WIFO

1030 WIEN, ARSENAL, OBJEKT 20
TEL. 798 26 01 • FAX 798 93 86

ÖSTERREICHISCHES INSTITUT FÜR WIRTSCHAFTSFORSCHUNG

Statistical Benchmarking as a Development Tool

An Introduction for Practitioners

Klaus S. Friesenbichler, Agnes Kügler

Research assistance: Alexandros Charos, Anna Strauss

February 2017

ÖSTERREICHISCHES INSTITUT FÜR WIRTSCHAFTSFORSCHUNG
AUSTRIAN INSTITUTE OF ECONOMIC RESEARCH

Statistical Benchmarking as a Development Tool

An Introduction for Practitioners

Klaus S. Friesenbichler, Agnes Kügler

February 2017

Austrian Institute of Economic Research
Commissioned by The World Bank
Internal review: Harald Oberhofer • Research assistance: Alexandros Charos, Anna Strauss

Abstract

This note provides an introduction to two prominent econometric benchmarking methods: Data Envelopment Analysis and Stochastic Frontier Analysis. It discusses the econometric techniques, provides a practical example using the World Bank's Enterprise Survey data, and offers conclusions for development practitioners.

Please refer to: klaus.friesenbichler@wifo.ac.at, agnes.kuegler@wifo.ac.at, alexandros.charos@wifo.ac.at, anna.strauss@wifo.ac.at

2017/045/S/WIFO project no: 2315

Medieninhaber (Verleger), Herausgeber und Hersteller: Österreichisches Institut für Wirtschaftsforschung,
1030 Wien, Arsenal, Objekt 20 • Tel. (+43 1) 798 26 01-0 • Fax (+43 1) 798 93 86 • http://www.wifo.ac.at/ • Verlags- und Herstellungsort: Wien

Verkaufspreis: 30,00 € • Kostenloser Download: http://www.wifo.ac.at/wwa/pubid/59303

Executive Summary

This note provides an introduction for practitioners to two prominent econometric benchmarking methods: Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA). Both techniques rely on the calculation or estimation of an efficiency frontier, which is used to identify efficient and inefficient units or firms, respectively. These methods were illustrated by a practical example using Enterprise Survey data. The results were then reflected against total factor productivity scores, and used in subsequent regression analysis.

Each method identifies different firms as efficient. What technique is preferable depends on the context-specific production function that is assumed. This also mirrors that benchmarking is more than a mere comparison of indicators, but requires knowledge about the underlying production mechanisms. In other words, there needs to be an understanding of the specific input-output relationships. Aggregate performance is then shaped by (i) the efficiency frontier and (ii) the average inefficiency, i.e. the mean distance to the frontier.

Benchmarking requires data on comparable units or firms, respectively. The World Bank's Enterprise Surveys data is used to give a practical example. In particular the questions on productivity are generally suitable to perform benchmarking analysis. However, there are limitations with regard to comparability of units. An Enterprise Survey is a firm-level survey of a representative sample of an economy's private sector. This implies that the samples for sector specific studies are often too small. Also, there may be sector specific production functions, which are only insufficiently captured by the standard modules of the questionnaires.

Implementing benchmarking in private sector development projects therefore requires additional information. To this end, the Enterprise Survey design should be adjusted to allow for the analysis of (sub-)sectors. Questions relevant for the context specific production functions should be included. The sampling should (i) consider context specific structures, and (ii) provide for sufficiently large sample sizes to allow for statistical analysis.

Table of Contents

Executive Summary *1*

1. Introduction *4*

2. Statistical benchmarking methods *5*

2.1. Data Envelopment Analysis (DEA) 5

2.2. Stochastic Frontier Analysis 8

3. Benchmarking Manufacturing Firms in Mauritania and Tanzania *11*

3.1. An Example using Enterprise Survey Data 11

3.2. Further Analysis of Efficiency Scores 14

3.3. Limitations 15

4. Conclusions *17*

References *18*

Table of Figures

Figure 1: Input-output combinations, the frontier and economies of scale 5
Figure 2: Farrell efficiency in the two-dimensional case 6
Figure 3: Principles of the Stochastic Frontier Analysis 8
Figure 4: Histograms of MFP and efficiency scores of SFA and DEA (CRS, VRS) 13

Table of Tables

Table 1: A comparison of DEA and SFA 10
Table 2: Results (mean values) of the Benchmarking Methods 14
Table 3: Regression Results 15

1. Introduction

Benchmarking is a prominent tool in development. The effective and efficient implementation of development policies is essential to achieve development goals under given resource constraints. In this respect benchmarking is an important instrument for the assessment of policies, and the term 'benchmarking' has become an often mentioned concept in economic development policy documents. This is in line with developments of performance assessment outside development. For instance, a study by the Global Benchmarking Network explored the role of twenty performance improvement tools in a global survey (Mann et al., 2010). The results showed that 77% of the surveyed firms reported mission and vision statements as well as client surveys. This was closely followed by SWOT analysis (72%) and informal benchmarking (68%). 49% of the respondents reported formal performance benchmarking (49%) and best practice benchmarking (39%).

This note provides an introduction into benchmarking as a tool of efficiency analysis. The objective of benchmarking is to compare different decision making units (e.g., countries, firms or individuals) that are entitled with similar resources or inputs (e.g., skills and capital) to produce similar outputs (e.g., sales). In other words, benchmarking refers to the comparison of performance metrics to similar units. Consequently, the use of benchmarking techniques allows identifying room for improvement, particularly regarding the distribution of scarce resources. By using especially efficient decision making units as roles models, a systematic benchmarking may stimulate learning, coordination and motivation effects. Successful benchmarking ideally leads to performance improvements, a greater understanding of underlying processes, and a refined implementation of policy and strategy instruments.

