JEL Classification: D24, L60, L70
Keywords: production, efficiency measurement, data envelopment analysis, small and medium-sized enterprises

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

Jan PRŮŠA – Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague (jan.prusa@ies-prague.org)

Abstract
This paper analyzes the efficiency of Czech small and medium-sized enterprises (SMEs). The main focus is on structural analysis of Czech SMEs in manufacturing based on their efficiency. We use sectoral data from 2002 to 2005 for 30 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, variance which is only partly removed by the robust DEA specification. We found that the majority of sectors operate below full efficiency, with only a few industries belonging to the top performers. The average efficiency lies between 50 and 70 percent of that of the best sectors. We conclude that only a minor proportion of Czech SMEs are able to generate high value added per unit of labor-capital.

1. Introduction

1.1 Aims of the Analysis and Related Literature

Small and medium-sized enterprises (SMEs) form a vital part of developed economies, as has been stressed in a growing body of literature (see e.g. Schiffer and Weder, 2001; Ayygari et al., 2007; Acs et al. (eds.), 1999; Taymaz, 2005; Yang and Chen, 2007). Research on Czech enterprises has stressed institutional factors related to the transition from a centrally planned economy to capitalism, such as the role of foreign direct investment (FDI) and institutions (examples include Djankov and Hoekman, 2000, and Marcinčin and Wijnbergen, 1997). However, the literature on SMEs in the Czech Republic is rather scarce.

To the best of our knowledge, this paper is the first attempt to measure the economic efficiency of Czech SMEs based on microeconomic principles using data envelopment analysis (DEA), with the main focus on structural analysis of Czech manufacturing SMEs based on their efficiency. Our text therefore complements previous results, which mostly relied on macroeconomic methods. The study by Benáček et al. (1997) is an exception, as the authors measured the efficiency of textile and clothing firms using distance functions. Thanks to detailed information on individual firms, they were even capable of separating technical and allocation efficiency.

* This research was supported by the Grant Agency of the Academy of Sciences of the Czech Republic, Grant No. IAA700280803 “Efficiency and Employment in the SME Sector,” Grant No. P402/12/0982 “Trade Flows in Times of Economic Boom and Slump: Modifying the Gravity Model for Country, Time and Product-specific Decision-Making,” and Grant No. P402/11/0948 “Developing an Analytical Framework for Energy Security: Time-Series Econometrics, Game Theory, Meta-Analysis and Theory of Regulation.” I would like to thank Mr. Vladimír Benáček and anonymous referees for valuable comments. Any remaining errors are of course entirely mine.

In a previous study, Průša (2010) provided general characteristics of the production process among small Czech companies. This paper, by contrast, will closely explore the structural characteristics of SMEs, in that we will perform a cross-sectional study of SME statistics. This way we offer the reader revealing insights into the industrial fundamentals of the Czech economy. Specifically, our model answers the following questions:

1. *How dispersed is the efficiency of individual sectors? Do most firms operate close to the efficiency frontier or away from it?*
   This is important for understanding the extent to which static equilibrium is a good approximation of the real economy.
2. *Which are the most efficient industries?*
   Information about the cross-sectional distribution of efficiency can guide profitable investment decisions which separate winners from losers.
3. *Are industries which are more concentrated and/or more regulated also more profitable?*
   This is especially useful from the regulatory policy point of view.
4. *Does FDI support higher efficiency of the sectors concerned?*
   Foreign investment is promoted as a crucial contributor to economic development. However, its impact is not straightforward.
5. *Are larger firms more efficient?*
   Finally, this is the famous question of production economics, which, from the theoretical viewpoint, is condensed into returns to scale. In practice, we can recognize much subtler points, such as efficient control in family businesses as compared to the embedded agency costs incurred in large corporations.

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

As is usual with empirical research, we are confronted with tensions between theory and practice. While the object–SMEs–is precisely defined, the statistics on SMEs 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 make a 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 SMEs. Next, we proceed to the methodology of our analysis. We review data envelopment analysis (DEA), an efficiency measurement technique which is commonly used in the economic literature. Since numerous modifications have been developed over the years, even the comprehensive handbooks listed in the bibliography of this paper (Cooper et al., 2002; Cooper et al., 2004; Coelli et al., 2005) are far from exhaustive. We focus on two specifications which we find suitable for our data and treat them in more detail.

