)	Model (4)
Intercept	82.71***	84.87***	84.22***	83.96***
	(4.062)	(2.620)	(2.645)	(2.632)
% female students	0.091***	0.0873***	0.086**	0.095***
	(0.033)	(0.033)	(0.033)	(0.034)
% not qualified teachers	0.061	0.055		
	(0.056)	(0.057)		
Yearly salary per teacher	0.000			
	(0.006)			
State	−0.634	−0.936	−0.772	
	(1.250)	(1.272)	(1.254)	
Vocational program	−5.309***	−5.534***	−5.269***	−4.981***
	(1.278)	(1.271)	(1.420)	(1.527)
% graduates attending a 6 or 8 year program	0.030***	0.026**	0.027**	0.029**
	(0.012)	(0.012)	(0.012)	(0.012)
School size	0.005***	0.006***	0.005**	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
Small town			0.909	
			(0.777)	
Middle sized town	−0.628	−0.648		
	(0.792)	(0.807)		
Large town	−2.026**	−1.128		
	(1.008)	(0.920)		
Prague	−4.125**	−1.737*	−0.685	
	(1.700)	(0.953)	(0.812)	
% university educated	0.164*			
	(0.093)			

Notes: *p<0.1; **p<0.05; ***p<0.01
White heteroskedasticity-consistent standard errors are reported in brackets

Source: author's computations.

Appendix B

Assumption of Homoskedasticity and Normality of Residuals

To test for homoskedasticity we use the special White test[1]. However, we test homoskedasticity on a sample of residuals that excludes the units of observation that reach full efficiency. The reason for restricting the sample is shown in Figure B.1.[2] The problem is that we do not observe the values of the latent variable for observations with efficiency scores equal to 100. Points above the upper limit are unknown as only the red points on the upper limit line are observed. The residuals u^* for the latent variable are the desirable residuals, but unfortunately, we are unable to compute them. We are only capable of calculating the residuals as the difference between the fitted values and observed variables depicted as $\tilde{u}$, that are clearly inappropriate.

The F-statistic p-values of white test on a restricted sample for model 1, model 2 model 3, and model 4 are 0.0206, 0.1292, 0.092, 0.038 respectively. Therefore, we do not reject homoskedasticity of residuals at 1% significance level for model 1 and 4, 10% for model 2, and 5% for model 3.

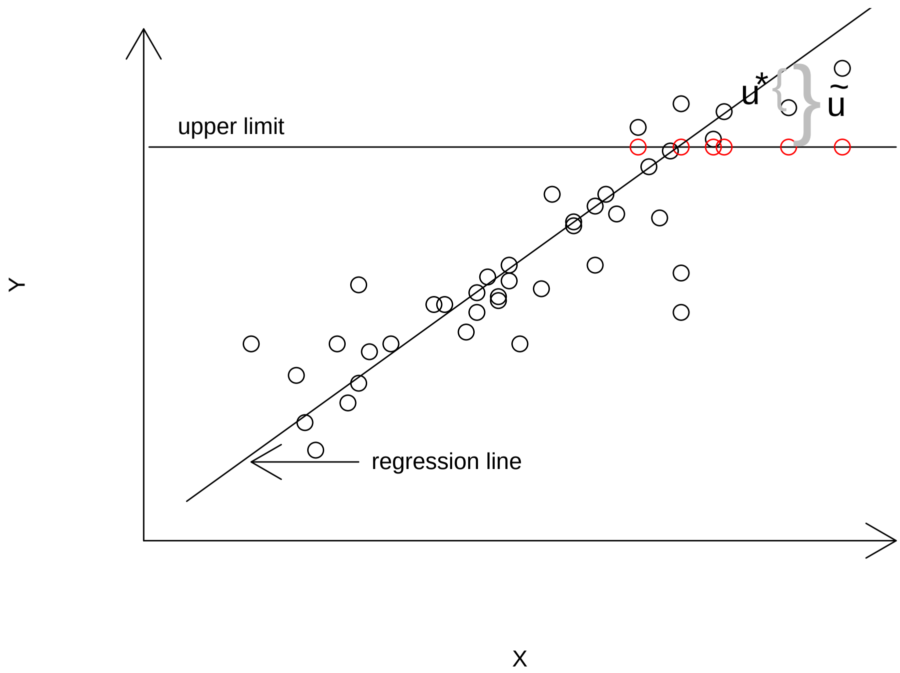
As far as normality of residuals is concerned, histogram of residuals together with a Q-Q plot, comparing the normal distribution with the sample distribution of residuals, for each model is show in Figure B.2. We observe that the distributions are not perfectly normal but according to our opinion, there is not a substantial departure from normal distribution. Therefore, we believe that the estimates are not inconsistent. We are aware that this approach for testing residuals for heteroskedasticity and normality in Tobit model is rather

[1] Squared residuals $\hat{u}$ are regressed on fitted values $\hat{y}$ and squared fitted values $\hat{y}^2$ (Wooldridge 2008).

