Univerzita Karlova v Praze
Fakulta sociálních věd
Institut ekonomických studií

DIZERTAČNÍ PRÁCE

2013 Jan Průša

Charles University in Prague
Faculty of Social Sciences
Institute of Economic Studies

DISSERTATION THESIS

Economic Efficiency, Competition and Equilibrium in Heterogeneous Production

Author:	**Jan Průša, M.Phil.**
Supervisor:	**doc. Ing. Vladimír Benáček, CSc.**
Academic year:	**2013/2014**

Declaration of Authorship

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

V Praze dne 16. září 2013

Jan Průša

Acknowledgement

I would like to sincerely thank to Mr. Vladimír Benáček for his permanent support throughout my studies.

Abstract

This thesis provides a bridge between two strands of efficiency literature. As we describe in the first part, the theory of efficiency is generally focused on equilibrium and mild deviations from it. In contrast, empirical studies document large variations in efficiency that are persistent in real economies.

We describe two theoretical concepts as driving forces behind fluctuating performance of companies. Firstly, efficiency is derived from competition and is dynamic by its nature. As production happens in time, changing supply and demand conditions induce the necessity to continuously adjust production processes. These changes are implemented under conditions of uncertainty, which directly leads to regular inefficiencies, implying that out-of-equilibrium situations are normal rather than rare. Secondly, standard models typically rely on price exogeneity to separate technical and allocative components of overall economic efficiency. We point out that this assumption is likely to fail due to extreme heterogeneity of the units of analysis. We elaborate in detail on the significance of heterogeneity in efficiency models, especially the heterogeneity of capital. As a result we demonstrate how various combinations of heterogeneous assets imply further swings in efficiency.

We show that integrating both phenomena into theoretical models provides reconciliation between the hitherto static view of efficiency and empirical studies reporting wild efficiency fluctuations. We further emphasize that the heterogeneity and time dimensions of production make the measurement of pure technical efficiency almost impossible in most applications. Instead, we propose to focus on monetary measurement of economic efficiency. We argue that this approach directly accounts for issues of heterogeneity and provides an empirical approximation of the profit function, which is the basic decision criterion of the entrepreneur.

The empirical part of the thesis provides applications of the proposed non-parametric, money-metric efficiency measurement to Czech and British small and medium-sized enterprises. In these studies we were most interested in the relationship between economic efficiency and size measured by number of employees. Our results confirm the large spread of efficiency scores, with mean efficiency ranging from 25% to 75%. While Czech industrial sectors show a positive impact of company size on efficiency, widening the sample to all sectors including services in the British case leads to a negligible negative effect which is not economically significant.

Keywords: economic efficiency; heterogeneity; profit function; small enterprises

JEL classification: D24; L25; L26; B53

Abstrakt

Tato dizertační práce propojuje dva proudy literatury o efektivnosti. Jak popisujeme v první části práce, teorie efektivnosti se většinou zaměřuje na rovnovážný stav a drobné odchylky od rovnováhy. Naopak empirické studie dokládají velké výkyvy v efektivnosti, které se vyskytují v reálných ekonomikách.

Popisujeme dva teoretické koncepty jako hlavní příčiny ve výkyvech výkonnosti firem. Za prvé, efektivnost je odvozena od ekonomické soutěže a je ze své povahy dynamická. Jelikož se výroba odehrává v čase, výrobní procesy se musí neustále přizpůsobovat měnícím se podmínkám nabídky a poptávky. Takové změny se provádí v podmínkách nejistoty, a to přímo vyvolává pravidelnou neefektivnost. Znamená to také, že nerovnovážné stavy jsou běžné spíše než vzácné. Za druhé, standardní modely většinou považují ceny za vnější parametr (mimo zkoumaný systém), aby mohly oddělit technickou a alokativní část efektivnosti z celkové ekonomické efektivnosti. Ukazujeme, že v praxi tento předpoklad velmi pravděpodobně selhává vzhledem k velké nesourodosti analyzovaných jednotek vstupů a výstupů. Detailně se věnujeme významu nesourodosti v modelech efektivnosti, především pak nesourodosti kapitálu. Jako výsledek ukazujeme, jak různé kombinace nesourodých kapitálových statků vedou k dalším výkyvům efektivnosti.

Ukazujeme, že zahrnutí obou jevů do teoretických modelů uvádí do souladu dosud převážně statický pohled na efektivnost a empirické studie dokumentující výrazné výkyvy v efektivnosti. Dále zdůrazňujeme, že nesourodost a časovost ve výrobě činí měření čistě technické efektivnosti téměř nemožné ve většině případů. Místo toho navrhujeme soustředit se na peněžní měření ekonomické efektivnosti. Argumentujeme, že tento přístup bere v potaz nesourodost a poskytuje empirický odhad ziskové funkce, která je základním rozhodovacím kritériem podnikatele.

Empirická část dizertační práce se věnuje aplikaci navrhovaného neparametrického peněžního měření efektivnosti na české a britské malé a střední podniky. V těchto studiích se zajímáme především o vztah mezi ekonomickou efektivností a velikostí měřenou počtem zaměstnanců. Naše výsledky potvrzují velké rozpětí efektivnosti, s průměrnou efektivností v rozsahu 25 až 75 procent. Zatímco české průmyslové sektory vykazují pozitivní vliv velikosti firmy na efektivnost, rozšíření vzorku na všechny sektory včetně služeb v britském případě vede k zanedbatelnému negativnímu vlivu, který není významný z ekonomického hlediska.

Keywords: ekonomická efektivnost; nesourodost; zisková funkce; malé podniky

JEL classification: D24; L25; L26; B53

Contents

Chapter 1

Executive Summary

1.1 The Premises

Competition belongs to the most powerful ideas in economics. Being able to compare economic units—be it individual agents, firms, industries or national economies—against each other implies that we are able to provide direct insights into wealth creation. Such analysis of productivity renders motivation for improvement among economic agents and thus drives development of the economy and, ultimately, of the society. The related concepts of comparative advantage, competitiveness, productivity and efficiency have provided economists with tools to measure economic performance at both microeconomic and macroeconomic level.

When we started our research in this area, we found that these tools were applied to the Czech economy mostly on the macroeconomic level. The respective studies could be broadly classified into two groups: static comparative analysis and institutional analysis. Examples from the first group include Havlík [59], who descriptively examined statistical data on labour costs, and the aggregate study by Klacek and Vopravil [67], who investigated total factor productivity growth in manufacturing industries. Institutional papers considered the relationship between performance and institutional factors such as firm ownership and foreign direct investment, including the policies promoting FDI, for example Hanousek, Kočenda and Mašika [58]. Further research on Czech enterprises focused on institutional questions related to the transition, such as privatization (e.g. Marcinčin and van Wijnbergen [84]) or financing (e.g. Klapper *et al.* [68]).

Our intention was to enhance the existing research in two directions: Firstly, by looking at strictly **microeconomic structures**, especially firm size measured by the number of employees; and secondly by building on **the dynamic view of production and efficiency**. In particular we examined the role of capital in evolution of firms' production capabilities: how efficiency changes with investment in capital. To the best of our knowledge, this alone provided a unique view on the Czech economy.

In the course of our research efforts we realized that previous literature related to theory and applications of efficiency analysis often relied on *ad hoc* solutions that were not rigorously underpinned by explicit assumptions. It was also our impression that once formalized, these implicit premises such as price exogeneity turned out to be rather strong and restrictive. We therefore devoted extensive portions of our studies to theoretical considerations related to definitions of economic efficiency and its measurement under dynamic conditions.

Even though the presented thesis had been motivated and emerged backwards (the last chapters first), we offer the reader the convenience to read it in the standard direction from the most general part to the empirical case studies.

1.2 Who Gets the Gold Medal?

One elementary concept in economics of production is the definition of economic efficiency, its separation into the technical and allocative component and the related separation of prices and quantities. While these definitions are perfectly consistent and easily understandable in theory, their application in real world situations is not straightforward.

Let us frame the problem with a simple example. Suppose that two athletes virtually compete in an Olympic sport, say rowing. Suppose the personal best time for a standard 2km indoor race for the first rower is below six minutes (5:50.9), but due to reasons unimportant for our case he performs only 6:30, which is more than 11% below his potential. The second rower has a much better day: He beats his previous personal best time of 7:02 by 10 seconds. Even though we could claim that the latter rower performed 100% during the race, the winner is quite clearly the former rower by a large margin. There would hardly be any dispute whom to award with the gold medal.

Thus as a first principle, it appears reasonable that **efficiency has to be applied to all competitors equally**. What matters in economic competition is not only the internal efficiency (personal best time), but comparative efficiency vis-à-vis other players (overall rank). For the sake of wealth creation and general welfare, market competition knows no mercy. Even though a firm improves internally, it might not be enough for business success in the market.

We could then ask what would be the measure if we wanted to compare results in, say, rowing and swimming. Obviously, comparisons of performance in different sports are hardly needed. Yet the converse is true for economic activities, and as long as these activities are fundamentally heterogeneous, we have to focus on another measure than technical efficiency. The markets naturally offer a generally applicable, cross-sectional indicator—namely prices and the derived money-valued measurement. It is here where the allocative component and overall economic efficiency gain importance.

It is unquestionable that technical efficiency is of utmost importance to the firm internally from the management viewpoint. Nonetheless, many studies concentrate on technical efficiency even in cross-sectional studies under the assumption of comparable inputs and outputs and constant exogenous prices. In this thesis we provide extensive arguments which imply that these assumptions are likely not to be satisfied. Instead we propose that **the general measurement of economic efficiency should be based on monetary values** and that separation into technical and allocative components can be applied only in special cases.

1.3 And Yet It Moves

With a paraphrase of Galileo we highlight two concepts that make measurement of separate quantities and prices which would be comparable extremely difficult.

The first reason is **the high degree of heterogeneity in all dimensions of production**. While we provide references to the extensive literature on heterogeneity below, we know from daily experience that specialization has been increasing immensely and that hardly any two products are alike. This implies that without accounting for the respective prices, variability in efficiency measured in physical (technical) units may simply just reflect the revealed technical difference between inputs or outputs.

For the benefit of the reader let us include a slightly longer quote from Griliches and Mairesse [56] which perfectly highlights our idea:

> We started our work in this area [of production function identification] with the hope that micro-data may be the answer to the various difficulties encountered at the aggregate level, primarily because this is the level which our theories claim to comprehend, and because we believed that this will reduce multicollinearity and provide us with more identifying variance. We also thought that one could reduce aggregation biases by reducing the heterogeneity as one goes down from some general mixtures such as 'total manufacturing' to something more coherent, such as 'petroleum refining' or the 'manufacture of cement'. But something like Mandelbrot's fractals phenomenon seems to be at work here also: **the observed variability-heterogeneity does not really decline as we cut our data finer and finer**. There is a sense in which different bakeries are just as much different from each other, as the steel industry is from the machinery industry.
>
> This paradox arises, in part, from the fact that our theories, while denominated in micro language of the firm or plant, have really been designed with macro questions in mind. They deal with reasonable crude aggregates: output, labour, capital which turn out to be rather vague concepts when we go down to the micro level and have to face the large number of products, labour types, machines and technologies. We have neither the data or a convenient language to describe all this variability effectively.
>
> (Griliches and Mairesse [56, p. 23], emphasis added)

The second reason is that **production happens in time**. Changing consumer preferences and technologies lead to changes of goods within their categories, the fade of some products and the emergence of new ones. This again complicates their definitions and physical measurement, let alone any meaningful comparability. We see that under dynamic conditions the role of prices as weights to decide between alternative production plans is simply irreplaceable.

The dynamic nature of production leads to another phenomenon — the role of entrepreneurs in organizing heterogeneous assets over time. It is acknowledged in management studies that to organize inputs so as to produce outputs, or **to set up production plans requires entrepreneurial judgement**. For example, K. Foss *et al.* [48] argue that in reality useful properties of heterogeneous capital are not known and have to be discovered by entrepreneurs:

> [M]ost assets have unspecified, unknown future attributes, and an important function of entrepreneurship is to create or discover these attributes. (K. Foss *et al.* [48, p. 1172])

We do not discuss the nature of entrepreneurship in detail because it goes beyond the scope of this thesis. But it is crucial to realize that profits serve as the ultimate compass guiding entrepreneurs in their business activities. Entrepreneurs organize production and prices provide them with a measure for comparison across sections and time. This measure is nothing other than profits or money-valued economic efficiency. Profitability is key—without it the entrepreneurs would be blind. Profitability and entrepreneurship are two sides of the same coin.

As such our work is a refinement of efficiency analysis which stresses the role of economic efficiency expressed by profits and the importance of monetary accounting in production.

1.4 Results and Conclusions

Finally it remains to consider some of the results shown in the empirical chapters — two on Czech small and medium enterprises and one on British enterprises.

All three studies confirm the immense variance in efficiency that is visible in the data. Typical efficiency scores range from 25 to 75% of the best performers, even after applying robust methods of estimation and accounting for extreme observations. We take this as an illustration of the enormous competitive pressure that firms face in the market.

When looking at firm size measured by the number of employees, the results were ambiguous. In the Czech case we covered manufacturing industries and we found a significant but mild effect of 7% per employee group (groups were defined as 1-9, 10-19, 20-49, 50-99, 100-250). The British study covered all industries including services and found no sizeable and statistically significant relationship between efficiency and size. It seems that dependence of efficiency on size may be limited to selected sectors only, rather to be present throughout the economy.

In the British study we also could not report any significant changes in economic efficiency over time. Because our measure was money-valued, it implies that any changes in profitability were equally shared between labourers and capitalists. In the case of Czech enterprises, we found a negative time effect of 15% over the period 2002–2005. This result appears surprising but might possibly be attributed to the specifics of economic transition from a centrally planned to free market economy, and to the accession of the Czech Republic to the EU.

These and other interesting conclusions can be derived from efficiency analysis applied to real world data. Even though the tools of researchers in efficiency are quite powerful, they must be used with caution. Most of all, researchers should always bear in mind the theoretical foundations of production economics and the logic of market competition. A lot of work still needs to be done to derive an encompassing analysis of dynamic efficiency, profitability and their drivers. Our thesis attempts to be a step in these efforts.

PART I

THEORETICAL CONSIDERATIONS ON ECONOMIC EFFICIENCY

Chapter 2

Principles of Analysis of Economic Efficiency

2.1 Introduction

2.1.1 Motivation

Efficiency analysis is already an established field in both economic theory and practice. It is commonly used to compare relative performance of economic units on different levels of aggregation — ranging from single production lines to plants to firms to whole industrial sectors. Hundreds or even thousands of empirical studies employed efficiency measurement to compile rankings of various producers, while theoretical papers are predominantly looking for the best computational methods that would yield statistically reliable results.

The need for efficiency analysis can be traced back to the idea of competition. **Competition** belongs to the most powerful concepts of economics. It captures the idea of consumer choice among alternatives, some of which are subjectively valued more than the others. Those alternatives that are perceived as better satisfying consumer preferences will be rewarded accordingly more, and competition is the market mechanism which ensures selection of the most preferred choices. Efficiency then emerges as a measure of competitive performance against a given benchmark. For now we ignore the question if this benchmark should be absolute or relative.

The attentive reader will quickly realize that for entrepreneurs there is another obvious indicator of their success, namely the profitability of their enterprise. The reader might then ask himself why economists work with efficiency instead of profitability. The answer consists of at least three points: (1) Profit is a one-dimensional business indicator, while efficiency estimators attempt to capture systematic patterns in production, such as the employment of factors of production. (2) Efficiency, albeit somewhat modified, can be evaluated even in the cases when standard profit accounting is not available, e.g. for certain publicly owned and/or provided services such as analysis of hospital performance (see e.g. Biørn *et al.* [14]). (3) As we will show below, a full model of economic efficiency is nothing else than an estimator of the profit function. As far as this last point is concerned, we have to admit that the objection of the reader is valid and that the terminology of economists is indeed somewhat confusing.

Even a brief look at the literature reveals that empirical studies about efficiency analysis have long been one step ahead of economic theory. Practical researchers simply adopted the idea of variation in efficiency as a matter of fact (Reifschneider and Stevenson [97] write about **systematic inefficiency**) and focused on two agendas: (1) How do we best measure efficiency, and (2) what do the results say about the economy; the former question clearly overlapping with theory but rather statistical than economic. In contrast, mainstream microeconomic models centre around the equilibrium where room for inefficiency is very limited. The field of industrial organisation developed a lot of models which depart from the simplest notion of perfect competition. However the very nature of equilibrium as the balanced state cannot naturally accommodate systematic and large fluctuations of economic performance which we observe in practice.

2.1.2 Plan of Work

This essay builds on our previous empirical work where we analysed the efficiency of Czech and British small and medium enterprises. We will highlight our findings and formulate lessons for the theory of efficiency analysis. Besides empirical results, our work contributes towards theoretical models of efficiency in several aspects.

Above all, the approach that we propose provides the basis to reconcile efficiency and profitability. In most instances of efficiency analysis economists stressed the technical production function, which however is not only almost unobservable to the researcher, but generally also irrelevant. Instead we believe that economists should care mainly about indicators of economic exchange values, that is monetary indicators with some prices always attached. We further show how the focus on monetary values can lead to a clearer treatment of efficiency, which is moreover much closer to profitability.

As a by-product, we pave the way to the reconciliation of empirical studies and theoretical models. Empirical results suggest that the theory has to pay much more attention to the dynamics of production. The time dimension in production leads to a more sophisticated concept of equilibrium in entrepreneurial activities, which is able to accommodate varying conditions and existing inefficiencies as the standard outcome.

The rest of the paper is organized as follows. We first describe the concept of efficiency and provide an extensive survey of the existing microeconomic theories which incorporate inefficiency. We then proceed to a general model of economic efficiency which is in accord with the world of inefficiency. As the final part we provide a summary of previous empirical results that are presented in separate parts of this thesis.

2.1.3 Subject of Analysis

For the rest of the thesis, the smallest unit of analysis is **the firm**. The modelling is based on market interaction of firms as single decision making units (DMU). This is in contrast to models of firms which describe the interaction of agents within firms—most notably owners, managers and workers. We acknowledge that internal organisation within firms is one of the main drivers of efficiency, if not the most important one. Therefore, even though these processes enter our analysis only exogenously, we always bear them in mind as an explanatory factor of observed inefficiencies.

Finally, before we dive into microeconomic analysis of efficiency, it appears useful to recall the generic meaning of efficiency. **Efficiency** generally represents **successful execution**. The term implies that there exists a known or pre-defined plan that is to be fulfilled, a benchmark that is to be matched. Efficiency is the ability to achieve: Certain aim or target is assumed or expected to be possible and we evaluate whether it was achieved and with what effort. Throughout the thesis we shall keep these concepts in mind.

2.2 Microeconomic Paradigms of Production

The production side of the economy has to solve the following task: How to employ available resources most efficiently to satisfy human needs? In the end it is consumer preferences which determine all human action. Nevertheless specialization and highly advanced production methods form a complex system of trading relationships of its own. Goods which consumers demand are in most cases delivered in the last step of a long chain of trades.

As a consequence, understanding production is far from trivial. This has to be reflected in microeconomic analysis. We shall now introduce three paradigms along which current microeconomics of production is evolving.

2.2.1 Theory and Empirical Studies

One possible approach to understanding production is to construct an accurate **theoretical description**. Industrial organization is the field of economic theory which builds models on interaction of companies in the markets. These models cover different market structures in the neoclassical range from perfect competition to monopoly. Shy [103, p. 61] presents this framework in a simple chart where market competition is classified according to the following criteria:

- perfect competition (price taking behaviour), oligopoly, duopoly (both facing residual demand) or monopoly (facing market demand),
- non-cooperative or cooperative interaction,
- simultaneous, sequential or repeated interaction,

- the subject of business decisions is either quantity or price.

In this list we see all the main categories which represent the building blocks of current theoretical models of production. It is apparent from this classification that only little attention is paid to the time structure or heterogeneity, the two issues that we will concentrate on below.

On the other hand, empirical research is largely focused on the measurement of relative performance of companies. We leave aside managerial sciences, which are rather based on *ad hoc* practical solutions for decision making. A large body of empirical literature is devoted to productive efficiency, which is also the main topic of our thesis. We will treat this concept in extensive detail below. Suffice it to say here that the underlying idea is to compare companies according to their output-to-input ratio.

The fact remains that even advanced theoretical models of market structures do not allow for inefficiency. The reason is that they work with competitive equilibria, which by definition assume efficiency at least at the firm level.

2.2.2 Decomposition of Efficiency

2.2.2.1 The Concept of Technical and Allocative Efficiency

Efficiency in each enterprise has two dimensions. The first one is related to physical production and focuses on the employment of technology. This component is captured in **technical efficiency**. It shows whether machines are operated close to their nominal capacity. The second one is related to company sales and focuses on the performance of the sales force. This component is called **allocative efficiency**. It is based on the gains from exchange and it shows whether products are sold at a profit. A company can only achieve allocative efficiency if it is able to match its production plan to the demand from customers. Both components together yield **economic efficiency**.

Both components of efficiency are depicted in the left part of figure 2.1. The figure shows an output isoquant curve of a production function with two inputs $y = f(x_1, x_2)$, and the isocost line with slope $-\frac{w_1}{w_2}$, where w_1, w_2 are the respective input prices. The efficient producer is located at the point of tangency between output isoquant and isocost lines. The producer F is inefficient:

1. F is producing output y with quantities of inputs x_1 and x_2 larger than necessary, both quantities could be decreased while keeping the same output y. **Technical inefficiency** of F can be measured as the distance BF.

2. Even at B the firm would not be using the cheapest combination of inputs to produce y. The firm could substitute the more expensive input x_2 for the cheaper input x_1 (moving along the output isoquant curve) and thus decrease its costs. **Allocative inefficiency** of F can be measured as the distance AB.

The overall **economic inefficiency** is then equal to the distance AF. A firm can achieve full economic efficiency if and only if it is both technically and allocative efficient at the same time.

It must be noted that the decomposition of efficiency assumes that prices and quantities can be separated. This is equivalent to assuming that we can identify a non-monetary, purely technical production function in physical units. The case where this assumption is not satisfied is thoroughly analysed in section 2.4 below.

2.2.2.2 Relation between Technical and Economic Efficiency

It is important to understand how both types of efficiency relate to the behaviour of the firm. Because technical efficiency represents capacity utilization in physical units, it is determined solely by internal factors. In this sense technical efficiency measures the **efficiency within the firm**. On the contrary, allocative efficiency of a firm is driven by the interaction of that firm with other agents in the market. Measurement of allocative efficiency is based on prices which can only exist after the firm has negotiated with its suppliers and customers and after it has taken into account the behaviour of its competitors. Allocative efficiency therefore incorporates the **efficiency across firms**. Perhaps more precisely, allocative efficiency measures how internal processes and decisions within the firm position this firm in the market when facing interaction with other agents.

This distinction is paramount: An enterprise can evaluate its technical efficiency on a stand-alone basis without any market reference, e.g. from nominal capacity of its machines. Consider however that a firm might be technically efficient but if its product are not demanded by anyone, the resources are completely wasted. Here comes the crucial function of allocation through market interaction which takes into account various demands on the one hand and scarce resources on the other hand. Unless the firm goes to the market to demand inputs and offer its products, nothing can be said about its allocative (and thus also economic) efficiency.

Even though technology *per se* is not the subject of economics, much research has been devoted to the study of technical efficiency, often to the point where allocative (and economic) efficiency is disregarded. Farrell [47], whose paper deserves credit for sparking the interest in systematic development of feasible efficiency measures, devoted a large part of his article to technical efficiency. About allocative efficiency (which he called **price efficiency**) he observed:

> The price efficiency of a firm also depends on the measurement of inputs, but in a rather complex way, so that such problems are best discussed *ad hoc*. [...] Thus, price efficiency is a measure that is both unstable and dubious of interpretation; its virtue lies in leaving technical efficiency free of these faults, rather than in any intrinsic usefulness. (Farrell [47, p. 260–261])

Farrell correctly understood the complicated nature of evaluating economic efficiency as a whole and decided to leave the problem aside. Ever since Farrell, the models of efficiency were generally derived from the production function, looking primarily at physical inputs and outputs and taking prices simply as exogenous labels for technology. This approach widely persists to this day.

This contrasts with the needs of company owners, who are primarily interested in profitability of their enterprise. In the real world we see that technical efficiency can be achieved relatively easily. Modern technologies allow plant managers to see the output in real time (even remotely). This information can be immediately compared to nominal capacity and actions can be taken in the case of disruptions. What matters much more is the purchasing and selling abilities of the firm, that is the exchange for monetary values. In fact we observe that technical inefficiency in the form of unused capacity is the effect, and not the cause, of allocative inefficiency in the form of not being able to sell.

2.2.3 Exogeneity of Prices

The relationship between the technical and allocative components of economic efficiency which we just described is an integral part of a wider debate on exogeneity of prices. Efficiency models often assume that prices are parameters that can be treated as if they were independent of the evaluated system. In other words, prices can be fully separated from quantities and applied across all observations. This amounts to assuming that in the competitive equilibrium the Law of One Price (LoOP) holds, i.e. that the same price applies to all units in one category.

Following the tradition of focus on technical efficiency, empirical papers on efficiency still assume that prices are simply exogenous variables separated from quantities. This assumption is kept even by researchers who want to measure economic efficiency, for example whenever aggregate price indices are used. In fact recent literature develops the idea of replacing unknown prices with prices estimated from the data, such as in Kuosmanen *et al.* [76]. In this study the authors attach to observed (physical) input and output vectors their "shadow prices" calculated as the "most favourable prices" for each company. These are such prices which assign to the given firm its best profitability ranking within the sample. Leaving aside the credibility of this "most favourable" principle, we immediately see the exogeneity in play: A randomly selected price vector is used to calculate profit of all firms, regardless of possible individual price differences which firms were able to achieve in reality.

However it is difficult to justify such a degree of exogeneity in empirical studies. As noted already by Farrell [47, p. 261], a linear isoprofit line implies perfect price elasticity. That is, assuming LoOP is tantamount to assuming that the price of a good (either an input or an output) does not change whatever the amount of that good. Still, this could at least be justified on the grounds that the LoOP holds approximately for the quantities within the analysed sample. In other words, the range of quantities within the sample might be narrow so that price might be approximately the same for whatever quantity.

A more serious objection is that the LoOP could in fact be reformulated as **the Law of One Good**. Namely, all units within any given category would have to be perfectly interchangeable (homogeneous), so that one price might be applied equally to all units. We already see the snag: The few categories of goods (e.g. two inputs and one output) that are commonly used in empirical studies are surely too wide for the Law of One Good to apply to them. To the contrary, it seems reasonable to assume that given the heterogeneity of goods even within one category, each price will be quite unique to the single observation.

2.2.4 Summary: Renaissance of Economic Intuition

The presented paradigms illustrate that the advances in microeconomic theory of production entail many detailed structures that are only rarely applicable in empirical studies. This concerns not only the models of competitive interaction among firms, but especially the separation of prices and quantities and the focus on production technologies.

On the other hand, common theoretical models do not incorporate other observed features of production. Here we have in mind above all the large variations in efficiency, as well as the heterogeneity of goods, as a consequence of which there cannot be unique prices which would hold universally throughout the sample. In the latter case it is not so much the violation of the LoOP which applies to **identical** goods but rather the fact that the goods themselves (especially within the broadly defined observed categories) are not identical in the first place, meaning that the LoOP cannot hold **ex ante**.

These shortcomings trickled down to empirical comparative studies and aggregation, which were often based on trial and error. As was noted by Blaug [16, p. 171]:

> Much of this empirical work [on aggregate production functions] was little more than “measurement without theory”.

We would add that the same holds for many studies on efficiency.

In the next section we turn to theoretical models which take inefficiency into account.

2.3 The Case for Inefficiency

2.3.1 Theories of Inefficiency

2.3.1.1 A Neoclassical Firm

Before we present some of the main theoretical contributions that allow for inefficiency, we have to briefly introduce the simplest neoclassical model of production, such as the one in chapter 5 of Mas-Colell *et al.* [86].

Its logic is based on the approximation of the long-run equilibrium where firms know their technology represented by a production function. By assumption of profit maximizing or cost minimizing behaviour, firms attain both technical and allocative efficiency. Hence, allocation at the firm level is efficient. If all firms face the same technology and the same prices, as is the case in the long-run, they will all lie on the same aggregate production frontier. This means that both types of inefficiency (technical and allocative) occurring either within firms or across firms are assumed away.

This framework is useful to illustrate the basic principles of market interaction. Its inapplicability to observed data stems not so much from the ignorance of short-term variability which is explicitly excluded at the outset by assumption, but more from the idea of a single terminal long-term equilibrium. Production possibilities are clearly changing in time, so that we encounter firms at various stages of their development. This effect is important because both our daily experience and the available empirical evidence show that there are big differences in production abilities among firms.

2.3.1.2 Introducing Frictions in the Neoclassical Paradigm

Varying economic performance implies that a realistic economic theory has to model efficiency differentials. As shown in section 2.2.1, one possible approach is to attribute them to market interactions of firms and to focus on detailed market structure. In this set-up higher profits are associated with more monopolistic structure. Such analysis is certainly valid, but it cannot account for differences among similar firms, that is firms which operate in the same market with comparable products. In other words, the theory needs to depart from the symmetry assumption.

In 1937, Ronald Coase [26] posed himself the following question, immediately suggesting an answer:

> Our task is to attempt to discover why a firm emerges at all in a specialized exchange economy. [...] The main reason why it is profitable to establish a firm would seem to be that there is a cost of using the price mechanism. (Coase [26, 390])

In retrospect this observation may appear almost trivial, yet this contribution finally made economists accept what they had seemed to ignore: Whatever smoothness might be assumed in theory, it does not hold in practice. Coase prepared the ground for inefficiency to become widely accepted among economists.

Following this idea of frictions or transaction costs, more realistic models of firms have been developed. One stream incorporated a more dynamic view of firm's capital, which explicitly takes into account different **vintages of capital**. This term, used e.g. by Johansen [64], was later generalized to **technology**, but the original literal description of **vintages** is quite instructive about the nature of firms in reality. The model assumes decreasing efficiency of capital in time (depreciation) and continuous investment of firms in new vintages of capital enabled by technical progress. Such a high-level model could be empirically implemented on the macroeconomic level, see e.g. Wickens [114] who estimated the aggregate U.S. production function. Recent macroeconomic growth models have also included investments in capital of different vintages, such as Jovanovic and Yatsenko [65].

Another field that was derived from the work of Coase is the theory of transaction costs and institutional economics, as described in Moschandreas [90, chapter 3]. The main idea covers the conflict between the profit seeking of the owner and the rent seeking of the managers, where managers and employees in general might choose less than efficient decisions as long as it enhances their well-being (e.g. excessive travel expenses). This approach was further developed into the theory of incomplete contracts, imperfect monitoring and the principal-agent problem. All of these concepts are now well developed and understood in order to provide solid theoretical backing for empirical inefficiency measurement.

2.3.1.3 X-Efficiency

A completely different paradigm was offered by Harvey Leibenstein [80] in 1966. Leibenstein argued that a significant proportion of empirically documented inefficiencies stem from sources other than technical and allocative inefficiency. He introduced a new term: **X-efficiency**, and developed a theory based on this definition.

Frantz [52] points out that the difference between X-efficiency and the neoclassical paradigm lies in the main assumption: While the latter assumes maximising behaviour in all circumstances, the former allows for situations where individuals are consciously not optimising. In Leibenstein's [80, p. 407] own words:

> The simple fact is that neither individuals nor firms work as hard, nor do they search for information as effectively, as they could.

Leibenstein's article was followed by an intensive discussion. Stigler [108], De Alessi [39] and others defended the neoclassical paradigm, arguing that it developed enough tools to handle inefficiencies (see section 2.3.1.2). Yet as noted by Frantz [52], X-efficiency lies outside the neoclassical paradigm, and hence cannot be refuted by neoclassical arguments.

The theory of X-efficiency amounts merely to **inefficiency by assumption** which does not offer much room for explanation. Nevertheless by calling the emperor naked, it marks an important step in debunking the concept of **full efficiency by assumption**.

2.3.1.4 Austrian Theory of Production

If we put aside managerial sciences, then the Austrian school provides perhaps the most detailed analysis of entrepreneurial activities. The school concentrates on dynamics of the economy and regards entrepreneurs as those who pursue arbitrage. This approach implies two important lines of thought regarding inefficiency:

1. The plan of production is not explicitly known in advance but is uncertain and has to be discovered. Some entrepreneurs are better at this process than others. In the words of Foss and Klein [50], these entrepreneurs have better **judgement** about uncertain future production conditions.

2. The optimal production plan is changing in time. The continuous dynamic adjustment is driven by entrepreneurs who exploit profitable opportunities as they emerge.

The adjustment process together with uncertainty about the future imply certain natural volatility of economic performance.

In comparison to this paradigm, the neoclassical stereotype suffers from the static equilibrium-always view. Sautet [101, p. 10] calls this the **market theory problem**, which is:

> the **inconsistency** involved in trying to answer questions that would not exist in an equilibrium-always world. (emphasis original)

Inefficiency can be regarded as one example of the market theory problem because inefficiency is simply not admissible in equilibrium. With regard to efficiency, Sautet writes (p. 49 *ibid*):

> Understanding competition as a process helps explain empirical phenomena that cannot be explained by standard neoclassical theory, such as the persistent dispersion of returns that is wider among firms of the same industry than across industries (Rumelt 1984, 1987) and the different rates of growth among firms of the same industry (Penrose 1995 [1959]).

Thus it can be concluded that the Austrian school incorporates realistic assumptions about the production process, which then give rise to differences in performance. Inefficiency is understood as an inherent component of entrepreneurship without negative connotation *per se*, just as the runner-ups in sports championships are not necessarily bad sportsmen.

2.3.1.5 Summary: Acknowledging Inefficiency

Above we list several models that feature inefficiency as an essential part of economic reality. The neoclassical textbook model is known to all economists as the basis for microeconomic analysis of production. Nonetheless, it has to be properly understood as a starting point from which the analysis departs in order to develop more realistic models, rather than the gauge against which all empirical situations are measured or compared.

The focus on equilibrium made economists look for **efficiency in all situations, at all times**, an approach resembling almost an efficiency mantra. But the presented theories imply that **inefficiency has to be systematically analysed** as an essential component of the dynamic adjustment process.

2.3.2 Heterogeneity and Inefficiency

It is our impression that one phenomenon is severely underrated when evaluating efficiency, namely the immense heterogeneity of all economic activities. Even though we as consumers carefully consider several distinctive options when making a purchase, we as economists tend to treat highly differentiated goods as homogeneous groups. To put it simply, we compare things that are not really comparable, and not surprisingly this creates high variance in results. The inefficiency is then nothing more than **revealed difference** within categories that we compare.

2.3.2.1 Heterogeneity of Products

If products are heterogeneous, one of the main assumptions of perfect competition is violated, which implies a different form of market interaction. This is why the typical approach to product heterogeneity in modern economics follows the models of imperfect (monopolistic) competition that was developed by Joan Robinson in her work The Economics of Imperfect Competition and by Edward Chamberlin in The Theory of Monopolistic Competition. These authors proposed that even if there were many firms in the market, it would still not be uncommon that each of them faced a downward sloping demand curve. Firms would therefore perceive a price-quantity relationship which they could use to adjust their sales.

As explained by Brakman and Heidra [17, introduction], imperfect competition was precisely formulated in the Dixit-Stiglitz model in 1977. The model introduces a composite differentiated good that has N varieties which are imperfect substitutes for each other (*ibid*, equation 1.2). By assumption the model is completely symmetric, so that the price of all varieties is the same in the equilibrium. This means that the focus of the Dixit-Stiglitz model is again on the balance instead of the variance in products.

A detailed model that accounts for product heterogeneity is the much overlooked contribution developed by Lancaster [78], [79].

Lancaster's Model of Product Characteristics[1]

In standard consumer theory preferences are defined on the space of goods, i.e. consumer utility function takes a vector of goods as its argument. In contrast, Lancaster postulates that consumers in fact choose between goods according to their characteristics, i.e. consumer utility function takes a vector of characteristics as its argument. The link between goods and characteristics is established by means of two axiomatic principles stemming from everyday experience:

1. A single good may possess several different valuable characteristics.

2. A number of different goods may have some of their characteristics exactly the same.

[1] This section follows our previous working paper written jointly with Pavel Ryska [99].

In formal terms Lancaster [79] defines the transformation from the goods space ($\mathcal{G}$-space) to characteristics space ($\mathcal{C}$-space):

$$
\begin{aligned}
\boldsymbol{z} &= \mathbb{B}\boldsymbol{x}, \\
\mathbb{B} &\in M(r \times n), \\
\text{where } z_i &= \sum_{j=1}^{n} b_{ij} x_j,
\end{aligned}
\tag{2.1}
$$

where b_{ij} is the quantity of the i-th characteristic possessed by a unit amount of the j-th good, z_i is the quantity of the i-th characteristic, x_j is the quantity of j-th good, r is the number of characteristics and n is the number of goods. $\mathbb{B}$ is the consumption technology matrix with elements b_{ij} and it describes all goods on the market in terms of their characteristics. Note that equation (2.1) is based on two further assumptions:

1. **Linearity.** $z_p = b_{pj} x_j$.

2. **Additivity.** $z_q = b_{qj} x_j + b_{qk} x_k$.

The consumer optimisation problem then takes the form:

$$
\begin{aligned}
&\max_{\boldsymbol{x}} \; u(\boldsymbol{z}) \\
\text{s.t. } \boldsymbol{z} &= \mathbb{B}\boldsymbol{x} \\
\boldsymbol{x} &\geq \boldsymbol{0} \\
\boldsymbol{p}'\boldsymbol{x} &\leq Q
\end{aligned}
\tag{2.2}
$$

where the last line is the budget constraint. The choice variable is still the vector of goods $\boldsymbol{x}$ but it enters the objective function only through the consumption technology transformation. The vector of characteristics $\boldsymbol{z}$ is what matters to the consumer.