Overall efficiency of a system is more than a mere input-output relationship. A simple comparison of input and output is likely to lead to biased results.[1] A country could be superior with respect to some indicators (partial efficiency), but at the same time be less efficient from an overall perspective. This highlights the relevance of sub-processes, or as Bogetoft - Otto (2010) phrased it, '*The reason is that to do well in total, it is not only important to do well in the different sub-processes - it is also important to make use of the sub-processes that have relatively higher productivities than others*'. In addition, simultaneously including various input and output factors allow for different combinations of input and output factors that might be efficient and, thus provide a more flexible instrument of efficiency measurement.

Two widely used concepts are compared: Stochastic Frontier Analysis and Data Envelopment Analysis. As a result various optimization algorithms have been developed to calculate or estimate the efficiency frontier with respect to different firms, sectors, countries. This note compares two benchmarking concepts which are prominent in the development literature,

[1] E.g., Edquist - Zabala-Iturriagagoitia, 2015, simply divide linearly aggregated input factors by aggregated output. However, not all inputs have the same effects on every output factor. Various processes transforming inputs into outputs exits and those have different effects on total efficiency (e.g., the use of financial capital versus human resources).

the *Data Envelopment Analysis* (DEA) based on the work of Charnes - Cooper - Rhodes, 1978, and the *Stochastic Frontier Analysis* (SFA). The chosen unit of analysis is the firm.

2. Statistical benchmarking methods

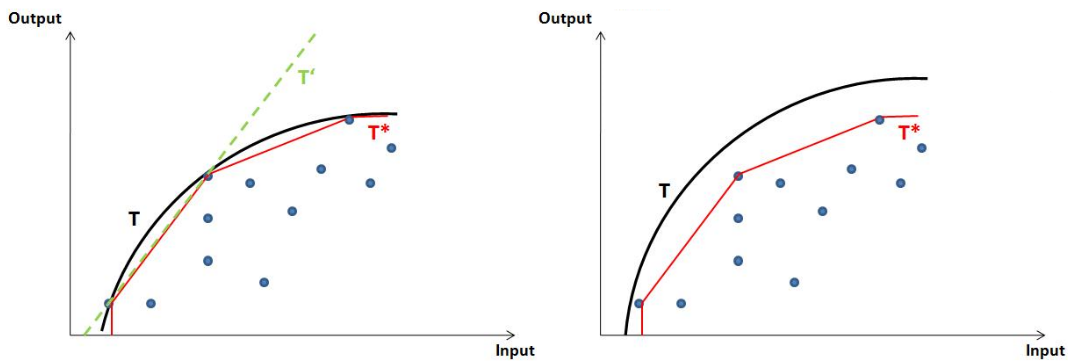
Firms are benchmarked against an efficiency frontier. The 'distance to the frontier' is a measure for inefficiency. The presently discussed methods are 'frontier' models. This is, a unit's performance is measured in terms of its distance from the efficient frontier. The frontier is the based on a function depicting economic behavior, and indicates the maximum attainable level of output corresponding to a given quantity of predefined inputs. The frontier is estimated based on inputs and outputs, which may be chosen based on a microeconomic production function. The distance to the frontier is then an indicator for inefficiency (Chen - Delmas - Lieberman, 2015).

2.1. Data Envelopment Analysis (DEA)

Data Envelopment Analysis computes an efficiency frontier using a set of predetermined inputs and outputs. In a first step, DEA computes an efficiency frontier. It draws on a pre-defined production process and uses the combination of inputs and outputs to extrapolate an efficiency frontier that is a linear surface or 'piecewise hyperplane'. DEA relies on the "minimal extrapolation principle". This is, the frontier represents the smallest set of input-output combinations that satisfies the imposed production assumption, i.e., the inputs assumed to be required to produce outputs (see Bogetoft - Otto, 2010; Jiang - Takeuchi - Lepak, 2013).

The curve (T) in Figure 1 indentifies the theoretical efficiency frontier in the two-dimensional space, i.e. the highest possible output that can be achieved by using one input factor. The observations (the dots) are totally embedded by the theoretical efficiency frontier, even though it is not required that the theoretical efficiency frontier has been reached in reality (see right hand side of Figure 1). This is because DEA approximates the theoretical frontier by using mathematical maximization algorithms based on the observed data points.

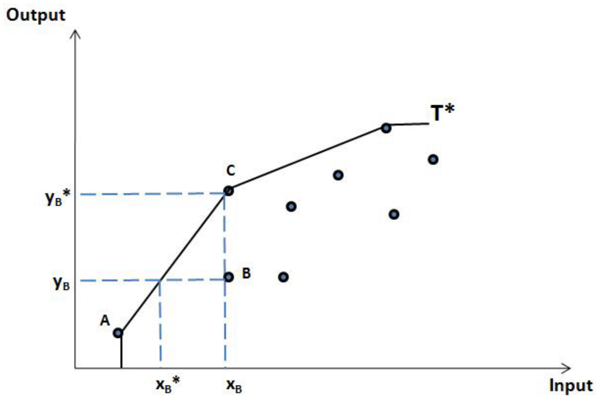
Figure 1: Input-output combinations, the frontier and economies of scale



A prerequisite is that a functional form of the approximated frontier is chosen ex ante. The left hand side of Figure 1 depicts two different functional forms based on assumptions regarding economies of scale. The curve (T*) represents decreasing economies of scale, while the dotted line (T') assumes constant returns to scale, which results in a linear approximation frontier. The (true) theoretical frontier encases the observed data. However, it should be noted that such approximations might over- or underestimate the true efficiency frontier[2]. Hence, the DEA only refers to relative efficiencies with respect to the most efficient observed observation, not to absolute efficiencies.

