Charles University in Prague

Faculty of Social Sciences

Institute of Economic Studies

BACHELOR THESIS

2010 **Michal Lebovič**

Charles University in Prague

Faculty of Social Sciences

Institute of Economic Studies

BACHELOR THESIS

Efficiency of public spending

Author: **Michal Lebovič**

Supervisor: **MRes PhDr. Jan Zápal**

Academic Year: **2009/2010**

Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a použil pouze uvedené prameny a literaturu.

V Praze dne 15. 5. 2010

Michal Lebovič

Acknowledgments

I would like to thank my consultant, Jan Zápal, for his useful comments and valuable feedback.

Abstract

This thesis aims to offer a comprehensive introduction into the topic of efficiency measurement in the public sector. Firstly, usual definitions and concepts of efficiency are introduced. Attention is then turned to the description of various factors and problems specific for public sector that are crucial to efficiency measurement. It is shown that these factors preclude the use of general (private sector) efficiency measurement methods or demand their modification. The most common methods of analysis are then introduced and their relative advantages and disadvantages in the environment of public sector are explained. Finally the thesis outlines the possible uses and benefits of efficiency measurement, including the use in the economic policy-making, but also points out the limits inherent to this analysis in the current stage of development.

Abstrakt

Tato práce si klade za cíl nabídnut čtenáři vyčerpávající úvod do problematiky měření efektivity ve veřejném sektoru. Po představení obvyklých definic a konceptů efektivity je pozornost obrácena k veřejnému sektoru a k jeho specifickým rysům a problémům, které jsou z hlediska měření efektivity klíčové. Ukáže se, že běžné přístupy používané v soukromém sektoru je nutno výrazně modifikovat. Dále jsou představeny běžné metody analýzy efektivity veřejných výdajů a jsou porovnány jejich výhody a nevýhody. Na závěr tato práce nastiňuje možnosti využití měření efektivity ve veřejném sektoru, včetně otázky tvorby ekonomických politik, ale také upozorňuje na omezení, která jsou této analýze v současné fázi vývoje vlastní.

Table of contents

List of figures:

List of tables

1. Introduction

Why is efficiency of the public sector important? In most countries public sector is large enough to constitute a significant part of the economy as a whole. It implies that public sector performance has profound effects on the performance of aggregate economy. First the performance of the private sector is affected by the capability of public sector to ensure a solid infrastructure and a fair, predictable economic environment. Secondly, the activities of a relatively large public sector must be financed by correspondingly high taxes - and these affect the decision making process in the private sector and cause distortions in the economy as they reallocate the resources from private uses to (arguably less effective) public uses.

Improving efficiency means either producing larger (or better) output using the same resources or (viewed from the other side) reducing the amount of resources used to produce the same output. In short it is an improvement in "value for money". Thus high efficiency is indeed a desirable property. Yet pursuing a high efficiency in the public sector means to identify and analyze its inefficiencies in the first place. That exposes one of the reasons why we should pay attention to public sector efficiency - it is a potentially productive exercise that might generate significant returns. Only by studying efficiency and isolating its effect from the effects of exogenous environment in which the producers operate we can explore the hypothesis on the sources of variance of efficiency and productivity. This in turn might help us formulate reasonable economic policies leading to improvements in efficiency, putting scarce public resources to much better use.

We can also study efficiency as a possible measure of performance in the public sector. In private sector we usually compare the performance of companies on the basis of profitability; however there is no such easy yardstick in public sector where output is often unpriced and profit might be completely irrelevant. There is no doubt that some measure of performance is necessary in order to provide transparency to the democratic debate on the advantages and disadvantages of public sector service and its benefits to society - notion of profitability is thus often substituted by efficiency.

But past decades of research have revealed that constituting a proper methodological framework allowing to quantify efficiency (or inefficiency) in the public sector services is a very hard conceptual challenge. I should stress here that the focus of my paper is set on introduction of general issues of public sector efficiency measurement rather than on technical description of various techniques. This is driven by my belief that the credibility of results of efficiency assessment exercise is constrained more by general problems tied with unclear definitions and problematic measurement of some aspects of public sector provision rather than by exactness of the technical apparatus alone. Amongst other things it is now clear that credible and systematic efficiency measurement is impossible without an elementary cooperation of researchers with the public sector itself - a prerequisite of any reasonable analysis is a solid dataset, yet as I show in this paper obtaining plausible data is often extremely difficult and thus the possibilities of application and empirical testing of methods are restricted.

The remainder of the paper is organized as follows. Section 2 introduces the concepts of productivity, efficiency and effectiveness as they are apprehended in efficiency measurement literature. Section 3 is devoted to various specific characteristics of public sector that distinguish it significantly from the private sector when it comes to efficiency. Section 4 extends section 3 in the sense that it explains the impact of these factors on the efficiency measurement exercise and introduces the main problems associated with this kind of analysis. In section 5 the most commonly used techniques are briefly introduced and discussed and finally section 6 presents the possibilities and limitations of use of the results.

2. Efficiency and productivity

2.1 Introduction

In this part of the text an intuition (rather nontechnical) behind the concepts of efficiency and productivity will be provided and the relationship between the two described. It is quite common when comparing two or more producers to refer to them as more or less productive, or more or less efficient. But what exactly is meant by these notions and to what degree they refer to the same attribute of the producer? For the purposes of this paper there will also be another area of interest – how do these concepts translate into the public sector and what the possible benefits of analysing efficiency and productivity of public expenditure are.

As I will later describe productivity and efficiency both describe firm's ability to transform inputs into outputs. Understanding this process has always been one of the important areas of economic research. Identifying the sources of waste can be viewed as a tool that contributes to rationalizing the allocation of scarce resources to satisfying various (unlimited) needs. Also in the public sector the efficient use of resources is increasingly recognized as extremely important.

2.2 Productivity

The definition of producer's (or service provider's) productivity is rather straightforward and in its most general form it can be simply stated as a ratio of its inputs to its outputs. If we think about a simple case of a producer (further in the text, by term „producer" I will refer to any agent that takes a set of inputs and transforms them to a set of outputs – I stick to this very general definition to encompass all various public sector producers and public service providers generally) with a single input and a single output this should also be fairly easy to calculate – provided that both inputs and outputs are easily quantifiable and measurable (but there is no need for the input and output to be measured in the same units as the ratio only tells us how much output is produced per unit of input). It offers us a shallow insight on the producer's ability to transform inputs to outputs without taking any specific production-related factors in consideration. We can also easily identify the productivity growth as a difference between the growth of output and the growth of input. In the more realistic case of multiple inputs and outputs these must be aggregated in some meaningful way in order to calculate productivity (so with input and output being a weighted sum of its components it remains a ration of two scalars).

Critical question remains – what are the potential sources of productivity variance of two different organizations in the public sector (operating in the same field, e.g. providing the same service in different jurisdictions)? Smith (1988) suggests the following dimensions of differences:

1) Organizations might pursue different goals and objectives.

2) Particular needs of an area of operation might vary, even when the objectives are set the same.

3) Organizations can face different costs because of differences in operating environment.

4) Organizations can display a difference in managerial competence in the use of resources.

5) There might be a mistake in the data.

Note that 1) is of particular importance in public sector, for example when we compare producers in different jurisdictions, who might have certain degree of freedom in choice of extent and quality of services provided. Points 2) and 3) overlap, but they are not exactly the same. Simpson (2008) notices that in contrast with private sector, for the public sector managers the choice of area of operation is not at their discretion. It is reasonable to believe that private firms will take environment–specific circumstances under consideration when forming their business strategy and their success in doing so should be well reflected in financial statement. On the other hand for public sector producers it is often necessary to operate in conditions that are somewhat ungenerous for their business and they do not have an option to move the operation elsewhere (or even stop it completely), instead they are expected to deliver certain standard level of service no matter what. Point 2) notes on the fact that specific environment circumstances can advocate for a different mix of inputs used and outputs produced in order to deliver certain required standard. Taking police as an example, costs and resources used for assuring certain level of crime prevention will depend heavily on the socio-economic conditions in the region. Other area specifics might include for example quality of infrastructure, density of population or other demographic or geographic condition. 3) then suggests that - given the facts stated above - services must often be provided given costs of labour and capital prevailing in the area of operation, which might differ dramatically (thus might as well result in different mix of capital and labour used).

Obviously 4) is the core of our interest. It is an element we would like to isolate from the others. While usually not much can be done in other four dimensions this one calls for immediate improvements when it is revealed as a source of variance of performance. There might be more sources of inefficiency than this and I will discuss them in more detail later. For now it is enough to say that this is what I will later describe as technical efficiency.

Finally point 5) refers to various problems in data collection, measurement and credibility. Ignoring various errors, data can be distorted for example by creative accounting or (more innocently) by varying accounting standards and practices and data collection and processing methods in different countries and jurisdictions.

From this point of view it is interesting to compare the difficulty of decomposing performance variance in public and private sector. Point 1) is usually negligible as the objective of vast majority of private enterprises is profit maximization. Also points 2) and 3) do not require a direct attention of an analyst because as I already hinted these are endogenous in private sector and can be subsumed under point 4) – they are already reflected in firm’s operating decisions. 5) also cannot be fully avoided but it is a potential problem of any empirical research. In private sector it is at least alleviated by the fact that most of the data are related to profitability thus they can be measured in monetary units which is not always the case in the public sector where we often miss information on prices (especially for the outputs that are, by definition, not sold on the market). So the analysis of financial statement of private company is mainly interested in 4) and 5) and even for this task the range and complexity of methods is immense. Similar analysis in public sector presumes that (in this case) exogenous problems of 1), 2) and 3) are controlled for and even so there are further perils that I will describe in later chapters.

Smith's approach is very useful in listing various exogenous factors affecting public sector producers' productivity. Nevertheless it does not tell a lot about a specific form that inefficiencies of production might take as all endogenous factors are summarized under a bit vague point 4). A study of Guellec and Potterie (2001) for OECD uses a different approach that is more specific in addressing various inefficiency factors. It attributes the sources of variation of productivity to:

a) Differences in production technology.
b) Differences in the scale of operation.
c) Differences in operation efficiency.
d) Differences in operating environment.

This approach takes the previous one and turns it on its head: it takes points 1), 2) and 3) and summarizes them under point d) which is now supposed to include all exogenous factors. Then it focuses attention on point 4) and tries to identify more possible endogenous sources of inefficiency. We would be naturally interested in isolating such factors that are (at least to a certain degree) under the discretion of management or policy makers. Point c) is probably closest to what Smith addressed by „managerial competence“ and while production technology and scale of operation might not always be easily and quickly changed motivating a higher competence of managers and seeking the best practice could lead to an easy improvement in value-for-money (thus it should be in the interest of policy makers).

2.3 Efficiency and effectiveness

As outlined above efficiency is about comparing observed and optimal values of inputs and outputs and can be viewed as one of the dimensions of productivity, perhaps as a "comparative dimension". In later text I will address several kinds of efficiency – namely technical efficiency, allocative efficiency and scale efficiency.

