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

Jan Průša †

October 3, 2010

Abstract

This paper analyses 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 classification: D24, L60, L70.

1 Introduction

1.1 Aims of the Analysis and Related Literature

Small and medium enterprises (hereinafter SMEs) form a vital part of developed economies,
as has been stressed in a growing body of literature, see e.g. Schiffer & Weder [20], Ayygari
et al. [3], Acs et al. (eds.) [1], Taymaz [21], Yang & Chen [23]. Research on Czech enterprises
stressed especially institutional factors related to transition from a centrally planned economy
to capitalism, such as the role of foreign direct investment (FDI) and institutions (examples
include Djankov & Hoekman [12] and Marcinéin & Wijnbergen [16]). However literature on
small and medium enterprises in the Czech Republic is rather scarce.

---

*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.

†Institute of Economic Studies, Charles University in Prague, Czech Republic. E-mail: jan.prusa@ies-prague.org.
To the best of our knowledge, this paper is the first attempt to measure economic efficiency of Czech SMEs based on microeconomic principles using data envelopment analysis (hereinafter DEA). Our text therefore complements previous results which mostly relied on macroeconomic methods. The study by Benáček et al. [4] is an exception where the authors measured efficiency of textile and clothing firms by distance functions. Thanks to detailed information on individual firms, Benáček et al. were even capable of separating technical and allocation efficiency.

In a previous study Pruša [17] provided general characteristics of the production process among Czech small companies. This paper will by contrast closely explore structural characteristics of SME, in that we will perform a cross-sectional study of SME statistics. This way we offer the reader revealing insights into the industrial fundamentals of the Czech economy. Specifically our model answers the following questions:

1. How dispersed is the efficiency of individual sectors? Do most firms operate close to the efficiency frontier or away from it?
2. Which are the most efficient industries?
3. Are industries which are more concentrated and/or more regulated also more profitable?
4. Does FDI support higher efficiency of the related sectors?
5. Are larger firms more efficient?

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

The rest of the paper is organized as follows: First we give the reader basic definition of SMEs. Next we proceed to the methodology of our analysis. We review data envelopment analysis (DEA), a practice for efficiency measurement which is commonly used in economic literature. Since lots of modifications were developed over the years, even the comprehensive handbooks (Cooper et al. [9], Cooper et al. [8], Coelli et al. [6]) 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 analyse 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 structural patterns within Czech SME industrial sectors.

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.

<table>
<thead>
<tr>
<th>Enterprise Category</th>
<th>Headcount</th>
<th>Turnover</th>
<th>Balance Sheet Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro</td>
<td>&lt; 10</td>
<td>( \leq €2) million</td>
<td>( \leq €2) million</td>
</tr>
<tr>
<td>Small</td>
<td>&lt; 50</td>
<td>( \leq €10) million</td>
<td>( \leq €10) million</td>
</tr>
<tr>
<td>Medium-sized</td>
<td>&lt; 250</td>
<td>( \leq €50) million</td>
<td>( \leq €43) million</td>
</tr>
</tbody>
</table>

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

Switzerland or the USA chooses 500 employees as the cutoff.

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

2 Measurement of Efficiency

2.1 The Concept of Efficiency

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

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

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

Accordingly explanations of efficiency emerged as separate (though not always isolated) theories. It is not the purpose of this study to present them thoroughly, nevertheless we list here the main contributions in this field:

1. **Vintage models**—It is assumed that although aggregate technology is available to all producers, it is evolving over time and thus different producers at different times of investment acquire different vintages of technology. This implies heterogeneity of production capabilities. Before investment is made, the production set (defined later) is the same for all producers: \( (Y_i|\beta_i) \), \( \beta_i \) being the vector of parameters which characterize the technology. After investment is made, each producer has his own specific production capabilities: \( (Y_i|\beta_i) \). See e.g. Johansen [13].

2. **Institutional economics**—assumes frictions stemming from transactions and other ‘soft’ costs. Inefficiencies may 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. 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.

