

Průša, Jan

Working Paper

The most efficient Czech SME sectors: An application of robust data envelopment analysis

IES Working Paper, No. 3/2009

Provided in Cooperation with:

Charles University, Institute of Economic Studies (IES)

Suggested Citation: Průša, Jan (2009) : The most efficient Czech SME sectors: An application of robust data envelopment analysis, IES Working Paper, No. 3/2009, Charles University in Prague, Institute of Economic Studies (IES), Prague

This Version is available at:

<https://hdl.handle.net/10419/83408>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

The Most Efficient Czech SME Sectors: An Application of Robust Data Envelopment Analysis

Jan Průša

IES Working Paper: 3/2009



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

[UK FSV – IES]

**Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>**

**Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze**

**Opletalova 26
110 00 Praha 1**

**E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>**

Disclaimer: The IES Working Papers is an online paper series for works by the faculty and students of the Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Czech Republic. The papers are peer reviewed, but they are *not* edited or formatted by the editors. The views expressed in documents served by this site do not reflect the views of the IES or any other Charles University Department. They are the sole property of the respective authors. Additional info at: ies@fsv.cuni.cz

Copyright Notice: Although all documents published by the IES are provided without charge, they are licensed for personal, academic or educational use. All rights are reserved by the authors.

Citations: All references to documents served by this site must be appropriately cited.

Bibliographic information:

Průša, J. (2009). “ The Most Efficient Czech SME Sectors: An Application of Robust Data Envelopment Analysis ” IES Working Paper 3/2009. IES FSV. Charles University.

This paper can be downloaded at: <http://ies.fsv.cuni.cz>

The Most Efficient Czech SME Sectors: An Application of Robust Data Envelopment Analysis

Jan Průša*

*Faculty of Economics,
University of Cambridge and
IES, Charles University Prague
E-mail: jan.prusa@gmail.com

January 2009

Abstract:

This paper analyzes the efficiency of Czech small and 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 employ standard and advanced robust data envelopment analysis (DEA) to obtain cross-sectional rankings of individual industries. The results reveal substantial variance in the efficiency scores, which is only partly removed by the robust DEA specification. We found that the majority of firms operate below full efficiency; with only a few companies (industries) belonging to top performers. Average efficiency lies between 50 to 70 per cent of the best sectors. We conclude that only a minor proportion of Czech SME concentrate on high value added production.

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

JEL: D24, L60, L70

Acknowledgements:

This research was supported by the Grant Agency of Academy of Sciences, grant no. IAA700280803 "Efficiency and employment in the SME sector". I would like to thank to Mr Vladimír Benáček and an anonymous referee for valuable comments. Any remaining errors are of course entirely mine.

1 Introduction

1.1 Aims of the Analysis and Related Literature

In this paper we perform a cross-sectional study on the efficiency of small and medium enterprises (SME). Hence we offer the reader revealing insights into the industrial fundamentals of the Czech economy. We compare the performance of industrial sectors of Czech SME 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 rest of the paper is organized as follows: First we give the reader a basic definition of small and medium enterprises. Next we proceed to the methodology of our analysis. We review data envelopment analysis (DEA), a practice for efficiency measurement which has been commonly used in economic literature. Since lots of modifications were developed over the years, even the comprehensive handbooks (Cooper *et al.* [7], Cooper *et al.* [6], Coelli *et al.* [5]) listed in the bibliography of this paper are far from exhaustive. We focus on two specifications which we find suitable for our data and which are treated in more detail.

Finally section 3 forms the core of our genuine research. We analyze the dataset on Czech small and medium enterprises for the period 2002 to 2005. DEA is used to obtain industry-specific efficiency scores. This allows us to unveil the structural patterns in the Czech SME sector.

To the best of our knowledge, this paper 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.

While macroeconomic phenomena such as the impact of foreign direct investment have received extensive research coverage, literature on small and medium enterprises in the Czech Republic is rather scarce. One pioneering study is by Benáček *et al.* [2], 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 no longer be possible in any European Union country.