Koopmans (1951) defined so called technical efficiency (although he originally called it „productive efficiency" the phenomenon was later referred to as technical efficiency to be distinguished from other kinds of efficiency) in the following way - being technically efficient either means that we produce a maximal possible amount of output from given amount of input or that we use a minimal possible amount of input to achieve certain desired level of output (thus efficiency can be understood as output-enhancing or input-preserving). Put another way an efficient producer operates on the production possibility frontier and (1) it cannot increase any output without also increasing inputs and (2) it cannot decrease any input without also decreasing outputs. But the true foundation of modern efficiency measurement theory was laid by the work of Debreu (1951) and then Farrell (1957). Farrell worked with Koopman's definition of technical efficiency and suggested a measure for it – he wrote that (in)efficiency could be measured by the firm's distance from the production possibility frontier or more precisely he said that it is a maximum equiproportionate (i.e. radial) reduction in all inputs feasible with given technology and outputs (from the input-preserving view). Although Farrell was originally referring to private sector his thoughts also became an inspiration for the efficiency measurement in the public sector. Since that time the methods have been significantly improved and the investigation of efficiency became a frequent tool of economic analysis.

For a comprehensive discussion on various concepts of efficiency I recommend Forsund, Lovell and Schmidt (1980). They offer a complete survey on what different authors mean by term "efficiency", describe all forms it can take and how it can be decomposed. A more technical introduction can be found e.g. in Farsi, Filippini and Greene (2006) who illustrate on an example of electricity distribution sector how these problems can be tackled from the econometrics point of view. Such discussions are beyond the scope of this text - it should be emphasized here that the focus of this paper as well as the focus of majority of efficiency measurement literature is mainly set on the technical efficiency derived from Koopmans (1951): *A producer is technically efficient if and only if it is not possible to improve any input or output without worsening some other input or output.* For a basic insight, however, I will offer in this section a description of other types of efficiency most typically assumed in the literature.

Using what was written above measurement of technical efficiency could be broken into three steps:

1) Define and measure inputs and outputs.
2) Define what is optimal (or define firm's potential – or in Farrell's words find a relevant production possibility frontier).
3) Compare observed and optimal values.

However even this very simple approach disguises several conceptual problems. First - what inputs and outputs to include and how to measure them? This issue was touched by Knight (1933, 1965) who noted that if the operation of producers was defined precisely, accounting for *all possible forms* of output they produced (including variables such as in-job leisure of employees) and inputs they used, there would be no point in assessing productivity or efficiency – since neither matter or energy can be destroyed or created all producers would achieve the same productivity (or efficiency) score of 1. Naturally the provider would be 100% efficient in producing its unique set of outputs thus Knight redefined productivity as a ratio of *relevant* outputs to *relevant* inputs. It is not our intention to find out whether the provider is efficient in production of all of its outputs (it obviously is) but if it is efficient in production of certain outputs that are stated as its objectives and/or that the society values. However defining such outputs (and inputs) can be a very difficult task in case of many public services. Public expenditure often has other broader objectives than only achieving certain amount of observable or physical outputs (in an example of hospitals it can be a number of patients treated). It can pursue some socio-economic or demographic goals (like decreasing the infant mortality rate or making the service equally accessible to all socio-economic groups of citizens), this kind of outputs will be later referred to as outcomes. Further also quality of the service matters (increase in output/input ratio can be undesirable when it comes on the expense of quality). This should all somehow be incorporated in the analysis.

If we are able to identify and measure all of the relevant variables second problem immediately arises – how shall we define the weights to combine these variables when the information on prices or valuation is missing? This issue is one of the most controversial and I will discuss it in more detail in later chapters. Finally our third problem is how to define what is optimal or feasible. Because it is extremely problematic to find the absolute potential of producer efficiency is usually defined in a comparative fashion – the producer's inputs and outputs are compared to those of other producers and thus potential is described by the best practice. Our situation is even more problematic when there is only a single producer operating in a given field – in such cases we often have to turn to international comparisons or to analyze time series data of a single producer. Nevertheless it is the treatment of the second and third problem in which the different methodologies for estimating efficiency substantially differ.

Increase in the technical efficiency can be interpreted as a move towards a production possibility frontier which gives the maximum attainable output given the amount of inputs. Being technically efficient is of course a desirable property of a producer. It ensures that given the chosen mix of inputs and outputs the performance cannot be improved i.e. there is no waste in the production process. Technical efficiency however does not tell us anything about the optimality of chosen mix of inputs and outputs and because not every technically efficient production necessarily makes economic sense it should be examined as well. This takes us to another concept of efficiency called „allocative efficiency" which looks for an optimal mix (or allocation) of resources and outputs taking into account the costs and benefits adherent to each production possibility. High degree of technical efficiency of production does not guarantee an effective allocation of public sector spending when a higher output could be easily achieved by choosing a different set of inputs (for instance in education a balanced mix of inputs such as teachers and books is needed – using only one and omitting the other would probably not lead to very good results no matter how efficiently the input was used). Assessing allocative efficiency requires a deeper understanding of a service in question and also the information on input prices so that the costs and benefits can be sensibly identified and compared.

Yet another concept of efficiency focuses on the scale of operation – from here „scale efficiency“. Scale of production is often somewhat exogenous for companies in public sector (for example it might be given by the number of inhabitants of a jurisdiction). Nevertheless the differences in this aspect might lead to advantages or disadvantages of individual producers. Thus when comparing their efficiency it is a factor that should be controlled for if a researcher believes that scale of production matters or in other words that returns to scale are not constant.

Regarding the efficiency of the public spending there is one more important issue to be reminded in this section. Many of the items of public spending are meant to achieve some broader final objectives than just generating certain amount of quantifiable outputs – e.g. they might be meant to improve some socio-economic or demographic indicators affecting the society. Education expenditure produces some direct outputs which can be the number of graduates and their final examination scores but what really matters at the end is the overall level of literacy in the population or composition and qualification of labour force. Most often these final objectives are related to welfare, growth or equity. The analysis of public sector expenditure should definitely take into account to what degree is the expenditure successful in fulfilling such objectives, which are of course subject to a political choice. Impacts of the spending on these objectives are usually denoted as „outcomes“ in the public sector efficiency literature and the relation between inputs/outputs and outcomes is denoted as effectiveness (rather than efficiency, which usually refers to the relation between inputs and outputs).

To conclude this section I present a figure from Dierx, Ilzkovitz and Mandl (2008, pp.3) that represents the notation and structure of efficiency framework most commonly assumed in the literature and summarizes the relation of concepts introduced in this section:

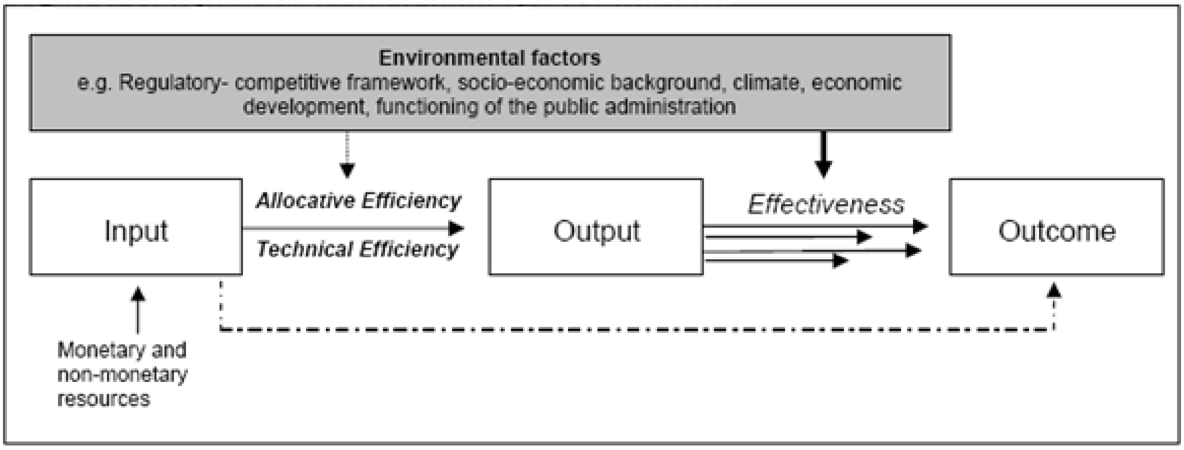

Figure 2.1. Conceptual framework of efficiency and effectiveness. Source: Dierx et al. (2008).

3. Public sector and efficiency

3.1 Introduction

First important step in our exercise before we proceed to more technical description of efficiency measurement is to understand what makes public sector so special in assessing performance and efficiency. There are several differences to private sector that make the measurement of public sector efficiency a more complicated task and that make the standard private sector measurement methods difficult or even impossible to implement. Full discussion of public spending, its structure, rationale and scope is outside of the focus of this paper and can be reviewed elsewhere – Tanzi and Schuknecht (2000) offer an excellent discussion on the role and development of public expenditure over the past century. Here I will only briefly conclude and concentrate on the aspects that are somehow related to performance, efficiency or effectiveness as I explained them in previous section. Aim of this chapter is to go through the substantial differences between the two sectors, explain how exactly they affect the efficiency assessment and outline how they can be dealt with. But I think it is also helpful to understand the problem from the other way round, i.e. how can the factors in question help to explain inefficiency, so I will also devote some space to this perspective as well. On the most general level, we could summarize the most important differences between private enterprises and public service providers into four components (and as I will show later they often interlap in their effects): different ownership structure (subsection 3.2), different competitive conditions and different objectives and constraints (subsection 3.3). These differences result in a severe practical complication for efficiency measurement in public sector – there is a lack of a single objective measure for evaluation of a producer's activity (this phenomenon is treated separately in subsection 3.4).

3.2 Ownership

In private sector enterprises are owned by their stockholders, whereas the "ownership" of public sector entities is dispersed amongst the public, which has several implications. First this form of ownership suggests - given the greater dispersion and nontransferability of ownership rights - that monitoring of performance and efficiency might be more problematic and thus performance incentives are weaker than in private sector, as Alchian (1965) and other authors argue. The principal-agent problem seems to be even more severe here than in the private sector.

In economics, principal-agent problem describes the relationship and incentives of those, whose capital is used to finance an activity (principal) and those who act on their behalf (agents). Agents are hired by principals to fulfil their interests but in the situation of information asymmetry there might be space for them to follow the interests of their own, and these might be more or less inconsistent with the ones of a principal.

In private sector this problem is represented by the relationship of an investor and a management. Investors hire managers and want them to ensure that their capital is put in the best use possible. Some form of communication then must be established to allow the principal to have some degree of control over agent's actions, to be able to judge his performance and to impose sanctions whenever necessary. The most important communication of this kind is the periodic finance statement. Analysing this and other information investor can decide whether to buy, hold or sell equity of a firm (where selling as a withdrawal of funds is a sanction for poor performance).

Wintrobe (1997) documents that in public sector system of accountability is much more complicated (if defined at all) and even less transparent. The voters who elect the government and pay the taxes to finance it might be considered as the principals. Government should in principle implement their wishes and thus can be treated as agent accountable to the electorate. But principal's sole sanction is a vote in the ballot or in some cases also moving to another municipality, country etc., both of these being very far from perfect controlling mechanisms. Situation gets more complicated as the government must hire managements to carry out its agenda (service provision) - and that is another line of accountability. Finally, the citizen enters the system second time as the clients of the provider of a service. Information asymmetry in this relationship is much more persistent in public sector than in private sector, where it is assumed to be reduced by advertisement, experience, competition etc. (so it is not paid much attention in private sector).

The key conclusion here is that with the dispersion of ownership rights and consequent difficulty or impossibility of organizing the "owners" to impose immediate and effective sanctions the civil servants are probably less likely to be punished for their under-performance than their private sector counterparts which creates weaker incentives for pursuing efficiency. Also the motivation of the public to monitor the efficiency of public spending is low given the costs and limited possibility to impose sanctions so the difference between the pressure on public and private sector managers is magnified (stakeholders hire financial analysts to do the job for them, but in the public sector this is only done sporadically by academic researchers or in some countries by specialized government agencies – but only for specific services). The practical consequence on efficiency measurement is quite straightforward – given the low motivation the demand for monitoring the activity of public producers is low which in turn often results in lack or poor accessibility of data. For more detailed discussion on the agency problems in public sector see Wintrobe (1997).