3. **Austrian economics**—concentrate on entrepreneurs as discoverers of market opportunities. In this view, the economy is always developing and never achieves static equilibrium. See e.g. Sautet [19].

4. **X-efficiency** developed by Leibenstein [15].

### 2.2 The Plain Vanilla Model of Efficiency

#### 2.2.1 Technical Efficiency

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

Following the exposition by Daraio & Simar [10], the production set \( \mathcal{Y} \) is defined as all feasible input-output vectors \( (x, y) \) from the set of nonnegative real numbers \( \mathbb{R}_{0,+}^r \times \mathbb{R}_{0,+}^s \):

\[
\mathcal{Y} = \left\{ (x, y) : x \in \mathbb{R}_{0,+}^r, y \in \mathbb{R}_{0,+}^s \mid (x, y) \text{ is feasible} \right\}.
\]

\(^2\)For detailed discussion on standard assumptions on technology see e.g. Kogiku [14].
We can further define the technologically efficient production frontier \( \text{Eff}(\mathcal{Y}) \):

\[
\text{Eff}(\mathcal{Y}) = \{ (x, y) \in \mathcal{Y} | \forall [x^1 \leq x, y^1 \geq y, (x^1, y^1) \neq (x, y)] : (x^1, y^1) \not\in \mathcal{Y} \}.
\]

Then a producer will be technically efficient if and only if they operate on \( \text{Eff}(\mathcal{Y}) \).

\[3\]

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 \( p \) and of inputs \( w \) into our analysis. In the simplest neoclassical case of perfect competition, prices are assumed to be exogenous, so that the profit function can be derived.

**Definition 1** A profit function \( \Pi(\cdot) \) is a general solution to the profit maximisation problem:

\[
\Pi(p, w) = \arg \max_{(x, y)} \{ p'y - w'x | (x, y) \in \mathcal{Y} \}.
\]

This is by a contradiction argument equivalent to:

\[
\Pi(p, w) = \arg \max_{(x, y)} \{ p'y - w'x | (x, y) \in \text{Eff}(\mathcal{Y}) \}.
\]

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

2.3 Measuring Efficiency in Monetary Units

The separation of the two components of efficiency poses the main snag for any efficiency measurement. The technical part is captured in data in physical units. If we assign certain prices to these volumes, we can trace the economic part. The ideal statistic would contain all these pieces of information for a large number for individual producers; this is however rarely available (and in most situations even not sensible).

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

Thirdly, we can define a framework which explicitly allows for price exogeneity and product heterogeneity. Průša [18, section 4] suggested to use money-metric production frontiers, where

\[3\] As we have seen, assuming \( \beta \) away is equivalent to saying that all firms with the same products use the same transformation of inputs. This would be the case of perfect competitions where producers are identical (in terms of technology), or in the long run when all producers can adopt the most efficient technology. In the short run however, which will be the framework for our data analysis, differences in \( \beta \) will be one explanatory factor of inefficiency.
definitions in equations (1) and (2) are in monetary units (see Průša [18, equation 2]). In other words, equation (2) tracks the ‘profit frontier’, meaning that the impact of imperfect competition and product heterogeneity is already incorporated in money-denominated datapoints.

Money-metric efficiency frontiers trade in separation of technical and allocation efficiency for clear economic interpretation. The resulting efficiency score directly captures overall economic efficiency. It must be stressed that the first and the third options are computationally equivalent—since we plug in the data we have. However, it seems more straightforward to assume price endogeneity, especially with cross-sectional data. Therefore, in the following sections, we assume the third approach: Input vectors $x$ and output vectors $y$ denote data in monetary units, unless otherwise stated.