1.2 Definition of SME

Small and medium enterprises, abbreviated as *SME*, are defined as companies not exceeding specific size limits. The official definition by the European Union is given in table 1. It is not a clearly disjunctive definition, if related to employment only. The complication emanated from the fact that in the EU SME has become an important tool for economic policy measures. Note that a firm must satisfy the first condition and either one of the last two conditions at the same time in order to be classified as SME. Lots of countries created their own definitions, e.g. Switzerland or the USA chooses 500 employees as the cutoff.

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

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

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

Table 1: Definition of SME according to the EU legislation.

proper analysis.

2 Production Process and Efficiency

2.1 A Model of Production

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

First we deal with technology, which can be captured in two ways: Either as a production function, or in a more general way using the set theory. Following the exposition by Daraio & Simar [8], the production set is defined as all feasible input-output vectors $[\mathbf{x}, \mathbf{y}]$ from the set of nonnegative real numbers $\mathfrak{R}_{0,+}^p$:

$$\Psi = \{[\mathbf{x}, \mathbf{y}] \mid \mathbf{x} \in \mathfrak{R}_{0,+}^p, \mathbf{y} \in \mathfrak{R}_{0,+}^r \text{ is feasible}\}.$$

The output correspondence set consists of all possible output vectors \mathbf{y} which a firm can produce from various possible input vectors \mathbf{x} :

$$P(\mathbf{x}) = \{\mathbf{y} \in \mathfrak{R}_{0,+}^r \mid [\mathbf{x}, \mathbf{y}] \in \Psi\}.$$

Further we consider isoquants of $P(\mathbf{x})$, which correspond to level curves of production functions:

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

For assumption on the technology see Kogiku [10].

2.2 Productive Efficiency

2.2.1 Technical Efficiency

Lovell [12, reading 15] 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 necessarily be strictly convex. For this reason we define the efficient subset of an isoquant as:

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

It is easily recognized that $\text{Isoq}P(\mathbf{x}) = \text{Eff}P(\mathbf{x})$ if and only if the isoquant is strictly convex.

Then a producer will be technically efficient if and only if $\mathbf{y} \in \text{Eff}P(\mathbf{x})$.

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. Therefore we want to include market prices of outputs \mathbf{p} and of inputs \mathbf{w} into our analysis. We consider the easiest case when they are handled as exogenous.

The regularity conditions ensure that maximizing revenue or minimizing cost yield the same maximum profit. Hence the effort to maximize revenue can be described as:

$$R(\mathbf{x}, \mathbf{p}; \boldsymbol{\beta}) = \max_{\mathbf{y}} \{\mathbf{p}^T \mathbf{y} : \mathbf{y} \in P(\mathbf{x}; \boldsymbol{\beta})\},$$

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

2.2.3 Overall Efficiency

Accordingly a producer is called 'efficient' when he reaches the maximum revenue on the efficient subset of the production set:

$$\text{Eff}R(\mathbf{x}, \mathbf{p}; \boldsymbol{\beta}) = \max_{\mathbf{y}} \{\mathbf{p}^T \mathbf{y} : \mathbf{y} \in \text{Eff}P(\mathbf{x}; \boldsymbol{\beta})\}. \quad (1)$$

Note that the derivation could be easily reformulated for cost minimization.

Such analysis contends that prices are an exogenous factor, but in reality 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 paper.

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 these volumes, we can trace the economic part. The ideal statistic would contain all these pieces of information for a large number for individual producers; this is however rarely available.

2.2.4 Sources of Inefficiency

The real world is different from economic models. Due to various frictions, not all producers will always be efficient.²In our paper, instead of modelling a perfect situation, we are trying to track the best practice and estimate the scale of inefficiency. Before we turn to that, let us briefly point out possible explanations for inefficiency.

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. But future development of markets cannot be completely anticipated due to the complexity of interactions, so that suboptimal situations are the outcome.

For similar reasons, different companies have technologies of different vintages at their disposal. This also results in the varying performance of compared firms. Other bottlenecks may stem from inappropriate institutional settings. The more the state interferes in entrepreneurial activities, the higher the risk that something will go wrong.

²The first explicit study on inefficiency other than allocative inefficiency in economic literature was by Leibenstein [11].

Further 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 a different amount of goods is produced on different days due to unexpected failures and complications.