The distance to the efficiency frontier is an indicator for relative inefficiency. After the empirical determination of the frontier it is possible to identify the distance of observations to the frontier. The most commonly used efficiency indicator was proposed by Farrell, 1957. Depending on the perspective that the specific DEA model takes, the Farrell indicator either measures the potential proportional reduction of inputs at constant outputs, or the potential increase in outputs given constant inputs, respectively. The Farrell efficiency measure can be intuitively explained by Figure 2 in the case of two dimensions. It depicts the observed input-output combinations as well as a calculated efficiency frontier, T*. The input-output combination of decision-making unit B is located below the efficiency frontier. The given output level y_B could also be achieved by using a lower input level x_B^*. Then the input efficiency can be calculated by the ratio of the optimal to the used input level, $E_I = x_B^*/x_B$. The output efficiency can be calculated analogously. Using the input level x_B the decision-making unit B should be able to produce output y_B^* instead of only y_B. The output efficiency of observation B results from the ratio between the optimal output level and the actually achieved output level, i.e. $E_O = y_B^*/y_B$.

Figure 2: Farrell efficiency in the two-dimensional case



[2] This might be the case when the theoretical frontier is characterized by a concave functional form but for the calculations a linear approximation is assumed. Probably the real potential is not yet reached and thus the calculated frontier based on empirical data is situated below the theoretically highest possible output (right hand side of Figure 1).

In the general case of multiple dimensions, the Farrell *input-efficiency* identifies the highest possible proportional reduction of all inputs to achieve given output levels. For example, an efficiency score of $E_I = 0.8$ indicates that 20°% of all input factors could be reduced without a simultaneous reduction of outputs. Vice versa, an output efficiency score of $E_O = 1.2$ would point to a potential increase in output of 20°% maintaining the given input levels. Efficient decision-making units are characterized by efficiency scores of one. Hence DEA computes inefficiency scores in a rather straightforward way. These can be interpreted as an indicator of X-inefficiency, i.e. the difference between units on the efficiency frontier as a proxy efficient behavior of businesses assumed by theory and their observed performance in practice (Leibenstein, 1966). However, the analysis of the reasons for operational efficiency requires an auxiliary regression model.

DEA is a data driven and flexible tool, whose input factors should be chosen carefully. DEA is a highly flexible instrument, which does not require assumptions about the distribution of indicators. In addition, it allows relaxing the assumption of constant economies of scale by drawing a non-linear efficiency frontier. This implies splitting the efficiency component score into (i) size (scale) efficiency and (ii) technical efficiency. In addition to being able to relax the returns to scale assumption, DEA is also flexible in its indicator choice. While economic efficiency typically focuses on financial indicators (e.g., value added, sales per employee), the approach can be extended towards non-financial measures. Hence, DEA can incorporate a variety of dimensions, including innovation, organizational structures or production processes. However, the number of factors used strongly influences the results of a DEA. Generally the more input and output factors included in the analysis, the more units will be identified as efficient. Thus, indicators should be chosen carefully, and the focus should be on the most important factors only, especially if the number of observations is limited. Another important aspect concerns the free disposability of the chosen inputs, or how rational the choice of input-output mix is, respectively. In other words, it asks the question if a shift from one production factor to another is possible, even if it leads to a higher efficiency level.

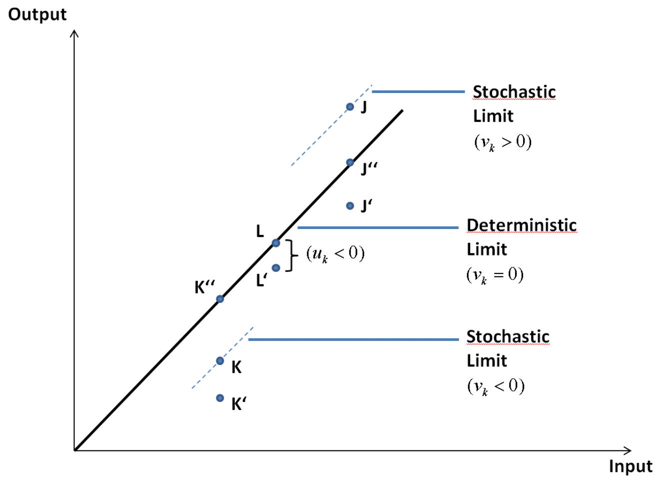
Efficiency scores hinge on the efficiency frontier, which is susceptible to outliers and noise. The original DEA concept is a deterministic one, i.e. it does not consider a random element. Hence, the computation of the efficiency frontier renders it susceptible to outliers, which would otherwise be absorbed by the error component. The constitution of the sample determines the results immensely, i.e. the shape of the efficiency frontier might be changed by excluding/including only one decision-making unit. Generally, the assumption that data are not affected by noise is subject to critique, and poses a major limitation. If data are possibly biased, for instance, due to exogenous shocks or data collection issues, the obtained results of the DEA may not be valid, because the random deviations from the frontier would be interpreted as inefficiencies. In contrast, stochastic approaches like the Stochastic Frontier Analysis (SFA) incorporate a stochastic error term.

2.2. Stochastic Frontier Analysis

Stochastic Frontier Analysis is regression based, and incorporates stochastic noise. Other than DEA, SFA specifies a production (or cost function) with stochastic noise, i.e. an error term. A parametric function of inputs and outputs is assumed, $y_k = f(x_{1k}, x_{2k}, \dots, x_{mk}, u_k, v_k)$. The term u_k represents the technical efficiency of the k decision-making units and v_k represents the stochastic 'noise' component; together they constitute a compound error term. Thus, the SFA allows differentiating between (i) a random component including stochastic variation (e.g., measurement errors) and (ii) a systematic term identifying the inefficiencies (Aigner - Lovell - Schmidt, 1977, and Meeusen - Van den Broeck, 1977).

