Univerzita Karlova v Praze
Fakulta sociálních věd
Institut ekonomických studií

# BAKALÁŘSKÁ PRÁCE

2008 Jan Průša

Univerzita Karlova v Praze
Fakulta sociálních věd
Institut ekonomických studií

**BACHELOR THESIS**

# Estimating the Efficiency of Small and Medium Enterprises in the Czech Republic

Author: Jan Průša
Tutor: doc. ing. Vladimír Benáček, CSc.
Academic year: 2007/2008

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

V Praze dne 21. května 2008

Jan Průša

I would like to thank to Mr. Vladimír Benáček for his permanent support.

**Abstract.**

This thesis analyzes the efficiency of Czech small medium enterprises. We use the data from 2002 to 2005 of thirty manufacturing industries, each divided into five subgroups according to the number of employees. We combine two types of methods: We employ standard and advanced robust data envelopment analysis (DEA) to obtain cross-sectional rankings; and stochastic frontier analysis (SFA) to make statistical inference on the production process.

The results reveal substantial variance in the efficiency scores, which is only partly removed by the robust DEA specification. Both SFA and DEA suggest an inverse relationship between size in terms of employees and efficiency, but this relationship is not statistically significant. On the contrary, our tests strongly support the presence of a systematic gap between common practice and best practice.

Our results further demonstrate that Czech SME depend more on labour than on capital. The impact of investment or intangible assets such as software or patents is negligible. Finally a simple test for time effect shows that between 2003 and 2005 Czech SME moved towards higher efficiency.

**Abstrakt.**

Práce analyzuje efektivnost českých malých a středních podniků. Použita jsou data z let 2002 až 2005 o třiceti průmyslových odvětvích, z nichž každé je rozděleno do pěti podskupin podle počtu zaměstnanců. V práci kombinujeme dva metodologické postupy: Pomocí standardní a pokročilé robustní obálkové metody (data evelopment analysis, DEA) je určeno vzájemné pořadí sektorů; stochastická analýza produkčních funkcí (stochastic frontier analysis, SFA) umožňuje statistické testy o produkčním procesu.

Výsledky odhalují značné rozdíly v efektivnosti, které jsou jen zčásti odstraněny použitím robustní specifikace DEA. Obě metody, SFA i DEA, vykazují inverzní vztah mezi velikostí charakterizovanou počtem zaměstnanců a efektivností, ale tento vztah není statisticky signifikantní. Naopak naše testy silně podporují přítomnost systematického rozdílu mezi nejlepší a běžnou úrovní efektivnosti.

Naše výsledky dále potvrzují, že české malé a střední podniky závisí více na využití práce než kapitálu. Vliv investic nebo nehmotných aktiv jako softwaru nebo patentů je zanedbatelný. Konečně jednoduchý test na přítomnost časového efektu ukazuje, že mezi roky 2003 a 2005 se podniky posunuly směrem k větší efektivnosti.

Le savant doit ordonner; on fait la science avec des faits comme une maison avec des pierres; mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison.

The Scientist must set in order. Science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house.

*Henri Poincaré* *

*Quoted from Wikiquote.org: [http://en.wikiquote.org/wiki/Henri_Poincaré], original quote from Chapter 9 of Hypotheses in Nature, as translated by George Bruce Halsted (1913).

# Contents

# List of Tables

# Introduction
# A Road Map

This thesis offers the reader revealing insights into the industrial fundamentals of the Czech economy. Not that the fundamentals were, or even had to be, industrial; it only means that we focus exactly on the part of the Czech economy which is industrial.

The text contains a couple of mathematical expressions, hence it is crucial to keep in mind the guiding thought of this analysis. We want to compare the performance of industrial sectors of Czech small and medium enterprises (SME). Accordingly we present an empirical cross-sectional inquiry, which we obtained by applying the methodology of efficiency measurement on Czech statistics.

As is usual for empirical research, we are confronted with tensions between theory and reality. While the object—SME—is precisely defined, the statistics on SME are not so precisely measured and not completely available. Although 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 structure of the thesis follows from the initial research idea. In chapter 1, we give the reader a basic briefing on small and medium enterprises in general and particularly in the Czech Republic.

Chapters 2 and 3 deal with the methodology of our analysis. The former chapter estabilishes a conceptual framework for the microeconomic analysis of production and describes the reasoning behind the term "efficiency". The latter chapter reviews techniques which are available for efficiency measurement. Our overview is complex in that it covers the foundations of the two alternative approaches which have been used in economic literature: stochastic frontier analysis (SFA); and data envelopment analysis (DEA). Since however lots of modifications were developed over years, even the comprehensive handbooks (Kumbhakar & Lovell (2000), Cooper *et al.* (2002), Cooper *et al.* (2004), Coelli *et al.* (2005)) listed in the bibliography of this thesis are by far not exhaustive. We proceed to two specifications which we find suitable for our data and which are treated in more detail.

Finally chapter 4 forms the core of our genuine research. We analyze the dataset on Czech small and medium enterprises for the period 2002 to 2005.[2] Both DEA and SFA are used to obtain industry-specific efficiency scores. This allows us to unveil the structural patterns in the Czech SME sector. Standard regressions are computed in order to formulate a general production function of SME and to estimate the sensitivity of the factors of production. Also by means of SFA we test whether size influences efficiency.

To the best of our knowledge, this thesis is the first application of the data envelopment analysis on Czech statistics. In the case of the Czech Republic, the analysis of a fairly detailed dataset is itself rare for two reasons: (1) the statistics are not widely available, and (2) the methodology of data collection on the micro level changed several times. This also explains why we were not able to consistently compare a longer time period.

Estimating the production function on the macroeconomic level is a common economic exercise, and appeared in several papers on the Czech economy. To name just one example: Hájek (2005) tracked the determinants of economic growth given by the model of Solow. Yet our approach differs from that of Hájek and similar studies, since we are deriving the model from microeconomics. By plugging in aggregated data, the model of course shifts towards macroeconomics, but that is a pinch of compromise which belongs to the recipe.

While macroeconomic phenomena such as the impact of foreign direct investment have received extensive research coverage, literature on small and medium enterprises is rather scarce. One pioneering study is by Benáček *et al.* (1997), who measured efficiency of textile and clothing firms by distance functions. In fact this idea serves as the basis for the formulation of a standard DEA model. Thanks to detailed information on individual firms, Benáček *et al.* were even capable of separating technical and allocation efficiency. Due to data protection, this will probably never more be possible in any European Union country.

Still this thesis shall demonstrate that efficiency analysis can be meaningful even with aggregated data, the most representative example of which is the list of top performing industries: such information might be useful for investment of any kind.

[2] The dataset was kindly provided to us by the Czech Statistical Office. Refer to section 4.1 and appendices A and B.

# Chapter 1

# Small and Medium Enterprises

## 1.1 Definition & Background

In economics the distinction between small, medium and large firms appears trivial, at least without considering relevant market structures. It is therefore useful to recall the roots of such terminology.

The particular characteristics of the European economies tell one part of the story. When the government plays an important role and favours the so called 'national champions', smaller entrepreneurs will attempt to lobby for their interests with still more determination. Since the group counts lots of voters, 'small and medium enterprises' quickly enter the latest political newspeak. This could be observed recently in the whole European Union, which itself acts as a self-appointed SME's guardian.

The other plot turns our attention from daily newspapers to respected journals. Besides management science, the size of companies has been under close scrutiny of economists in terms of internal organization of the firm or various joint applications of law and economics. Yet examination of contracts or legal types of companies goes far beyond the simple size parameter included in the definition of SME.

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 the term SME.

*Small and medium enterprises*, abbreviated as *SME*[1], are defined as companies not exceeding specific size limits. The simplest classification, such as that of the World Bank, relies solely on the number of employees—the WB uses 250 as the limit. In the European Union SME have become an important tool for economic policy, so that the definition is accordingly sophisticated, see table 1.1. Note that a firm must satisfy the first condition and either one of the last two conditions at the same time

[1]Sometimes the abbreviation 'SME' stands for 'small and medium entrepreneurs'.

in order to be classified as SME.

| 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 1.1: Definition of SME according to the EU legislation.

Lots of countries created their own definitions, eg Switzerland or the USA take 500 employees as the cutoff (Ayyagari *et al.*, 2007).[2]

## 1.2 SME around the World

### 1.2.1 The Importance 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, SME matter because they form an economy's fundamentals. They can be regarded as ants, who impact little individualy but hugely altogether. Small entrepreneurs build the economy from the bottom, so that they are the true discoverers of market gaps that call for filling.

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

From table 1.2 follows that in modern market economies, SME employ between one half to three quarters of the workforce in manufacturing. It is true that we create an artificial breakdown point, nonetheless it becomes apparent that size of businesses matters—to people employed there.

Exceptions are twofold. Post-soviet countries who have not yet undergone full transition show negligible SME sectors, eg Belarus, Georgia or Ukraine. However in

[2]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.

| Country | GDP/capita | SME 250 | Country | GDP/capita | SME 250 |
|---|---|---|---|---|---|
| Austria | 29619.35 | 66.10 | Italy | 19218.46 | 79.70 |
| Belarus | 2522.94 | 4.59 | Japan | 42520.01 | 71.70 |
| Belgium | 27572.35 | 69.25 | Luxembourg | 45185.23 | 70.90 |
| Brazil | 4326.55 | 59.80 | Netherlands | 27395.01 | 61.22 |
| Bulgaria | 1486.74 | 50.01 | Poland | 3391.08 | 63.00 |
| Croatia | 4453.72 | 62.00 | Portugal | 11120.81 | 79.90 |
| Czech Republic | 5015.42 | 64.25 | Romania | 1501.08 | 37.17 |
| Denmark | 34576.38 | 68.70 | Russian Federation | 2614.38 | 13.03 |
| Estonia | 3751.59 | 65.33 | Slovak Republic | 3651.45 | 56.88 |
| Finland | 26813.53 | 59.15 | Spain | 15361.80 | 80.00 |
| France | 27235.65 | 67.30 | Sweden | 27736.18 | 61.30 |
| Georgia | 736.79 | 7.32 | Taiwan, China | 12474.00 | 68.60 |
| Germany | 30239.82 | 59.50 | Turkey | 2864.80 | 61.05 |
| Greece | 11593.57 | 86.50 | Ukraine | 1189.84 | 5.38 |
| Hungary | 4608.26 | 45.90 | United Kingdom | 19360.55 | 56.42 |
| Ireland | 19528.13 | 67.20 | | | |

GDP/capita = Real GDP per capita in USD. SME 250 = Share of the SME sector in 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 1.2: Share of SME on employment in selected countries (Ayyagari *et al.*, 2007).

this countries lots of prospective entrepreneurs take part in the informal economy, not captured by official statistics, so that the true percentage is a bit higher.

The other exception is the USA, where the share on labour force is 52.54 per cent, but with 500 employees as the yardstick. It illustrates that the world's biggest economy has quite different dimensions from Europe.

### 1.2.2 What Determines the Size of the SME Sector?

According to a widespread argument, a strong sector of competitive small and medium enterprises heavily depends on the quality of business environment. State bureaucracy is only one of a dozen factors selected by the World Bank to be measured in its global survey of business environment.[3] 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 playing field adds 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 (2001) explored the hypothesis that size explains part of the variance in responses to the above mentioned survey of the World Bank. 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 did large firms. The significance of this strong finding declined (though not disappeared) when they split up the sample to regional groups. In particular, in the then 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 at the time of the survey in 1990's.

Schiffer & Weder conclude that size matters with respect to the perceived obstacles and do not elaborate much on the presence or absence of this effect in the subsamples. But they might indeed overlook the quite considerable implication of their results: namely that more of the "free market" leads to less "size discrimination". In other words, this would buttress the argument that liberal market reforms do equalize conditions for market players and that SME deserve special political treatment—in the sense of cutting red tape.

Another study by Ayyagari *et al.* (2007) tested two hypotheses. Large SME sectors may stem from high exit costs and government subsidies, so that they are prevented to grow or to disappear (negative reasoning). The other explanation argues that large amount of SME results 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.

[3]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 (Schiffer & Weder, 2001, p. 14).

We can translate their investigation to a simple imperative: Governments must not crack down on natural entrepreneurship if their 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; but in this case we would hardly expect the opposite.

Our second look at SME would consider the vital importance of small firms to economic advancement. A few paragraphs above we noted that entrepreneurs often start from scratch and thus embody the ability and will to learn and create. In his case study on Turkey, Taymaz (2005) concludes that "most firms start small", moreover at their distinct disadvantage: both their scale and efficiency are suboptimal. It follows that these businesses have to achieve higher rates of growth in order to survive. That small firms grow faster is exactly the finding of the recent study by Mohnen & Nasev (2005), 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 is not the subject of our thesis and we skip more details at this point.

## 1.3 Foundations so Tiny

### 1.3.1 Czech SME Sector

In this section we present a general survey on Czech small and medium enterprises.

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 Vladimír Benáček (1994). 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 1.3 quotes statictics on SME published by the Ministry of industry and trade in its "Report on the development of SME and its support in 2006".[4]

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

| 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 B. Included are all entrepreneurial activities in manufacturing, construction, commerce and a part of services.

Table 1.3: Share of SME on macroeconomic indicators.

Selected indicators are depicted in figure 1.1. 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 breaktrough 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.

In our analysis will not deal with international trade of SME, but we will assess the impact of investment on their efficiency.

### 1.3.2 Czech Business Environment

The World Economic Forum compiles an extensive report on business environment (Lopez-Carlos, 2006). Based on several economic indicators as well as proxies for business environment and responses to the Executive Opinion Survey, the authors defined the 'Global Competitiveness Index' and accordingly ranked between 80 to

125 countries of the world. The yearly reporting changed significantly between 2002 and 2006, in order to provide a more comprehensive set of indicators. However positive evolvement, it makes comparison of a country's performance over time tiresome and imprecise.[5]

| **Notable comptetitive advantages** | | **Notable competitive disadvantages** | |
|---|---|---|---|
| Description / Rank 2005 (2003) | | Desription / Rank 2005 (2003) | |
| FDI and technology transfer | 3 (4) | Real effective exchange rate | 108 (85) |
| Quality of math and science education | 6 (17) | Efficiency of the tax system | 83 (75) |
| Cellular telephones | 7 (8) | Government surplus/deficit | 81 (61) |
| Extent of the bureaucratic red tape | 10 | Wastefulness of government spending | 78 (80) |
| Pay and productivity | 16 (11) | Hiring and firing practices | 77 |
| Centralization of economicy policymaking | 19 (18) | Soundness of banks | 73 (78) |
| Quality of electric supply | 20 | Government prioritization of ICT | 73 (62) |
| Private sector employment of women | 22 | Effectiveness of law-making bodies | 69 (67) |
| Foreign ownership restrictions | 22 | Tax burden | 67 |
| Capacity for innovation | 23 | Public trust of politicians | 66 (65) |

Rank is out of 117 countries in 2005 and 102 countries in 2003. In the report only extreme indicators are reported. When some indicators for 2003 are not shown, it means that Czech Republic placed close to its average and thus different, more outlying indicators were published instead.

Table 1.4: Czech business environment according to the World Economic Forum, selected indicators (Lopez-Carlos, 2005), (Sala-i-Martin, 2003).

To get some picture about the Czech business environment, we reproduce table 1.4 based on two of these reports, and figure 1.2, which depicts the answers of Czech businessmen to the World Bank survey mentioned in the previous section and where higher score means more severe obstacle to doing business.

[5] The current index itself is a was created by merging two former main rankings reported before 2006, which seemingly led to some lapses. The report 2006–2007 says that Czech Republic placed 29th in the GCI index (Lopez-Carlos, 2006, p. xvii), the same position as in the year before. Yet in the report 2005–2006, Czech Republic ranked 38th, allegedly according to the same index (Lopez-Carlos, 2005, p. xvii). In fact Czech Republic was 27th in the 'Business Competitiveness Index' and 38th in the 'Growth Competitiveness Index' in 2005–2006.

Their content does not surprize: In 2005–2006 Czech economy was supposed to be FDI- and technology-ready, but featured three major shortcomings: (1) weak fiscal governance, (2) rigit labour market and (3) Czech crown appreciating at a cracking pace. Moreover these factors mostly did not improve, quite the other way round, they either stagnated or even worsened (yet we have to take into account that the number of countries included in the report also increased).

One particular difficulty perceived by entrepreneurs in the Czech Republic is the inefficiency of the tax system. This fact does not come down just to the tax rate, which has been gradually decreasing over the past seven years or so. The report by CERGE-EI (Dušek & Žigić, 2005, p. 41) documents the rising perplexities of the Czech income tax law. Between March 1993 and January 2005 the number of words in the law quadrupled from less than 20,000 words to more than 81,000. The phrase "with exception of" appeared 254 times in the law, compared to zero in 1993. No wonder that Czech Republic was reported as the country with the highest costs of tax collection among OECD members. The ratio of administrative cost to revenue collections was 2.08 per cent in 2002, compared to 1.46 in Slovakia, 1.32 in Poland or 0.52 in the USA (Dušek & Žigić, 2005, p. 42).

### 1.3.3 Detecting Common Patterns

Unfortunately with the available statistics we are not able to examine hypotheses about institutional factors which we just mentioned. The purpose of section 1.3.2 rather was to visualize the vast range of factors which contribute to the general economic environment, to which small and medium enterprises are especially sensitive.

The following chapters will focus on cross-sectional analysis. Besides structural results, we are particularly concerned with what stands behind the table 1.3. 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 1.3 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 thesis. However we are also aware that figure 1.1 is related to a broader group of SME than that represented by our dataset.

After the descriptive exposition, we can turn to the methodology.

Figure 1.1: Share of SME on macroeconomic indicators, table 1.3.