2.3 Data Envelopment Analysis

2.3.1 Basic Model Structure

In this paper we use *data envelopment analysis* (DEA) to analyze technical and economic efficiency. Building a DEA model has two stages: the first stage constructs a plane around the dataset, with points lying on the plane being technically efficient and points within the space being inefficient. The second stage adds prices to the data and identifies economic efficiency.

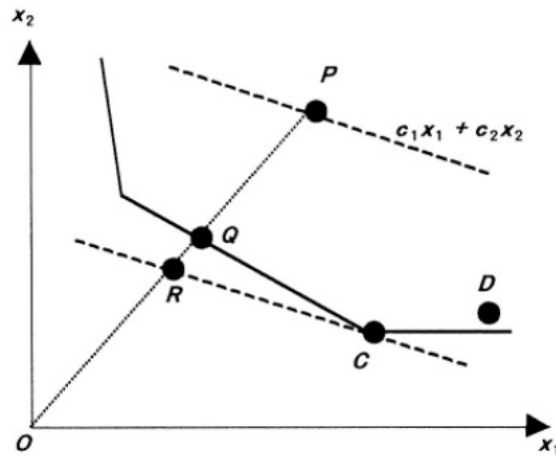


Figure 1: Technical and allocation efficiency, figure from Cooper *et al.* [7, p. 223].

The decomposition of these two stages is shown in figure 1, where x_1, x_2 are two inputs, and letters C, D, P, Q denote firms producing the same level of output \mathbf{y}_0 . The solid line depicts the input requirement set³ for \mathbf{y}_0 . Once estimated, the solid line represents the piecewise linear envelope corresponding to the empirical best practice in technical efficiency. The broken line passing through R and C is the isocost line at the cost level which minimises the cost of producing \mathbf{y}_0 . If we employ the radial measure of efficiency, then technical efficiency of firm P is represented by the ratio $\frac{|OQ|}{|OP|}$ and allocative efficiency by $\frac{|OR|}{|OQ|}$.⁴ Thus overall economic efficiency equals to $\frac{|OQ|}{|OP|} \times \frac{|OR|}{|OQ|} = \frac{|OR|}{|OP|}$.

We start with the first stage and return to the second stage in section 2.3.3. We already listed reference books on DEA in our introduction to this paper. Here we depict the basic model and proceed to a recent robust specification. For technical efficiency we can write a simple input-oriented DEA problem in matrix notation as follows:

$$\begin{aligned}
 & \min_{\lambda, \theta} \theta & (2) \\
 & \text{subject to } \theta \mathbf{x}_i \geq X\lambda \\
 & Y\lambda \geq \mathbf{y}_i \\
 & \lambda = (\lambda_1, \dots, \lambda_n) \geq 0,
 \end{aligned}$$

³ $L(\mathbf{y})$ is the input requirement set mirroring $P(\mathbf{x})$: $L(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}_{0,+}^p \mid (\mathbf{x}, \mathbf{y}) \in \Psi\}$.

⁴One can also use a general measure of distance $d(\cdot)$: $\frac{d(O, Q)}{d(O, P)}$.

which is known as the CCR model, since it was formulated by Charnes, Cooper and Rhodes.⁵

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

2.3.2 Returns to Scale

Model (2) does not impose any additional conditions on λ , 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 constraint $\sum_{i=1}^n \lambda_i = 1$ to the CCR model. Similarly, the specification of $\sum_{i=1}^n \lambda_i \leq 1$ would result in non-increasing returns to scale.

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

2.3.3 Prices and Units Invariance

So far our DEA considerations covered production in terms of physical units. In order for the DEA models to accommodate economic efficiency (which we called the second stage), we want to include prices of inputs w and outputs p , assuming them to be exogenous. The following theorem in Cooper *et al.* [7, p. 24] steers towards one solution.

Theorem 2.1 Units Invariance Theorem. *The optimal values of $\max \theta = \theta_{(x_i, y_i)}^*$ in (2) are independent of the units in which the inputs and outputs are measured, provided these units are the same for every decision making unit (DMU).*

Because we do not have separate data on physical and monetary units, 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'. Therefore we are measuring both technical and allocation efficiency at the same time. In terms of figure 1, we are not able to construct the isocost line. In addition, the solid envelope now represents the efficiency frontier in monetary units, so that the relevant economic efficiency measure is the distance $\frac{|OQ|}{|OP|}$.