[2] For illustrative purposes, we assume simple regression model in Figure B.1.

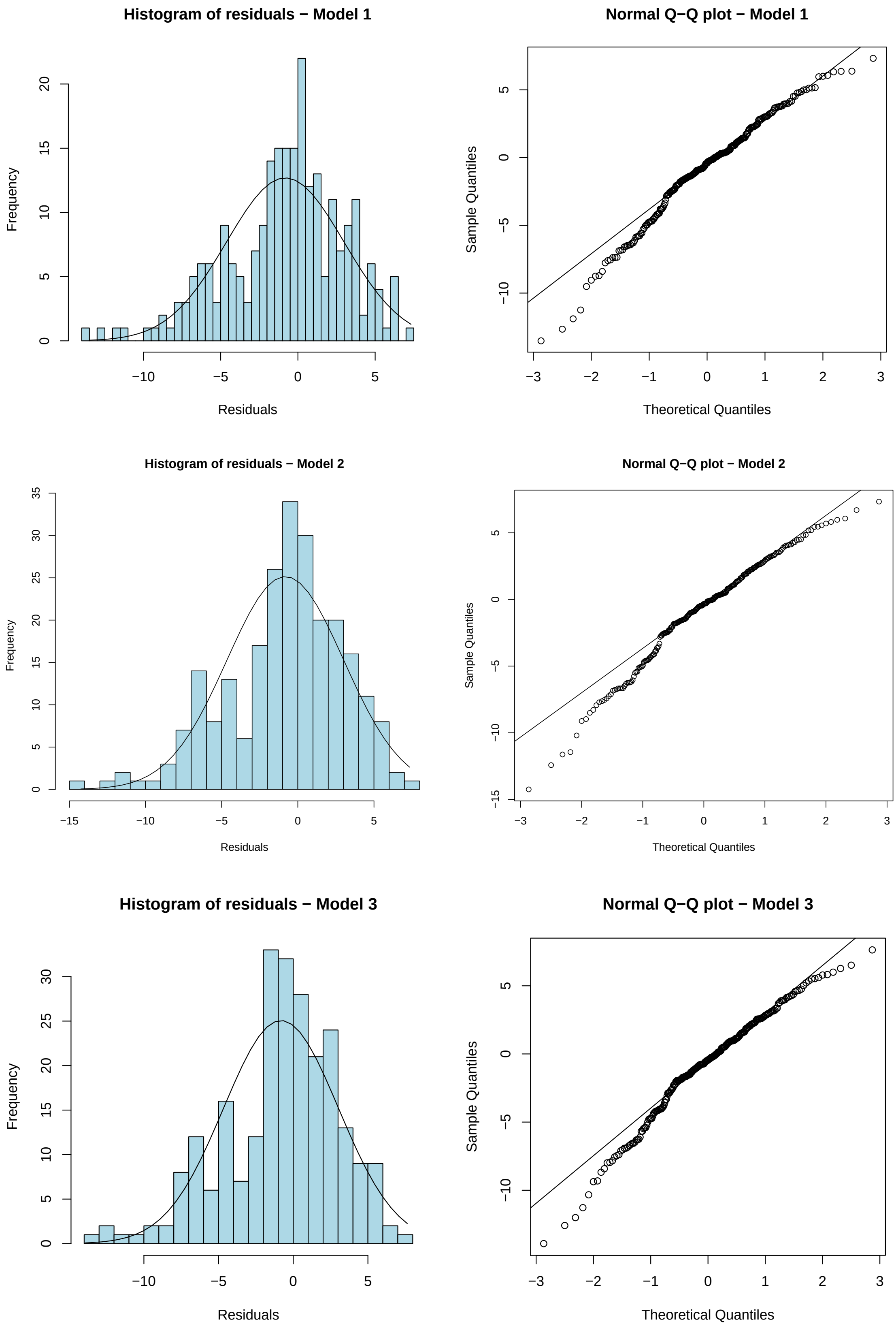
unconventional but there is not a standard commonly accepted testing strategy for these assumptions in the Tobit model.