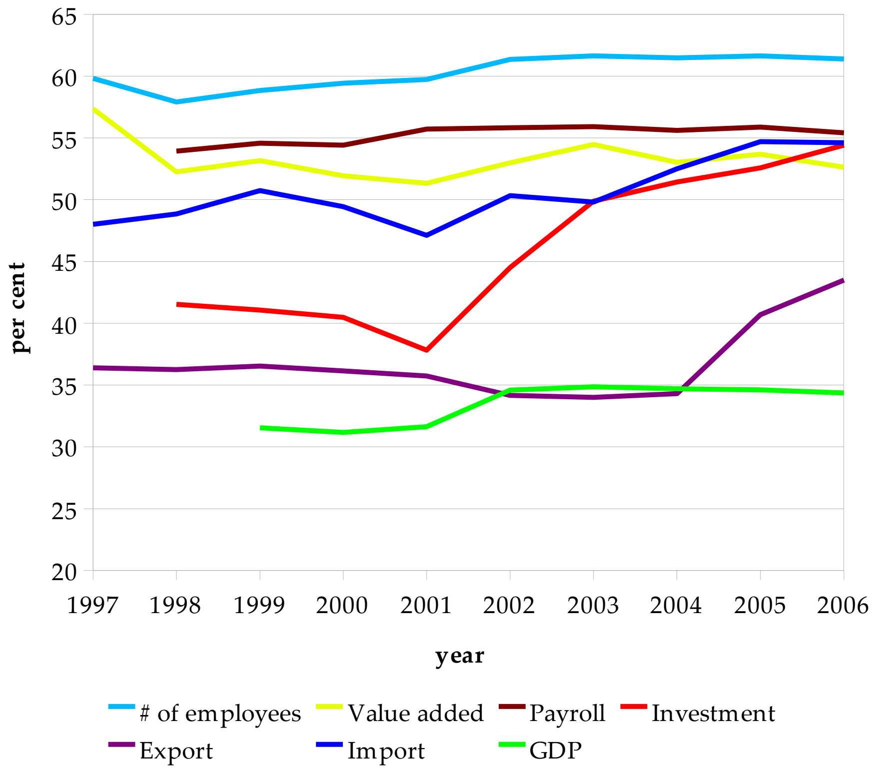
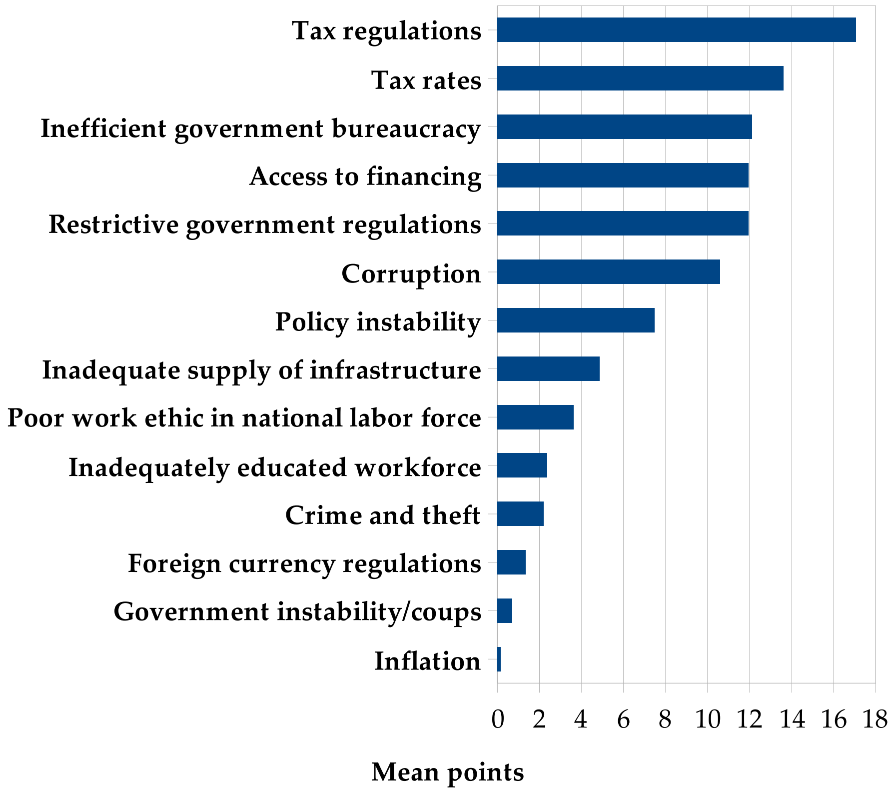



Figure 1.2: The Most problematic factors for doing business (Lopez-Carlos, 2005).



# Chapter 2

# Production Theory

## 2.1 A Model of Production

Our objective is to analyze the sector of small firms in the Czech Republic. Over years microeconomics has developed sophisticated ways to describe firms' behaviour. Before we turn to detailed concepts of efficiency, it is useful to formulate a general framework.

The starting point of production analysis is set to be profit maximization, which is formulated eg by Varian (2006, p. 334):

> [...] the firm chooses a production plan so as to maximize its profits.

Under some regularity conditions which we do not list here it does not matter which side of the subtraction we solve: either revenue maximization or cost minimization both lead to maximum profit.

For our thesis a slightly different point of view is more important. The production proces 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.

### 2.1.1 Production Function

We briefly recall the neoclassical production function, here outlined as in Nicholson (1992).[1] Assume there to be two productive inputs, labour $L$ and capital $K$, which are used in production of a particular good. The maximum amount that can be produced from a certain amount of $(K, L)$ is given by the functional form

$$y = f(K, L).$$

[1]Detailed concepts of production functions can be found in Sato (1975) and Johansen (1972).

The production function captures technical parameters and is usually assigned several important properties. We will list them later when we set up a more general desription of the production technology, here we mention only one of them. Additional output from one additional unit of input, the other input held constant, is positive but decreasing with more input:

$$MP_K = \frac{\partial f}{\partial K} > 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial K^2} < 0\,,$$
$$MP_L = \frac{\partial f}{\partial L} > 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial L^2} < 0\,,$$

where we assume that these derivatives exist and $MP$ stands for marginal product.

We define returns to scale as a response of output to an equiproportionate increase in all inputs by a factor $a > 1$, which we write as

$$f(aK, aL) \lesseqgtr a \cdot f(K, L).$$

We will refer to the respective relationship as decreasing, constant or increasing returns to scale.

More generally we can extend inputs to a $p$-dimensional vector $\boldsymbol{x} \in \Re^p_{0,+}$, so that $f(\boldsymbol{x})$ represents the complete technical relationship between inputs. It characterizes the technology available to the firm and describes how all inputs inclusive land, materials, goodwill etc. are transformed into the $r$-dimensional output vector $\boldsymbol{y} \in \Re^r_{0,+}$. Other properties of $f(\cdot)$ remain the same as hereinbefore.[2]

So far we were concerned with technology. However the economy is governed by prices and firms attempt to maximize their economic profits. They have to adjust the technological knowledge to economic conditions and to choose the profit maximizing set of inputs and outputs. We do not elaborate in detail on profit maximization. As a brief reminder, let us mention the case of perfect competition, where for a small firm the price of the good $P$ is given on the market. Denote the price of labour $w$ (wage) and the price of capital $r$ (rent). Then the link to economic parameters is desribed by a simple profit function where the firm maximizes the profit as the difference between total revenue and total cost. This yields the first order condition that in equilibrium marginal revenue from each the unit of output equals its marginal cost:

$$\begin{aligned} \max_y \Pi &= \max_y \big(\underbrace{P \cdot f(K, L)}_{TR} - \underbrace{(wL + rK)}_{TC}\big), \\ MR = \frac{\partial TR}{\partial Q} &= \frac{\partial TC}{\partial Q} = MC. \end{aligned}$$

[2] There are however definition extensions in the multidimensional case, eg for elasticity of substitution. (Nadiri, 1982)

### 2.1.2 Production Set

Instead of exact mathematical formulae we can define firms' production in a more general way using the set theory. The production set is defined as all feasible vectors $(\boldsymbol{x}, \boldsymbol{y})$ (Daraio & Simar, 2007a):

$$\Psi = \{(\boldsymbol{x}, \boldsymbol{y}) |\, \boldsymbol{x} \in \Re^{p}_{0,+}, \boldsymbol{y} \in \Re^{r}_{0,+} \text{ is feasible}\}.$$

By analogy to production functions we can dissect this multidimensional set and find its borders (isoquants). The output correspondence set consists of all possible output vectors $\boldsymbol{y}$ which a firm can produce from various possible input vectors $\boldsymbol{x}$:

$$P(\boldsymbol{x}) = \{\boldsymbol{y} \in \Re^{r}_{0,+} | (\boldsymbol{x}, \boldsymbol{y}) \in \Psi\}.$$

Several assumptions on the characteristics of the underlying technology are connected with this definition (Kogiku, 1971):

- ⋆ Continuity: $P(\boldsymbol{x})$ is closed.
- ⋆ Possibility of inaction: $\mathbf{0} \in P(\boldsymbol{x})$.
- ⋆ Impossibility of free production: $P(\mathbf{0}) = \mathbf{0}$.
- ⋆ Irreversibility of production, ie $\boldsymbol{x}$ and $\boldsymbol{y}$ cannot switch roles.
- ⋆ Free disposal.
- ⋆ Constant returns to scale: $P(\boldsymbol{x})$ is convex.

The last one is often relaxed, implications for practical application will be discussed later. The isoquant is the border of the production set, which we write as:

$$\text{Isoq}P(\boldsymbol{x}) = \{\boldsymbol{y} : \boldsymbol{y} \in P(\boldsymbol{x}), \forall\, \lambda \in (1, +\infty) : \lambda \boldsymbol{y} \notin P(\boldsymbol{x})\}.$$

This isoquant is directly related to the production function introduced in the previous section: If it is a smooth curve, then Isoq$P(\boldsymbol{x})$ corresponds to the contour line of $f(\boldsymbol{x})$. Also in this case the firm will maximize the profit due to the same principle as hereinbefore. Yet we want to examine more general (and more realistic) shapes of Isoq$P(\boldsymbol{x})$.

We will now use the set description not only to accommodate economic parameters, but also to define efficiency.

## 2.2 Productive Efficiency

The production function is an ideal concept where no frictions exist. In real world inefficiencies occur and not all producers are able to reach the maximum possible output: $\boldsymbol{y} < f(\boldsymbol{x})$, meaning that the firm will operate somewhere inside the production set. Also once a firm achieves the maximum output, it can still be inefficient in terms of prices, since it may use a combination of inputs and/or produce a vector of outputs that do not maximize the profit for given prices.

It follows that productive efficiency has two components, as in Lovell (Université Catholique de Louvain: CORE, 1992, reading 15):

1. Technical efficiency, which measures whether production reaches the maximum output given by available technology.

2. Economic efficiency, which measures whether the combination of inputs and outputs is the most profitable according to the current market prices.

### 2.2.1 Technical Efficiency

Lovell (cf.) quotes the definition of technical efficiency by Koopmans:

> A producer is technically efficient if an increase in any output requires a reduction in at least one other output or an increase in at least one input, and if a reduction in any input requires an increase in at least one other input or a reduction in at least one output.

To rewrite this definition in terms of production sets, we have to introduce one additional definition. Generally only some parts of the isoquant are technically efficient by the definition of Koopmans, since the isoquant need not be necessarily strictly convex. For this reason we define the efficient subset of an isoquant as:

$$\mathrm{Eff}P(\boldsymbol{x}) = \{\boldsymbol{y} : \boldsymbol{y} \in P(\boldsymbol{x}), \forall \boldsymbol{y'} > \boldsymbol{y} : \boldsymbol{y'} \notin P(\boldsymbol{x})\}.$$

Then it is easily recognized that $\mathrm{Isoq}P(\boldsymbol{x}) = \mathrm{Eff}P(\boldsymbol{x})$ iff the isoquant is strictly convex.

Then a producer will be technically efficient iff $\boldsymbol{y} = f(\boldsymbol{x})$ or $\boldsymbol{y} \in \mathrm{Eff}P(\boldsymbol{x})$.

### 2.2.2 Economic Efficiency

It would not make much sense for a firm to produce goods at a cost or for a price that nobody buys them. Instead producers need to adapt to market prices of outputs $\boldsymbol{p}$

and of inputs $\mathbf{w}$, which are graphically represented as isorevenue and isocost lines. Until now we took output as the target variable, which is linked to revenues. Similar considerations can be applied to inputs, and the duality theorem ensures that the outcome will be the same.

The effort to maximize revenue or minimize profit can be described as functions:

$$\begin{aligned} R(\boldsymbol{x}, \boldsymbol{p}; \boldsymbol{\beta}) &= \max_{\boldsymbol{y}} \{\boldsymbol{p}^T \boldsymbol{y} : \boldsymbol{y} \in P(\boldsymbol{x}; \boldsymbol{\beta})\}, \\ C(\boldsymbol{y}, \boldsymbol{w}; \boldsymbol{\beta}) &= \min_{\boldsymbol{x}} \{\boldsymbol{w}^T \boldsymbol{x} : \boldsymbol{x} \in L(\boldsymbol{y}; \boldsymbol{\beta})\},^{3} \end{aligned}$$

where $\boldsymbol{\beta}$ is the vector of parameters characterizing the production technology. The presence of this factor indicates that different firms might have different technologies installed. We can assume $\boldsymbol{\beta}$ away and suppose all firms with the same products to 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.

### 2.2.3 Overall Efficiency

Accordingly a producer is called 'efficient' iff he reaches the maximum revenue on the efficient subset of the production set, or equivalently on the production function:

$$\begin{aligned} \mathrm{Eff}R(\boldsymbol{x}, \boldsymbol{p}; \boldsymbol{\beta}) &= \max_{\boldsymbol{y}} \{\boldsymbol{p}^T \boldsymbol{y} : \boldsymbol{y} \in \mathrm{Eff}P(\boldsymbol{x}; \boldsymbol{\beta})\}, && (2.1) \\ \mathrm{Eff}R(\boldsymbol{x}, \boldsymbol{p}; \boldsymbol{\beta}) &= \max_{\boldsymbol{y}} \{\boldsymbol{p}^T \boldsymbol{y} : \boldsymbol{y} = f(\boldsymbol{x}; \boldsymbol{\beta})\}, && (2.2) \end{aligned}$$

where $f(\cdot)$ is characterized by the same vector of parameters $\boldsymbol{\beta}$ as in the first line. This could again be reformulated for cost minimization. For graphical illustration see figure 3.1.

Such analysis contends that prices are an exogenous factor, but their nature can change due to market structure. Surely bargaining the best prices belongs to key business skills and as such greatly contributes to efficiency. Yet enquiring theoretically into price dynamics goes beyond the scope of this thesis.

We realize that the main snag of any efficiency measure is to separate the two components of efficiency. The technical part is captured in data about production given in some physical units. If we assign certain prices to this volumes, we can trace the economic part. The ideal statistic would contain all these pieces of information

[3] L($\mathbf{y}$) is the input requirement set mirroring $P(\boldsymbol{x})$: $L(\boldsymbol{y}) = \{\boldsymbol{x} \in \Re^{p}_{0,+} | (\boldsymbol{x}, \boldsymbol{y}) \in \Psi\}$.

for a large number for indivitual producers; such data is however rarely available, less so in the Czech Republic.

## 2.3 Sources of Inefficiency

Real world is different from economic models, since assumptions do not always hold and other deficiencies occur. Keynes started the wave of a more realistic description of the world, which evolved into looking for second best solutions. Before we measure the distance between the best practice and the second best, we shall explain several reasons why firms are not perfectly efficient.[4]

### 2.3.1 External Factors

One concept of inefficiency concerns the dynamics of production. Especially with advanced technologies, production units invest a large share of capital in fixed assets which cannot be easily turned into other production procedures. Then proper investment planning should ensure efficient production in the future. But future development of markets cannot be completely anticipated due to the complexity of interactions. Because firms will not able to adjust long term investment to the actual economic situation, uncertainty about the future will lead to suboptimal results and cause inefficiencies.

Another bottleneck may stem from inappropriate institutional setting. The more the state interferes in entrepreneurial activities, the higher the risk that something will go wrong. It is impossible to estimate exactly in advance the fallouts of a legal ruling. Before the adaptive process of firms is finished, inefficiencies prevail. Moreover, entrepreneurs cannot know what kind of legislation will by adopted by the parliament in the following years, which hampers their planning even more.

### 2.3.2 Internal Factors

The simplest explanation of technical inefficiency is that different companies have technologies of different vintages at their disposal. This is a direct outcome of the fixed nature of one part of a firm's assets. Such situation is only natural in a dynamic environment and with the current pace of technology change becomes the more significant.

[4] The first explicit study on inefficiency other than allocative inefficiency in economic literature was by Liebenstein (1966).

Other inefficiencies result from the internal organization of the firm. Management techniques will crucially influence a firm's performance, as will the staff and their behaviour. Even in the same firm different amount of goods is produced on different days due to unexpected failures and complications. Cooperation between employees is not faultless and such transaction costs increase with the size of the firm. SME are affected least, but certainly not negligibly.

The scale of inefficiency will be estimated in the following chapters. It is noteworthy that some schools of thought heavily criticize the concept of equilibrium in economics. On the way to the state of balance, their argument goes, the original conditions change and so does the equilibrium itself. Hence any 'perfect equilibrium' remains permanently out of reach, which constitutes no pathology.

From this point of view, studying efficiency may not appear reasonable. Nevertheless, our approach incorporates one strong advantage: As any method that works with real-world data, instead of modelling a perfect situation, we are trying to track the best practice. Companies ranked efficient in one country do not necessarily have to be efficient in another. Such flexibility seems a harsh bargain, but it underlines that we are far from any 'planning' in the socialist sense of the word.