SFA splits the error term into an efficiency indicator and a stochastic error component. SFA first estimates an average production function (i.e., the deterministic frontier), which is represented by the black line in Figure 3. The decision-making unit L' is located below the efficiency limit. The target output would be point L. In this case the efficiency term is negative and the stochastic component is zero, i.e. this observation is inefficient, but not affected by stochastic influences. Observation K' is characterized by a negative stochastic noise component. Therefore, the value is corrected. This is, the target value K'' (deterministic limit) needs to be adapted; the new target value K (stochastic limit) is located below the black line. Without stochastic shocks ($v_k = 0$) the target output would still be K''. In addition to the noise term v_k also the efficiency term u_k is negative. Therefore observation K' is not only positioned below K but also below K''. As a result, the whole deviation of K' from K'' can be resolved into an efficiency component and a stochastic component.

Figure 3: Principles of the Stochastic Frontier Analysis



Observation J' represents the case of a positive stochastic shock. Therefore the target value has to be shifted above the black line to the new target value J. Without stochastic shocks, i.e. statistical noise, the target value would be J''. One finds a positive stochastic noise component v_k, and a negative efficiency component u_k. In other words, considering stochastic shocks a larger improvement (J' to J) is necessary than without considering stochastic elements (J' to J'').

SFA is based on large sample theory. In contrast to the DEA, which is based on a pure mathematical optimization algorithm, SFA is a statistical method relying on asymptotic assumptions like (frequentist) regression techniques. Therefore, a reasonably large sample is needed to conduct a SFA and statistical tests that go along with it, like t-test and significance levels for coefficients. The efficiency term u_k is negative if the decision-making unit operates inefficiently, otherwise its zero. In contrast, the random part of the compound error term v_k can either be positive, negative or zero. Therefore, the combined error term can be positive or negative and can be treated analogously to the error term in regression analyses. The two components can be determined by estimating the distribution of their variances.

SFA requires assumptions about the distributions of noise and efficiency. To estimate the distribution of variances, one needs to make assumptions regarding the distribution of both noise and efficiency (v_k and u_k). Often noise v_k is assumed to be normally and efficiency u_k to be half-normally or exponentially distributed. In most cases, to estimate the coefficients of the two components of the error term Maximum-Likelihood estimation (ML) is applied. However, assumptions regarding the functional form of the production function (linear, semi-log, etc.) have to be made in accordance to the assumed distributions of the error components. Because of the assumptions about the distributions of v_k and u_k it is possible to draw conclusions about the two components of the error term for every decision-making unit separately.

Wrong assumptions can lead to biased results. If an inappropriate functional form is chosen, deviations might not root in inefficiencies, but in inaccurate model specifications. The same holds true for the error term. In practice, if the data are not following one of the assumed distributions of the efficiency term, the SFA can lead to implausible results. In such cases the application of a DEA would be an alternative approach as DEA does not need any assumptions about distributions. The application of benchmarking methods requires comparable units, which may be problematic when comparing firms where estimating identical functional parameters is conceptually unfeasible. In the SFA, the frontier represents an efficient reference technology for the transformation of inputs into outputs for all observations. Since SFA uses uniform functional parameters, heterogeneity regarding the used technology of decision-making units is not considered. However, if the sample is large enough, one could group the decision-making units according to their technologies and apply SFA to the resulting latent classes that are more homogenous (Orea - Kumbhakar, 2004). Then a comparison between decision-making units is only possible within a group but not between groups.

This comparison has important implications for the choice of the benchmarking method. A comparison of the both methods is provided in Table 1.

Table 1: A comparison of DEA and SFA

Implementation	**DEA**	**SFA**
Frontier shape	The DEA frontier is a piece-wise linear surface	The SFA frontier follows a specific functional form (e.g., Cobb-Douglas, translog).
Applicable to multiple outputs	DEA allows for multiple outputs in the production function. However, including additional outputs may decrease the discrimination power.	Production frontier model requires that output be specified as a single measure; cost frontier models can accomodate multiple ouptuts
Statistical assumptions	DEA is a deterministic approach and therefore does not require assumptions about the probability distributions of parameters.	SFA requires an ex ante specification of the model, including the distribution form of the ineffiency term.
Sampling errors	The DEA efficiency score is confounded with statistical noise and inefficiency; it is more susceptible to the influcence of sampling errors and outliers.	SFA incorporates a statistical error term in the formulation.
Panel data structure	Panel data can be incorporated with assumptions on total productivity changes.	SFA can make use of the panel data structure.
Hypothesis tests for the impacts of inputs and exogenous factors on outputs	DEA generates the efficiency score only. To estimate the impacts (coefficients) of inputs and exogenous factors on outputs, it is necessary to find an auxiliary regression model that uses the DEA efficiency score as the dependent variabe.	SFA can estimate the marginal influence of each input and exogenous factors on the output.
Computation	The DEA efficiency score can be easily obtained by solving a number of linear programming problems.	SFA relies on maximum likelihood estimation; ill-structured data or misspecification of the SFA model can lead to numercial problems when estimating the coefficients.

Source: Adapted from Chen - Delmas - Lieberman, 2015.

3. Benchmarking Manufacturing Firms in Mauritania and Tanzania

3.1. An Example using Enterprise Survey Data

Enterprise Survey data basically represent a feasible data source for a benchmarking analysis. In the following, the World Bank's Enterprise Survey data is used. The dataset provides an establishment (henceforth firm) level survey of a representative sample of an economy's private sector. The surveys cover a broad range of business environment topics including access to finance, corruption, infrastructure, crime, competition, and performance measures. The data generation process followed a stratified random sampling strategy. The stratification considers i) firm size classes (micro: 1-5, small: 6-19, medium: 20-99, large: >99; large firms were oversampled), ii) regions at the district level, and iii) 22 industries (ISIC Rev. 3.1, 2-digit). In the Appendix we provide the references for the respective commands in the statistical software package 'Stata'.[3]

The present analysis focuses on the manufacturing sector in Mauritania and Tanzania. We use data from surveys conducted in Mauritania and Tanzania in 2012 and 2013, respectively. All monetary values have been converted into US dollars, and pecuniary data for Mauritania have been deflated to make it comparable using the same base year. The sample consists of a total of 163 observations, of which 23 are in Mauritania and 140 in Tanzania. We consider manufacturing firms, with the largest sectors being food (ISIC Rev. 3.1: 15; 51 observations), furniture (ISIC: 36; 38 observations) and garments (ISIC: 18) and fabricated metal products (ISIC: 28, 11 observations each).