We need to highlight the separation mechanism between $\mathbb{B}$ and $u(\bullet)$. The consumption technology matrix $\mathbb{B}$ is assumed to be objectively observable at least in theory. This is equivalent to producer theory where production technology is assumed to be hypothetically observable by all producers. $\mathbb{B}$ represents information that all consumers can agree on: area of a flat, engine power of a car, or weight of a laptop. Subjective perceptions about goods, that is how individual characteristics are relatively valuable to a given consumer, are still contained in the utility function $u(\bullet)$.

Implications and Applications of Lancaster's Characteristics Model

Since in a general case $r = n$ cannot be guaranteed, properties of a solution to (2.2) are not straightforward. In particular, it appears reasonable to assume that the number of goods will be higher than the number of considered characteristics ($r < n$), since producers will try to develop goods combining various characteristics to satisfy as many consumers as possible.

In this case Lancaster [79] shows that consumer with utility function $u(\bullet)$ will not consume all n goods but at most r goods and often even less than r goods. In other words, some of the goods will not be purchased at all by this consumer. They will be those goods which do not have an attractive combination of price-discounted characteristics so as to satisfy the given preferences. Lancaster explains that corner solutions to the optimisation problem become general (occurring much more often), as compared to textbook consumer theory. Another important property of the characteristics framework is that small changes in prices of goods that are not consumed do not affect equilibrium choice of $\boldsymbol{x}^*$.

Lancaster's model illustrates that the implications of properly defined product differentiation are very different from the Dixit-Stiglitz symmetry. It also suggests that the study of heterogeneity deserves more attention. In this respect it is useful to mention that when it comes to recognizing the importance of product heterogeneity, empirical economics seems to be one step ahead of theory. The characteristics model has been applied in frequently cited empirical studies on demand systems for differentiated products. Perhaps the most influential is the paper by Berry, Levinsohn & Pakes (BLP [12]) who estimated demand parameters in the U.S. car market.

In order to make the concept of characteristics operational and econometrically sound, BLP worked with observed and unobserved characteristics of cars and estimated the respective demand slopes for the observed (or better: measured) ones. BLP enhanced their model by second choice data (BLP [13]). Other papers include Bresnahan, Stern & Trajtenberg (BST [18]), who estimate the demand system for personal computers. These studies demonstrate that heterogeneity of products is a relevant and recognized fact in empirical economics, but much less so in theoretical economics.

2.3.2.2 Heterogeneity of Labour

Having seen the importance of product differentiation for models of consumer demand, we turn to heterogeneity of inputs into production. It is our claim that input heterogeneity has crucial impact on productive efficiency.

Economics of production traditionally works with two main aggregate inputs: labour and capital. It is not difficult to realize that such aggregation is not appropriate for microeconomic analysis. Several empirical studies recently pointed out that incorrect aggregation may lead to severe biases of results even on macroeconomic level. Alonso-Borrego [4] divided workers in his Spanish data into two subgroups — production (blue collar) and non-production (white collar) workers; and showed that adjustment costs of hiring and firing each type of worker are highly significant, being lower for the presumably less skilled production type.

Bresson *et al.* [19] present a still more important finding. Even when the estimate of an aggregate model of labour demand appears satisfactory, the aggregation can hide misspecification problems which become apparent only if the underlying disaggregated model is estimated. Their conclusion is straightforward:

> It seems that one should forsake the hope of working on total employment with

> aggregate variables even at the firm level. On the contrary, one should estimate Euler equations for different skill levels and **specify general functional forms** for adjustment costs and **production functions**. (Bresson *et al.* [19, p. 166], emphasis added.)

We see that the importance of differentiation of labourers by their skill is an acknowledged fact in empirical economics.

Market for Labour Characteristics

In our previous work with Pavel Ryska [99] we argued that Lancaster's characteristics model introduced in the previous section 2.3.2.1 can be successfully applied to the labour market. In our opinion, the characteristics model easily accommodates the main features of labour markets which can be observed in practice: Above all the advanced level of workers' specialization and the elaborate screening during job matching process for the majority of vacancies. We showed that the demand for characteristics can explain observed unemployment as a **wage-characteristic mismatch of the unemployed worker**, meaning that:

> The crucial conclusion of the model concerns the importance of relative prices of labour characteristics: If a low-effort or low-skilled worker is unemployed, either he must lower his wage significantly (this shifts the vertex of the $\mathcal{G}$-tetrahedron away from the origin), or he must change the characteristic offered to high effort or high skill as expressed by coefficients of matrix $\mathbb{B}$. [...] Another option for the worker is to look for an employer with a different utility function. (Ryska and Průša [99, p. 15])

Below we derive the impact of worker heterogeneity on firm efficiency.

Assume that x_j's represent n labourer **characteristic-types** who offer their work on the market, and each of the workers has r characteristics z_i. Again it seems natural to assume that $r < n$, since firms generally look for a limited bundle of knowledge and skills. Thus employers solve the optimisation problem in (2.2).

The two important properties of any solution to (2.2) described in the previous section apply. Given workers' characteristics captured in $\mathbb{B}$, firms will not consider all candidates for their vacancies but rather search for those with the most favourable combination of desired characteristics discounted by their prices.

We illustrate this in Figure 2.2 with $n = 3$ and $r = 2$, where the budget constraint in the labourer $\mathcal{G}$-space is transformed by $\mathbb{B} \in M(2 \times 3)$ to the characteristics $\mathcal{C}$-space. On the left hand side (LHS), the tetrahedron represents combinations (consumption bundles) of three characteristic-types of labourers $(x_1, x_2, x_3)'$ which the firm can afford given budget Q. Since the firm is interested in characteristics, it projects the tetrahedron into $\mathcal{C}$-space. In terms of geometry, the four vertices on the LHS correspond to the four vertices on the RHS in Figure 2.2. The tetrahedron is convex, hence the four vertices determine the envelope of the convex quadrilateral on the RHS, which is the budget constraint in $\mathcal{C}$-space.

Implications of Worker Heterogeneity for Productive Efficiency

Figure 2.2 depicts the optimum of an employer with the utility function $u(\mathbf{z})$ conditional on budget constraint Q. The optimal choice lies on one of the edges of the quadrilateral, so that a combination of two out of the three characteristic-types of candidates will be chosen for the job. Going back to $\mathcal{G}$-space, the optimum will lie on one of the edges, say between x_1 and x_2, while labourers of characteristic-type 3 will not be considered ($x_3^* = 0$). Note that this is the crucial implication for unemployment: As is apparent from the $\mathcal{C}$-quadrilateral in figure 2.2, the characteristic-type corresponding to the leftmost non-zero vertex is not competitive and will not be demanded because his price-discounted combination of characteristics is too expensive.

Consider what happens if the firm is small and wants to hire one employee only. Ideally, the firm would like to find a worker who combines both characteristics, with the optimal choice being e.g. $\frac{3}{4}$ of worker type 1 and $\frac{1}{4}$ of worker type 2. In reality workers are not perfectly divisible, if only for the reason that they prefer to find one full-time job instead of several part-time jobs. Moreover, transaction costs for the firm of employing more part-time workers might be prohibitive. The firm is then likely to hire one worker of type 1, foregoing the characteristics associated with type 2. Therefore we see that **the efficiency of a firm will be influenced by its ability to match the need for certain characteristics with the supply of workers**, a matching process that will be distorted by discrete choice among imperfectly divisible workers.

Another complication for firms arises from the frequency of corner solutions. If corner solutions occur more often in the characteristics framework, as is suggested by Lancaster, firms will more often face decisions that do not result from smooth optimisation but rather from second-best solutions on the edge between alternatives. Two firms with different slopes of utility isoquants may still choose to hire the same type of worker. As seen from the perspective of labour supply, two workers of the same type might be doing different jobs to which they are imperfectly suitable, and moreover with a different degree of misfit. Again, **the occurrence of choices on the vertices of budget constraints will considerably impact productive efficiency**.

We believe that the reader will deem our arguments compelling for the following conclusion: It appears hardly acceptable to measure labourers simply as unit inputs in the production function (e.g full time employee equivalents), disregarding their skill and other differences.

2.3.2.3 Heterogeneity of Capital

With capital we arrive at the most complicated productive input. Unlike land or labour, capital does not have an intuitive general representation. The word capital does not even represent one concept in economics. The two most frequent uses are (1) capital as the produced means of production, and (2) capital as the money that finances production. And yet capital is **the everyday word** of economists and businessmen alike.

Even though there are many economic theories of capital, they all accept its heterogeneity. Lachmann [77, p. 2] writes:

> All capital resources are heterogeneous. The heterogeneity which matters is here, of course, not physical heterogeneity, but heterogeneity in use.

Where all capital theories diverge is how this heterogeneity should be treated in economic models.

Cambridge Capital Controversies

Cambridge capital theory controversies stand for a series of disagreeing papers on the relevance of capital heterogeneity for economic theory. The common neoclassical solution to heterogeneity is to aggregate all capital into one sum K. The subject of the controversies was whether this is a possible and valid approach.

The aggregation does not only circumvent the underlying structure of capital, but it also ignores any frictions that arise in dynamic models from investments and divestments in different sectors of the economy. The importance of capital adjustment costs was shown in many empirical studies. Recent results such as those by Cooper and Haltiwanger [31] indicate that adjustment costs have distinctly different forms (e.g. aspects of convexity, non-convexity, irreversibility and disruption of production), thus adding a further dimension to the heterogeneity and complexity of capital.

In addition, there is a more important theoretical inconsistency. Already when the debate started, it had long been known that heterogeneous capital invalidates some of the results of neoclassical models of production. Heterogeneous capital cannot be measured in physical units but has to be aggregated by its **value**. This value of capital equals the cost of its production, or alternatively the present value of future outputs which the capital goods generate. However, both of these measures include time and hence the rate of interest, which in a standard neoclassical model depends on the quantity of capital itself. This implies circularity of the model and causes the so-called Wicksell effects in the form of inventory revaluations (price Wicksell effects) or differences in the physical stock of capital (real Wicksell effects).

As a result of Wicksell effects two situations can occur in neoclassical models with heterogeneous capital. The first one is **reswitching**, which means that a given technique of production can be preferred at two different interest rates while a different technique is optimal at an intermediate interest rate. The second one is **capital reversing**, a situation in which a **lower** capital-labour ratio is associated with a **lower** interest rate.

Cohen and Harcourt [30], who provide a summary of the controversies, write that attempts to extend neoclassical results from one-dimensional to multi-dimensional models failed because

> Wicksell effects made the links between capital and interest rate bidirectional rather than one-way.

While neoclassical economists generally acknowledge issues related to heterogeneity of capital in their models, they quickly add that these are not significant and that since we do not have an adequate replacement, the current standard model of production function with one-dimensional capital input shall continue to be used. But the critics from Cambridge, England, claim that reswitching and capital-reversing are game changers which occur precisely when the theory passes the point of generalization from simple to more complex models with heterogeneous capital.

To this day, the controversy remained unresolved, as Cohen and Harcourt [30] note:

> The Cambridge controversies were **not** a tempest in a teapot. [...] Major issues—explaining (and justifying) the return to capital, visions of accumulation, limitations of equilibrium tools—were and are at stake. While many of the key Cambridge, England, combatants stopped asking questions because they died, the questions have not been resolved, only buried.

This is the reason why we attempt to revive heterogeneity and implement it in efficiency measurement.

Implications for Productive Efficiency

From the point of microeconomic efficiency analysis, the lack of a physical unit of capital distorts the concept of the technical production function. Efficiency studies found two obvious ways around this issue:

1. Plugging the value of capital into the technical production function.

2. Working on a sufficiently detailed micro-level where specific physical units of capital can be identified.

By now the reader will certainly have foreseen our objections against both of these approaches. The first proposal entails a plain and severe misspecification. For once we decide to attempt a strict distinction between technical and allocative efficiency, it can hardly be justified that the technical production function should take the value of capital as its argument. Still more often that not this is the tacitly acknowledged partial way out.

The second proposal points in the right direction of what we would ideally have for a proper understanding of production on the micro level. Unfortunately this strand of research usually commits the mistake of overconfidence as it neglects the true extent of specialization in the economy. Experience tells us that higher-order capital equipment above a certain threshold is nowadays largely custom-made for the given firm and then fine-tuned during installation at the plant in order to be fully adjusted to the unique product. Even capital units at different plants of a single company might be hardly comparable.

To illustrate this point, consider the plant data that we collected in our previous study (Průša, Klimešová, Janda [96]) on six Czech photovoltaic power plants. Table 2.1 shows the production of the plants as compared to their nominal production capacity, and also the resulting capacity usage ratio. All of the plants are of the same type installed in similar conditions, nevertheless we observe a large variance in their productivity: The ratio of the worst to the best plant equals 0.84, or 84%.

This example is intended to highlight the main point: Capital is a complicated input in terms of its structure, but that only makes it impossible to ignore its main property — heterogeneity. Two consequences follow for efficiency measurement: (1) In most cases capital units that could be meaningfully compared within a (technical) production function framework cannot be well defined. **Due to heterogeneity counting capital in physical units is simply not tenable** for most empirical applications. (2) The choice of specific capital equipment will directly impact productive efficiency of a firm. Therefore, **from the point of efficiency capital heterogeneity is extremely important**.

2.3.2.4 Heterogeneity of Entrepreneurs

Once we recognize the heterogeneity of inputs and outputs, it appears legitimate to investigate also the heterogeneity of entrepreneurs. We already mentioned the importance of entrepreneurs for the production process in section 2.3.1.4. An entrepreneur can be defined as someone who turns the production function **formula** into a real **production process**, and not surprisingly some people will achieve better results than others.

The variance of entrepreneurial quality was already incorporated in Austrian macroeconomic models. Evans and Baxendale [45] propose that during an economic boom, monetary expansion attracts less able entrepreneurs through availability of cheap credit to projects of low quality. This in turn results in the expansive cycle that ends up with a bust. Engelhardt [44] demonstrated this mechanism of adverse entrepreneurial selection with U.S. housing market data from the past 25 years.

It is simple logic to look at entrepreneurial quality as another source of efficiency fluctuations. If inputs and outputs were homogeneous, there would not be much room for the entrepreneur to deviate from the best practice. However, considering the differences across all input-output dimensions that we demonstrated in the previous sections, **the effects of heterogeneity in every dimension are multiplicatively compounded**.

2.3.2.5 Summary: The Missing Common Denominator of Measurement

The purpose of this section was to recall the extent of heterogeneity in the real economy. Although heterogeneity is generally acknowledged by economists, for convenience it is often tacitly ignored in economic models. Despite the fact that heterogeneity can turn the results of the model upside down.

The very essence of efficiency is by definition connected with heterogeneity. Not only do differences in inputs and outputs impact efficiency and lead to differentials in relative performance, but they also complicate its measurement and analysis. Thus we can draw two main conclusions from the above:

1. In most cases heterogeneity makes the comparison of physical units with each other almost impossible, because no common denominator of measurement can be well defined. Within a technical production function framework this holds for products and for labour, let alone for capital.

2. Heterogeneity entailed in all dimensions of the production process is among the main drivers of inefficiency fluctuations. Efficiency analysis is a field that has to pay special attention to heterogeneity.

Both conclusions bring us back to the central idea that we mentioned in section 2.2.2. It appears reasonable that, at least in the general base case, economic analysis should move from too much focus on technical production function to focus on money-valued profit measurement.

2.3.3 Efficiency Analysis and Time: Beyond Equilibrium

2.3.3.1 The Concept of Equilibrium

One of the core questions in the history of economics has been distribution theory. The elegant, both a simple and a powerful solution was finally delivered by Jevons and was later dubbed the marginal revolution. This theory explained how the last — or **marginal** — units (marginal utility, marginal product or marginal costs) determine the conditions of exchange on the market. In addition, the marginal theory provides a logical underpinning for market equilibria which necessarily have to arise from marginality conditions.

For the purpose of efficiency analysis however it appears necessary to reconsider two distinct features of the neoclassical equilibrium.

From Static to Dynamic Equilibrium

It was perhaps the endless accusations of market-caused injustice which led economists to stress equilibria as a natural outcome of the markets, as if equilibrium was related to justice in any sense. The equilibrium quickly became the first line of defence against any critique of the prevailing market and social order. Unfortunately this supported the erroneous idea which equates equilibrium and reality. Hence whenever economists encounter a new situation, they tend to adjust the set of equilibrium conditions, so that the result corresponds to the observed outcome.

Much too often the mechanics of adjustment are completely ignored. Most economic models nowadays use sophisticated mathematics with indefinite, continuous variables (such as the letter p for price) instead of definite quantities (price $= 100$). Paradoxically researchers seem to forget that **variables do indeed vary**, and that what we observe is not a static equilibrium situation, but rather a continually changing environment.

In fact, any equilibrium in efficiency analysis is relevant only to the extent that we understand the dynamic process of adjustment and of reaching the equilibrium. Inefficiency measures the deviation from the possibility frontier, i.e. from the hypothetical equilibrium. This equilibrium serves as a benchmark which economic agents strive to approach, and precisely the competition for leadership is of crucial interest for economists. Moreover, the frontier itself is not static but constantly changing, so that economists have to analyse the responses of the agents to such productivity shocks.

A Strictly Positive View of Equilibrium

A further controversy arises when economists adopt the approach of the **normative equilibrium** instead of the positive equilibrium. In the former case the standard equilibrium logic is turned upside down: First the desired outcome which is viewed as correct and right is defined. Then it is attempted to reach this outcome and to describe possible ways of doing so. As Blaug [16] reminds us, John Neville Keynes even distinguished between a positive science, a normative or regulative science and an art. In the words of Keynes:

> The object of a positive science is the establishment of **uniformities**, of a normative science the determination of **ideals**, and of an art the formulation of **precepts**. (Blaug [16, p. 122], emphasis original)

Even though most economists would at least in theory agree that positive and normative analysis should be strictly separated, most would at the same time admit that normative tendencies have become quite powerful so that researchers submit to them almost unconsciously. We see that **policy recommendations** are often a standard feature in many economic research papers. However it would be perhaps more precise and more transparent to call them **political actions**, since this is what they really are. It is our belief that as such they should be confined to **policy papers** of think tanks with defined political agendas, so as not to be mixed with the neutral, positive research.

For sure it is not the purpose of this text to engage in a complex critique of the normative approach in economics. We would merely like to stress that we shall stick to the distinction between positive and normative analysis.

Two Guiding Principles for Equilibrium Analysis

From the above we would like to derive two principles which guide our approach to efficiency analysis. Firstly, **equilibrium is not fixed but it evolves over time.** The process of adjustment towards the equilibrium is regarded as much more important than the equilibrium itself. This is because production takes place in time, equilibrium conditions are constantly changing and so must be the hypothetical steady state.

Secondly, whenever we talk about equilibrium, we refer to **equilibrium as a tool of descriptive analysis**, rather than a normative target to be reached.

2.3.3.2 Production and Time

All production takes place in time. As Rothbard [98, p. 13] remarked,

> there must always be more than one scarce factor of production.

Otherwise the product would miraculously transfer itself from one stage (e.g. the yoghurt in the shop or the yoghurt in the fridge) to the final consumption stage (yoghurt being eaten at a table during breakfast). By this simple contradiction argument it is clear that any production will require at least (1) the good itself from the previous stage of production, (2) time, (3) space, (4) any other cooperating factors of production.

Time is the crucial dimension of all human activities. With respect to production, time implies that entrepreneurs will face changing conditions in the market and will have to adjust their original production plans according to shifts in preferences and technology and according to competitive pressure. **Shifts in preferences** affect the demand for firm's products as well as the supply of factors of production. **Technological progress** may render the production process of the firm or even its product as such obsolete.

The most visible of these dynamic influences on the firm is certainly the **competitive pressure** of other firms. As soon as a product or a service is successful, others will attempt to emulate this success by a similar and usually improved product offering of their own. This will put downward pressure on the firm's demand, prices and ultimately on profitability. Competition is directly embodied in the functioning of the free market economy. While the two former aspects of dynamic change (shifts in preferences and technology) are exogenous shocks that cannot be exactly predicted by the entrepreneur, actions of competitors must be anticipated in advance and must be reflected in the business plan whenever possible.

We can immediately draw two lessons from the above. Firstly, the entrepreneur has to constantly look for signals of change in order to be able to adjust his production as quickly as possible. The reader surely knows from his experience that with the flood of information available in our age, the true entrepreneurial skill with respect to information has changed—it lies more in filtering relevant information rather than purely looking for it. And yet, ever since Hayek's masterpiece **The Use of Knowledge in Society** [60] we know that from the many sources of information one is the superior one: namely **prices**, which have the unique property of conveying a huge amount of underlying knowledge in an extremely condensed datum. Prices, when not manipulated, serve as the major compass for the entrepreneur.

Secondly, introducing the factor of time in production automatically implies variations in efficiency. If a firm simply repeats its production process day by day, it will sooner or later realize that the efficient frontier has been shifting away. Some firms will be chasing the efficient frontier, while others will be those driving the expansion of production possibilities, while still others will be jumping in and out at random places within (or outside) the frontier. This is in a stark contrast to the static view of efficiency, where the efficiency frontier simply exists as a definite goal.

2.3.3.3 Dynamic Efficiency

From the static point of view, being efficient means to reach the most productive point on the frontier which is available at a given point in time. It corresponds to the best exploitation of available means to achieve known ends.

Yet as we have seen the possibility frontier shifts over time. Therefore dynamic efficiency has to embody adjustment to continuous change in economic conditions as described in the previous section. Often this adjustment is regarded as a passive reaction to exogenous shocks. The ability to adequately respond to unexpected situations certainly belongs to important components of efficiency. But the prevalence of passive adjustment would still not be enough to establish a dynamically efficient system.

The primary purpose of all entrepreneurial activities is **the creative discovery of profit opportunities**. This implies that factor combinations and production plans are set up which previously did not occur to other market participants but which are desired by consumers. It is the **active** adjustment which lies at the heart of dynamic efficiency and is superior to the static concept:

> [We] can affirm that the dynamic aspect of efficiency is the most important. Even though an economic system may not have achieved a point on the production possibility frontier, all of its agents may profit if entrepreneurial creativity constantly shifts the curve outward and hence improves everyone's possibilities with a continuous, creative flow of new ends and means which, prior to their entrepreneurial discovery, had yet even to be envisioned. (Huerta de Soto [62, p. 11])

Huerta de Soto explains that previous approaches to dynamic efficiency included some partial factors but never the whole dynamic creative process. Schumpeter, for example, concentrated only on the **creative destruction**, while North focused on **adaptive efficiency** (implicitly passive). Both ignored the actively creative component that Huerta de Soto regards as the fundamental and overarching principle.

We would like to highlight that entrepreneurial creative discovery is not necessarily conditioned by advances in technology. Quite to the contrary, a strong case could be made for the claim that most of the shifts of the possibility frontier are due to new combinations of **existing** factors. Opening a new café or launching a new fashion brand certainly does not require any new technologies. In fact, in many cases the expansion of possibilities may include the luxury of returning from new mass production technologies back to the old ones, as is especially visible in the field of luxurious food items (think hand made pralines). As Foss [49, p. 162] points out:

> innovations have many other sources than the RandD function, and they include process innovations and innovations of management and organisation.[2]

[2] RandD stands for research and development.

It also follows that vintage capital models mentioned in section 2.3.1.2 cannot provide an exhaustive representation of capital accumulation and the resulting economic growth, since vintage (age) of capital is nothing more than a proxy to its real function and usefulness in production.

Here we have to recall again the concept of heterogeneity described in section 2.3.2, and capital heterogeneity in particular. Once we take into account the fact that capital items are functionally different and can be combined in various distinct structures, it directly follows that **alternative capital structures imply alternative production possibilities**. Not only can capital restructuring alter the production frontier, e.g. when we find new use for old machines. If moreover we add to the total stock of capital, we can extend the specialization of some capital items and thus discover new production possibilities. According to Lachmann this change in the composition of capital is the typical source of economic growth because it allows us to escape diminishing returns:

> As capital accumulates there takes place a 'division of capital', a specialization of individual capital items, which enables us to resist the law of diminishing returns. As capital becomes more plentiful its accumulation does not take the form of multiplication of existing items, but that of a change in the composition of capital combinations. (Lachmann [77, p. 79])

This gives us a fresh picture of continuous change in the economy: When entrepreneurs chase new profitable opportunities, they regroup heterogeneous capital and form new combinations, either by using existing resources or by employing newly accumulated capital. This shifts the efficient frontier and it consequently impacts the efficiency of all companies. The competitors have to react as quickly as possible and move towards the frontier, not just following it (as in the static perspective), but trying to become the drivers of the frontier themselves. To assume economic growth is simply to assume that the whole process repeats indefinitely.

2.3.3.4 The Importance of Dynamic Adjustments

While the theoretical concept of dynamic efficiency as the continuous adjustment of the production process is simple, we have to consider its relevance to the real world economy. In our view it appears that the extent of dynamic adjustments is highly significant.

Even though studies on fluctuations are typically oriented towards the macroeconomic data, there is persuasive evidence that firms face considerable competitive pressure on the micro level. A recent study by di Giovanni, Levchenko and Méjean [41] decomposed aggregate output fluctuations into three components: (1) macro shock FM, (2) sector shock FS and (3) firm-specific shock FF. They analysed French firm-level data from tax and customs authorities from the period 1990–2007, containing 2.3 million individual observations. Their result was unambiguous:

$$(FM + FS) : FF = 47 : 53 \sim 50 : 50.$$

Firm-level fluctuations are equally important as macroeconomic and sectoral fluctuations together.

Apparently, the shocks that firms face cannot be simply explained by some general omnipresent trend. In fact, there is much more going on on the micro level than a mere reflection of the common trend. This result is fully in line with the idea of dynamic efficiency and it supports the view that reallocation of **existing** tangible and intangible assets can be equally important as creation of new assets. In one illustration, for example, Foss [49, p. 161] references a study by Foster, Haltiwanger and Krizan [51] who estimate that

> competitive dynamics through **reallocation of productive assets** account for about 50 percent of aggregate productivity growth. (Foss [49, p. 161, emphasis added])

Many authors also observed that the degree of variability does not decline even if we take extremely narrow parts of the data. It was noted that whatever the level of detail we look at, the underlying micro level fluctuations remain the same. It appears that the behaviour of microeconomic agents with respect to variability shows some patterns corresponding to fractals. The conclusion reached by Griliches and Mairesse [56] illustrates this beautifully:

> We also thought that one could reduce aggregation biases by reducing the heterogeneity as one goes down from some general mixtures such as 'total manufacturing' to something more coherent, such as 'petroleum refining' or the 'manufacture of cement'. But something like Mandelbrot's fractals phenomenon seems to be at work here also: the observed variability-heterogeneity does not really decline as we cut our data finer and finer. There is a sense in which different bakeries are just as much different from each other, as the steel industry is from the machinery industry. (Griliches and Mairesse [56, p. 23])

Our own empirical studies presented below suggest that the level of efficiency fluctuations within the economy is rather large.

2.3.3.5 Summary: Economy as a Living System

If we introduce time into the dynamic analysis in its narrow sense, it can simply mean either an infinite repetition of the same process, or convergence to the final steady state. Taking the latter concept, efficiency is then viewed as approaching the fixed or given frontier. Inefficiency is a **mild fluctuation**. When Aigner, Lovell and Schmidt [3] defined the stochastic production function, they attempted to distinguish the two-directional statistical noise from the one-directional inefficiency factor, where inefficiency was viewed as a slight vibration of the frontier.

On the other hand, the concept that we advanced in this section takes into account the degree of complexity inherent in the fully dynamic production process. The time dimension is coupled with the actively creative component (as opposed to passive chasing of the frontier), which turns inefficiency into a **wild fluctuation** (Taleb [109]) and **systematic inefficiency** (Reifschneider and Stevenson [97]). We would like to contrast the image of a vibration of the frontier, as is commonly assumed in the static or narrowly dynamic efficiency analysis, with the image of **the production space as a beehive**.

Once we understand the principle of continuous creative adjustment and (re)-inventions of production plans, we also have to accept inefficiency as a prevalent, natural consequence of these dynamic processes.

2.3.4 Summary: The Case for Inefficiency

Typically there is not much room for inefficiency in any economic analysis that is based on equilibrium. However, as we explained in the preceding sections, equilibrium has to be viewed as a theoretical, hypothetical concept that is rarely reached in the real world. If we want to understand the complexity of production processes, we have to focus on the dynamic adjustment that causes frequent departures from the efficient frontier. Accordingly we prefer to use the term **inefficiency** rather than efficiency.

We described in detail two factors that impact efficiency fluctuations and their measurement:

1. **Heterogeneity** or specificity of inputs and outputs. It is one thing to operate existing resources at full (efficient) capacity, but it is quite another issue to select the suitable labour, capital and products in the first place. Differences within these categories might be subtle but still economically highly significant, so that observed efficiency fluctuates wildly. Furthermore, we pointed out that in practice it can be very difficult to find units which are mutually comparable, precisely because no common denominator (other than money) exists. This implies one more obstacle for researchers.

2. **Time** which captures the dynamic, ever changing nature of production. Due to creative adjustments inefficiencies arise in the form of wild fluctuations and the production space appears as a beehive. Inefficiency is inherent in dynamic functioning of the economy and has to be incorporated in the analysis of production as a standard feature.

Heterogeneity and time imply that a pure technical production function defined across sectors will be spurious. There will always be an economic component in what is attempted to be measured as 'technical efficiency', simply because in most cases we cannot distinguish if lower or slower output was due to higher quality requirements or due to plain underperformance. **The only way to fully reflect the heterogeneity and time dimensions of production is to measure economic efficiency in prices.**

Static efficiency analysis with exogenous prices cannot and does not take heterogeneity and time fully into account. Nonetheless, we believe to have persuaded the reader of their importance in modern production processes, of which they are essential components. In the next section we propose a simple generalization of efficiency frontier which can at least partly accommodate these features.

2.4 Implications for Efficiency Analysis

In the previous section we advanced theoretical approaches to production that significantly expand the dimensions of the static homogeneous structure which is common in simple efficiency models. In what follows we proceed to specify a model that we believe is more suitable to capture complex production processes. We proposed this framework in Průša [93] as presented in chapter 3. Here we sketch the principle in order to link it to theoretical considerations in section 2.3.

2.4.1 Production Accounting

2.4.1.1 Production Technology

We maintain the familiar notation: The input vector $\boldsymbol{x}$ has a corresponding price vector $\boldsymbol{w}$, and for simplicity we consider only one output y with price p.

The set of all feasible pairs $(\boldsymbol{x}, y)$ is the space of production possibilities $\mathscr{Y}$. We define the **production frontier** $\mathrm{Eff}(\mathscr{Y})$ as the subset of input-output pairs where:

- the production process achieves the maximum possible output for any given input vector, or conversely
- the production process achieves the minimum possible input vector for any given output.[3]

As long as a valid functional form exists, we can write the production function $y = f(\boldsymbol{x})$. In general this mathematical formula captures all technological knowledge on transformation of available resources into demanded consumer goods.

2.4.1.2 Production and Exchange

The closer a firm is to the production frontier, the better will be its performance. There is not much more that the economist has to say about technical efficiency.

From economic perspective the technology frontier becomes interesting when we connect it with prices which emerge from market exchange of products. In the simplest case prices are exogenous for individual producers, who then select inputs and output in order to maximize profits. Formally we can write the profit function:

$$\Pi(p, \boldsymbol{w}) = \arg\max_{\{\boldsymbol{x}, y\}} \{py - \boldsymbol{w}'\boldsymbol{x} | \ (\boldsymbol{x}, y) \in \mathrm{Eff}(\mathscr{Y})\},$$

[3]For mathematical formulation see section 3.2.1.1.

which determines the optimal amount of inputs and output given any possible market prices.

It would indeed not require sophisticated judgement for the entrepreneur if he could, knowing the admissible prices, just choose the optimal input-output vector. Such a model is theoretically instructive, however it is bound to fail when applied to real world data. We argue in this thesis (sections 2.2.3 and 3.2.2.1) at length why: Experience tells us that (1) firms operate in highly differentiated, asymmetric markets; (2) they face downward sloping demand curves; (3) both their demand (consumer preferences) and supply (competition and technology) conditions change quickly and frequently.

2.4.2 Functional Specification

Before we turn to non-parametric estimation of efficiency, for reference we would like to briefly mention parametric specification.

Parametric estimation of production functions is called **stochastic frontier analysis** (SFA) in the literature and is extensively covered e.g. by Kumbhakar & Lovell [72]. For mathematical simplicity the Cobb-Douglas production function is the most popular formulation. The debate on adequacy of other specifications re-emerged especially in the context of macroeconomic growth models, as is shown in the papers by Duffy & Papageorgiou [43] for the constant elasticity of substitution function, and by Kneller & Stevens [70] for the translog specification. The reader is referred to these and other studies for further discussion.

Given that there always exists a certain amount of rivalry between parametric and non-parametric approaches, it is interesting to note that recent developments include efforts towards the integration of both methodologies, following the comparative parametric/non-parametric studies such as Bardhan *et al.* [7]. Most notably, Kuosmanen [74] proposed a method called **stochastic non-parametric envelopment of data**, combining components of non-parametric data envelopment analysis (DEA) and parametric SFA. This method was applied to electricity distribution networks in Kuosmanen [75]. Other approaches include Tsionas [112] using Bayesian statistics, and Kumbhakar *et al.* [73] using local maximum likelihood estimation.

Much more important for our thesis is the fact that besides SFA, the parametric research developed detailed models of production addressing endogeneity of prices and other related issues. Systems of equations were proposed by Marschak & Andrews [85, see eqs. 1.29-1.31] and since then have grown to complex demand–supply models for differentiated products as in Berry *et al.* [12]. There is a significant body of literature devoted to highly detailed micro-level modelling of heterogeneous products. Other notable studies which use parametric models include Klette & Griliches [69], Melitz [87], Levinsohn & Petrin [81] and De Loecker [40]. It is the more striking that non-parametric efficiency studies did not fully reflect these advances in parametric models.

Once major heterogeneity is present, detailed modelling making ever finer cuts through the data is one possible way of analysis. Our approach goes in the opposite direction, relying on the most general money-valued measurement for comparison of arbitrarily large units.

2.4.3 Money-metric Production Frontiers

2.4.3.1 Formulation of the Money-metric Principle

If we recognize all difficulties connected to measurement of genuinely heterogeneous inputs and outputs, it is a simple proposition to measure production in monetary units. After all this principle is applied in business accounting and is ultimately the only criterion of success of any particular enterprise.

Let us define the input-output space in terms of vectors $(\boldsymbol{wx}, py)$ where each input x and output y is multiplied by its respective price. Those combinations of $\boldsymbol{wx}$ and py that are feasible in the market form the **money-metric production set** $\mathscr{M}$ defined in equation (3.3). In plain words, every observation represents a possible cost and revenue situation of a firm, so that we might also speak about the **cost-revenue opportunity set**.

In analogy to the production frontier, we can define the frontier of this cost-revenue opportunity set as in equation (3.4):

$$\begin{aligned}\mathrm{Eff}(\mathscr{M}) \quad = \quad & \{(\boldsymbol{wx}, \boldsymbol{py}) \in \mathscr{M} \,| \\ & \forall [\boldsymbol{wx}^1 \leq \boldsymbol{wx},\ \boldsymbol{py}^1 \geq \boldsymbol{py},\ (\boldsymbol{wx}^1, \boldsymbol{py}^1) \neq (\boldsymbol{wx}, \boldsymbol{py})] : \\ & (\boldsymbol{wx}^1, \boldsymbol{py}^1) \notin \mathscr{M}\}.\end{aligned}$$

Once we establish this frontier, mathematical principles of efficiency measurement can be simply applied to $\mathscr{M}$.

It is important to highlight the crucial difference between the production possibility set $\mathscr{Y}$ and the revenue-cost opportunity set $\mathscr{M}$. The former is fully determined by technology, it entails the complete technical know-how and can be therefore regarded as the **engineer's knowledge**. As such it is only relevant to economics to the extent that it has to be taken into account as a constraint of economic decisions.

On the contrary, the money-metric production set is driven by economic interaction of producers and consumers. Technological limitations are in the background (they are given) and the focus is instead on behavioural aspects of production — on entrepreneurship. $\mathscr{M}$ can be thus viewed as the **entrepreneur's ability**.

This framework is in fact very similar to the approach used in consumer choice theory to construct the money-metric utility functions. Utility is not interpersonally comparable, but the consumer can express his utility in monetary terms: at given prices, how much money does he need to reach the utility level corresponding to a certain bundle of consumption goods. We can proceed in the same way in production analysis: Even though capital, labour or products are not perfectly comparable, we can still label those in terms of costs and revenues to see the level of efficiency applicable to all firms.

2.4.3.2 Properties of the Money-metric Frontier

The frontier Eff($\mathscr{M}$) is by definition an approximation of the profit function. The fundamental principle of the frontier is that **higher profits are equivalent to higher economic efficiency**. The company can become more efficient either by increasing revenue or by decreasing costs. As such the frontier appears highly plausible and natural especially to business accountants.

The main advantage of this approach lies not only in its intuitive appeal, but in its versatility. Money serves as the numeraire which allows us to compare completely different production plans, e.g. the production of bottled mineral water and the production of flavoured carbonated drinks. Both these business lines are close enough so that their comparison is of interest from the business point of view. On the other hand, the respective technologies are sufficiently different to create significant problems when measuring the technical production function and comparing technical efficiency. Money-metric production frontiers facilitate aggregation and are therefore universally suitable for large cross-sections as well as time series. We discuss the methodology of pooling observations in detail in section 3.4.2.3.

We pointed out above that the money-metric frontier directly incorporates the bargaining abilities of the entrepreneur. This implies that any observed point in the cost-revenue set has to be regarded as unique in the sense that the firm might not be able to achieve it again. Quite to the contrary, as long as demand and supply conditions are changing, it is more probable that the result in the next period will be different from the previous one. This also implies that we cannot assume $\mathscr{M}$ to be convex and have to assume a free disposal hull frontier instead (refer to figure 2.1). The convexity of the technical production possibility set is justified by replication. For reasons explained here, we cannot automatically presuppose replication in monetary terms: When a firm moves from a certain cost-revenue position, many business parameters change immediately, so that an equivalent change in profitability is possible but not guaranteed.