3.3 Specific objectives and constraints, competitiveness

Another dimension that distinguishes public sector from private one is a different set of objectives and constraints that the managements face. It is because the standard profit maximization objective that private firms’ managers (should) pursue is irrelevant in vast majority of public sector cases. Fox (1999) also argues that public sector managers have less control over a combination of outputs than their private sector counterparts (combination and quantities of services they provide) and thus have limited possibilities to find a way of efficient allocation of resources and services (or put in another way, output vector might be treated as somewhat exogenous). In public sector various other constraints than budget constraints are often active (various community service obligations etc) and it is not always straightforward to incorporate all of them in the model. Correct identification of relevant objectives and constraints is however a very important premise of any efficiency analysis. This point also warns us of the difficulties to develop a fair theoretical framework to compare public and private sector efficiency and suggests that conventional (private sector) methods of measurement would not work.

Also the competitiveness of the market has implications on companies' behaviour with respect to efficiency. Economic theory predicts that greater competition in the market should lead to better performance of competing firms (or that better performers survive). Although public sector companies may also face some competition (in the form of providers from other jurisdictions or even private sector) it is common that they operate in less competitive markets than private companies. In many cases given the lack of competition and (or) the nature of services provided these companies under the patronage of government find themselves in a position similar to monopoly. It raises an important question – what is the implication for efficiency? Hicks (1935) wrote in his Theory of Monopoly that firms dominating the market „...are likely to exploit their advantage much more by not bothering to get very near the position of maximum profit, than by straining themselves to get very close to it. The best of all monopoly profits is a quiet life." This quotation suggests that producers might not be only profit maximizers but that they can also optimize their behaviour with respect to some other objectives. Even though Hicks was originally referring to private sector monopolists this notion is quite well applicable on public sector producers in uncompetitive markets. Hick's thought was further developed by Alchian and Kessel (1962) who introduced a concept of utility maximization (instead of profit maximization) where monopolists' apparent lack of efficiency is attributed to their selection of more unobserved output (leisure) where competitive firms select more of observed output (profit), but both can be considered equally effective in maximizing utility. In case of public providers it would probably not be a choice between profit and leisure - this was only to illustrate that lack of competition creates a space for pursuing various secondary objectives that are often inconsistent with the objective of efficiency.

With enough maneuvering space the executives will maximize their own utility function where performance, as Williamson (1964) argues, is only one of the variables, or more likely, only a constraint on the pursuit of other objectives (and as I outlined above public service provision can be just this kind of environment). Niskanen (1971) applied this proposition to public sector and thought of public sector managers as utility-seeking budget maximizers (instead of output maximizers). This and other authors questioned the effectiveness of decision making in the public sector and believed that it was a common endeavour of managers to benefit from the lack of competitive pressures and imperfections of controlling and sanctioning mechanisms. They can do so e.g. by intentionally creating budgetary slacks or by choosing inefficient combinations of inputs and outputs that satisfy their own objectives. This discussion obviously coincides with the discussion of ownership above and they both lead to very similar conclusions.

3.4. Lack of objective measures

This finally gets us to the most pervasive problem of assessing efficiency in public sector. There are a lot of ways of assessing the performance or efficiency scores of private companies and while there is no agreement on which method is the best they all have a common feature - they all work with one basic variable, the profit or profitability. For a typical enterprise profit is an ultimate measure of its success so assessing efficiency score should not be that hard in principle in private sector - having reliable accounting data, such as the financial statements, in hand. On the other hand there is not such a measure in the public sector. Profit might be completely irrelevant and we have no information on output and (sometimes) input prices (and if so, it is usually not market-driven). Further - going back to principal-agent problem as mentioned above - while in a private company the principal's desires are quite clear (they want a maximal profit), in the public sector different parts of the society might have vastly different ideas about optimal combinations and quantities of outputs and there is no definite guidance on which of these ideas is the right one (for example different voters want different levels of national defence). This is a serious complication for efficiency measurement because it needs to take into account some notions of optimality and these are necessarily ambiguous. This is then primarily a matter of political choice.

All of these problems make the measurement of efficiency in public sector a hard challenge for researchers and pose a question whether there can even exist a unique plausible measure of efficiency or performance (not to mention its correct assessment).

For readers with deeper interest in comparison of performance of the two sectors I recommend a nice summary of the topic in Bozeman and Rainey (2000) who also offer references on a large body of empirical evidence. Much of what has been written above and much of what can be deduced from economic theory leads to an expectation that performance is likely to be more efficient in private than in public sector. However – and it is somewhat puzzling and perhaps it is given by the data limitation or inadequate methodology – the empirical research brought only an ambiguous evidence.

4. Main issues in efficiency measurement

4.1. Introduction, data limitations

Before we proceed to a description of the most frequent methods of efficiency assessment and their specific issues it is worthwhile to note on the problem that is common to all of them. For any assessment to be done one first needs to have a reasonable dataset. Nevertheless obtaining sufficiently wide, plausible and unambiguous data can be a serious problem in the public sector. First of all it need not be available – this should not be a problem on the most aggregated levels of public spending but finding detailed disaggregated data on a specific service might not always be possible and it depends a lot on the country's public expenditure monitoring standards. That is only the beginning though. Not only that we face various problems in data measurement and collection but also it is often problematic to even identify the information relevant to our research – namely the definition of service's output is not straightforward as I will show in this chapter.

As I suggested earlier many problems depicted in this section is caused by the lack of a definite objective that all the public spending would follow. In private sector it is the profit-maximization that suits this purpose and it provides both a solid ground for a comparison of various producers across the sector and also a safe guidance on what data we need to make such a comparison (costs, revenues and other profit-related data is usually enough here). Whole problem was aptly commented by Mintzberg (1996, 79) in what he called the myth of measurement: „Many activities are in the public sector precisely because of measurement problems: If everything was so crystal clear and every benefit so easily attributable, those activities would have been in the private sector long ago."

The measurement should also take in account the matter of quality of the service. Quality is a very important aspect of a service, yet it is hard to define and to measure.

Consequently in the public sector quality indicators are reported even less frequently than quantity indicators and are even less reliable. While most researchers are aware of the importance of quality when comparing different providers of the service, they often fall victim to data constraint in this matter. In private sector this is not a big problem as different quality of the products or services should be well reflected in prices – which in the well functioning competitive environment should reflect both the marginal cost and the consumer’s marginal valuation of the product.

Kevin J. Fox (2002;23) summarizes the list of elements the ideal dataset should have in order to be a basis for a proper assessment of public sector performance:

"i) a quantity vector of services provided;
ii) one or more service provision quality indicators;
iii) a quantity vector of resources consumed, or an operating budget or both;
iv) a vector of service prices if prices are actually charged; and
v) a vector of resource prices."

Note i) could be further extended to contain all other factors we consider relevant for our analysis – for example the final outcomes that are targeted by the expenditure, like equity of access or infant mortality rate. Ideally – to make the analysis more robust – we would have a regularly updated time series of the data. Dataset should also cover all providers and should be complete in the sense it should not omit any important outputs or resources both across providers and time. Unfortunately there is often a blatant lack of data supply and collection mechanisms that would enable a convenient evaluation of public sector efficiency and its changes - but such evaluation can be of crucial importance for public sector as it is the only way to justify various spending items or evaluate impacts of reforms on efficiency. It creates a possibility to gain more control over public sector performance and ultimately contribute to a more effective use of scarce public resources. However realistic empirical assessment of public sector performance is achievable only with a quality and sufficient data. Only an allocation of additional resources to data monitoring and collection would make this more possible.

Throughout this section I will address various issues related to the data required for an efficiency analysis in the public sector. Specifically I will pay attention to problems of measuring inputs and outputs and to the problem of specification of the scrutinized service.

4.1. Specification of the services

A very important step in any efficiency analysis is to reasonably specify the service under the scrutiny. Such a specification includes an identification of all relevant inputs, outputs, outcomes, objectives and constraints of producers in question. A number of decisions must be made at this point and often there is no other guidance than researcher's own judgement. Many decisions will be derived from the intended scope of measurement or the level of aggregation. Generally more aggregated analysis are easier to be done due to better data availability but their results can be less informative – aggregation can conceal some of inefficiencies that might be otherwise revealed.

4.1.1 Identifying the objectives and measuring the outputs

Simpson (2008) suggests that when we want to measure an output of a public service it should ideally contain all factors that are valued by the society. Not only the quality and quantity of services provided matter here. There are various further objectives that make the public service provision so different from the one of the private sector. Outputs of public expenditure are many and have very different characteristics and some of them are almost impossible to measure (although we know they probably matter in the efficiency analysis). To illustrate the diversity of outputs I will provide several examples. Besides ordinary quantifiable outputs we might want to consider the accessibility or ease of use for the consumer. Society can also value the equity of access of individuals with different characteristics (different education level or health condition etc.) to the services provided by public sector. Further there are outputs consumed by individuals (education) and those provided to a society as a whole and consumed collectively (e.g. defence). For some services, both of these two features will be present - in the case of police, we expect both the prevention of the crime (as a service provided to whole society) and the investigation of individual criminal cases - and for reasonable measure of output, we need to take in account both of these. Sometimes we want to consider the value of some specific parts of the output (as accessibility) independently on the number of consumers of the service - easy accessibility of the hospital is valued also by the people who do not need it in a given moment. Then there might be various outcomes as effects of the expenditure on socio-economic or demographic characteristics. What if certain policy affects only certain group of individuals - does the efficiency increase or decrease? This only shows a difficulty of creating a universal efficiency measure.

Due to its overwhelming complexity most of the applications cease the attempts to include all the outputs and benefits – such analysis can still be very informative but we should keep in mind that if some of the relevant output was not included efficiency scores might be consequently not reflecting the reality. To sort things out a bit individual results produced from the inputs could be generally divided into three categories - activities, outputs and outcomes. For example in education sector, activity would be e.g. a number of teacher-hours, number of graduates would be an example of output, while the future employment and income of graduates would be the outcomes. Each of these categories offers useful information about the overall output and optimal measure of output should take all of these into account.

Sometimes the measures of output are based on the simple counts of activities, mainly due to the fact that these are relatively easy to be quantified. In most cases though, if not accompanied by other measures these are not comprehensive and objective enough to capture the overall output of the service provided. In case of the police, example of such measure would be the number of criminal incidents successfully solved over a time period. While this is definitely easier to measure than crime prevention, objective measure of the output should take in account both. Using only the measurable categories may yield misleading conclusions about the changes in efficiency over time or about the efficiency of different service providers. If a police in one region was exceptionally effective in crime prevention it would consequently need to settle lower number of criminal accidents which would incorrectly result in a lower overall output if the number of criminal cases settled was taken as the measure. Technological changes would pose the similar problem here - development of new treatment procedures might decrease the number of treatments needed to cure the illness and if the count of treatments was the measure it would result in decreasing output. General conclusion of the latter problem is that simple counts of activities are not effective in capturing the quality of the service provided. The solution of these problems is not straightforward. In case of the hospitals, we could rather use the patients as the unit of analysis, rather than count of activities. But this "output-based" approach is much more demanding and costly as it requires us to collect the data on patients (to survey their health condition, number of treatments taken etc.) instead of working simply with the count of treatments (which is easily available). In practice data availability often forces the researchers to stick to more easily quantifiable outputs such as flows of intermediate services or products (number of teacher-hours or number of absolvents in education or number of fires put away by firemen) and use them as proxies of more complex and hardly measurable outputs (overall education level or fire-prevention provided). It is extremely difficult to construct useful output indicators and the possibility of misleading conclusions on efficiency is inherent to our inability to include all relevant aspects of the service in the model.