Table 1 Definition of SME According to the EU Legislation

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 2 Czech Macroeconomic Indicators 2001–2006

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

Source: Czech Statistical Office

Finally, section 3 forms the core of our genuine research. We analyze sectoral data on Czech SMEs for the period 2002 to 2005. DEA is used to obtain industry-specific efficiency scores. This allows us to uncover structural patterns within Czech industrial SME sectors.

1.2 SME Definition

Small and medium-sized enterprises, abbreviated as *SMEs*, are defined as companies that do not exceed specific size limits. The official definition of the European Union is given in *Table 1*. It is not a clearly disjunctive definition if related to employment only. A complication emanates from the fact that in the EU SMEs have 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 an SME.

Many countries have created their own definitions. For example, Switzerland and the USA choose 500 employees as the cut-off.

In the Czech Republic, SMEs account for one third of Czech GDP and close to two thirds of employment. This share was more or less stable over the ten-year period 1997–2006. This holds for accounting value added as well, which stayed close to 53 percent throughout the same period.[1] It confirms that SMEs form the backbone of the Czech economy and deserve proper analysis.

1.3 Macroeconomic Environment

Before we turn to the analysis of SME efficiency, we provide a basic macroeconomic overview for the period 2001–2006 in *Table 2.* There is a minor slowdown visible in 2002 following a downturn in global conditions (the dot-com bubble), but

[1] Statistics on SMEs published by the Ministry of Industry and Trade in its "Report on the Development of SMEs and Their Support in 2006," downloaded from http://www.mpo.cz/dokument32006.html on January 5, 2008.

overall this was a time of both prosperity and increasing productivity for the Czech economy.

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

2. Measurement of Efficiency

2.1 The Concept of Efficiency

Competition is one of 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 a motivation for improvement and thus drives the development of the economy and, ultimately, of society.

The related concepts of comparative advantage, competitiveness, productivity, and efficiency have provided economists with tools to measure economic performance at both the microeconomic and macroeconomic level. Since this paper concentrates on the former, this section provides the 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 than by advances in the pure theory of production. The core of neoclassical economic analysis relies mostly on static equilibrium, which without doubt provides insightful illustrations of market principles, but which cannot properly account for systematic departures from what is perceived as the efficient frontier.

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

Vintage models assume that although aggregate technology is available to all producers, it is evolving over time and thus different producers at different times of investment acquire different vintages of technology. This implies heterogeneity of production capabilities, i.e., a certain time structure of capital. Before an investment is made, the production set (defined later) is the same for all producers i : $\left(\boldsymbol{\Psi}_i \mid \boldsymbol{\beta}\right)$, $\boldsymbol{\beta}$ being the vector of parameters which characterize technology. After the investment is made, each producer has its own specific production capabilities: $\left(\overline{\boldsymbol{\Psi}}_i \mid \overline{\boldsymbol{\beta}}_i\right)$. See e.g. Johansen (1972).

Institutional economics assumes frictions which arise for each exchange transaction, be it exchange on the market (buying or selling at a price) or a non-market transaction (e.g. interaction within an organization). Inefficiencies may result from the internal organization of the firm. Management techniques (termed corporate governance by institutional economists) will crucially influence a firm's performance, as will the staff and their behavior. Even in the same firm a different amount of goods is produced on different days due to unexpected failures and complications. Other bottlenecks may stem from inappropriate institutional settings. The more the state interferes in entrepreneurial activities, the higher the risk that something will go wrong. Ménard (2005) offers an up-to-date summary of the institutionalist view of organizations.

Austrian economics concentrates on entrepreneurs as discoverers of market opportunities. In this dynamic view, the economy is always developing and never achieves static equilibrium. The main stress is put on the importance of time in the production process. Therefore, this stream is somewhat related to the vintage models and the time structure of capital. For a modern overview of Austrian production theory, see e.g. Sautet (2000).

Finally, let us mention the view which was developed by Leibenstein (1966). He coined the term *X*-efficiency and his theory directly assumes inefficiency to be an inherent property of all human activities. Because his approach to inefficiency is axiomatic and does not offer much room for explanation, this theory remains peripheral.

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 revenues less costs. If we are to find out which decision-making unit performs best, 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, while the latter are governed by preferences and scarcity. Accordingly, we formalize the production process and the concept of efficiency.