# Chapter 3

# Efficiency Measures

Theoretical concepts of production introduced in the previous chapter rest on the idea of "maximum amount of outputs that can be produced out of given inputs". Determining this maximum feasible amount is fundamental for all practical applications. The more so that neoclassical tradition reckons only efficient players will persevere in the market, and accordingly rather prescinds from inefficiencies. But we would like to look at the state of affairs and try to identify those likely to persevere and those inefficient, and for that matter find the maximum.

Introductory microeconomic tasks start with explicit production or cost functions which allegedly describe these maxima. These formulations define the best production per se, and hence they do not guarantee that these production patterns cannot be improved.

It seems necessary to bear in mind the governing thought of the subsequent methods of efficiency measurement. We assign the cetral role to competition without which the term efficiency would make only little sense. In sports there is no best performance without competitors, only the race appoints medal winners. The same holds for competitive efficiency: We describe methods for performing comparative analysis in order to find out best practice and award medals, which in this case mean prospects of profitability.

We shall note that the described models will not consider dynamics of production, ie any changes of production patterns occuring in time. That is, we will handle cross-sectional models, which can further be split into groups (Daraio & Simar, 2007a)

1. according to the statistical nature:

    - ⋆ **Deterministic**—No randomness in the model.
    - ⋆ **Stochastic**—Random variables and disturbances included.

2. according to the treatment of technical parameters of production:
   - ⋆ **Parametric**—Tracking parameters which govern the production process.
   - ⋆ **Non-parametric**—Efficiency is measured without explicitely knowing the parameters.

Finally let us introduce notation:

$$\begin{array}{rll} \text{DMU}_i & i = 1, \dots, n & \text{decision making units;} \\ x_j & j = 1, \dots, p & \text{inputs;} \\ y_k & k = 1, \dots, r & \text{outputs.} \end{array}$$

Assuming all vectors to be columns, this yields input matrix $X \in M(p \times n)$ and output matrix $Y \in M(r \times n)$. By 'dataset' we refer to these two matrices $X, Y$. Each decision making unit $i$ is characterized by a pair of vectors

$$[\boldsymbol{x}_i, \boldsymbol{y}_i] = \left[ (x_{1i}, \dots, x_{pi})^T, (y_{1i}, \dots, y_{ri})^T \right].$$

## 3.1 Parametric Approach

### 3.1.1 Estimator of Technical Efficiency

One straightforward method to assess efficiency is to apply regression analysis on the observed data. This analysis aims at (2.2) and primarily attempts to extract an estimate $\hat{\boldsymbol{\beta}}$ from $X, Y$. Then each DMU can be compared with this estimate and any deviations can be tested statistically for significance.

Let us consider the basic model for technical efficiency (Kumbhakar & Lovell, 2000, equation 3.2.18):

$$\begin{array}{rcl} y_i & = & f(\boldsymbol{x_i}, \boldsymbol{\beta}) \cdot \exp\{\nu_i - \tau_i\} \, ^1 \\ \nu_i & \dots & \text{random disturbance term,} \\ \tau_i \geq 0 & \dots & \text{inefficiency term.} \end{array} \tag{3.1}$$

The idea is that a general unknown production function is adjusted to differently productive firms by a multiplicative inefficiency term $\tau$ which is of one sign only.

This approach was pioneered in the commonly quoted paper by (Aigner & Chu, 1968). The authors noted that standard regression did not work well for efficiency analysis, because the obtained estimates tracked some kind of "average" production

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

function, a result lacking theoretical underpinning. For this reason they introduced the idea of bounded residuals and derived a purely deterministic model in the form of equation (3.1) without $\nu_i$.

The stochastic term $\nu_i$ was introduced nine years later by Aigner, Lovell and Schmidt and Meeusen and van den Broeck (Kumbhakar & Lovell, 2000, c.f.). Assume the initial functin to be:

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

Then we can rewrite the model (3.1) as the log-linear Cobb-Douglas production function to obtain:

$$\begin{aligned} \log y_i &= \beta_0 + \sum_{j=1}^{p} \beta_j \log x_{ij} + (\nu_i - \tau_i) \qquad (3.2) \\ \nu_i &\ \ldots\ iid,\ \mathcal{L}(\nu_i) \sim \mathcal{N}(0, \sigma_\nu^2). \end{aligned}$$

Ordinary least squares (OLS) yield estimates of $\beta_j$s, but we would also like to separate $\beta_0, \nu_i, \tau_i$ to obtain producer-specific efficiency scores. In order to get these, we need two additional assumptions:[2]

- ⋆ $\tau_i$ is *iid*, $\mathcal{L}(\tau_i) \sim \mathcal{N}^+(0, \sigma_\tau^2)$, so that we have the half-normal density function

$$g(\tau) = \frac{1}{\sigma_\tau}\sqrt{\frac{2}{\pi}} \cdot \exp\left\{-\frac{\tau^2}{2\sigma_\tau^2}\right\}.$$

- ⋆ $\tau_i$ and $\nu_i$ are independent on each other and on the regressors.

Denoting $\epsilon_i = \nu_i - \tau_i$, we can write the following densities:

$$\begin{aligned} g(\tau, \nu) &= \frac{2}{2\pi\sigma_\tau\sigma_\nu} \cdot \exp\left\{-\frac{\tau^2}{2\sigma_\tau^2} - \frac{\nu^2}{2\sigma_\nu^2}\right\}, \\ g(\tau, \epsilon) &= \frac{2}{2\pi\sigma_\tau\sigma_\nu} \cdot \exp\left\{-\frac{\tau^2}{2\sigma_\tau^2} - \frac{(\epsilon+\tau)^2}{2\sigma_\nu^2}\right\}, \\ g(\epsilon) &= \int_0^\infty g(\tau, \epsilon)\, \mathrm{d}\tau = \\ &= \frac{2}{\sigma\sqrt{2\pi}} \cdot \left[1 - \Phi\left(\frac{\epsilon\lambda}{\sigma}\right)\right] \cdot \exp\left\{-\frac{\epsilon^2}{2\sigma^2}\right\} = \\ &= \frac{2}{\sigma} \cdot \phi\left(\frac{\epsilon}{\sigma}\right) \cdot \Phi\left(-\frac{\epsilon\lambda}{\sigma}\right), \\ \sigma &= \sqrt{\sigma_\tau^2 + \sigma_\nu^2}\,, \\ \lambda &= \frac{\sigma_\tau}{\sigma_\nu}\,. \end{aligned}$$

[2] The following derivation is adapted from (Kumbhakar & Lovell, 2000, p. 72-81).

$\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal density and distribution functions.

This leads to the following maximum likelihood estimator (MLE):

$$\begin{aligned} \mathscr{L}(\epsilon, \sigma^2, \lambda) &= \prod_{i=1}^{n} \frac{2}{\sigma\sqrt{2\pi}} \cdot \exp\left\{-\frac{\epsilon_i^2}{2\sigma^2}\right\} \cdot \Phi\left(-\frac{\epsilon_i \lambda}{\sigma}\right), \\ \log \mathscr{L}(\epsilon, \sigma^2, \lambda) &= n \cdot \log \frac{2}{\sqrt{2\pi}} - n \cdot \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \epsilon_i^2 + \sum_{i=1}^{n} \log \Phi\left(-\frac{\epsilon_i \lambda}{\sigma}\right). \end{aligned} \tag{3.3}$$

Similarly it is possible to derive the conditional distribution of $\tau_i$ given $\epsilon_i$, 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 $\tau_i$ can be obtained as either the mean $E(\cdot)$ or the median $M(\cdot)$:

$$\begin{aligned} E(\tau_i|\epsilon_i) &= \mu_{*i} + \sigma_* \left[\frac{\phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)}{1 - \Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)}\right] = \\ &= \sigma_* \left[\frac{\phi\left(\frac{\epsilon_i \lambda}{\sigma}\right)}{1 - \Phi\left(\frac{\epsilon_i \lambda}{\sigma}\right)} - \left(\frac{\epsilon_i \lambda}{\sigma}\right)\right], \\ M(\tau_i|\epsilon_i) &= \begin{cases} -\epsilon_i \left(\frac{\sigma_\tau^2}{\sigma^2}\right) & \text{if } \epsilon_i \le 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

It only remains to transform the estimate back to the exponential form of (3.1) to obtain the estimate of technical efficiency $\widehat{TE}_i$:

$$\widehat{TE}_i = \exp\left\{-E(\tau_i|\epsilon_i)\right\},$$

the same holds for the mode. Yet there is still one complication. In our derivation, we assumed that inefficiency has multiplicative form, which we then transformed by taking logarithm. Thus it makes more sense to construct an enstimate including efficiency which is already transformed back. In other words we can write:

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

This problem was pointed out by Battese & Coelli and is mentioned by Kumbhakar & Lovell (2000). Accordingly they proposed an improved point estimator:[3]

$$E(\exp\left\{-\tau_i\right\}|\epsilon_i) = \left[\frac{1 - \Phi\left(\sigma_* - \frac{\mu_{*i}}{\sigma_*}\right)}{1 - \Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)}\right] \cdot \exp\left\{-\mu_{*i} + \frac{1}{2}\sigma_*^2\right\}. \tag{3.4}$$

[3] In the original article (Battese & Coelli, 1988), the authors defined the estimator for panel data, where $\tau_i$ was constant over time but $\nu_{it}$ was allowed to vary among periods.

If we want to estimate efficiency from data using equation (3.4), we need to estimate $\hat{\sigma}^2_\tau, \hat{\sigma}^2_\nu$ and obviously $\hat{\epsilon}_i$s themselves. We could solve (3.3), which requires an iterative algorithm (Aigner *et al.*, 1977).[4] Aigner *et al.* also noted that it is possible to find consistent estimates of variances from OLS residuals by the method of moments. From the density $g(\epsilon)$ we know:

$$\begin{aligned} E(\epsilon) = -E(\tau) &= -\sqrt{\frac{2}{\pi}} \cdot \sigma_\tau, \\ var(\epsilon) = var(\tau) + var(\nu) &= \left(\frac{\pi-2}{\pi}\right)\sigma^2_\tau + \sigma^2_\nu, \qquad (3.5) \\ E\left(\epsilon - E(\epsilon)\right)^3 &= \frac{2\sigma^2_\tau}{\sqrt{2\pi}} \cdot (1-4\pi). \qquad (3.6) \end{aligned}$$

Because $E\left(\hat{\epsilon}_i^{OLS}\right) = 0$, we have to use the second and third sample central moments, denote them $m_2$ and $m_3$, to solve the equations for $\hat{\sigma}^2_\tau, \hat{\sigma}^2_\nu$. If we further denote $\sigma_2$ and $\sigma_3$ the (theoretical) central moments, it holds:

$$\begin{aligned} E(m_2) &= \frac{n-1}{n} \cdot \sigma_2, \\ E(m_3) &= \frac{(n-1)(n-2)}{n^2} \cdot \sigma_3. \end{aligned}$$

Finally we use

$$\begin{aligned} s^2 &= \frac{1}{n-1} \cdot \sum_i \left(\epsilon_i - \frac{1}{n}\sum_i \epsilon_i\right)^2, \\ s^3 &= \frac{n}{(n-1)(n-2)} \cdot \sum_i \left(\epsilon_i - \frac{1}{n}\sum_i \epsilon_i\right)^3, \end{aligned}$$

as the estimators for the left hand sides of equations (3.5) and (3.6), where moreover $\frac{1}{n}\sum_i \hat{\epsilon}_i^{OLS} = 0$. This yields:

$$\begin{aligned} \hat{\sigma}^2_\tau &= \frac{\sqrt{2\pi}}{2-8\pi} \cdot \frac{n}{(n-1)(n-2)} \sum_i \left(\hat{\epsilon}_i^{OLS}\right)^3, \\ \hat{\sigma}^2_\nu &= \frac{1}{n-1} \sum_i \left(\hat{\epsilon}_i^{OLS}\right)^2 - \frac{\pi-2}{2} \cdot \hat{\sigma}^2_\tau. \end{aligned}$$

[4] We can adapt analytical first-order derivatives of (3.3) from Aigner *et al.* (1977) to our log-linear model:

$$\begin{aligned} \frac{\partial \log \mathscr{L}}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n \left(\log y_i - \boldsymbol{\beta}^T \log \boldsymbol{x}_i\right)^2 + \frac{\lambda}{2\sigma^3}\sum_{i=1}^n \frac{\phi_*}{1-\Phi_*}\left(\log y_i - \boldsymbol{\beta}^T \log \boldsymbol{x}_i\right), \\ \frac{\partial \log \mathscr{L}}{\partial \lambda} &= -\frac{1}{\sigma}\sum_{i=1}^n \frac{\phi_*}{1-\Phi_*}\left(\log y_i - \boldsymbol{\beta}^T \log \boldsymbol{x}_i\right), \\ \frac{\partial \log \mathscr{L}}{\partial \beta} &= \frac{1}{\sigma^2}\sum_{i=1}^n \left(\log y_i - \boldsymbol{\beta}^T \log \boldsymbol{x}_i\right) \cdot \log \boldsymbol{x}_i + \frac{\lambda}{\sigma}\sum_{i=1}^n \frac{\phi_*}{1-\Phi_*}\left(\log \boldsymbol{x}_i\right). \end{aligned}$$

Once we have these values at hand, we can easily compute the point estimates of efficiency by plugging $\hat{\sigma}_\tau^2$ and $\hat{\sigma}_v^2$ into (3.4).

### 3.1.2 The Economic Dimension

The deliberations so far took into account just the world of technology. In section 2.2.2 we stressed that economics primarily focuses on allocation efficiency. Hence we ought to extend the theory to revenue and cost functions, which would allow us to track the economic performance of analyzed production units.

The underlying idea for the design of an estimator is similar to what we presented in the previous chapter, only the algebra is more complicated. A comprehensive overview can be found in Kumbhakar & Lovell (2000). We do not pursue their derivation for a simple reason: we cannot apply them anyway, and we shall explain this point.

Our available data described later in the thesis are measured in monetary units. Thus the proxies for inputs and outputs contain the information on both technical and allocation efficiency at the same time.[5] With such data we cannot construct profit functions and decompose overall efficiency into the two parts. Instead we have to give up some of the microeconomic details and in part rely on macroeconomics.

This move is less drastic that it appears. In fact we already recalled in section 2.1.1 the assumption of perfect competition, on which the formulation of the standard profit function is based. Yet we cannot expect that all of the conditions will be fulfilled when we perform cross-sectional analysis. 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.

One way is to devise ever more complicated models encompassing greater variety of (micro)economic factors. This approach is valuable and desirable in theory, but not quite so for empirical analysis. Practice requires compromise and as a consequence there is the possibility to make a step towards macroeconomics. We go for this solution. We showed the microeconomic underpinnings of estimating technical efficiency in production functions. Then we can plug in our data, so that our estimate which we denoted $\widehat{TE}_i$ will in fact include both components and hence will measure overall efficiency. Admittedly this is exactly the point where micro and macro approaches get combined. In other words, we estimate the profit function,

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

although our starting point (3.2) is not a proper profit function, but a simple production function. Yet our claim is that the compromise is worth it and we use this method as one possibility to assess efficiency levels of industrial sectors.

## 3.2 Non-Parametric Approach

### 3.2.1 Basic Model Structure

This method is heading to (2.1). In the first stage it constructs a plane around the dataset, with points lying on the plane being technically efficient and points within the space being inefficient. Hence the first model constructed this way was called data envelopment analysis or DEA. The envelope can have different shapes according to the exact mathematical specification; plenty of variants has been developed since the formulation of DEA.[6] The second stage adds prices to the data and identifies economic efficiency. Decomposition of these two stages is shown in figure 3.1. If

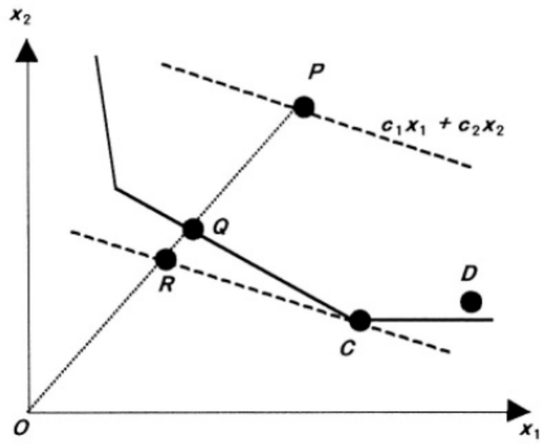



Figure 3.1: Technical and allocation efficiency. (Cooper *et al.*, 2002, p. 223)

we employ the radial measure of efficiency, then technical efficiency is represented by the ratio $\frac{|OQ|}{|OP|}$ and allocative efficiency by $\frac{|OR|}{|OQ|}$.[7] Thus overall efficiency equals to $\frac{|OQ|}{|OP|} \times \frac{|OR|}{|OQ|} = \frac{|OR|}{|OP|}$.

The presence of $\boldsymbol{\beta}$ in (2.1) is not contradicting the section heading 'non-parametric approach'. The concept of being 'non-parametric' only suggests that DEA does not deal with $\boldsymbol{\beta}$ directly. Instead of estimating $\boldsymbol{\beta}$ and then looking for concord of DMUs

[6]A very condensed overview of DEA models can be found in (Daraio & Simar, 2007a).

[7]One can also use a general measure of distance $d(\cdot)$: $\frac{d(O,Q)}{d(O,P)}$.

with the estimate $\hat{\boldsymbol{\beta}}$, DEA puts out a direct estimate of efficiency and considers $\boldsymbol{\beta}$ as the governing parameter behind the scenes.