The money-metric approach has an obvious drawback: In the cost-revenue opportunity set we cannot distinguish decisions on prices from decisions on quantities. Such a distinction might be desirable in certain business contexts, but the analysis would have to be limited to a very narrow industrial sub-segment. The money-metric approach is not structural. This contrasts with managerial sciences, where the focus is a single firm which is compared to very few of its closest competitors. Nonetheless, as we argued above, in most cases the degree of heterogeneity will be too high to permit meaningful comparisons in physical units. Then we can always turn to the money-metric framework.

2.4.4 Summary: The General Framework

Theories of efficiency measurement are usually built on the assumption of separated technical and allocative efficiency. This is very useful if we want to analyse detailed theoretical structures of production of one firm. For empirical measurement however this assumption is too restrictive. Researchers often overcome this issue implicitly by mixing purely technical dimensions of production with economic (money-valued) ones, which immediately raises the question of the economic substance of such measurements.

We propose the money-metric approach which can easily account for heterogeneity in production processes and for their dynamic, ever-changing nature. The money-metric frontier is not to be seen as a replacement of the standard technical-allocative efficiency decomposition. We intend $\mathrm{Eff}(\mathscr{M})$ to provide an alternative, theoretically consistent modelling set-up for empirical efficiency measurement. To the extent that it is approximates the profit function, it is highly intuitive and widely applicable. Unlike detailed structural models, it is simple and general, suited for looking at the big picture.

2.5 Selected Empirical Evidence

The essays presented in chapters 3 to 5 show empirical efficiency analysis of small and medium enterprises in the United Kingdom and in the Czech Republic. The methods as well as data used in each of the essays are different, which allows us to draw more general lessons from their comparison. Here we would like to summarize key results that we obtained in these studies.

2.5.1 Variance of Efficiency

Regardless of what angle we take when looking at the data, and regardless of what advanced smoothing or robust methodology we use, the variance of efficiency scores is tremendous. In all three studies mean efficiency scores range from 25 to 75%, confirming the evidence from section 2.3.3.4. This observation revealed to us the importance of heterogeneity and its omnipresence in market competition.

In chapter 4 we dubbed this phenomenon 'the clear distinction between leaders and stragglers' (see proposition 4.3.3). In section 2.3.3.5 we pictured the production space as the beehive. The idea is still the same: The variance of economic performance is immense and the analysis of production should not only describe an unattainable equilibrium, but also the continuous adjustment process.

2.5.2 Returns to Scale

In each of the essays we grouped the observations according to the number of employees into size clusters. We wanted to see if firm size significantly influenced firm efficiency. Statistical results are contradictory. While the non-parametric estimation for Czech data does not show a clear trend (table 4.5), the parametric estimation found a significant positive size effect of 7% (section 5.4.3.3). Perhaps the clearest picture was given in the British study, which revealed a statistically significant, negative but negligible size effect of merely 0.85% per 100 employees (section 3.5.3.2).

One conclusion is therefore clear: If a size effect is present at all, its magnitude is certainly much lower than the magnitude of overall dispersion of efficiency scores. The evidence that we gathered suggests that there does not exist a general size–efficiency relationship that would be present throughout the economy. Apparently size matters in some business situations, and it does not matter in others.

2.5.3 Time Effects

Given that our data span over several years, we could additionally investigate the changes of economic efficiency in time. It is crucial to remember that the money-metric frontier in fact approximates profitability. If we look at the overall rate of profit in the economy, it should remain relatively stable as it represents the distribution of income between labour and capital.

This is what we confirmed for British enterprises: There we could not find a statistically significant impact of time on economic efficiency scores. Thus we claimed that during the years 1998-2007 the income from production was distributed evenly among capitalists and workers. In the case of Czech enterprises, we found a negative time effect of 15% over the period 2002-2005 (section 5.4.3.4), which appears surprising but might possibly be attributed to the specifics of economic transition from a centrally planned to free market economy.

2.5.4 Unit Wages and Number of Employees

Let us mention the specific example of the Czech production function where we separated prices and quantities for the labour input. In this case we departed from the purely money-metric approach. We measured the effect of higher wages in contrast to higher number of employees and how these two parameters jointly affected profitability.

Our results were straightforward (see section 5.4.2.3):

> We find that high average wage has positive influence on value added, while large number of employees impacts production adversely.

To the extent that higher wages reflect higher quality of the workforce, we can deduce from the quoted result that Czech enterprises optimize their value added (and profitability) by choosing a lower number of high skilled workers (to whom they pay higher wages) rather than employing more low skilled workers. This brings us back to section 2.3.2.2 where we elaborated in depth on heterogeneity of labour and the relationship between wages and quality. Our result confirms precisely this effect of worker selection based on price discounted quality characteristics.

2.6 Conclusions

In this essay we argued that variability of economic performance — profitability — is a natural feature of all production processes. For any firm inefficiency will be the rule, and only a few best performers will ever reach the frontier.

We discussed in detail two main factors behind such variance. Firstly, all production happens in time, so that the firm has to face changing supply and demand conditions. What appears optimal now might not work in the next period and the firm has to constantly adjusts its production plans. Secondly, all dimensions of production are deeply heterogeneous. This concerns not only inputs and outputs but also entrepreneurs themselves. Both these factors jointly imply that making the right decisions in changing conditions requires significant effort and poses huge profit opportunities as well as risk of losses. In the end the entrepreneur might find himself in a completely different situation than he had intended at the beginning — for better or worse.

We pointed out that once the importance of these factors is recognized, the production space can be seen as a lively beehive: Some firms pushing the frontier outward, some falling back and still others moving in between. We proposed that the simplest and most intuitive way of capturing this creatively dynamic process is to utilize the money-metric production possibility set. This model in fact approximates the profit function in the cost-revenue space, so that it can directly overcome the obstacles of heterogeneity in physical units. If we couple this model with non-parametric estimation, it provides the most general way of measuring economic efficiency.

This is not to say that the proposed model is superior to the traditional technical–allocative efficiency decomposition which should consequently be abandoned. The latter is useful both as a theoretical concept and as a tool for detailed model building. Instead we offer the money-metric perspective as a consistent estimation approach in cases where researchers might be lost due to inherent problems in unit measurements, simply because physical measurement is impossible (as is the case for capital) or because the observed categories are too broad (as might be the case for workers or products). We believe this will contribute to further advances in understanding production and profitability of enterprises in the globalized world.

Figure 2.1: Efficiency decomposition and Free Disposal Hull frontier.

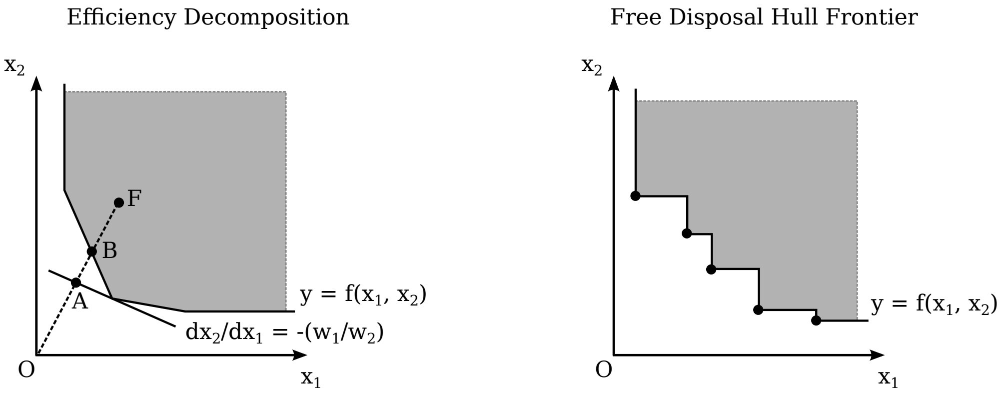


Figure 2.2: An illustrative transformation of budget constraint from $\mathcal{G}$-space to $\mathcal{C}$-space

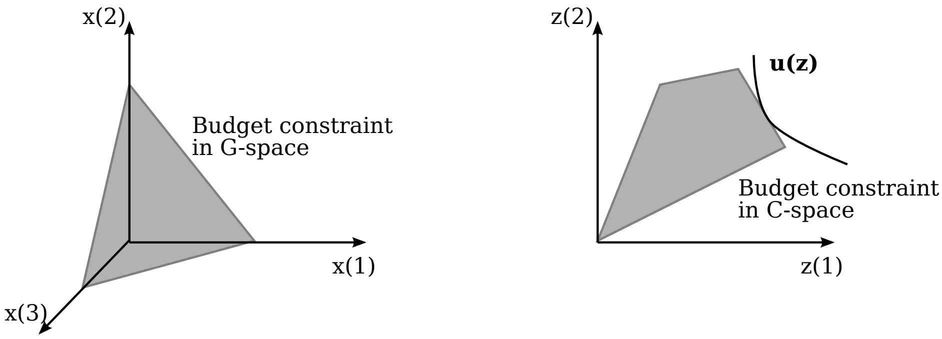


Table 2.1: Production of sample PV plants in 2011.

	Legal form	Production capacity	Annual production	Capacity usage	
		MW_e	*MWh net*	*MWh/MW_e*	%
3L Invest	a.s.	38.3	39,962	1,043	11.91%
FVE Czech Novum	s.r.o.	35.1	40,383	1,151	13.13%
Gentley	a.s.	29.9	32,533	1,088	12.42%
AREA-GROUP CL	a.s.	17.5	17,629	1,007	11.50%
DOMICA FPI	s.r.o.	16.0	18,365	1,148	13.10%
ŽV - SUN	s.r.o.	13.0	13,051	1,004	11.46%
Divalia	a.s.	10.2	12,185	1,195	13.64%
All plants in 2011		121.7	134,146	1,099	12.54%

Note: Hours computed as MWh/MW_e.

Source: Průša, Klimešová, Janda [96, table 1].

PART II

EMPIRICAL CASE STUDIES ON ECONOMIC EFFICIENCY

Chapter 3

The Money-metric Production Frontier with an Application to British SME

This article was published as Průša, J. A General Framework to Evaluate Economic Efficiency with an Application to British SME. IES Working Paper 25/2009. Institute of Economic Studies, Charles University in Prague. [93].

Abstract

This article formalises the idea of **money-metric production frontiers**, which we propose as a general framework for nonparametric evaluation of **economic** efficiency. As we show in our methodological discussion, this improves the flexibility and economic interpretation of our model.

The empirical part is the first attempt to test the existence of a size-efficiency relationship among small businesses in the United Kingdom. It is based on a unique panel both with respect to size — ranging from agriculture to services — and to the ten year time span. We employ statistically robust methods to estimate and analyse sectoral efficiency.

Our analysis yields three main insights: (1) Average sectors are expected to be two to four times less efficient than those on the efficient frontier. Great dispersion of efficiency scores highlights the importance of dynamic out-of-equilibrium modelling. (2) There is no evidence of a general economy-wide size-efficiency relationship. (3) Economic efficiency remained constant over the decade 1998–2007.

Keywords: small and medium enterprises; economic efficiency; firm size; robust efficiency estimation.

JEL classification: D24, L25, L26.

3.1 Introduction

Research on small and medium enterprises (henceforth SME) has recently received much attention. The question that has been examined most intensively can be posed with Yang & Chen [118]: Are small firms less efficient? The authors list nine studies (table 1 *ibid*) which find a positive size-efficiency relationship, although their own results for Taiwan's electronics industry are heterogeneous. Research has also pointed to dynamic aspects of SME efficiency. Recent studies on efficiency dynamics include Taymaz [110] who analyses Turkish manufacturing industries. He confirms that higher efficiency implies higher probability of survival.

Our case study of the United Kingdom contributes to research on efficiency of SME. We extend previous studies by two main features.

Firstly, numerous articles on the efficiency-size relationship were motivated by technical efficiency and returns to scale (e. g. Alvarez & Crespi [5])and derived their models from static microeconomic framework. The shortcomings of these simplifications are well understood and have been extensively covered in parametric applications. Yet not much discussion has been devoted to overcoming them in nonparametric estimation. Therefore we focus on **economic** efficiency and propose a more general solution which we contend is more suitable to evaluate economic efficiency.

Secondly, we use a large dataset based on firm-level survey and compare most of the sectors in the economy, which allows us to test whether previous results were sector-specific or whether they extend to the whole economy. We analyze efficiency scores and test if they are size or time dependent.

The most important institution conducting research specifically aimed at SME in the United Kingdom is the Centre for Business Research at the University of Cambrigde.[1] Most of the recent papers are concerned with institutional and structural issues, such as financing (Hughes [63], Cosh *et al.* [34]), innovation (overview by Hoffman *et al.* [61], Cosh *et al.* [35]), or subcontracting (Wynarczyk & Watson [117]). To the best of our knowledge, no recent article examined SME efficiency. This offers room for our analysis.

The rest of the paper is organised as follows: In section 3.2 we first review the familiar framework for microeconomic modelling of production, and then discuss several issues related to identification of variables in the model. In section 3.3 we define the money-metric production frontier that we use for measurement of economic efficiency. We also provide a detailed discussion on its properties. Finally in sections 3.4 and 3.5 we employ this model to evaluate economic efficiency of British SME: We first describe the computation procedures and then present the results.

[1] [http://www.cbr.cam.ac.uk/].

3.2 Empirical Methodology of Efficiency Measurement

3.2.1 Modelling Production

Let us first recall the basic concepts for microeconomic analysis of production.

3.2.1.1 Technology

The production set is defined as all feasible input-output vectors $(\boldsymbol{x},\boldsymbol{y})$, as in Tulkens & Eeckaut [113]:

$$\mathscr{Y} = \{(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{x} \in \Re^{r}_{0,+}, \boldsymbol{y} \in \Re^{s}_{0,+} |\ (\boldsymbol{x},\boldsymbol{y}) \text{ is feasible}\}.^{2}$$

The points that are technically efficient are given by:

$$\text{Eff}(\mathscr{Y}) = \{(\boldsymbol{x},\boldsymbol{y}) \in \mathscr{Y} |\ \forall[\boldsymbol{x}^1 \leq \boldsymbol{x},\ \boldsymbol{y}^1 \geq \boldsymbol{y},\ (\boldsymbol{x}^1,\boldsymbol{y}^1) \neq (\boldsymbol{x},\boldsymbol{y})] :\ (\boldsymbol{x}^1,\boldsymbol{y}^1) \notin \mathscr{Y}\}. \tag{3.1}$$

$\text{Eff}(\mathscr{Y})$ is known as the production frontier. If we can find a functional form, we have:

$$(\boldsymbol{x},\boldsymbol{y}) \in \text{Eff}(\mathscr{Y}) \iff \mathscr{T}(\boldsymbol{x},\boldsymbol{y}) = 0,$$

where $\mathscr{T}(\cdot)$ is the transformation function. For scalar output, this simplifies to the production function $f(\boldsymbol{x}) = y$. Models based on specific functional forms are called parametric, while our focus will be on **nonparametric** models derived from equation (3.1).

3.2.1.2 Simple Cost and Profit Functions

In order to arrive at economic efficiency we need to add prices to the production function. This can be done very easily in parametric models once we assume that input prices $\boldsymbol{w}$ and output prices $\boldsymbol{p}$ are fixed. Refining the definition by Greene [55, p. 142], we can write a cost function for scalar output y as:

$$\mathscr{C}(\boldsymbol{w},y) = \arg\min_{\{\boldsymbol{x}\}}\{\boldsymbol{w}'\boldsymbol{x}|\ f(\boldsymbol{x}) \geq y\}.$$

More generally, optimization over both inputs and outputs yields the profit function:

$$\Pi(\boldsymbol{p},\boldsymbol{w}) = \arg\max_{\{\boldsymbol{x},\boldsymbol{y}\}}\{\boldsymbol{p}'\boldsymbol{y} - \boldsymbol{w}'\boldsymbol{x}|\ (\boldsymbol{x},\boldsymbol{y}) \in \mathscr{Y}\},$$

which is by a contradiction argument equivalent to:

$$\Pi(\boldsymbol{p},\boldsymbol{w}) = \arg\max_{\{\boldsymbol{x},\boldsymbol{y}\}}\{\boldsymbol{p}'\boldsymbol{y} - \boldsymbol{w}'\boldsymbol{x}|\ (\boldsymbol{x},\boldsymbol{y}) \in \text{Eff}(\mathscr{Y})\}.$$

[2] We explicitly include 0 in the notation to indicate that vectors $\boldsymbol{x},\boldsymbol{y}$ are nonnegative.

3.2.1.3 Efficiency Decomposition

Along the lines of the definitions above, efficiency can be decomposed into technical and allocative component. Technical efficiency is intuitively straightforward and means operating on the frontier $\mathrm{Eff}(\mathscr{Y})$, as opposed to points within the production set $\mathscr{Y}$ (refer to 3.1). The only issue is to specify whether the feasibility of $\mathscr{Y}$ is with respect to a specific firm or whether we consider aggregate technology.

Being restricted to the most productive points of $\mathscr{Y}$, a firm achieves allocative efficiency if and only if it chooses the most profitable point of $\mathrm{Eff}(\mathscr{Y})$. This corresponds precisely to the profit function defined in the preceding section.

However, mathematical representation of this optimal choice by cost or profit functions requires strong structural assumptions. As indicated by the arguments of $\mathscr{C}(\cdot)$ or $\Pi(\cdot)$, these functions are derived for exogenous prices, but research suggests that this exogeneity is rare in practice. For example, Fabiani *et al.* [46] present evidence on pricing behaviour of $11,000$ firms. The result relevant for our discussion is that 54% of firms use markup pricing, while only 27% of firms use competitors' price as the main price setting factor (section 3.1 *ibid*). To arrive at a valid framework for empirical application, we need to examine the structural assumptions in more depth.

3.2.2 Empirical Variables for Models of Production

3.2.2.1 Exogeneity of Prices

Although it is possible in theory to separate technical and allocative efficiency, identification of these components requires detailed data on quantities and prices. Such level of detail is usually unavailable, and even if it were, exogeneity of prices would still be questionable.

Suppose that a researcher has detailed data on both quantities and prices. Under standard regularity conditions, cost or profit functions are sufficient statistics for technology. Yet this holds only for fixed prices, that is when firms treat prices as exogenous (given in the market) for their economic decisions. In fact, the same conclusion applies to aggregation. Mas-Colell *et al.* [86, p. 149] note:

> If firms maximise profits taking prices as given, then the production side of the economy aggregates beautifully.

Estimation of $\mathscr{C}(\cdot)$ or $\Pi(\cdot)$ implicitly relies on this structural assumption about price setting. However, when analysing large cross-sections of firms, exogeneity assumptions are likely to be too restrictive: Not only because we are uncertain about how prices are actually formed, but especially because the structure of price setting will certainly differ across firms, sectors and markets.

Indeed, if the situation was symmetric, we could just work with a representative firm. Yet it is precisely the **asymmetry** assumption that justifies efficiency analysis; that is, asymmetry for which economists developed a variety of explanations, ranging from transaction costs to entrepreneurship. This asymmetry implies that we should attempt to incorporate prices endogenously in our model whenever possible.

3.2.2.2 Specification of Variables

The starting point of any analysis of efficiency is the selection of inputs and outputs. Because technologies are simply too complex, economics came up with the concept of basic factors of production: capital, labour, materials, energy, and land (which is commonly omitted). This concept is widely accepted, not least because it simplifies aggregation. We take the most general form from Burnside [23, equation 2.1]:

$$y = f(\text{capital, labour, energy, materials, technology}). \tag{3.2}$$

Burnside provides a thorough discussion of assumptions which underlie specific choices of variables or pricing structures. His treatment provides a link between micro and macro level production so that the equation represents output at the industry level. The specification in (3.2) is however commonly used for firm-level studies, one example being Yasar *et al.* [119], even though they call technology as *total factor productivity*.

Researchers usually take sales or value added (possibly in logs) for output, depending on whether materials and energy have been subtracted. Other measures which are applied to assess performance of companies are surveyed by Murphy *et al.* [91, table 2]. As regards inputs, researchers employ tangible and intangible assets for capital; and a combination of employees, hours worked and wages for labour. Dynamic models include investment in the form of acquisition of assets and depreciation. Technology is treated by separate models, and these go beyond the scope of our article.

Due to their aggregate nature, general factors of production induce a measurement problem, which is especially apparent for capital. Capital is supposed to represent machinery, but since it is a term too broad, any measure of capital suitable for comparison must be monetary.

This created a considerable amount of confusion. Studies on production often combine data in physical and monetary units without proper discussion. In his study on production functions, Johansen [64, chapter 9] analyses output of Norwegian tankers. While he measures output as tonne-miles per day, the inputs — fuel and labour per day — are measured in Norwegian **kroner**. This approach could be justified, say if output of all tankers was traded at the same price. Whether this explanation would be reasonable or not we leave aside. More surprising is that the author does not attempt at all to explain this specification.

More recently Biørn *et al.* [15] specify their micro-based **production** function as follows: Output in tonnes, and inputs as capital and materials in Norwegian **kroner**, labour in man-hours and energy in kWh (appendix B9.2 *ibid*). Their justification is rather anecdotal. The authors prefer tonnes to **kroner** for output because of possible mismeasurement in sales. On the other hand, they do not mind using arbitrary constant depreciation rates for capital and transforming fuels to kWh using "estimated average energy content" (*ibid*).

These examples reveal that empirical studies have not made a clear distinction between technical and allocative efficiency. In our view this results in dubious interpretation. We propose a solution to these inconsistencies in section 3.3.

3.3 Money-metric Production Frontiers

3.3.1 Definition

In previous sections we noted that in empirical literature on efficiency it is the exception rather than the rule to discuss the step from Eff$(\mathscr{Y})$ to $\Pi(\cdot)$ and the underlying assumptions. This holds especially for the price exogeneity assumption, which affects the whole model building procedure, and it also applies to combining data on quantities and prices in a single equation.

Our treatment rests on a more general approach that addresses some of the issues explained in the preceding exposition. Let us define the **money-metric production possibility set** $\mathscr{M}$ in the r-input, s-output space:

$$\begin{aligned} \mathscr{M} &= \{(\boldsymbol{wx}, \boldsymbol{py}), \boldsymbol{wx} \in \Re^r_{0,+}, \boldsymbol{py} \in \Re^s_{0,+} \mid (\boldsymbol{wx}, \boldsymbol{py}) \text{ is feasible}\}, \qquad (3.3) \\ \boldsymbol{wx} &= (w_1x_1, \ldots, w_rx_r)', \\ \boldsymbol{py} &= (p_1y_1, \ldots, p_sy_s)'. \end{aligned}$$

By analogy to the production frontier Eff$(\mathscr{Y})$ in (3.1) we can define the **money-metric production frontier** Eff$(\mathscr{M})$:

$$\begin{aligned} \text{Eff}(\mathscr{M}) = \{&(\boldsymbol{wx}, \boldsymbol{py}) \in \mathscr{M} \mid \\ &\forall [\boldsymbol{wx}^1 \le \boldsymbol{wx},\ \boldsymbol{py}^1 \ge \boldsymbol{py},\ (\boldsymbol{wx}^1, \boldsymbol{py}^1) \neq (\boldsymbol{wx}, \boldsymbol{py})] : \\ &(\boldsymbol{wx}^1, \boldsymbol{py}^1) \notin \mathscr{M}\}. \end{aligned} \qquad (3.4)$$

The notation indicates that firms participated in some form of bargaining, so that inputs and outputs are money-valued. Yet no explicit structure is placed on the bargaining process, so that the set $\mathscr{M}$ is a more general and flexible basis for efficiency measurement.

3.3.2 Derivation and Properties

3.3.2.1 A Plain Vanilla Efficiency Score

Let us recall that the intuitive idea of an efficiency score θ_i of firm i can be summarized as:

$$\theta(\boldsymbol{x}_i, \boldsymbol{y}_i) \propto \frac{\boldsymbol{v}_i'\boldsymbol{y}_i}{\boldsymbol{u}_i'\boldsymbol{x}_i},$$

where **u** and **v** are vectors of weights to be specified, as in equation 2.13 in Cooper *et al.* [32]. The weights u_r and v_s indicate how much individual inputs and outputs contribute to the total **technical** efficiency score $\theta(\boldsymbol{x}_i, \boldsymbol{y}_i)$.

In the simplest data envelopment specification, optimal weights $\boldsymbol{v}^*$ and $\boldsymbol{u}^*$ can be found by solving a linear programming problem. The optimization criterion to select the weights is to maximize $\theta^*(\boldsymbol{x}_i, \boldsymbol{y}_i)$ (see Cooper *et al.* [32, section 2.5]).

3.3.2.2 Money-metric Efficiency Scores

A money-metric efficiency score can be interpreted along the same lines:

$$\theta(\boldsymbol{wx}_i, \boldsymbol{py}_i) \propto \frac{\boldsymbol{v}_i'(\boldsymbol{py})_i}{\boldsymbol{u}_i'(\boldsymbol{wx})_i},$$

where weights **u** and **v** indicate how much each individual cost and revenue component contributes to the total **economic** efficiency score $\theta(\boldsymbol{wx}_i, \boldsymbol{py}_i)$.

This adds significant flexibility to the efficiency score especially with regard to heterogeneity across observations. Consider the following example[3]: Two firms i and j use the same amount of a single input $x_i = x_j$ and produce different outputs. For simplicity we assume $s = 2$ and $y_{2,i} = y_{1,j} = 0$. Because their output sets are disjoint and the relationship $y_{1,i} \lesseqgtr y_{2,j}$ for the nonzero components of output vectors is not defined, the efficiency comparison is meaningless and they are both ranked as efficient because both observations lie on the efficient frontier $\mathrm{Eff}(\mathscr{Y})$.

Now suppose you also observe monetary outputs (revenues) py_i and py_j. This effectively reduces the dimension of the output set to $s = 1$ and makes the observations perfectly comparable. Hence the money-metric formulation overcomes the problem of disjoint technical production possibility sets. Theoretically we can define multidimensional monetary output, so that the problem of disjoint sets might arise for the money-metric formulation as well. The responsibility of meaningfully specifying dimensions of $\mathscr{M}$ lies with the researcher: Most components of input-output vectors should be nonzero for most observations.

3.3.2.3 What Determines Efficiency

We saw in the preceding paragraph that money-metric production frontiers can be viewed as a nonparametric approximation of the profit function. While models based on equation (3.1) capture only technical efficiency, the money-metric approach in equation (3.4) encompasses also the allocative component and captures **economic** efficiency. It immediately follows that a firm has essentially two ways of improving its efficiency score: Either increasing revenues, or reducing costs.

In a model based on $\mathscr{M}$, it makes no difference whether the adjustment involves price or quantity components of input and output vectors because both variables are endogenous and inseparable. In practice we can observe that while increasing revenue is typical during periods of economic expansion, cost reduction is sought especially in times of crisis or recession. Adjustments in both directions are equally effective in terms of economic efficiency.

[3]We are grateful to an anonymous reviewer for pointing out to us this example.

We already described in section 3.2.2.2 the practice of combining physical and monetary variables in production models. These applications usually declare that they aim to estimate technical efficiency. But it is clear that as long as there are monetary variables involved in the estimation, the resulting efficiency score will contain **some** portion of the allocative component and will not be purely technical. On the other hand, it is also clear that unless **all** of the variables are expressed in monetary terms, the efficiency score will not capture the complete allocative component and will therefore represent a mix of technical and allocative efficiency.

The monetary measure that we propose here has therefore the advantage of focusing purely on economic efficiency. The bottom line of our model is plain and simple: **More profits mean higher economic efficiency score.**

3.3.2.4 Relationship Between $\mathscr{M}$ and Neoclassics

It is important to mention the connection between $\mathscr{M}$ and neoclassical economics. The flexibility of our framework comes at a cost. Once prices are endogenous to the system, there is no anchoring point to define feasible derivatives, nor to implement the implicit function theorem and other analytical tools of neoclassical economics.

Consider an example when output price in a sector increases, ceteris paribus. According to neoclassical analysis, the firm is probably no longer economic efficient unless the firm does not change its behaviour, i.e. unless it adjusts the quantities of inputs and outputs. In the money-metric framework, however, this firm will be more efficient because it will generage — ceteris paribus — higher profits. This can account for situations in imperfect markets, e.g. when the price inceased precisely **because of** the behaviour of the firm, such as a successful marketing campaign.

The same logic can be applied to a general input price increase between two periods t and $t+1$: The firms might still be efficient **within** the year $t+1$, but the efficiency score has to go down relatively to the previous year, everything else remaining equal. The reason is that in the previous period t the firms were able to produce the given output with lower costs than they are producing it now. We employ this idea later when we discuss pooling of observations over time.

3.3.3 Interpretation: Knowing What We Do Not Know

3.3.3.1 Measuring Economic Efficiency

Several comments regarding interpretation and application of definition (3.4) are in order.

Firstly, some studies in fact adopt the approach in (3.4), without explicitly stating it. Yang & Chen [118] use a production frontier kernel which is strictly money-valued, to which they add other regressors (see table 2 *ibid*). Paradoxically, they consistently use the term "technical efficiency", which highlights the lack of discussion of the underlying methodology. But we already stressed that as long as the equation contains money-valued terms, the resulting efficiency measure will always contain some portion of **economic** efficiency.

Our contribution is that by properly defining $\mathscr{M}$ we formalize the existing idea of money-valued production **and** give it a theoretical underpinning. As we saw in section 3.2.2.2, the explanation for interchanging physical and monetary units has so far been neglected. Unlike the mixture models from section 3.2.2.2, definition in equation (3.4) provides a valid and consistent framework for analysis of economic efficiency.

3.3.3.2 The Flexibility—Detail Trade-off

Secondly, $\mathscr{M}$ and its frontier provide the most general description of production. Because it is purely empirical, we contend that it has relevance to real economies. Research on efficiency attempted to uncover both technical and allocation processes in firms and to separate these two components. As a consequence, questions concerning interpretation or validity of assumptions faded into the background.

The focus on technical efficiency is itself surprising, since technology per se is not the subject of economics, although with the words of Sautet [101, p. 4] "it has some influence on economic issues". It is true that one possible way to overcome the problems arising from separation of technical and allocative efficiency is a more detailed structure building. One of the most prominent examples is the complete demand–supply model by Berry *et al.* [12] for automobile industry. This approach can however lead to highly complex structures applicable only to very narrow segments. Note carefully that the specification of Berry *et al.* is only for one product (=automobiles), not even for the automotive industry as a whole.

On the other hand, our generalized approach abandons identification of efficiency components for the reasons that we discussed above: (i) data are not easily measurable and hence not available; (ii) structural assumptions required for identification are questionable for large cross-sections or aggregated datasets; (iii) identification in empirical studies has been handled with neither good precision nor with great success; and finally (iv) we want to learn mainly about the economic process, not the technique relation, since what matters in the end is the money-valued outcome. Note that some production schemes might be technically possible but economically infeasible: Think of re-building the Egyptian pyramids. These are of no interest from the economic point of view.

3.3.3.3 Aggregation

Thirdly, the crucial assumption for aggregation is to view the economic environment as a technology pool **and** a market pool. Just as technology can be regarded as the aggregate omnipresent knowledge that is available to everybody, so are market opportunities and negotiations available to all entrepreneurs. It follows that the observations are all drawn from one $\mathscr{M}$, rather than each firm or sector having its own $\mathscr{M}_i$.

In fact, if we only had one (cross-section) or a few (time series) observations for each

$\mathscr{M}_i$, we could not estimate much. Aggregation across sectors therefore enables more precise estimation by increasing the number of relevant observations. Note that aggregation in the case of $\mathscr{M}$ can be justified precisely on the grounds of no specific pricing structure, which would normally differ across companies.

Fourthly, we cannot expect $\mathscr{M}$ to be convex. The replication argument which justifies convexity of $\mathscr{Y}$ is likely to fail here: The feasibility of a given $(\boldsymbol{wx}, \boldsymbol{py})$ relies not only on general availability of aggregate technology and its replicability (as for $\mathscr{Y}$), but also on the unique bargaining abilities of the entrepreneur. This point will be important for computational implementation.

3.3.3.4 Panel Data

Finally it is necessary to bear in mind one crucial property of efficiency estimation: In a cross-section N efficiency scores are estimated from N observations, although the whole sample is used when estimating an individual efficiency score (i.e. one is looking for an ordinal ranking of evaluated firms). This complicates statistical inference for individual firms. Moreover, it is improbable that panel data could provide a remedy: Long time series for single firms are generally not available, and further questions arise with a dynamic specification of efficiency.

One would then ask why the feasible set should remain constant over time. Greene [54, p. 277] notes:

> For panels which involve more than a very small number of periods, this [time invariant efficiency] is a significant and possibly unreasonable assumption.

Therefore, it seems natural to treat inefficiency as time varying instead. We return to this issue in section 3.4.2.3.

3.4 Application to SME in the UK

3.4.1 The Data

The data that we use to test our hypotheses are extracted from the Annual Business Inquiry organised by the Office for National Statistics.[4] Compared to the publicly available version, our data are sizebanded according to the number of employees to distinguish different classes of SME.

The dataset can be summarised:

- Four-digit Standard industrial classification (SIC) which includes all sectors from agriculture to services.

[4] Detailed information about this product is found at [http://www.statistics.gov.uk/abi/]. Data are collected at firm level and aggregated into SIC sectors.

- Sizebands 1-10, 11-25, 26-50, 51-100, 101-250, and more than 250 employees.
- Variables: Number of firms; number of employees; wage costs; total employment costs ($EMPCOST$); net capital expenditure (NCE); turnover; gross value added (GVA).
- Years 1998–2007.

Because of data confidentiality, about a third of observations involved missing information, and these had to be omitted. We further deleted observations with negative GVA or NCE. Still, the resulting dataset contained $N_0 = 16,826$ observations, with more than $1,500$ datapoints for each year.

3.4.2 General Model Specification

3.4.2.1 Fitting a Model of Production

We specify our model of production as:

$$GVA = h(NCE; EMPCOST), \tag{3.5}$$

so that $\boldsymbol{wx} = (NCE, EMPCOST)$ and $py = GVA$. Note that when defining a one-dimensional output, we employed the fact how the money-metric framework can overcome the issue of disjoint production possibility sets, as described in section 3.3.2.2. This directly follows equation (3.2) as a widely accepted formulation in the literature. Energy and material costs do not enter (3.5) because they were already subtracted from total sales, yielding GVA. We are aware that subtracting costs of energy and materials is in effect a parametric operation. Nevertheless, there are at least two reasons why it should not influence information in the data significantly: (1) In this case the parametrization is intuitive; and (2) prices can be reasonably treated as given for energy and materials.

To this formulation we apply the robust nonparametric efficiency estimator described in section 3.4.3. The method for analysis of efficiency scores is outlined in section 3.4.4. Prior to that, we discuss our model building procedure.

3.4.2.2 Measurement of Capital

Measurement poses a challenge especially for capital. Studies in efficiency analysis commonly use data on fixed assets. However, economists know that from the viewpoint of economic calculation what we ideally want to measure is a flow variable; that is: How much does the use of given capital assets cost? Or equivalently: How much would it cost to hire these capital assets for the time required for production? Researchers are aware of this problem, and some of them use depreciation to extract a flow proxy from the stock of capital. Nonetheless, if this is accomplished by a constant depreciation rate, as in Biørn *et al.* [15], it does not bring any additional information.

Although it seems that 'net capital expenditure' could be adequate as a flow variable, NCE has its own shortcomings. It captures one-off acquisitions and disposals of capital, and hence offers no guide as to how these values are distributed over time. Because it is in fact the sum of positive and negative investments, it is very volatile, and in addition it can result in spuriously low or even negative values. The best option would be to combine one stock measure and one flow measure of capital. Since the former is not available to us, we have to continue with the latter, bearing in mind its flaws.

3.4.2.3 Pooling Observations

Pooling Across Sections

We attempt to provide an economy-wide analysis of efficiency of SME. To accomplish this, we utilize maximum available information and pool observations across sectors. If a vector $(\boldsymbol{wx}, py)$ is observed, it has to be feasible by definition. Hence once an observation is made, it is only natural to add it to the money-metric production set $\mathscr{M}$ defined in (3.3). Contrary to models involving detailed structure, our version provides framework which is flexible enough that pooling across heterogeneous sectors is economically meaningful.

Pooling Over Time

Using observations from different time periods raises major difficulties in any econometric analysis, because they cannot be regarded as independent.

Methodology for nonparametric methods is provided in Tulkens & Eeckaut [113]. They consider two main approaches: Either we construct a frontier for each year separately, or we update the frontier every year with new observations. This offers a decomposition of frontier shift and firm specific efficiency change.

We decided to use 2007 as the reference year against which efficiency is measured, so that the number of reference observations to construct $\widehat{\mathscr{M}}_{2007}$ was $1,526$.[5] This involves a significant computational simplification (see section 3.4.3.3), but only a minor loss of information. The relative efficiency ranking across years and across sectors remains the same, we only forfeit the absolute efficiency ranking for each given year. This simplification further makes comparison of efficiency scores more intuitive than in Tulkens & Eeckaut [113].

To clarify the proposed concept, consider two time periods $t = 0$ and $t = 1$. Let us define, slightly abusing the notation:

$$\begin{aligned} EY_t &= \text{efficiency frontier at time } t \\ Y_t &= \text{output of a firm at time } t \\ \theta_t &= \text{efficiency score of a firm at time } t \end{aligned}$$

[5] Outliers are excluded from this figure.

Then conceptually we can write:

$$\begin{aligned}\theta_t &= \frac{Y_t}{EY_t} \text{ for } t \in 0,1 \\ \theta_0^* &= \frac{Y_0}{EY_1}\end{aligned}$$

In this paper we use θ_0^* rather than θ_0, because only data from 2007 are used to construct the efficiency frontier, and all observations are compared against this EY_1. Then we can derive:

$$\begin{aligned}\frac{\theta_1}{\theta_0^*} &= \frac{Y_1}{EY_1} \times \frac{EY_1}{Y_0} = \overbrace{\frac{Y_1}{Y_0}}^{\text{output growth}} = \\ &= \frac{EY_1}{EY_0} \times \frac{Y_1}{EY_1} \cdot \frac{EY_0}{Y_0} \\ \frac{\theta_1}{\theta_0^*} = \frac{Y_1}{Y_0} &= \underbrace{\frac{EY_1}{EY_0}}_{\text{economic growth}} \times \underbrace{\frac{\theta_1}{\theta_0}}_{\text{ranking growth}} \\ \text{output growth} &= \text{economic growth} \times \text{ranking growth}\end{aligned}$$

This gives a clear interpretation of time-dependent efficiency scores which we estimate. The ratio of the most recent score θ_1 to scores from previous years (measured by the recent frontier) θ_0^* includes two components. First, it includes the relative performance growth: the improvement in the ranking as it would be measured by efficiency frontier in each given year (e.g. 80% efficiency in year 1 compared to 60% efficiency in year 0 implies an improvement by a factor of $\frac{1}{3}$). Second, it also captures the dynamic component of changing efficiency frontier which we call the economic growth component and which is represented by the frontier shift $\frac{EY_1}{EY_0}$. At the same time the ratio is equivalent to output growth of the given firm.