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. On the other hand, we forfeit the possibility to decompose inefficiency into its two parts. In our paper this bargain is necessary due to the nature of our data, therefore we will not elaborate further on allocation DEA models.⁶

2.4 Statistical Methods in Non-Parametric Approach

In this section we select one modification of DEA which surmounts two big obstacles of the basic model: (1) deterministic and non-statistical nature; (2) influence of outliers and extreme

⁵Let us discuss the intuition behind this mathematical problem. The vector λ attaches weights to single producers: In the third line, λ selects certain firms, which are called 'reference' producers of the evaluated decision making unit DMU_j . These 'reference' producers, weighed together by λ , produce at least as many outputs as DMU_j . λ then scales the input matrix X to see whether it is possible to cut down inputs at DMU_j by some coefficient θ .

⁶Incorporating 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 amount of outputs produced. For examples of allocation efficiency models, see eg Coelli [4] or Cooper *et al.* [6, section 1].

values (Daraio & Simar [8, p. xviii]).

2.4.1 Probabilistic Production Process

The CCR model from section 2.3 is fully deterministic in that it assumes $\Pr([\mathbf{x}_i, \mathbf{y}_i] \in \Psi) = 1$, where $\Pr(\cdot)$ denotes probability. This time inputs and outputs are a pair of independent and identically distributed (*iid*) multidimensional random variables $(\mathcal{X}, \mathcal{Y})$, although for individual observation it still holds $\Pr([\mathbf{x}_i, \mathbf{y}_i] \in \Psi) = 1$. Following the derivation of Daraio & Simar [9], this yields a joint probability measure characterized by the function

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

For the DMU $[\mathbf{x}, \mathbf{y}]$ 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 know the probability that once the firm produces less, it also uses more inputs. Thus we consider the conditional distribution function

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

where we assume $S_{\mathcal{Y}}(\mathbf{y}) > 0$. This can be empirically estimated by computing

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

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

2.4.2 Order- m Estimator

This estimator was introduced by Cazals *et al.* [3]. The idea is simple: Suppose we have an observation $[\mathbf{x}_0, \mathbf{y}_0]$. As in the CCR model (2), we take observations with larger output. From this set of observations we draw randomly with replacement X_1, \dots, X_m , which is distributed according to $F_{\mathcal{X}|\mathcal{Y}}(\cdot|\mathbf{y})$, as follows from the previous section. We construct the production possibility set as in Daraio & Simar [9, c. f.]:

$$\tilde{\Psi}_m(\mathbf{y}_0) = \left\{ [\mathbf{x}, \mathbf{y}] \in \mathbb{R}_+^{p+r} \mid \mathbf{x} \geq X_i, \mathbf{y} \geq \mathbf{y}_0 \right\}.$$

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

$$\tilde{\theta}_{(\mathbf{x}_0, \mathbf{y}_0)}^m = \inf \{ \theta \mid (\theta \mathbf{x}_0, \mathbf{y}_0) \in \tilde{\Psi}_m(\mathbf{y}_0) \} \quad (3)$$

and take expectations

$$\theta_{(\mathbf{x}_0, \mathbf{y}_0)}^m = E_{\mathcal{X}|\mathcal{Y}}(\tilde{\theta}_{(\mathbf{x}_0, \mathbf{y}_0)}^m | \mathcal{Y} \geq \mathbf{y}). \quad (4)$$

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 a setting.

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

$$\hat{\theta}_{(\mathbf{x}_0, \mathbf{y}_0)}^m = \hat{E}_{\mathcal{X}|\mathcal{Y}}(\tilde{\theta}_{(\mathbf{x}_0, \mathbf{y}_0)}^m | \mathcal{Y} \geq \mathbf{y}) = \int_0^{\infty} (1 - \hat{F}_{\mathcal{X}|\mathcal{Y}}(u\mathbf{x}|\mathbf{y}))^m du.$$

Unfortunately this integration can not be carried out analytically. Cazals *et al.* [3] proposed a four step Monte-Carlo algorithm, which we quote as in Daraio & Simar [8]:

- [1] Draw a sample with replacement among X_i such that $Y_i \geq \mathbf{y}_0$ and denote this sample $(X_{1,b}, \dots, X_{m,b})$.
- [2] Compute $\tilde{\theta}_{(\mathbf{x}_0, \mathbf{y}_0)}^{m,b} = \min_{i=1, \dots, m} \left\{ \max_{j=1, \dots, p} \left(\frac{X_{i,b}^j}{x_i^j} \right) \right\}$.
- [3] Redo [1]-[2] for $b = 1, \dots, B$, where B is large.
- [4] $\hat{\theta}_{n,(\mathbf{x}_0, \mathbf{y}_0)}^m = \frac{1}{B} \sum_{b=1}^B \tilde{\theta}_{(\mathbf{x}_0, \mathbf{y}_0)}^{m,b}$.

2.4.3 Convex order- m frontier

Most of section 2.3 deals with efficiency estimates based on convex technology. The only exception is FDH, briefly mentioned in 2.3.2. Since the order- m frontier is based on FDH, it is not convex. FDH is derived from the approximation of production technology (Daraio & Simar [9]):

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

Daraio & Simar recall that usual DEA scores can be easily obtained from FDH results: It suffices to multiply observed inputs \mathbf{x} by $\hat{\theta}_{(\mathbf{x}, \mathbf{y})}^{FDH}$ and then run the respective linear program on the transformed data, which can be for example equation (2).

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

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

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

$$\begin{aligned} \hat{\theta}_{(\mathbf{x}_i, \mathbf{y}_i)}^{m,C} &= \min_{\lambda, \theta} \theta & (5) \\ \text{subject to } \theta \mathbf{x}_i &\geq \sum_{i=1}^n \lambda_i \hat{\mathbf{x}}_{m,i}^\partial \\ Y \boldsymbol{\lambda} &\geq \mathbf{y}_i \\ \sum_{i=1}^n \lambda_i &= 1 \\ \lambda_1, \dots, \lambda_n &\geq 0. \end{aligned}$$

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

3 Efficiency of Czech SME

3.1 Data Description

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

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

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

1. thirty-item two-digit OKEC⁸ classification, including OKEC codes 10 to 41⁹, ie agriculture and services are not included;
2. size classification with breakdowns at the following number of employees: 0-10-20-50-100-250;
3. eleven economic indicators: output, sales revenue, accounting value added, tangible assets, intangible assets, acquisition of tangible and intangible assets, number of employees, average number of employees, payroll and other personnel expenses;
4. years 2002 through 2005.

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

The usage of the economic indicators deserves several comments. The indicators can be regarded as aggregated accounting figures. Sales revenue tracks all goods and services that the company was able to vend on the market. Output adds goods that were already produced but not yet sold to the sales revenue. Finally, when the cost of materials is subtracted, we get 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 the number of employees. The latter captures the sum of employees at one particular day, which are then recalculated on the basis of days worked to get the former. It follows that the average captures all the fluctuation of employees, which is exactly what we need.

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

3.2 Envelopes I: Standard DEA Results

Units invariance theorem (2.1) 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 λ will sort the data itself. We could indeed plug in the data matrix described in section 3.1 right as it is.

Yet there is a snag: One major drawback of DEA, its sensitivity to outliers, becomes more pronounced with more variables. Therefore 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 (2) with the additional constraint $\sum_{i=1}^n \lambda_i = 1$ introducing variable returns to scale. As the vector of inputs 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 program by T. Coelli [4].

Industries with $\theta_{(x_i, y_i)}^* = 1$ which are thus lying on the efficient frontier are indicated in table 2, the subscript denotes the upper bounds of the respective size subgroup. These points also mostly act as “peers”: points at the edges of the envelope, against which slacks are computed.

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

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

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					
\cap (2002-05)							28 ₉	31 ₉					
3 out of 4 years							13 ₉	16 ₉	28 ₉	31 ₉			

Table 2: Efficient industries according to the BCC model.

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
\cap (2002-05)				35 ₂₅₀						
2 out of 4 years				10 ₂₅₀	17 ₁₉	17 ₄₉	23 ₉	27 ₁₉	35 ₉₉	

Table 3: The least efficient industries according to the CCR model.