One introductory textbook on DEA is by (Cooper *et al.*, 2002), where we also find the simplest classification of DEA models, roughly corresponding to profit maximization and cost minimization:

⋆ Output-oriented DEA: While using no more than the observed amount of any input, the largest possible increase in outputs is computed.

⋆ Input-oriented DEA: While producing at least the given output levels, the largest possible decrease in inputs is computed.

Combinations of both are also possible. For analytical purpose there is no difference in the type applied, since efficiency rank will remain unaffected. Managerial decisions connected with adjusting either inputs or outputs will of course differ significantly. Given the availability of software described later, we decided for the input-oriented models.

Mathematical formulation of DEA has been developed from input-output ratios and evolved to linear programming problems.[8] For technical efficiency we can write a simple input-oriented DEA problem in matrix notation as follows:

$$\begin{aligned} \min_{\boldsymbol{\lambda},\theta} \quad & \theta \\ \text{subject to } \theta \boldsymbol{x}_i \quad & \geq \quad X\boldsymbol{\lambda} \\ Y\boldsymbol{\lambda} \quad & \geq \quad \boldsymbol{y}_i \\ \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n) \quad & \geq \quad 0.^9 \end{aligned} \tag{3.7}$$

It is known as the CCR-I model, since it was formulated by Charnes, Cooper and Rhodes. Note that in (3.7) the vectors and matrices are ordered so as to be economically suggestive; standard linear programming would order them differently

[8] As a reference book for linear programming, we used that by Vanderbei (2001) available online.

[9] The extended notation of matrices reads:

$$\theta \cdot \begin{pmatrix} x_{1i} \\ \vdots \\ x_{pi} \end{pmatrix} \geq \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{p1} & \dots & x_{pn} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix},$$

$$\begin{pmatrix} y_{1i} \\ \vdots \\ y_{ri} \end{pmatrix} \leq \begin{pmatrix} y_{11} & \dots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{r1} & \dots & y_{rn} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

Linear programming problems can also be expressed in its dual (multiplier) form which we do not investigate here.

to fit into 'dictionary' notation which is used to solve the problem. Vector $\boldsymbol{\lambda}$ attaches weights to single producers: In the third line, $\boldsymbol{\lambda}$ selects firms, which are called 'reference' producers of the evaluated firm. They, weighed together by $\boldsymbol{\lambda}$, produce at least as many outputs as the evaluated DMU$_i$. By this $\boldsymbol{\lambda}$ we scale $X$ and see whether it is possible to cut down inputs at DMU$_i$ by a coefficient $\theta$.

This problem must be solved $n$ times for all producers to obtain each firm's technical efficiency score, which is an estimate $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} \in [0,1]$. This result measures the distance of the evaluated producer from the efficient frontier.

In terms of section 2.2.1, $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$ is the Farrell input-oriented measure of technical efficiency, defined as (Cazals *et al.*, 2002):

$$\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} = \inf\left\{\theta | \theta \boldsymbol{x}_i \in L(\boldsymbol{y}_i)\right\} = \inf\left\{\theta | \left(\theta \boldsymbol{x}_i, \boldsymbol{y}_i\right) \in \Psi\right\}.$$

Sometimes this measure is also called Debreu-Farrell, eg in Lovell (1992, c.f. reading 15, p. 13). Lovell also reminds of the Shepard input distance function which is an alternative measure defined as:

$$D_{I,(\boldsymbol{x}_i,\boldsymbol{y}_i)} = \sup\left\{\theta \left| \left(\frac{\boldsymbol{x}_i}{\theta}\right) \in L(\boldsymbol{y}_i)\right.\right\} = \sup\left\{\theta \left| \left(\frac{\boldsymbol{x}_i}{\theta}, \boldsymbol{y}_i\right) \in \Psi\right.\right\},$$

so that it is the inverse of $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$. Before we turn to slacks and prices, we introduce constraints defining returns to scale.

### 3.2.2 Returns to Scale

Model (3.7) does not impose any addtional conditions on $\boldsymbol{\lambda}$, so that technical efficiency is computed under the assumption of constant returns to scale. Variable returns to scale (RTS) were introduced in the BCC model by Banker, Charnes and Cooper who added the convexity constraint $\sum_{i=1}^{n} \lambda_i = 1$ to the CCR model. A summary of DEA variations with respect to RTS is presented in the following table:

$$\begin{aligned}
\boldsymbol{e} = (1,\ldots,1)^T \quad & \ldots \quad \text{vector of ones,} \\
\boldsymbol{\lambda} \text{ free} \quad & \ldots \quad \text{constant returns to scale,} \\
\boldsymbol{e}^T\boldsymbol{\lambda} = \sum_{i=1}^{n} \lambda_i = 1 \quad & \ldots \quad \text{variable returns to scale,} \\
\boldsymbol{e}^T\boldsymbol{\lambda} = \sum_{i=1}^{n} \lambda_i \leq 1 \quad & \ldots \quad \text{non-increasing returns to scale,} \\
\boldsymbol{e}^T\boldsymbol{\lambda} = \sum_{i=1}^{n} \lambda_i \geq 1 \quad & \ldots \quad \text{non-decreasing returns to scale,} \\
\left(\sum_{i=1}^{n} \lambda_i = 1\right) \wedge \left(\forall i : \lambda_i \in \{0,1\}\right) \quad & \ldots \quad \text{free disposal hull (FDH).}
\end{aligned}$$

These modifications can be applied in two ways. If we believe that we have sufficient information on the type of RTS in the data, we can specify it generally for all firms prior to the computation. On the other hand, we can compute efficiency scores under VRTS and then determine RTS for each producer individually. A sequential approach to perform this is thoroughly explained in (Cooper *et al.*, 2002, chapter 5), along with the respective mathematical theorems.

Free disposal hull is not connected to RTS and it differs from both CCR and BCC models in that it draws an envelope that is not convex, as is depicted in figure 3.2. $\lambda_i$ is an integer variable, so that it allows for either one or none reference producer. Apparently FDH is the closest envelope to the data. The FDH efficient frontier is not

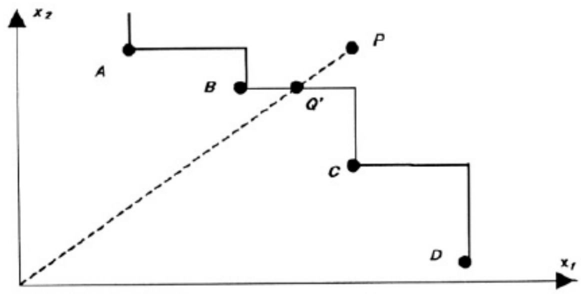



Figure 3.2: Input-oriented FDH. (Cooper *et al.*, 2002, p. 105)

continuous, which might be attractive in some applications where there are reasons to assume discrete inputs or outputs. Obviously this requires detailed information on the character of examined firms. However we will deal with aggregated data, so that we will not be able to make use of this specification.

### 3.2.3 Slacks and the Additive Model

Frontiers as in figure 3.1 are by natural sense called piecewise linear. Such envelopes satisfy $\boldsymbol{x} \in \mathrm{Isoq}L(\boldsymbol{y})$, but not necessarily $\boldsymbol{x} \in \mathrm{Eff}L(\boldsymbol{y})$ (see section 2.2.1). Consider the company $P$ in figure 3.2: So far we would have measured its efficiency as compared to $Q'$; yet it is still possible to decrease input $x_1$ without impacting the amount of output as represented by point $B$. In other words there is an 'input slack' in $x_1$.

Slacks in input $j \in \{1, \ldots, p\}$ and in output $k \in \{1, \ldots, r\}$ for the input-oriented model in (3.7) can be residually computed as (Daraio & Simar, 2007a):

$$
\begin{aligned}
s_j^- &= x_{ji} \cdot \theta^*_{(\boldsymbol{x}_i, \boldsymbol{y}_i)} - \sum_{i=1}^{n} x_{ji}\lambda_i, \\
s_k^+ &= \sum_{i=1}^{n} y_{ki}\lambda_i - y_{ki},
\end{aligned}
$$

which in matrix notation yields:

$$\begin{aligned} \boldsymbol{s}^- &= \theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}\boldsymbol{x}_i - X\boldsymbol{\lambda}, \\ \boldsymbol{s}^+ &= Y\boldsymbol{\lambda} - \boldsymbol{y}_i, \\ \boldsymbol{s}^-,\boldsymbol{s}^+ &\geq 0. \end{aligned}$$

The evaluated producer is technically efficient iff $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} = 1$ and $\boldsymbol{s}^- = \boldsymbol{s}^+ = 0$.

Because this specification is not included in the solving procedure of (3.7), such residual calculation may not disclose full inefficiency. It is possible to account for maximum slacks in (3.7) by adding a second phase objective, which is solved separately.[10]

The alternative is to solve the additive problem (Cooper *et al.*, 2002):

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{s}^-,\boldsymbol{s}^+} \quad z &= \boldsymbol{e}^T\boldsymbol{s}^- + \boldsymbol{e}^T\boldsymbol{s}^+ \\ \text{subject to} \quad X\boldsymbol{\lambda} + \boldsymbol{s}^- &= \boldsymbol{x}_i \\ Y\boldsymbol{\lambda} - \boldsymbol{s}^+ &= \boldsymbol{y}_i \\ \boldsymbol{e}^T\boldsymbol{\lambda} &= 1 \\ \boldsymbol{\lambda},\boldsymbol{s}^-,\boldsymbol{s}^+ &\geq 0, \end{aligned} \tag{3.9}$$

where we already added the convexity constraint accounting for variable RTS. Note that the slack $\boldsymbol{s}^-$ in (3.9) is modified vis-a-vis the previous definition. Comparison between the models (3.7) and (3.9) gives an instantaneous interpretation of the latter: In order to directly measure the magnitude of inequalities in (3.7), we plug in two variables $\boldsymbol{s}$ for each of the inequalities. These variables replace the sole score coefficient $\theta$.

As a consequence, we directly detect maximum slacks in all dimensions, which have to be converted to a single score. We also note that the slack measure is distant from the radial measure described in section 3.2.1: it is more comprehensive at once.

[10] Knowing $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$ we solve the second phase:

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{s}^-,\boldsymbol{s}^+} \quad z &= \boldsymbol{e}^T\boldsymbol{s}^- + \boldsymbol{e}^T\boldsymbol{s}^+ \\ \text{subject to} \quad \boldsymbol{s}^- &= \theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} \cdot \boldsymbol{x}_i - X\boldsymbol{\lambda} \\ \boldsymbol{s}^+ &= Y\boldsymbol{\lambda} - \boldsymbol{y}_i \\ \boldsymbol{\lambda},\boldsymbol{s}^-,\boldsymbol{s}^+ &\geq 0. \end{aligned} \tag{3.8}$$

Note that unlike in the residual definition, the second phase allows $\boldsymbol{\lambda}$ to vary again. Thus given the distance from the frontier we obtain maximum slacks as compared to the new set of DMUs selected by $\boldsymbol{\lambda}$ (which may or may not be different from the first stage solution), and so an estimate of total technical inefficiency.

### 3.2.4 Prices and Units Invariance

For DEA it holds the very same idea which we already outlined in section 3.1.2 for the parametric approach. Namely: So far our DEA considerations covered production in terms of physical units. With very detailed data such productivity analysis will attract all engineers who deal directly with designing the technological production process. However, what matters to us is economic efficiency, ie profitability.

In order for the DEA models to accommodate an economic meaning, we want to include prices of inputs $\boldsymbol{w}$ and outputs $\boldsymbol{p}$, assuming them to be exogenous. The following theorem (Cooper *et al.*, 2002, p. 24) steers towards one solution.

**Theorem 3.1** ***Units Invariance Theorem.*** *The optimal values of* $\max\theta = \theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$ *in (3.7) and of* $\max z = z^*$ *in (3.9) are independent of the units in which the inputs and outputs are measured, provided these units are the same for every DMU.*

We can take advantage of this fact and plug in the data expressed in units of money. This setup changes our models into measuring the distance of single DMUs to the empirical unit value isoquant, or rather 'unit value envelope'. It is a significant simplification, but for the reasoning we refer to what we already said in section 3.1.2.

It makes things both better and worse. Money as a universal unit of measurement offers direct comparisons, while it is troublesome to compare lots of inputs recorded in different units.[11] On the other hand, we forfeit the possibility to decompose inefficiency.

Apparently with higher level of economy under scrutiny, we have to forgo data in physical units, because it hampers both data collection and analysis. The use of detailed inputs and outputs will be tractable only if we were to inspect individual producers in one sectorof the economy. Otherwise we have to settle for a simple capital–labour framework.

### 3.2.5 Allocation Efficiency Models

Incorporating prices in DEA is done 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 amout of outputs produced.

Let us illustrate this with two examples. One simple modification of DEA linear

[11] How do pieces of heavy machinery compare to, say, computers used in a bank?

programs reads (Coelli, 1996a):

$$
\begin{aligned}
\min_{\boldsymbol{\lambda},\boldsymbol{x}_i^*} \quad & \boldsymbol{w}^T\boldsymbol{x}_i^* && (3.10)\\
\text{subject to } \boldsymbol{x}_i^* &\geq X\boldsymbol{\lambda}\\
Y\boldsymbol{\lambda} &\geq \boldsymbol{y}_i\\
\boldsymbol{e}^T\boldsymbol{\lambda} &= 1\\
\boldsymbol{\lambda} &\geq 0.
\end{aligned}
$$

This model evaluates the minimal cost attainable at DMU$_i$: The respective output $\boldsymbol{y}_i$ is sifted through the best practice grid set by $\boldsymbol{\lambda}$ to find out the theoretical most favourable amount of inputs $\boldsymbol{x}_i^*$.

It is important to bear in mind one crucial feature of (3.10): While looking for the optimal input vector, proportions of its searched-for components $x_{ji}^*$ are allowed to vary. Hence this procedure tracks slacks as well and yields overall cost efficiency as a ratio

$$\theta^{*,CE}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} = \frac{\boldsymbol{w}_i^T\boldsymbol{x}_i^*}{\boldsymbol{w}_i^T\boldsymbol{x}_i}.$$

This score can be decomposed residually if we evaluate technical efficiency separately by (3.7) and use the fact that $CE = TE \times AE$. Slacks will be assigned to allocative efficiency.

We can extend this model to revenue and profit as well (Cooper *et al.*, 2004, chapter 1). To track overall efficiency, we can use the following additive formulation:

$$
\begin{aligned}
\max_{\boldsymbol{\lambda},\boldsymbol{s}^+,\boldsymbol{s}^-} \pi &= \boldsymbol{p}^T\boldsymbol{s}^+ + \boldsymbol{w}^T\boldsymbol{s}^- && (3.11)\\
\text{subject to } X\boldsymbol{\lambda} - \boldsymbol{s}^- &= \boldsymbol{x}_i\\
Y\boldsymbol{\lambda} - \boldsymbol{s}^+ &= \boldsymbol{y}_i\\
\boldsymbol{e}^T\boldsymbol{\lambda} &= 1\\
\boldsymbol{\lambda} &\geq 0\\
\boldsymbol{s}^-, \boldsymbol{s}^+ &\ \ldots \ \text{free}
\end{aligned}
$$

where $\pi$ stands for total profit foregone. Note that the slack variables are now unbounded. Also in this case solutions are not units invariant. For decomposition it is again necessary to solve the additive problem (3.9) for technical efficiency.

## 3.3 Statistical Methods in Non-Parametric Approach

Daraio & Simar (2007a, p. xviii) summarize four major shortcomings of envelopment analysis:

- ⋆ deterministic and non-statistical nature;
- ⋆ influence of outliers and extreme values;
- ⋆ lack of parameters for the economic interpretation;
- ⋆ unsatisfactory techniques for the introduction of environmental or external variables in the measurement (estimation) of the efficiency.

They synthesize recent efforts to surmount these obstacles. We select one possibility to improve estimators (described in section 3.2), which deal with the first two drawbacks and which we implement in our data analysis. Prior to that, we have to formalize the underlying probabilistic production process.

### 3.3.1 Probabilistic Production Process

In section 3.2 we presented fully deterministic measures of efficiency. These assume that $\Pr((\boldsymbol{x}_i, \boldsymbol{y}_i) \in \Psi) = 1$. The parametric approach outlined in section 3.1 also assumes that observations are deterministic, nevertheless it contains random noise outside the production function. Technical inefficiency $\tau$ is itself considered an *iid* random variable.

This time, inputs and outputs are a pair of *iid* multidimensional random variables $(\mathcal{X}, \mathcal{Y})$, although for individual observation it still holds $\Pr((\boldsymbol{x}_i, \boldsymbol{y}_i) \in \Psi) = 1$. Following the derivation of Daraio & Simar (2007b), this yields a joint probability measure characterized by the function

$$H_{\mathcal{X}\mathcal{Y}}(\boldsymbol{x}, \boldsymbol{y}) = \Pr(\mathcal{X} \leq \boldsymbol{x}, \mathcal{Y} \geq \boldsymbol{y}).$$ [12]

For DMU $[\boldsymbol{x}, \boldsymbol{y}]$ it captures the probability that this firm will perform worse than others, ie that it will use more inputs and produce less output. Further we want to

[12] Daraio & Simar (2007b) note that the support of $H_{\mathcal{X}\mathcal{Y}}(\cdot, \cdot)$ is $\Psi$, where support of a probability density $g_{\theta \in \Theta}(\boldsymbol{x})$, $\theta$ being the parameter from the parameter space $\Theta$, is defined as:

$$S = \{\boldsymbol{x} \in \Re^p : g_{\theta \in \Theta}(\boldsymbol{x}) > 0\}.$$

know the probability that once the firm produces less, it also uses more inputs. Thus we consider 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 $S_{\mathcal{Y}}(\boldsymbol{y}) > 0$. This can be empirically estimated by computing

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

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

### 3.3.2 Order-$m$ Estimator

This estimator was introduced by Cazals *et al.* (2002). The idea is simple: Suppose we have an observation $[\boldsymbol{x}_0, \boldsymbol{y}_0]$. As in the CCR model (3.7), we take observations with larger output, this time without scaling. From this set of observations we draw randomly with replacement $X_1, \ldots, X_m$, which is distributed according to $F_{\mathcal{X}|\mathcal{Y}}(\cdot|\boldsymbol{y})$, as follows from the previous section. We construct the production possibility set (Daraio & Simar, 2007b, c. f.)