It must be noted that once a subset of observations is used as reference set, some observations might get **super-efficient**, that is they might achieve scores higher than one.

3.4.2.4 Data Processing

Outliers

When we run a simple free disposal hull efficiency measure from Cooper *et al.* [32, equation 4.69] as a preliminary test with all 16,826 observations, most of the efficiency scores lay within a reasonable interval of three standard deviations. However a small number of scores were wildly away, such that in a few cases the computer was effectively attaching them zero efficiency on a standardized interval $[0, 1]$.

Therefore we decided to employ outlier detection as suggested by Wilson [115] and we trimmed $\simeq 0.5\% \leftrightarrow 80$ observations. The principle of Wilson's measure is to compare the volume spanned by the whole dataset to the volume spanned by a subset where one or more points are deleted. For technical details see Wilson [115]. The number of observations was cut to $N = 16,746$.

Heteroscedasticity

A convenient property of nonparametric estimators is that data do not need any standardization. Nonparametric estimators automatically deal with heteroscedasticity, and consequently we do not have to scale the data as, for example, cost per unit of value added. This in turn means that we do not have to adopt any prior parametric assumption, which would normally be required for scaling.

This property has another important implication: The dataset does not have to be deflated by an inflation index. Given that the same deflating measure would be applied to all observations in a given year, it would not change the relative position of an observation as compared to other observations from that year. Deflating the data would only affect the relative spread of efficiency scores over time, yet this effect would be spurious because an aggregate inflation index does not reflect relative inflation in each sector.

3.4.3 Evaluation of Efficiency

3.4.3.1 Order-m Estimator

We employ the nonparametric estimator of efficiency by Cazals *et al.* [24]. The estimator is based on Assumption 4.2.1 of Simar & Wilson [106], which can be modified to our framework: Money-valued inputs and outputs are a pair of independent and identically distributed (*i.i.d.*) multidimensional random variables $(\mathcal{W}\mathcal{X}, \mathcal{P}\mathcal{Y})$ with a probability density on the support $\mathscr{M}$, with the property $\Pr((\boldsymbol{wx}, py) \in \mathscr{M}) = 1$ so that there is no statistical noise.

The robustness of this estimator comes from the fact that we are comparing an observation $(\boldsymbol{wx}_0, py_0)$ not to the whole sample, but to a randomly drawn subset of the sample. Averaging over the subset-dependent efficiency scores gives expected efficiency.

3.4.3.2 Convexity considerations

To evaluate efficiency relative to the money-metric frontier $\mathrm{Eff}(\mathscr{M})$, we must decide whether the empirical counterpart to equation (3.3) is a convex or nonconvex hull of available observations.[6]

We claim that the non-convex approach (FDH) is preferable. Convexity of the reference frontier is based on the replication argument. But $\mathscr{M}$ incorporates a variety of factors: technology, market structure, negotiation, managerial abilities etc. Reasoning based on replication is likely to fail here, and FDH is more appropriate. Moreover, convexity is only important in small samples: When the number of observations grows, approximation of the true frontier in $\widehat{\mathscr{M}}$ will approach strict convexity even for FDH.[7]

[6] The convex approach is called “data envelopment analysis” (DEA), the non-convex is called “free disposal hull” (FDH).

[7] It must be noted that the use of FDH was questioned by Thrall [111]. The criticism regards efficiency de-

3.4.3.3 Monte-Carlo Simulation

Order-m expected efficiency can be estimated as integration 3.5 in Cazals *et al.* [24], which does not have an analytical solution. Cazals *et al.* *ibid* proposed a four step Monte-Carlo algorithm. We take the computation from Daraio & Simar [37, p. 72] and adjust it to our money-metric frontier:

[1] Draw a sample (denoted sample b) with replacement among $\boldsymbol{wx}_i$ of size m such that $py_i \geq py_0$ and denote this sample $(\boldsymbol{wx}_{1,b}, \ldots, \boldsymbol{wx}_{m,b})$.

[2] Compute

$$\tilde{\tilde{\theta}}_{OM,b}(\boldsymbol{wx}_0, py_0) = \min_{q=1,\ldots,m} \left\{ \max_{j=1,\ldots,r} \left(\frac{wx^j_{q,b}}{wx^j_0} \right) \right\}.^{8}$$

[3] Redo [1]-[2] for $b = 1, \ldots, B$, where B is large.

[4] $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_0, py_0) = \frac{1}{B} \sum_{b=1}^{B} \tilde{\tilde{\theta}}_{OM,b}(\boldsymbol{wx}_0, py_0)$.

The simulated efficiency estimator $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_0, py_0)$ lies in $(0, 1)$ for inefficient observations.

After experimenting with the behaviour of $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_0, py_0)$ in smaller subsamples, and taking into account computational aspects, we finally specified the sample size of the draws $m = 150 \simeq 10\%$ and the number of iterations $B = 100$. The computational burden is considerable: Our specification required 75 minutes to compute.[9] This is one of the factors why we use only one year to construct the reference set $\widehat{\mathscr{M}}_{OM,2007}$.

3.4.4 Analysis of Efficiency Scores

3.4.4.1 Two-Stage Regressions

In our analysis, we would like to go further and find regular patterns in efficiency scores. Regressing estimates on explanatory variables other than those included in the production process — we shall denote them $\boldsymbol{z}$ — is widespread.

This practice was heavily criticised by Simar & Wilson [106, ch. 4.6]. The problem with a second stage regression is that estimates of efficiency are biased and serially correlated, and by construction induce dependence between the error term and explanatory variables in the second stage regression.

composition: Because FDH frontiers are not convex, some points on the 'efficient' frontier will necessarily be allocatively inefficient. But firstly this conclusion of Thrall was opposed by Cherchye *et al.* [25]; and secondly once our frontier is money-valued, this concern is irrelevant. Another complication of FDH is identification of returns to scale, but solutions are now available (e.g. Soleimani-damaneh & Reshadi [107]).

[8] Note that wx^j is the j-th element of vector $\boldsymbol{wx}$. This min-max algorithm is computationally equivalent to eq. 4.69 in Cooper *et al.* [32], see eqs. 2.26-2.27 in Daraio & Simar [37, p. 37].

[9] On a computer with 3 GHz processor and 2GB RAM.

A full statistical model which incorporates second stage analysis of efficiency scores, and which mitigates the above shortcomings, was developed by Simar & Wilson [105]. Their model is based on the assumption that a vector of additional variables $\boldsymbol{z}$ directly influences efficiency, so that for the joint distribution holds $G(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \neq G(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z})$. This form of statistical dependence is crucial, since as Simar & Wilson [105, p. 39] argue:

> otherwise, there would be no motivation for the second-stage regression.

3.4.4.2 Reformulation for Money-Metric Frontiers

Our model of efficiency frontier $\mathrm{Eff}(\mathscr{M})$, as we formulated it in section 3.3, is concerned with overall economic efficiency. This raises the question which variables belong to $\boldsymbol{z}$ in the distribution $G(\boldsymbol{wx}, py|\boldsymbol{z})$.

The measure of performance in our model is strictly monetary, so that it attempts to approximate profitability. Hence conditioning (i.e. environmental) variables $\boldsymbol{z}$ must be economic concepts concerning both external and internal environment in which firms operate. The former (external) could be captured by information on market structure, e.g. concentration indices. The latter (internal) are related to organization, management and entrepreneurship. For example, Man *et al.* [83] developed a conceptual model of entrepreneurial success, which consists of (1) Competitive scope, (2) Entrepreneurial competencies, and (3) Organizational capabilities (see figure 4 *ibid*). We treat firm size as a special case in section 3.4.5.

Nevertheless analysis of these factors lies beyond the scope of this article, not least because no such information is present in our dataset.

3.4.4.3 Ex-post Analysis

The data described in section 3.4.1 does not include any direct environmental variables, but we still would like to understand if some sectors show better performance than others, or whether efficiency improved over time. Obviously, by no reason should time or sectoral classification influence efficiency in the economic sense; this information is only collected ex-post.

It could be argued that the profitability of a sector influences the entrepreneur's decision to start his business, creating a link between sectoral classification and economic efficiency. But at the same time several mechanisms will work in the opposite direction to weaken this correlation. We doubt that potential entrepreneurs dispose of detailed information on profitability of sectors according to standard industrial classification. Rather, only some parts of the efficiency distribution within a given sector will be visible to them, leading to biased choices. Even if potential entrepreneurs had complete information on profitability of sectors, their decision will be driven by other factors such as their knowledge, skills and tastes, capital intensity and availability or regulatory obstacles. Most of all, given that the majority of businesses do not survive the early period of their existence, those who indeed decided based on the sectoral classification will be randomly mixed with those who decided based on other factors. We thererefore work with the assumption that sectoral classification and economic efficiency scores are statistically independent.

We can now return to the sceptical view of Simar & Wilson [105]: Does it mean that we cannot infer anything about efficiency patterns in this case? We want to see if efficiency score can be significantly explained along a sectoral classification. Therefore what we attempt is a decomposition motivated by 'unobserved components' class of models. A regression based on separation efficiency effects across three dimensions — time, sector, and firm size — cannot be justified in the sense of Simar & Wilson [105]. However we contend that it can still be useful from the empirical viewpoint, as a complement to pure descriptive analysis of efficiency scores.

3.4.4.4 Regression Specification

The model we employ in the second stage reads:

$$\begin{aligned} \hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_{i,t}, py_{i,t}) &= \alpha + \beta \cdot NEF_{i,t} + \sum_{t=1999}^{2007} \delta_t \cdot YEAR_t + \\ &+ \sum_{u \in \{SIC\}} \delta_u \cdot SIC_i + \\ &+ \sum_{v \in \{EG\}} \delta_v EG_{i,t} + \xi_i + \epsilon_{i,t}, \end{aligned} \tag{3.6}$$

where NEF is average number of employees per firm in the given four-digit SIC sector i (computed as Total number of employees divided by Number of firms), $YEAR$ is year dummy, SIC is sector dummy based on one-digit aggregated SIC[10], and EG is dummy for number of employees, grouped as 1-10, 11-25, 26-50, 51-100, 101-250, and >250.

To estimate this model, we interpret ξ_i as random effects. Because the variables included in (3.6) represent ex-post clustering, it is reasonable to assume zero correlation between ξ_i and regressors, as discussed in the previous section.

Further, contrary to Simar & Wilson [105], we use a robust measure of efficiency where scores are distributed on both sides of the efficient frontier. Therefore there is no need to compute truncated normal regression, instead ϵ is viewed as Gaussian.

The most pressing problem in equation (3.6) is the degree of correlation among the scores $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_{i,t}, py_{i,t})$. To obtain a meaningful covariance matrix we applied bootstrapping. We did not use the algorithm proposed by Simar & Wilson [104]. The shortcoming of their procedure is that they add new information into the sample, because estimates are updated each time based on draws from truncated normal distribution. This yields consistency if the underlying model of truncated normal distribution is correctly specified. Yet it also decreases the robustness of such an approach when the underlying distribution is not close to truncated normal. Hence we used simple bootstrap available directly in **STATA**, where covariance matrix is computed for repeatedly drawn subsamples from the data.

[10] One-digit SIC has more detailed classes A-O, but we could not regroup our observations this way. Instead, we created groups according the first digit of the four-digit SIC, which yielded ten clusters.

3.4.5 Conditional Efficiency Scores

3.4.5.1 Definition

It could be argued that firm size, as represented by number of employees (NEF), is an environmental variable which influences efficiency scores *ex ante*. The statistical model for efficiency is then characterized by the conditional distribution

$$G(\boldsymbol{wx} = (NCE, EMPCOST), py = GVA | z = NEF).$$

Instead of explaining efficiency by firm size in the second stage regression, we can directly evaluate conditional efficiency scores. Procedures for conditional estimation were developed by Daraio & Simar [36] and further enhanced by Bădin, Daraio & Simar [20]. While the latter model can work with multidimensional vectors $\boldsymbol{z}$, for our purpose the former model is fully sufficient. For recent overviews of conditional efficiency scores refer to Bădin, Daraio & Simar [21], [22].

3.4.5.2 Estimation of the Size Effect

We extend our previous analysis by the Monte-Carlo simulation proposed in Daraio & Simar [36, section 4.2]. The algorithm is similar to the simulation 3.4.3.3 above, except that in the first step there is an additional condition placed on the draws: The sample is drawn with probability

$$K((z - z_0)/h) / \sum_{i=1}^{N} K((z - z_i)/h).$$

Here $K(\cdot)$ is a probability kernel and h is a bandwidth. We decided to use Gaussian kernel.

We follow Daraio & Simar [36, section 4.3] and select for each observation with $Z_i = NEF_i$ a local bandwidth h_{Z_i} such that there exist k points Z_j verifying $|Z_j - Z_i| \leq h_{Z_i}$. The number of neighbourhood points k is chosen such that it maximizes the likelihood cross validation criterion. In our case $k = 1144$.

This procedure yields the conditional order-m efficiency score $\hat{\hat{\theta}}_{ZOM}(\boldsymbol{wx}_0, py_0 | z_0)$. Finally it remains to check if the size effect is significant. This can be done by regressing NEF_i on the ratio

$$\omega_{i,t} = \frac{\hat{\hat{\theta}}_{ZOM}(\boldsymbol{wx}_{i,t}, py_{i,t} | z_{i,t})}{\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_{i,t}, py_{i,t})}. \tag{3.7}$$

According to Daraio & Simar [36, section 4.4], the regression coefficient will be positive if the effect of $Z = NEF$ on production is negative, and conversely $\beta_{NEF} < 0$ if the effect of NEF on production is positive.

We cannot compare the result of this regression directly with results from equation (3.6). Nevertheless the ratio ω serves as a cross-validation of our second stage regression.

3.5 Results

3.5.1 First Stage

Computations were implemented in the statistical package **R**, using library FEAR by Wilson [116]. See table 3.2 for a summary.

Efficiency scores $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i)$ have to be regarded as relative ratios against the efficient level equal to one. Hence in table 3.2 the mean of $\simeq 0.477$ means that average sectors are less than half efficient as compared to best performers. Recall that this result holds even after accounting for outliers, who are ranked as 'superefficient'.

Interpreting efficiencies is not clear cut. Our measure is monetary, so that the only driving factor is costs per unit of value added. We built our model so that we are not able to distinguish technical and allocative efficiency. However, the great advantage of our approach is that it directly accounts for quality as it is perceived by buyers, because all production is priced.

Table 3.2 conveys one fundamental message. The wide dispersion of efficiency scores implies the need for more dynamic models of short-run out-of-equilibrium adjustment. Static equilibrium analysis helps us define and understand concepts of efficiency. Nonetheless our results suggest that imposing equilibrium conditions in empirical work on sectors that are not narrowly defined could potentially be misleading.

Visualising data with number of observations this large would require sophisticated tools and more space, because standard scatterplot matrix proved to be disorderly. Due to limited space, we illustrate only the most important relationship between efficiency and size of companies. In figure 3.1, we use the method of hexagon binning[11] to approximate the two-dimensional distribution, where the colour of each hexagon represents the number of observations in its area. Displayed are $13,871$ observations restricted to satisfy $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i) \in [0,1]$ and $NEF_i \in (0,250]$. From the clusters in the figure it is apparent that the majority of observations do not achieve full efficiency.

3.5.2 Second Stage

Results from the previous section still suffered from extreme points, with the farthest observation being 100 times more efficient than the unit reference frontier. Wilson's method to detect outliers ex-ante, as described in section 3.4.2.4, proved unsatisfactory, so we omitted approximately 1% of observations before conducting the second stage analysis, yielding $N_2 = 16567$.[12]

Regression (3.6) was implemented in the package **STATA** using maximum likelihood estimation. The first dummy in each group was automatically left out due to perfect multi-

[11] Library hexbin for **R**, see Lewin-Koh [82].

[12] Precisely, we removed 82 observations with $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i) < 0.0485$, 80 observations with $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i) > 4.56$, and 17 observations where $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i)$ could not be computed.

collinearity. The results are summarised in table 3.3, in table 3.4 where NEF is dropped and table 3.5 where both NEF and SIC dummies are dropped.

3.5.2.1 Overall Significance

Although all three regression are significant as a whole according to the Wald test, most of the individual dummies are not. It must be noted that dropping either one of the three dummy groups from equation (3.6) resulted in insignificant regressions. Hence it appears that the clusters capture a good portion of information on distribution of the efficiency scores.

Nonetheless most of the effects alone do not move efficiency in a definite direction. Specifically, only three dummies reported in table 3.3 have their confidence interval with both limits of the same sign.[13]

3.5.2.2 Size Effect

The coefficient on average number of employees per firm, β, is not significant at 5%, and this result is robust to dropping EG dummies. We report in table 3.4 the results of regression without NEF. In this case the joint hypothesis that all nine industrial dummies are zero ($\delta_u = 0$) cannot be rejected, so that we also report in table 3.4 the results of regression without NEF and δ_u's.

The EG dummies are in fact the most dominant effects. The joint hypothesis that all five employee group dummies are zero ($\delta_v = 0$) is rejected in all three cases at 5% significance level with the corresponding p-values of 0.0178, 0.0458 (NEF excluded) and 0.0260 (NEF and δ_u excluded), respectively.

Individually, in all three regressions the coefficients δ_{EG5} and δ_{EG6} are negative at 5% significance level, while δ_{EG4} is close to significant at 5% and also negative. Note that this negative effect is relative to the base-case of $EG1$, which is the employee group with 1-10 employees already contained in the coefficient α. Moreover, the magnitude of the effect is increasing with the number of employees, which suggests that with more employees economic efficiency actually **worsens**.

We are aware that our finding with respect to size is not sufficiently significant, of small magnitude (unit percentage points of efficiency score per hundred of employees) and that it appears counterintuitive. The conclusion that we draw is more cautious: The results hint that a positive relationship between size and efficiency proposed by earlier studies is limited to certain sectors and does not apply to the economy as a global principle.

3.5.2.3 Time Effect

Two of the time effects are significant at 1% and 6% respectively for the first and second regression and at 3% for the third regression. Nevertheless the overall message is blurred

[13] We obtained the same result with dummies coding two-digit SIC groups.

as no clear direction of the effect over time can be seen and the confidence intervals cover both positive and negative numbers. We investigated this further by including a simple time trend $\gamma \cdot t$ in (3.6), where we followed Battese & Coelli [9]. The result was insignificant both with or without year dummies, so we do not report it here.

The hypothesis that **economic** efficiency changed over time was therefore strongly rejected. This statement must however be read in its positive sense, not normative. For example, in one possible underlying scenario technical efficiency might have improved due to growth of labour productivity, but this might have been compensated by higher wages, so that overall the effect cancelled out. Because GVA less wages and capital costs can be viewed as a proxy proportional to profits, our results reveal that the share of revenues from entrepreneurial activities going to equity shareholders remained constant over time.

3.5.2.4 Mean Efficiency

Finally, significance of α statistically confirms the outcome of table 3.2: Average efficiency of the base-case sector ($YEAR = 1998$, EG 1-10 employees, SIC 1) represented by α is expected to be between quarter to half of the best practice frontier (see confidence intervals in tables 3.3, 3.4 and 3.5). This once again underlines not only the dynamic nature of competition, but also the magnitude of competitive pressures in the markets.

3.5.3 Conditional Second Stage

3.5.3.1 Conditional Efficiency Estimation

We evaluated both efficiency scores in the ratio ω_i in equation (3.7), as well as the likelihood cross validation criterion for bandwidth selection, in the statistical package **OCTAVE**. We employed adapted **MATLAB** routines from Daraio & Simar [36].[14]

In the Monte-Carlo algorithm we kept $B = 100$. As we were interested in the effect of firm size on the aggregate production frontier, we pooled all observations when estimating $\widehat{\mathscr{M}}$. This means that rather than recycling $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_{i,t}, py_{i,t})$ scores from section 3.5.1, we re-calculated them in **OCTAVE**. This was also done to keep both scores in ω_i computationally consistent.

As in section 3.5.2, we eliminated some of the outliers and restricted efficiency scores to 20, so that the number of observations was $N = 16656$. Box-plot statistics for efficiency scores $\hat{\hat{\theta}}_{ZOM}(\boldsymbol{wx}_{i,t}, py_{i,t}|z_{i,t})$ are reported in table 3.6.

3.5.3.2 Conditional Size Effect

While Daraio & Simar [36] estimate a non-parametric regression, we run a simple parametric regression to test the magnitude of the size effect:

$$\omega_{i,t} = c + \beta_{NEF} \cdot NEF_{i,t} + \eta_{i,t} \tag{3.8}$$

[14] We are grateful to Cinzia Daraio for kindly providing me with the **MATLAB** routines.

In this equation time is not treated structurally, hence all observations were pooled. The regression was implemented in **R** and the results are reported in table 3.7.

The conditional efficiency scores are on average higher than the standard order-m scores. This is expected, because conditioning on NEF explains some of the variation in efficiency scores.

The coefficient β_{NEF} is significant at 0.1% and positive, which suggest a negative size-efficiency relationship. However the coefficient is negligibly small: Increasing firm size by 100 employees would decrease efficiency by mere 0.85%, as compared to the non-conditioned efficiency score. Given the reported variation in overall efficiency scores, this result is not substantial. Thus we conclude that the results for conditional scores are in line with those reported in the previous section.

3.6 Conclusions

In the previous section we presented detailed efficiency analysis of British SME. We would like to stress the robustness of our work and its complementarity to previous research. Both these advantages are based on these features of the article: Firstly, we proposed a general methodological framework for nonparametric evaluation of economic efficiency which we call **money-metric efficiency frontier** $\mathrm{Eff}(\mathscr{M})$. This clarifies and extends the approach of previous papers. Secondly, our dataset ranges from agriculture to services, and this allowed us to test economy-wide hypotheses which had not yet been examined. Thirdly, we employed state-of-the-art robust methods for efficiency estimation. The nonparametric nature seems especially suitable for our large dataset.

Our results are related to economic efficiency, which we modelled as creation of value added relative to costs of inputs. The findings can be summarised in the following stylised propositions:

1. Efficiency scores across observations are very dispersed with no systematic sectoral pattern, which implies great heterogeneity within the economy. Specifically, we contend that it calls for more focus on out-of-equilibrium competitive and adjustment processes in further research.

2. We do not find significant evidence of a substantial, economy-wide size-efficiency relationship. Small samples benefit from better defined structure, but our finding implies that previous studies' results documenting a positive size-efficiency relationship are specific to either technical efficiency or to narrow sectors. The negative effect that we report is negligibly small. The mild evidence that the largest firms create less value added per unit of costs might be due to reporting bias.

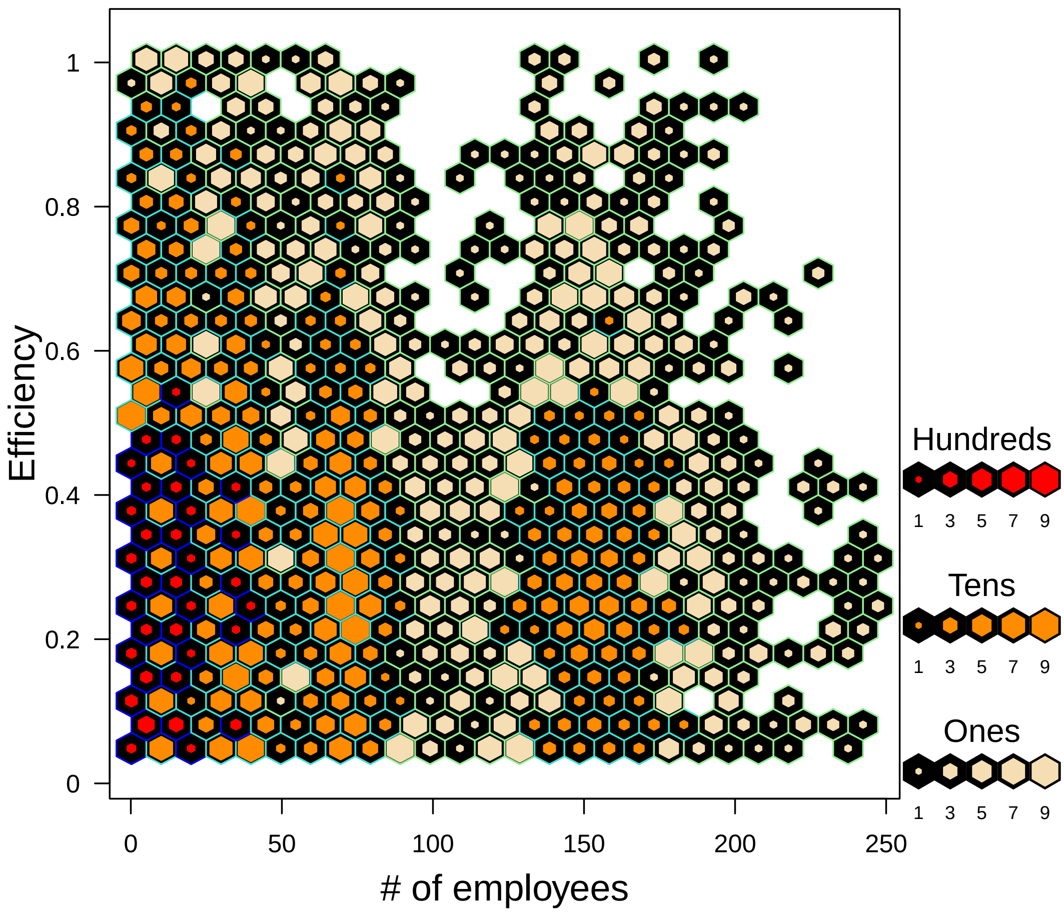
3. Economic efficiency remains relatively stable over time. In our view this constitutes evidence for the claim that wealth gains from presumed technology advances are

evenly distributed across stakeholders in firms (i.e. owners and providers of labour and capital).

4. Average sectors are expected to be two to four times less efficient than those on the efficient frontier. We interpret this as an indicator for the magnitude of competitive pressures in the markets.

Two extensions of our second stage analysis are straightforward: Firstly, we did not structurally address the dependence of efficiency scores between size groups (EG) within one SIC sector. This would require a more detailed three-level model, where we would consider possible combinations of interaction effects between the three levels time–sector–EG. Secondly, we could specify a dynamic regression with lagged efficiency score among explanatory variables $\mathbf{z}$, using GMM estimation. These extensions are left for further research.

Figure 3.1: Distribution of $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_{i,t}, py_{i,t})$ against $NEF_{i,t}$.



The figure shows clusters of efficiency scores depending on the number of employees. It appears that there is no straightforward pattern in the size-efficiency relationship. Upon detailed inspection three significant points of gravity can be identified: Small firms with 1–20 employees and efficiency between 0.1–0.4, firms with 65–75 employees and efficiency between 0.2–0.4 and finally firms with around 150 employees and efficiency of 0.2–0.3.

Table 3.1: Summary statistics of data from section 3.4.2.4.

	Mean	Std.dev.	1Q	Median	3Q
GVA	328,440	777,611	36,073	94,378	267,780
NCE	32,896	101,314	2,373	7,146	22,695
EMPCOST	177,957	413,022	19,842	52,672	150,730
NEF	160	416	15	37	130

$N = 16746$. GVA, NCE and $EMPCOST$ in thousand £.

Table 3.2: Box plot statistics for efficiency scores $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_{i,t}, py_{i,t})$.

min	1Q	median	3Q	max	mean
0.03958	0.18374	0.31484	0.47763	0.91753	0.46722

min $=$ max$\{$sample minimum; $1Q - 1.5(3Q - 1Q)\}$
max $=$ min$\{$sample maximum; $3Q + 1.5(3Q - 1Q)\}$
$\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i) = 1 \Rightarrow (\boldsymbol{wx}_i, py_i)$ is expected to be efficient according to the approximation of $\mathrm{Eff}(\widehat{\mathscr{M}}_{OM,2007})$.
$\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i) < 1 \Rightarrow (\boldsymbol{wx}_i, py_i)$ is inefficient.
$N = 16746$. Number of superefficient observations: $891 \simeq 5.3\%$.

Table 3.3: Maximum likelihood estimation of the model (3.6).

	Coefficient	Bootstrap Std. Err.	p-value	95% Conf. Interval	
β	0.0000131	0.0000124	0.291	−0.0000112	0.0000375
δ_{1999}	−0.0154784	0.0115287	0.179	−0.0380743	0.0071176
δ_{2000}	−0.014184	0.0139489	0.309	−0.0415233	0.0131552
δ_{2001}	−0.0292996	0.0113953	0.010	−0.051634	−0.0069652
δ_{2002}	−0.0127928	0.0161132	0.427	−0.0443741	0.0187886
δ_{2003}	−0.0099504	0.0129875	0.444	−0.0354054	0.0155046
δ_{2004}	−0.01348	0.0124282	0.278	−0.0378388	0.0108789
δ_{2005}	−0.0190152	0.0128486	0.139	−0.044198	0.0061676
δ_{2006}	−0.0297586	0.0156952	0.058	−0.0605206	0.0010033
δ_{2007}	−0.0071706	0.0143761	0.618	−0.0353472	0.021006
δ_{SIC1}	0.049848	0.0677879	0.462	−0.0830139	0.1827099
δ_{SIC2}	0.0597635	0.0682027	0.381	−0.0739114	0.1934384
δ_{SIC3}	0.0535592	0.0675033	0.428	−0.0787448	0.1858631
δ_{SIC4}	0.0388944	0.0676146	0.565	−0.0936279	0.1714167
δ_{SIC5}	0.0633962	0.0679409	0.351	−0.0697656	0.196558
δ_{SIC6}	0.098012	0.071844	0.172	−0.0427996	0.2388236
δ_{SIC7}	0.0611964	0.0673043	0.363	−0.0707175	0.1931103
δ_{SIC8}	0.0302315	0.0723416	0.676	−0.1115555	0.1720184
δ_{SIC9}	0.0436897	0.0684226	0.523	−0.0904161	0.1777955
δ_{EG2}	−0.0042957	0.012248	0.726	−0.0283013	0.0197099
δ_{EG3}	−0.0226775	0.0137472	0.099	−0.0496215	0.0042664
δ_{EG4}	−0.0285102	0.0149814	0.057	−0.0578732	0.0008529
δ_{EG5}	−0.0347542	0.0135295	0.010	−0.0612716	−0.0082368
δ_{EG6}	−0.0445577	0.0214728	0.038	−0.0866435	−0.0024718
α	0.3721508	0.0677007	0.000	0.2394598	0.5048418
σ_ξ	0.1115971	0.0055925		0.1011572	0.1231145
σ_ϵ	0.3720866	0.0085735		0.3556566	0.3892756
Wald χ^2 (df)	68.00		0.000		
♯ of obs.	16567				

Dependent variable is $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i)$.

Table 3.4: Maximum likelihood estimation of the model (3.6): NEF omitted.

	Coefficient	Bootstrap Std. Err.	p-value	95% Conf. Interval	
δ_{1999}	−0.0154453	0.0114328	0.177	−0.0378531	0.0069625
δ_{2000}	−0.0141136	0.0140524	0.315	−0.0416557	0.0134286
δ_{2001}	−0.0292714	0.0113942	0.010	−0.0516035	−0.0069392
δ_{2002}	−0.0127227	0.016114	0.430	−0.0443054	0.0188601
δ_{2003}	−0.009931	0.0130277	0.446	−0.0354647	0.0156028
δ_{2004}	−0.0133439	0.0123432	0.280	−0.037536	0.0108483
δ_{2005}	−0.0189297	0.0132626	0.153	−0.044924	0.0070646
δ_{2006}	−0.0297122	0.0157598	0.059	−0.0606009	0.0011765
δ_{2007}	−0.0069796	0.0144166	0.628	−0.0352357	0.0212765
δ_{SIC1}	0.0499005	0.0687583	0.468	−0.0848633	0.1846644
δ_{SIC2}	0.0595242	0.0695447	0.392	−0.0767808	0.1958293
δ_{SIC3}	0.0535214	0.0692095	0.439	−0.0821267	0.1891696
δ_{SIC4}	0.0394771	0.0684026	0.564	−0.0945895	0.1735437
δ_{SIC5}	0.0646025	0.0682038	0.344	−0.0690744	0.1982794
δ_{SIC6}	0.0990822	0.0717816	0.167	−0.0416072	0.2397715
δ_{SIC7}	0.0617157	0.0681998	0.366	−0.0719536	0.1953849
δ_{SIC8}	0.0303525	0.0739122	0.681	−0.1145128	0.1752179
δ_{SIC9}	0.0440441	0.0685649	0.521	−0.0903406	0.1784289
δ_{EG2}	−0.0040732	0.0123875	0.742	−0.0283522	0.0202058
δ_{EG3}	−0.022198	0.0136918	0.105	−0.0490334	0.0046375
δ_{EG4}	−0.0275403	0.0149815	0.066	−0.0569034	0.0018228
δ_{EG5}	−0.0326803	0.0134169	0.015	−0.058977	−0.0063836
δ_{EG6}	−0.0319063	0.0156711	0.042	−0.0626212	−0.0011915
α	0.3717171	0.0682933	0.000	0.2378647	0.5055696
σ_{ξ}	0.1115048	0.0056017		0.1010489	0.1230426
σ_{ϵ}	0.3721156	0.0083737		0.3560601	0.388895
Wald χ^2 (df)	44.67(23)		0.0044		
♯ of obs.	16567				

Dependent variable is $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i)$.

Table 3.5: Maximum likelihood estimation of the model (3.6): NEF, SIC omitted.

	Coefficient	Bootstrap Std. Err.	p-value	95% Conf. Interval	
δ_{1999}	-0.015449	0.0122945	0.209	-0.0395457	0.0086477
δ_{2000}	-0.0141482	0.0133259	0.288	-0.0402666	0.0119702
δ_{2001}	-0.0295187	0.013498	0.029	-0.0559743	-0.0030632
δ_{2002}	-0.0130009	0.0137133	0.343	-0.0398784	0.0138767
δ_{2003}	-0.010204	0.0134308	0.447	-0.0365278	0.0161198
δ_{2004}	-0.0135797	0.0109645	0.216	-0.0350697	0.0079103
δ_{2005}	-0.0194305	0.0129952	0.135	-0.0449006	0.0060397
δ_{2006}	-0.0300194	0.0120281	0.013	-0.0535941	-0.0064447
δ_{2007}	-0.0072851	0.014183	0.607	-0.0350833	0.0205132
δ_{EG2}	-0.0040838	0.0127173	0.748	-0.0290093	0.0208417
δ_{EG3}	-0.0221065	0.0147888	0.135	-0.0510921	0.006879
δ_{EG4}	-0.027305	0.0130397	0.036	-0.0528624	-0.0017477
δ_{EG5}	-0.0324329	0.0134949	0.016	-0.0588824	-0.0059833
δ_{EG6}	-0.0315033	0.0136279	0.021	-0.0582135	-0.004793
α	0.4293697	0.0126198	0.000	0.4046354	0.454104
σ_{ξ}	0.1123086	0.0045928		0.1036581	0.1216809
σ_{ϵ}	0.3721185	0.0111727		0.3508522	0.3946738
Wald χ^2 (df)	29.57(14)		0.0087		
♯ of obs.	16567				

Dependent variable is $\hat{\hat{\theta}}_{OM}(\boldsymbol{wx}_i, py_i)$.

Table 3.6: Box plot statistics for conditional efficiency scores from section 3.5.3.

	min	1Q	median	3Q	max	mean
$\hat{\hat{\theta}}_{OM,octave}(\boldsymbol{wx}_{i,t}, py_{i,t})$	0.00614	0.40669	0.60176	0.82104	1.44257	0.67944
$\hat{\hat{\theta}}_{ZOM}(\boldsymbol{wx}_{i,t}, py_{i,t}\|z_{i,t})$	0.00614	0.37456	0.63464	0.84351	1.54694	0.65412

$$\min = \max\{\text{sample minimum}; 1Q - 1.5(3Q - 1Q)\}$$
$$\max = \min\{\text{sample maximum}; 3Q + 1.5(3Q - 1Q)\}$$
$$N = 16656.$$

Table 3.7: Linear regression from equation (3.8).

	Coefficient	Std. Error	t-statistic	p-value
c	1.076	8.198e-03	131.242	$<$ 2e-16
β_{NEF}	8.547e-05	1.835e-05	4.659	3.2e-06
$\sharp$ of obs.	16656			

Residual standard error: 0.9873 on 16654 degrees of freedom.
Multiple R-squared: 0.001302, Adjusted R-squared: 0.001242.
F-statistic: 21.7 on 1 and 16654 DF, p-value: 3.205e-06.

Chapter 4

Non-Parametric Production Frontier of Czech Small and Medium Enterprises

This article was published as Průša, J. The Most Efficient Czech SME Sectors: An Application of Robust Data Envelopment Analysis. Czech Journal of Economics and Finance (2012), vol. 62, no. 1, pp. 44–65 [95].

Abstract

This paper analyzes the efficiency of Czech small and medium enterprises. Main focus is on structural analysis of Czech SME in manufacturing based on their efficiency. We use sectoral data from 2002 to 2005 of thirty manufacturing industries, each divided into five subgroups according to the number of employees. We employ standard and advanced robust data envelopment analysis (DEA) to obtain cross-sectional rankings of individual industries.

The results reveal substantial variance in the efficiency scores, which is only partly removed by the robust DEA specification. We found that the majority of sectors operate below full efficiency; with only a few industries belonging to top performers. Average efficiency lies between 50 to 70 per cent of the best sectors. We conclude that only a minor proportion of Czech SME are able to generate high value added per unit of labour-capital.

Keywords: production, efficiency measurement, data envelopment analysis, small and medium enterprises.

JEL classification: D24, L60, L70.

