The Potential Determinants of German Firms' Technical Efficiency: An Application Using Industry Level Data

by

Oleg Badunenko* and Andreas Stephan†

March, 2004

Abstract

Stochastic Frontier Analysis is employed to calculate technical efficiencies of German firms at the industry level. The data come from the German Cost Structure Census of manufacturing for the period 1995-2001. This survey is conducted by the German Federal Statistical Office (Statistisches Bundesamt). Aggregating 1995 to 2001 firm-level data yields an unbalanced panel with 241 cross-sections (industries). While the unbalanced nature of the data precludes some time-varying specifications, one can estimate the parameters of a time-invariant fixed-effects model. With only one industry being fully efficient, the rest perform poorly, having a technical efficiency mode of 0.32. To account for this outlier, one industry is dropped from the sample. In the reduced sample, the estimated mode of technical efficiency is 0.64.

The scores of technical efficiencies are negatively correlated to concentration indices, positively related to new firm formation and human capital proxy. The analysis shows that technical efficiency is not related to sales growth, R&D expenditures, capital intensity, proportion of East German firms in the industry and size of the firm. The straightforward continuation of this analysis is making use of available firm-level data.

*PhD Student, European University Viadrina, Faculty of Economics, Große Scharrnstraße 59, 15230 Frankfurt Oder, Germany. E-Mail: badunenko@euv.ffo-de; Phone: +49 335 5534 2943.

†Junior Professor of Economics, European University Viadrina, Faculty of Economics and Management Studies, Große Scharrnstraße 59, 15230 Frankfurt Oder, Germany, URL: http://oekonometrie.euv-frankfurt-o.de, E-Mail: astephan@euv.ffo-de; Phone: +49 335 5534 2203, Fax: +49 335 5534 2279.

Introduction

Analysis of a firm activity on a industry level involves comparison of its performance relative to that of others. By the same token, using frontier methodologies, the performance of each industry can be measured relative to 'best-practice' frontier, which is constructed based on the performance of other industries in the economy.

This paper focuses on the cost efficiency of firms at the industry-level; due to the lack of data on the input prices, the basic production frontier model is employed, which is estimated based on the assumption that firms are maximizing output, given inputs[1].

The usefulness of the frontier approach to the measuring efficiency of industries is twofold. Firstly, it provides managers of the firms with answers to the questions regarding cost minimization, organization and distribution systems. Secondly, frontier methodology offers guidance to regulators and policy makers as for solving and mitigating problems in particular industry and economy in general.

Concept of in-Efficiency

On the following figure hypothetical one input (y) – one output (x) production process is depicted.

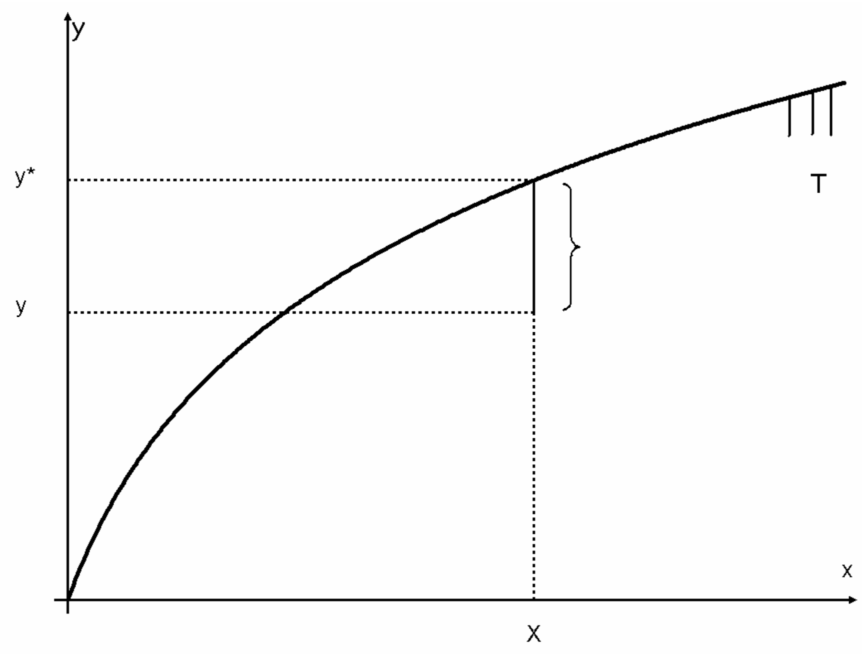

Figure 1: Explanation of in-Efficiency concept: one-input-one-output case

One particular industry uses x input intensity and produces output y. It can be well seen that within given 'best-practice' technology T this particular industry produces inefficiently. With current input intensity x, when efficiently, it could have produced y^*. The distance between y^*, the potential output, and y, the observed output, stands in the literature for the inefficiency.

[1]This definition is referred to as *output-oriented* frontier.

Methodology: Time-Invariant Technical Efficiency

The survey of nowadays applied frontier methods to measure the efficiency maybe found in Kumbhakar and Lovell (2000). In this paper the simplest version of existing panel data efficiency models, **the fixed-effects model** (Schmidt and Sickles, 1984) is employed.