2.4 Data Envelopment Analysis

2.4.1 Basic Model Structure

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

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

$$\min_{\lambda, \theta} \theta$$

subject to

$$\theta x_i \geq X\lambda$$

$$Y\lambda \geq y_i$$

$$\lambda = (\lambda_1, \ldots, \lambda_n) \geq 0,$$

which is known as the CCR model, since it was formulated by Charnes, Cooper and Rhodes.\(^4\)

The problem must be solved $n$ times for all producers to obtain each firm’s efficiency score, which is an estimate $\theta^*_i(x_i, y_i) \in [0, 1].$\(^5\)

2.4.2 Returns to Scale

Model (3) does not impose any additional conditions on $\lambda$, so that technical efficiency is computed under the assumption of constant returns to scale. Variable returns to scale (RTS) were

---

\(^4\)Let us discuss the intuition behind this mathematical problem. The vector $\lambda$ attaches weights to single producers: In the third line, $\lambda$ selects certain firms, which are called ‘reference’ producers of the evaluated decision making unit $DMU_i$. These ‘reference’ producers, weighed together by $\lambda$, produce at least as many outputs as $DMU_i$. $\lambda$ then scales the input matrix $X$ to see whether it is possible to cut down inputs at $DMU_i$ by some coefficient $\theta$.

\(^5\)Instead of assuming data in monetary units, prices can be incorporated into DEA by assigning value to the objective function, leaving constraints unchanged. This requires strong assumptions, above all that prices remain constant for any amount of inputs consumed and any amount of outputs produced. For examples of allocation efficiency models, see eg Coelli [7] or Cooper et al. [8, section 1].
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 \( \sum_{i=1}^{n} \lambda_i = 1 \land (\forall i : \lambda_i \in \{0, 1\}) \), 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.

### 2.5 Statistical Methods in Non-Parametric Approach

In this section we select one modification of DEA which surmounts two big obstacles of the basic model: (1) deterministic and non-statistical nature; (2) influence of outliers and extreme values (Daraio & Simar [10, p. xviii]).

#### 2.5.1 Probabilistic Production Process

The CCR model from section 2.4 is fully deterministic in that it assumes \( \Pr ([x_i, y_i] \in \mathcal{Y}) = 1 \), where \( \Pr(\cdot) \) denotes probability. This time inputs and outputs are a pair of independent and identically distributed (i.i.d) multidimensional random variables \((X, Y)\), although for individual observation it still holds \( \Pr ( [x_i, y_i] ) \in \mathcal{Y} ) = 1 \). Following the derivation of Daraio & Simar [11], this yields a joint probability measure characterized by the function

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

For the DMU \([x, y]\) it captures the probability that this firm will perform worse than others, i.e. 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

\[
F_{X|Y}(x|y) = \Pr(X \leq x|Y \geq y) = \frac{\Pr(X \leq x, Y \geq y)}{\Pr(Y \geq y)} = \frac{H_{X,Y}(x, y)}{S_Y(y)},
\]

where we assume \( S_Y(y) > 0 \). This can be empirically estimated by computing

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

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

#### 2.5.2 Order-\( m \) Estimator

This estimator was introduced by Cazals et al. [5]. The idea is simple: Suppose we have an observation \([x_0, y_0]\). As in the CCR model (3), we take observations with larger output. From this set of observations we draw randomly with replacement \( X_1, \ldots, X_m \), which is distributed
according to $F_{X|Y}(\cdot | y)$, as follows from the previous section. We construct the production possibility set as in Daraio & Simar [11, c. f.]

$$\mathcal{Y}_m(y_0) = \left\{ [x, y] \in \mathbb{R}^{p+r}_+ | x \geq X_i, y \geq y_0 \right\}.$$  

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

$$\hat{\theta}_m(x_0, y_0) = \inf \left\{ \theta | (\theta x_0, y_0) \in \mathcal{Y}_m(y_0) \right\}$$  

and take expectations

$$\hat{\theta}^m(x_0, y_0) = E_{X|Y}(\hat{\theta}_m(x_0, y_0) | Y \geq y).$$  

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 $\hat{F}_{X|Y}$, the score can be estimated as:

$$\hat{\theta}_m(x_0, y_0) = \hat{F}_{X|Y}(\hat{\theta}_m(x_0, y_0) | Y \geq y) = \int_0^\infty \left( 1 - \hat{F}_{X|Y}(ux | y) \right)^m du.$$  