4.1.2 Setting the relative weights

Measuring the individual elements of outputs is not the only problem we face. Let us consider now that we found a convenient way to measure all the individual outputs. Next issue we come across is finding the appropriate way to weight the different outputs together. Normally if market prices would be available they would serve as a natural set of weights to indicate relative values of individual outputs. In most of the applications these will not be available as outputs of public spending are by definition not traded in the market. Thus even if we had a reasonable quantity indicators, lack of prices as the weights of these indicators would make it impossible to obtain a real value of public sector output to construct even the most simple productivity index.

Finding a solution to this problem is one of the biggest conceptual challenges in efficiency measurement. Ideally the weights should be based on the marginal social benefits that are associated with each respective output. Unfortunately, not only we do not have the information on these values but we can also expect that they differ substantially for different groups of individuals. We can either try to find some useful proxy values or look for a different method to weight the outputs.

Very common way to do this is to use the relative costs related to individual outputs, as in Afonso et al. (2006). This is used in practice quite often because of the data availability. There are several difficulties though. First, this method requires very accurate assessment of costs to each individual output. It might be problematic in the cases where a single input is used to produce more than one type of output (take labour input for instance - the employees need not to specialize in producing a single output, so the labour costs might be shared across large number of individual outputs). Further, using costs as weight means, that relatively cheaper outputs are given smaller weights than the more expensive ones. We can find two implications worth considering - first, changing the structure of the outputs towards relatively cheaper outputs will result in a smaller aggregate output, and second, decrease in the cost of any output will result in a smaller aggregate output. Both of these implications need not match the reality. Let us consider again the improved medical treatment. If the cost of the treatment decreased while all of the other significant characteristics remain the same, we would hardly say that the aggregate output of healthcare system decreased. This system of weights is only as appropriate as the relative costs of outputs match their respective relative marginal social benefits, but easy data availability still remains this method's big advantage.

The problems above refer mainly to construction of aggregate output indices that are often used as output data in empirical applications. Many approaches to efficiency measurement partially circumvent the weight selection problem as they do not require an a-priori choice of weights. Instead they are determined as a by-product of an analysis. While on the first glance this might look like very elegant there are in fact many perils hidden in this - I will discuss this further in section 5.

4.1.3. Measuring inputs

Measuring inputs is less problematic. We need to take in account labour inputs having in hand the information on wages - but here we have problems of attributing the costs to a particular output, as the labour is often shared between more outputs. Then we need to account for capital - we have information on value (value measure) and using the most disaggregated price indices we can transform this into volume measures. Price indices should adjust for quality changes, otherwise changes in quality of inputs might overstate the productivity (increase in outputs can be attributed to increased productivity, instead of increased quality of inputs).

Especially for cross-country comparisons (or comparison across providers) quality of inputs must be taken into consideration. We can incorrectly associate the difference in input/output ratios to different productivity where it is caused only by different quality of inputs.

Sometimes a decision must be made whether to treat a specific factor as an input or as an exogenous factor. For example is the wage-setting mechanism under the control of management or should we consider it as given? This has often a connection to the scope of analysis. Sticking to the wage-setting example when judging on the efficiency of a particular service provider it might be given by higher authorities but when measuring the efficiency of the public sector as a whole it may be one of the crucial explaining variables.

4.2. Dynamic effects

Another problem arises from the dynamic nature of certain factors related to efficiency. A general survey of dynamic efficiency measurement can be found in Silva and Stefanou (2007). From the dynamic point of view a producer’s performance is affected not only by the immediate decisions of the management but also by the inheritances from the past – historical context, past investments in human capital etc.

simply matter. Thus also the production process should be modelled in a dynamic way to take these effects in account.

Viewed from the other way around there might be some outputs that yield no significant measurable immediate benefits because there might be directed towards future attainments. These outputs should in principle be incorporated in the analysis (as ignoring will result in understating efficiency scores) but in reality they are extremely difficult to capture – their future effects can be hardly included in the analysis not to mention the fact that investment itself includes uncertainty and possibly inefficiency.

In the efficiency measurement literature there were some attempts to account for dynamic effects however none of the methods offered so far proved satisfactory. The complexity of the problem in the nonparametric analysis is shown in Fare and Grosskopf (1996) and for parametric approach see e.g. Bond (2002) who demonstrates a possible approach having suitable panel data in hand. Nevertheless the applicability of approaches above seems to be very narrow and so far no generally accepted „ready-made“ solution has been made available.

The question of dynamics is also concerned with weights. If we are working with a time series of data sometimes it is reasonable to believe that the preferences of the society (or the targets set for the public spending) have been changing during the period in question and our model should ideally reflect that.

5. Overview of approaches to efficiency measurement

5.1. Introduction

Core of efficiency measurement lies in comparing the observed performance of a decision making unit (further DMU as producers under scrutiny are often referred to in the efficiency literature) with the optimal (maximal feasible) performance. I already depicted many problems hidden in this concept. In this section I will describe the most widely used approaches to efficiency analysis. There is a crucial question that leads to split of the efficiency literature into two wide approaches - how to define the optimality or potential of the DMU, i.e. how to construct the production possibility frontier? Since the true potential of a DMU can never be known it is usually substituted by the achievements of best performers, who serve as benchmarking units. The main difference between the two approaches to efficiency measurement is in the way in which the production possibility frontier is derived from the observed data.

A detailed survey of efficiency measurement history and development can be found in Battese, Coelli and Rao (1998). These authors claim that the first possible approach to the modelling of production possibility frontier can be tracked at least 80 years into the past, when Cobb and Douglas started with empirical estimating of production functions – it is the econometric modelling of the production frontiers that is based on the least squares methods which fit the functions into the data and estimate the mean performance. Battese et al. however note that standard econometric modelling approach needed to be modified in order to become a suitable tool for efficiency measurement. Speaking of efficiency we are naturally interested in the extreme values that we observe whereas standard econometric approach tends to neglect these unusual observations in favour of more probable but less (in)efficient ones as the focus is set on the average trends. Much effort has been devoted to accommodation of econometric methods to efficiency framework and parametric estimation became one of the two basic methodologies of efficiency measurement known as SFA or Stochastic Frontier Analysis. Name is inspired by notation of Aigner, Lovell and Smith (1977) who proposed a „stochastic frontier model“ that became a basis of contemporary parametric efficiency analysis.

In contrast with the parametric approach which involves an a-priori selection of the functional form of the production frontier (and also of the inefficiency factor) there are the non-parametric methods that rely on enveloping the observed data. Production frontier is constructed as an envelope of all observations (from here the name Data Envelopment Analysis or DEA) hence its shape is determined by outliers. Unlike in the previous case no functional form must be assumed here. It is the efficiency analysis where similar methods based on mathematical programming made a comeback into economics. In most areas of economic research these are commonly considered inferior to econometric approach for their manifest weakness – they do not incorporate the statistical noise which is a base of econometric approach. This weakness has two important implications – first the effects of noise can be easily attributed to efficiency and second this framework at its basic form does not allow for the statistical inference.

Also Fried, Lovell and Schmidt (2008) summarize two obvious differences in the two approaches in the following way: econometric approach is stochastic and can isolate the effects of statistical noise from inefficiency but it conceals potential adverse effects of functional form misspecification of both production frontier and efficiency component whereas nonparametric approach is somewhat immune to model misspecification but it cannot distinguish between (in)efficiency and statistical noise. But they also emphasize the fact that it would be wrong to interpret DEA as a strictly nonstochastic and SFA as rigidly parameterized - much of the research of past decades was aimed at making both methods more robust to both specification errors and statistical noise. The boundaries between two approaches are getting blurred and new methods offer (though still in limited extent) the basis for inference in DEA and for the use of semiparametric or nonparametric techniques in econometric approach.

5.2. Data envelopment analysis

To illustrate the data envelopment analysis I will first introduce the first and most basic model formulated by Charnes, Cooper and Rhodes (1978) („CCR“ model for simplicity). Term „Data envelopment analysis“ or DEA was first used by the same authors in their paper „A Data Envelopment Analysis Approach to Evaluation of the Program Follow Through Experiments in U.S. Public School Education“ (1978). To retain their notation I will also refer to organizations under scrutiny as DMUs (decision making units) to emphasize that (pp.429) „the interest is centred on non-profit-entities (...) and the data is not readily weighted by market prices.“ This nicely shows the intention with which the model was created – to answer the problems of weighting inputs and outputs arising when market prices are either unavailable or distorted which is a typical case in public sector. In later paper authors state that incentives for development of this alternative approach were created by unsatisfactory results generated by using standard econometric models to evaluate the reforms of American educational system in 70s.

5.2.1. CCR (Charnes, Cooper and Rhodes) model

In contrast with econometrics where data are fitted into the model DEA treats data in a purely empirical fashion, (with a bit of exaggeration) it derives the model from the data. Instead of selecting a-priori weights CCR model (and DEA generally) generates the sets of weights as a by-product of an analysis. CCR model computes the efficiency θ_k of a DMU_k by the following mathematical program:

$$\max_{\mathbf{u},\mathbf{v}} \theta_k = \frac{\sum_{r=1}^{s} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}}$$

$$s.t. \quad \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \ \text{ for } j = 1, \dots, k, \dots, n$$

$$u_r, v_i \geq 0 \text{ for } r = 1, \dots, s;\ i = 1, \dots, m$$

where n is the number of DMUs compared, $\mathbf{x_j} \in R^m$ is a vector of inputs that DMU_j uses to produce the vector of outputs $\mathbf{y_j} \in R^s$ and $\mathbf{u} \in R^s, \mathbf{v} \in R^m$ are the vectors of weights that maximize the efficiency score of DMU_k relative to all other DMUs. This way we can compute the efficiency score for all DMUs. Since all input and output variables are also assumed to be nonnegative (and „useful") we get that $0 \leq \theta_k \leq 1$. From the equations above it is clear that each DMU will be generally assigned a different set of weights with which it will be compared to other DMUs (so for n DMUs we will obtain n efficiency scores and n sets of weights). The weights are chosen so that they maximize the efficiency of given DMU thus each DMU is shown in the best possible light. DMU for which θ =1 is called efficient. This efficiency is identical to that of Koopmans (1951), based on Pareto-optimality, i.e. a DMU is efficient if and only if it cannot make an improvement in any input or output without worsening another input or output. Also note that this formulation of the model is output-oriented in the sense it expresses the inefficiency as a shortfall in output (or as a possible increase in output given the inputs so that the DMU in question would be projected on an efficient frontier). It could be easily reformulated to an input-oriented measure that looks for the excesses in inputs (in this case we try to **minimize** the efficiency score which is defined here as a possible reduction in all inputs that is feasible given the output, i.e. that would project the DMU on a production frontier).

If $\theta_k < 1$ there must be at least one other DMU for which the first constraint turns to equality (otherwise $\theta_k$could be increased). Such DMUs are efficient (under the choice of weights optimal for DMU_k) and are refered to as reference set or peer group to DMU_k .