Variable definitions. In this example, observations on value added, capital, and labor are used to estimate a Cobb–Douglas production function. All variables are log-transformed. This approach is in line with private sector development, which often focuses on firms and therefore allows drawing on micro-econometric literature for conceptual thoughts.

Value added. Value added has been defined as sales (the variable code in the Enterprise Survey questionnaire is d2) minus intermediate inputs (this is the total annual cost of raw materials and intermediate goods used in production; n2e).

Capital costs. This indicator is defined the replacement value of machinery, vehicles and equipment (n7a), which is based on the question, 'Hypothetically, if this establishment were to purchase the assets it uses now, in their current condition, how much would they cost?'.

Labour. Full time equivalents are used as the labor stock. This indicator is the sum of two components. The first component consists of full time employees (i), who are defined as all paid employees that are contracted for a term of one or more fiscal years and/or have a guaranteed renewal of their employment contract and that work full-time (variable code: l1). The second component (ii) comprises full-time temporary employees (l8) weighted by their

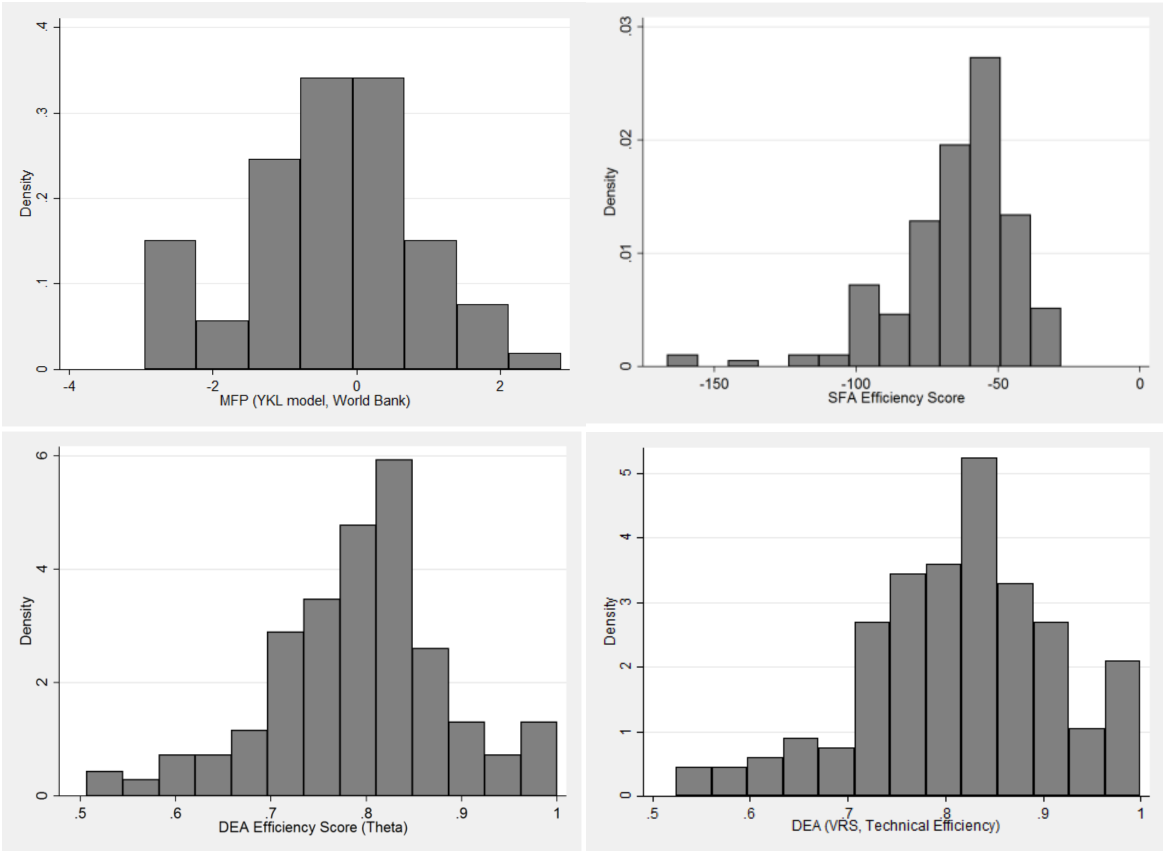
[3] For further details proofs, see Ji - Lee, 2010 and http://www.stata.com/manuals13/rfrontier.pdf.

average length of employment (l8). These were added to the permanent, full-time employees.

We compare the results from benchmarking with multifactor productivity. We use a pooled sample to estimate firm-specific efficiency scores. First, we establish a reference point that puts the benchmarking results into perspective. We use data on multi-factor-productivity (MFP) provided by the World Bank (Cusolito et al. 2016), which is based on a Cobb-Douglas production function. The underlying regression is estimated at the sector level across countries with value added as output and labor and capital as input factors. The productivity indicator follows Syverson (2011), and defines MFP as the sum of the residual, the intercept and a fixed effect. We matched productivity information with the Enterprise Survey data by the unique firm identifier (idstd). This procedure led to a total of 73 matched observations in Tanzania; productivity information on firms in Mauritania's food industry is lacking. Second, we estimate a stochastic frontier analysis using the same three variables. Third, we estimate two output-oriented DEA models, one with constant returns to scale and one with variable returns to scale.