4.1 Introduction

4.1.1 Aims of the Analysis and Related Literature

Small and medium enterprises (hereinafter SMEs) form a vital part of developed economies, as has been stressed in a growing body of literature, see e.g. Schiffer & Weder [102], Ayygari *et al.* [6], Acs *et al. (eds.)* [1], Taymaz [110], Yang & Chen [118]. Research on Czech enterprises stressed especially institutional factors related to transition from a centrally planned economy to capitalism, such as the role of foreign direct investment (FDI) and institutions (examples include Djankov & Hoekman [42] and Marcinčin & Wijnbergen [84]). However literature on small and medium enterprises in the Czech Republic is rather scarce.

To the best of our knowledge, this paper is the first attempt to measure economic efficiency of Czech SMEs based on microeconomic principles using data envelopment analysis (hereinafter DEA), with main focus on structural analysis of Czech SME in manufacturing based on their efficiency. Our text therefore complements previous results which mostly relied on macroeconomic methods. The study by Benáček *et al.* [11] is an exception where the authors measured efficiency of textile and clothing firms by distance functions. Thanks to detailed information on individual firms, Benáček *et al.* were even capable of separating technical and allocation efficiency.

In a previous study Průša [94] provided general characteristics of the production process among Czech small companies. This paper will by contrast closely explore structural characteristics of SME, in that we will perform a cross-sectional study of SME statistics. This way we offer the reader revealing insights into the industrial fundamentals of the Czech economy. Specifically our model answers the following questions:

1. *How dispersed is the efficiency of individual sectors? Do most firms operate close to the efficiency frontier or away from it?*
   This is important to understand to which extent is static equilibrium a good approximation of real economy.

2. *Which are the most efficient industries?*
   Information about cross-sectional distribution of efficiency can guide profitable investment decisions which separate winners from losers.

3. *Are industries which are more concentrated and/or more regulated also more profitable?*
   This is useful especially from the regulation policy point of view.

4. *Does FDI support higher efficiency of the respective sectors?*
   Foreign investments are publicised as crucial contributors to economic development. However their impact is not straighforward.

5. *Are larger firms more efficient?*
   Finally this is the famous question of production economics, which from the theoretical viewpoint is condensed in returns to scale. In practice we can recognize much

subtler points, such as efficient control in family businesses as compared to embedded agency costs incured in large corporations.

This paper will focus on questions 1, 2 and 5. Although we do not attempt to provide rigorous analysis of questions 3 and 4, we are able to give several stylized facts as reference points for further investigation. We would also like to highlight here that sectoral classification does not serve in our analysis as an explanatory variable for inefficiency of its own. While understanding how efficiency varies across sectors ex post can be useful for practical reasons, analytically it is merely a descriptive statistic which adds detail to our results.

As is usual with empirical research, we are confronted with tensions between theory and practice. While the object—SME—is precisely defined, the statistics on SME are not so precisely measured and not completely available. While the methods are exactly defined, their application requires some assumptions to be loosened or disregarded. Thus we devote conscious effort to discuss how we proceed from theory to practice.

The rest of the paper is organized as follows: First we give the reader basic definition of SMEs. Next we proceed to the methodology of our analysis. We review data envelopment analysis (DEA), a practice for efficiency measurement which is commonly used in economic literature. Since lots of modifications were developed over the years, even the comprehensive handbooks (Cooper *et al.* [32], Cooper *et al.* [33], Coelli *et al.* [29]) listed in the bibliography of this paper are far from exhaustive. We focus on two specifications which we find suitable for our data and which are treated in more detail.

Finally section 4.3 forms the core of our genuine research. We analyze sectoral data on Czech small and medium enterprises for the period 2002 to 2005. DEA is used to obtain industry-specific efficiency scores. This allows us to unveil structural patterns within Czech SME industrial sectors.

4.1.2 Definition of SME

Small and medium enterprises, abbreviated as *SME*, are defined as companies not exceeding specific size limits. The official definition by the European Union is given in table 4.1. It is not a clearly disjunctive definition, if related to employment only. The complication emanated from the fact that in the EU SME has become an important tool for economic policy measures. Note that a firm must satisfy the first condition and either one of the last two conditions at the same time in order to be classified as SME. Lots of countries created their own definitions, e.g. Switzerland or the USA chooses 500 employees as the cutoff.

In the Czech Republic, SME account for one third of the Czech GDP and for close to two thirds of employment. This share remained more or less stable over the last ten years 1997-2006. This holds for the accounting value added as well, which stayed close to 53 per cent throughout the ten years.[1] It confirms that SME form the fundamentals of Czech economy,

[1]Statistics on SME published by the Ministry of industry and trade in its "Report on the development of SME and their support in 2006", downloaded at [http://www.mpo.cz/dokument32006.html] on January 5, 2008.

which are worth a proper analysis.

4.1.3 Macroeconomic Environment

Before we turn to analysis of SME efficiency, we provide basic macroeconomic overview for the period 2001–2006 in table 4.2. There is a minor slowdown visible in 2002 following adverse global conditions (dot-com bubble), however overall this was a time of both prosperity and increasing productivity for the Czech economy.

For most of the period inflation measured by the consumer price index (CPI) was moderate, while the producer price index (PPI) experienced wider fluctuations. The most interesting with respect to our topic is labour productivity, defined as GDP divided by employment. Even though productivity grew at a fast pace, this was partially offset by increased real wages. Labour costs for firms did not rise so dramatically, however the effect on profitability is not straightforward as costs were increasing along with productivity. These considerations provide yet another reason to investigate firm efficiency in great detail.

4.2 Measurement of Efficiency

4.2.1 The Concept of Efficiency

Competition belongs to the most powerful ideas in economics. Being able to benchmark economic units (individual agents, firms, whole economies) against each other implies that economists are able to provide direct insights into wealth creation. Such analysis of productivity renders motivation for improvement and thus drives development of the economy and, ultimately, of the society.

The related concepts of comparative advantage, competitiveness, productivity or efficiency have provided economists with tools to measure economic performance both at microeconomic and at macroeconomic level. Since this paper concentrates on the former, this sections provides microeconomic framework for efficiency measurement.

Although efficiency analysis is now an established field of microeconomics, it must be noted that this was driven more by necessity and observations about reality rather than by advances in pure theory of production. The core of neoclassical economic analysis is mostly relying on static equilibrium which without doubt provides insightful illustrations of market principles, but which cannot properly account for systematic departures from what is perceived as the efficient frontier.

Accordingly explanations of efficiency emerged as separate (though not always isolated) theories. It is not the purpose of this study to present them thoroughly, nevertheless let us mention here major streams in this field.

Vintage models assume that although aggregate technology is available to all producers, it is evolving over time and thus different producers at different times of investment acquire different vintages of technology. This implies heterogeneity of production capabilities, i.e.

certain time structure of capital. Before an investment is made, the production set (defined later) is the same for all producers i: $(\mathscr{Y}_i|\boldsymbol{\beta})$, $\boldsymbol{\beta}$ being the vector of parameters which characterize the technology. After the investment is made, each producer has his own specific production capabilities: $(\overline{\mathscr{Y}}_i|\overline{\boldsymbol{\beta}}_i)$. See e.g. Johansen [64].

Institutional economics assumes frictions which arise for each exchange transaction, be it exchange on the market (buying or selling for a price) or a non-market transaction (e.g. interaction within an organization). Inefficiencies may result from the internal organization of the firm. Management techniques (termed corporate governance by institutional economists) will crucially influence a firm's performance, as will the staff and their behaviour. Even in the same firm a different amount of goods is produced on different days due to unexpected failures and complications. Other bottlenecks may stem from inappropriate institutional settings. The more the state interferes in entrepreneurial activities, the higher the risk that something will go wrong. Ménard [88] offers an up-to-date summary of institutionalist view of organizations.

Austrian economics concentrates on entrepreneurs as discoverers of market opportunities. In this dynamic view, the economy is always developing and never achieves static equilibrium. The main stress is put on the importance of time in the production process. Therefore this stream is somewhat related to the vintage models and the time structure of capital. For a modern overview of Austrian production theory see e.g. Sautet [101].

Finally let us mention the view which was developed by Leibenstein [80]. He coined the term X-efficiency and his theory directly assumes inefficiency as an inherent property of all human activities. Because his approach to inefficiency is axiomatic and does not offer much room for explanation, this theory remains peripheral.

4.2.2 The Plain Vanilla Model of Efficiency

4.2.2.1 Technical Efficiency

The starting point of modern production analysis is profit maximization, profits being defined as the difference of revenues less cost. If we are to find out which decision making unit performs best at this decision, we have to recall that the production process links together two distinct worlds: technical parameters and economic parameters. The former determine the capability to produce large quantities of outputs, the latter are governed by preferences and scarcity. Accordingly we formalize the production process and the concept of efficiency.

Following the exposition by Daraio & Simar [37], the production set $\mathscr{Y}$ is defined as all feasible input-output vectors $[\boldsymbol{x},\boldsymbol{y}]$ from the set of nonnegative real numbers $\Re^r_{0,+} \times \Re^s_{0,+}$:

$$\mathscr{Y} \quad = \quad \{(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{x} \in \Re^r_{0,+}, \boldsymbol{y} \in \Re^s_{0,+} |\ (\boldsymbol{x},\boldsymbol{y}) \text{ is feasible}\}.^2 \tag{4.1}$$

[2] For detailed discussion on standard assumptions on technology see e.g. Kogiku [71].

We can further define the technologically efficient production frontier $\mathrm{Eff}(\mathscr{Y})$:

$$\begin{aligned}\mathrm{Eff}(\mathscr{Y}) \quad = \quad & \{(\boldsymbol{x},\boldsymbol{y}) \in \mathscr{Y} | \\ & \forall[\boldsymbol{x}^1 \leq \boldsymbol{x},\ \boldsymbol{y}^1 \geq \boldsymbol{y},\ (\boldsymbol{x}^1,\boldsymbol{y}^1) \neq (\boldsymbol{x},\boldsymbol{y})] : \\ & (\boldsymbol{x}^1,\boldsymbol{y}^1) \notin \mathscr{Y}\}. \end{aligned} \tag{4.2}$$

Then a producer will be technically efficient if and only if they operate on $\mathrm{Eff}(\mathscr{Y})$.[3]

4.2.2.2 Economic Efficiency

Even if the firm was technically efficient, it would not make much sense for the firm to produce goods at a cost or for a price that nobody buys them. It is the key task for the firm to allocate resources according to the willingness of consumers to pay for produced goods. The ability of firms to choose from the technical possibilities the one which suits most its customers is called allocative efficiency.

The tool which allows firms to achieve allocative efficency are the prevailing market prices, which directly embody information on customer preferences. Therefore we want to include market prices of outputs $\boldsymbol{p}$ and of inputs $\boldsymbol{w}$ into our analysis. In the simplest neoclassical case of perfect competition, prices are assumed to be exogenous from the point of view of a single firm[4], so that the profit function can be derived.

Definition 4.2.1 *A profit function $\Pi(\cdot)$ is a general solution to the profit maximisation problem:*

$$\Pi(\boldsymbol{p},\boldsymbol{w}) = \arg \max_{\{\boldsymbol{x},\boldsymbol{y}\}} \{\boldsymbol{p}'\boldsymbol{y} - \boldsymbol{w}'\boldsymbol{x} |\ (\boldsymbol{x},\boldsymbol{y}) \in \mathscr{Y}\}.$$

This is by a contradiction argument equivalent to:

$$\Pi(\boldsymbol{p},\boldsymbol{w}) = \arg \max_{\{\boldsymbol{x},\boldsymbol{y}\}} \{\boldsymbol{p}'\boldsymbol{y} - \boldsymbol{w}'\boldsymbol{x} |\ (\boldsymbol{x},\boldsymbol{y}) \in \mathrm{Eff}(\mathscr{Y})\}. \tag{4.3}$$

Naturally, for a producer to achieve overall efficiency, they have to be both technically and allocatively efficient.

4.2.3 Measuring Efficiency in Monetary Units

The separation of the two components of efficiency poses the main snag for any efficiency measurement. The technical part is captured in data in physical units. If we assign certain prices to these volumes, we can trace the economic part. The ideal statistic would contain

[3] As we have seen, assuming the technology parameter $\boldsymbol{\beta}$ away is equivalent to saying that all firms with the same products use the same transformation of inputs. This would be the case of perfect competitions where producers are identical (in terms of technology), or in the long run when all producers can adopt the most efficient technology. In the short run however, which will be the framework for our data analysis, differences in $\boldsymbol{\beta}$ will be one explanatory factor of inefficiency.

[4] Under perfect competition prices are determined following interaction of a large number of firms and consumers who have complete information, thus a single firm cannot change the prevailing market price.

all these pieces of information for a large number for individual producers; this is however rarely available (and in most situations even not sensible).

If a researcher has data in monetary units at hand, he is left with three options. Firstly, we can assume exogenous and hence *constant* prices across the dataset, which is the perfect competition case. Then prices are just labels for technology and technical efficiency can be measured directly. Secondly, which amounts to assuming the same, we can adjust the data for prices manually—this means that we divide each observation by an aggregate price index. This way we can get from monetary back to technical units.

Thirdly, we can define a framework which explicitely allows for price exogeneity and product heterogeneity. Průša [93, section 4] suggested to use *money-metric production frontiers*, where definitions in equations (4.1) and (4.2) are in monetary units (see Průša [93, equation 2]). In other words, equation (4.2) tracks the 'profit frontier', meaning that the impact of imperfect competition and product heterogeneity is already incorporated in money-denominated datapoints.

Money-metric efficiency frontiers trade in separation of technical and allocation efficiency for clear economic interpretation. The resulting efficiency score directly captures overall economic efficiency: In terms of equation (4.3) higher revenues per unit of costs are regarded as equivalent to higher economic efficiency. Moreover, the beauty of the monetary computation lies in the fact that this 'profit frontier' logic holds irrespective of technology. It must be stressed that the first and the third options are computationally equivalent—since we plug in the data we have. However, it seems more straightforward to assume price endogeneity, especially with cross-sectional data. Therefore, in the following sections, we assume the third approach: Input vectors $\boldsymbol{x}$ and output vectors $\boldsymbol{y}$ denote data in monetary units, unless otherwise stated.

4.2.4 Data Envelopment Analysis

4.2.4.1 Basic Model Structure

In this paper we use *data envelopment analysis* (DEA) to analyze economic efficiency. A DEA model constructs a hyperplane around the dataset, with points lying on the plane being efficient and points within the space being inefficient. Efficiency is then measured as the distance of a given observation to the efficient frontier.

We already listed reference books on DEA in our introduction to this paper. Here we depict the basic model and proceed to a recent robust specification. We can write a simple input-oriented DEA problem in matrix notation as follows:

$$\begin{aligned} \min_{\boldsymbol{\lambda},\theta} \quad & \theta \\ \text{subject to } \theta \boldsymbol{x}_i \quad & \geq \quad X\boldsymbol{\lambda} \\ Y\boldsymbol{\lambda} \quad & \geq \quad \boldsymbol{y}_i \\ \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \quad & \geq \quad 0, \end{aligned} \tag{4.4}$$

which is known as the CCR model, since it was formulated by Charnes, Cooper and Rhodes. The intuition behind this mathematical problem is as follows: The vector $\boldsymbol{\lambda}$ attaches weights to single producers: In the third line, $\boldsymbol{\lambda}$ selects certain firms, which are called 'reference' producers of the evaluated decision making unit DMU$_i$. These 'reference' producers, weighed together by $\boldsymbol{\lambda}$, produce at least as many outputs as DMU$_i$. $\boldsymbol{\lambda}$ then scales the input matrix X to see whether it is possible to cut down inputs at DMU$_i$ by some coefficient θ.

In other words, given that producers selected by $Y\boldsymbol{\lambda}$ have greater output than $\boldsymbol{y}_i$ (third line), then DMU$_i$ should certainly not use more inputs than $X\boldsymbol{\lambda}$ (second line). θ_i measures by how much inputs of DMU$_i$ can be decreased before they reach the boundary of $X\boldsymbol{\lambda}$.

The problem must be solved n times for all producers to obtain each firm's efficiency score, which is an estimate $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} \in [0,1]$.[5]

4.2.4.2 Returns to Scale

Model (4.4) does not impose any additional conditions on $\boldsymbol{\lambda}$, so that technical efficiency is computed under the assumption of constant returns to scale, see Cooper *et al.* [32, chapter 4]. Variable returns to scale (RTS) were introduced in the BCC model by Banker, Charnes and Cooper who added the constraint $\sum_{i=1}^{n} \lambda_i = 1$ to the CCR model. Similarly, the specification of $\sum_{i=1}^{n} \lambda_i \leq 1$ would result in non-increasing returns to scale.

One further specification is derived from a similar constraint: if we add the constraint $(\sum_{i=1}^{n} \lambda_i = 1) \ \wedge \ (\forall i : \lambda_i \in \{0,1\})$, we change DEA to the free disposal hull (FDH) model. FDH is not connected to returns to scale and it differs from both CCR and BCC models in that it draws an envelope that is not convex.[6] We will need this specification later for the statistical modification of DEA.

4.2.5 Statistical Methods in Non-Parametric Approach

In this section we select one modification of DEA which surmounts two big obstacles of the basic model: (1) deterministic and non-statistical nature; (2) influence of outliers and extreme values (Daraio & Simar [37, p. xviii]).

4.2.5.1 Probabilistic Production Process

The CCR model from section 4.2.4 is fully deterministic in that it assumes $\Pr([\boldsymbol{x}_i, \boldsymbol{y}_i] \in \mathscr{Y}) = 1$, where $\Pr(\cdot)$ denotes probability. This time inputs and outputs are a pair of independent and identically distributed (*iid*) multidimensional random variables $(\mathcal{X}, \mathcal{Y})$, although for

[5]Instead of assuming data in monetary units, prices can be incorporated into DEA by assigning value to the objective function, leaving constraints unchanged. This requires strong assumptions, above all that prices remain constant for any amount of inputs consumed and any amount of outputs produced. For examples of allocation efficiency models, see e.g. Coelli [27] or Cooper *et al.* [33, section 1].

[6]Convex technology means that if there are two input combinations c_1 and c_2 that generate a certain level of output y, then any convex combination of c_1 and c_2 will also produce the same level of output y.

individual observation it still holds $\Pr([\boldsymbol{x}_i, \boldsymbol{y}_i]) \in \mathscr{Y}) = 1$. Following the derivation of Daraio & Simar [38], this yields a joint probability measure characterized by the function

$$H_{\mathcal{X}\mathcal{Y}}(\boldsymbol{x}, \boldsymbol{y}) = \Pr(\mathcal{X} \leq \boldsymbol{x}, \mathcal{Y} \geq \boldsymbol{y}).$$

For the DMU $[\boldsymbol{x}, \boldsymbol{y}]$ this function captures the probability that this firm will perform worse than others, i.e. that it will use more inputs ($\mathcal{X} \leq \boldsymbol{x}$) and at the same time produce less output ($\mathcal{Y} \geq \boldsymbol{y}$).

For this probability measure we can derive the probability that once the firm produces less, it also uses more inputs. This is the conditional probability that the firm uses more inputs ($\mathcal{X} \leq \boldsymbol{x}$) conditional on producing less output ($\mathcal{Y} \geq \boldsymbol{y}$) and can be written as the conditional distribution function:

$$F_{\mathcal{X}|\mathcal{Y}}(\boldsymbol{x}|\boldsymbol{y}) = \Pr(\mathcal{X} \leq \boldsymbol{x}|\mathcal{Y} \geq \boldsymbol{y}) = \frac{\Pr(\mathcal{X} \leq \boldsymbol{x}, \mathcal{Y} \geq \boldsymbol{y})}{\Pr(\mathcal{Y} \geq \boldsymbol{y})} = \frac{H_{\mathcal{X}\mathcal{Y}}(\boldsymbol{x}, \boldsymbol{y})}{S_{\mathcal{Y}}(\boldsymbol{y})},$$

where we assume $S_{\mathcal{Y}}(\boldsymbol{y}) > 0$. Notice how this corresponds to the idea behind the minimization problem in (4.4): There the computation also selects dominant producers with greater output than the analyzed DMU (third line in the linear program 4.4) and looks by how much inputs of the analyzed DMU are greater than those of the dominant reference producers (second line).

This conditional probability can be empirically estimated by computing

$$\widehat{F}_{\mathcal{X}|\mathcal{Y},n}(\boldsymbol{x}|\boldsymbol{y}) = \frac{\sum_{i=1}^{n} I(X_i \leq x, Y_i \geq y)}{\sum_{i=1}^{n} I(Y_i \geq y)},$$

$I(\cdot)$ is the indicator function, and X_i, Y_i are individual observations.

4.2.5.2 Order-m Estimator

Once we established the conditional probability measure in the previous section, it remains to compute efficiency based on this probabilistic production process. This can be done by the order-m estimator introduced by Cazals *et al.* [24].

The idea is simple: Suppose we have an observation $[\boldsymbol{x}_0, \boldsymbol{y}_0]$. As in the CCR model (4.4), we select those observations with larger output. From this subset of observations satisfying $Y \geq \boldsymbol{y}_0$ we draw randomly with replacement $X_1, \ldots, X_m$. These draws are then distributed according to the conditional distribution function $F_{\mathcal{X}|\mathcal{Y}}(\cdot|\boldsymbol{y})$, as follows from the previous section.

We construct the production possibility set as in Daraio & Simar [38, c. f.]:

$$\widetilde{\mathscr{Y}}_m(\boldsymbol{y}_0) = \left\{ [\boldsymbol{x}, \boldsymbol{y}] \in \Re_+^{p+r} | \boldsymbol{x} \geq X_i, \boldsymbol{y} \geq \boldsymbol{y}_0 \right\}.$$

The set $\widetilde{\mathscr{Y}}_m(\boldsymbol{y}_0)$ captures the trivial fact that once input X_i selected by the random draws is sufficient to produce output $\boldsymbol{y} \geq \boldsymbol{y}_0$, then any greater amount of input $\boldsymbol{x} \geq X_i$ must also be able to produce output $\boldsymbol{y} \geq \boldsymbol{y}_0$.

Then we measure the efficiency of our firm against production possibility set $\widetilde{\mathscr{Y}}_m(\boldsymbol{y}_0)$ as the expected minimum efficiency score. We first compute

$$\widetilde{\theta}^m_{(\boldsymbol{x}_0,\boldsymbol{y}_0)} = \inf\left\{\theta|(\theta\boldsymbol{x}_0,\boldsymbol{y}_0) \in \widetilde{\mathscr{Y}}_m(\boldsymbol{y}_0)\right\} \tag{4.5}$$

and take expectations

$$\theta^m_{(\boldsymbol{x}_0,\boldsymbol{y}_0)} = E_{\mathcal{X}|\mathcal{Y}}(\widetilde{\theta}^m_{(\boldsymbol{x}_0,\boldsymbol{y}_0)}|\mathcal{Y} \geq \boldsymbol{y}). \tag{4.6}$$

Notice that equation (4.5) nicely corresponds to the second constraint in the linear program (4.4): In both cases θ determines by how much it is possible to contract inputs of DMU_0 before we reach the 'minimum input requirement' boundary which is set by firms producing at least as much output as DMU_0.

It is equation (4.6) which differentiates the probabilistic approach from section 4.2.4.1. Here we compare our DMU to randomly drawn subsets of larger producers (i.e. those with higher output), effectively evaluating the CCR model (4.4) for each draw, and then look at the efficiency score we can statistically expect over a large number of randomly drawn subsets. That is, instead of computing the efficiency score once deterministically, we compute it many times for smaller subsets of observations (against which our DMU is compared), and then calculate the average score. This procedure is designed to smooth potential outliers or data errors. It is precisely this idea that makes the order-m estimator much more robust than the standard CCR model.

Finally equation (4.6) has to be turned into an operational procedure for computation. Using the empirical distribution function $\widehat{F}_{\mathcal{X}|\mathcal{Y}}$, and recalling that statistical expectation is simply the integral over the distribution function, it can be shown that the score equals to:

$$\widehat{\theta}^m_{(\boldsymbol{x}_0,\boldsymbol{y}_0)} = \widehat{E}_{\mathcal{X}|\mathcal{Y}}(\widetilde{\theta}^m_{(\boldsymbol{x}_0,\boldsymbol{y}_0)}|\mathcal{Y} \geq \boldsymbol{y}) = \int_0^\infty \left(1 - \widehat{F}_{\mathcal{X}|\mathcal{Y}}(u\boldsymbol{x}|\boldsymbol{y})\right)^m \mathrm{d}u.$$

Unfortunately this integration can not be carried out analytically. Instead Cazals *et al.* [24] proposed a four step Monte-Carlo algorithm, which we quote as in Daraio & Simar [37]:

[1] Draw a sample with replacement among X_i such that $Y_i \geq \boldsymbol{y}_0$ and denote this sample $(X_{1,b}, \ldots, X_{m,b})$.

[2] Compute $\widetilde{\theta}^{m,b}_{(\boldsymbol{x}_0,\boldsymbol{y}_0)} = \min_{i=1,\ldots,m}\left\{\max_{j=1,\ldots,p}\left(\frac{X^j_{i,b}}{x^j}\right)\right\}$.

[3] Redo [1]-[2] for $b = 1, \ldots, B$, where B is large.

[4] $\widehat{\theta}^m_{n,(\boldsymbol{x}_0,\boldsymbol{y}_0)} = \frac{1}{B}\sum_{b=1}^{B}\widetilde{\theta}^{m,b}_{(\boldsymbol{x}_0,\boldsymbol{y}_0)}$.

4.2.5.3 Convex order-m frontier

Most of section 4.2.4 deals with efficiency estimates based on convex technology. The only exeption is FDH, briefly mentioned in 4.2.4.2. Since the order-m frontier is based on FDH, it is not convex. Therefore in this section we add convexity to the order-m model from 4.2.5.2.

FDH is derived from the approximation of production technology (Daraio & Simar [38]):

$$\begin{aligned}\widehat{\mathscr{Y}}_{FDH} &= \left\{[\boldsymbol{x},\boldsymbol{y}] \in \Re_+^{p+r} | \boldsymbol{x} \geq \boldsymbol{x}_i, \boldsymbol{y} \leq \boldsymbol{y}_i, i = 1,\ldots,n\right\},\\ \widehat{\theta}^{FDH}_{(\boldsymbol{x}_0,\boldsymbol{y}_0)} &= \inf\left\{\theta | (\theta\boldsymbol{x}_0, \boldsymbol{y}_0) \in \widehat{\mathscr{Y}}_{FDH}\right\}.\end{aligned}$$

Daraio & Simar recall that usual convex DEA scores can be easily obtained from FDH results: It suffices to multiply observed inputs $\boldsymbol{x}$ by the FDH efficiency scores $\widehat{\theta}^{FDH}_{(\boldsymbol{x},\boldsymbol{y})}$ and then run the respective convex linear program on the transformed data, which can be for example the CCR minimization problem as defined in (4.4).

They use this feature to convexify the order-m estimate in the same way. They construct transformed data by

$$\widehat{\boldsymbol{x}}^{\partial}_{m,i} = \widehat{\theta}^{m}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} \cdot \boldsymbol{x}_i$$

and propose the linear program for the convex order-m efficiency estimator (hereinafter referred to as COM):

$$\begin{aligned}\widehat{\theta}^{m,C}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} &= \min_{\boldsymbol{\lambda},\theta} \theta\\ \text{subject to } \theta\boldsymbol{x}_i &\geq \sum_{i=1}^{n} \lambda_i \widehat{\boldsymbol{x}}^{\partial}_{m,i}\\ Y\boldsymbol{\lambda} &\geq \boldsymbol{y}_i\\ \sum_{i=1}^{n}\lambda_i &= 1\\ \lambda_1,\ldots,\lambda_n &\geq 0.\end{aligned} \tag{4.7}$$

This is the final formulation which we will use in our data analysis.

4.3 Efficiency of Czech SME

4.3.1 Data Description

The dataset is based on a statistical enquiry by the Czech Statistical Office, which covers all firms with 100 or more employees, 55 per cent of companies with 10–99 employees and about 2.6 per cent of the micro-segment (below 10 employees). Certain part of the aggregated data is published in the yearly summary on economic activity of Czech small and medium enterprises.[7]

Our data were obtained directly from the Czech Statistical Office and they are slightly more detailed than in the publicly available booklet. The dataset has four dimensions:

[7] The publication can be found under reference number 8007-[xx], where xx are the last two digits of the corresponding year. The 2008 version is available at: [http://www.czso.cz/csu/2008edicniplan.nsf/p/8007-08].

1. thirty-item two-digit OKEC[8] classification, including OKEC codes 10 to 41[9], i.e. agriculture and services are not included;

2. size classification with breakdowns at the following number of employees: 0-10-20-50-100-250;

3. eleven economic indicators: output, sales revenue, accounting value added, tangible assets, intangible assets, acquisition of tangible and intangible assets, number of employees, average number of employees, payroll and other personnel expenses;

4. years 2002 through 2005.

The data implies the main characteristics of the analysis. Items under point 3 are fitted to the standard economic labour-capital-output framework. Points 1 and 2 are used as the basis for cross-section computations. Together they yield $30 \times 5 = 150$ observations, less some empty rows each year. Finally we get $n^{(2002)} = 135$, $n^{(2003)} = 135$, $n^{(2004)} = 134$ and $n^{(2005)} = 136$, totalling 540 observations.

4.3.2 Model Specification

4.3.2.1 Dimensions of the Frontier

Specifications of production functions generally follow the 'KLEM' approach, where gross output y_{gross} is given by a function as defined in Burnside [23, equation 2.1]:

$$y_{\text{gross}} = f(\text{capital, labour, energy, materials; technology}). \tag{4.8}$$

The abstract notion of "technology" does not enter the model ex ante; rather it is the result of estimation in the form of the Solow residual. We are dealing with manufacturing industries only, hence land can be neglected without serious distortions of our model.

We can subtract nonproductive intermediate inputs from equation (4.8) and in so doing arrive at a second possible specification where output is measured as value added y_{net}:

$$y_{\text{net}} = f(\text{capital, labour; technology}). \tag{4.9}$$

In this paper we prefer the latter approach for both theoretical and practical reasons. The theoretical justification is that we are interested in *productive efficiency*, that is in efficient employment of productive inputs: namely capital and labour.[10] Eficient use of non-productive inputs is certainly significant from the managerial point of view, but it is not in the scope

[8] European Union uses the abbreviation NACE: Nomenclature Générale des Activités Économiques dans les Communautés Européennes.

[9] OKEC 12 is not included. Full list of industries is available at [http://www.czso.cz/csu/klasifik.nsf-/i/odvetvova_klasifikace_ekonomickych_cinnosti_(okec)] in Czech or at [http://ec.europa.eu/comm/-competition/mergers/cases/index/nace_all.html] in English.

[10] As we noted above we neglect land in this model.

of this paper. Referring back to section 4.2.3, in our model higher value added per unit of monetary inpus implies higher economic efficiency.

The practical reason stems from the sensitivity of DEA to outliers, an issue which becomes more pronounced with more variables.[11] Specification in (4.9) should further improve the robustness of estimation results due to lower dimension of the model.

The step from equation (4.8) to (4.9) places a strong parametric assumption on how energy and materials enter the production process. Let us recall from section 4.2.3 that we measure efficiency in monetary units.[12] Then however (4.8) and (4.9) represent transformations of a profit function in which all components are naturally additive. Therefore in our specification the frontiers as defined in equations (4.8) and (4.9) are equivalent. Clearly the resulting efficiency scores will be slightly different, because the latter kapital-labour (KL) efficiency (4.9) neglects the efficiency components in energy and materials. Yet as we noted above these non-productive inputs are not the focus of our paper; we concentrate on productive efficiency.

Based on the preceding discussion we specify as the vector of inputs:

$$\boldsymbol{x} = [\text{assets, investment, employees, wages}]',$$

while output is represented by accounting value added. Before we proceed to a detailed discussion of the model structure in the next section, we define the input variables here: 'Assets' are totalled tangible and intangible assets; 'wages' are wage outlays plus other personal expenses—both summations were done in order to decrease the number of explanatory variables. 'Investment' is acquisition of tangible and intangible assets. 'Employees' is the average number of employees, the single non-monetary input.[13]

4.3.2.2 Economic Meaning of the Model

The usage of the economic indicators deserves several comments. The indicators can be regarded as aggregated accounting figures. Sales revenue tracks all goods and services that the company was able to sell on the market. Output adds goods that were already produced but not yet sold to the sales revenue. Finally, when the cost of materials is subtracted, we get accounting value added. This should approximately express how much a firm is able to produce from its *flow* of capital and labour, since the cost of these is not included in the sum of materials.[14]

The average number of employees is more preferable to the number of employees. The latter captures the sum of employees on each particular day, which is then recalculated on

[11] The speed of convergence in probability of DEA estimators decreases exponentially with their dimension, while it increases only linearly with the number of observations.

[12] The only exception being labour; see below.

[13] Wage outlays are highly correlated with the number of employees. As was pointed to us by a referee, in an econometric setting this would lead to multicollinearity and would have to be accounted for. However with DEA this issue does not cause any problems.

[14] Output can be considered a proxy for y_{gross} in equation (4.8) and value added a proxy for y_{net} in (4.9).

the basis of days worked to get the former. It follows that the average captures all fluctuation of employees, which is exactly what we need.

The reason to include both the number of employees and total wage outlays is that we want to account for the firm size effect. We cannot use average wages instead of total wage outlays, because statistical data only measure total wages directly. The average is then computed from the total by dividing by the number of employees, and this division would algebraically create perfect multicollinearity between average wages and number of employees.

We include 'investment' even though it is a forward-looking variable. Variables in a production function should represent **flows** but 'assets' is a **stock** variable. Ideally we would like to include the real cost of capital to the firm, which is however unknown and we are not aware of any precise measure for this variable. This is why we assume 'investment' to be a good proxy for depreciation, the more so that we use aggregate data on sectoral level which smooths the effect of one-off investments on the firm level. In turn we consider depreciation in itself a plausible approximation for the real cost of capital. Rather than deleting assets altogether from the model, combining 'assets' and 'investment' should provide us with a reasonable picture of how efficiently firms employ their capital. Moreover investment can also be interpreted as a proxy for the willingness of firms to innovate. Thus we argue that it will help us unveil the importance of innovation for productive abilities of Czech SME. footnoteThe comment on multicollinearity from the previous footnote applies to assets and investment as well.

We refrain from deflating the money values, for which we find two reasons. Firstly, if the adjustment should add any useful information, we would require detailed separate data on input *and* output prices across various sectors. However such data are not available and there is no reason to assume that deflating by aggregate CPI and PPI would improve the results, quite on the contrary. All sectors would be deflated by the same figure and this would only distort the results even more. Secondly, since we are measuring value added, and assuming that inflation on both input and output side of the production equation are similar across sectors, neglecting inflation should not significantly impact our cross-sectional results.[15]

It remains to note that panel research is limited by the short time span—only four consecutive years. Therefore we do not explicitely account for technological change. Any technological advances are entrenched nonparametrically in the efficiency scores.

4.3.3 Envelopes I: Standard DEA Results

Consider the BCC model, i.e. equation (4.4) with the additional constraint $\sum_{i=1}^{n} \lambda_i = 1$ introducing variable returns to scale. We implemented this computation for each year separately

[15] If price shocks are not evenly distributed across sectors, there will be time-series bias in the efficiency scores. Rather than distorting the data ex-ante, we prefer to look at the results ex-post and to see if jumps in efficiency are correlated with asymmetric price shocks. As the number of employees is measured in physical units, this variable effectively dampens the asymmetric inflation bias. We are grateful to a referee for this suggestion.

via **DEAP**, a freely available program by Coelli [27].

To get an overview of the distribution of efficiency, we computed box plot statistics given in table 4.3, where Q stands for quartile. The true maximum of $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$ is of course always equal to one, nevertheless in this case statistics defines maximum as the upper quartile plus 1.5-times the quartile spread $(3Q - 1Q)$. Points above this outside bar (or below the respective bar for minimum) are taken as outliers.

For all years the mean of scores is higher then the median, meaning that the estimated efficiency distribution is skewed to lower scores. Average efficiency amounts to a mere 25 per cent of the best industries, a feeble performance. This demonstrates the sensitivity of DEA to outliers and calls for correction by means of a more advanced model.

Our analysis concentrates on groups of firms defined by size, so we break down our results with respect to number of employees (table 4.4). It seems that average efficiency is increasing with more employees, but this relationship starts only at the second size group (10-19 labourers). The smallest firms do best in every year, and moreover by a considerable gap.

Proposition 4.3.1 ***Preliminary results.*** *The BCC model unveiled the following:*

- *Distribution of efficiency results is heavily skewed to lower scores. It seems that there are outliers which exercise considerable influence on overall results.*
- *Larger firms tend to be more efficient on average, with one surprising exception: The smallest entrepreneurs rank first in every observed year.*

From this proposition we can deduce what to do next. Firstly, we will apply a statistically based DEA model in order to control for significant outliers. With refined results at hand, we will observe what the impact on efficiency distribution and its skewness will be, if any.

Secondly, we will analyze the sectoral structure. To make our conclusions more precise, we take 25 best and 25 worst industries in every year. In other words, we classify close to twenty percent of the observations as frontier points, among which we look for the intersection in at least three years.

4.3.4 Envelopes II: Robust DEA Results

In this section we report results of the convex order-m estimator (COM). We obtained the scores thanks to the package **FEAR** by Paul Wilson [116], where both the Monte-Carlo simulation from section 4.2.5.2 and the solution of equation (4.7) are available.

First we had to specify the computational aspects: parameters m and B. Cazals *et al.* [24, theorem 2.3] show that as $m \to \infty$, we have the convergence $\widehat{\theta}^m_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} \to \widehat{\theta}^{FDH}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$, and similarly $\widehat{\theta}^{m,C}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} \to \widehat{\theta}^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$. With higher m fewer observations will lie above the efficient frontier and the estimator gets less robust. Based on trial and error, we chose $m = 50$ (i.e. $\simeq 10\%$ of observations) as the level of robustness. With lower numbers of reference observations (e.g.

$m = 20$), there was unusually high ratio of super efficient firms with scores higher than unity, namely more than two thirds, which we assessed implausible. For $m = 50$ this ratio fell little below 50%. As for the number of replications, we used $B = 200$. More replications did not bring remarkably different results, only the computation time grew rapidly.