Similarly in table 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. We do not observe much overlapping: If a certain industry was top efficient (say 13₉), we do not find the same industry of a different size group among the worst performers (i.e. 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, e.g. in the only year when 16₉ were not fully efficient—namely 2002, it ended up worst.

In table 3 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, where Q stands for quartile. The true maximum of $\theta^*(x_i, y_i)$ is of course always equal to one, nevertheless in this case statistics defines maximum as the upper quartile plus 1.5-times the quartile spread ($3Q - 1Q$). Points above this outside bar (or below the respective bar for minimum) are taken as outliers.

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

Our analysis concentrates on groups of firms defined by size, so we break down our results

	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: Box plot statistics for efficiency scores $\theta_{(x_i, y_i)}^*$.

with respect to number of employees (table 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.

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

Table 5: Mean efficiency score $\theta_{(x_i, y_i)}^*$ according to size group and year.

Proposition 3.1 Preliminary results. *The BCC model unveiled the following:*

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

From this proposition we can 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 analyze the sectoral structure in more depth. For this purpose we will modify the proceeding used in tables 2 and 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.

3.3 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 [13], where both the Monte-Carlo simulation from section 2.4.2 and the solution of equation (5) are available.

First we had to specify the computational aspects: parameters m and B . Cazals *et al.* [3, theorem 2.3] show that as $m \rightarrow \infty$, we have the convergence $\hat{\theta}_{(x_i, y_i)}^m \rightarrow \hat{\theta}_{(x_i, y_i)}^{FDH}$, and similarly $\hat{\theta}_{(x_i, y_i)}^{m, C} \rightarrow \hat{\theta}_{(x_i, y_i)}^*$. With higher m fewer observations will lie above the efficient frontier and the estimator gets less robust. Based on trial and error, we chose $m = 50$ as the level of robustness. With lower numbers of reference observations (eg $m = 20$), there was unusually high ratio

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

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

Table 6: Box plot statistics for efficiency scores $\hat{\theta}_{(x_i, y_i)}^{m, C}$.

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

Recalling Aigner & Chu [1] and their criticism of average production functions, it could seem that we only moved to a certain “average” production plan. Yet histograms which we do not reproduce here disclose that the results are far from resembling normal distribution, because there are two peaks. Moreover the estimates are still skewed to the left, so that while having used the flexible measure, apparently we did not lose large parts of information contained in the data.

Table 6 tracks in more detail the distribution of efficiency scores. When confronted with the initial results in table 5, we conclude that any direct relation between efficiency and size formulated in proposition 3.1 is weakened by the COM model. If we trust COM in that it suppressed the influence of outliers, we may conclude that the strong mean efficiency of the smallest enterprises (as reported in table 5) was a result given by the presence of favourable extreme observations.

As noted in section 3.1, our measure of output is the accounting value added, which is defined as output less cost of materials used in manufacturing.¹⁰ The efficiency estimate there-

¹⁰Output = Sum of: (1) sales revenue from own products, (2) gross profit on merchandise sold (3) received leasing

fore says how much of value added a firm is able to produce from a certain stock of capital and employed labour, and it is normed relative to the best practice. Hence lower efficiency score means less value added per unit of capital-labour.

Proposition 3.2 *Distributional results.*

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

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 7: Best and worst industries according to $\hat{\theta}_{(x_i, y_i)}^{m, C}$.

Let us repeat what we achieved by COM: Due to the small number of observations, we did not leave out extreme points. As a consequence, we smoothed the efficient frontier, but our structural results should not greatly differ from those in section 3.2.

installements, (4) change in inventories and (5) self-constructed asset revenue.

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

In table 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 four we classify as structural leaders and structural 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 3.3 Structural results.

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

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

4 Conclusions

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

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

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

The resulting list of leaders and stragglers as in proposition 3.3 does not suggest any clear-cut outperforming or losing clusters; though we can still identify the chain *metal ores—fabricated metal products—machinery—electrical machinery*. What becomes apparent is that the large scale boom of big factories is not necessarily passed on to SME suppliers—e.g. automotive; coal & lignite; crude petroleum & natural gas.