Unfortunately this integration can not be carried out analytically. Cazals et al. [5] proposed a four step Monte-Carlo algorithm, which we quote as in Daraio & Simar [10]:

1. Draw a sample with replacement among $X_i$ such that $Y_i \geq y_0$ and denote this sample $(X_{1,b}, \ldots, X_{m,b})$.
2. Compute $\hat{\theta}_m^m(x_0, y_0) = \min_{i=1, \ldots, m} \left\{ \max_{j=1, \ldots, p} \left( \frac{x_j}{x_i} \right) \right\}$.
3. Redo [1]-[2] for $b = 1, \ldots, B$, where $B$ is large.
4. $\hat{\theta}^m_n(x_0, y_0) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_m^m(x_0, y_0)$.

### 2.5.3 Convex order-$m$ frontier

Most of section 2.4 deals with efficiency estimates based on convex technology. The only exeption is FDH, briefly mentioned in 2.4.2. Since the order-$m$ frontier is based on FDH, it is not convex. FDH is derived from the approximation of production technology (Daraio & Simar [11]):

$$\mathcal{Y}_{FDH} = \left\{ [x, y] \in \mathbb{R}^{p+r}_+ | x \geq x_i, y \leq y_i, i = 1, \ldots, n \right\},$$

$$\hat{\theta}_{FDH}^m(x_0, y_0) = \inf \left\{ \theta | (\theta x_0, y_0) \in \mathcal{Y}_{FDH} \right\}.$$  

Daraio & Simar recall that usual DEA scores can be easily obtained from FDH results: It suffices to multiply observed inputs $x$ by $\hat{\theta}_{FDH}^m(x, y)$ and then run the respective linear program on the transformed data, which can be for example equation (3).
They use this feature to convexify the order-\(m\) estimate in the same way. They construct transformed data by

\[
\hat{x}_{m,i}^{\theta} = \hat{\theta}_{m,y_i}^{\theta} \cdot x_i
\]

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

\[
\hat{\theta}_{m,C}^{\theta,x,y} = \min_{\lambda,\theta} \theta \\
\text{subject to } \theta x_i \geq \sum_{i=1}^{n} \lambda_i \hat{x}_{m,i}^{\theta} \\
Y \lambda \geq y_i \\
\sum_{i=1}^{n} \lambda_i = 1 \\
\lambda_1, \ldots, \lambda_n \geq 0.
\] (6)

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.\(^6\)

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

1. thirty-item two-digit OKEC\(^7\) classification, including OKEC codes 10 to 41\(^8\), i.e. agriculture and services are not included;
2. size classification with breakdowns at the following number of employees: 0-10-20-50-100-250;
3. eleven economic indicators: output, sales revenue, accounting value added, tangible assets, intangible assets, acquisition of tangible and intangible assets, number of employees, average number of employees, payroll and other personnel expenses;

\(^6\)The publication can be found under reference number 8007-[xx], where xx are the last two digits of the corresponding year. The 2008 version is available at: [http://www.czso.cz/csu/2008ediciplan.nsf/p/8007-08].

\(^7\)European Union uses the abbreviation NACE: Nomenclature Générale des Activités Economiques dans les Communautés Européennes.

\(^8\)OKEC 12 is not included. Full list of industries is available at [http://www.czso.cz/csu/klasifik.nsf/i/-odvetvova_klasifikace_ekonomickych_cinnosti_(okec)] in Czech or at [http://ec.europa.eu/comm/competition/-mergers/cases/index/nace_all.html] in English.
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.

3.2 Model Specification

The usage of the economic indicators deserves several comments. The indicators can be regarded as aggregated accounting figures. Sales revenue tracks all goods and services that the company was able to sell on the market. Output adds goods that were already produced but not yet sold to the sales revenue. Finally, when the cost of materials is subtracted, we get 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.