Distribution of individual efficiency estimates appears more favourable than in the simple CCR model. Scores for 2004 and 2005 shifted most visibly, so that we do not observe 75% of the data below 30%-level of top efficiency any more. The probabilistic approach suppressed super efficient outliers and the obtained estimates represent the true efficiency level of individual observations more accurately. We actually applied a flexible measure, which we expanded in the middle and stripped at the extreme values. Still variation of efficiency scores remains high even for the robust estimator and this volatility appears to be a robust result itself.

Recalling Aigner & Chu [2] and their criticism of average production functions, it could seem that we only moved to a certain "average" production plan. Yet histograms which we do not reproduce here disclose that the results are far from resembling normal distribution, because there are two peaks. Moreover the estimates are still skewed to the left, so that while having used the flexible measure, apparently we did not lose large parts of information contained in the data.

Table 4.5 tracks the distribution of efficiency scores in more detail. When confronted with the initial results in table 4.4, we conclude that any direct relation between efficiency and size formulated in proposition 4.3.1 is weakened by the COM model. If we trust COM in that it suppressed the influence of outliers, we may conclude that the strong mean efficiency of the smallest enterprises (as reported in table 4.4) was a result given by the presence of favourable extreme observations.[16]

As noted in section 4.3.1, our measure of output is the accounting value added, which is defined as output less cost of materials used in manufacturing.[17] The efficiency estimate therefore says how much of value added a firm is able to produce from a certain stock of capital and employed labour, and it is normed relative to the best practice. Hence lower efficiency score means less value added per unit of capital-labour.

Taking the example of capital, in practice this can be interpreted as follows. A firm can have a few or a lot of machines. Remember that all our computations are per unit of input, say per one machine. Thus in textile industry value added per sewing machine can be either high (jeans sold for higher price, or more jeans produced, or both) or low. Our results mean that in most cases the value added produced by a sewing machine will be rather low.

Proposition 4.3.2 *Distributional results.*

[16]These in turn may have been caused by favourable sample selection.

[17]Output = Sum of: (1) sales revenue from own products, (2) gross profit on merchandise sold (3) received leasing installements, (4) change in inventories and (5) self-constructed asset revenue.

Cost of materials = Sum of (1) the value of purchased and already used material, energy and of supplied materials which are not storable, and (2) of the value of purchased services.

- *Although the robust specification of DEA mitigated the skewness caused by outliers, variation of efficiency scores remains high.*
- *COM estimator results are skewed towards lower efficiency. The majority of firms operate below full efficiency, while only a few companies (industries) belong to top performers. Average efficiency lies between 50 to 70 per cent of the best sectors.*
- *Since value added was used as a proxy for output, we conclude that only a minor proportion of Czech SME are able to generate high value added per unit of labour-capital.*

Let us repeat what we achieved by COM: Due to the small number of observations, we did not leave out extreme points. As a consequence, we smoothed the efficient frontier, but our structural results should not greatly differ from those in section 4.3.3.

In table 4.6, we list 25 best and worst industries for each year, which is nearly one fifth of the data. Those items which were on the list in at least three years out of the four we classify as structural leaders and structural losers of the beginning of the twenty first century. In each of the groups we further distinguish between those oriented towards processing of raw materials and those in advanced manufacturing.

Proposition 4.3.3 ***Structural results.***

- ***Leaders.*** *Most top efficient industries belong to sophisticated manufacturing: food; tobbaco products; fabricated metal products; machinery; electrical machinery; radio, television and communication equipment. Yet there are also some commodities among the most profitable: electricity, gas, steam and hot water supply, which might stem from the monopolistic nature in this segment; and further wood & cork; metal ores.*
- ***Stragglers.*** *Just two items do not deal with raw materials: office machinery & computers; automotive. The rest of those losing out are more or less connected to commodities: leather; pulp & paper; coke, refined petroleum products and nuclear fuel; basic metals; recycling; water supply; coal & lignite; crude petroleum & natural gas. The latter two are surprising, given the rising energy prices.*
- *We identify one strong chain: metal ores—fabricated metal products—machinery—electrical machinery.*
- *That the automotive, coal & lignite and crude petroleum & natural gas sectors place among the worst performers means that gains on a large scale (e.g. due to FDI) are not always passed on to suppliers among SME.*

The last point is a strong result: It confirms that even in booming sectors supported by influx of FDI, smaller companies do not have the negotiating leverage necessary to reap more profits and grow rapidly.

To sum up, we were able to identify at least three key patterns in the course of our analysis: (1) There is significant variation of efficiency scores. (2) Czech SME are not able to generate high value added per unit of labour-capital. (3) Finally we identified the best performing SME sectors.

4.4 Conclusions

At the beginning we set the aim of analyzing cross-sectional efficiency of Czech small and medium enterprises, which are grossly defined as companies with less than 250 employees.

Data envelopment analysis (DEA) constructs the boundary of the multidimensional set of observations and measures the distance of firms from this efficient frontier. It is derived from microeconomic framework. The statistics from the Czech Statistical Office do not represent individual producers, so that we took a careful step towards aggregation. However given the detailed breakdown of the industries and size groups, even so we did not touch the level of aggregation commonly applied in macroeconomics.

By construction DEA is particularly suitable for cross-sectional rankings. Therefore we let it unveil structural lags among industries. We first observed unreasonably high variance of individual efficiency scores. For this reason we applied the probabilistic DEA, which made the efficiency measure more flexible. Right at the beginning, we made the assumption of variable returns to scale; this simplification has been widely recognized in literature by the frequent use of the Banker-Charnes-Cooper specification.

The resulting list of leaders and stragglers as in proposition 4.3.3 does not suggest any clear-cut outperforming or losing clusters; though we can still identify the chain *metal ores—fabricated metal products—machinery—electrical machinery*. What becomes apparent is that the large scale boom of big factories is not necessarily passed on to SME suppliers—e.g. automotive; coal & lignite; crude petroleum & natural gas.

Moreover we find that the majority of sectors operate below full efficiency, while only a few industries belong to top performers. Average efficiency lies between 50 to 70 per cent of the best sectors. In our computations we used value added as a proxy for output. Therefore we derive that only a minor proportion of Czech SME are able to generate high value added per unit of labour-capital. That is, most industries do not generate as much value added from their *flow* of capital and labour as the best ones. This result is not very surprising, just as it is not very encouraging.

Table 4.1: Definition of SME according to the EU legislation.

Enterprise Category	Headcount	Turnover	Balance Sheet Total
Micro	< 10	$\leq$ €2 million	$\leq$ €2 million
Small	< 50	$\leq$ €10 million	$\leq$ €10 million
Medium-sized	< 250	$\leq$ €50 million	$\leq$ €43 million

Table 4.2: Czech macroeconomic indicators 2001–2006.

Indicator		2001	2002	2003	2004	2005	2006
Real GDP growth	*%, y-o-y*	3.1	2.1	3.8	4.7	6.8	7.0
Inflation (CPI)	*%, y-o-y avg.*	4.7	1.8	0.1	2.8	1.9	2.5
Inflation (PPI)	*%, y-o-y avg.*	2.8	-0.6	-0.4	5.5	3.1	1.5
Unemployment	*%, avg.*	8.1	7.3	7.8	8.3	7.9	7.1
Labour productivity	*%, y-o-y*	3.2	0.9	5.1	5.0	5.0	6.6
Unit labour costs	*%, y-o-y*	4.0	3.9	1.3	2.5	-1.1	-0.4
Average real wage	*%, y-o-y*	3.9	6.1	5.7	3.4	3.0	4.0

Source: Czech Statistical Office.

Table 4.3: Box plot statistics for efficiency scores $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$.

	min	$1Q$	median	$3Q$	max	mean
2002	0.1500	0.4155	0.4910	0.6290	0.9410	0.5534
2003	0.020	0.370	0.498	0.691	1.000	0.5279
2004	0.031	0.064	0.133	0.299	0.604	0.2282
2005	0.0420	0.0995	0.1660	0.3630	0.6690	0.2743

Table 4.4: Mean efficiency score $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$ according to size group and year.

♯ of employees	2002	2003	2004	2005
<10	0.754	0.629	0.358	0.390
10-19	0.482	0.496	0.115	0.142
20-49	0.485	0.486	0.169	0.253
50-99	0.485	0.478	0.209	0.264
100-250	0.541	0.540	0.268	0.311

Table 4.5: Box plot statistics for efficiency scores $\widehat{\theta}^{m,C}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$.

	$\sharp$ of employees	min	1Q	median	3Q	max	mean
2002	<10	0.248	0.542	0.681	0.929	1.000	0.694
	10-19	0.122	0.457	0.541	0.664	1.000	0.572
	20-49	0.293	0.467	0.548	0.659	0.991	0.575
	50-99	0.399	0.522	0.564	0.656	0.922	0.587
	100-250	0.217	0.495	0.582	0.785	1.000	0.618
2003	<10	0.335	0.493	0.685	0.847	1.000	0.682
	10-19	0.188	0.397	0.497	0.599	1.000	0.535
	20-49	0.302	0.470	0.617	0.680	1.000	0.605
	50-99	0.139	0.429	0.529	0.651	1.000	0.546
	100-250	0.141	0.524	0.645	0.799	1.000	0.639
2004	<10	0.075	0.196	0.355	0.748	1.000	0.478
	10-19	0.087	0.161	0.276	0.363	0.816	0.317
	20-49	0.116	0.290	0.388	0.549	0.771	0.412
	50-99	0.093	0.266	0.340	0.620	1.000	0.437
	100-250	0.162	0.347	0.457	0.676	0.988	0.517
2005	<10	0.075	0.222	0.410	0.625	1.000	0.474
	10-19	0.095	0.195	0.270	0.484	0.949	0.383
	20-49	0.117	0.284	0.429	0.681	1.000	0.492
	50-99	0.080	0.244	0.398	0.657	1.000	0.476
	100-250	0.126	0.396	0.475	0.767	1.000	0.546

Table 4.6: Best and worst industries according to $\widehat{\theta}^{m,C}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$.

Best industries					**Worst industries**				
2002	2003	2004	2005	$\bigcap$	2002	2003	2004	2005	$\bigcap$
11_{250}	11_{9}	11_{9}	10_{99}	13_{9}	10_{49}	10_{19}	10_{9}	10_{9}	10_{250}
13_{9}	13_{9}	11_{99}	14_{250}	15_{250}	10_{99}	10_{99}	10_{250}	10_{19}	11_{99}
14_{250}	14_{19}	13_{9}	15_{49}	16_{9}	10_{250}	10_{250}	14_{9}	10_{49}	19_{9}
15_{99}	15_{250}	15_{250}	15_{99}	20_{9}	11_{9}	11_{99}	15_{19}	11_{9}	21_{9}
15_{250}	16_{9}	16_{9}	15_{250}	28_{9}	11_{99}	11_{250}	17_{99}	11_{99}	23_{9}
16_{9}	16_{99}	18_{49}	16_{9}	28_{250}	14_{19}	19_{9}	19_{9}	11_{250}	27_{9}
18_{9}	18_{9}	18_{250}	16_{49}	29_{250}	16_{49}	21_{9}	20_{19}	13_{9}	30_{250}
19_{9}	18_{19}	20_{9}	17_{49}	31_{9}	16_{250}	22_{19}	20_{99}	16_{99}	34_{9}
19_{19}	18_{49}	22_{9}	20_{9}	32_{250}	21_{9}	23_{9}	21_{9}	19_{9}	34_{19}
20_{9}	18_{250}	23_{99}	21_{99}	40_{250}	21_{19}	24_{19}	22_{19}	19_{19}	37_{19}
22_{9}	19_{250}	24_{9}	22_{19}		23_{9}	25_{19}	22_{99}	20_{99}	41_{19}
25_{49}	20_{9}	25_{19}	23_{99}		24_{9}	26_{19}	23_{9}	21_{9}	41_{99}
26_{250}	23_{19}	25_{99}	25_{19}		27_{9}	27_{19}	24_{49}	23_{9}	
28_{9}	23_{49}	26_{250}	26_{19}		27_{250}	27_{99}	24_{99}	23_{19}	
28_{49}	28_{9}	27_{99}	27_{99}		28_{19}	30_{19}	27_{9}	23_{49}	
29_{9}	28_{250}	28_{9}	28_{9}		30_{49}	30_{49}	27_{19}	27_{9}	
29_{99}	29_{9}	28_{250}	28_{250}		30_{250}	30_{250}	30_{250}	27_{19}	
29_{250}	29_{19}	29_{250}	29_{250}		32_{49}	34_{9}	31_{19}	30_{49}	
30_{19}	29_{250}	31_{9}	31_{9}		34_{9}	34_{99}	34_{9}	30_{99}	
31_{9}	31_{9}	32_{9}	32_{250}		34_{19}	35_{99}	34_{19}	30_{250}	
32_{9}	31_{250}	32_{250}	34_{9}		34_{49}	37_{19}	35_{19}	33_{250}	
33_{9}	32_{250}	33_{49}	35_{250}		35_{250}	37_{49}	37_{19}	34_{19}	
35_{9}	40_{9}	33_{250}	36_{19}		37_{19}	37_{250}	41_{9}	35_{19}	
36_{9}	40_{49}	34_{99}	40_{49}		41_{19}	41_{19}	41_{49}	41_{19}	
40_{250}	40_{250}	40_{9}	40_{250}		41_{99}	41_{49}	41_{99}	41_{99}	

$\bigcap$ indicates that the industry was among the best/worst in at least three years.

Table 4.7: Selected NACE classification: Nomenclature Générale des Activités Économiques dans les Communautés Européennes.
See [http://ec.europa.eu/comm/competition/mergers/cases/index/nace_all.html].

Code	Description
10	Mining of coal and lignite; extraction of peat
11	Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction, excluding surveying
13	Mining of metal ores
14	Other mining and quarrying
15	Manufacture of food products and beverages
16	Manufacture of tobacco products
17	Manufacture of textiles
18	Manufacture of wearing apparel; dressing and dyeing of fur
19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
21	Manufacture of pulp, paper and paper products
22	Publishing, printing and reproduction of recorded media
23	Manufacture of coke, refined petroleum products and nuclear fuel
24	Manufacture of chemicals and chemical products
25	Manufacture of rubber and plastic products
26	Manufacture of other non-metallic mineral products
27	Manufacture of basic metals
28	Manufacture of fabricated metal products, except machinery and equipment
29	Manufacture of machinery and equipment not elsewhere classified (n.e.c.)
30	Manufacture of office machinery and computers
31	Manufacture of electrical machinery and apparatus n.e.c.
32	Manufacture of radio, television and communication equipment and apparatus
33	Manufacture of medical, precision and optical instruments, watches and clocks
34	Manufacture of motor vehicles, trailers and semi-trailers
35	Manufacture of other transport equipment
36	Manufacture of furniture; manufacturing n.e.c.
37	Recycling
40	Electricity, gas, steam and hot water supply
41	Collection, purification and distribution of water

Chapter 5

Parametric Production Function of Czech Small and Medium Enterprises

This article was published as Průša, J. Productivity of Czech Small and Medium Enterprises: Lagging Behind Their Potential. Journal of Industry, Competition and Trade (2010), vol. 10, no. 3-4, pp. 343–363 [94].

Abstract

This paper analyzes microeconomic production functions of Czech small medium enterprises. We use the data from 2002 to 2005 of thirty manufacturing industries (agriculture and services are not included), each divided into five subgroups according to the number of employees. We employ stochastic frontier analysis (SFA) to make statistical inference on the production process.

Our results demonstrate that Czech SME depend in their functioning more on labour than on capital. The impact of tangible or intangible assets such as software or patents is negligible, while the effect of investment is negative. SFA strongly supports the presence of a systematic gap between common practice and best practice: the majority of firms significantly differ from top performers. Finally a simple test for time effect shows that between 2003 and 2005 Czech SME moved towards higher efficiency.

Keywords: production, efficiency measurement, stochastic frontier analysis, small and medium enterprises.

JEL classification: D24, L25.

5.1 Introduction

The term small and medium enterprises (SME) has recently gained more attention in general media, eventually reflecting the key contribution of SME to a healthy economy. In the Czech Republic however, where small entrepreneurs had to build from scratch after 1989, research in this field has remained largely untouched. Our paper aims to partly fill this gap, in that it captures main characteristics of the production function of SME.

We can regard small and medium businesses from two points of view: static and dynamic. Firstly, we look at their structural position in the economy. Although general public better knows giant brands, the SME matter because they form an economy's fundamentals. They can be compared to ants, who impact little individually but hugely altogether. Small entrepreneurs build the economy from the bottom, so that they are the true discoverers of market niches that call for filling.

Besides their economic impact on the creation of value (GDP), they play a key social role as well. Although many of them start as self-employed, later on as they grow they eventually become important local or national employers. Usually SME lack sufficient sources of capital and rely on more labour intensive production processes, or even they concentrate in industries which are inherently labour intensive. This biases their productivity in terms of value added per employee towards worse ranking (which can be misleading), yet it leads to their prominent position as dynamic and flexible job creators.

It follows from table 5.1 that in modern market economies, SME employ between one half to three quarters of the workforce in manufacturing. It is true that the breakdown point at 250 employees (or any other number) is artificial, nonetheless it becomes apparent that size of businesses matters—not least to people employed there.[1]

These figures clearly illustrate the interest of SME for economists. In the rest of our paper, we first present several characteristics of the SME sector. It is clear that there are economies to scale and also its resources need not be allocated in an optimal structure. Section 5.3 outlines the methodology used to estimate the production function and efficiency of SME: stochastic frontier analysis (SFA). The results are presented in section 5.4.

Estimating the production function on the macroeconomic level is a common econometric exercise, and appeared in several papers on the Czech economy. To name just one example: Hájek [57] tracked the determinants of economic growth given by the model of Solow. Yet our approach differs significantly from that of Hájek and similar studies, since we are deriving the model from microeconomics and use the enterprise data. By plugging in data aggregated into industries, the model of course shifts towards macroeconomics, but we will argue that this is both a reasonable and necessary simplification.

5.2 Small and Medium Enterprises

5.2.1 Definition

Simple though it appears, the mere size criterion can still be relevant for economic analysis. We shall illustrate it in the following section, but prior to that we give proper definition of

[1] Exceptions are twofold. Post-soviet countries that have not yet undergone full transition show negligible SME sectors, eg Belarus, Georgia or Ukraine. However in this countries lots of prospective entrepreneurs take part in the informal economy, not captured by official statistics, so that the true percentage is higher. The other exception is the USA, where the share on labour force is 53 per cent, but with 500 employees as the yardstick. It illustrates that the world's biggest economy has quite different dimensions than Europe.

the term SME.

Small and medium enterprises, abbreviated as *SME*[2], are defined as companies not exceeding specific size limits. The official definition of the European Union is given in table 5.2. It is not a clearly disjunctive definition, if related to employment only. The complication emanated from the fact that in the EU SME have become an important tool for economic policy measure. Note that a firm must satisfy the first condition and either one of the last two conditions at the same time in order to be classified as SME.

The simplest classification, such as that of the World Bank (WB), relies solely on the number of employees—the WB uses 250 as the limit. Lots of countries created their own definitions, eg Switzerland or the USA take 500 employees as the cutoff.[3]

5.2.2 SME around the World

According to a widespread argument, a strong sector of competitive small and medium enterprises heavily depends on the quality of business environment. As is known, economic institutions can play a double role: they can both improve and impede the efficiency of markets and thus increase or decrease the competitiveness of firms. Some sources point to the negative role of state bureaucracy—but bureaucracy is only one of a dozen factors selected by the World Bank that impinge on the efficiency of firms.

We can think of several intuitive arguments which support this hypothesis: In the first place, complicated bureaucracy acts as a sure deterrent to start up a business at all, just as does persistence of organized crime or anti-competitive practices. Secondly, additional costs incurred due to obstacles to business form the larger share of a company's costs the smaller the company actually is, so that smaller firms are harmed more. Thirdly, chaotic economic institutions add to overall uncertainty in doing business, against which it is harder to hedge for smaller firms than for larger firms.

Two studies have addressed the issue of firm's size and institutional setting on the global scale. Schiffer & Weder [102] explored the hypothesis that size explains part of the variance in responses to the World Bank global survey of business environment. Companies were asked to judge the severity of the following obstacles: (1) Financing, (2) infrastructure, (3) taxes and regulations, (4) policy instability or uncertainty, (5) inflation, (6) exchange rate, (7) functioning of the judiciary, (8) corruption, (9) street crime, theft or disorder, (10) organized crime or mafia, and (11) anti-competitive practices by government or private enterprises.

In the overall sample of roughly 10.000 firms, the authors found that small firms on average viewed the obstacles to doing business as more severe than large firms, ie SME perceived more obstacles than large firms. The significance of this strong finding declined

[2]Sometimes the abbreviation 'SME' stands for 'small and medium-sized entrepreneurs'.

[3]Large portions of subsidies which are distributed to SME each year in the EU can truly lead to heated political debates over the definition of size categories. Recently France proposed to extend the mark to 500 employees, presumably to create a loophole for state aid to larger companies—which is otherwise banned by the European law.

(though not disappeared) when they split up the sample to regional groups[4]. In particular, in the ten OECD countries firms report the same level of obstacles irrespective of their size. Yet the effect remained significant in two regions: Latin America and the Carribean; and transition economies, where Czech Republic belonged to at the time of the survey in 1990's.

Schiffer & Weder do not elaborate much on the fact that the significance of the finding varied within the regional subsamples. But they might have overlooked the quite considerable implication of their results: namely that more of the "free market" leads to less "size discrimination". In other words, this would support the argument that liberal market reforms do equalize conditions for market players and that SME deserve special political treatment—how to cut the red tape in order to open SME to competitive (and more efficient) markets.

Another study by Ayyagari *et al.* [6] tested two mutually exclusive hypotheses. Firstly, large SME sectors may stem out from high exit costs and government subsidies, so that they are prevented to grow or to disappear (negative reasoning). Alternatively they argued that large amount of SME could result from low barriers to entry and better credit availability (positive reasoning). The authors test a large cross-country dataset from the 1990's. They do not find any conclusive support for the former hypothesis, but a significant backing for the latter.

We can translate the result of their investigation to a simple imperative: governments must not crack down on natural entrepreneurship if they want to foster a thriving SME sector. Their finding also contains another dimension: The study suggests that financial aspects (entry costs and credit availability) matter more for creation of SME than other institutional factors do.

It is of our concern throughout this paper to quantify the conditions of efficiency, which are of vital importance for the advancement of small and medium-sized firms in the environment of globalised competition. I.e. in cases when competition to SME comes both from outside and inside.

Entrepreneurs often start from scratch and thus embody the ability and will to learn and create. In his case study on Turkey, Taymaz [110] concludes that "most firms start small", moreover, they are most often challenged there by their distinct systemic disadvantage: both their scale and efficiency are suboptimal. It follows that these businesses have to achieve higher rates of growth in order to survive, notwithstanding their resultant lower profits or lower wages. That small firms grow faster is exactly the finding of the recent study by Mohnen & Nasev [89], who analyzed German SME. Taymaz notes that the Schumpeterian selection process is quite drastic, given high mortality rate among entrants.

The rate of technical change gained much attention in economic literature, but productivity dynamics of this kind is not the subject of our paper and we skip more details at this point.

[4]Groups were as follows: OECD countries; transition economies; Latin America & the Carribean; East Asia & Pacific; South Asia; Africa.

5.2.3 Czech SME Sector: Foundations so Tiny

Until 1989, Czechoslovakia had one of the toughest regime concerning private enterprise among the communist countries. Private businesses were violently nationalized or collectivized in the 1950s. Any entrepreneurial activities were forced to the informal economy. The prompt revival of the SME sector in Czechoslovakia in the first years after the fall of the "iron curtain" is thoroughly analysed in the study by Benáček [10]. Benáček claims that from the start the impact of the emerging small ventures, both legal and informal, was largely underestimated by official statistics and substantially contributed to an economically smooth transition.

Table 5.3 quotes statistics on SME published by the Ministry of industry and trade in its "Report on the development of SME and its support in 2006".[5] We make several straightforward observations. SME account for one third of the Czech GDP and for close to two thirds of employment. This share remained more or less stable over the last ten years. This holds for the accounting value added as well, which stayed close to 53 per cent throughout the ten years.

On the contrary, three indicators changed significantly and suggest that the SME sector has come through an intensive consolidation. On one hand its share on exports and imports has gone up seven and six percentage points respectively, meaning that SME are now more involved in international trade. Moreover the breakthrough appears around the years 2004/2005, when Czech Republic entered the EU. On the other hand, SME invest relatively more, or they are rather correcting the underinvestment from the earlier period.

The following chapters will focus on cross-sectional analysis. Besides structural results, we are particularly concerned with what stands behind the table 5.3. We estimate the SME-specific production function, derived from the microeconomic background, to reveal the sensitivity of productive inputs. Above all, we investigate the relationship between labour, capital and investment. We noted in relation with table 5.3 that SME have recently experienced a massive investment surge, which should result in higher capital endowment and better productivity. This effects on efficiency scores are handled by a separate model in the last part of the paper.[6]

5.3 A Stochastic Model of the Production Function

5.3.1 A Model of Production

The starting point of our analysis is the neoclassical production function.[7] We consider a p-dimensional vector of inputs $\boldsymbol{x} \in \Re^p_{0,+}$ and an r-dimensional vector of outputs $\boldsymbol{y} \in \Re^r_{0,+}$.

[5] Downloaded at [http://www.mpo.cz/dokument32006.html] on January 5, 2008.

[6] However we are also aware that table 5.3 is related to a broader group of SME than that represented by our dataset.

[7] Detailed concepts of production functions can be found in Sato [100] and Johansen [64], a condensed overview in Nadiri [92].

Production function characterizes the technology available to the firm and describes how all inputs inclusive capital, labour, land, materials, know-how etc. are transformed into outputs. This is written as

$$\boldsymbol{y} = f(\boldsymbol{x}),$$

so that $f(\boldsymbol{x})$ represents the complete technical relationship between inputs as well as between inputs and outputs. We assume $f(\boldsymbol{x})$ to have all standard properties; we mention their application later in the text.

Production function defines only one part of the economic world: It constitutes the constraint subject to which every firm has to operate. The other part consists of preferences and scarcity and is captured in prices of inputs and outputs. Because firms act so as to maximize their profits, knowledge of $f(\boldsymbol{x})$ would not be sufficient for economic analysis. The way we track prices is discussed in section 5.3.3

Production function is an ideal concept when no frictions exist. In real world inefficiencies occur and not all producers are able to reach the maximum possible output: $\boldsymbol{y} \leq f(\boldsymbol{x})$, meaning that some firms will operate inside the area constrained by the production function. Furthermore once a firm achieves maximum output it can still be inefficient in terms of costs, revenues or profits, since it may use a combination of inputs and/or produce a vector of outputs that do not maximize the profit at prevailing market prices.

Both technical and economic inefficiency are now widely used concepts in economics, but their exposition is not the purpose of this paper. In our analysis we mark those industries as inefficient which simply do not achieve the best practice.

5.3.2 Estimator of Technical Efficiency

In the next sections, we build a framework to estimate $f(\boldsymbol{x})$ and the extent of inefficiency among SME. The method we use is called stochastic frontier analysis, or SFA. We adapt the model for technical efficiency from Kumbhakar & Lovell [72, equation 3.2.18]:

$$y_i \;=\; f(\boldsymbol{x_i}, \boldsymbol{\beta}) \cdot \exp\left\{\nu_i - \tau_i\right\}, \tag{5.1}$$ [8]

where $\boldsymbol{\beta}$ is the vector of parameters of $f(\cdot)$, ν_i is the random disturbance term and $\tau_i \geq 0$ is the inefficiency term. A general unknown production function is adjusted to differently productive firms by a multiplicative inefficiency term τ which is of one sign only. With data in the form of matrices X, Y, we look for estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\tau}_i$.

This approach was pioneered by Aigner & Chu [2], Aigner, Lovell & Schmidt and Meeusen & van den Broeck and is thoroughly depicted in Kumbhakar & Lovell [72, p. 72-81]. The initial production function is assumed to be:

$$f(\boldsymbol{x_i}, \boldsymbol{\beta}) = \exp\left\{\beta_0\right\} \cdot \prod_{j=1}^{p} x_{ij}^{\beta_j}.$$

[8] Throughout this section we assume only one output.

Then we can rewrite the model (5.1) as the log-linear Cobb-Douglas production function to obtain:

$$\begin{aligned} \log y_i &= \beta_0 + \sum_{j=1}^{p} \beta_j \log x_{ij} + (\nu_i - \tau_i) \\ \nu_i &\ \ldots\ iid,\ \mathcal{L}(\nu_i) \sim \mathcal{N}(0, \sigma_\nu^2). \end{aligned} \tag{5.2}$$

Ordinary least squares (OLS) yield estimates of β_js, but we would also like to separate β_0, ν_i, τ_i to obtain producer-specific efficiency scores. In order to get these, we need two additional assumptions: (1) τ_i is *iid*, $\mathcal{L}(\tau_i) \sim \mathcal{N}^+(0, \sigma_\tau^2)$, and (2) τ_i and ν_i are independent on each other and on the regressors. Denoting $\epsilon_i = \nu_i - \tau_i$, $\sigma = \sqrt{\sigma_\tau^2 + \sigma_\nu^2}$ and $\lambda = \frac{\sigma_\tau}{\sigma_\nu}$, the following maximum likelihood estimator (MLE) can be derived:

$$\mathscr{L}(\epsilon, \sigma^2, \lambda) = \prod_{i=1}^{n} \frac{2}{\sigma\sqrt{2\pi}} \cdot \exp\left\{-\frac{\epsilon_i^2}{2\sigma^2}\right\} \cdot \Phi\left(-\frac{\epsilon_i \lambda}{\sigma}\right), \tag{5.3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal density and distribution functions. Further it is possible to derive the conditional distribution of τ_i given ϵ_i, which is truncated normal:

$$g(\tau_i|\epsilon_i) \sim \mathcal{N}^+\left(\mu_{*i} = -\frac{\epsilon_i \sigma_\tau^2}{\sigma^2}; \sigma_*^2 = \frac{\sigma_\tau^2 \sigma_\nu^2}{\sigma^2}\right).$$

From this distribution point estimators of τ_i can be obtained as either the mean $E(\tau_i|\epsilon_i)$ or the median $M(\tau_i|\epsilon_i)$. Both results have to be transformed back to the exponential form of (5.1) to obtain the estimate of technical efficiency $\widehat{\mathcal{E}}_i = \exp\{-E(\tau_i|\epsilon_i)\}$, the same holds for the mode. One more complication is that we assumed inefficiency to have multiplicative form, which we then transformed by taking logarithm. Thus it makes more sense to construct an estimator which is based on efficiency already transformed back, ie in exponential form. In other words we can write:

$$\exp\{-E(\tau_i|\epsilon_i)\} \neq E(\exp\{-\tau_i\}|\epsilon_i).$$

This problem was resolved by Battese & Coelli and is mentioned by Kumbhakar & Lovell [72]. Instead of mean or mode, they proposed an improved point estimator:[9]

$$E(\exp\{-\tau_i\}|\epsilon_i) = \left[\frac{1 - \Phi\left(\sigma_* - \frac{\mu_{*i}}{\sigma_*}\right)}{1 - \Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)}\right] \cdot \exp\left\{-\mu_{*i} + \frac{1}{2}\sigma_*^2\right\}. \tag{5.4}$$

5.3.3 The Economic Dimension

Deliberations in section 5.3.2 deal just with technology. In section 5.3.1 we stressed that economics primarily focuses on allocation efficiency. Hence we ought to analyze revenue and cost functions to capture economic performance of SME.

[9] In the original article [8], the authors defined the estimator for panel data, where τ_i was constant over time but ν_{it} was allowed to vary among periods.

The underlying idea for derivation of an estimator is similar to the previous section, only the algebra is more complicated. A comprehensive overview can be found in Kumbhakar & Lovell [72]. But we do not pursue their exposition because we do not have sufficient data to apply it, since our dataset is given in monetary units[10]. With such data we cannot construct profit functions and decompose overall efficiency into technical and allocation efficiency. Instead we have to give up some of the microeconomic detail.

This move is less drastic that it appears. It is difficult to imagine that even with the most detailed data for individual firms we could reasonably use the standard construct of uniform exogenous input and output prices. Today with incredibly diverse forms of capital we always have to aggregate to a certain degree.

As a consequence our empirical analysis will plug in aggregated data, so that our estimate which we denoted $\widehat{\mathcal{E}}_i$ includes both components, ie it measures overall efficiency. In other words, we estimate a profit function, although our starting point (5.2) is not a proper profit function, but a production function. We show that this simplification can still deliver interesting results.

5.4 Efficiency of Czech SME

5.4.1 Data Description

Czech Statistical Office (CSU) publishes a yearly summary on economic activity of Czech small and medium enterprises, which can be found under reference number 8007-[year]. This publication contains several indicators along with condensed size and sector groups.

These data are obtained by a statistical enquiry, which covers all firms with 100 or more employees, 55 per cent of companies with 10–99 employees and about 2,6 per cent of the micro-segment (below 10 employees). Individual data are aggregated and are not made available.

Following an official request, CSU provided us with slightly more detailed data than one can find in the publicly available booklet. Our dataset has four dimensions:

1. thirty-item two-digit OKEC[11] classification, including OKEC codes 10 to 41[12], i.e. agriculture and services are not included;

2. size classification with breakdowns at the following number of employees: 0-10-20-50-100-250;

[10] The only exception is one of the proxies for labour in the production function: number of employees.

[11] European Union uses the abbreviation NACE: Nomenclature Générale des Activités Économiques dans les Communautés Européennes.

[12] OKEC 12 is not included. Full list of industries is available at [http://www.czso.cz/csu/klasifik.nsf-/i/odvetvova_klasifikace_ekonomickych_cinnosti_(okec)] in Czech or at [http://ec.europa.eu/comm/-competition/mergers/cases/index/nace_all.html] in English.

3. eleven economic indicators listed in table 5.4, for complete definitions of indicators refer to appendix 5.6;

4. years 2002 through 2005, so that the short time span restricts us to cross-sectional analysis, ie we will assume that technology did not change in time.

The dataset turns our initial estimation idea to:

$$\begin{aligned} y_i &= f(\boldsymbol{x_i};\boldsymbol{\beta}) \\ VAD_i &= f(TAS_i, IAS_i, INV_i, AEM_i, PAY_i, OPE_i;\boldsymbol{\beta}), \end{aligned} \tag{5.5}$$

with $n^{(2002)} = 135$, $n^{(2003)} = 135$, $n^{(2004)} = 134$ and $n^{(2005)} = 136$, totalling 540 observations.

5.4.2 Estimation of the Parameters: SFA Results

5.4.2.1 Indentifying a Model

Stochastic frontier analysis can yield twofold distinct results: (1) Sensitivity of factors of production, ie estimation of $\boldsymbol{\beta}$; and (2) estimation of individual efficiency scores. We first concentrate on the former point, which can be consistently solved by standard ordinary least squares (OLS). The latter point is investigated in section 5.4.3 by means of MLE.

Equation (5.5) for our variables gives the formula:

$$VAD_i = \beta_0 \cdot TAS_i^{\beta_1} \cdot IAS_i^{\beta_2} \cdot INV_i^{\beta_3} \cdot AEM_i^{\beta_4} \cdot PAY_i^{\beta_5} \cdot OPE_i^{\beta_6} \cdot \exp\{v_i - \tau_i\}, \tag{5.6}$$

which has to be adjusted in several ways. Firstly, we expect absolute values to result in heteroskedasticity, i.e. non-constant variance. This was confirmed by our preliminary tests, hence we normed the variables by output.

Secondly, INV include acquisition of assets whose period of usage is longer than one year and that usually take some time to be realized. Accordingly we use lagged values $INV_{-1,i}$, so that we will only be able to model just three years out of the four, 2003 through 2005.

By construction $TAS_{t,i}$ and $IAS_{t,i}$ include lagged investment $INV_{t-1,i}$, however we consider it meaningful to include investment in equation (5.6). From the theory point of view, our approach is based on a microeconomic model of production, where investment captures readiness of the firm to upgrade its assets and hence to build and/or maintain its *competitive advantage*. Our idea is to interpret INV as a general indicator of innovation, and not as a component of 'capital' in the macroeconomic sense.[13]

Statistically, such dependence between regressors might produce multicollinearity, i.e. near-singularity of the matrix of regressors. However this adverse effect is mitigated because the relationship between $TAS_{t,i}$, $IAS_{t,i}$ and lagged investment is additive, while equa-

[13] Note that all business professionals talk about the firm being able to find a niche market segment where they have enough pricing power; product homogeneity is strictly avoided. Our model asks how successful SMEs are at this effort.

tion (5.6) is multiplicative-exponential. It follows that there is no straightforward linear relation between the regressors.[14]

We tested multicollinearity by means of matrix condition number, as suggested by Greene [53, p. 57]. For 2004 data, lagged investment only added approximately 8% to the dataset condition number and we concluded that INV should not cause multicollinearity issues in the dataset.[15] At this point, we also tested possible multicollinearity due to similarity of PAY and OPE. Since OPE increased the condition number by 850%, we decided to drop it from equation (5.6).[16]

Finally, we can think of PAY as the multiple of AEM and average pay per employee, AWG.[17] By including $\left(\frac{AEM_i}{OUT_i}\right)^{\beta_4} \cdot \left(\frac{PAY_i}{OUT_i}\right)^{\beta_5}$, we would in fact count AEM twice. Instead of PAY, we include two separate variables AEM and AWG because we want to separate the effect of the amount of employed labour β_4 and the effect of its quality β_5.

These considerations yield the following model:

$$\frac{VAD_i}{OUT_i} = \beta_0 \cdot \left(\frac{TAS_i}{OUT_i}\right)^{\beta_1} \cdot \left(\frac{IAS_i}{OUT_i}\right)^{\beta_2} \cdot \left(\frac{INV_{-1,i}}{OUT_{-1,i}}\right)^{\beta_3} \cdot \left(\frac{AEM_i}{OUT_i}\right)^{\beta_4} \cdot \cdot \left(\frac{AWG_i}{OUT_i}\right)^{\beta_5} \cdot \exp\{\epsilon_i\}.^{18} \tag{5.7}$$

This production function is of course estimated in its log-linear form.