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

Then we measure the efficiency of our firm against this subset as the expected minimum efficiency score. We first compute

$$\tilde{\theta}^m_{(\boldsymbol{x}_0,\boldsymbol{y}_0)} = \inf \left\{ \theta | (\theta \boldsymbol{x}_0, \boldsymbol{y}_0) \in \tilde{\Psi}_m(\boldsymbol{y}_0) \right\} \tag{3.12}$$

and take expectations

$$\theta^m_{(\boldsymbol{x}_0,\boldsymbol{y}_0)} = E_{\mathcal{X}|\mathcal{Y}}(\tilde{\theta}^m_{(\boldsymbol{x}_0,\boldsymbol{y}_0)} | \mathcal{Y} \geq \boldsymbol{y}). \tag{3.13}$$

In other words, we compare our DMU to random subsets of larger producers (ie with higher output) and look at the efficiency score we can statistically expect in such setting.

Using the empirical distribution function $\hat{F}_{\mathcal{X}|\mathcal{Y}}$, the score can be estimated as:

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

Unfortunately this integration can not be carried out analytically. Cazals *et al.* (2002) proposed a four step Monte-Carlo algorithm, which we quote as in Daraio & Simar (2007a):

[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 $\tilde{\theta}^{m,b}_{(\boldsymbol{x}_0,\boldsymbol{y}_0)} = \min_{i=1,\ldots,m} \left\{ \max_{j=1,\ldots,p} \left( \frac{X^j_{i,b}}{x^j} \right) \right\}$.[13]

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

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

### 3.3.3 Convex order-$m$ frontier

Most of section 3.2 deals with efficiency estimates based on convex technology. The only exeption is FDH, briefly mentioned in 3.2.2. Since the order-$m$ frontier is based on FDH, it is not convex. FDH is derived from the approximation of production technology (Daraio & Simar, 2007b):

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

Daraio & Simar recall that usual DEA scores can be easily obtained from FDH results: It suffices to multiply observed inputs $\boldsymbol{x}$ by $\hat{\theta}^{FDH}_{(\boldsymbol{x},\boldsymbol{y})}$ and then run the respective linear program on the transformed data, which can be for example equation (3.7).

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

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

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

$$\begin{aligned} \hat{\theta}^{m,C}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} &= \min_{\boldsymbol{\lambda},\theta} \theta \\ \text{subject to } \theta \boldsymbol{x}_i &\geq \sum_{i=1}^{n} \lambda_i \hat{\boldsymbol{x}}^{\partial}_{m,i} \\ Y\boldsymbol{\lambda} &\geq \boldsymbol{y}_i \\ \sum_{i=1}^{n} \lambda_i &= 1 \\ \lambda_1, \ldots, \lambda_n &\geq 0. \end{aligned} \tag{3.14}$$

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

[13] This min-max algorithm is the easy way to obtain the standard FDH efficiency score.

# Chapter 4

# Efficiency of Czech SME

## 4.1 Data Description

Czech Statistical Office publishes a yearly summary on economic activity of Czech small and medium enterprises, which can be found under reference number 8007-0[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: For the sake of data security, the law requires some information contained in the original data to be destroyed.

Following an official request, the Czech Statistical Office provided us with slightly more detailed data than one can find in the publicly available booklet. The datased has four dimensions:

1. thirty-item two-digit OKEC classification, see appendix A;
2. size classification with breakdowns at the following number of employees: 0-10-20-50-100-250;
3. eleven economic indicators listed in table 4.1, for complete definitions of indicators refer to appendix B;
4. years 2002 through 2005.

The data imply main characteristics of the analysis. Items under point 3 are fitted to the standard economic labour-capital-output framework, which is indicated in table 4.1. 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

**Indicator and the corresponding variable in the model**

| | | |
|---|---|---|
| Number of active firms | | |
| $OUT$ | Output | Output |
| $REV$ | Sales revenue | Output |
| $VAD$ | Accounting value added | Output |
| $TAS$ | 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 4.1: Indicators on SME provided by the Czech Statistical Office.

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

The usage of the economic indicators deserves several comments. Items of table 4.1 can be regarded as aggregated accounting figures. Sales revenue tracks all goods and services that the company was able to vend on the market. Output adds goods that were already produced but not yet sold to the sales revenue. Finally, when cost of materials is subtracted, we get the accounting value added. This should approximately express how much a firm is able to produce from its stock of capital and labour, since the cost of these is not included in the sum of materials. Further, the average number of employees is more preferable to $EMP$. $EMP$ captures the sum of employees at one particular day, which are then recalculated on the basis of days worked to get $AEM$. It follows that $AEM$ captures all the fluctuation of employees, which is exactly what we need.

These considerations turn our initial estimation idea to:

$$\begin{aligned} y_i &= f(\boldsymbol{x_i}; \boldsymbol{\beta}) \\ VAD_i &= f(TAS_i, IAS_i, INV_i, AEM_i, PAY_i, OPE_i; \boldsymbol{\beta}), \end{aligned} \tag{4.1}$$

considering either the parametric or the non-parametric approach.

It remains to note that panel research is limited by short time span—only four years, which are moreover consecutive. Primarily we will assume that we can neglect the differences in installed technology over these four years, ie that technology did not change in time. We will then test for systematic technological progress in regression analysis.

## 4.2 DEA Results

### 4.2.1 Envelopes I: Standard DEA Results

Units invariance theorem (3.1) is an extremely powerful property of DEA. Practically it means that we do not have to care about any prior sort-out of the data, since vector $\boldsymbol{\lambda}$ will sort the data itself. We could indeed plug in the data matrix described in section 4.1 right as it is.

As it goes, there is a snag: One major drawback of DEA, its sensitivity to outliers, becomes more pronounced with more variables. Unlike in regressions, where we want to asses separate effects of lots of factors, here we have to rationalize the number of variables. This task is not too onerous, because we can add up related items.

Consider the BCC model, ie equation (3.7) with the additional constraint $\boldsymbol{e}^T\boldsymbol{\lambda} = \sum_{i=1}^{n} \lambda_i = 1$ introducing variable returns to scale. As the vector of inputs $\boldsymbol{x}$ we take [assets, investment, employees and wages], where 'assets' are totalled tangible and intangible assets and 'wages' are wage outlays plus other personal expenses; output will be represented by value added.

We implemented this computation for each year separately via DEAP, a freely available programme by T. Coelli (Coelli, 1996a). We also run the second stage from equation (3.8) in order to obtain slacks and to see where does the inefficiency come from.

Industries with $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} = 1$ which are thus lying on the efficient frontier are indicated in table 4.2, the subscript denotes the upper bound of the respective size subgroup. These points also mostly act as "peers": points at the edges of the envelope, against which slacks are computed.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2002 | $10_9$ | $11_{250}$ | $13_9$ | $14_9$ | $20_9$ | $28_9$ | $30_{19}$ | $31_9$ | $32_9$ | $33_9$ |
| | $35_9$ | $36_9$ | $40_9$ | | | | | | | |
| 2003 | $11_9$ | $13_9$ | $16_9$ | $16_{99}$ | $18_9$ | $18_{250}$ | $23_{19}$ | $28_9$ | $31_9$ | $35_{19}$ |
| | $40_9$ | | | | | | | | | |
| 2004 | $11_9$ | $11_{99}$ | $13_9$ | $16_9$ | $23_{99}$ | $28_9$ | $31_9$ | | | |
| 2005 | $10_{99}$ | $16_9$ | $16_{49}$ | $23_{99}$ | $28_9$ | $29_{250}$ | $31_9$ | $40_{49}$ | | |

| | | | | |
|---|---|---|---|---|
| $\bigcap$ (2002-05) | $28_9$ | $31_9$ | | |
| 3 out of 4 years | $13_9$ | $16_9$ | $28_9$ | $31_9$ |

Table 4.2: Efficient industries according to the CCR model.

The most striking finding of the first result is: Out of the 39 fully efficient industries, exactly two thirds belong to the smallest grouping with nine or less employees. This prompts us to ask whether there is a significant size effect, and in this case tending to the unexpected side.

We can identify four segments which were effective in at least three years out of the four, all of them being the smallest entrepreneurs. Somewhat surprisingly, mining of metal ores (13) qualified for this prime sample. So did manufacture of tobacco products (16), of metal products except machinery (28) and of electrical machinery (31). The latter two sectors can be viewed as more technically advanced, yet it still contradicts the intuitive expectation which one would probably have here: that of larger machinery manufacturers, automotive suppliers and alike.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2002 | $35_{250}$ | $10_{250}$ | $30_{250}$ | $23_{9}$ | $34_{9}$ | $16_{9}$ | $20_{19}$ | $27_{250}$ | $21_{19}$ | $17_{49}$ |
| 2003 | $23_{49}$ | $10_{250}$ | $10_{19}$ | $30_{49}$ | $23_{9}$ | $11_{99}$ | $14_{19}$ | $35_{250}$ | $10_{9}$ | $41_{19}$ |
| 2004 | $27_{19}$ | $35_{250}$ | $17_{49}$ | $34_{99}$ | $36_{19}$ | $17_{19}$ | $41_{99}$ | $35_{99}$ | $32_{250}$ | $25_{19}$ |
| 2005 | $27_{19}$ | $23_{9}$ | $35_{19}$ | $22_{19}$ | $35_{250}$ | $19_{250}$ | $30_{250}$ | $32_{49}$ | $35_{99}$ | $17_{19}$ |

| | | | | | | |
|---|---|---|---|---|---|---|
| $\bigcap$ (2002-05) | $35_{250}$ | | | | | |
| 2 out of 4 years | $10_{250}$ | $17_{19}$ | $17_{49}$ | $23_{9}$ | $27_{19}$ | $35_{99}$ |

Table 4.3: The least efficient industries according to the CCR model.

Similarly in table 4.3 we summarize the ten least efficient industries according to our first DEA model. Industries are listed in descending order of their respective efficiency score. Here size differences are not so clear. Moreover we do not observe much overlapping: If a certain industry was top efficient (say $13_9$), we do not find the same industry of a different size group among the worst performers (ie there is no $13_x$ listed in the "worst" table). This is a welcome result, since we expect members of the same industry to be rather close to each other. Still exceptions are present, eg in the only year when $16_9$ was not fully efficient—namely 2002, it ended up worst.

This time we take industries which ranked least efficient in at least two years out of the four: mining of coal and lignite (10), manufacture of textiles (17), which has been on decline ever since the velvet revolution, manufacture of coke, refined petroleum products and nuclear fuel (23), manufacture of basic metals (27) and of other transport equipment (35). On the one hand, processing of raw materials (coal, coke, basic metals) in principle does not create lots of value added. On the other hand, the data are a little bit old to capture the steep rise in commodity prices.

To get an overview of the distribution of efficiency, we computed box plot statistics given in table 4.4, 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.

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

The result is plotted in figure 4.1, where outliers are shown as circles. For all years the mean of scores is higher then the median, meaning that the estimated efficiency distribution is skewed to lower scores. Especially for 2004 and 2005, the graph exemplifies noticeable skewness. Average efficiency amounts to 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 4.5). 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 4.1** ***Preliminary results.*** *The CCR 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 derive a list of 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 test the presence of a size effect, ie the hypothesis that higher efficiency is connected with certain size group.

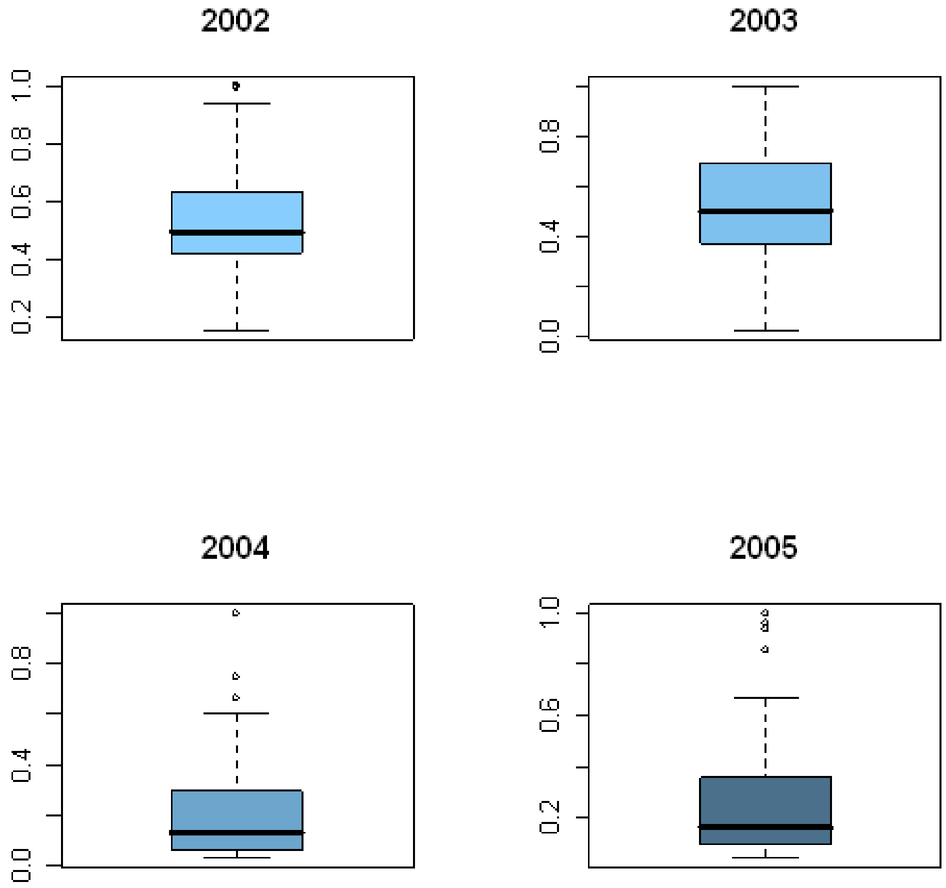



Figure 4.1: Box plots of efficiency scores $\theta^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$.

Thirdly, we will analyze the sectoral structure in more depth. For this purpose we will modify the proceeding used in tables 4.2 and 4.3. 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.

### 4.2.2 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 (2007), where both the Monte-Carlo simulation from section 3.3.2 and the solution of equation (3.14) are available.

First we had to specify the computational aspects: parameters $m$ and $B$. As

| | ♯ of employees | min | 1$Q$ | median | 3$Q$ | max | mean |
|---|---|---|---|---|---|---|---|
| | <10 | 0.200 | 0.593 | 0.810 | 1.000 | 1.000 | 0.754 |
| | 10-19 | 0.282 | 0.404 | 0.442 | 0.530 | 0.718 | 0.482 |
| 2002 | 20-49 | 0.299 | 0.373 | 0.472 | 0.537 | 0.784 | 0.485 |
| | 50-99 | 0.318 | 0.412 | 0.448 | 0.517 | 0.675 | 0.485 |
| | 100-250 | 0.150 | 0.428 | 0.490 | 0.686 | 1.000 | 0.541 |
| | <10 | 0.112 | 0.376 | 0.653 | 0.832 | 1.000 | 0.629 |
| | 10-19 | 0.044 | 0.357 | 0.470 | 0.591 | 0.941 | 0.496 |
| 2003 | 20-49 | 0.020 | 0.391 | 0.488 | 0.606 | 0.908 | 0.486 |
| | 50-99 | 0.124 | 0.345 | 0.428 | 0.600 | 0.983 | 0.478 |
| | 100-250 | 0.037 | 0.395 | 0.494 | 0.710 | 1.000 | 0.540 |
| | <10 | 0.049 | 0.095 | 0.146 | 0.604 | 1.000 | 0.358 |
| | 10-19 | 0.031 | 0.060 | 0.076 | 0.129 | 0.233 | 0.115 |
| 2004 | 20-49 | 0.039 | 0.069 | 0.140 | 0.190 | 0.372 | 0.169 |
| | 50-99 | 0.040 | 0.056 | 0.108 | 0.272 | 0.597 | 0.209 |
| | 100-250 | 0.032 | 0.098 | 0.269 | 0.356 | 0.743 | 0.268 |
| | <10 | 0.049 | 0.132 | 0.278 | 0.532 | 1.000 | 0.390 |
| | 10-19 | 0.042 | 0.082 | 0.103 | 0.162 | 0.281 | 0.142 |
| 2005 | 20-49 | 0.069 | 0.099 | 0.165 | 0.321 | 0.654 | 0.253 |
| | 50-99 | 0.070 | 0.109 | 0.184 | 0.348 | 0.705 | 0.264 |
| | 100-250 | 0.063 | 0.119 | 0.268 | 0.406 | 0.838 | 0.311 |

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

$m \to \infty$, we have the convergence $\hat{\theta}^m_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} \to \hat{\theta}^{FDH}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$ (Cazals *et al.*, 2002, theorem 2.3), and similarly $\hat{\theta}^{m,C}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)} \to \hat{\theta}^*_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$ so that with higher $m$ less observations will lie above the efficient frontier and the estimator gets less robust. Based on trial and error, we chose $m = 50$ as the level of robustness. With lower numbers of reference observations (eg $m = 20$), there was unusualy high ratio of superefficient 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.