Stochastic Frontier Analysis and Multi Factor Productivity generate similar results. DEA hinges on a different efficiency concept and selects a different set of firms. The first two models exhibit similar results. Firms in Mauritania are not only larger than firms in Tanzania, but also exhibit higher multifactor productivity scores. Accordingly, the inefficiency scores for firms in Tanzania are – on average – slightly higher than for firms in Mauritania. The multifactor productivity scores are highly correlated with the inefficiency scores (ρ: -0.89). It is likely that the values obtained are similar due to the underlying statistical philosophy – both estimators are based on deviations from the averages. The Data Envelopment Analysis approaches efficiency differently. It does not consider deviation from an average, but draws an efficiency frontier using outlier observations and then computes deviations from this curve. It therefore selects a different set of firms as efficient.

Figure 4: Histograms of MFP and efficiency scores of SFA and DEA (CRS, VRS)



Note: The sample is dominated by firms in Tanzania, which make for 93% of all observations. Firms in Mauritania report – on average – more output, capital and labor.4 The SFA suggests higher inefficiency scores in Tanzania, and the DEA reports higher efficiency scores in Mauritania. However, the small size of Mauritania's subsample implies that these results are not representative and to be interpreted with caution. Poolability is assumed in the efficiency estimates.

Figure 4 shows the histogram for each efficiency indicator. Histograms are graphical illustrations to roughly assess the probability distribution of the pooled efficiency scores. All values are put into "bins" covering the range of values. This is, the entire range of values is divided set into a series of non-overlapping intervals. The illustration then shows the frequencies of each interval. The inefficiency scores of the SFA have been transformed into an efficiency index, so that higher values indicate greater efficiency levels. Both constant and variable returns to scale are shown for the data envelopment analysis (CRS and VRS, respectively). Table 2 provides descriptive statistics of the performance indicators and the underlying variables.

[4] These patterns are statistically significant at the 10% level.

Table 2: Results (mean values) of the Benchmarking Methods

	Obs.	Value added (nat. log)	F.T.E. (nat. log)	Capital stock (nat. log.)	MFP (YKL)	Inefficiency Score (SFA)	DEA, Efficiency Score (Theta)	DEA, VRS (Tech. Efficiency)	DEA, VRS (Scale Efficiency)
Mauritania	12	18.822	17.944	4.409	n.a.	0.557	0.827	0.868	0.954
Tanzania	170	16.052	14.996	3.203	-0.381	0.651	0.790	0.812	0.974
Total	182	16.235	15.190	3.282		0.645	0.792	0.815	0.973

Note: This table shows the descriptive statistics for the data used and various performance indicators. MFP: Multifactor Productivity; SFA: Stochastic Frontier Analysis; DEA: Data Envelopment Analysis (Theta denotes the performance indicator for constant returns to scale); VRS: variable returns to scale).

3.2. Further Analysis of Efficiency Scores

Efficiency scores can be used for further analysis. The efficiency scores obtained can be used in regression analysis. The efficiency scores of both DEA and SFA are censored data, i.e. DEA scores take on a maximum value of one, and inefficiencies have a lower bound of zero. Hence tobit regressions which consider the censored structured of the data are typically used. In addition, DEA scores are obtained from a non-parametric method, which is why bootstrapped regressions are typically implemented to consider the non-stochastic nature of the efficiency terms.

In the following, we use several possible explanatory variables which reflect outcomes of common development policies. These are education and the use of emails in firms' correspondence as proxies for the quality both human and physical capital, as well as innovation and quality certificates as proxies for firms' positioning on the market.

Secondary education. This is Percentage of full time permanent workers who completed secondary school.

Quality certification. This is a dichotomous variable taking on the value of one if the firm reports using an internationally recognized quality certification such as ISO 9000 or 14000, or HAPC, and zero otherwise (b8).

Email. This is a dichotomous variable taking on the value of one if the firm reports using emails in their correspondence with clients or suppliers, and zero otherwise (c22a).

Innovation. This is a dichotomous variable taking on the value of one if the firm reports either product (h1) or process innovation (h3), and zero otherwise.

Using emails for corresponding with clients and suppliers as well as quality certificates are associated with higher efficiency. The efficiency scores obtained from the benchmarking can be used in a regression analysis. The results point at slightly different aspects related to technology. The results from the stochastic frontier analysis indicate that a greater share of staff with secondary education exerts a positive influence on efficiency. The DEA relates efficiency to the presence of quality certificates and the use of email in a firm's correspondence. Also multifactor productivity is positively associated with quality certificates.

Table 3: Regression Results

Outcome Variable	(1) MFP (YKL)	(2) SFA	(3) DEA (CRS)	(4) DEA (VRS, TE)
Secondary education (share)	-0.06	14.32*	-0.00	0.00
	(0.621)	(6.007)	(0.026)	(0.029)
Quality certification (dummy)	0.61+	0.26	0.05*	0.05**
	(0.373)	(3.824)	(0.022)	(0.021)
Email (dummy)	-0.45	1.53	0.05*	0.06**
	(0.404)	(3.630)	(0.022)	(0.020)
Innovation (dummy)	-0.49	2.14	-0.02	-0.01
	(0.420)	(4.013)	(0.015)	(0.016)
Tanzania (dummy)		-9.21	-0.02	-0.04
		(7.024)	(0.029)	(0.027)
Constant	0.06	-66.33**	0.80**	0.82**
	(0.450)	(7.621)	(0.030)	(0.030)
Observations	72	174	174	174
R-squared	0.042	0.006	-0.123	-0.180

Note: This table shows the regression results of the impact of education, quality certifications, the use of email and innovation on multifactor productivity (1), SFA (2) and DEA efficiency scores with constant (3) and variable (4) returns to scale. Quality certificates are associated with a higher level of MFP (1), and secondary education exerts a positive influence on efficiency obtained from the SFA (2). Quality certificates and the use of email are linked to higher efficiency scores obtained from the DEA analyses. All regressions with bootstrapped standard errors (500 replications); standard errors in parentheses; Weights were not used due to the bootstrapped standard errors. Significance levels: ** p<0.01, * p<0.05, + p<0.1