Following the exposition by Daraio and Simar (2007), the production set $\mathbf{\Psi}$ is defined as all feasible input-output vectors $(\mathbf{x},\mathbf{y})$ from the set of non-negative real numbers $\mathbf{R}^r_{0,+} \times \mathbf{R}^s_{0,+}$:[2]

$$\mathbf{\Psi} = \left\{(\mathbf{x},\mathbf{y}), \mathbf{x} \in \mathbf{R}^r_{0,+}, \mathbf{y} \in \mathbf{R}^s_{0,+} | (\mathbf{x},\mathbf{y}) \text{ is feasible}\right\} \tag{1}$$

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

$$\mathrm{Eff}(\mathbf{\Psi}) = \left\{(\mathbf{x},\mathbf{y}) \in \mathbf{\Psi} \mid \forall\left[\mathbf{x}^1 \le \mathbf{x}, \mathbf{y}^1 \ge \mathbf{y}, \left(\mathbf{x}^1,\mathbf{y}^1\right) \ne (\mathbf{x},\mathbf{y})\right] : \left(\mathbf{x}^1,\mathbf{y}^1\right) \notin \mathbf{\Psi}\right\} \tag{2}$$

A producer will then be technically efficient if and only if it operates on $\mathrm{Eff}(\mathbf{\Psi})$.[3]

[2] For a detailed discussion of the standard assumptions on technology, see e.g. Kogiku (1971).

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

2.2.2 Economic Efficiency

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

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

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

$$\Pi(\mathbf{p},\mathbf{w}) = \arg \max_{\{\mathbf{x},\mathbf{y}\}} \left\{ \mathbf{p}'\mathbf{y} - \mathbf{w}'\mathbf{x} \mid (\mathbf{x},\mathbf{y}) \in \boldsymbol{\Psi} \right\}$$

This is by a contradiction argument equivalent to:

$$\Pi(\mathbf{p},\mathbf{w}) = \arg \max_{\{\mathbf{x},\mathbf{y}\}} \left\{ \mathbf{p}'\mathbf{y} - \mathbf{w}'\mathbf{x} \mid (\mathbf{x},\mathbf{y}) \in \mathrm{Eff}(\boldsymbol{\Psi}) \right\} \tag{3}$$

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

2.3 Measuring Efficiency in Monetary Units

Separating 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 of individual producers; this is, however, rarely available (and in most situations not even sensible).

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

Third, we can define a framework which explicitly allows for price endogeneity and product heterogeneity. Průša (2009) suggested using *money-metric production frontiers*, where the definitions in equations (1) and (2) are in monetary units (see Průša, 2009). In other words, equation (2) tracks the "profit frontier," meaning that the impact of imperfect competition and product heterogeneity is already incorporated into money-denominated data points.

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

Money-metric efficiency frontiers trade in the separation of technical and allocation efficiency for clear economic interpretation. The resulting efficiency score directly captures overall economic efficiency. In terms of equation (3), higher revenues per unit of costs are regarded as equivalent to higher economic efficiency. Moreover, the beauty of the monetary computation lies in the fact that this "profit frontier" logic holds irrespective of technology. It must be stressed that the first and third options are computationally equivalent, since we plug in the data we have. However, it seems more straightforward to assume price endogeneity, especially with cross-sectional data. Therefore, in the following sections we assume the third approach: input vectors $\boldsymbol{x}$ and output vectors $\boldsymbol{y}$ denote data in monetary units unless otherwise stated.

2.4 Data Envelopment Analysis

2.4.1 Basic Model Structure

In this paper we use *data envelopment analysis* (DEA) to analyze economic efficiency. A DEA model constructs a hyper plane 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 from the efficient frontier.

We listed the 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:

$$\min_{\boldsymbol{\lambda},\theta} \theta \tag{4}$$

$$\text{subject to } \theta \mathbf{x}_i \geq \mathbf{X\lambda}$$

$$\mathbf{Y\lambda} \geq \mathbf{y}_i$$

$$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n) \geq 0$$

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

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

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

[5] Instead of assuming data in monetary units, prices can be incorporated into DEA by assigning a value to the objective function, leaving constraints unchanged. This requires strong assumptions, above all that prices remain constant for any amount of inputs consumed and any amount of outputs produced. For examples of allocation efficiency models, see e.g. Coelli (1996), and Cooper et al. (2004, section 1).