Figure 4.2 shows boxplots for yearly efficiency scores. Distribution of individual efficiency estimates appears more favourable than in figure 4.1. 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. Probabilistic approach suppressed superefficient outliers

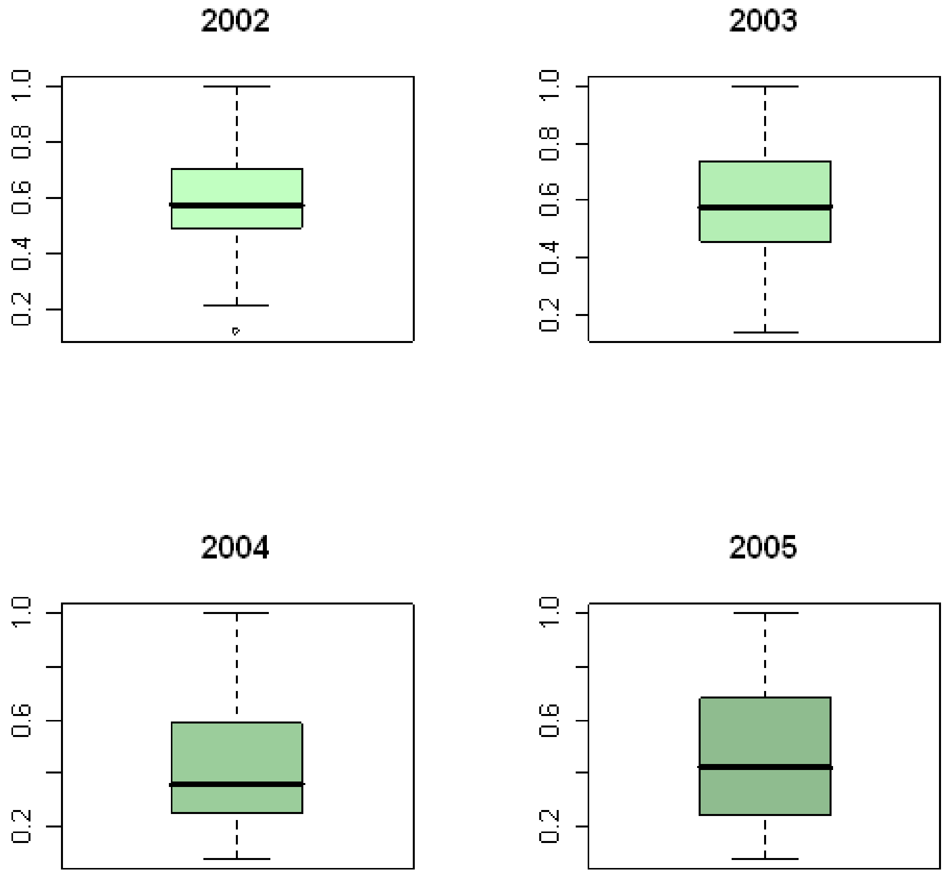



Figure 4.2: Box plots of efficiency scores $\hat{\theta}^{m,C}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$.

and the obtained estimates represent 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.

Recalling Aigner & Chu (1968) and his criticism of average production functions quoted in section 3.1 once more, it could seem that we only moved to certain "average" production plan. Yet histograms which we do not reproduce here disclose that the results do not by far resemble normal distribution, because there are two peaks.[1] Moreover the estimates are still skewed to the left, so that while having used the flexible measure, apparently we did not lose large part of information contained in the data.

[1]Although seemingly superfluous, we run statistical tests for studentized COM scores, which strongly rejected the null of normality.

| | ♯ of employees | min | 1$Q$ | median | 3$Q$ | 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: Box plot statistics for efficiency scores $\hat{\theta}^{m,C}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$.

Table 4.6 tracks in more detail the distribution of efficiency scores. When confronted with the intitial results in table 4.5, we conclude that any direct relation between efficiency and size formulated in proposition 4.1 is weakened by the COM model. If we trust COM in that it supressed the influence of outliers, we may conclude that strong mean efficiency of the smallest enterprises (as reported in table 4.5) was a result given by the presence of favourable extreme observations. We get back to this issue later in section 4.4.3.

Let us repeat what we achieved by COM: Due to 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 4.2.1.

In table 4.7, 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

| **Best industries** | | | | | **Worst industries** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2002 | 2003 | 2004 | 2005 | 3 out of 4 years | 2002 | 2003 | 2004 | 2005 | 3 out of 4 years |
| $11_{250}$ | $11_{9}$ | $11_{9}$ | $10_{99}$ | $13_{9}$ | $10_{49}$ | $10_{19}$ | $10_{9}$ | $10_{9}$ | $10_{250}$ |
| $13_{9}$ | $13_{9}$ | $11_{99}$ | $14_{250}$ | $15_{250}$ | $10_{99}$ | $10_{99}$ | $10_{250}$ | $10_{19}$ | $11_{99}$ |
| $14_{250}$ | $14_{19}$ | $13_{9}$ | $15_{49}$ | $16_{9}$ | $10_{250}$ | $10_{250}$ | $14_{9}$ | $10_{49}$ | $19_{9}$ |
| $15_{99}$ | $15_{250}$ | $15_{250}$ | $15_{99}$ | $20_{9}$ | $11_{9}$ | $11_{99}$ | $15_{19}$ | $11_{9}$ | $21_{9}$ |
| $15_{250}$ | $16_{9}$ | $16_{9}$ | $15_{250}$ | $28_{9}$ | $11_{99}$ | $11_{250}$ | $17_{99}$ | $11_{99}$ | $23_{9}$ |
| $16_{9}$ | $16_{99}$ | $18_{49}$ | $16_{9}$ | $28_{250}$ | $14_{19}$ | $19_{9}$ | $19_{9}$ | $11_{250}$ | $27_{9}$ |
| $18_{9}$ | $18_{9}$ | $18_{250}$ | $16_{49}$ | $29_{250}$ | $16_{49}$ | $21_{9}$ | $20_{19}$ | $13_{9}$ | $30_{250}$ |
| $19_{9}$ | $18_{19}$ | $20_{9}$ | $17_{49}$ | $31_{9}$ | $16_{250}$ | $22_{19}$ | $20_{99}$ | $16_{99}$ | $34_{9}$ |
| $19_{19}$ | $18_{49}$ | $22_{9}$ | $20_{9}$ | $32_{250}$ | $21_{9}$ | $23_{9}$ | $21_{9}$ | $19_{9}$ | $34_{19}$ |
| $20_{9}$ | $18_{250}$ | $23_{99}$ | $21_{99}$ | $40_{250}$ | $21_{19}$ | $24_{19}$ | $22_{19}$ | $19_{19}$ | $37_{19}$ |
| $22_{9}$ | $19_{250}$ | $24_{9}$ | $22_{19}$ | | $23_{9}$ | $25_{19}$ | $22_{99}$ | $20_{99}$ | $41_{19}$ |
| $25_{49}$ | $20_{9}$ | $25_{19}$ | $23_{99}$ | | $24_{9}$ | $26_{19}$ | $23_{9}$ | $21_{9}$ | $41_{99}$ |
| $26_{250}$ | $23_{19}$ | $25_{99}$ | $25_{19}$ | | $27_{9}$ | $27_{19}$ | $24_{49}$ | $23_{9}$ | |
| $28_{9}$ | $23_{49}$ | $26_{250}$ | $26_{19}$ | | $27_{250}$ | $27_{99}$ | $24_{99}$ | $23_{19}$ | |
| $28_{49}$ | $28_{9}$ | $27_{99}$ | $27_{99}$ | | $28_{19}$ | $30_{19}$ | $27_{9}$ | $23_{49}$ | |
| $29_{9}$ | $28_{250}$ | $28_{9}$ | $28_{9}$ | | $30_{49}$ | $30_{49}$ | $27_{19}$ | $27_{9}$ | |
| $29_{99}$ | $29_{9}$ | $28_{250}$ | $28_{250}$ | | $30_{250}$ | $30_{250}$ | $30_{250}$ | $27_{19}$ | |
| $29_{250}$ | $29_{19}$ | $29_{250}$ | $29_{250}$ | | $32_{49}$ | $34_{9}$ | $31_{19}$ | $30_{49}$ | |
| $30_{19}$ | $29_{250}$ | $31_{9}$ | $31_{9}$ | | $34_{9}$ | $34_{99}$ | $34_{9}$ | $30_{99}$ | |
| $31_{9}$ | $31_{9}$ | $32_{9}$ | $32_{250}$ | | $34_{19}$ | $35_{99}$ | $34_{19}$ | $30_{250}$ | |
| $32_{9}$ | $31_{250}$ | $32_{250}$ | $34_{9}$ | | $34_{49}$ | $37_{19}$ | $35_{19}$ | $33_{250}$ | |
| $33_{9}$ | $32_{250}$ | $33_{49}$ | $35_{250}$ | | $35_{250}$ | $37_{49}$ | $37_{19}$ | $34_{19}$ | |
| $35_{9}$ | $40_{9}$ | $33_{250}$ | $36_{19}$ | | $37_{19}$ | $37_{250}$ | $41_{9}$ | $35_{19}$ | |
| $36_{9}$ | $40_{49}$ | $34_{99}$ | $40_{49}$ | | $41_{19}$ | $41_{19}$ | $41_{49}$ | $41_{19}$ | |
| $40_{250}$ | $40_{250}$ | $40_{9}$ | $40_{250}$ | | $41_{99}$ | $41_{49}$ | $41_{99}$ | $41_{99}$ | |

Table 4.7: Best and worst industries according to $\hat{\theta}^{m,C}_{(\boldsymbol{x}_i,\boldsymbol{y}_i)}$.

four we classify as structural leaders and structurals losers of the beginning of the first century. In each of the groups we further distinguish between those oriented towards processing of raw materials and those in advanced manufacturing.

**Proposition 4.2** ***Structural results.***

- ⋆ ***Leaders.*** *Most top efficient industries belong to sophisticated manufacturing: Food. Tobbaco products. Fabricated metal products. Machinery. Electrical machinery. Ra-*

*dio, 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. 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 surprizing, given the rising energy prices.*
- ⋆ *We identify one strong chain: metal ores—fabricated metal products—machinery—electrical machinery.*
- ⋆ *That the sectors automotive, coal & lignite and crude petroleum & natural gas place among the worst performers means that gains on a large scale are not always passed on to suppliers among SME.*

## 4.3 Estimation of the Parameters: SFA Results

### 4.3.1 Indentifying a Model

In the rest of the thesis, we implement several ideas which we described in secion 3.1. Regression analysis can yield twofold distinct results: (1) Sensitivity of the factors of production, ie estimation of the vector of parameters $\boldsymbol{\beta}$, and (2) estimation of the individual efficiency scores. Both are derived from the equation (3.1), nevertheless each of the aims requires slightly different setup. While the first problem can be consistently solved by standard ordinary least squares (OLS), the second requires a specific method.

We will first concentrate on the former point, the latter is investigated in section 4.4. Surely we could proceed immediately to the more advanced maximum likelihood estimation. However we decided to include ordinary least squares estimates because they offer fast routine computations and—more importantly—because OLS are accompanied by a set of standardized "quality checks". This way we establish a suitable model which we then pass on to the maximum likelihood.

If we apply the most widely used Cobb-Douglas production function, it turns equation (4.1) into:

$$\begin{aligned} VAD_i &= \beta_0 \cdot TAS_i^{\beta_1} \cdot IAS_i^{\beta_2} \cdot INV_i^{\beta_3} \\ &\quad \cdot AEM_i^{\beta_4} \cdot PAY_i^{\beta_5} \cdot OPE_i^{\beta_6} \cdot \exp\{\nu_i - \tau_i\}\,. \end{aligned} \tag{4.2}$$

Our objective is to make the best possible statistical inference about $\boldsymbol{\beta}$, which promts us to forget for a while about the distributional aspects of the term $\exp\{\nu_i - \tau_i\}$ which we made in secion 3.1; for convenience we will write $\epsilon_i = \nu_i - \tau_i$. Instead we want to arrive at approximately normally distributed residuals, so that the $t$-statistics in the regression model becomes accurate.

Because absolute figures enter the previous formula, we have good reason to suppose that the model will not satisfy homoskedasticity, ie constant variance of disturbances. We can expect large sectors to have larger variance among themselves, on the contrary to smaller sectors. To tackle this, we norm the variables to relative size—the most natural choice seems output. We get to the modified model:

$$\begin{aligned}\frac{VAD_i}{OUT_i} &= \beta_0 \cdot \left(\frac{TAS_i}{OUT_i}\right)^{\beta_1} \cdot \left(\frac{IAS_i}{OUT_i}\right)^{\beta_2} \cdot \left(\frac{INV_i}{OUT_i}\right)^{\beta_3} \cdot \left(\frac{AEM_i}{OUT_i}\right)^{\beta_4} \cdot \\ &\cdot \left(\frac{PAY_i}{OUT_i}\right)^{\beta_5} \cdot \left(\frac{OPE_i}{OUT_i}\right)^{\beta_6} \cdot \exp\{\epsilon_i\}.^{2} \qquad (4.3)\end{aligned}$$

There are two more point that deserve adjustment. $INV$ consist of acquisition of assets whose period of usage is longer than one year. These investments often take some time to be realized, thus in financial analyses they are taken as lagged values, $INV_{-1,i}$. Accordingly we will only be able to model just three years out of the four, 2003 through 2005.

The second remark concerns employees. We can think of $PAY$ as the multiple of $AEM$ and average pay per employee, $APAY$. By including $\left(\frac{AEM_i}{OUT_i}\right)^{\beta_4} \cdot \left(\frac{PAY_i}{OUT_i}\right)^{\beta_5}$, we are in fact counting $AEM$ twice. Therefore our resulting starting point for regression is the following model:

$$\begin{aligned}\frac{VAD_i}{OUT_i} &= \beta_0 \cdot \left(\frac{TAS_i}{OUT_i}\right)^{\beta_1} \cdot \left(\frac{IAS_i}{OUT_i}\right)^{\beta_2} \cdot \left(\frac{INV_{-1,i}}{OUT_{-1,i}}\right)^{\beta_3} \cdot \left[\left(\frac{AEM_i}{OUT_i}\right)^{\beta_4}\right] \cdot \\ &\cdot \left(\frac{PAY_i}{OUT_i}\right)^{\beta_5} \cdot \left(\frac{OPE_i}{OUT_i}\right)^{\beta_6} \cdot \exp\{\epsilon_i\}, \qquad (4.4)\end{aligned}$$

where the square brackets indicate that we will explicitely handle the submodel which omits $AEM$. This model production function (4.4) is of course estimated in its log-linear form.

[2]We could further assume $\sum_{p=1}^{6} \beta_p = 1$, then (4.3) would be an equivalent expression of the first raw model. This restriction would however carry with it computational complications, so that we take (4.3) as a new model instead.

### 4.3.2 Standard Regression

We evaluate ordinary least squares regression for (4.4) in the statistical package $\mathcal{R}$.[3]

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.3249 | 0.6681 | 0.49 | 0.6276 |
| $TASn.03$ | 0.0596 | 0.0399 | 1.49 | 0.1375 |
| $IASn.03$ | −0.0127 | 0.0186 | −0.68 | 0.4955 |
| $INVn.02$ | 0.0410 | 0.0365 | 1.12 | 0.2636 |
| $AEMn.03$ | 0.1559 | 0.1185 | 1.32 | 0.1907 |
| $PAYn.03$ | −0.9431 | 0.1330 | −7.09 | 0.0000 |
| $OPEn.03$ | 1.2971 | 0.1785 | 7.27 | 0.0000 |

Residual standard error: 0.2859 on 125 degrees of freedom
Multiple R-Squared: 0.4354, Adjusted R-squared: 0.4083
F-statistic: 16.06 on 6 and 125 DF, p-value: 1.237e-13

Table 4.8: OLS for the model (4.4), data for 2003.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 2.0138 | 0.3081 | 6.54 | 0.0000 |
| $TASn.04$ | 0.0697 | 0.0288 | 2.42 | 0.0169 |
| $IASn.04$ | 0.0364 | 0.0130 | 2.81 | 0.0058 |
| $INVn.03$ | 0.0353 | 0.0341 | 1.04 | 0.3017 |
| $AEMn.04$ | 0.3888 | 0.0522 | 7.45 | 0.0000 |
| $PAYn.04$ | −0.5764 | 0.0942 | −6.12 | 0.0000 |
| $OPEn.04$ | 0.6820 | 0.0662 | 10.30 | 0.0000 |

Residual standard error: 0.2535 on 125 degrees of freedom
Multiple R-Squared: 0.6269, Adjusted R-squared: 0.609
F-statistic: 35.01 on 6 and 125 DF, p-value: < 2.2e-16

Table 4.9: OLS for the model (4.4), data for 2004.

The obtained results are presented in tables 4.8, 4.9 and 4.10. We added $n$ to the names of explanatory variable to indicate that they were normed by $OUT$ before taking logarithms.

R-squared indicates that we are able to explain between 40 and 60 per cent of the original data variation. This does not seem much, and less so since estimating the model (4.2) with lagged investment yields R-squared above 90 percent, as in

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

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 1.5003 | 0.4124 | 3.64 | 0.0004 |
| $TASn.05$ | 0.1431 | 0.0313 | 4.57 | 0.0000 |
| $IASn.05$ | 0.0019 | 0.0150 | 0.13 | 0.8968 |
| $INVn.04$ | −0.0806 | 0.0293 | −2.76 | 0.0067 |
| $AEMn.05$ | 0.4121 | 0.0751 | 5.49 | 0.0000 |
| $PAYn.05$ | −0.5513 | 0.1309 | −4.21 | 0.0000 |
| $OPEn.05$ | 0.4907 | 0.0682 | 7.20 | 0.0000 |

Residual standard error: 0.2753 on 125 degrees of freedom

Multiple R-Squared: 0.4789, Adjusted R-squared: 0.4538

F-statistic: 19.14 on 6 and 125 DF, p-value: 9.91e-16

Table 4.10: OLS for the model (4.4), data for 2005.