3.3. Limitations

Even though benchmarking tools are powerful, there are conceptual limitations. Benchmarking is an effective tool that allows for the identification of inefficiencies, which may promote the effectiveness and efficiency of developing policies. Hence, benchmarking reaches beyond the identification of the average performance of firms, and the relative importance of inputs such as (skilled) labor or physical capital. Benchmarking tools identify (international) top performers. The underlying methods rely on previously collected data, and show an efficiency frontier in a given market. The distance to the frontier mirrors relative inefficiency, or vice versa, firms on the frontier exhibit relative efficiency. However, this does not mean that the efficiency frontier itself cannot be shifted upwards. Perhaps some firms on the frontier have not reached their idiosyncratic efficiency level. This may render efficiency scores as conservative estimates of true, yet unobserved efficiency. Also, the computed frontier hinges on observed data, which does not automatically imply that the estimated frontier is what is theoretically possible.

Benchmarking concerns single units, and not (economic) systems. In the bigger picture of private sector development, the performance of the single firm may be less relevant, however. It may be more important how relative firm performance can be replicated and spill-over to wider sectoral growth. From this perspective the distribution of (in-)efficiencies may be more relevant than identifying single firms. On a similar note, benchmarking is restricted to comparable units. As a result, either tool is likely to be insufficient to analyze complex systems such as value chains or industrial clusters with different types of firms.

Sectoral productivity changes are mirrored by shifts of the efficiency frontier and the average distance of each observed unit from the frontier. The goal of development projects typically is to shift the aggregate productivity of a given industry, and not the productivity of a single firm. Official productivity indices are typically available at the industry level, and are based on aggregating inputs and outputs. These are not directly comparable with DEA or SFA results, which are based on the identification of an efficiency frontier that is shaped by which best-practice firms are selected in a given industry. Changes in overall multi factor productivity are obtained as a compound effect of the shift in the technical frontier and changes in the average distance of each production unit from the frontier (Milana, Nascia, and Zeli 2008). From a slightly different perspective, one could argue that the distribution of efficiency scores across the entire population of firms matters more than the efficiency of a single firm.

There are critical factors for successful benchmarking. According to a study by the Global Benchmarking Network Mann et al., 2010, the key factors for the successful implementation of benchmarking practices are high level support and clear objectives on the one hand, and on the other hand a thorough understanding of mechanisms processes in which benchmarking outcomes are linked to specific measures. Asked about the effectiveness of benchmarking, only about two thirds of the respondents identified benchmarking as having a moderate to high effect. Reversely, a third perceived benchmarking as ineffective. Possible explanations for these are inadequately trained staff, the improper implementation of benchmarking techniques, lacking documentation of the benchmarking process (aim, scope, sponsor and members), which may lead to the misidentification of best practices.

Effective benchmarking goes beyond the mere collection of indicators. While collecting data is a prerequisite for any analytical work, benchmarking goes beyond collecting single indicators, which are subsequently compared to each other. While this is justifiable given uncertainty about complex interactions and relationships between indicators, the question arises whether the collected data can be interpreted in a conclusive manner. For instance, a report produced for the World Economic Forum presents a wide array of measures on 'inclusive growth' (Samans et al., 2015). However, the comprehensive list of indicators makes is difficult to draw unequivocal conclusions. Samans et al., 2015 recognize that many determinants are thought to influence both outcomes and their distributions. While the chosen indicator set is likely to be important for ensuring inclusiveness, the variables used are often not mutually exclusive. A similar point could be made about the World Bank's Ease of

Doing Business Indicators. These may stimulate reform efforts, but it remains unclear whether specific reforms (e.g., start-up procedures) effectively contribute to the achievement of development goals.

Benchmarking requires a conceptual understanding of how indicators are interrelated, thereby relaxing the unrealistic assumption of equal weights. An understanding of how indicators are interrelated is typically lacking in indicator sets which complicates their interpretation. In addition, indicator comparisons are likely to be distorted, since they typically do not impose an explicit hierarchy and therefore do not implement a weighting of indicators. On the contrary, this implies an equal weighting of indicators, and therefore making the unrealistic assumption that all indicators are equally important for achieving outcomes. This weakens the results obtained from indicator sets, because not all processes are equally important. For instance, a firm could rank as more efficient in many overall processes, yet rank as less efficient in its overall score. In other words, doing well in total is not only restricted to processes that have relatively higher productivities than others (Bogetoft - Otto, 2010).

4. Conclusions

This note provided an introduction to data envelopment analysis and stochastic frontier analysis. In private sector development, the measurement of firm performance has become a standard tool in the design, prioritization and evaluation of development policies. This note provided an introduction to two widely used statistical tools – (i) data envelopment analysis (DEA) and (i) stochastic frontier analysis (SFA). Both methods have advantages and disadvantages, and both methods' results hinge on how critical assumptions are made (e.g., about the distribution of inputs). The results were reflected against multi factor productivity scores provided by the World Bank, and eventually used in further regression analysis. Each method identifies different firms as (in-)efficient, and which technique is to be preferred depends on the context-specific production function that is assumed ex-ante. This indicates that benchmarking is more than a comparison of a collection of indicators. It requires an understanding of underlying mechanisms.