2.4.2 Returns to Scale

Model (4) does not impose any additional conditions on $\boldsymbol{\lambda}$, so that technical efficiency is computed under the assumption of constant returns to scale. Variable returns to scale (RTS) were introduced into 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 $\left(\sum_{i=1}^{n}\lambda_i = 1\right) \wedge \left(\forall i : \lambda_i \in \{0,1\}\right)$, we change DEA into the free disposal hull (FDH) model. FDH is not connected to returns to scale and it differs from both the CCR and BCC models in that it draws an envelope that is not convex.[6] We will need this specification later for the statistical modification of DEA.

2.5 Statistical Methods in the Non-Parametric Approach

In this section we select one modification of DEA which surmounts two big obstacles of the basic model: (1) its deterministic and non-statistical nature; (2) the influence of outliers and extreme values (Daraio and Simar, 2007a, p. xviii).

2.5.1 Probabilistic Production Process

The CCR model from section 2.4 is fully deterministic in that it assumes $Pr\left(\left(\mathbf{x}_i, \mathbf{y}_i\right) \in \mathbf{\Psi}\right) = 1$, where $Pr(\cdot)$ denotes probability. This time inputs and outputs are a pair of independent and identically distributed (iid) multi-dimensional random variables (X, Y), although for an individual observation it still holds that $Pr\left(\left(\mathbf{x}_i, \mathbf{y}_i\right) \in \mathbf{\Psi}\right) = 1$. Following the derivation of Daraio and Simar (2007b), this yields a joint probability measure characterized by the function

$$H_{XY}(\mathbf{x}, \mathbf{y}) = Pr\left(X \leq \mathbf{x}, Y \geq \mathbf{y}\right)$$

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

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

$$F_{X|Y}(\mathbf{x} \mid \mathbf{y}) = Pr\left(X \leq \mathbf{x} \mid Y \geq \mathbf{y}\right) = \frac{Pr\left(X \leq \mathbf{x}, Y \geq \mathbf{y}\right)}{Pr\left(Y \geq \mathbf{y}\right)} = \frac{H_{XY}(\mathbf{x}, \mathbf{y})}{S_Y(\mathbf{y})}$$

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

where we assume $S_Y(\mathbf{y}) > 0$. Notice how this corresponds to the idea behind the minimization problem in (4). There, the computation also selects dominant producers with greater output than the analyzed DMU (the third line in linear program 4) and examines by how much the inputs of the analyzed DMU are greater than those of the dominant reference producers (the second line).

This conditional probability can be empirically estimated by computing

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

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

2.5.2 Order-*m* Estimator

Having established the conditional probability measure in the previous section, all that remains is to compute efficiency based on this probabilistic production process. This can be done by the order-m estimator introduced by Cazals et al. (2002).

The idea is simple. Suppose we have an observation $\left(\mathbf{x}_0, \mathbf{y}_0\right)$. As in the CCR model (4), we select those observations with larger output. From this subset of observations satisfying $Y \geq \mathbf{y}_0$ we draw randomly with replacement $X_1, \ldots, X_m$. These draws are then distributed according to the conditional distribution function $F_{X|Y}(\cdot \mid \mathbf{y})$, as follows from the previous section.

We construct the production possibility set as in Daraio and Simar (2007b):

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

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

Then we measure the efficiency of our firm against the production possibility set $\tilde{\mathbf{\Psi}}_m\left(\mathbf{y}_0\right)$ as the expected minimum efficiency score. We first compute

$$\tilde{\theta}^m_{(\mathbf{x}_0, \mathbf{y}_0)} = \inf\left\{\theta \mid \left(\theta \mathbf{x}_0, \mathbf{y}_0\right) \in \tilde{\mathbf{\Psi}}_m\left(\mathbf{y}_0\right)\right\} \tag{5}$$

and take expectations

$$\theta^m_{(\mathbf{x}_0, \mathbf{y}_0)} = E_{X|Y}\left(\tilde{\theta}^m_{(\mathbf{x}_0, \mathbf{y}_0)} \mid Y \geq \mathbf{y}\right) \tag{6}$$

Notice that equation (5) nicely corresponds to the second constraint in the linear program (4): In both cases θ determines by how much it is possible to contract the inputs of DMU_0 before we reach the "minimum input requirement" boundary which is set by firms producing at least as much output as DMU_0.