Figure 4.1 below demonstrates the situation in the most simple case with a single input X and a single output Y (weight issues are irrelevant here):

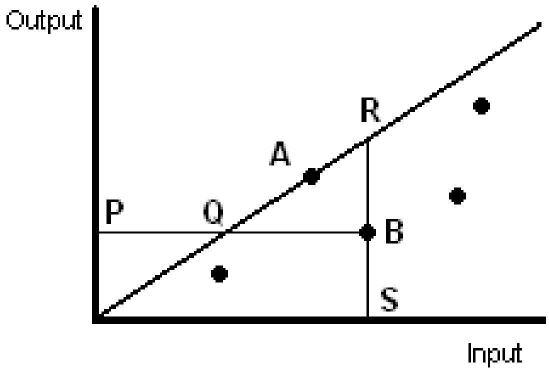

Figure 4.1. CCR model for single input and single output.

Out of five DMUs on the picture (each DMU is represented by a black dot indicating its input and output volumes) only DMU A is identified as efficient ($\theta_1 = 1$) - line passing through A and an origin is a production frontier to which all other DMUs are compared (it also defines a production possibility set as the production frontier and area below it). Note that CCR model implicitly assumes constant return to scale, i.e. if $A = (X_A; Y_A)$ is an element of the production possibility set then for any $c \geq 0$, $A' = (cX_A; cY_A)$ also lies within a production possibility set. Efficiency of DMU B then would be measured by input oriented CCR model as a ratio of two distances $\theta_B = \frac{|PQ|}{|PB|} < 1$ and if e.g. $\theta_B = 0{,}5$ it means that hypothetical DMU Q on an efficient frontier would need only 50% of B's inputs to produce the same output. Alternatively output oriented model would look for a shortage of B in output in comparison with hypothetical DMU R using the same inputs and B's efficiency would be evaluated as $\theta_B = \frac{|BS|}{|RS|} < 1$.

Advantages of this approach are clear. Weights are generated automatically (although this can be also viewed as a disadvantage, as I will argue later) and also the reference group of more efficient producers is determined in the process for each DMU. The fact that weights are chosen to maximize the efficiency of DMU in question has another important consequence - $\theta_k < 1$ actually means that DMU_k is strictly outperformed by some (or all) of the members of the reference group *no matter what the weights are* (in contrast to econometric approach which identifies the outperformed DMUs only with a given set of weights). For any weights we choose, there is always at least one DMU that is more efficient than DMU_k. This is important as it helps to safely identify the best performers and the worst performers and such information might help to unveil the factors affecting efficiency. One of the most important advantages of this approach is that it can handle large numbers of both inputs and outputs (opposed to stochastic models that are usually designed to account for many inputs but a single output or vice versa). By using DEA we also avoid some problems with model specification – when we have chosen an appropriate model our only problem is to choose which variables to include (unlike in parametric approach where at least the functional form of the frontier and distribution of error and efficiency term must be specified). This measure of efficiency is also unit invariant (efficiency score does not change with the change of units in which some or all of inputs and outputs are measured). See e.g. Cooper, Seiford and Tone (2007) for more detailed theoretical underpinnings.

Viewed from the other side it is not hard to identify some shortcomings too. First due to a different set of weights assigned to each evaluated entity this approach is totally unfit to construct a rankings table of DMU's efficiency scores. Different weights might be interpreted as different objectives and it does not make a lot of sense to rank DMUs by their efficiency in pursuing different objectives. Moreover the design of the model tends to overestimate the efficiency scores and with no other restrictions on weights than non-negativity we can often see that the efficiency of the DMU is maximized by assigning maximum weight to an output in which the evaluated DMU is relatively efficient and zero weight to all other outputs (this result is encountered very often in empirical applications). Nevertheless it is this property that makes DEA so useful for identification of some strongly underperforming entities – assuming that all relevant inputs and outputs are included if DEA efficiency score that judges the DMUs most leniently is low it indicates a serious slack in that DMU's overall performance (while low efficiency assigned by some other methods might still be a product of wrong identification of objectives or other specification problems). Finally the nonstochastic nature of the model does not allow to account for statistical noise and does not provide ground for statistical inference.

5.2.2. DEA development, other DEA models

Naturally there are lots of extensions and different approaches to the basic model described in the previous paragraphs. First of all assumption of constant returns to scale in CCR model seems to be too restrictive. Fortunately it is only one of many ways various DEA models construct the production possibility frontier. DEA was extended in various applications to account for all possible returns to scale assumptions. For example model of Banker, Charnes and Cooper (1984) (further BCC model) accounts for variable returns to scale by defining the production frontier as a convex hull of all DMUs (variable returns to scale in the sense that there is a part of production frontier with increasing, decreasing and possibly constant returns to scale) that envelopes the data „more tightly“, as shown in figure 4.2. The difference in production frontiers between CCR and BCC model naturally results in different efficiency scores – it generally holds that $\theta_{CCR} \leq \theta_{BCC}$ (as ilustrated in figure 4.2. where $\frac{|PQ|}{|PB|} < \frac{|PQ'|}{|PB|}$. DMU can be efficient in BCC model but at the same time inefficient in CCR model. I am mentioning this because the ratio $\frac{\theta_{CCR}}{\theta_{BCC}} = \theta_{SC} \leq 1$ can be used as a simple measure of scale efficiency (denoted as θ_{SC}), if needed, as I described it in section 3. Decomposition of efficiency on „pure technical efficiency“ (θ_{CCR}) and „scale efficiency“ is commonly used in applications.

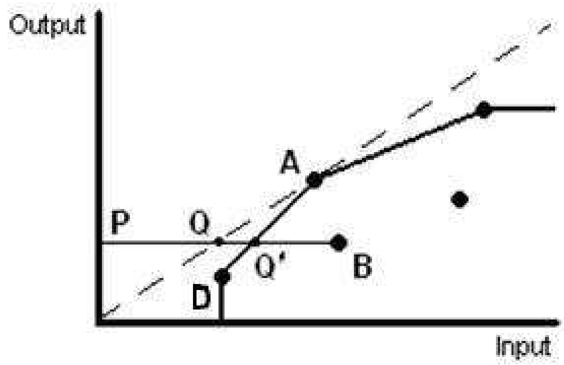

Figure 4.2. Variable returns to scale in DEA.

In figure 4.2. DMU A for which $\frac{\theta_{CCR}}{\theta_{BCC}} = 1$ operates on the most productive scale size (i.e. is scale-efficient). On the other hand DMU D is purely technically efficient but not scale efficient, which means that whole its inefficiency is caused by suboptimal scale of operation (rather than by wasting resources).

Another way of modelling the production frontier that enjoys a wide use in applications can be found in so called „free disposal hull model“ first formulated by Deprins, Simar and Tulkens (1984). This model represents the most general form of DEA since it is developed from only the basic assumptions on disposability of inputs and outputs and makes no presumptions about the returns to scale or shape of production frontier other than what is directly observed. Hypothetical linear or quasiconcave production frontier is substituted by a step function as shown in Figure 4.3. This frontier encloses the set of all production possibilities that can be generated directly from the data and the single assumption that it is possible to produce same output using more inputs (which is consistent with our everyday experience) and it is this assumption modesty that makes FDH model attractive. Some empirical studies suggest that quasiconcavity assumption implicit to other DEA models might often be violated (see Cherchye et al. (2000) for a discussion).

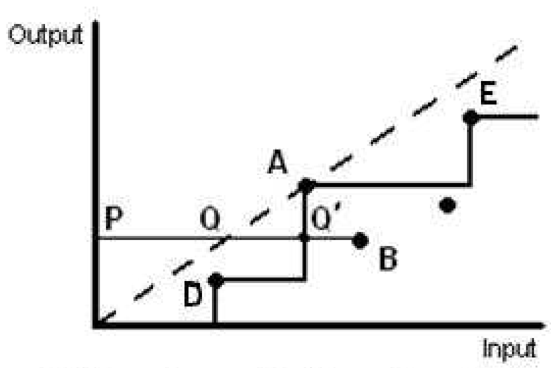

Figure 4.3. Free disposal hull model production frontier.

Under the FDH model all of DMUs D, A and E have been found efficient. Since we have no assumption on returns to scale the only inefficient DMUs are those that are "strictly outperformed" by another DMU in the data set, e.g. B is inefficient because it was strictly outperformed by A that needed less input to produce larger output.

5.2.3. Further issues

One of the unfortunate characteristics of early DEA models was a problematic inclusion of exogenous factors (in section 2 I explained how important it is to account for factors that potentially affect the efficiency but are not under control of the management). After decades of development there are several approaches available to offset this shortcoming but none of them seems to be applicable generally. One approach follows the work of Banker and Morey (1986) who tried to modify the CCR model to include these variables and treated them as „fixed inputs“. Other applications would try to compensate for the exogeneities by manually adjusting the data (thought this might raise the question of plausibility). However more common approach in recent applications is to use a two-stage approach when in the first stage DEA efficiency scores are computed ignoring the exogenous factors and then in second stage the scores are adjusted by econometric regression on the factors of interest.

Some modification also exist to solve the „optimal weights“ problem depicted above. With no restriction on weights and the freedom of each DMU to choose its own objectives (i.e. to set the best vector of weights that shows the DMU in best possible light) optimal vector of weights shows many zeros for most DMUs. But in many cases assigning a zero weight to certain output or input included in the analysis does not make economic sense – for some reason we might think that particular variable is important for evaluating all DMUs no matter how it fits in their scheme. That leads to conclusion that the model might be too benevolent and that the efficiency scores are overestimated. This observation was well reflected in the development of DEA and several modified models called „restricted multipliers models“ were developed to make the weights less flexible allowing to fix certain weights absolutely or relatively to other weights or to let them fluctuate only within limited bounds (e.g. by imposing additional constraint $a \leq \frac{v_1}{v_2} \leq b$ on weights v_1, v_2).

In the beginning of the section I mentioned that DEA models are immune to model misspecification in the sense that functional form of the efficient frontier and weights need not to be specified and also no assumptions on inefficiency factor are needed (which is not the case in parametric approach, as I will show in next section). However from what I have written above it is clear that there is a danger of model misspecification in DEA. First, different models bring different assumptions about returns to scale. Second and more important issue is a choice of inputs and outputs to be included in the analysis. In econometrics model specification is a subject of great interest and there is a substantial battery of tools that has been developed to test whether our model is correctly specified. But in DEA, the choice is completely up to researcher and the basic problem is that standard DEA models do not offer any device to review the suitability of the chosen model other than researcher's own judgement. The problem of robustness in DEA is a matter of ongoing research, a good introduction into the issue is offered in Smith (1997) who analyses (yet only empirically) a sensitivity of various DEA models in relation to sample size, variations in number of inputs, correlations between inputs and other factors. Despite the general knowledge that DEA results might be very sensitive to choices above this problem remains unsolved in DEA theory and is subject to researcher's opinion.

For a complete discussion on both theory and development of DEA with a broad body of references see Cooper, Seiford and Tone (2007) or Fried, Lovell and Schmidt (2008) (the latter treats both nonparametric and parametric approaches).