Avoiding comparisons of apples with oranges requires detailed information on sub-sectors. Developed countries often provide access to official statistics containing relevant information. This is not possible in developing countries. Statistical benchmarking methods require data on comparable units or firms, respectively. In many cases the direct comparison of units from different sectors might be biased due to different technologies or regulations. The use of samples of specific sub-sectors (e.g., Nace three or four digit data) can help overcome these issues. There are many applications of benchmarking techniques on specific sectors, such as hospitality, energy or banking (Pulina, Detotto, and Paba 2010; Azadeh et al. 2007; Brown 2006). Most of these studies draw on data on developed countries, where very detailed survey data is provided by statistical authorities. Given confidentiality issues, not all statistical agencies offer data access, but anonymized data can be accessed in many countries (e.g., Sweden, the Netherlands or France). Access is granted either via a remote online access, or in a safe center. In developing countries, data that could be used for benchmarking is poorly available, and large scale surveys, such as the World Bank's Enterprise Surveys, do not collect data which are explicitly intended to be used for benchmarking.

Additional data on sub-sectors are required to embed benchmarking into development projects. This note used the World Bank's Enterprise Surveys data to give a practical example. In particular, the questions on firm productivity are suitable to perform benchmarking analysis. However, there are limitations with Enterprise Survey data, which also became evident in this note. An Enterprise Survey is a firm-level survey of a representative sample of an economy's private sector. This implies that the samples used for sector specific studies are often too small. Also, there may be sector specific production functions, which are only insufficiently captured by the standard modules of the questionnaires. These shortcomings especially concern sector-specific analysis, which is not the main objective of this type of survey. Nevertheless, such issues restrict the general applicability of benchmarking techniques. Implementing benchmarking in development projects therefore requires additional information.

Both questionnaires and the collection methodology can be based on the World Bank's Enterprise Surveys. However, the survey design needs to accommodate the needs of a context-specific benchmarking analysis. The Enterprise Survey instrument is a suitable starting point for benchmarking techniques. The questionnaires provide a multitude of critical variables, especially in the productivity section. The questionnaires should be expanded to allow for an analysis of (sub-)sectors. The information collected should allow for the estimation of context-specific production functions, and thus improve the consideration of relevant production factors. Also, the sample size should increase (wherever possible) to allow for statistical estimations.

References

Aigner, Dennis, CA Knox Lovell, and Peter Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6 (1): 21–37.

Azadeh, Amalnick, MS Amalnick, SF Ghaderi, and SM Asadzadeh. 2007. "An Integrated DEA PCA Numerical Taxonomy Approach for Energy Efficiency Assessment and Consumption Optimization in Energy Intensive Manufacturing Sectors." *Energy Policy* 35 (7): 3792–3806.

Bogetoft, Peter, and Lars Otto. 2010. *Benchmarking with DEA, SFA, and R*. Vol. 157. Springer Science & Business Media.

Brown, Rayna. 2006. "Mismanagement or Mismeasurement? Pitfalls and Protocols for DEA Studies in the Financial Services Sector." *European Journal of Operational Research* 174 (2): 1100–1116.

Charnes, Abraham, William W Cooper, and Edwardo Rhodes. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research* 2 (6): 429–44.

Chen, Chien-Ming, Magali A. Delmas, and Marvin B. Lieberman. 2015. "Production Frontier Methodologies and Efficiency as a Performance Measure in Strategic Management Research: Production Frontier Methodologies in Strategic Management Research." *Strategic Management Journal* 36 (1): 19–36. doi:10.1002/smj.2199.

Cusolito, Ana Paula, David C. Francis, Nona Karalashvili, and Jorge Rodriguez Meza. 2016. "Firm Level Productivity Estimates: Methodological Note." *The World Bank*.

Edquist, Charles, and Jon Mikel Zabala-Iturriagagoitia. 2015. "The Innovation Union Scoreboard Is Flawed: The Case of Sweden–not Being the Innovation Leader of the EU." Lund University, CIRCLE-Center for Innovation, Research and Competences in the Learning Economy.

Ji, Yong-bae, and Choonjoo Lee. 2010. "Data Envelopment Analysis." *The Stata Journal* 10 (2): 267–80.

Leibenstein, Harvey. 1966. "Allocative Efficiency vs. 'X-Efficiency.'" *American Economic Review* 56 (3): 392–415.

Mann, Robin S., Ahmed Abbas, Holger Kohl, Ronald Orth, and Mario Görmer. 2010. "Global Survey on Business Improvement and Benchmarking." *Global Benchmarking Network*, 1–40.

Meeusen, Wim, and Julien Van den Broeck. 1977. "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review*, 435–44.

Milana, Carlo, Leopoldo Nascia, and Alessandro Zeli. 2008. "Changes in Multifactor Productivity in Italy from 1998 to 2004: Evidence from Firm-Level Data Using DEA." *EU-KLEMS Working paper,(33)*. http://euklems.net/pub/no33.pdf.

Orea, Luis, and Subal C Kumbhakar. 2004. "Efficiency Measurement Using a Latent Class Stochastic Frontier Model." *Empirical Economics* 29 (1): 169–83.

Pulina, Manuela, Claudio Detotto, and Antonello Paba. 2010. "An Investigation into the Relationship between Size and Efficiency of the Italian Hospitality Sector: A Window DEA Approach." *European Journal of Operational Research* 204 (3): 613–620.

Samans, Richard, Jennifer Blanke, Gemma Corrigan, and Margareta Drzeniek. 2015. "Benchmarking Inclusive Growth and Development." In *World Economic Forum, Discussion Paper, 36p*. http://www3.weforum.org/docs/WEF_Inclusive_Growth_Development2014.pdf.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49 (2): 326–65. doi:10.1257/jel.49.2.326.

Data

The data, questionnaires and implementation notes are publicly available at the World Bank's Enterprise Survey portal: www.enterprisesurveys.org

Syntax

The statistical syntax of the software package Stata is available online for both the stochastic frontier analysis (SFA) and the data envelopment analysis (DEA):

SFA (official manual): http://www.stata.com/manuals13/rfrontier.pdf

DEA (user written): http://www.stata-journal.com/sjpdf.html?articlenum=st0193