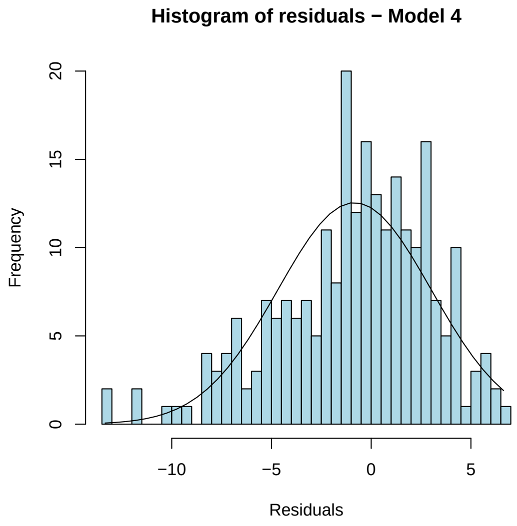
Figure B.1: Latent vs observed residuals

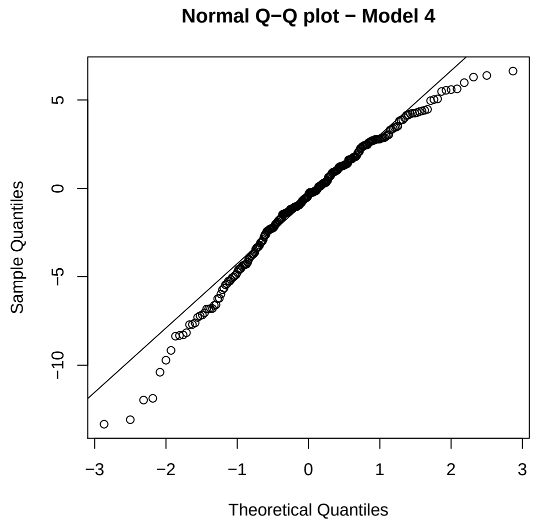


Source: author's layout.

Figure B.2: Histograms and Q-Q plots of residuals







Source: author's computations

Appendix C

Bachelor Thesis Proposal

Author	Kateřina Červenková
Supervisor	PhDr. Lenka Šťastná, Ph.D.
Proposed topic	Efficiency Analysis of Grammar Schools in the Czech Republic

Each secondary school operates with a certain amount of resources, such as quality and quantity of teaching staff, that are exploited in the educational process. However, the efficiency of utilized resources, meaning the results achieved by its students when amount of inputs is taken into consideration, differs across educational institutions. The thesis focuses on evaluation of relative efficiency of Czech grammar schools and aims to seek sources of inefficiency.

The first part of the thesis introduces basic methods for assessment of relative efficiency of comparable decision making units (data envelopment analysis, stochastic frontier analysis etc.), ordinarily used for evaluation efficiency of educational institutions, and mentions their advantages and shortcomings. The empirical part deals with application of selected methods to the most recent data provided by the Ministry of Education, Youth and Sports to evaluate relative efficiency of grammar schools in the Czech Republic. Furthermore, it discusses the comparability of the results acquired from different methods. Then the inefficiency is quantified and analyzed. The last part aims to identify common characteristics of units at similar level of inefficiency e.g. size of institution, state or private founder, geographical location etc. to capture possible sources of inefficiency.

Outline

1. Introduction

2. Description of Efficiency Analysis Methods
3. Application of Selected Methods to Czech Grammar School Data
4. Evaluation and Comparison of the Results
5. Identification of Possible Sources of Inefficiency
6. Conclusion

Core bibliography

1. Coelli, T. J., D. S. Prasada Rao, C. J. O'Donnel& G. E. Battese (2005): " An Introduction to Efficiency and Productivity Analysis." 2nd Edition (Springer, USA)

2. Cooper, W. W., L. M. Seiford, & K. Tone (2007): " Data Envelopment Analysis: A Comprehensive Text With Models, Applications, References and DEA-Solver Software." 2nd Edition (Springer, USA)

3. Cooper, W. W., S. Li, L. M. Seiford& K Tone (2001): " Sensitivity and stability analysis in DEA: Some recent developments." *Journal of Productivity Analysis* **15(3)**: pp. 217–246.

4. Hanushek, E. (1987): "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature* **24(3)**: pp. 1141–1177.

5. Kirjavainen, T. & H., A. Loikkanen (1998): " Efficiency differences of finnish senior secondary schools: an application of DEA and tobit analysis." *Economics of Education Rewiev* **17(4)**: pp. 337–394.

6. Kumbhakar, S. C. & C., K. Lovell (2000): " Stochastic Frontier Analysis." Cambridge University Press, New York.

7. Ruggiero, J. & D., F. Vitaliano (1999): " Assessing the efficiency of public schools using data envelopment analysis and frontier regression." *Contemporary Economic Policy* **17(3)**: pp. 321–331.

Author   Supervisor