Moreover we find the majority of firms operate below full efficiency, while only a few industries belong to top performers. Average efficiency lies between 50 to 70 per cent of the best

sectors. In our computations we used value added as a proxy for output. Therefore we derive that only a minor proportion of Czech SME concentrates on high value added production. That is, most industries do not generate as much value added from their stock of capital and labour as the best ones. This result is not very surprising, just as it is not very encouraging.

We recognize that there remains 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. These are left for further research.

References

- [1] D. J. Aigner and S. F. Chu. On estimating the industry production function. *The American Economic Review*, 58:826–839, 1968.
- [2] V. Benáček, D. Shemetilo, and A. Petrov. Restructuring and measurement of efficiency in firms in transition. In S. Sarma, editor, *Restructuring Eastern Europe*, pages 58–72. Cheltenham: Elgar, 1997. ISBN: 1-858-985-765.
- [3] C. Cazals, J.-P. Florens, and L. Simar. Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106:1–25, 2002.
- [4] T. J. Coelli. A guide to deap version 2.1: A data envelopment analysis (computer) program. Working paper 8/96, The University of New England: Centre for Efficiency and Productivity Analysis, 1996. Available at [<http://www.uq.edu.au/economics/cepa/index.htm>]. ISBN: 1-863-89-4969.
- [5] T. J. Coelli, P. Rao, Ch. J. O’Donnell, and G. E. Battese. *An Introduction to Efficiency and Productivity Analysis*. New York: Springer, 2005.
- [6] W. W. Cooper, L. M. Seiford, and J. Zhu, editors. *Handbook on Data Envelopment Analysis*. New York: Kluwer, 2004. eBook ISBN: 1-4020-7798-X.
- [7] W.W. Cooper, L.M. Seiford, and K. Tone. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. New York: Kluwer, 2002. eBook ISBN: 0-306-47541-3.
- [8] C. Daraio and L. Simar. *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. Studies in Productivity and Efficiency. New York: Springer, 2007. ISBN: 0-387-35155-8.
- [9] C. Daraio and L. Simar. Conditional nonparametric frontier models for convex and non-convex technologies: A unifying approach. *Journal of Productivity Analysis*, 28:13–32, 2007.
- [10] K. C. Kogiku. *Microeconomic Models*. New York: Happer & Row, 1971. SBN: 06-043746-4.
- [11] H. Leibenstein. Allocative efficiency vs. “x-efficiency”. *The American Economic Review*, 56:392–415, 1966.
- [12] Charles University in Prague: CERGE Université Catholique de Louvain: CORE. 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. 1992.

- [13] Paul W. Wilson. 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>], 2007.

A Industrial Classification of Economic Activities

The dataset at hand consists of rows for aggregated industries. We list complete Czech OKEC¹¹ definitions for industries that we have available. The following is a transcript of definitions used by the Czech Statistical Office¹² and their English equivalents used by the European Union¹³.

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 (<i>neobsaženo</i>)	Mining of uranium and thorium ores (<i>not included</i>)
13	Těžba a úprava ostatních rud	Mining of metal ores
14	Těžba a úprava ostatních nerostných surovin	Other mining and quarrying
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

Table 8: Selected OKEC/NACE classification.

¹¹Odvětvová klasifikace ekonomických činností.

¹²See [[http://www.czso.cz/csu/klasifik.nsf/i/odvetvova_klasifikace_ekonomickych_cinnosti_\(okec\)](http://www.czso.cz/csu/klasifik.nsf/i/odvetvova_klasifikace_ekonomickych_cinnosti_(okec))].

¹³Nomenclature Générale des Activités Économiques dans les Communautés Européennes, or NACE, see [http://ec.europa.eu/comm/competition/mergers/cases/index/nace_all.html].

¹⁴N.e.c. = Not elsewhere classified.

Code	Czech description	English description
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
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. ¹⁴
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 9: Selected OKEC/NACE classification (continued).