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. Therefore, as the vector of inputs we specify:

$$x = \text{assets, investment, employees and wages}.$$  

‘Assets’ are totalled tangible and intangible assets; ‘wages’ are wage outlays plus other personal expenses. ‘Investment’ is a forward-looking variable, but we assume this to be a good proxy for depreciation, which in itself is a plausible measure of the real cost of capital to the firm. ‘Employees’ is the average number of employees, the single non-monetary input.\(^9\) Output is represented by accounting value added

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

3.3 Envelopes I: Standard DEA Results

Consider the BCC model, i.e. equation (3) with the additional constraint $\sum_{i=1}^n \lambda_i = 1$ introducing variable returns to scale. We implemented this computation for each year separately via DEAP, a freely available program by Coelli [7].

\(^9\)Wage outlays are highly correlated with the number of employees. As was pointed to us by a referee, in an econometric setting this would lead to multicollinearity and would have to be accounted for. However this with DEA this issue does not cause any problems.
To get an overview of the distribution of efficiency, we computed box plot statistics given in table 2, where $Q$ stands for quartile. The true maximum of $\theta^*_i(x, y)$ 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.

<table>
<thead>
<tr>
<th></th>
<th>min</th>
<th>1Q</th>
<th>median</th>
<th>3Q</th>
<th>max</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>0.1500</td>
<td>0.4155</td>
<td>0.4910</td>
<td>0.6290</td>
<td>0.9410</td>
<td>0.5534</td>
</tr>
<tr>
<td>2003</td>
<td>0.020</td>
<td>0.370</td>
<td>0.498</td>
<td>0.691</td>
<td>1.000</td>
<td>0.5279</td>
</tr>
<tr>
<td>2004</td>
<td>0.031</td>
<td>0.064</td>
<td>0.133</td>
<td>0.299</td>
<td>0.604</td>
<td>0.2282</td>
</tr>
<tr>
<td>2005</td>
<td>0.0420</td>
<td>0.0995</td>
<td>0.1660</td>
<td>0.3630</td>
<td>0.6690</td>
<td>0.2743</td>
</tr>
</tbody>
</table>

Table 2: Box plot statistics for efficiency scores $\theta^*_i(x, y)$.

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

Our analysis concentrates on groups of firms defined by size, so we break down our results with respect to number of employees (table 3). 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.

<table>
<thead>
<tr>
<th># of employees</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10</td>
<td>0.754</td>
<td>0.629</td>
<td>0.358</td>
<td>0.390</td>
</tr>
<tr>
<td>10-19</td>
<td>0.482</td>
<td>0.496</td>
<td>0.115</td>
<td>0.142</td>
</tr>
<tr>
<td>20-49</td>
<td>0.485</td>
<td>0.486</td>
<td>0.169</td>
<td>0.253</td>
</tr>
<tr>
<td>50-99</td>
<td>0.485</td>
<td>0.478</td>
<td>0.209</td>
<td>0.264</td>
</tr>
<tr>
<td>100-250</td>
<td>0.541</td>
<td>0.540</td>
<td>0.268</td>
<td>0.311</td>
</tr>
</tbody>
</table>

Table 3: Mean efficiency score $\theta^*_i(x, y)$ according to size group and year.

**Proposition 1** Preliminary results. The BCC model unveiled the following:

- Distribution of efficiency results is heavily skewed to lower scores. It seems that there are outliers which exercise considerable influence on overall results.

- Larger firms tend to be more efficient on average, with one surprising exception: The smallest entrepreneurs rank first in every observed year.
From this proposition we can deduce what to do next. Firstly, we will apply a statistically based DEA model in order to control for significant outliers. With refined results at hand, we will observe what the impact on efficiency distribution and its skewness will be, if any.