It is equation (6) which differentiates the probabilistic approach from section 2.4.1. Here, we compare our DMU to randomly drawn subsets of larger producers (i.e., those with higher output), effectively evaluating the CCR model (4) for each draw, and then look at the efficiency score that we can statistically expect over a large number of randomly drawn subsets. That is, instead of computing the efficiency score once deterministically, we compute it many times for smaller subsets of observations (against which our DMU is compared) and then calculate the average score. This procedure is designed to smooth potential outliers or data errors. It is precisely this idea that makes the order-m estimator much more robust than the standard CCR model.

Finally equation (6) has to be turned into an operational procedure for computation. Using the empirical distribution function $\hat{F}_{X|Y}$, and recalling that statistical expectation is simply the integral over the distribution function, it can be shown that the score equals:

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

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

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

2.5.3 Convex Order-*m* Frontier

Most of section 2.4 deals with efficiency estimates based on convex technology. The only exception is FDH, briefly mentioned in 2.4.2. Since the order-*m* frontier is based on FDH, it is not convex. Therefore, in this section we add convexity to the order-*m* model from 2.5.2.

FDH is derived from the approximation of production technology (Daraio and Simar, 2007b):

$$\hat{\mathbf{\Psi}}_{FDH} = \left\{\left(\mathbf{x},\mathbf{y}\right) \in \mathbf{R}^{r+s}_{+} \mid \mathbf{x} \geq \mathbf{x}_i, \mathbf{y} \leq \mathbf{y}_i, i = 1,\ldots,n\right\}$$

$$\hat{\theta}^{FDH}_{(\mathbf{x}_0,\mathbf{y}_0)} = \inf\left\{\theta \mid \left(\theta\mathbf{x}_0,\mathbf{y}_0\right) \in \hat{\mathbf{\Psi}}_{FDH}\right\}$$

Daraio and Simar recall that usual convex DEA scores can be easily obtained from FDH results: it suffices to multiply the observed inputs $\mathbf{x}$ by the FDH efficiency scores $\hat{\theta}^{FDH}_{(\mathbf{x},\mathbf{y})}$ and then run the respective convex linear program on the transformed data, for example the CCR minimization problem as defined in (4).

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

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

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

$$\hat{\theta}^{m,C}_{(\mathbf{x}_i,\mathbf{y}_i)} = \min_{\boldsymbol{\lambda},\theta} \theta \tag{7}$$

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

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

3. Efficiency of Czech SMEs

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 percent of companies with 10–99 employees, and about 2.6 percent of the micro-segment (less than 10 employees). A part of the aggregated data is published in the yearly summary on the economic activity of Czech SMEs.[7]

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

1. the 30-item two-digit OKEC[8] classification, including OKEC codes 10 to 41[9], i.e., agriculture and services are not included;
2. a size classification with breakdowns at the following numbers 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. the years 2002 through 2005.

The data implies the main characteristics of the analysis. The items under point 3 are fitted to the standard economic labor-capital-output framework. Points 1

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

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

[9] OKEC 12 is not included. The 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.

and 2 are used as the basis for cross-sectional 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 $n^{(2005)} = 136$, totaling 540 observations.

3.2 Model Specification

3.2.1 Dimensions of the Frontier

The specifications of production functions generally follow the "KLEM" approach, where gross output y_{gross} is given by a function defined in Burnside (1996, equation 2.1):

$$y_{gross} = f(\text{capital, labor, energy, materials; technology}) \tag{8}$$

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

We can subtract non-productive intermediate inputs from equation (8) and in so doing arrive at a second possible specification where output is measured as value added y_{net}:

$$y_{net} = f(\text{capital, labor, technology}) \tag{9}$$

In this paper we prefer the latter approach for both theoretical and practical reasons. The theoretical justification is that we are interested in productive efficiency, that is, in efficient employment of productive inputs, namely, capital and labor.[10] Efficient use of non-productive inputs is certainly significant from the managerial point of view, but it is not in the scope of this paper. Referring back to section 2.3, in our model higher value added per unit of monetary inputs implies higher economic efficiency.

The practical reason stems from the sensitivity of DEA to outliers, an issue which becomes more pronounced with more variables.[11] The specification in (9) should further improve the robustness of the estimation results due to the lower dimension of the model.