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 3.8521 | 0.5733 | 6.72 | 0.0000 |
| $TAS.05$ | 0.2100 | 0.0370 | 5.67 | 0.0000 |
| $IAS.05$ | 0.0174 | 0.0145 | 1.20 | 0.2320 |
| $INV.04$ | 0.0045 | 0.0344 | 0.13 | 0.8955 |
| $AEM.05$ | 0.5636 | 0.0957 | 5.89 | 0.0000 |
| $PAY.05$ | −1.1023 | 0.1756 | −6.28 | 0.0000 |
| $OPE.05$ | 1.2652 | 0.2155 | 5.87 | 0.0000 |

Residual standard error: 0.3711 on 125 degrees of freedom

Multiple R-squared: 0.9569, Adjusted R-squared: 0.9548

F-statistic: 462.3 on 6 and 125 DF, p-value: < 2.2e-16

Table 4.11: OLS for the model (4.2) with lagged $INV$, data for 2005.

table 4.11 for the year 2005. But after performing some basic diagnostics, (4.4) appears more favourable than the equation in the non-normed form, at least in terms of statistical inference.

The most important tool for inference is normality of residuals. Among the plenty of tests available, we decided for Kolmogorov-Smirnov and Shapiro-Wilk.[4] We do not detect any normality in models (4.2, lagged) for any year, since $p$-values are of order $10^{-5}$ to $10^{-6}$. This also holds for 2003 for (4.4), but we get $p$-values 0.078/0.017 for 2004 and 0.078/0.021 for 2005. Thus the transformation by $OUT$

[4]It may be of interest that the $\chi^2$-test, which is very popular as well, returned $p$-values two orders higher than the other five tests which we implemented in $\mathcal{R}$. Due to this strange behaviour, we disregarded its $p$-values.

shifted the results by a considerable amount towards normality.

Another property we are expecting is homoscedasticity, which we already touched above. We assume that nonspherical disturbances could stem from the size variation and use Breusch-Pagan to test for this effect.[5] The difference between (4.2, lagged) and (4.4) is not so clear, (4.4) does not always perform better: it yields $p$-values 0.096 compared to 0.016 in 2005 but 0.228 versus 0.467 in 2004. We infer that the source of the presumably largest problems with non-constant variance is not statistically significant. However we are aware that more detailed study of heteroscedasticity would be useful.

We could also be concerned with multicollinearity, ie additional dependence among explanatory variables. Kennedy (1992, p. 181) mentiones exactly the example of the 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, which is just our case (except for the insignificant variables of course). Having discussed basic diagnostics of our model, we can proceed with interpretation of the estimates.

### 4.3.3 Production Function of Czech SME

Following the results from the previous tables, we can write for the year 2005 the estimate of production function as:

$$\begin{aligned} \frac{\widehat{VAD_i}}{OUT_i} &= 4.483 \cdot \left(\frac{TAS_i}{OUT_i}\right)^{0.143} \cdot \left(\frac{IAS_i}{OUT_i}\right)^{0.001} \cdot \left(\frac{INV_{-1,i}}{OUT_{-1,i}}\right)^{-0.081} \cdot \\ &\cdot \left(\frac{AEM_i}{OUT_i}\right)^{0.412} \cdot \left(\frac{PAY_i}{OUT_i}\right)^{-0.551} \cdot \left(\frac{OPE_i}{OUT_i}\right)^{0.491}, \end{aligned}$$

where however the presence of the expression $\frac{IAS_i}{OUT_i}$ is not significant. The same can be done for 2004 and 2003.

One point of view is the significance of estimated parameters, which is tenable at least for 2004 and 2005 due to normality of residuals. There are two candidates for variables that could be left out: lagged investment and intangible assets. Moreover $INVn.05$ got a negative coefficient. We know that investment, defined as acquisition of long term assets, is at first a considerable expense, which should later turn profitable. The fluctuation of the coefficient between positive and negative values

[5]For estimations whose residuals are not normally distributed we use the studentized Breusch-Pagan test (Kennedy, 1992, p. 130).

suggests that the one-year lag is not always enough for the investment to generate positive revenue.

We find the insignificance of intangible assets more discouraging, because it in fact means that no high-tech revolution occured in the past years in the SME sector. Since intangible assets capture immaterial results of R&D, it seems 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 posseses any of these, we would recognize it as an advantage.

Table 1.3 documents the investment surge among small and medium enterprises, which in conjunction with our results implies two possible hypotheses: (1) SME invest in standard "hard" technologies, while research is not stressed enough. (2) Investment turns profitable only after a few years, which makes it especially troublesome for SME to spend money there.

We tested the second hypothesis by lagging $INV$ by two and three years for 2005, other things equal. With the three year lag, we finally arrived at a positive coefficient, yet none of the lags was statistically significant (with $p$-values 0.842757 for $INV_{-2}$ and 0.2707 for $INV_{-3}$).

The other aspect is the sign of the estimated coefficients, which must be interpreted carefully. We find that payroll has negative influence on value added. It is true that wages are costs. Yet the theory also says that differences in wages should reflect quality of the workforce, and hence higher salaries should result in more value added. Our estimate cannot confirm this mechanism. On the contrary, from the joint result for $AEM$ and $PAY$ we conclude that SME rely on low-skilled and thus low-cost labourers.

The interpretation of other personnel expenses is rather vague. The positive coefficient could prompt us to deduce that outsourcing tasks beyond regular contracts is efficient. Yet in the indicator several types of expenses is mixed, among others expert testimonies or royalties, which in fact slightly weakens our previous point that intangible assets are unimportant. Hence we cannot draw any strong conclusion from this variable. Finally, the results for tangible assets and the number of employees are as expected.

The discussion yields the following summary:

**Proposition 4.3** ***Production function characteristics.***

- ⋆ *Our regressions confirm that Czech SME depend more on labour than on capital. Their profitability is determined by the ability to employ lots of people and to pay them little.*

- ⋆ *SME seem not to be able to reap the benefits of intangible assets, such as software or patents.*
- ⋆ *The effect of previous investment turns out hazy: it influences the production either negatively or not at all.*
- ⋆ *The above points mean that SME fundamentals of the Czech economy have not yet converted to an innovation based production process on a large scale.*

## 4.4 Effects of Size and Time

### 4.4.1 Model Specification

In this section we focus on the estimation of efficiency of individual industries by using the parametric approach. In other words, we replicate what was done in section 4.2 by another method, and we have two reasons: (1) to compare the outcome of both methods, and (2) to test the effects of size and time on efficiency levels.

As shown in section 3.1.1, the starting point is now the equation (3.3). The solution of this maximum likelihood maximization is implemented in the freely available program FRONTIER by T. Coelli (Coelli, 1996b), which moreover offers several extension to this basic model. The other possibility is to estimate the moments given by (3.5) and (3.6) from OLS residuals.

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

$$\begin{aligned} y_{it} &= \beta_0 + \sum_{j=1}^{p} \beta_j \cdot x_{ijt} + (\nu_{it} - \tau_{it}) \qquad (4.5)\\ \nu_{it} &\ \ldots\ iid,\ \mathcal{L}(\nu_{it}) \sim \mathcal{N}(0, \sigma_\nu^2),\\ \tau_{it} &= \tau_i \cdot \exp\left\{-\eta \cdot (t - T)\right\},\\ \tau_i &\ \ldots\ iid,\ \text{truncations at zero of } \mathcal{N}(\mu_\tau, \sigma_\tau^2). \end{aligned}$$

As in the previous text, we would plug in logarithms of the data rows. We can further specify that we have only one time period, that the efficiency time-variation parameter $\eta$ be zero and the non-centrality parameter of the truncated normal distribution $\mu_\tau$ be zero, so that we get our ininitial model (3.2). The computational mechanism consists of three steps.[6] Its brief description, as well as further refer-

[6](1) OLS, (2) grid search for initial values, which are passed on to the iteration, and (3) Davidon-Fletcher-Powell Quasi-Newton iterative mechanism which maximizes the log-likelihood function.

ences which treat computational aspects, are mentioned by Coelli in the FRONTIER manual (Coelli, 1996b, c.f.).

At first we compare the MLE result with the estimates by sample moments. We take the six normed variables with lagged investment as in (4.4) for the year 2004 (ie we have only one time period, $\eta = 0$). Denote $\gamma = \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma_\nu^2}$. Then if we use OLS residuals, we obtain $\hat{\sigma}^2 = 0.0614$ and $\hat{\gamma} = 0.0028$, while maximum likelihood returns $\hat{\sigma}^2 = 0.1354$ and $\hat{\gamma} = 0.7900$. In the case of gamma we see a difference of order two, which is implausible. Due to this variation in results, in the rest of the section we work with the maximum likelihood output given by FRONTIER, because this approach has been extensively used in literature and therefore has strong theoretical backing.

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

$$\begin{aligned} y_{it} &= \beta_0 + \sum_{j=1}^{p} \beta_j \cdot x_{ijt} + (\nu_{it} - \tau_{it}) \\ \nu_{it} &\dots\ iid,\ \mathcal{L}(\nu_{it}) \sim \mathcal{N}(0, \sigma_\nu^2), \\ \tau_i &\dots\ iid,\ \text{truncations at zero of } \mathcal{N}(\xi_{it}, \sigma_\nu^2), \\ \xi_{it} &= \delta_0 + \sum_{h=1}^{d} \delta_h \cdot z_{it}, \end{aligned} \tag{4.6}$$

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 $\delta$ 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. Because $\tau$ is defined as a random variable, we have the random framework for modelling: $E(\tau|\boldsymbol{z}) = \xi = \delta_0 + \boldsymbol{z}^T\boldsymbol{\delta}$.

Coelli (1996b) remarks that the models (4.5) and (4.6) are not nested, so they cannot be tested against each other.

What deserves special attention is the distribution of the inefficiency term $\tau$. In section 3.1.1 we assumed $\mathcal{L}(\tau_i) \sim \mathcal{N}^+(0, \sigma_\tau^2)$, so that the distribution was half normal. In equations (4.5) and (4.6) 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. On the contrary, if the underlying density is modelled as non-central with $\mu_\tau > 0$, 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. This is a bold idea, but with regard to the boxplots (figures 4.1 and 4.2) we conlude that it is worth a proper statistical test.

| | ♯ of employees | min | $1Q$ | median | $3Q$ | max | mean |
|---|---|---|---|---|---|---|---|
| | Not restricted | 0.3310 | 0.7257 | 0.8034 | 0.8684 | 0.9697 | 0.7878 |
| | <10 | 0.3310 | 0.7793 | 0.8878 | 0.9330 | 0.9697 | 0.8457 |
| 2003 | 10-19 | 0.5401 | 0.7267 | 0.8073 | 0.8683 | 0.9313 | 0.7886 |
| to | 20-49 | 0.5185 | 0.7348 | 0.8049 | 0.8483 | 0.9003 | 0.7768 |
| 2005 | 50-99 | 0.5793 | 0.6869 | 0.7949 | 0.8269 | 0.9364 | 0.7647 |
| | 100-250 | 0.5446 | 0.6926 | 0.7669 | 0.8144 | 0.9148 | 0.7560 |

Table 4.12: Box plot statistics for maximum likelihood efficiency scores using (4.5).

Before we turn to tests of hypotheses, we list table 4.12, which contains box plot statistics for the MLE estimates of efficiency scores using (4.5), where $\mu_\tau$ and $\eta$ are set to zero. This time the scores are joint for 2003 to 2005; the reason why we merged the years is explained in the next chapter.

### 4.4.2 Estimating the Common Practice

Before we turn to the desired statistical test, we have to consider some computational aspects of MLE. One particular difficulty arises concerning the assumption on the underlying distribution of $\epsilon_i = \nu_i - \tau_i$. Simar & Wilson (2005) pointed out two possible problems of the half-normal distribution: (1) $\sigma_\tau^2 = 0$, and (2) $\hat{\epsilon}$ have positive skewness. Both lead to wrong computations. The latter case means that the random part of the frontier would shift most producers above the efficient level (since exponential of a positive number is higher than one). Simar & Wilson firmly warn from the intuitive interpretation that it further means that the model is misspecified. They argue that positive skewness in some cases will from time to time occur in finite samples and that the right remedy is to increase the number of observations.[7]

Among the three inspected years 2003 to 2005,[8] we indeed encounter just this case: For 2005 the OLS residuals are positively skewed, so that to use MLE for this

[7]To mention one numerical example, in their simulation Simar & Wilson (2005) find that: "If $\lambda^2 = 0.5$—which is far from implausible in applied studies—even with 1,000 observations, one should expect that about 22.5 percent of samples will have positive skewness."

[8]Recall that we left out 2002 since we cannot use lagged investment for this year.

subset of observations is not plausible.[9]

Other complications with small samples were shown by Ritter & Simar (1997). They presented the situation where a combination of different distributions of $\nu$ and $\tau$ is chosen (their combination is normal-gamma). Yet Simar & Wilson (2005, c. f.) claim that the same should be expected for the truncated non-central normal distribution of the inefficiency term $\tau$, which is exactly the case of equations (4.5) and (4.6).

Therefore we found two compelling reasons to apply the maximum likelihood equation on the joint dataset, ie all the three years together: (1) wrong skewness of residuals in 2005, and (2) employment of the non-central truncated normal distribution for $\tau$. By this increase of the number of observations, the quality of the resulting estimate should substantially improve and offer a solid basis for inference.

We apply the model (4.6), where we specify $x$ to be the same six variables as in (4.4). We use $t = 2003, 2004, 2005$. In order 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.

| | Estimate | Std. Error | t ratio |
|---|---|---|---|
| $\mu_\tau$ | 0.39522 | 0.06441 | 6.13572 |
| (Intercept) | 1.86460 | 0.25560 | 7.29485 |
| $TASn$ | 0.08546 | 0.01813 | 4.7144 |
| $IASn$ | 0.01036 | 0.00816 | 1.26993 |
| $INVn_{-1}$ | −0.01923 | 0.01780 | −1.08301 |
| $AEMn$ | 0.36763 | 0.04514 | 8.14354 |
| $PAYn$ | −0.52141 | 0.07247 | −7.19485 |
| $OPEn$ | 0.54343 | 0.43216 | 12.57470 |
| sigma-squared | 0.07813 | 0.01216 | 6.42307 |
| gamma | 0.49984 | 0.05249 | 9.52275 |

Table 4.13: MLE result for $\boldsymbol{\beta}$ and $\mu_\tau$, equation (4.5), data for 2003-2005.

The result, along with the coefficients of all explanatory variables, is reported in table 4.13. We get estimated efficiency $E\widehat{\mathcal{E}_i} = \exp\{-\hat{\mu}_\tau\} = \exp\{-0.39522\} = 0.677$. This yields the expected mean efficiency level of two thirds.

[9] Empirical skewness computed from sample second and third central moments given in section 3.1.1 was equal to

$$\frac{[s^3]}{[s^2]^{3/2}} = 0.06437.$$

Let us formulate the first hypothesis to test: Under the null $\mu_\tau = 0$, under the alternative $\mu_\tau \neq 0$. In FRONTIER we solve (4.5), where we set $\eta = 0$.

For the test itself we use the likelihood-ratio test, which is specified as in Battese & Coelli (1995):[10]

$$-2 \cdot \{\log \mathscr{L}(H_0) - \log \mathscr{L}(H_A)\} \overset{\text{asymptotically}}{\longrightarrow} \chi^2(J),$$

where the degrees of freedom $J$ are equal to the number of restrictions, ie we have $J = 1$. $\log \mathscr{L}$ are the values of the log-likelihood function for the respective case. We get:

$$\hat{\chi}^2 = -2 \cdot (-23.30802 + 19.71990) = 7.17624 \rightsquigarrow p = 0.01478.^{11}$$

Whether the hypothesis is rejected depends on the significance level. Given the previous results of DEA and the already mentioned boxplots, we conclude that on the weaker level $\alpha = 0.05$ we reject the null.

**Proposition 4.4** ***Systematic inefficiency.*** *The production function of Czech small and medium enterprises (4.4) is likely to contain systematic inefficiency, meaning that the common practice is significantly different from the best practice.*

This conclusion could be questioned on the basis that we already stated that the presence of outliers probably influences the overall distribution of efficiency scores (proposition 4.1). Nevertheless if we compare table 4.6 with the the mean of efficiency scores estimated by (4.5), which equals to 0.7878, and the spread respective spreads between the first and the third quartile, which are 0.4278 by COM versus 0.1426 by MLE, we can easily challenge this objection.

### 4.4.3 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. Table 4.12 suggests that the relationship indeed is inverse, which is a bit perplexing. One reason could be that if Czech SME lack capital due to low past investment, with additional labour the efficiency has to decrease. But this explanation is rather artificial, because in fact we would expect the smallest firms to have invested least.

We use $EGR$, the group according to the number of employees, as the single $z$ variable. Hence $EGR \in \{1, 2, 3, 4, 5\}$, as corresponds to the breakdown 0-10-20-50-100-250 (refer to section 4.1). In table 4.14, we report results for the estimate of $\boldsymbol{\delta}$.

[10] Two other options could be used: the Wald test and the Lagrange multiplier test (Kennedy, 1992, p. 61).