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

3.4 Envelopes II: Robust DEA Results

In this section we report results of the convex order-$m$ estimator (COM). We obtained the scores thanks to the package FEAR by Paul Wilson [22], where both the Monte-Carlo simulation from section 2.5.2 and the solution of equation (6) are available.

First we had to specify the computational aspects: parameters $m$ and $B$. Cazals et al. [5, theorem 2.3] show that as $m \to \infty$, we have the convergence $\hat{\theta}_m^{(x,y)}(\theta_{\text{FDH}}^{(x,y)})$, and similarly $\hat{\theta}_m^{C,(x,y)}(\theta_{\text{FDH}}^{(x,y)})$. With higher $m$ fewer observations will lie above the efficient frontier and the estimator gets less robust. Based on our analysis, we chose $m = 50$ (i.e. ≃ 10% of observations) 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.

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

Table 4 tracks in more detail the distribution of efficiency scores. When confronted with the initial results in table 3, we conclude that any direct relation between efficiency and size formulated in proposition 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 3) was a result given by the presence of favourable
Table 4: Box plot statistics for efficiency scores $\hat{\theta}_{m,C}^{\hat{m},C}(x,y)$.

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

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

**Proposition 2 Distributional results.**

- Although the robust specification of DEA mitigated the skewness caused by outliers, variation of efficiency scores remains high.

---

10These in turn may have been caused by favourable sample selection.

11Output = Sum of: (1) sales revenue from own products, (2) gross profit on merchandise sold (3) received leasing installements, (4) change in inventories and (5) self-constructed asset revenue.

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

<table>
<thead>
<tr>
<th>Best industries</th>
<th>Worst industries</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002 2003 2004 2005</td>
<td>2002 2003 2004 2005</td>
</tr>
<tr>
<td>11,250 11,19 10,49</td>
<td>10,49 10,19 10,9 10,6 10,250</td>
</tr>
<tr>
<td>13,9 13,9 11,19 14,250</td>
<td>10,9 10,99 10,250 10,19 11,99</td>
</tr>
<tr>
<td>14,250 14,19 13,9 15,49</td>
<td>10,250 10,250 14,9 10,49 19,9</td>
</tr>
<tr>
<td>15,99 15,250 15,250 15,49</td>
<td>15,250 15,250 15,250 15,250 15,250</td>
</tr>
<tr>
<td>16,9 16,99 18,49 16,9</td>
<td>14,19 19,9 19,9 11,250 27,9</td>
</tr>
<tr>
<td>18,9 18,19 18,250 16,49</td>
<td>16,49 21,9 20,19 13,9 30,250</td>
</tr>
<tr>
<td>19,9 18,19 20,9 17,49</td>
<td>16,250 22,19 20,99 16,99 34,9</td>
</tr>
<tr>
<td>19,9 18,49 22,9 20,9</td>
<td>16,250 22,19 20,99 16,99 34,9</td>
</tr>
<tr>
<td>20,9 18,250 23,99 21,99 40,250</td>
<td>21,9 21,9 22,19 19,9 37,19</td>
</tr>
<tr>
<td>22,9 19,250 24,9 22,9</td>
<td>23,9 25,19 22,19 20,99 41,19</td>
</tr>
<tr>
<td>25,49 20,9 25,19 23,9</td>
<td>24,9 26,19 23,9 21,9 41,99</td>
</tr>
<tr>
<td>26,250 23,19 25,99 25,19</td>
<td>27,9 27,19 24,49 23,9</td>
</tr>
<tr>
<td>28,9 23,49 26,250 26,19</td>
<td>27,250 27,99 24,99 23,19</td>
</tr>
<tr>
<td>28,49 28,9 27,99 27,99</td>
<td>28,19 30,19 27,9 23,49</td>
</tr>
<tr>
<td>29,9 28,250 28,9 28,9</td>
<td>30,49 30,49 27,19 27,9</td>
</tr>
<tr>
<td>29,99 29,9 28,250 28,250</td>
<td>30,250 30,250 30,250 27,19</td>
</tr>
<tr>
<td>29,250 29,19 29,250 29,250</td>
<td>32,49 34,9 31,19 30,49</td>
</tr>
<tr>
<td>30,19 29,250 31,9 31,9</td>
<td>34,9 34,99 34,9 30,99</td>
</tr>
<tr>
<td>31,9 31,9 32,9 32,250</td>
<td>34,19 35,99 34,19 30,250</td>
</tr>
<tr>
<td>32,9 31,250 32,250 34,9</td>
<td>34,49 37,19 35,19 33,250</td>
</tr>
<tr>
<td>33,9 32,250 33,49 35,250</td>
<td>35,250 37,49 37,19 34,19</td>
</tr>
<tr>
<td>35,9 40,9 33,250 36,19</td>
<td>37,19 37,250 41,9 35,19</td>
</tr>
<tr>
<td>36,9 40,49 34,99 40,49</td>
<td>41,19 41,19 41,19 41,19</td>
</tr>
<tr>
<td>40,250 40,250 40,9 40,250</td>
<td>41,99 41,49 41,99 41,99</td>
</tr>
</tbody>
</table>