The step from equation (8) to equation (9) places a strong parametric assumption on how energy and materials enter the production process. Let us recall from section 2.3 that we measure efficiency in monetary units.[12] Then, however, (8) and (9) represent transformations of a profit function in which all components are naturally additive. Therefore, in our specification the frontiers as defined in equations (8) and (9) are equivalent. Clearly the resulting efficiency scores will be slightly different, because the latter capital-labor (KL) efficiency (9) neglects the efficiency components in energy and materials. Yet as we noted above, these non-productive inputs are not the focus of our paper; we concentrate on productive efficiency.

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

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

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

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

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

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

3.2.2 Economic Meaning of the Model

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

The average number of employees is more preferable to the number of employees. The latter captures the sum of employees on each particular day, which is then recalculated on the basis of days worked to get the former. It follows that the average captures all fluctuation of employees, which is exactly what we need.

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

We include “investment” even though it is a forward-looking variable. The variables in a production function should represent flows, but “assets” is a stock variable. Ideally we would like to include the real cost of capital to the firm, but this is unknown and we are not aware of any precise measure for it. This is why we assume “investment” to be a good proxy for depreciation, the more so because we use aggregate data on the sectoral level, which smooths the effect of one-off investments on the firm level. In turn, we consider depreciation in itself a plausible approximation for the real cost of capital. Rather than deleting assets altogether from the model, combining “assets” and “investment” should provide us with a reasonable picture of how efficiently firms employ their capital. Moreover, investment can also be interpreted as a proxy for the willingness of firms to innovate. Thus, we argue that it will help us reveal the importance of innovation for the productive abilities of Czech SMEs.[15]

We refrain from deflating the money values, for which we find two reasons. Firstly, if the adjustment is to add any useful information, we require detailed

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

[14] Output can be considered a proxy for y_{gross} in equation (8) and value added a proxy for y_{net} in (9).

[15] The comment on multicollinearity from footnote 13 applies to assets and investment as well.

Table 3 Box Plot Statistics for Efficiency Scores $\theta^*_{(x_i, y_i)}$

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

separate data on input and output prices across various sectors. However, such data are not available, and there is no reason to assume that deflating by the aggregate CPI and PPI would improve the results. On the contrary, all sectors would be deflated by the same figure and this would only distort the results even more. Secondly, since we are measuring value added, and assuming that inflation on both the input and output side of the production equation is similar across sectors, neglecting inflation should not significantly affect our cross-sectional results.[16]

It remains to note that panel research is limited by the short time span–only four consecutive years. Therefore, we do not explicitly account for technological change. Any technological advances are entrenched non-parametrically in the efficiency scores.

3.3 Envelopes I: Standard DEA Results

Consider the BCC model, i.e., equation (4), with the additional constraint $\sum_{i=1}^{n} \lambda_i = 1$ introducing variable returns to scale. We implemented this computation for each year separately via *DEAP*, a freely available program by Coelli (1996).

To get an overview of the distribution of efficiency, we computed the box plot statistics given in *Table 3*, 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 the statistics define the maximum as the upper quartile 1.5 times the quartile spread (3*Q*–1*Q*). Points above this outside bar (or below the respective bar for the minimum) are taken as outliers.

For all years the mean of the scores is higher than the median, meaning that the estimated efficiency distribution is skewed toward lower scores. The average efficiency amounts to a mere 25 percent 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 the number of employees (*Table 4*). It seems that the average efficiency is increasing with more employees, but this relationship starts only at the second size group (10–19 laborers). The smallest firms do best in every year and, moreover, by a considerable gap.

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

Table 4 Mean Efficiency Score $\theta^*_{(x_i, y_i)}$ According to Size Group and Year

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

Proposition 1: Preliminary results. The BCC model revealed the following:

- the distribution of the efficiency results is heavily skewed toward lower scores. It seems that there are outliers which exercise a considerable influence on the 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 the refined results at hand, we will observe what the impact is on the efficiency distribution and its skewness, if any.

Secondly, we will analyze the sectoral structure. To make our conclusions more precise, we take the 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 intersection in at least three years.

3.4 Envelopes II: Robust DEA Results

In this section we report the results of the convex order-*m* estimator (COM). We obtained the scores using the FEAR package by Paul Wilson (2008), where both the Monte-Carlo simulation from section 2.5.2 and the solution of equation (7) are available.