[11] $p$-value was computed as: $p = 2 \times \min\{\Pr(\mathcal{C} \geq \hat{\chi}^2), \Pr(\mathcal{C} < \hat{\chi}^2)\}$, where the random variable $\mathcal{C}$ is governed by $\chi^2(J)$.

| | Estimate | Std. Error | t ratio |
|---|---|---|---|
| $\delta_0$ | $-10.92826$ | 5.44829 | $-2.00581$ |
| $\delta_1$ | 0.83951 | 0.35894 | 2.33890 |
| $\log \mathscr{L}(H_0)$ | $-43.07255$ | $\log \mathscr{L}(H_1)$ | $-40.67602$ |

Table 4.14: MLE result for $\boldsymbol{\delta}$, equation (4.5), data for 2003-2005.

Since we estimate $\xi$ as the mean of inefficiency, the computed coefficient means that larger firms perform actually worse. We get estimated efficiency $E\widehat{\mathcal{E}}_i = \exp\left\{-\hat{\xi}\right\} = \exp\left\{-0.06569 \cdot EGR\right\}$. This means that the efficiency level goes down from 94% for the smallest firms to about 74% for the largest firms.

To test the significance of this result, our hypothesis is as follows: Under the null $\delta_1 = 0$, under the alternative $\delta_1 \neq 0$.

The test statistics is the same as in the previous section:

$$\hat{\chi}^2 = -2 \cdot (-43.07255 + 40.67602) = 4.79306 \rightsquigarrow p = 0.05715.$$

Contrary to the previous section, we cannot reject the null, so that the effect of $EGR$ is insignificant.

**Proposition 4.5** ***Size effect.*** *Although the suggested relationship between a firm's size and its efficiency is inverse, we do not find it statistically significant.*

### 4.4.4 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. Our idea is to relate this test to figure 4.2, where the shifts of mean scores are unclear and where we even observe a dip between 2003 and 2004.

We use the specification from equation (4.5). The log-likelihoods are in table 4.15. The test is again the likelihood-ratio. It makes sense to test the null of $\eta = 0$ against the alternative of $\eta \neq 0$ for two cases: once with $\mu_\tau = 0$ and once with $\mu_\tau \neq 0$.

In the former case, ie testing the second column, we get $p = 0.01174$, and in the latter (third column) we compute $p = 0.01105$. Because we already concluded that $\mu_\tau \neq 0$ (proposition 4.4) and because we reject the null on the significance level $\alpha = 0.05$, we derive that the best model is with $\hat{\eta} = 0.06441$ and $\hat{\mu}_\tau = 0.36851$ (note that this estimate of $\mu_\tau$ is very close to that in table 4.13).

| | $\mu_\tau = 0$ | $\mu_\tau \neq 0$ |
|---|---|---|
| $\eta = 0$ | $\log \mathscr{L} = -23.30802$ | $\log \mathscr{L} = -19.71990$ |
| $\eta \neq 0$ | $\log \mathscr{L} = -19.51304$ | $\log \mathscr{L} = -15.87048$ |
| $\eta$ | 0.12238 | 0.06441 |
| (Std. Error) | 0.04548 | 0.04002 |

Table 4.15: MLE result for $\eta$ and $\mu_\tau$, equation (4.5), data for 2003-2005.

To get an idea about the magnitude of this estimated effect, we use equation (4.5) which yields that the inefficiency index fell from $1.12 \times \tau_i$ in 2003 to $1 \times \tau_i$ in 2005. The estimated time effect was about twelve per cent.

**Proposition 4.6** ***Time effect.*** *During the three observed years, efficiency of Czech SME increased; nonetheless this effect is only weakly significant.*

# Conclusions

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

The methods for efficiency measurement we used are 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.

Our toolbox consisted of two compartments: Data envelopment analysis (DEA) constructs the boundary of the multidimensional set of observations and measures the distance of firms from this efficient frontier. 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 margin between fitted and observed values.

We were able to gather several propositions about Czech SME in the form of stylized facts. The nature of our results reflects the different ways in which the two methods work—we moreover expected subtle variations between the two outcomes.

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 turned 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 propostion 4.2 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 large scale boom of big factories is not necessarily passed on to SME suppliers—eg automotive; coal & lignite; crude petroleum & natural gas.

SFA estimates production and profit functions and due to its statistical formulation, it offers better procedures for testing hypotheses. We used ordinary least squares and the accompanying standardized diagnostics to find out an acceptable specification which we then hand over to maximum likelihood estimators.

We confirmed that Czech SME are more dependent on labour employment rather than on capital usage. Further, the investment surge documented in figure 1.1 was not mirrored in our estimates. There are more feasible explanations: (1) It takes more time for the investments to start generating profits, so that our dataset is too old to capture this effect. (2) Investment is not evenly spread across industries; while it improves efficiency of successful and growing companies, it turns out globally insignificant. If true, this reasoning means that recent investments lead to divergence in efficiency score and that there is a dramatic shift in the structure of Czech SME on the way. (3) If we asses the contribution of past investment to value added as insignificant, it might stem from the fact that some of the investment projects were and are not prosperous. In other words we can imagine that SME have not yet learnt to invest effectively. (4) Figure 1.1 contains a broader group of SME than our dataset, namely the construction industry, commerce and part of services. It is possible that these sectors cared for a larger part of the observed jump in investment. Finally we note that these points are not mutually exclusive, hence in fact all of them can be partially true.

Similar conclusions can be made for the role played by intangible assets, which include goodwill, software, patents, copyrights, trademarks and tradenames. That their presence in the production function is insignificant means that Czech SME are not yet innovation driven on a large scale.

At last we performed three specific tests: We found the presence of systematic inefficiency highly significant. By this we mean than instead of being close to the effcient frontier, the majority firm operates on a considerably lower level of efficiency.

However we were not able to accept the hypothesis which would explain this inefficiency by the size of enterprises. Even if confirmed, the size effect would work in the opposite direction than expected: larger firm performing worse. Our conclusion is compatible with that of DEA: There we see the size effect to be mixed and to become more indistinct with the robust specification. At large it hints that the structure of inefficiency effects is much more complicated in reality and that it would deserve a more detailed treatment, eg by analyzing clusters of related industries.

As with every empirical study, we are well aware of the fact that practice requires compromise. One that we encountered throughout the thesis 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 put a lot of effort to present a comprehensive analysis of the production patterns of Czech SME. However we recognize that there remains vast room for further enhancements of this analysis. Besides detailed inspection of the distribution of efficiency scores across industries and identification of clusters of industries, one could run exactly defined tests on the underlying type of returns to scale. Both might help to explain better the relationship between size and efficiency. Similarly we barely touched any deliberations on efficiency change over time. All of these are left for further research.

# Appendix A

# Industrial Classification of Economic Activities

The dataset at hand consists of rows for aggregated industries. We list complete Czech OKEC[1] definitions for industries that we have available. The following is a transcript of definitions used by the Czech Statistical Office[2] and their English equivalents used by the European Union[3].

| Code | Czech description | English description |
|---|---|---|
| 10 | Těžba uhlí, lignitu a rašeliny | Mining of coal and lignite; extraction of peat |
| 11 | Těžba ropy, zemního plynu a související činnosti kromě průzkumných vrtů | Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction, excluding surveying |
| 12 | Těžba a úprava uranových a thoriových rud *(neobsaženo)* | Mining of uranium and thorium ores *(not included)* |
| 13 | Těžba a úprava ostatních rud | Mining of metal ores |
| 14 | Těžba a úprava ostatních nerostných surovin | Other mining and quarrying |

Table A.1: Selected OKEC/NACE classification.

[1] Odvětvová klasifikace ekonomických činností.

[2] See [http://www.czso.cz/csu/klasifik.nsf/i/odvetvova_klasifikace_ekonomickych_cinnosti_(okec)].

[3] Nomenclature Générale des Activités Économiques dans les Communautés Européennes, or NACE, see [http://ec.europa.eu/comm/competition/mergers/cases/index/nace_all.html].

[4] N.e.c. = Not elsewhere classified.

| Code | Czech description | English description |
|---|---|---|
| 15 | Výroba potravinářských výrobků a nápojů | Manufacture of food products and beverages |
| 16 | Výroba tabákových výrobků | Manufacture of tobacco products |
| 17 | Výroba textilií a textilních výrobků | Manufacture of textiles |
| 18 | Výroba oděvů, zpracování a barvení kožešin | Manufacture of wearing apparel; dressing and dyeing of fur |
| 19 | Činění a úprava usní, výroba brašnářských a sedlářských výrobků a obuvi | Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear |
| 20 | Zpracování dřeva, výroba dřevařských, korkových, proutěných a slaměných výrobků kromě nábytku | Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials |
| 21 | Výroba vlákniny, papíru a výrobků z papíru | Manufacture of pulp, paper and paper products |
| 22 | Vydavatelství, tisk a rozmnožování nahraných nosičů | Publishing, printing and reproduction of recorded media |
| 23 | Výroba koksu, jaderných paliv, rafinérské zpracování ropy | Manufacture of coke, refined petroleum products and nuclear fuel |
| 24 | Výroba chemických látek, přípravků, léčiv a chemických vláken | Manufacture of chemicals and chemical products |
| 25 | Výroba pryžových a plastových výrobků | Manufacture of rubber and plastic products |

Table A.2: Selected OKEC/NACE classification (continued).

| Code | Czech description | English description |
|---|---|---|
| 26 | Výroba ostatních nekovových minerálních výrobků | Manufacture of other non-metallic mineral products |
| 27 | Výroba základních kovů a hutních výrobků | Manufacture of basic metals |
| 28 | Výroba kovových konstrukcí a kovodělných výrobků (kromě strojů a zařízení) | Manufacture of fabricated metal products, except machinery and equipment |
| 29 | Výroba a opravy strojů a zařízení j. n. | Manufacture of machinery and equipment n.e.c.[4] |
| 30 | Výroba kancelářských strojů a počítačů | Manufacture of office machinery and computers |
| 31 | Výroba elektrických strojů a zařízení j. n. | Manufacture of electrical machinery and apparatus n.e.c. |
| 32 | Výroba rádiových, televizních a spojových zařízení a přístrojů | Manufacture of radio, television and communication equipment and apparatus |
| 33 | Výroba zdravotnických, přesných, optických a časoměrných přístrojů | Manufacture of medical, precision and optical instruments, watches and clocks |
| 34 | Výroba motorových vozidel (kromě motocyklů), výroba přívěsů a návěsů | Manufacture of motor vehicles, trailers and semi-trailers |
| 35 | Výroba ostatních dopravních prostředků a zařízení | Manufacture of other transport equipment |
| 36 | Výroba nábytku; zpracovatelský průmysl j. n. | Manufacture of furniture; manufacturing n.e.c. |
| 37 | Recyklace druhotných surovin | Recycling |
| 40 | Výroba a rozvod elektřiny, plynu a tepelné energie | Electricity, gas, steam and hot water supply |
| 41 | Shromažďování, úprava a rozvod vody | Collection, purification and distribution of water |

Table A.3: Selected OKEC/NACE classification (continued).

# Appendix B

# Data Definition

We give complete definitions of the data obtained from the Czech Statistical office. These definitions are available online.[1] 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 installements, (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.

[1] [http://dw.czso.cz/pls/metis/TUCUK_N.ZAC].

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 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.

# References

AIGNER, D. J., & CHU, S. F. 1968. On Estimating the Industry Production Function. *The American Economic Review*, **58**, 826–839.

AIGNER, D. J., LOVELL, C. A. K., & SCHMIDT, P. 1977. Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics*, **6**, 21–37.

AYYAGARI, M., BECK, T., & DEMIRGUC-KUNT, A. 2007. Small and Medium Enterprises across the Globe. *Small Business Economics*, **29**, 414–434.

BATTESE, G. E., & COELLI, T. J. 1988. Prediction of Firm-level Technical Efficiencies with a Generalized Frontier Producion Function and Panel Data. *Journal of Econometrics*, **38**, 387–399.

BATTESE, G. E., & COELLI, T. J. 1995. A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics*, **20**, 325–332.

BENÁČEK, V. 1994. *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. Downloaded at [http://www.cerge-ei.cz/pdf/wp/Wp53.pdf].

BENÁČEK, V., SHEMETILO, D., & PETROV, A. 1997. Restructuring and Measurement of Efficiency in Firms in Transition. *Pages 58–72 of:* SARMA, S. (ed), *Restructuring Eastern Europe*. Cheltenham: Elgar. ISBN: 1-858-985-765.

CAZALS, C., FLORENS, J.-P., & SIMAR, L. 2002. Nonparametric Frontier Estimation: A Robust Approach. *Journal of Econometrics*, **106**, 1–25.

COELLI, T. J. 1996a. *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. Available at [http://www.uq.edu.au/economics/cepa/index.htm]. ISBN: 1-863-89-4969.

COELLI, T. J. 1996b. *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. Available at [http://www.uq.edu.au/economics/cepa/index.htm]. ISBN: 1-86389-4950.

COELLI, T. J., RAO, P., O'DONNELL, CH. J., & BATTESE, G. E. 2005. *An Introduction to Efficiency and Productivity Analysis*. New York: Springer.

COOPER, W. W., SEIFORD, L. M., & ZHU, J. (eds). 2004. *Handbook on Data Envelopment Analysis*. New York: Kluwer. eBook ISBN: 1-4020-7798-X.

COOPER, W.W., SEIFORD, L.M., & TONE, K. 2002. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. New York: Kluwer. eBook ISBN: 0-306-47541-3.

DARAIO, C., & SIMAR, L. 2007a. *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Studies in Productivity and Efficiency. New York: Springer. ISBN: 0-387-35155-8.

DARAIO, C., & SIMAR, L. 2007b. Conditional Nonparametric Frontier Models for Convex and Nonconvex Technologies: A Unifying Approach. *Journal of Productivity Analysis*, **28**, 13–32.

DUŠEK, L., & ŽIGIĆ, K. (eds). 2005. *Czech Republic 2005: Year After*. P. O. Box 882, Politických vězňů 7, 11121 Prague, Czech Republic: CERGE-EI. ISBN: 80-7343-058-4.

HÁJEK, M. 2005. *Economic Growth and Total Factor Productivity in the Czech Republic from 1992 to 2004*. Working paper 5/2005. Cetrum ekonomických studií VŠEM.

JOHANSEN, L. 1972. *Production Functions: An Integration of Micro and Macro, Short Run and Long Run Aspects*. 1st edn. Amsterdam: North-Holland. ISBN 0-7204-3175-1.

KENNEDY, P. 1992. *A Guide to Econometrics*. Oxford (UK): Blackwell. ISBN: 0-631-18211-X.

KOGIKU, K. C. 1971. *Microeconomic Models*. New York: Happer & Row. SBN: 06-043746-4.

KUMBHAKAR, S. C., & LOVELL, C. A. K. 2000. *Stochastic Frontier Analysis*. Cambridge University Press. ISBN: 0-521-48184-8.

LIEBENSTEIN, H. 1966. Allocative Efficiency vs. "X-efficiency". *The American Economic Review*, **56**, 392–415.

LOPEZ-CARLOS, A. (ed). 2005. *The Global Competitiveness Report 2005–2006*. Hampshire: Palgrave MacMillan. World Economic Forum, ISBN-13: 978-1-4039-9844-6.

LOPEZ-CARLOS, A. (ed). 2006. *The Global Competitiveness Report 2006–2007*. Hampshire: Palgrave MacMillan. World Economic Forum, ISBN-13: 978-1-4039-9636-7.

SALA-I-MARTIN, X. (ed). 2003. *The Global Competitiveness Report 2003–2004*. Hampshire: Palgrave MacMillan. World Economic Forum, ISBN: 0-19-517360-0.

MOHNEN, A., & NASEV, J. 2005 (October). *Growth of Small and Medium-Sized Firms in Germany*. Available at SSRN: [http://ssrn.com/abstract=852785].

NADIRI, M. I. 1982. Producers Theory. *Chap. 10, pages 431–490 of:* ARROW, K. J., & INTRILIGATOR, M. D. (eds), *Handbook of Mathematical Economics, vol. II*. Amsterdam: North-Holland. ISBN: 0-444-86127-0.

NICHOLSON, W. 1992. *Microeconomic Theory: Basic Principles and Extensions*. 5th edn. The Dryden Press. ISBN 0-03-055043-2.

RITTER, CH., & SIMAR, L. 1997. Pitfalls of Normal-Gamma Stochastic Frontier Models. *Journal of Productivity Analysis*, **8**, 167–182.

SATO, K. 1975. *Production Functions and Aggregation*. 1st edn. Amsterdam: North-Holland. ISBN 0-7204-3197-2.

SCHIFFER, M., & WEDER, B. 2001. *Firm Size and the Business Environment: Worldwide Survey Results*. Discussion paper 43. World Bank/International Finance Corporation. ISBN: 0-8213-5003-X.

SIMAR, L., & WILSON, PAUL W. 2005. *Estimation and Inference in Cross-sectional Stochastic Frontier Models*. Technical report 0541. IAP Statistics Network. [http://www.stat.ucl.ac.be/IAP/].

TAYMAZ, E. 2005. Are Small Firms Really Less Productive? *Small Business Economics*, **25**, 429–445.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN: CORE, CHARLES UNIVERSITY IN PRAGUE: CERGE. 1992. Seminar on Efficiency Analysis: A Project Supported by the "Action for Cooperation in Economics" Program of the Commission of the European Communities. Prague, Czechoslovakia, October 5-10, 1992.

VANDERBEI, R. J. 2001. *Linear Programming: Foundations and Extensions*. Downloaded at [http://www.princeton.edu/~rvdb/LPbook/] on July 26, 2007.

VARIAN, H. R. 2006. *Intermediate Microeconomics: A Modern Approach*. New York: W. W. Norton. ISBN: 0-393-92702-4.

WILSON, PAUL W. 2007. *FEAR: A Package for Frontier Efficiency Analysis with $\mathcal{R}$*. Socio-Economic Planning Sciences, *forthcoming*. Software & documentation available at [http://www.clemson.edu/economics/faculty/wilson/-Software/FEAR/fear.html].