\( \cap \) indicates that the industry was among the best/worst in at least three years.

Table 5: Best and worst industries according to \( \hat{\theta}_{m,C}^{(x,y)} \).

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 3.3.

In table 5, 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 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 (e.g. due to FDI) are not always passed on to suppliers among SME.

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

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

Moreover we find that the majority of 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


### A Standard Industrial Classification

The dataset at hand consists of rows for aggregated industries. We list complete Czech OKEC\(^\text{12}\) definitions for industries that we have available. The following is a transcript of definitions used by the Czech Statistical Office\(^\text{13}\) and their English equivalents used by the European Union\(^\text{14}\).

---

\(^{12}\)Odvětvová klasifikace ekonomických činností.

\(^{13}\)See [http://www.czso.cz/cs/klasifik.nsf/i/odvetvova_klasifikace_ekonomickych_cinnosti_(okec)].

\(^{14}\)Nomenclature Générale des Activités Économiques dans les Communautés Européennes, or NACE, see [http://ec.europa.eu/comm/competition/mergers/cases/index/nace_all.html].

\(^{15}\)N.e.c. = Not elsewhere classified.
<table>
<thead>
<tr>
<th>Code</th>
<th>Czech description</th>
<th>English description</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Těžba uhlí, lignitu a rašeliny</td>
<td>Mining of coal and lignite; extraction of peat</td>
</tr>
<tr>
<td>11</td>
<td>Těžba ropy, zemního plynu a související činnosti kromě průzkumných vrtů</td>
<td>Extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction, excluding surveying</td>
</tr>
<tr>
<td>12</td>
<td>Těžba a úprava uranových a thoriových rud (<em>neobsazeno</em>)</td>
<td>Mining of uranium and thorium ores (<em>not included</em>)</td>
</tr>
<tr>
<td>13</td>
<td>Těžba a úprava ostatních rud</td>
<td>Mining of metal ores</td>
</tr>
<tr>
<td>14</td>
<td>Těžba a úprava ostatních nerostných surovin</td>
<td>Other mining and quarrying</td>
</tr>
<tr>
<td>15</td>
<td>Výroba potravinářských výrobků a nápojů</td>
<td>Manufacture of food products and beverages</td>
</tr>
<tr>
<td>16</td>
<td>Výroba tabákových výrobků</td>
<td>Manufacture of tobacco products</td>
</tr>
<tr>
<td>17</td>
<td>Výroba textilií a textilních výrobků</td>
<td>Manufacture of textiles</td>
</tr>
<tr>
<td>18</td>
<td>Výroba odevů, zpracování a barvení kožešin</td>
<td>Manufacture of wearing apparel; dressing and dyeing of fur</td>
</tr>
<tr>
<td>19</td>
<td>Činění a úprava usní, výroba brašnářských a sedlářských výrobků a obuvi</td>
<td>Tanning and dressing of fur; manufacture of luggage, handbags, saddlery, harness and footwear</td>
</tr>
<tr>
<td>20</td>
<td>Zpracování dřeva, výroba dřevařských, korkových, proutěných a slaměných výrobků kromě nábytku</td>
<td>Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials</td>
</tr>
<tr>
<td>21</td>
<td>Výroba vlákliny, papíru a výrobků z papíru</td>
<td>Manufacture of pulp, paper and paper products</td>
</tr>
<tr>
<td>22</td>
<td>Vydavatelství, tisk a rozmnožování nahraných nosičů</td>
<td>Publishing, printing and reproduction of recorded media</td>
</tr>
<tr>
<td>23</td>
<td>Výroba koksu, jaderných paliv, rafinérské zpracování ropy</td>
<td>Manufacture of coke, refined petroleum products and nuclear fuel</td>
</tr>
<tr>
<td>24</td>
<td>Výroba chemických látek, přípravků, léčiv a chemických vláken</td>
<td>Manufacture of chemicals and chemical products</td>
</tr>
<tr>
<td>25</td>
<td>Výroba pryžových a plastových výrobků</td>
<td>Manufacture of rubber and plastic products</td>
</tr>
</tbody>
</table>