5.3. Parametric approach

Second approach to efficiency analysis closely relates to traditional econometric methods. The original idea that started this kind of efficiency analysis was the suggestion of Farrell (1957) that technical efficiency of a producer (as defined by Koopmans (1951), saying that a producer is fully efficient if and only if it cannot improve its output without increasing some input or vice versa) could be measured as a realized deviation from a hypothetical production frontier isoquant. Fried, Lovell and Schmidt (2008) give notice that even at that time econometric estimation of production functions (or cost or profit functions) was a standard statistical exercise. Nevertheless the aim was always to estimate an average technology which is not of a particular interest in efficiency analysis. Farrell's idea originated a way of research that shifted attention from average to best practice and that tries to estimate the idealized potential production frontier and measures efficiency of producers in terms of deviations from this frontier. It logically follows that we are interested not only in an estimated function but even more so in the residuals of specific observations – in fact residuals are often the only phenomena of direct interest which is in sharp contrast with traditional econometrics where the concern in residuals is usually limited to verifying the model assumptions (Fried, Lovell and Schmidt (2008)).

In the context of public sector efficiency production functions and cost functions are the functions of interest (again depending on whether we want to measure efficiency as a shortfall in output or a waste of inputs). Contemporary approaches to this topic are based on the seminal papers of Aigner, Lovell and Smith (1977) and Meeusen and van den Broeck (1977) that both came with almost identical formulation of so called „stochastic frontier model“ – for this reason parametric approaches to efficiency analysis are usually referred in the literature as stochastic frontier analysis (SFA).

Prior to these papers parametric analysis was mostly represented by COLS models – corrected ordinary least squares – which estimated the production function by OLS method and then „corrected“ it (shifted up) to pass through the observation with the highest positive residual. Because *an entire residual was interpreted as an inefficiency factor* such observation was considered fully efficient and inefficiency was then calculated for all other observation by their realized deviation from the COLS frontier. Figure 4.4. shows a logarithmic functional form of the frontier estimated by COLS where observation A is found to be efficient and inefficiency of other DMUs is measured as a full value of their corrected residual.

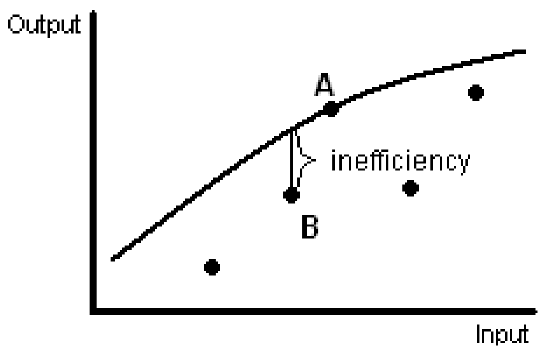

Figure 4.4. Inefficiency in the COLS approach.

Kumbhakar and Lovell (2000) note that the COLS approach was almost completely discarded for its apparent incapability of including other factors determining the residual than inefficiency – which is an issue this method shares with DEA while lacking some attractive feature that DEA has. It was substituted by more flexible stochastic frontier models that are based on the idea of decomposition of the error term where it is explained partially by inefficiency and partially by statistical noise. Fried, Lovell and Schmidt (2008) say that SFA models have undergone a great deal of development in past decades and they now offer a great flexibility that allows tight interconnection of modelling reality with economic theory. They allow for comparisons of efficiency of various producers and also for assessing changes in efficiency over time by analysing time series data[1].

5.3.1. Basic stochastic frontier model

Stochastic frontier model (or sometimes „composed error model“) estimates a production function in a general form

$$y_i = f(x_i, \beta) + \varepsilon_i; \, i = 1, \dots, n$$

where y_i is an (single) output of DMU_i, x_i and β are vectors of inputs of i-th DMU and vector of estimated parameters respectively. What is a new is a definition of error term

$$\varepsilon_i = u_i - \theta_i$$

where $u_i \sim N(0, \delta_u^2)$ is a standard two-sided (meaning unrestricted in sign) normally distributed error term (e.g. variations in productivity caused by random unexpected expenditures) and $\theta_i \geq 0$ is a non-negative term symbolizing the technical inefficiency of i-th observation. So the residual (deviation of a DMU from an estimated stochastic production frontier) is divided into two components – statistical noise and inefficiency.

[1] Treatment of time series data is technically possible also in DEA, but it is very controversial because of underlying variable weight mechanism – does it make economic sense to allow output and input weights of a single producer to change over time, often radically? See Cooper, Seiford and Tone (2007) for a discussion.

Assuming that $u_i = 0$ inefficiency is given as a difference between the observed value of output and a maximal attainable output for given inputs (which is analogical to DEA approach).

Any standard forms of production function can be used in the model but typically a log-linear (Cobb-Douglass) function is employed so the model looks like this (for two inputs):

$$y_i = \beta_0 x_{i1}^{\beta_1} x_{i2}^{\beta_2} e^{u_i - \theta_i}.$$

Having at hand the residuals $\hat{\varepsilon}$ we are still facing the problem how to separate the two terms they are composed of. It might tell us something about average inefficiency (might be estimated e.g. by an average value of $\hat{\varepsilon}$) but in most cases our prime concern is in estimates of inefficiency of *all individual observations* so that we can evaluate and compare individual DMUs we are analysing. Thus another assumption is needed on a conditional distribution of θ_i given ε_i, which must be explicitely specified and parameters estimated. Error term u_i is usually assumed to be normally distributed but in case of $\theta_i$there are more alternatives. Regarding the non-negativity of θ_i researchers usually choose between half-normal, truncated normal, exponential and gamma distributions (Smith and Street (2005)) that are all one-sided. According to Schmidt (1985) there is no theoretical guidance on which choice of these distributions is optimal.

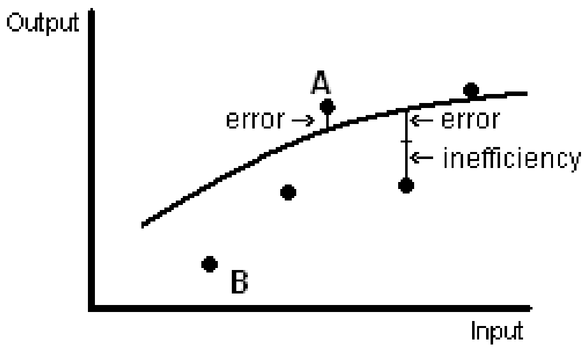

Figure 4.5. Inefficiency in stochastic frontier approach.

Besides the case above where we are estimating an efficient production frontier we can alternatively formulate the model to estimate the minimum cost function as function of outputs (so this is a single-input multiple-output form of FSA). Employing the most commonly used logarithmic form the model would look like

$$\log x_i = \log\{f(y_i, \beta)\} + u_i + \theta_i$$

where x_i is a single output (=cost), y_i is a vector of outputs of i-th DMU and β is a paremeter vector to be estimated along with inefficiency term θ_i. Note that here inefficiency would translate to the increase in costs thus $\theta_i \geq 0$.

5.3.2. Further issues

There are several works trying to cover the whole theory of parametric approaches to efficiency measurement, amongst them a booklenght treatise of Kumbhakar and Lovell (2000) which provides various methods survey and a huge amount of references to other literature on theory, history of thought and applications.

As I already suggested one drawback of stochastic frontier analysis is that it cannot account for multiple inputs and multiple outputs simultaneously like DEA. Obviously this is a problem mainly if we are estimating a production function. If we are interested in cost function and have information on total cost further decomposition is usually not necessary (though it might hold some useful information like identification of the most inefficiently used inputs etc.). Only way to account for more outputs involves aggregating them in one composite output indicator. It brings many problems (in public sector environment it is especially finding an appropriate way to weight outputs together, which is an issue I already discussed) but often it is an only way to obtain some useful measure of (in)efficiency.

On the plus side, Kumbhakar and Lovell emphasize that SFA excels in flexibility when it comes to accounting for exogenous factors (as described in section 2.3). As in standard econometrics, we can control for any factor that we believe to affect the dependant variable by simply including it in the model. We can include both continuous variables (e.g. density of population) or discrete ones (e.g. the dummy variable for rural or urban area) and „level the playing field" for all DMUs under scrutiny right in the model.

Another merit (and maybe the most significant one) of this approach is that it is based on a century-old research of econometrics that has developed a very deep theory and immense number of instruments to solve any potential problem that can be encountered (unlike about 30 years old DEA that also receives only a marginal interest in general research). Many aspects like testing of assumptions, statistical inference, model specification issues etc. have been already solved in general econometrics and now are readily used in SFA.

5.4 Discussion of the two approaches

I tried to cover the most considerable issues in previous text when explaining individual methods but several points are worth developing a bit deeper. First there is a question of automatic generation of relative weights as a by-product of an analysis. The system of weights assessment in DEA has both its advocates and opponents and remains to be one of the most controversial factors of the approach. Some see merit in this arguing that it is advantageous that exact relation between inputs and outputs need not to be explicitly specified and that it can vary between DMUs reflecting the local preferences and circumstances (Cooper et al., 2007). Yet opponents doubt that there can be some rightful self-defining measures determinable without any reference to context. One needs to ask what are the societal preferences or objectives of a given service proclaimed by government (or whatever we consider to determine the relative values of various inputs and outputs, even if it would be only a common judgement) and how they are reflected in our model. Some researchers even see the weight variation as a major shortcoming that is inevitable in DEA – see for example Stone (2002) who concludes that „the idea that some sort of exogenous valuation is necessary to resolve matters surfaces embarrassingly throughout the DEA literature, e.g. in the restricted weights idea...“

The same argument applies to some extent to SFA technique where in a cost-estimation model in linear form the relative weights (parameter estimates $\hat{\beta}$) are determined for each output as a statistical estimate of the sample average cost of an additional unit of that output, again based solely on the data without reference to any exogenous judgements. Some researchers, see e.g. Afonso et al.(2005), try to circumvent this issue by first aggregating the data into a single composite output and composite input indicators and then applying the simplest one-input one-output form of one of the models (in this case free disposal hull model is used). Though this can be criticized just as well because there is no guarantee that the weights are chosen reasonably in their composite indicators, so the suspicion arises that this way the problem does not disappear, it just takes on a different form.

I have written much about the merits and demerits of both approaches throughout the previous section. One more natural question remains to be answered – which approach is favourable for the measurement of public sector efficiency? The answer depends on an application specific factors and the relation of the approaches can be studied from several viewpoints.

First line of thought might be represented by Gong and Sickels (1992) or Cubbin and Tzanidakis (1998) who treat data envelopment analysis and parametric approaches as competing methods, or even mutually exclusive (although these and other authors report rather ambiguous findings on methods' applicability). However other authors have tried to use both techniques in a complementary fashion to cross-check each other as in Ferrier and Lovell (1990) or Spottiswoode (2000) – authors claim that they want to avoid the methodological bias this way, which might be particularly important when addressing important policy problems. Finally several applications have also attempted to combine both approaches in a conjoint analysis creating various „hybrid" models – see for example Arnold et al. (1994). We can also quite commonly come across various two-stage models that estimate DEA efficiency score in stage 1 and subsequently use this information in some regression procedure in stage 2. A common reason for this treatment is the prevalence of various exogenous factors that strongly affect the performance of different DMUs but are not controlled for by DEA.

6. Using the findings

6.1. Variations in efficiency

Two crucial questions remain to be answered when we have successfully performed the efficiency scores assessment - what is the explanation of the performance variance and what is to be done with the findings?