IES Working Paper Series

2008

1. Irena Jindrichovska, Pavel Körner : *Determinants of corporate financing decisions: a survey evidence from Czech firms*
2. Petr Jakubík, Jaroslav Heřmánek : *Stress testing of the Czech banking sector*
3. Adam Geršl : *Performance and financing of the corporate sector: the role of foreign direct investment*
4. Jiří Witzany : *Valuation of Convexity Related Derivatives*
5. Tomáš Richter : *Použití (mikro)ekonomické metodologie při tvorbě a interpretaci soukromého práva*
6. František Turnovec : *Duality of Power in the European Parliament*
7. Natalie Svarciva, Petr Svarc : *Technology adoption and herding behavior in complex social networks*
8. Tomáš Havránek, Zuzana Iršová : *Intra-Industry Spillovers from Inward FDI: A Meta-Regression Analysis*
9. Libor Dušek, Juraj Kopecsni : *Policy Risk in Action: Pension Reforms and Social Security Wealth in Hungary, Czech Republic, and Slovakia*
10. Alexandr Kuchynka : *Volatility extraction using the Kalman filter*
11. Petr Kadeřábek, Aleš Slabý, Josef Vodička : *Stress Testing of Probability of Default of Individuals*
12. Karel Janda : *Which Government Interventions Are Good in Alleviating Credit Market Failures?*
13. Pavel Štika : *Možnosti analytického uchopení reciprocity v sociálních interakcích*
14. Michal Bauer, Julie Chytilová: *A Model of Human Capital, Time Discounting and Economic Growth*
15. Milan Rippel, Petr Teplý : *Operational Risk – Scenario Analysis*
16. Martin Gregor : *The Strategic Euro Laggards*
17. Radovan Chalupka, Petr Teplý : *Operational Risk Management and Implications for Bank's Economic Capital – a Case Study*
18. Vít Bubák : *Value-at-Risk on Central and Eastern European Stock Markets: An Empirical Investigation Using GARCH Models*
19. Petr Jakubík, Petr Teplý : *The Prediction of Corporate Bankruptcy and Czech Economy's Financial Stability through Logit Analysis*

20. Elisa Gaelotti : *Do domestic firms benefit from geographic proximity with FDI? Evidence from the privatization of the Czech glass industry*
21. Roman Horváth, Marek Rusnák : *How Important Are Foreign Shocks in Small Open Economy? The Case of Slovakia*
22. Ondřej Schneider : *Voting in the European Union - Central Europe's lost voice*
23. Fabricio Coricelli, Roman Horváth : *Price Setting and Market Structure: An Empirical Analysis of Micro Data*
24. Roman Horváth, Kamila Koprnická : *Inflation Differentials in EU New Member States: An Empirical Evidence*
25. Michal Franta, Branislav Saxa, Kateřina Šmídková : *Inflation Persistence: Is It Similar in the New EU Member States and the Euro Area Members?*
26. Jakub Seidler : *Implied Market Loss Given Default: structural-model approach*
27. Radovan Chalupka, Juraj Kopecsni : *Modelling Bank Loan LGD of Corporate and SME Segments: A Case Study*
28. Michal Bauer, Julie Chytilová, Jonathan Morduch: *Behavioral Foundations of Microcredit: Experimental and Survey Evidence From Rural India*
29. Jiří Hlaváček, Michal Hlaváček : *Mikroekonomické modely trhu s externalitami, zobrazení Coaseho teorém*
30. Václav Hausenblas, Petr Švarc : *Evoluční dynamika vězňova dilematu: Vliv topologie interakcí a imitace na vývoj kooperativního chování*
31. Peter Marko, Petr Švarc : *Firms formation and growth in the model with heterogeneous agents and monitoring*
32. Jan Průša : *Productivity of Czech Small and Medium Enterprises: Lagging Behind Their Potential*
33. Kamila Fialová : *Wage Differentiation and Unemployment in the Districts of the Czech Republic*
34. Ibrahim L. Awad : *Switching to the Inflation Targeting Regime: Does it necessary for the case of Egypt?*

2009

1. František Turnovec : *Fairness and Squareness: Fair Decision Making Rules in the EU Council?*
2. Radovan Chalupka : *Improving Risk Adjustment in the Czech Republic*
3. Jan Průša : *The Most Efficient Czech SME Sectors: An Application of Robust Data Envelopment Analysis*

All papers can be downloaded at: <http://ies.fsv.cuni.cz>