Table 6: Selected OKEC/NACE classification.
<table>
<thead>
<tr>
<th>Code</th>
<th>Czech description</th>
<th>English description</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>Výroba ostatních nekovových minerálních výrobků</td>
<td>Manufacture of other non-metallic mineral products</td>
</tr>
<tr>
<td>27</td>
<td>Výroba základních kovů a hutních výrobků</td>
<td>Manufacture of basic metals</td>
</tr>
<tr>
<td>28</td>
<td>Výroba kovových konstrukcí a kovodělných výrobků (kromě strojů a zařízení)</td>
<td>Manufacture of fabricated metal products, except machinery and equipment</td>
</tr>
<tr>
<td>29</td>
<td>Výroba a opravy strojů a zařízení j. n.</td>
<td>Manufacture of machinery and equipment n.e.c.</td>
</tr>
<tr>
<td>30</td>
<td>Výroba kancelářských strojů a počítačů</td>
<td>Manufacture of office machinery and computers</td>
</tr>
<tr>
<td>31</td>
<td>Výroba elektrických strojů a zařízení j. n.</td>
<td>Manufacture of electrical machinery and apparatus n.e.c.</td>
</tr>
<tr>
<td>32</td>
<td>Výroba rádiových, televizních a spojových zařízení a přístrojů</td>
<td>Manufacture of radio, television and communication equipment and apparatus</td>
</tr>
<tr>
<td>33</td>
<td>Výroba zdravotnických, přesných optických a časoměrných přístojů</td>
<td>Manufacture of medical, precision and optical instruments, watches and clocks</td>
</tr>
<tr>
<td>34</td>
<td>Výroba motorových vozidel (kromě motocyklů), výroba přívěsů a návěsů</td>
<td>Manufacture of motor vehicles, trailers and semi-trailers</td>
</tr>
<tr>
<td>35</td>
<td>Výroba ostatních dopravních prostředků a zařízení</td>
<td>Manufacture of other transport equipment</td>
</tr>
<tr>
<td>36</td>
<td>Výroba nábytku; zpracovatelský průmysl j. n.</td>
<td>Manufacture of furniture; manufacturing n.e.c.</td>
</tr>
<tr>
<td>37</td>
<td>Recyklace druhotných surovin</td>
<td>Recycling</td>
</tr>
<tr>
<td>40</td>
<td>Výroba a rozvod elektriny, plynu a tepelné energie</td>
<td>Electricity, gas, steam and hot water supply</td>
</tr>
<tr>
<td>41</td>
<td>Shromažďování, úprava a rozvod vody</td>
<td>Collection, purification and distribution of water</td>
</tr>
</tbody>
</table>

Table 7: Selected OKEC/NACE classification (continued).