[14]While TAS_t is linearly proportional to INV_{t-1}, $\log TAS_t$ is not linearly related to $\log INV_{t-1}$. For a more general mathematical interpretation see below the footnote at equation (5.7).

[15]Note that Kennedy [66, p. 181] mentiones the example of Cobb-Douglas production function, where

> the inputs capital and labour are highly collinear, but nonetheless good estimates are obtained.

He recommends to do nothing if t-statistics are greater than 2.

[16]Data for 2003 and 2005 yielded the same results. I am grateful to one anonymous referee for this suggestion.

[17]We obtained AWG as PAY divided by AEM.

[18]It must be stressed that the proposed equation is *not* derived from a macroeconomic production function, which gives output in monetary terms. As the discussion in section 5.3.3 points out, our initial model is based on purely microeconomic model of production, where the production function returns physical units of output. In this framework, any applied work would then aim at a profit function in the ideal case. However, in practice we have to combine this microeconomic basis with an aggregate approach where the function is money-valued. Above all, we use the multiplicative Cobb-Douglas production function.

The proposed equation reflects the cost structure of firms. Accounting value added is the value of output less cost of materials (see appendix 5.6). The inputs on the right hand side are: (1) share of capital on output; (2) share of labour on output, which is also money-valued, ie multiplied by payroll; and finally (3) investment.

The third component captures the share of investment on the output. The idea here is to provide a rough measure of how much firms innovate. Therefore an alternative specification would include the share of investment on capital (ie capital renewal) rather than on output, so that we would like to have $\left(\frac{K}{Y}\right)^a \left(\frac{I}{K}\right)^b$ instead of $\left(\frac{K}{Y}\right)^a \left(\frac{I}{Y}\right)^b$. But consider the following algebra with obvious simplified notation:

$$\left(\frac{K}{Y}\right)^a \left(\frac{I}{Y}\right)^b = \left(\frac{K}{Y}\right)^a \left(\frac{I}{Y}\right)^b \left(\frac{K}{K}\right)^b = \left(\frac{K}{Y}\right)^{a+b} \left(\frac{I}{K}\right)^b,$$

so that the results of our specification can also be readily interpreted in this manner. Of course this derivation is inaccurate for past investment, since past investment is counted in present capital. But our claim is that we can stick to the simple specification because of the simple intuition behind and because it performs reasonably

5.4.2.2 Standard Regression

We evaluate ordinary least squares regression for (5.7) using 2005 data and perform standard diagnostics. Then we proceed to panel data regression. Models are evaluated in the statistical package $\mathcal{R}$.[19] In order to get a balanced panel, i.e. to have exactly the same industries in each of the years, we had to cross some more rows to get $3 \times 131 = 393$ observations.

Results are presented in table 5.5. R-squared indicates that we are able to explain 22 per cent of the original data variation. We rejected the hypothesis of normally distributed residuals.[20] As for homoskedasticity, we assumed that nonspherical disturbances could stem from the size variation and use Breusch-Pagan to test for this effect. In the case when residuals are not normally distributed, we use the studentized Breusch-Pagan test, as advised by Kennedy [66, p. 130]. Model (5.7) yields p-value 0.4118 in 2005. We infer that the source of the presumably largest problems with non-constant variance is not statistically significant.

We estimate equation (5.7) in its panel form for years 2003 through 2005. We first evaluate both fixed effects and random effects specifications[21] and use the Hausman test to choose the preferred model. Under the null hypothesis the more efficient random effects model is selected. For our data p-value is lower than $2 \cdot 10^{-16}$, therefore the null is strictly rejected.

In addition, fixed effects can be justified on theoretical grounds: In the case of production function it is most likely that individual effects are correlated with the regressors, meaning that random effects are not suitable for estimation. We acknowledge one important disadvantage of the fixed effects approach which concerns large number of group dummies relative to the number of time periods. For completeness we report the random effects specification in table 5.7.

The fixed effects estimation is shown in table 5.6. We analyse individual efficiency scores in more detail in the following sections, therefore we do not report group effects; let us merely note here that 23 group dummies are significant at least at the 5% level.[22]

5.4.2.3 Production Function of Czech SME

One way to interpret table 5.6 is the significance of estimated parameters. Two variables are of lower significance: tangible and intangible assets.

Insignificance of intangible assets means that no high-tech revolution occurred in the

well with the data. Finally it must be noted that there is vast room for testing of numerous changes in the model specification.

[19] For more information refer to the webpage [http://www.r-project.org/].

[20] Among the plenty of tests available, we decided for Kolmogorov-Smirnov and Shapiro-Wilk. p-values are 0.000342 and 0.000029 respectively.

[21] We use the plm package in $\mathcal{R}$.

[22] By comparing the results of OLS and fixed effects estimation, it appears that what OLS shows to be the significant effect of tangible assets (β_{TAS}) is then absorbed into the significant group effects. However, the fixed effects estimation seems to give a more precise estimates for the impact of labour on efficiency (β_{AEM} and β_{AWG}). This means that efficiency of capital assets tends to be group (industry) specific while efficiency of labour is largely driven on firm level.

past years in the SME sector, and that SME do not properly exploit patents or trademarks. It is quite clear that the production of vast majority of small and medium companies is not based on leading-edge or even innovative technology. Nonetheless, we would expect that once a company possesses any of these, we would recognize it as an advantage.

We find more surprising that tangible assets are insignificant. From a simple perspective it means that SME are more dependent on labour than on capital. Our result may also point to inefficient use of capital, since we must distinguish between accumulated capacity and relative employment of this capacity.

Nevertheless we have to bear in mind two problems with the definition capital as such. Ideally we would like to measure capital as a flow, but our variable captures accumulated stock of capital. Moreover measurement of capital is likely to contain inaccuracies, especially for the smallest businesses. This is because the relative weight of fixed write-offs and similar accounting practices is higher and omissions more probable than for larger firms.

The other point of view is the sign of the estimated coefficients, which must be interpreted carefully. We know that investment, defined as acquisition of long term assets, is at first a considerable expense, which should later turn profitable. The negative coefficient on lagged investment implies two possible hypotheses: (1) SME have not managed to create competitive advantage out of investment, i.e. their investment is not efficient. (2) Investment turns profitable only after a few years, which makes it especially troublesome for SME to spend money there. In other words, the investment surge among small and medium enterprises documented in table 5.3 has not yet generated positive revenue. To distinguish between the two scenarios we would need a longer panel, but we believe that the real scenario consists of both effects.

We find that high average wage has positive influence on value added, while large number of employees impacts production adversely. Since the theory says that differences in wages should reflect quality of the workforce, higher salaries should result in more value added. Hence the mechanism we expect follows the relationship high skill — high wage — high value added and vice versa. Profit maximizing firms have to find a balance between high wages and high value added, because wages are costs. In our multiplicative specification of the cost function where labour input is money valued, both the number of employees and their payroll contribute to the value added. The employer wants of course to keep both AEM and AWG low while maintaining high value added, and we want to find out which one is kept relatively lower.

Our regression cannot confirm that higher wages alone imply higher value added. On the contrary, the joint result on AEM and AWG tells us much about firms' cost minimization technique: The coefficient of AEM is negative and of AWG is positive, and hence (following the result above) we conclude that firms rather choose small numbers of high-wage workers for a given level of value added.

However the same comment applies here as in the case of capital: Since AWG is a cumulative yearly wage and not a standardized hourly wage, it cannot be considered a proper flow variable. Therefore our result might simply mean that for a firm it can be cheaper to pay its current employees a low marginal wage for overtime than hiring a new employee with high additional labour costs.

We sum up the discussion in a proposition:

Proposition 5.4.1 ***SME production function characteristics.***

- *Our regressions confirm that value generation of Czech SME depends more on labour than on capital. Tangible assets are not significant in production, however this result might be biased due to measurement and reporting errors.*
- *SME are not able to reap the benefits of intangible assets, such as software or patents.*
- *SME are optimizing labour costs by paying higher wages to less employees rather than the opposite.*
- *The effect of previous investment turns out negative: Investment is either inefficient, or requires a longer payback period, or both.*
- *The above points mean that by large, SME fundamentals of the Czech economy have not yet converted to an innovation based production process.*

5.4.3 Effects of Size and Time

5.4.3.1 Model Specification

In this section we focus on the estimation of efficiency of individual industries by using the parametric approach. As shown in section 5.3.2, the starting point is now the equation (5.3).[23] The solution of this maximum likelihood maximization is implemented in the freely available program FRONTIER by T. Coelli [28], which moreover offers several extension to this basic model.

FRONTIER is able to compute two specifications. The first reads for time $t = 1, \ldots, T$:

$$\begin{aligned} y_{it} &= \beta_0 + \sum_{j=1}^{p} \beta_j \cdot x_{ijt} + (\nu_{it} - \tau_{it}) \\ \nu_{it} &\ldots\ iid,\ \mathcal{L}(\nu_{it}) \sim \mathcal{N}(0, \sigma_\nu^2), \\ \tau_{it} &= \tau_i \cdot \exp\{-\eta \cdot (t - T)\}, \\ \tau_i &\ldots\ iid,\ \text{truncations at zero of } \mathcal{N}(\mu_\tau, \sigma_\tau^2). \end{aligned} \tag{5.8}$$

As in the previous text, we would plug in logarithms of the data rows.

[23] Note that the maximum likelihood estimation in this section is based on strict distributional assumptions, in particular that the random disturbance term and the efficency term in equation (5.3) are not correlated.

The second specification available in FRONTIER for $t = 1, \ldots, T$:

$$\begin{aligned} y_{it} &= \beta_0 + \sum_{j=1}^{p} \beta_j \cdot x_{ijt} + (\nu_{it} - \tau_{it}) \\ \nu_{it} &\ldots\; iid,\ \mathcal{L}(\nu_{it}) \sim \mathcal{N}(0, \sigma_\nu^2), \\ \tau_i &\ldots\; iid, \text{ truncations at zero of } \mathcal{N}(\xi_{it}, \sigma_\nu^2), \\ \xi_{it} &= \delta_0 + \sum_{h=1}^{d} \delta_h \cdot z_{it}, \end{aligned} \tag{5.9}$$

the data again being logarithms. This specification allows the inefficiency to be modelled by other factors than time, meaning that z are variables influencing efficiency and δ is the respective vector of parameters to be estimated. The idea is simple: once we have estimated efficiency, we would like to explain it and run a second-step regression. Yet a more efficient procedure is to estimate both parameter vectors in a single step, as is done by FRONTIER.

Coelli [28] remarks that the models (5.8) and (5.9) are not nested, so they cannot be tested against each other.

What deserves special attention is the distribution of the inefficiency term τ. In section 5.3.2 we assumed $\mathcal{L}(\tau_i) \sim \mathcal{N}^+(0, \sigma_\tau^2)$, so that the distribution was half normal. In equations (5.8) and (5.9) we specify the distribution to be truncation of a non-central normal distribution, though still at zero. The impact on implementation is modest, since it only results in a more complicated likelihood function. Yet it considerably modifies the modelling framework.

With $\mu_\tau = 0$, most firms should lie on, or be close to, the efficient frontier, since inefficiency is close to white noise. On the contrary, if the underlying density is modelled as non-central with $\mu_\tau \neq 0$, the centre of gravity is moving towards inefficiency. By this we in fact allege that there is systematic inefficiency, which we can track either by time or by particular explanatory variables z_d. In other words, we claim that the best practice and the common practice are not identical.

Before we turn to tests of hypotheses, we list table 5.8, which contains quartile statistics for the MLE estimates of efficiency scores using (5.8), where μ_τ and η are included in the estimation. The scores are joint for 2003 to 2005.

5.4.3.2 Estimating the Common Practice

We apply the model (5.8), where we specify x to be the same six variables as in (5.7). We use $t = 2003, 2004, 2005$. Let us formulate the first hypothesis to test: Under the null $\mu_\tau = 0$, under the alternative $\mu_\tau \neq 0$, i.e. inefficiency is significantly different from white noise. In FRONTIER we solve (5.8), where we also include η.

We get $\widehat{\mu}_\tau = -1.3484$ and estimated standard error 0.4906. The resulting t-ratio -2.7483 yields p-value 0.006 and we reject the null on 0.001 significance level.[24]

[24] p-value was computed as: $p = \Pr(\mathcal{T} \geq |\hat{t}|)$, where the random variable $\mathcal{T}$ is governed by t distribution.

From $\widehat{\mu}_\tau$ we can compute estimated mean inefficiency as the mean of the truncated normal distribution $\mathcal{N}(\widehat{\mu}_\tau, \widehat{\sigma}_\tau^2, a = 0)$, where a is the point of truncation and $\widehat{\sigma}_\tau^2 = 0.4546$. From Greene [53, p. 759] we have:

$$\begin{aligned} \mathrm{E}\left[\tau_i\right] = \mathrm{E}\left[\mathcal{N}(\widehat{\mu}_\tau, \widehat{\sigma}_\tau^2, a = 0)\right] &= \widehat{\mu}_\tau + \widehat{\sigma}_\tau \cdot \frac{\phi\left(\frac{a-\widehat{\mu}_\tau}{\widehat{\sigma}_\tau}\right)}{1 - \Phi\left(\frac{a-\widehat{\mu}_\tau,}{\widehat{\sigma}_\tau}\right)} \\ &= 0.2516, \\ \mathrm{E}\left[\widehat{\mathcal{E}}_i\right] &= \exp\{-0.2516\} = 0.7776, \end{aligned}$$

so that the expected mean efficiency level is 78%.

Proposition 5.4.2 ***Systematic inefficiency.*** *The production function of Czech small and medium enterprises (5.7) is likely to contain systematic inefficiency, meaning that the common practice is significantly different from the best practice.*

5.4.3.3 Testing Size Effects

We would like to test the hypothesis that there is a significant relationship between size and efficiency. By size we mean the SME definition in terms of employees.

We use EGR, the group according to the number of employees, as the single z variable in model (5.9). Hence $EGR \in \{1, 2, 3, 4, 5\}$, as corresponds to breakdown points 0-10-20-50-100-250. Under the null, $\delta_1 = 0$, so that the effect of EGR is not significant. In FRONTIER we compute $\hat{\delta}_1 = 0.9620$ with corresponding standard error 0.2691 and t-statistic 3.5755. Using the same formula as in the previous tests, the p-value is 0.0004 and we reject the null on 5% significance level.

The coefficient means that jumping from a smaller size group to a larger one increases ξ by 0.962. The impact on mean inefficiency is not straightforward, because the mean has to be computed as in section 5.4.3.2. The impact due to change in ξ is $\exp\{-0.962\} = 0.07$, i.e. seven per cent. We omit detailed computation of the derivative because we believe that one should not rely too much on estimated magnitudes of coefficients.

Proposition 5.4.3 ***Size effect.*** *Larger firms tend to be more efficient.*

5.4.3.4 Testing Time Effects

Our last test checks the presence of a significant time effect. We are aware that our specification does not rely on an advanced model for technical change. Moreover since we use data in monetary units, we are estimating the combined effect of technical *and* allocation efficiency anyway. We only want to test whether efficiency scores are different among years.

We use the specification from equation (5.8) as in the previous section, where we include μ_τ. Under the null, $\eta = 0$. The result is $\hat{\eta} = 0.0771$ with standard error 0.0348 and t-ratio

Contrary to the fixed effects panel regression, there are only ten parameters in the maximum likelihood estimation, so there are 393 − 10 degrees of freedom.

2.2114. As in the previous section, we compute p-value 0.028 and reject the null on 0.05 significance level.

To get an idea about the magnitude of this estimated effect, we use the estimates in equation (5.8). The log inefficiency index fell from $\exp\{2 \cdot 0.0771\} \times \tau_i$ in 2003 to $1 \times \tau_i$ in 2005, i.e. by about 17%. The estimated time effect from 2003 to 2005 was $\frac{\exp\{\tau_i\}}{\exp\{1.17\tau_i\}} = \exp\{(1-1.17)\} \approx 0.85$, which means that inefficiency factor approximately 15%.

Proposition 5.4.4 ***Time effect.*** *In the course of the three observed years, inefficiency of Czech SME decreased by about 15 per cent.*

5.5 Conclusions

Stochastic frontier analysis (SFA) works as an enhanced regression; it looks for the parameters which govern the production process and then estimates efficiency as the distance between fitted and observed values. We used it to find out the characteristics of the production function of Czech small and medium manufacturing enterprises (i.e., agriculture and services are not included).

We were able to gather several propositions about Czech manufacturing SME in the form of stylized facts. SFA estimates production and profit functions and due to its statistical formulation, it offers procedures for testing hypotheses. We used ordinary least squares and the accompanying standardized diagnostics to find out an acceptable specification which we then handed over to panel data regression and maximum likelihood estimators.

We confirmed that Czech manufacturing SME are more dependent on labour employment rather than on capital usage. Tangible assets do not contribute to value added in our model, and the same holds for intangible assets, which include goodwill, software, patents, copyrights, trademarks and tradenames. That their presence in the production function is insignificant means that Czech manufacturing SME are not yet innovation driven on a large scale. Further, we found that investment has a negative impact on production value, so the investment surge documented in table 5.3 was not mirrored in our estimates. There are at least two feasible explanations: (1) It takes more time for the investments to start generating profits, so that we would need longer dataset to capture this effect. (2) SME have not managed to create competitive advantage out of investment, i.e. their investment is not efficient. We note that these points are not mutually exclusive, hence in fact both of them can be partially true.

At last we performed three specific tests: We found the presence of systematic inefficiency highly significant. By this we mean that instead of being close to the efficient frontier, the majority of firms operate on a lower level of efficiency. We explained this effect by the size of enterprises; measured in terms of number of employees, larger firms tend to be more efficient. Finally our results suggest that inefficiency of Czech SMEs decreased between 2003 and 2005.

As with every empirical study, we are well aware of the fact that practice requires compromise. One that we encountered throughout the paper was as follows: The methods are constructed to trace technical and allocation efficiency separately, but we have to use data in monetary units which disables the separation of the two effects. Still we argued that it does not hinder us from using them and that they are capable of yielding meaningful results.

We are aware that this analysis could well be extended, an example of which is an explicit treatment of technological progress. This issue is left for further research.

5.6 Appendix: Data Definition

We give complete definitions of the data obtained from the Czech Statistical office. These definitions are available online.[25] For reference purposes, we list both the Czech expression and the English translation.

Gross profit on merchandise sold = revenue from goods acquired for resale less costs of resold merchandise.

1. **Number of active firms** (počet aktivních podniků).
   Number of firms which were active at least on one day during the reference period.
2. **Output** (výkony celkové).
   Sum of: (1) sales revenue from own products, (2) gross profit on merchandise sold (3) received leasing installments, (4) change in inventories and (5) self-constructed asset revenue.
3. **Sales revenue** (tržby za vlastní výkony a zboží).
   Sum of: (1) sales revenue from own products and (2) revenue from merchandise sold.
4. **Accounting value added** (účetní přidaná hodnota).
   Output less cost of materials used in manufacturing. The latter consists of (1) the value of purchased and already used material, energy and of supplied materials which are not storable, and (2) of the value of purchased services.
5. **Tangible assets** (dlouhodobý hmotný majetek).
   Includes mainly land, plants, capital equipment, orchards and vineyards, herd and draught animals and all other assets with supposed period of usage longer than one year.
6. **Intangible assets** (dlouhodobý nehmotný majetek).
   Immaterial assets worth more than 60 thousand CZK and with supposed period of usage longer than one year. Above all this indicator includes goodwill, software, patents, copyrights, trademarks and tradenames.
7. **Acquisition of tangible and intangible assets inclusive land save financial assets** (pořízení dlouhodobého majetku včetně pozemků bez dlouhodobého finančního majetku celkem).
   Includes purchased assets, expenses connected with self-constructed long-term assets, and the value of assets obtained by voluntary conveyance.
8. **Number of employees** (počet zaměstnaných osob).
   Number of people who are permanently or temporarily employed by the firm, irrespective

[25] [http://dw.czso.cz/pls/metis/TUCUK_N.ZAC].

of their country of citizenship. Employment means that employers perform continuous work for the employer. Generally, all employers who receive regular pay are included here, and emloyers who temporarily left their job and do not receive any wage at the same time (eg parental leave) are not counted.

9. **Average number of employees** (Průměrný evidenční počet zaměstnanců).
   The previous item recalculated in order to capture fluctuations. Number of employees on individual days of one month is divided by the number of days in the respective month, and this monthly figure is averaged for to obtain the yearly indicator.

10. **Payroll (without other personnel expenses)** (mzdy bez ostatních osobních nákladů).
    Salaries and payments in kind provided to employers belonging to the item "number of employees". Includes regular pay, supplementary pay, bonuses and other components of salaries. Gross wages are indicated, ie before social and health insurance contribution and income tax is deducted.

11. **Other personnel expenses** (ostatní osobní náklady).
    Payments that are not connected with regular employment contract, indicated as gross payments. These will typically be: remuneration for work contracted beyond the employment contract, remuneration for expert testimonies or for intermediation, royalties and other patent fees, severation or termination pays, salaries of judges.

Table 5.1: Share of SME on employment in selected countries (Ayyagari *et al.* [6]).

Country	GDP/capita	SME 250	Country	GDP/capita	SME 250
Austria	29619	66	Italy	19218	80
Belarus	2523	5	Japan	42520	72
Belgium	27572	69	Luxembourg	45185	71
Brazil	4327	60	Netherlands	27395	61
Bulgaria	1487	50	Poland	3391	63
Croatia	4454	62	Portugal	11121	80
Czech Republic	5015	64	Romania	1501	37
Denmark	34576	69	Russian Federation	2614	13
Estonia	3752	65	Slovak Republic	3651	57
Finland	26814	59	Spain	15362	80
France	27236	67	Sweden	27736	61
Georgia	737	7	Taiwan, China	12474	69
Germany	30240	60	Turkey	2865	61
Greece	11594	87	Ukraine	1190	5
Hungary	4608	46	United Kingdom	19361	56
Ireland	19528	67			

GDP/capita = Real GDP per capita in USD. SME 250 = Share of the SME sector on the total formal labour force in manufacturing when 250 employees is taken as the cutoff for the definition of an SME. Data are for the 1990's. Unfortunately for some other important world economies the cutoff 250 is not available.

Table 5.2: Definition of SME according to the EU legislation.

Enterprise Category	Headcount	Turnover	Balance Sheet Total
Micro	< 10	$\leq$ €2 million	$\leq$ €2 million
Small	< 50	$\leq$ €10 million	$\leq$ €10 million
Medium-sized	< 250	$\leq$ €50 million	$\leq$ €43 million

Table 5.3: Share of SME on macroeconomic indicators.

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
♯ of firms	99.77	99.78	99.80	99.71	99.81	99.85	99.81	99.84	99.85	99.85
♯ of employees	59.82	57.91	58.84	59.42	59.73	61.34	61.63	61.48	61.63	61.38
Gross output	52.91	53.03	53.63	51.53	51.44	52.46	52.79	52.29	52.42	51.45
Value added	57.36	52.25	53.17	51.93	51.33	52.98	54.46	53.02	53.68	52.62
Payroll	n/a	53.93	54.57	54.42	55.72	55.82	55.90	55.61	55.88	55.41
Investment	n/a	41.53	41.06	40.48	37.81	44.52	49.88	51.43	52.57	54.42
Export	36.40	36.25	36.54	36.15	35.74	34.16	34.0	34.3	40.7	43.5
Import	48.00	48.84	50.74	49.43	47.12	50.33	49.8	52.5	54.7	54.6
GDP	n/a	n/a	31.54	31.17	31.63	34.59	34.86	34.69	34.60	34.36

100% = Czech economy of the given year. Investment = Acquisition of tangible and intangible assets, refer to appendix 5.6. Included are all entrepreneurial activities in manufacturing, construction, commerce and a part of services.

Table 5.4: Indicators on SME provided by the Czech Statistical Office.

Indicator and the corresponding variable in the model		
Number of active firms		
OUT	Output	Production value
REV	Sales revenue	Production value
VAD	Accounting value added	Production value
TAS	Tangible assets	Capital
IAS	Intangible assets	Capital
INV	Acquisition of tangible and intangible assets	Capital
EMP	Number of employees	Labour
AEM	Average number of employees	Labour
PAY	Payroll	Labour
OPE	Other personnel expenses	Labour

Table 5.5: OLS for the model (5.7), data for 2005.

	Estimate	Std. Error	t value	Pr($>\|t\|$)
(Intercept)	-0.3654	0.2704	-1.35	0.1790
β_1 (TAS)	0.1414	0.0389	3.64	0.0004
β_2 (IAS)	-0.0241	0.0178	-1.35	0.1780
β_3 (INV_{t-1})	-0.0977	0.0354	-2.76	0.0067
β_4 (AEM)	0.1293	0.0338	3.82	0.0002
β_5 (AWG)	0.0111	0.0163	0.68	0.4977

Residual standard error: 0.3352 on 126 degrees of freedom

Multiple R-squared: 0.2214, Adjusted R-squared: 0.1905

F-statistic: 7.166 on 5 and 126 DF, p-value: 6.222e-06

Table 5.6: Fixed effects regression for the model (5.7), data for 2003-2005.

	Estimate	Std. Error	t value	Pr($>\|t\|$)
β_1 (TAS)	-0.0075	0.0318	-0.2364	0.8133
β_1 (IAS)	0.0126	0.0100	1.2628	0.2078
β_1 (INV_{t-1})	-0.0411	0.0220	-1.8675	0.0630
β_1 (AEM)	-0.1563	0.0414	-3.7781	0.0002
β_1 (AWG)	0.3787	0.0708	5.3486	0.0000

Total Sum of Squares: 16.033, Residual Sum of Squares: 13.716

F-statistic: 8.68528 on 5 and 257 DF, p-value: 1.2662e-07

Table 5.7: Random effects regression for the model (5.7), data for 2003-2005.

	Estimate	Std. Error	t value	Pr($>\|t\|$)
(Intercept)	-0.3716	0.1753	-2.1192	0.0347
β_1 (TAS)	0.0369	0.0244	1.5106	0.1317
β_1 (IAS)	0.0194	0.0096	2.0220	0.0439
β_1 (INV_{t-1})	-0.0409	0.0211	-1.9367	0.0535
β_1 (AEM)	0.0752	0.0200	3.7522	0.0002
β_1 (AWG)	0.0249	0.0140	1.7866	0.0748

Total Sum of Squares: 26.645, Residual Sum of Squares: 24.152

F-statistic: 7.98702 on 5 and 387 DF, p-value: 3.5169e-07

Table 5.8: Quartile statistics for maximum likelihood efficiency scores using (5.8).

	♯ of employees	min	1Q	median	3Q	max	mean
	Not restricted	0.1550	0.7167	0.8231	0.8909	0.9633	0.7868
	<10	0.1550	0.7385	0.8819	0.9174	0.9633	0.8164
2003	10-19	0.4887	0.7388	0.8178	0.8820	0.9373	0.8023
to	20-49	0.2808	0.7384	0.8496	0.8812	0.9390	0.7771
2005	50-99	0.4247	0.6796	0.8135	0.8798	0.9394	0.7699
	100-250	0.3590	0.6719	0.8067	0.8752	0.9592	0.7652

Table 5.9: Summary statistics of data from section 5.4.1.

	Mean	Std.dev.	1Q	Median	3Q
VAD	0.316220	0.103518	0.255206	0.315597	0.370825
TAS	0.451007	0.605088	0.198210	0.275333	0.434844
IAS	0.011951	0.051019	0.001882	0.004244	0.007865
INV	0.089241	0.127736	0.040562	0.057091	0.087235
$AEMP$	0.000766	0.000470	0.000466	0.000670	0.000954
AWG	0.000581	0.003842	0.000019	0.000041	0.000145

$N = 528$. All variables divided by OUT.

Bibliography

[1] ACS, Z. J., CARLSSON, B., AND KARLSSON, C., Eds. *Entrepreneurship, Small and Medium-sized Enterprises and the Macroeconomy*. Cambridge University Press, New York, 1999. ISBN: 0-521-62105-4.

[2] AIGNER, D. J., AND CHU, S. F. On estimating the industry production function. *The American Economic Review 58* (1968), 826–839.

[3] AIGNER, D. J., LOVELL, C. A. K., AND SCHMIDT, P. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics 6* (1977), 21–37.

[4] ALONSO-BORREGO, C. Demand for labour inputs and adjustment costs: Evidence from Spanish manufacturing firms. *Labour Economics 5*, 4 (1998), 475–497.

[5] ALVAREZ, R., AND CRESPI, G. Determinants of technical efficiency in small firms. *Small Business Economics 20* (2003), 233–244.

[6] AYYAGARI, M., BECK, T., AND DEMIRGUC-KUNT, A. Small and medium enterprises across the globe. *Small Business Economics 29* (2007), 414–434.

[7] BARDHAN, I. R., COOPER, W. W., AND KUMBHAKAR, S. C. A simulation study of joint uses of data envelopment analysis and statistical regressions for production function estimation and efficiency evaluation. *Journal of Productivity Analysis 9* (1998), 249–278.

[8] BATTESE, G. E., AND COELLI, T. J. Prediction of firm-level technical efficiencies with a generalized frontier producion function and panel data. *Journal of Econometrics 38* (1988), 387–399.

[9] BATTESE, G. E., AND COELLI, T. J. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis 3* (1992), 153–169.

[10] BENÁČEK, V. Small businesses and private entrepreneurship during transition: The case of the Czech Republic. Working paper 53, CERGE-EI, P. O. Box 882, Politických vězňů 7, 11121 Prague, Czech Republic, 1994. http://www.cerge-ei.cz/pdf/wp/Wp53.pdf.

[11] BENÁČEK, V., SHEMETILO, D., AND PETROV, A. Restructuring and measurement of efficiency in firms in transition. In *Restructuring Eastern Europe*, S. Sarma, Ed. Cheltenham: Elgar, 1997, pp. 58–72. ISBN: 1-858-985-765.

[12] BERRY, S., LEVINSOHN, J., AND PAKES, A. Automobile prices in market equilibrium. *Econometrica 63*, 4 (1995), 841–890.

[13] BERRY, S., LEVINSOHN, J., AND PAKES, A. Differentiated products demand systems from a combination of micro and macro data: The new car market. *The Journal of Political Economy 112*, 1 (2004), pp. 68–105.

[14] BIØRN, E., HAGEN, T. P., IVERSEN, T., AND MAGNUSSEN, J. The effect of activity-based financing on hospital efficiency: A panel data analysis of DEA efficiency scores 1992–2000. *Health Care Management Science 6* (2003), 271–283.

[15] BIØRN, E., SKJERPEN, T., AND WANGEN, K. R. Can random coefficient Cobb-Douglas production function be aggregated to similar macro functions? In *Panel Data Econometrics: Theoretical Contributions and Empirical Applications*, B. H. Baltagi, Ed., Contributions to Economic Analysis. Elsevier, Oxford (UK), 2006, pp. 229–258. ISBN: 978-0-444-521172-0.

[16] BLAUG, M. *The Methodology of Economics or How Economists Explain*, 2nd ed. Cambridge University Press, 1992. ISBN-13: 978-0-521-43678-6.

[17] BRAKMAN, S., AND HEIJDRA, B. J., Eds. *The Monopolistic Competition Revolution in Retrospect*. Cambridge University Press, Cambridge (UK), 2004.

[18] BRESNAHAN, T. F., STERN, S., AND TRAJTENBERG, M. Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND Journal of Economics 28*, 1 (1997), 17–44.

[19] BRESSON, G., KRAMARZ, F., AND SEVESTRE, P. Heterogeneous labor and the dynamics of aggregate labor demand: Some estimations using panel data. *Empirical Economics 17*, 1 (1992), 153–168.

[20] BĂDIN, L., DARAIO, C., AND SIMAR, L. Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European Journal of Operational Research 201*, 2 (2010), 633–640.

[21] BĂDIN, L., DARAIO, C., AND SIMAR, L. Explaining inefficiency in nonparametric production models: The state of the art. *Annals of Operations Research* (2012), 1–26.

[22] BĂDIN, L., DARAIO, C., AND SIMAR, L. How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research 223*, 3 (2012), 818–833.

[23] BURNSIDE, C. Production function regressions, returns to scale and externalities. *Journal of Monetary Economics 37* (1996), 177–201.

[24] CAZALS, C., FLORENS, J.-P., AND SIMAR, L. Nonparametric frontier estimation: A robust approach. *Journal of Econometrics 106* (2002), 1–25.

[25] CHERCHYE, L., KUOSMANEN, T., AND POST, T. What is the economic meaning of FDH? A reply to Thrall. *Journal of Productivity Analysis 13* (2000), 263–267.

[26] COASE, R. H. The nature of the firm. *Economica 4*, 16 (1937), 386–405.

[27] COELLI, T. J. A guide to DEAP version 2.1: A data envelopment analysis (computer) program. Working paper 8/96, The University of New England: Centre for Efficiency and Productivity Analysis, 1996. http://www.uq.edu.au/economics/cepa/index.htm. ISBN: 1-863-89-4969.

[28] COELLI, T. J. A guide to FRONTIER version 4.1: A computer program for stochastic frontier production and cost function estimation. Working paper 7/96, The University of New England: Centre for Efficiency and Productivity Analysis, 1996. http://www.uq.edu.au/economics/cepa/index.htm, ISBN: 1-86389-4950.

[29] COELLI, T. J., RAO, P., O'DONNELL, C. J., AND BATTESE, G. E. *An Introduction to Efficiency and Productivity Analysis*. New York: Springer, 2005.

[30] COHEN, A. J., AND HARCOURT, G. C. Whatever happened to the Cambridge capital theory controversies? *Journal of Economic Perspectives 17*, 1 (2003), 199–214.

[31] COOPER, R. W., AND HALTIWANGER, J. C. On the nature of capital adjustment costs. *Review of Economic Studies 73*, 3 (2006), 611–633.

[32] COOPER, W. W., SEIFORD, L. M., AND TONE, K. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Kluwer Academic Publishers, New York, 2002. ISBN: 0-306-47541-3.

[33] COOPER, W. W., SEIFORD, L. M., AND ZHU, J., Eds. *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishing, New York, 2004. ISBN: 1-4020-7798-X.

[34] COSH, A., CUMMING, D. J., AND HUGHES, A. Outside entrepreneurial capital. *Economic Journal, Forthcoming* (17 Aug. 2007). http://ssrn.com/abstract=663841.

[35] COSH, A., HUGHES, A., AND WOOD, E. Innovation in UK SMEs: Causes and consequences for firm failure and acquisition. In *Entrepreneurship, Small and Medium-sized Enterprises and the Macroeconomy*, Z. J. Acs, B. Carlsson, and C. Karlsson, Eds. Cambridge University Press, New York, 1999, ch. 12, pp. 329–366.

[36] DARAIO, C., AND SIMAR, L. Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis 24* (2005), 93–121.

[37] DARAIO, C., AND SIMAR, L. *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Studies in Productivity and Efficiency. Springer, New York, 2007. e-ISBN-13: 978-0387-35231-1.

[38] DARAIO, C., AND SIMAR, L. Conditional nonparametric frontier models for convex and nonconvex technologies: A unifying approach. *Journal of Productivity Analysis 28* (2007), 13–32.

[39] DE ALESSI, L. Property rights, transaction costs, and X-efficiency: An essay in economic theory. *The American Economic Review 73*, 1 (1983), 64–81.

[40] DE LOECKER, J. Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity. Princeton University, NBER and CEPR, Jan. 2009.

[41] DI GIOVANNI, J., LEVCHENKO, A. A., AND MÉJEAN, I. Firms, destinations and aggregate fluctuations. RSIE Discussion Papers 630, Aug. 2012.

[42] DJANKOV, S., AND HOEKMAN, B. Foreign investment and productivity growth in Czech enterprises. *World Bank Economic Review 14*, 1 (2000), 49–64.

[43] DUFFY, J., AND PAPAGEORGIOU, C. A cross-country empirical investigation of the aggregate production function specification. *Journal of Economic Growth 5* (2000), 87–120.

[44] ENGELHARDT, L. Expansionary monetary policy and decreasing entrepreneurial quality. *The Quarterly Journal of Austrian Economics 15*, 2 (2012), 172–194.

[45] EVANS, A. J., AND BAXENDALE, T. Austrian business cycle theory in light of rational expectations: The role of heterogeneity, the monetary footprint, and adverse selection in monetary expansion. *The Quarterly Journal of Austrian Economics 11*, 2 (2008), 81–93.

[46] FABIANI, S., ET AL. The pricing behaviour of firms in the euro area: New survey evidence. *ECB Working Paper*, 535 (2005). http://ssrn.com/abstract_id=825628.

[47] FARRELL, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General) 120*, 3 (1957), 253–290.

[48] FOSS, K., FOSS, N. J., KLEIN, P. G., AND KLEIN, S. K. The entrepreneurial organization of heterogeneous capital. *Journal of Management Studies 44* (2007), 1165–1186.

[49] FOSS, N. J. The continuing relevance of austrian capital theory. *The Quarterly Journal of Austrian Economics 15*, 2 (2012), 151–171.

[50] FOSS, N. J., AND KLEIN, P. G. *Organizing Entrepreneurial Judgment: A New Approach to the Firm*. Cambridge University Press, Cambridge (UK), 2012. ISBN: 978-0-521-69726-2.

[51] FOSTER, L., HALTIVANGER, J. C., AND KRIZAN, C. J. The link between aggregate and micro productivity growth: Evidence from retail trade. NBER Working Paper 9120, National Bureau of Economic Research, 2002.

[52] FRANTZ, R. S. *X-efficiency : Theory, Evidence and Applications*. Kluwer, Boston, 1988. ISBN: 0-8983-8242-4.

[53] GREENE, W. H. *Econometric Analysis*. Prentice Hall, 2003.

[54] GREENE, W. H. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics 126* (2005), 269–303.

[55] GREENE, W. H. The econometric approach to efficiency analysis. In *The Measurement of Productive Efficiency and Productivity Growth*, H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, Eds. Oxford University Press, New York, 2008, ch. 2, pp. 92–250. ISBN: 0-1951-8352-5.