First, we had to specify the computational aspects: parameters m and B. Cazals et al. (2002, theorem 2.3) show that as $m \to \infty$, we have the convergence $\hat{\theta}^{m}_{(\mathbf{x}_i,\mathbf{y}_i)} \to \hat{\theta}^{FDH}_{(\mathbf{x}_i,\mathbf{y}_i)}$, and similarly $\hat{\theta}^{m,C}_{(\mathbf{x}_i,\mathbf{y}_i)} \to \hat{\theta}^{*}_{(\mathbf{x}_i,\mathbf{y}_i)}$. With higher m, fewer observations will lie above the efficient frontier and the estimator gets less robust. Based on trial and error, we chose $m = 50$ (i.e., 10 percent of the observations) as the level of robustness. With lower numbers of reference observations (e.g. $m = 20$), there was an unusually high ratio (more than two thirds) of super-efficient firms with scores higher than unity, which we assessed as implausible. For $m = 50$ this ratio fell to just 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.

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

Table 5 Box Plot Statistics for Efficiency Scores $\hat{\theta}^{m,C}_{(x_i,y_i)}$

	♯ 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

and stripped at the extreme values. Still, the variation of the efficiency scores remains high even for the robust estimator and this volatility appears to be a robust result itself.

Recalling Aigner and Chu (1968) 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 the normal distribution, because there are two peaks. Moreover, the estimates are still skewed to the left, so that despite having used the flexible measure, apparently we did not lose large parts of the information contained in the data.

Table 5 tracks the distribution of the efficiency scores in more detail. When confronted with the initial results in *Table 4*, 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 4*) was a result given by the presence of favorable extreme observations.[17]

As noted in section 3.1, our measure of output is accounting value added, which is defined as output less the cost of materials used in manufacturing.[18] The efficiency estimate therefore says how much value added a firm is able to produce from a certain stock of capital and labor, and it is normalized relative to the best practice. Hence, a lower efficiency score means less value added per unit of capital-labor.

Taking the example of capital, in practice this can be interpreted as follows. A firm can have few or many machines. Remember that all our computations are per

[17] These in turn may have been caused by favorable sample selection.

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

unit of input, say per one machine. Thus, in the textile industry value added per sewing machine can be either high (jeans sold for a higher price, or more jeans produced, or both) or low. Our results mean that in most cases the value added produced by a sewing machine will be rather low.

Proposition 2: Distributional results

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

Let us repeat what we achieved using 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 differ greatly from those in section 3.3.

In *Table 6*, we list the 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 the structural leaders and structural losers of the beginning of the twenty first century. In each of the groups we further distinguish between those oriented toward processing of raw materials and those in advanced manufacturing.

Proposition 3: Structural results

- *Leaders.* Most of the top efficient industries belong to sophisticated manufacturing sectors: food, tobacco products, fabricated metal products, machinery, electrical machinery, and 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 of this segment), as well as wood and cork, and metal ores.
- *Stragglers.* Just two items do not deal in raw materials: office machinery and computers, and the automotive industry. The rest of those losing out are more or less connected to commodities: leather, pulp and paper, coke, refined petroleum products and nuclear fuel, basic metals, recycling, water supply, coal and lignite, and crude petroleum and 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 and lignite, and crude petroleum and natural gas sectors rank among the worst performers means that gains on a large scale (e.g. due to FDI) are not always passed on to suppliers among SMEs.

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

Table 6 Best and Worst Industries According to $\hat{\theta}^{m,C}_{(x_i,y_i)}$

Best industries					Worst industries				
2002	2003	2004	2005	⋂	2002	2003	2004	2005	⋂
10_{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}	

Note: ⋂ indicates that the industry was among the best/worst in at least three years..

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

4. Conclusions

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

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

By construction, DEA is particularly suitable for cross-sectional rankings. Therefore, we let it uncover structural lags among industries. We first observed an unreasonably high variance of the individual efficiency scores. For this reason we applied 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 the literature by the frequent use of the Banker-Charnes-Cooper specification.

The resulting list of leaders and stragglers 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. in the automotive, coal and lignite, and crude petroleum and natural gas segments.

Moreover, we find that the majority of sectors operate below full efficiency, while only a few industries belong to the top performers. The average efficiency lies between 50 and 70 percent 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 SMEs are able to generate high value added per unit of labor-capital. That is, most industries do not generate as much value added from their flow of capital and labor as the best ones. This result is not very surprising, just as it is not very encouraging.