Regarding the first issue, we should always be aware of the fact that there are two potential sources of variation in obtained efficiency scores - differences in managerial performance (inefficiency) or differences in the operating environment in which the services are provided (exogenous factors). Assessment of efficiency cannot be credible unless and until the variation in the operating environment is either ruled out or controlled for. Only when we adjust our measurement this way we can compare like with like. For example it would be clearly misleading to compare the performance of two schools - let us say we would use the graduates' education level as an output - with vastly different socioeconomic background of student without taking this into account. But this of course requires the influential characteristics of operating environment to be measurable and/or observable. I mentioned earlier that econometric methods can deal with this problem simply by including the relevant variables in the analysis – even if characteristic is not measurable but is observable and can be characterized by a polytomous variable (e.g. rural or suburban or urban) then it can be incorporated in the performance evaluation by adding dummy variables for each kind of environment in the econometric analysis. This causes more severe problems in data envelopment analysis where it forces either radical changes of the method or a second-stage regression of obtained efficiency scores on variables of interest. But some influential environment factors may end up omitted for various reasons – the information on these might be unavailable or they need not to be observable at all, or the researcher might misspecify the model by not realizing that certain factors affect the performance in a given sector. Also the scope of analysis plays an important role in this matter, generally the more aggregated sample we have the more troubles we have with controlling the variance in environment. In any case it would be very ambitious to assume that all of the exogenous factors were effectively controlled for in the analysis and that they do not further distort the efficiency scores and it should be realized before drawing any final conclusions.

6.2. Remarks on cross-country comparisons

Many applications are concerned with cross-country comparisons of efficiency, e.g. comparing efficiency of education or health system or a global efficiency of a public sector. This scope of an analysis however raises several thorny issues to be addressed. First, as I already argued, efficiency measurement techniques are very sensitive to the data used, and in this case it might be quite complicated to obtain a confidential data-set. For different countries different kind and range of data is obtainable – data on certain inputs or outputs can be missing for some countries or can be provided at a different level of aggregation. For this reason many researchers, e.g. Afonso et al. (2006) or Guellec and Potterie (2001), use the datasets constructed by large international organizations - such as World Bank's „Governance Indicators" found in Kaufman, Kraay and Mastruzzi (2008) – to at least reduce the differences in data processing methodology. But of course that possibilities of such datasets are limited in terms of covering all the possible variables researchers might need and also they might still not answer the important problem mentioned above - in short different traditions and organization of the public sector along with varying standards in data collection and measurement are most likely to result in heterogeneities in the data, raising concerns about their comparability. Further having at hand the information on amounts of outputs or inputs in each country a question of quality should be considered. Sticking to the education example, if number of teacher-hours is taken as an output, factors like qualification of the teacher or average number of students in the class are more than likely to affect a „value" of the service and our analysis should reflect that (it seems reasonable to believe that such variations are larger across countries than across schools within one country or region). Quality adjustments of inputs and outputs seem to be necessary for constructing a reasonable efficiency measures but the lack of appropriate data often leave the researchers helpless in this matter. Also the theory does not offer any generally accepted procedure to involve quality adjustments in the efficiency analysis and the researchers are deemed to look for application-specific solutions. In practice, differences in quality are often assumed away, despite the fact that it is in contradiction with even the most casual observation.

Another problem arises from possible differences in priorities across countries. I argued earlier that efficiency and effectiveness are related to certain objectives. Finding a measure of efficiency to make comparisons between countries when these objectives and priorities substantially differ in those countries proves to be very problematic and it might lead to some very misleading results. In this situation using the variable relative weights of individual inputs and outputs as derived from data envelopment analysis can be a justifiable way of accounting for the differences in priorities, but again, this approach almost completely excludes any attempts to construct efficiency-ranking tables or similar comparisons.

All problems mentioned in this section might have negative effects on the correctness of final efficiency scores. Efficiency would be underestimated for certain country if e.g. its higher quality of outputs would not be taken in account or if its expenditure follows some alternative objectives that are ignored in the model.

6.3. Remarks on policy responses

A lot of empirical works document that the choice of the method might have crucial influence on the resulting efficiency values. Even in the framework of one approach results vary substantially depending on the model specification as e.g. Smith (1997) or Jacobs (2001) show. The latter used many different DEA and SFA models on the same dataset concerning efficiency of 232 National Health Service hospitals in the UK and summarized the correlation coefficients of efficiency scores produced by different model specifications in the following Pearson correlation coefficient matrix:

Table 1: **Correlation of results of various DEA and SFA specifications.**

	DEA-1	DEA-2	DEA-3	DEA-4	DEA-5	SFA-1	SFA-2	SFA-3	SFA-4	SFA-5
DEA-1	1.0000									
DEA-2	0.2298	1.0000								
DEA-3	0.3729	0.6340	1.0000							
DEA-4	0.7575	0.3513	0.5372	1.0000						
DEA-5	0.4722	0.6062	0.8352	0.6149	1.0000					
SFA-1	**0.4274**	0.4667	0.5946	0.5166	0.5756	1.0000				
SFA-2	0.0957	**0.6209**	0.4231	0.1831	0.4038	0.6354	1.0000			
SFA-3	0.2154	0.4318	**0.5975**	0.3165	0.4852	0.8297	0.6917	1.0000		
SFA-4	0.4192	0.4835	0.6583	**0.5543**	0.5998	0.8763	0.6815	0.8065	1.0000	
SFA-5	0.3399	0.5195	0.6557	0.4633	**0.6343**	0.9496	0.6535	0.8731	0.8217	1.0000

Pearson correlation coefficient matrix of efficiency scores obtained from 5 DEA model specifications and their SFA counterparts (correlations of results based on the identical data are indicated in **bold**).
Source: Jacobs (2001).

Any of the 10 models could be justified but their mapping of efficiency across organizations and their implications on policy substantially differ. Several technical decisions must be made while there is often no theoretical guidance or tool to learn whether our decision is correct – any particular specification can be readily challenged and our chances to arrive at any definitive and correct conclusions on efficiency using the technical apparatus alone are indeed very low. Though I believe that both techniques can provide a very useful insight to any dataset, the results should be treated with extreme care regardless of the model we choose, being conscious about its limitations. This is also a reason why most of the authors would advocate some form of synergic use of several approaches in any application (although there is not a definitive consensus on this as I mentioned above), given the fact that the results of our analysis usually only represent the trends rather than absolute conclusions and serve mostly as a basis for further investigation rather than for direct policy responses.

This is to say that measuring efficiency in the public sector involves a great deal of responsibility and the word of warning on using the results is in place. One can easily imagine possible consequences of incorrect assessment of efficiency when used in a public sector policy design. In the first place incorrect identification of efficient and inefficient decision making units might lead to promotion of sub-optimal practice marking certain inefficient DMUs as examples of perfection or alternatively forcing efficient DMUs to adopt unnecessary changes. Subsequent adoption of contraproductive policy changes is an implicit danger of any efficiency evaluation in a public sector, not to mention the fact that technical apparatus can be abused to disguise the political motives (table 1 shows how drastically can the method selection affect the results). Borge et al. (2008) warn that this inconvenience is magnified by the fact that bad policies are sometimes hard to remedy owing to the limitations of legislative process, namely time and uncertainty of approval in the political debate.

Many studies are concerned with the use of efficiency indicators as a controlling mechanism enhancing the accountability in the public sector. According to Smith (1990) the uses to which efficiency or performance indicators are put from the managerial point of view can be divided into two broad categories: identification of problems and provision of supporting evidence for policy decisions. First making the information on efficiency publicly available can increase pressure on the ineffective providers to increase efficiency though as I argued in section 3 this argument collides with rather imperfect sanctioning mechanism and limited competition – in light of these factors the motivation of managements to increase efficiency is questionable at least. But we could optimistically presume that certain managements would try to take advantage of the information to seek for possible cost reductions or output augmentations. From the point of view of authorities identification of underperforming providers can indeed be a very useful information and a good reason for further investigation and suitable policy responses.

But there are hidden perils in using efficiency indicators for evaluation of performance of various providers. Smith (1990) provides an example, saying that if the method of measurement and specific relevant variables are well-established and known to DMUs under scrutiny (which is very likely in the case that it is used regularly to asses performance in a given sector) it might give rise to many kinds of distorted economic behaviour and possibly lead to adverse effects on the consumer’s utility. Above all we can easily imagine that managements would try to reallocate the resources to the activities covered by the indicator on the expense of those that are not included - maybe simply for the reason of problematic measurement, which might be the case of service quality or other important aspects. Smith and Street (2005) state that given the static nature of common efficiency-measurement methods we could also expect an underinvestment in activities that generate immediate costs but only accrue benefits in the future, such as investments in health education of citizens in healthcare sector, or investments in human capital generally. Such precautions of the management are not in any sense motivated by the pursuit of consumer's well-being but solely by optimizing the score on the particular indicator and the final result of our endeavour might be a socially inefficient situation.

7. Conclusion

Without an elementary cooperation of public sector institutions and researchers hardly any assessment of efficiency could be possible. Availability of reliable data, which presumes their subsequent measurement and collection in the public sector, is still the first and inevitable prerequisite of any analysis. Some of the countries have already recognized this and they are trying to devote more resources to improve the data collection mechanisms. Moreover in some countries the specialized agencies were created in the public sector whose function is not only to supervise the data collection process but also to monitor the efficiency of particular public expenditure. The importance of knowing how much additional output can be produced without absorbing further resources (or alternatively how much input can be saved without a decline in output) was pointed out by Koopmans (1951) and Farrell (1957) and it stands in the core of efficiency analysis. Detecting the sources of possible waste in the public sector can save a considerable amount of resources that can be diverted to satisfy other objectives elsewhere in the economy. Moreover showing that public spending is efficient can reinforce the credibility of the public sector as not doing so naturally leads many to a suspicion that part of the resources absorbed by distorting taxes is uselessly wasted. That is disturbing especially now in the times of economic recession when the pressures on the public budgets are particularly strong – improving the efficiency seems like an inexpensive and popular (from the point of view of voters) way of alleviating the public deficits.

I introduced the most commonly used techniques we can encounter in the efficiency measurement literature based on data envelopment analysis and econometric methods. I outlined their relative merits and possible difficulties of application on public sector data. I made a few notes on technical shortcomings of the methods but in my view the main problems of efficiency measurement do not lie in the technical apparatus itself but rather in various complications associated with specification and measurement of some aspects of public sector services, such as outputs, quality, objectives etc. If an acceptable solution to these issues cannot be found even the most accurate and theoretically flawless data processing technique will not generate satisfactory results.

Throughout the text I have shown that the complications caused by the nature of public sector services lead to several pitfalls and if the findings of an exercise are not used with adequate caution it might generate somewhat contraproductive results. Section 4 describes problems associated with obtaining a suitable data-set which is clearly not a trivial matter. I have also illustrated how the choice of method can influence the findings and I also argued that some decisions that need to be made are subject to a political choice, e.g. the definition of objectives of particular expenditure. I also noted on a danger of confusing inefficiency for specific environment characteristics that affect the performance of individual producers if we fail to account for them properly in the analysis.

Despite all of these issues I tried to make a point that with a development the efficiency analysis has undergone in past decades it may be a very useful tool for monitoring the performance of public sector service providers and it can greatly help with improving the „value-for-money" that public sector generates if we will not ignore the limitations inherent to this kind of analysis, in this kind of environment.