[56] GRILICHES, Z., AND MAIRESSE, J. Production functions: The search for identification. NBER Working Paper 5067, National Bureau of Economic Research, 1995.

[57] HÁJEK, M. Economic growth and total factor productivity in the Czech Republic from 1992 to 2004. Working paper 5/2005, Cetrum ekonomických studií VŠEM, 2005.

[58] HANOUSEK, J., KOČENDA, E., AND MAŠIKA, M. Corporate efficiency: Effect of ownership structures and financial indicators. *Politická Ekonomie*, 4 (2012).

[59] HAVLÍK, P. Trade and cost competitiveness in the Czech Republic, Hungary, Poland and Slovenia. World Bank Technical Paper 482, World Bank, Nov. 2000. ISBN: 0-8213-4796-9.

[60] HAYEK, F. A. The use of knowledge in society. *The American Economic Review 35*, 4 (Sept. 1945), 519–530.

[61] HOFFMAN, K., PAREJO, M., BESSANT, J., AND PERREN, L. Small firms, R&D, technology and innovation in the UK: A literature review. *Technovation 18*, 1 (1998), 39–55.

[62] HUERTA DE SOTO, J. *The Theory of Dynamic Efficiency*. Routledge, New York, 2009. ISBN: 978-0-415-42769-2.

[63] HUGHES, A. Finance for SMEs: A UK perspective. *Small Business Economics 9* (1997), 151–166.

[64] Johansen, L. *Production Functions: An Integration of Micro and Macro, Short Run and Long Run Aspects*, 1st ed. North-Holland, Amsterdam, 1972. ISBN: 0-7204-3181-6.

[65] Jovanovic, B., and Yatsenko, Y. Investment in vintage capital. *Journal of Economic Theory 147*, 2 (2012), 551–569.

[66] Kennedy, P. *A Guide to Econometrics*. Oxford (UK): Blackwell, 1992. ISBN: 0-631-18211-X.

[67] Klacek, J., and Vopravil, J. Total factor productivity in Czech manufacturing industry — KLEM framework. *Statistika*, 5 (2008), 414–428.

[68] Klapper, L. F., Sarria-Allende, V., and Sulla, V. Small- and medium-size enterprise financing in eastern europe. World Bank Policy Research Working Paper 2933, World Bank, Nov. 2002. http://ssrn.com/abstract=636295.

[69] Klette, T. J., and Griliches, Z. The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics 11*, 4 (1996), 343–361.

[70] Kneller, R., and Stevens, P. A. The specification of the aggregate production function in the presence of inefficiency. *Economic Letters 81* (2003), 223–226.

[71] Kogiku, K. C. *Microeconomic Models*. Happer & Row, New York, 1971. SBN: 06-043746-4.

[72] Kumbhakar, S. C., and Lovell, C. A. K. *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge (UK), 2000. ISBN: 0-521-48184-8.

[73] Kumbhakar, S. C., Park, B. U., Simar, L., and Tsionas, E. G. Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics 137*, 1 (2007), 1–27.

[74] Kuosmanen, T. Stochastic nonparametric envelopment of data: Combining virtues of SFA and DEA in a unified framework. *MTT Discussion Paper*, 3/2006 (June 2006). http://ssrn.com/abstract=905758.

[75] Kuosmanen, T. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. *Energy Economics 34*, 6 (2012), 2189–2199.

[76] Kuosmanen, T., Kortelainen, M., Sipiläinen, T., and Cherchye, L. Firm and industry level profit efficiency analysis using absolute and uniform shadow prices. *European Journal of Operational Research 202*, 2 (2010), 584–594.

[77] Lachmann, L. M. *Capital and Its Structure*. Sheed Andrews and McMeel, Kansas City, 1978. ISBN: 0-8362-0740-8.

[78] LANCASTER, K. J. A new approach to consumer theory. *Journal of Political Economy 74*, 2 (Apr. 1966), 132–157.

[79] LANCASTER, K. J. *Consumer Demand - A New Approach.* Columbia University Press, New York., 1971.

[80] LEIBENSTEIN, H. Allocative efficiency vs. "X-efficiency". *The American Economic Review 56*, 3 (1966), 392–415.

[81] LEVINSOHN, J., AND PETRIN, A. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies 70* (2003), 317–341.

[82] LEWIN-KOH, N. Hexagon binning: An overview. Tech. rep., http://www.bioconductor.org/packages/2.4/bioc/html/hexbin.html, 21 Apr. 2009.

[83] MAN, T. W. Y., LAU, T., AND CHAN, K. F. The competitiveness of small and medium enterprises: A conceptualization with focus on entrepreneurial competencies. *Journal of Business Venturing 17*, 2 (2002), 123–142.

[84] MARCINČIN, A., AND WIJNBERGEN, S. The impact of Czech privatization methods on enterprise performance incorporating initial selection-bias correction. *Economics of Transition 5*, 2 (1997), 289–304.

[85] MARSCHAK, J., AND ANDREWS, WILLIAM H., J. Random simultaneous equations and the theory of production. *Econometrica 12*, 3/4 (1944), 143–205.

[86] MAS-COLELL, A., WHINSTON, M. D., AND GREEN, J. R. *Microeconomic Theory*. Oxford University Press, New York, 1995. ISBN: 0-1950-7340-1.

[87] MELITZ, M. J. Estimating firm-level productivity in differentiated product industries. Department of Economics, Harvard University, 15 Dec. 2000.

[88] MÉNARD, C. *Handbook of New Institutional Economics*. Springer, 2005, ch. A New Institutional Approach to Organization, pp. 281–318.

[89] MOHNEN, A., AND NASEV, J. Growth of small and medium-sized firms in Germany. http://ssrn.com/abstract=852785, October 2005.

[90] MOSCHANDREAS, M. *Business Economics*. Thomson Learning, London, 2000. ISBN: 1-86152-538-9.

[91] MURPHY, G. B., TRAILER, J. W., AND HILL, R. C. Measuring performance in entrepreneurship research. *Journal of Business Research 36*, 1 (1996), 15–23. Entrepreneurship and New Firm Development.

[92] NADIRI, M. I. Producers theory. In *Handbook of Mathematical Economics, vol. II*, K. J. Arrow and M. D. Intriligator, Eds. North-Holland, Amsterdam, 1982, ch. 10, pp. 431–490. ISBN: 0-444-86127-0.

[93] PRŮŠA, J. A general framework to evaluate economic efficiency with an application to British SME. IES Working Papers 25/2009, IES FSV, Charles University, 2009.

[94] PRŮŠA, J. Productivity of Czech small and medium enterprises: Lagging behind their potential. *Journal of Industry, Competition and Trade 10*, 3–4 (2010), 343–363.

[95] PRŮŠA, J. The most efficient Czech SME sectors: An application of robust data envelopment analysis. *Czech Journal of Economics and Finance 62*, 1 (2012), 44–65.

[96] PRŮŠA, J., KLIMEŠOVÁ, A., AND JANDA, K. Consumer loss in Czech photovoltaic power plants in 2010–2011. *Energy Policy, forthcoming* (2013).

[97] REIFSCHNEIDER, D., AND STEVENSON, R. Systematic departures from the frontier: A framework for the analysis of firm inefficiency. *International Economic Review 32*, 3 (1991), 715–723.

[98] ROTHBARD, M. N. *Man, Economy, and State*. Ludwig von Mises Institute, Auburn, 2001. ISBN: 978-1-933550-27-5.

[99] RYSKA, P., AND PRŮŠA, J. Efficiency wages in heterogenous labour markets. Working Paper 28/2011, Institute of Economic Studies, Charles University in Prague, 2011.

[100] SATO, K. *Production Functions and Aggregation*, 1st ed. North-Holland, Amsterdam, 1975. ISBN 0-7204-3197-2.

[101] SAUTET, F. E. *An Entrepreneurial Theory of the Firm*. Foundations of the Market Economy. Routledge, London, 2000. ISBN: 0-415-22977-4.

[102] SCHIFFER, M., AND WEDER, B. Firm size and the business environment: Worldwide survey results. Discussion paper 43, World Bank/International Finance Corporation, 2001. ISBN: 0-8213-5003-X.

[103] SHY, O. *Industrial Organization: Theory and Applications*. The MIT Press, Cambridge, Massachusetts, 1996.

[104] SIMAR, L., AND WILSON, P. W. A general methodology for bootstrapping in nonparametric frontier models. *Journal of Applied Statistics 27* (2000), 779–802.

[105] SIMAR, L., AND WILSON, P. W. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics 136*, 1 (2007), 31–64.

[106] SIMAR, L., AND WILSON, P. W. Statistical inference in nonparametric frontier models: Recent developments and perspectives. In *The Measurement of Productive Efficiency and Productivity Growth*, H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, Eds. Oxford University Press, New York, 2008, ch. 2, pp. 92–250. ISBN: 0-1951-8352-5. Working paper version: http://www.clemson.edu/economics/faculty/wilson/Papers/sw_ch4_22-Dec-2006.pdf.

[107] SOLEIMANI-DAMANEH, M., AND RESHADI, M. A polynomial-time algorithm to estimate returns to scale in FDH models. *Computers & Operations Research 34*, 7 (2007), 2168–2176.

[108] STIGLER, G. J. The Xistence of X-efficiency. *The American Economic Review 66*, 1 (1976), 213–216.

[109] TALEB, N. N. *The Black Swan: The Impact of the Highly Improbable*. Penguin, London, 2008.

[110] TAYMAZ, E. Are small firms really less productive? *Small Business Economics 25* (2005), 429–445.

[111] THRALL, R. M. What is the economic meaning of FDH? *Journal of Productivity Analysis 11* (1999), 243–250.

[112] TSIONAS, E. G. Combining DEA and stochastic frontier models: An empirical Bayes approach. *European Journal of Operational Research 147* (2003), 499–510.

[113] TULKENS, H., AND EECKAUT, P. V. Non-parametric efficiency, progress and regress measures for panel data: Methodological aspects. *European Journal of Operational Research 80*, 3 (1995), 474–499.

[114] WICKENS, M. R. Estimation of the vintage Cobb-Douglas production function for the United States 1900–1960. *The Review of Economics and Statistics 52*, 2 (1970), 187–193.

[115] WILSON, P. W. Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business & Economic Statistics 11*, 3 (1993), 319–323.

[116] WILSON, P. W. FEAR: A software package for frontier efficiency analysis with R. *Socio-Economic Planning Sciences 42*, 4 (2008), 247–254.

[117] WYNARCZYK, P., AND WATSON, R. Firm growth and supply chain partnerships: An empirical analysis of UK SME subcontractors. *Small Business Economics 24* (2005), 39–51.

[118] YANG, C.-H., AND CHEN, K.-H. Are small firms less efficient? *Small Business Economics 32*, 4 (2007), 375–395.

[119] YASAR, M., NELSON, C. H., AND REJESUS, R. M. The dynamics of exports and productivity at the plant level: A panel data error correction model (ECM) approach. In *Panel Data Econometrics: Theoretical Contributions and Empirical Applications*, B. H. Baltagi, Ed., Contributions to Economic Analysis. Elsevier, Oxford (UK), 2006, pp. 279–305. ISBN: 978-0-444-521172-0.

Responses to Reviewer Comments
Dissertation Thesis

Original title:
Dynamic Efficiency Analysis

Amended title:
Economic Efficiency, Competition and Equilibrium in Heterogeneous Production

Jan Průša[a]

[a]*Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague. Opletalova 26, 11000 Prague, Czech Republic.*

Notes to the Reviewers

Dear Reviewers,

I would like to thank you for all comments which were invaluable in improving the thesis. Below please find the explanations of how I incorporated them.

Yours sincerely,

Jan Průša

---

1. Report by Vladimír Benáček

1.1. Suggestions for improvement

The whole research of Mr. Prusa deals with the allocations of various inputs as economic factors, two of which are labour and capital. In the economic theoretical debates there was one debate quite crucial and related to production functions and efficiency: the "Cambridge capital controversy". Even though we could add that at present the highly aggressive (and destructive) attack of neo-Ricardians on standard economics has settled down, it would not be a useless exercise to open that controversy in this work and explain in enlarged ch. 2

Email address: jan.prusa@ies-prague.org (Jan Průša)

how the problem of capital aggregation or capital comparison across firms and industries could be related to this research, (e.g. by reflecting on some of the questions raised by Cohen and Harcourt in Journal of Economic Perspectives, Vol. 17, pp. 199-214).

We included a discussion on Cambridge capital controversy in section 2.3.2.3 Heterogeneity of Capital. Above all we point to the importance of adjustment costs when capital switches employment from one sector to another, as documented by Cooper and Haltiwanger (2006).

Then there is perhaps my personal insufficiency — the Fig. 3.1 (p. 58) seems to me rather difficult to interpret (being used to contiguous patterns). Maybe it will be useful to place an explicit comment directly under the figure, explaining that efficiency is not a chaotically ordered pattern.

We added a comment to the figure in order to clarify its message: "The figure shows clusters of efficiency scores depending on the number of employees. It appears that there is no straightforward pattern in the size-efficiency relationship. Upon detailed inspection three significant points of gravity can be identified: Small firms with 1–20 employees and efficiency between 0.1–0.4, firms with 65–75 employees and efficiency between 0.2–0.4 and finally firms with around 150 employees and efficiency of 0.2–0.3."

2. Report by Barbara Gebicka

2.1. General comments

1. The dissertation would benefit from more connection and cross-referencing across the chapters. In the current version all four chapters appear to be completely independent. However, it appears natural to refer to Chapter 2 in each empirical chapter and comment how the principles of dynamic efficiency analysis are applied in a given empirical application. Additionally, as Chapters 3 and 4 present alternative efficiency analyses using the same data, the author should compare the results and comment on differences in conclusions of these two chapters.

Given that chapters 3-5 present independently written research articles, we take full advantage of the synthetic chapter 2 (and also of the new chapter 1 — executive summary) to provide all cross-references and comparisons here. This has been reflected in the structure of the dissertation, in that it is divided into two parts — a theoretical and an empirical part. We extensively compare methodologies as well as results of the empirical chapters in chapter 2.

2. While the author mentions the related literature in Chapter 2 and in introductory sections of the remaining three chapters, there is no clear comparison of the results presented in this dissertation with those from the up-to-date literature. Even if the methodology used by the author has not been used in any other study, it would be informative to see whether the obtained results are comparable to those coming from alternative methods and different data.

A broader set of references, as well as a clear explanation where the thesis fits in, is extensively discussed in chapter 1.

2.2. Chapter 2

1. This chapter is written using very strong claims. I would advise the author to avoid strong statements, like for example: "the focus on equilibrium led economists believe in the efficiency mantra" (p. 11), and use softer and mildly formulated arguments instead.

Many parts of chapter 2 were amended in order to improve the final version. For example, the sentence to which the referee points was changed to "The focus on equilibrium made economists look for *efficiency in all situations, at all times*, an approach resembling almost an efficiency mantra."

2. Some of the arguments presented in this chapter seem to be based on an incomplete review of literature. For example, the author criticizes the use of labor as one aggregate, while this factor of production is clearly heterogeneous. However, there exists a broad literature on task-skill model of the labor market with the idea that the production process is task-based and different skills are used to complete different tasks (see Acemoglu D., Autor, D. (2010) Skills, Tasks and Technologies: Implications for Employment and Earnings. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 4b, Amsterdam.)

The argument in our thesis is somewhat different: We argue that we can use labour as one aggregate (and thus neglect heterogeneity) only if we at the same time include the information on wages, which should capture a big portion of heterogeneity.

Further, we are well aware that heterogeneity of labour is well established *within* labour economics. In the efficiency literature however, the standard approach has been to count labour as the number of employees, irrespective of labour heterogeneity. For the sake of brevity we decided not to take a detour to macroeconomic models of labour markets that are not directly related to efficiency models.

Finally, the paper by Acemoglu and Autor to which the referee points introduces workers of different types. This is a simple form of heterogeneity and amounts merely to adding a further dimension to the vector of inputs. What we are interested in is a more complex form of heterogeneity, where the single unit (1 worker) can have multiple qualities. Therefore, the qualities of the worker cannot be simply separated to different dimensions of the input vector (as they are still carried by the single worker). But the worker also cannot be simply aggregated into a single labour input (as he carries more than one quality). This is a very different framework from that offered by Acemoglu and Autor. This underlines again our argument to use monetary measurement of labour (i.e. wages) jointly with the headcount.

3. While this chapter criticizes the aggregation approach in production frontier analysis and puts forward the money-metric approach, the following chapters use some aggregate values as explanatory variables in efficiency analysis. Specifically, labor is used as one aggregate input in Chapters 4 and 5.

In chapters 4 and 5 we use both the information on the number of employees *and* wages. This is perfectly in line with the arguments advanced in Chapter 2, where we argue that wages capture a significant portion of heterogeneity of workers.

See also the discussion at the previous comment.

4. The author argues that the money-metric approach is a remedy for the imprecision caused by the heterogeneity of inputs. Following this argument, one should compare the total value of all inputs with the total value of outputs when calculating efficiency.

Yes, and this is in fact what every business enterprise does at the top level in its profit and loss statement.

But it is obviously an interesting task to dig deeper, and we do not argue against a more detailed analysis per se. Our point is merely that we must proceed very carefully when identifying more detailed variables, and that in most cases, if we neglect prices and measure only technical units (as is the case in many efficiency studies), we will get spurious results due to heterogeneity of the measureed quantities. This is why we believe that, even at a very detailed level of analysis, prices are an indispensable part of efficiency measurement.

2.3. Chapter 3

1. While being interesting and clearly novel, this chapter is not well written. It would benefit from more informative introduction, clear literature review (short description of methodology used by other studies measuring efficiency of firms), better organized sections and spell-check. For example, section 3.2.5 is redundant given that the author applies nonparametric estimation; in Section 3.4.4.3 it would be informative to write which approach the author of this Chapter is using.

The chapter was extensively rewritten and improved. Section 3.2.5 was moved to chapter 2 where it fits better.

2. More detailed description of the data (e.g. summary statistics) is welcome. It is not clear whether the data is observed at firm level or at industry-firm size cell level.

Summary statistics was added. Data are collected at firm level and aggregated into SIC sectors — we added this remark into the footnote in section 3.4.1.

3. It is confusing to read in Chapter 2 that efficiency frontiers are dynamic and change over time and then read in Chapter 3 that the author uses one reference year against which efficiency is measured.

Selecting one year as the reference frontier allows to calculate efficiency scores which directly incorporate the dynamic component of changing frontiers. This is explained in section 3.4.2.3 - Pooling over time. We extended the exposition to make our approach clearer.

4. I do not agree with the claim that sectoral classification should not affect economic efficiency. An entrepreneur starting a company decides in which sector/industry to operate and this decision might affect efficiency of his firm.

This remark opens a highly interesting debate: What drives the entrepreneur's decision to set up an enterprise in a given industry? It is true that general profitability in a given industry will attract new entrants. However we see other factors working in the opposite direction:

- We doubt that potential entrepreneurs would have good and detailed information on profitability in single industries as per SIC (standard industrial classification) statistics.

- Other factors will impact the entrepreneur's decision, such as his knowledge and skills, capital intensity or regulatory obstacles.

- There will be efficiency variations within single industries, and only certain parts of the sample might be visible to the would-be entrepreneur. This might lead to biased selection (I see one or two extremely profitable automotive companies → I will enter the automotive industry, even though in reality most companies in the automotive segment operate at extremely low margins).

Due to these factors we believe that the overall dependency between sectoral classification and economic efficiency as an ex ante driver will be rather low (statistically insignificant). We added these remarks in section 3.4.4.3: "It could be argued that the profitability of a sector influences the entrepreneur's decision to start his business, creating a link between sectoral classification and economic efficiency. But at the same time several mechanisms will work in the opposite direction to weaken this correlation. We doubt that potential entrepreneurs dispose of detailed information on profitability of sectors according to standard industrial classification. Rather, only some parts of the efficiency distribution within a given sector will be visible to them, leading to biased choices. Even if potential entrepreneurs had complete information on profitability of sectors, their decision will be driven by other factors such as their knowledge, skills and tastes, capital intensity and availability or regulatory obstacles. Most of all, given that the majority of businesses do not survive the early period of their existence, those who indeed decided based on the sectoral classification will be randomly mixed with those who decided based on other factors. We thererefore work with the assumption that sectoral classification and economic efficiency scores are statistically independent."

5. The regression equation (3.6) includes a full set of industry dummies (SIC) and time dummies (YEAR) which makes any variable measured at industry-time level (such as NEF) perfectly explained. This results in insignificant estimates for β_{NEF}.

In equation 3.6 the industry dummies SIC_i do not cover the full 4-digit classification, but only the ten aggregate clusters determined by the first number of the 4-digit classification. Therefore there are many more observations than there are dummies and NEF is not fully determined by a combination of $YEAR$

and SIC. Moreover it should be noted that the three dummy groups $YEAR$, SIC and EG are independent by construction. As mentioned in section 3.5.2.1, dropping each one of the dummy groups results in insignificant regression, which is the reason why we keep all of them. Additionally we now provide results for regressions where NEF is omitted and where both NEF and SIC are omitted.

6. Could you be more specific about why you assume zero correlation between ξ_i and regressors? Would not the argument about an entrepreneur choosing in which sector to operate violate this assumption?

This issue is addressed in the response to remark 4.

7. Please, be more specific when formulating the conclusion that wide dispersion of efficiency implies the need for more dynamic models.

The conclusion is reformulated to "We contend that it calls for more focus on out-of-equilibrium competitive and adjustment processes in further research." This is a very general lesson drawn from the huge variability of efficiency scores. Chapter 3 is not intended to tackle this issue, but we discuss the dynamic aspects of efficiency in chapter 2, especially in section 2.3.3.

8. It is possible to directly test for a joint significance of a set of dummy variables (section 3.5.2.1). Do EG dummies prove to be jointly insignificant under this test?

Results for joint significance of the dummy variables were added to the corresponding section.

9. When the author writes "significant", he means "significant at 5% level". This should be clarified.

Remark added where appropriate.

10. The goal of this chapter is to confirm/reject the sector-specific findings of other studies that smaller firms are less efficient. Pooling all sectors together the author does not find this relationship, but we do not know that this is due to different methodology or due to using economy-wide approach. It would be helpful to repeat the analysis for single sectors or include interaction of sector and size variables to disentangle these effects.

This would certainly be an interesting question for further research, which however goes beyond the limited scope of the current thesis.

11. Note that the intercept does not correspond to the average efficiency score but to the efficiency of firms in year 1998, SIC_0 sector and of the minimum size.

Corrected.

12. The author concludes that efficiency scores across sectors are very dispersed, but regression results show that all industry dummies are statistically insignificant. The latter suggests that there is no difference across industries.

The results merely suggest that there is no *systematic* difference across industries. The variability of efficiency scores is not explained by sectoral classification. We reformulated the remark in section 'conclusions' to make our idea clear, saying that 'Efficiency scores across observations are very dispersed with no systematic sectoral pattern.'

Minor comments

13. The author writes that technical efficiency is intuitively straightforward (p.35/36), nevertheless it would be useful to shortly explain this concept.

Technical efficiency is defined in section 3.2.1.1 in equation 3.1. We reformulated the sentence to which the reviewer points so as to include the reference to this equation.

14. In section 3.4.1 the author writes about testing hypotheses, but these are not clearly formulated in the text of this chapter.

Section 3.5 was rewritten so as to make the formulation of the tested hypotheses clear.

15. I was confused by the use of the terms "sections", "industries", and "sectors". Do they all mean the same or is there any difference?

We consistently use "industry" in reference to industrial sectors only (i.e. do not include agriculture and services). But this is a matter of style and in practice the terms "sector" and "industry" can be used interchangeably. As a matter of common practice "section" is used in reference to econometric analysis of cross-sections, or to refer to parts of the text.

16. In the description of Monte-Carlo simulation (section 3.4.3.3) not all notation is well explained, which makes it difficult to comprehend. Should there not be m=150 instead of q=150?

Corrected; we added clear explanation of the notation used.

17. The variable YEAR in equation (3.6) should not have an i-index.

Corrected.

18. It is common practice to present all figures first and then all tables.

Corrected.

2.4. Chapter 4

Chapters 4 and 5 have been already published in refereed journals, thus there is much less to comment in their case.

1. I would say that putting no constraints on λ implies variable returns to scale, while constraining the sum of to be equal to 1 implies constant returns to scale.

Setting $\lambda = 1$ implies variable returns to scale. We added reference to Cooper et al. (2002, chapter 4) which clarifies this.

2. The author abstracts from technology when defining production functions in section 4.3.2.1. I agree with this approach, as technology is actually part of efficiency.

3. While the rationale behind including both the wage bill and number of employees in the model is understandable, it is confusing from the point of view of the arguments given in Chapter 2. The author writes that money-metric approach allows for comparability across different heterogeneous units at the cost of being able to decompose the source of (in)efficiency. Inclusion of the number of employees in the production function seems to be an ad-hoc solution to this latter disadvantage.

We argue in chapter 2 and 3 that separation of physical (quantity) and economic (price) units is difficult in practice since the data are usually not available. In chapter 4 we have separate data on number of employees and average wage. Hence we can compare the importance of each of these two dimensions for *overall economic efficiency.*

This however does not imply that the number of employees is related solely to technical efficiency and that average wage is related solely to allocative efficiency. For example, hiring workers at low wages might imply low quality of workers and thus *technical* inefficiencies in production, while hiring a low number of employees might impact *economic* efficiency (e.g. by increasing the average cost if the small number of workers are not able to specialize enough).

The point that we want to stress in chapter 2 is: Unless we deal with a pure technical production function (where none of the dimensions is accounted for in monetary units), there will *always* be an economic component in the estimated inefficiency, so that technical and allocative efficiency cannot be separated.

4. To clearly account for the real average number of employees, I would suggest using man-hours. This measure corrects for the fact that some employees *do not work full time.*

We agree that man-hours would be a more appropriate measure, however it was not available so that average number of employees was the best available proxy.

5. The author should comment more on the results of Envelopes I: specifically, what is the source of difference between the estimates for the years 2002 and 2003 as opposed to 2004 and 2005? There appears to be a huge drop of efficiency. Is this because the Czech Republic entered the European Union? After EU accession foreign, more efficient SMEs could have entered the Czech market pushing the efficiency frontier up. In this case smaller average efficiency for the latter two years would not mean that average Czech firm became worse, but that there appeared significantly more efficient firms in the market.

This is an appealing hypothesis but would have to be verified by data outside the current scope of the text. All observations that were used in the model are Czech manufacturing industries. What the results show is that compared to the best performers over the whole period (since the observations were pooled) average industries suddenly became less efficient in 2004 and 2005. Because the observations are compared to Czech peers (and not foreign companies), the presented model alone cannot verify this hypothesis.

Minor comments

6. What does β stand for in footnote 3?

We added the definition of β to the footnote.

7. There is a typo on the top paragraph on page 80: it should read twenty first century.

Corrected.

2.5. Chapter 5

1. Please, explain the meaning of each variable listed in equation (5.5) and provide a summary statistics.

Data definition is already provided in appendix 5.6 and notation of variables is shown in table 5.4. We added summary statistics for variables in table 5.9.

2. There could be another reason why the estimate for the coefficient by tangible assets is insignificant in the fixed effects model (while it is significant in OLS). Under fixed effects the major source of identification comes from variation within groups, i.e. variation over a 4-year period. There could be not *enough change in tangible assets over this short time period to allow for proper estimation of the respective coefficient. This could be checked by inspecting the summary statistics for TAS.*

The summary statistics reveal that this is not the case: The variation of TAS is similar to the remaining regression variables. Moreover the results of random effects and fixed effects regressions are similar in that TAS is not significant in both cases. This means that insufficient variation over time should not be the reason for insignificance of TAS.

3. The hypothesis that investments appear nonproductive at first but increase efficiency in a longer run could be tested. The author could repeat the basic regressions on years 2004 and 2005 only with a 3-year lag of INV.

While being true, this would also significantly reduce the sample size using only a one-year sample. This is the reason why we did not include this regression.

4. The author should spend more time on comparing the OLS and fixed effects results, as this is interesting per se.

We included a brief comparison in a footnote in section 5.4.2.2: "By comparing the results of OLS and fixed effects estimation, it appears that what OLS shows to be the significant effect of tangible assets (β_{TAS}) is then absorbed into the significant group effects. However, the fixed effects estimation seems to give a more precise estimates for the impact of labour on efficiency (β_{AEM} and β_{AWG}). This means that efficiency of capital assets tends to be group (industry) specific while efficiency of labour is largely driven on firm level."

5. The analysis presented in section 5.4.3 are based on the assumption that random disturbance term and efficiency term from equation (5.2) are not correlated with each other and the regressors. This is a very strong assumption, which has been proved to fail by the author himself. The comparison of fixed effects and random effects models suggests that individual effects (which are part of the disturbance term in 5.2) are correlated with the regressors. This conclusion could be further confirmed by comparing a pooled OLS with the fixed effects model.

The MLE model is by definition based on strict distributional assumptions. We acknowledge this issue in a footnote in section 5.4.3: "Note that the maximum likelihood estimation in this section is based on strict distributional assumptions, in particular that the random disturbance term and the efficency term in equation (5.3) are not correlated."

6. Even if fixed effects are used to deal with the correlation between individual effects and regressors, there still remains an issue of correlation between time-variable disturbance and regressors. For the discussion of correlation between the disturbance term and regressors in production function, see for example Levinsohn, J. and A. Petrin. (2003). Estimating production functions using inputs to control for unobservables. Review of Economic Studies 70(2): 317–342.

The issue of correlation between time-varying disturbances and regressors can be addressed by multiple methods. One possibility is to use investments as instrumental variables, which is in fact what we do in our estimation. The article by Levinsohn and Petrin proposes a different method where intermediate inputs are used as instrumental variables. This is in fact a complementary approach which could be potentially tested in an extension of the presented text.

3. Report by Peter Klein

One suggestion for improvement is that the connections to related literatures can be brought out even more strongly. For example, there are passing references to entrepreneurship but not a very close tie to the extant entrepreneurship literature (see some comments on this below). The links to Austrian economics can extended and deepened for instance, the Austrians have a very particular notion of capital heterogeneity in mind (related to the time structure of production), and it may be possible to capture this at a more microeconomic level.

We extended section 2.3.2.3 "Heterogeneity of Capital" to cover Cambridge capital controversies and the concepts of capital reswitching and reversing. Unfortunately we note that the limited scope of this thesis is not sufficient to cover the full theory of the time structure of production and capital.

More generally, while the dissertation is very strong, technically, it may have limited appeal for the non-specialist (i.e., economists and management scholars interested in innovation, technical change, entrepreneurship, and competition but not well versed in efficiency analysis, which is after all a fairly specialized set of techniques). I think the kind of analysis reported here has very important potential implications for the fields of entrepreneurship, innovation, and competition more generally. First, simply documenting the wide dispersion of efficiency scores challenges the standard neoclassical economics notion of competitive equilibrium, in which deviations from "optimal" behavior are assumed to be competed away quickly (Alchian, 1950; Friedman, 1953), and hence relatively *unimportant for analyzing and drawing welfare conclusions about markets. As the author notes, this finding "calls for more focus on out-of-equilibrium competitive and adjustment processes in further research" (p. 56). I agree, and would go one step further, and suggest that x-inefficiency represents an entrepreneurial opportunity (in the sense of Israel Kirzner's work), and that we would expect some kind of entrepreneurial response. Future work could look not at the causes, but at the consequences of such inefficiency. Are firms farther from the efficient frontier more likely than other firms to replace the top management team, to be acquired, or to be liquidated to make the assets available to other entrepreneurs (as in Schumpeterian creative destruction)? What kinds of firms are more likely to improve their efficiency scores over time – e.g., do firms with a stronger "entrepreneurial orientation" (a standard construct in the management entrepreneurship literature) take actions to identify and exploit opportunities to move closer to the frontier? A closer integration with the entrepreneurship literature would be a valuable future research project.*

The impact of observed inefficiencies on entrepreneurial opportunities and activities is a hugely interesting subject. We mention some references to the entrepreneurship literature in section 2.3.2.4 and in Chapter 1 (Executive Summary). A detailed analysis is left for further research.

The analysis of Czech SMEs, via non-parametric (Chapter 4) and parametric (Chapter 5) analysis, sheds considerable light on this important sector of the Czech economy. The wide variation in measured efficiency levels, found using both methods, appears to call into question the viability of this sector and its role in generating economy-wide growth. However, without a comparison to larger, more established firms, such a conclusion cannot be drawn. (The parametric analysis does show some improvements in efficiency over time, though the sample frame is short, just 2003 to 2005.)

The interesting comparison that can be made is that of micro firms versus medium-sized firms. Here it can be said that the micro firms perform reasonably well against their larger peers.

The argument could be strengthened, I think, with a more extensive discussion, for the nontechnical reader, of the differences among various efficiency approaches. Within the field of productivity analysis, there is an intense debate – some even talk about warring camps or factions – between those committed to nonparametric methods (Data Envelopment Analysis or DEA) and those who *prefer the parametric approach (Stochastic Frontier Analysis or SFA). Of course, tackling the same problem using both methods is appropriate, but I would have liked to see more explicit comparison of the findings across chapters 4 and 5, and some argument about how the evidence should be synthesized. For example, SFA is better suited for estimating the effects of individual covariates such as tangible and intangible assets. (The finding in chapter 5 that SMEs do not seem to reap efficiency gains from possession of intangible assets like software and patents is extremely interesting and runs counter to the extant entrepreneurship literature – some elaboration is desirable!)*

Discussion of results of all empirical papers is included in section 2.5, including the comparison of parametric and non-parametric results. We briefly compare parametric and nonparametric methodologies in section 2.4.2, above all we mention some of the recent efforts to merge these approaches: "Given that there always exists a certain amount of rivalry between parametric and non-parametric approaches, it is interesting to note that recent developments include efforts towards the integration of both methodologies, following the comparative parametric/non-parametric studies such as Bardhan et al. (1998). Most notably, Kuosmanen (2006) proposed a method called *stochastic nonparametric evelopment of data*, combining components of non-parametric data envelopment analysis (DEA) and parametric SFA. This method was applied to electricity distribution networks in Kuosmanen (2012). Other approaches include Tsionas (2003) using Bayesian statistics, and Kumbhakar et al. (2007) using local maximum likelihood estimation."

4. Report by Cinzia Daraio

4.1. Detailed comments and suggestions

1. We suggest to change the title in a more informative one regarding the content of the work, in something like "Efficiency, Competition and Equilibrium". We think that this new title better illustrates the economic aspects treated by the work. Indeed, the aspects related to time and the dynamics are really marginally discussed in this PhD thesis.

Following the suggestion of the Reviewer, we changed the title to "Economic Efficiency, Competition and Equilibrium in Heterogeneous Production".

2. A re-structuring of the first two chapters is necessary. At present, in the current version of the work there is a discrepancy between the first two chapters and the last two ones.

See the response below.

3. In particular at this stage it is not clear if the PhD thesis is a monograph or a collection of papers. In both cases, the first chapter should introduce, contextualize and summarize the contents of the following chapters. Hence, some work is required to rewrite the first chapter after having clarified if the work is a monograph or a collection of essays.

As is explained in the response to general comments by B. Gebicka, the thesis is a collection of 3 empirical essays supported by an extensive synthesizing theoretical chapter 2, which was rewritten and extended. We also added a summarizing chapter 1 — executive summary.

4. By reading the first two chapters we understood that the key concepts of the work are competition and equilibrium, and their relations with efficiency. This should be the core of chapter 2. In particular an effort to better analyse their interrelations should be done and more literature on these issues should be considered.

We significantly expanded chapter 2.3.3, where we extensively discuss the dynamic aspects of competition and equilibria in the economic system.

5. Some updates of the methodology could be acknowledged in Section 3.4.5.2 (Estimation, pag. 51); Section 3.4.5.3. (Analysis of the Size Effect, pag. 52) and Section 3.5.3.2 (Conditional Size Effect, pag. 55). In particular you might refer to Badin et al. (2012a, 2012b).

References to Bădin et al. (2012a) and Bădin et al. (2012b) were added to the corresponding sections of the thesis.

6. In the applications, how the value m for the robust (order-m) frontier estimation has been chosen?

As is mentioned in section 3.4.3.3, the value of m was selected mainly taking into account the ratio of m to the total sample and computational aspects (computer speed) for the simulation.

References

Bardhan, I.R., Cooper, W.W., Kumbhakar, S.C., 1998. A simulation study of joint uses of data envelopment analysis and statistical regressions for production function estimation and efficiency evaluation. Journal of Productivity Analysis 9, 249–278.

Bădin, L., Daraio, C., Simar, L., 2012a. Explaining inefficiency in nonparametric production models: The state of the art. Annals of Operations Research online, 1–26. doi:10.1007/s10479-012-1173-7.

Bădin, L., Daraio, C., Simar, L., 2012b. How to measure the impact of environmental factors in a nonparametric production model. European Journal of Operational Research 223, 818–833. URL: http://www.sciencedirect.com/science/article/pii/S0377221712004833, doi:10.1016/j.ejor.2012.06.028.

Cooper, R.W., Haltiwanger, J.C., 2006. On the nature of capital adjustment costs. Review of Economic Studies 73, 611–633.

Cooper, W.W., Seiford, L.M., Tone, K., 2002. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Kluwer Academic Publishers, New York. ISBN: 0-306-47541-3.

Kumbhakar, S.C., Park, B.U., Simar, L., Tsionas, E.G., 2007. Nonparametric stochastic frontiers: A local maximum likelihood approach. Journal of Econometrics 137, 1–27. URL: http://www.sciencedirect.com/science/article/B6VC0-4JVSWWW-2/2/565ae92c913f8662e98a4134a14894d8, doi:10.1016/j.jeconom.2006.03.006.

Kuosmanen, T., 2006. Stochastic nonparametric envelopment of data: Combining virtues of SFA and DEA in a unified framework. MTT Discussion Paper http://ssrn.com/abstract=905758.

Kuosmanen, T., 2012. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. Energy Economics 34, 2189–2199. URL: http://www.sciencedirect.com/science/article/pii/S0140988312000497, doi:10.1016/j.eneco.2012.03.005.

Tsionas, E.G., 2003. Combining DEA and stochastic frontier models: An empirical Bayes approach. European Journal of Operational Research 147, 499–510.