APPENDIX

Standard Industrial Classification:
Nomenclature Générale des Activités Économiques dans les Communautés Européennes

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

Note: See http://ec.europa.eu/comm/competition/mergers/cases/index/nace_all.html

REFERENCES

Acs ZJ, Carlsson B, Karlsson C (Eds) (1999): *Entrepreneurship, Small and Medium-Sized Enterprises and the Macroeconomy*. Cambridge University Press, New York. ISBN: 0-521-62105-4.

Aigner DJ, Chu SF (1968): On estimating the industry production function. The *American Economic Review*, 58:826–839.

Ayyagari M, Beck T, Demirguc-Kunt A (2007): Small and Medium Enterprises Across the Globe. *Small Business Economics*, 29:415–434.

Benáček V, Shemetilo D, Petrov A (1997): Restructuring and measurement of efficiency in firms in transition. In: Sarma S (Ed.): *Restructuring Eastern Europe.* Elgar, Cheltenham: 58–72. ISBN: 1-858-985-765.

Burnside C (1996): Production function regressions, returns to scale and externalities. *Journal of Monetary Economics*, 37:177–201.

Cazals C, Florens, JP, Simar L (2002): Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106:1–25.

Coelli TJ, Rao P, O'Donnell ChJ, Battese GE (2005): *An Introduction to Efficiency and Productivity Analysis*. Springer, New York.

Coelli TJ (1996): A guide to DEAP version 2.1: A Data Envelopment Analysis (computer) Program. *The University of New England, Centre for Efficiency and Productivity Analysis, Working paper*, no. 8/96, Available at: http://www.uq.edu.au/economics/cepa/index.htm. ISBN: 1-863-89-4969.

Cooper WW, Seiford LM, Zhu J (Eds) (2004): *Handbook on Data Envelopment Analysis*. New York, Kluwer. eBook ISBN: 1-4020-7798-X.

Cooper WW, Seiford LM, Tone K (2002): *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Kluwer, New York. eBook ISBN: 0-306-47541-3.

Daraio C, Simar L (2007a): *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications. Studies in Productivity and Efficiency*. Springer, New York. e-ISBN-13: 978-0387-35231-1.

Daraio C, Simar L (2007b): Conditional nonparametric frontier models for convex and nonconvex technologies: A unifying approach. *Journal of Productivity Analysis*, 28:13–32.

Djankov S, Hoekman B (2000): Foreign Investment and Productivity Growth in Czech Enterprises. *World Bank Economic Review*, 14(1):49–64.

Johansen L (1972): *Production Functions: An Integration of Micro and Macro, Short Run and Long Run Aspects*. North-Holland, Amsterdam. 1st edition. ISBN: 0-7204-3181-6.

Kogiku KC (1971): *Microeconomic Models*. Happer & Row, New York. ISBN: 06-043746-4.

Leibenstein H (1966): Allocative efficiency vs. "x-efficiency". The *American Economic Review,* 56(3):392–415.

Marcinčin A, Wijnbergen S (1997): The impact of Czech privatization methods on enterprise performance incorporating initial selection-bias correction. *Economics of Transition*, 5(2):289–304.

Ménard C (2005): A New Institutional Approach to Organization. In: Ménard C, Shirley M (Eds): *Handbook of New Institutional Economics*. Springer, New York: 281–318.

Průša J (2009). A General Framework to Evaluate Economic Efficiency with an Application to British SME. *(Charles University in Prague, Faculty of Social Sciences), IES Working Papers*, no. 25/2009.

Průša J (2010). Productivity of Czech Small and Medium Enterprises: Lagging Behind Their Potential. *Journal of Industry, Competition and Trade*, 10(3–4):343–363.

Sautet FE (2000): *An Entrepreneurial Theory of the Firm*. Foundations of the Market Economy. Routledge, London. ISBN: 0-415-22977-4.

Schiffer M, Weder B (2001): Firm size and the business environment: Worldwide survey results. *World Bank/International Finance Corporation, Discussion paper*, no. 43. ISBN: 0-8213-5003-X.

Taymaz E (2005): Are small firms really less productive? *Small Business Economics*, 25:429–445.

Wilson PW (2008): FEAR: A software package for frontier efficiency analysis with *R*. *Socio-Economic Planning Sciences*, 42(4):247–254.

Yang ChH, Chen KH (2007): Are small firms less efficient? *Small Business Economics*, 32(4): 375–395.