Bibliography:

- Aigner, D., Lovell, C. A. K. and P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6. rqpi.org.
- Afonso, A., Schuknecht, L. and Tanzi, V. (2005). Public sector efficiency: An international comparison. Public Choice, Springer.
- Afonso, A., Schuknecht, L. and Tanzi, V. (2006). Public Sector Efficiency, Evidence from New EU Member States and Emerging Markets. European Central Bank Working Paper Series (no. 581).
- Alchian, A. A. (1965). Some economics of property rights. Il Politico, 30: pp.816-829. Reprinted in A. A. Alchian (Ed.) (1977), Economic Forces at Work. Indianapolis: Liberty Press.
- Alchian, A.A. and Kessel, R.A. (1962). Competition, monopoly and the pursuit of money. In Aspects of labor economics (1962). Princeton: Princeton University Press.
- Arnold, V., Bardhan, I.R., Cooper, W.W. and Kumbhakar, S.C. (1994). New Uses of DEA and Statistical Regressions for Efficiency Evaluation and Estimation — With an Illustrative Application to Public Secondary Schools in Texas. Annals of Operations Research, 66, pp.255-278.
- Banker, R.D., Charnes, A. And Cooper, W.W. (1984). Some Models for estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science 30, pp.1078-1092.
- Banker, R.D. and Morey, R.C. (1986). Efficiency analysis for exogenously fixed inputs and outputs. Operations Research, 1986. jstor.org.
- Battese, T., Coelli, D.S.P. and Rao, G.E. (1998). An Introduction to Efficiency and Productivity Analysis. Boston: Kluwer Academic Publishers.
- Bond, S. (2002) Dynamic panel data models: a guide to micro data methods and practice. Port. Econ. J., 1, 141-162.
- Borge, L.E., Falch, T. and Tovmo, P. (2008). Public Sector Efficiency, The Roles of Political and Budgetary Institutions, Fiscal Capacity and Democratic Participation. Springer Science+Business Media, LLC 2008.
- Bozeman, H.G. and Rainey, B. (2000). Comparing Public and Private Organizations: Empirical Research and the Power of the A Priori. Journal of Public Administration Research and Theory, Vol. 10, No. 2: 447-470.
- Charnes, A., Cooper, W.W. and Rhodes, E. (1978a). Measuring the efficiency of decision making units. European Journal of Operational Research 2, pp.429-444.
- Charnes, A., Cooper, W.W. and Rhodes, E. (1978b). A Data Envelopment Analysis Approach to Evaluation of the Program Follow Through Experiments in U.S. Public School Education. Management Science Research Report No. 432. Pittsburgh: Carnegie-Mellon University.

- Cherchye, L., Kuosmanen, T. and Post, T. (2000). What is the Economic Meaning of FDH? A Reply to Thrall. Journal of Productivity Analysis 13, 263-267.
- Cooper, W.W., Seiford, L.M. and Tone, K. (2007). Data envelopment analysis: a comprehensive text with models, applications, references and DEA solver software. Springer Science+Business Media, LLC.
- Cubbin, J. and Tzanidakis, G. (1998) Regression versus data envelopment analysis for efficiency measurement: an application to the England and Wales regulated water industry. Utilities Policy, 7, 75-85.
- Debreu G. (1951). The Coefficient of Resource Utilization. Econometrica, 19, pp.273-292.
- Deprins, D., Simar, L. and Tulkens, H. (1984). Measuring Labor-Efficiency in Post Offices. In The Performance of Public Enterprises: Concepts and Measurement, chapter 10, 243–268. Amsterdam: North-Holland.
- Dierx, A., Ilzkovitz, F. and Mandl, U. (2008). The effectiveness and efficiency of public spending. European Communities, http://ec.europa.eu/economy_finance/publications
- Gong, B.-H. and Sickels, R.C. (1992). A Comparison Between Stochastic Frontier and Data Envelopment Methods Using Panel Data. Journal of Econometrics, 51, 259-84.
- Fare, R. and Grosskopf, S. (1996). Intertemporal Production Frontiers with Dynamic DEA. Dordrecht: Kluwer.
- Farrell, M.J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A (General).
- Farsi M., Filippini M. and Greene W.H. (2006). Application of Panel Data Models in Benchmarking Analysis of the Electricity Distribution Sector. Annals of Public and Cooperative Economics, Vol. 77, No. 3, pp. 271-290, September 2006.
- Ferrier, G.D. and Lovell, C.A.K. (1990). Measuring Cost Efficiency in Banking — Econometric and Linear Programming Evidence. Journal of Econometrics 6, pp.229-245.
- Forsund, F., Lovell, C.A.K. and Schmidt, P. (1980). A survey of frontier production functions and of their relationship to efficiency measurement. Journal of Econometrics, 1980. ideas.repec.org.
- Fox, K.J. (1999). Efficiency at different levels of aggregation: public vs. private sector firms. Economic Letters, Volume 65, Issue 2, 173-176.
- Fox, K.J. (2002). Efficiency of the Public Sector. Kluwer Academic Publishers, Springer.
- Fried, H., Lovell, C.A.K., and Schmidt, S. (2008). The Measurement of Productive Efficiency and Productivity Growth. Oxford: Oxford University Press.
- Guellec, D. and Potterie, B.P. (2001). R&D and productivity growth: panel data analysis of 16 OECD countries. www.oecd.org
- Hicks, J.R. (1935). Annual Survey of Economic Theory: The Theory of Monopoly. Econometrica, Vol. 3, No. 1, pp. 1-20.
- Jacobs R. (2001). Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis. Health Care Management Science, 2001 – Springer.
- Kaufman, D., Kraay, A. and Mastruzzi, M. (2008). Governance matters VII: Aggregate and Individual Governance Indicators 1996-2007. The World Bank Development Research Group

- Knight, F.H. (1933). Profit. In Encyclopaedia of the Social Sciences Vol. XII, pp. 481-486.
- Knight, F.H. (1965). Risk, Incertitude, and Profit. Boston: Houghton and Mifflin.
- Koopmans, T.C. (1951). Efficient allocation of resources. Econometrica, 19, pp.455-465.
- Kumbhakar, S., Lovell, C.A.K. (2000). Stochastic frontier analysis. Cambridge University Press: New York.
- Mintzberg, H. (1996). Managing government, governing management. Harvard Business Review, bdp.org.ar
- Meeusen, W. and J. van den Broeck (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18, 435-444.
- Niskanen, W.A. (1971). Bureaucracy and representative government. Aldine: Atherton
- Schmidt, P. (1985). Frontier production functions. Econometric Reviews, 4, pp.289-328.
- Silva E. And Stefanou S.E. (2007). Dynamic Efficiency Measurement: Theory and Application. American Journal of Agricultural Economics, Vol. 89, No. 2, pp. 398-419.
- Simpson, H. (2008). Productivity in Public Services. Journal of Economic Surveys (2009) Vol. 23, No. 2, pp. 250–276.
- Smith, P. (1988). Assessing competition among local authorities in England and Wales. Financial Accountability and Management, 4, pp.235-251.
- Smith, P. (1990). The Use of Performance Indicators in the Public Sector. Journal of the Royal Statistical Society, Series A (Statistics in Society), Vol. 153, No. 1, pp. 53-72.
- Smith, P. (1997). Model misspecification in Data Envelopment Analysis. Springer Science+Business Media, LLC.
- Smith, P. and Street, A. (2005). Measuring the Efficiency of Public Services: The Limits of Analysis. Journal of the Royal Statistical Society, Series A (Statistics in Society), Vol. 168, No. 2, pp. 401-417.
- Spottiswoode, C. (2000) Improving police performance: a new approach to measuring police efficiency. Public Services Productivity Panel, Her Majesty's Treasury, London.
- Stone, M. (2002). How Not to Measure the Efficiency of Public Services (And How One Might). Journal of the Royal Statistical Society, Series A (Statistics in Society), Vol. 165, No. 3, pp. 405-434.
- Tanzi, V. and Schuknecht, L. (2000). Public spending in the 20th century: a global perspective. Cambridge: Cambridge University Press.
- Williamson, O. (1964). The Economics of Discretionary Behaviour: Managerial Objectives in a Theory of the Firm. Englewood Cliffs, NJ: Prentice Hall.
- Wintrobe, R. (1997). Modern bureaucratic theory. In Mueller, D. (Ed.), Perspectives in public choice: a handbook. New York: Cambridge University Press.

UNIVERSITAS CAROLINA PRAGENSIS
založena 1348

Univerzita Karlova v Praze
Fakulta sociálních věd
Institut ekonomických studií

Opletalova 26
110 00 Praha 1
TEL: 222 112 330,305
TEL/FAX: 222 112 304
E-mail: ies@mbox.fsv.cuni.cz
http://ies.fsv.cuni.cz

Akademický rok 2008/2009

TEZE BAKALÁŘSKÉ PRÁCE

Student:	Michal Lebovič
Obor:	Ekonomické teorie
Konzultant:	MRes PhDr. Jan Zápal

Garant studijního programu Vám dle zákona č. 111/1998 Sb. o vysokých školách a Studijního a zkušebního řádu UK v Praze určuje následující bakalářskou práci

Předpokládaný název BP:

Efficiency of public spending

Charakteristika tématu, současný stav poznání, případné zvláštní metody zpracování tématu:

A lot has been said about the role and importance of public expenditure in the economy. In this paper, we are not addressing this discussion, we rather concentrate on one important aspect of public expenditures - the amount of the public spending does not always correspond to the benefits that a society gets from this expenditure. As an extreme example, we may take a look at the situation in the developing countries, where we can conclude, that despite huge increase in public spending (boosted by foreign aid and capital inflow) over the past decades, the effect of the public expenditures on the real economy and standards of living is nearly irrelevant in some cases. How is it possible? To answer this question, it is necessary to scrutinize the process of transformation of the public expenditure into actual outcomes in the society and identify the main factors that affect it. If we have an idea of this transformation process, we can pose another crucial question - how can we measure the amount of discrepancies between inputs and outcomes of public spending? Or in another words, how can we measure the efficiency and effectiveness of public spending?

It is a primary concern of this paper to summarize the measurement methods of spending efficiency, to outline the possibilities and limitations of their use and to conclude some theoretical implications. The crucial issues addressed are:

1) How to define effectiveness and efficiency of public spending, what are the factors determining the transformation of inputs into actual outcomes?

2) How to measure this efficiency and effectiveness?

3) How can we increase the efficiency and effectiveness of public spending?

Struktura BP:

1. Introduction
2. Public expenditure
    2.1. Basic definition, rationale and structure of public spending
    2.2. Aspects related to efficiency and effectiveness
3. Concepts of efficiency and effectiveness
    3.1. Definitions, basic concepts
    3.2. Methods of measurement
    3.3. Main determinants of efficiency
4. Cross-country and sub-country comparisons of public spending efficiency
    4.1. Basic concepts
    4.2. Limitations and shortcomings
5. Illustrative use of efficiency concepts
    5.1. Applying the outlined methods on Czech data
    5.2. Comparison of efficiency of educational spending of EU members
6. Concluding remarks

Seznam základních pramenů a odborné literatury:

- M.J. Farrell (1957). The measurement of productive efficiency, Journal of the Royal Statistical Society. Series A (General).
- M.S. Levitt, M.A.S. Joyce (1987). The growth and efficiency of public spending, Cambridge University Press.
- Afonso, Schuknecht, Tanzi (2005). Public sector efficiency: An international comparison, Public Choice.
- Afonso, Schuknecht, Tanzi (2006). Public Sector Efficinency, Evidence from New EU Member States and Emerging Markets, European Central Bank Working Paper Series (no. 581)
- In: H. Fried, C. A. Knox Lovell, & S. Schmidt (1993). The Measurement of Productive Efficiency:
- Techniques and Applications. New York: Oxford Univ. Press.
- Borge, Falch, Tovmo (2008). Public Sector Efficiency, The Roles of Political and Budgetary Institutions, Fiscal Capacity and Democratic Participation, © Springer Science+Business Media, LLC 2008.

Datum zadání:	červen 2009
Termín odevzdání:	červen 2010

Podpisy konzultanta a studenta:

V Praze dne