A general Cobb-Douglas production frontier with time-invariant technical efficiency can be written as follows:

$$\ln Y_{it} = \beta_0 + \sum_n \beta_n \ln C_{nit} + v_{it} - u_i \tag{1}$$

where v_{it} represents random statistical noise, and $u_i \geq 0$ represents technical efficiency; $i = 1, ..., I$ and $t = 1, ..., T$. Note that technical efficiencies remain constant over time. For the purpose of our analysis we make the following assumptions: (i) v_{it} are assumed to be iid $(0, \sigma_v^2)$ and uncorrelated with the regressors, and (ii) no distributional assumption is made on the u_i, and the u_i's are allowed to be correlated with the regressors or with v_{it}.

The u_i's are fixed effects and, therefore, are industry specific intercept. With simple substitution $\beta_{0i} = \beta_0 - u_i$ we apply OLS (Least Squares Dummy Variables, LSDV – from here and after) to the modified equation (1):

$$\ln Y_{it} = \beta_{0i} + \sum_n \beta_n \ln C_{nit} + v_{it} \tag{2}$$

After LSDV estimation[2] we obtain $\widehat{\beta_{0i}}$[3]. The maximum of industry specific fixed effects will suggest the most efficient industry. The u_i are proposed to be estimated as in (3)[4] and producer specific technical efficiencies as in (4):

$$\widehat{u_i} = \max_i(\widehat{\beta_{0i}}) - \widehat{\beta_{0i}} \tag{3}$$

$$TE_i = \exp(-\widehat{u_i}) \tag{4}$$

In this framework at least one industry is 100% efficient and others' efficiency scores are measured relative to technically efficient industry (or industries).

Data and Variables Definition

In this paper the data of the German Cost Structure Census of manufacturing for the period 1995-2001 is utilized. This survey is conducted by the German Federal Statistical Office (Statistisches Bundesamt).[5] It comprises almost all large German manufacturing firms with 500 or more employees. Firms with 20-499 employees are included as a random sample which is representative for the respective size category and industry. Firms with less than 20 employees are not sampled.

Output Output is measured by gross production, which comprises the turnover plus the net change of the stock of final products. We exclude turnover from activities that are classified as miscellaneous such as license fees, commissions, rents and leasing, etc. because we assume that such revenue can only be inadequately explained by means of a production function.

[2] The estimates of the β_n's are consistent as either $I \to \infty$ or $T \to \infty$ and the consistency property does not require that the u_i's be uncorrelated with regressors.

[3] The estimates of β_{0i} are consistent as $T \to \infty$, however consistency property requires both $I \to \infty$ and $T \to \infty$.

[4] The u_i's are ensured to be positive to satisfy our assumption.

[5] The firms are obliged to participate in the survey and are not eligible to disclose any required information.

Cost Structure The cost structure census contains information for a number of input categories. These categories are payroll, employers' contribution to the social security system, fringe benefits, expenditure on material inputs, self-provided equipment and goods for resale, energy, external wage-work, external maintenance and repair, tax depreciation of fixed assets, subsidies, rents and leases, insurance costs, sales tax, other taxes and public fees, interest payments as well as "other" costs such as license fees, bank charges, postage or expenses for marketing and transport.

Constructing cost groups Within this structure five following cost categories were aggregated: *material inputs* (intermediate material consumption plus commodity inputs plus energy consumption), *labor compensation* (salaries and wages plus employer's social insurance contributions), *user cost of capital* (depreciation plus rents and leases), *external services* (external services and external contract work) and *other inputs* related to production (e.g., transportation services, consulting or marketing). All input and output data series were deflated using the producer price index for the respective industry.

Additional Information Further industry-level information available in the Cost Structure Census includes[6] (i) Competition Rate (Herfindahl Concentration Index), (ii) Sales Growth, (iii) New Firm Formation Rate (mean annual number of new firms per employee at the 3-digit industry level 1992-2000 (%), (iv) R&D Expenditures, (v) Capital Intensity (mean annual depreciations plus expenditures for rents and leases over sales at industry level), (vi) Human Capital Intensity (number of employees with a university degree divided by number of untrained employees (%), (vii) Proportion of East German Firms (proportion of firms with headquarter in East Germany over all firms (%)), (vii) Average Firm Size (log of mean number of employees in respective industry from 1992 to 2000).

Results

In the preceding analysis it is implicitly assumed that all firms, and, therefore, industries face the same environment. Consequently, all the deviations from the frontier are only due to inefficiency.

Technical Efficiencies

The first step in calculating technical efficiency scores involves Least Squares Dummy Variables estimation. With all industries included in estimation, one outlier (tobacco industry) drives all the results: this industry is efficient, while the technical efficiency score's mode of the rest is only 0.32. Without this outlier, there are yet *few* efficient industries and technical efficiency score's mode of the rest increases up to 0.64. Additional dropping of outliers does not change the results considerably, i.e. as dropping of tobacco industry. In spite the fact that additional dropping of the outliers may yield better distribution of technical efficiencies it reduces the sample, which might make further analysis disadvantaged. Therefore, only one outlier[7] is decided to be dropped.

[6]The method of proxy construction is pointed in the parentheses.

[7]The output of the regression and resulted technical efficiencies indices' distribution for the full sample maybe found in Appendix.

Table 1. – Fixed-Effects regression, LSDV

$$\ln Prod_{it} = \beta_{0i} + \ln Mat_{it} + \ln Lab_{it} + \ln Ext_{it} + \ln Cap_{it} + \ln Oth_{it} + v_{it}$$

Number of groups = 236
Number of obs = 1573
Overall R^2 = 0.9953

lnProd	Coef.	Std. Err.	t	p-value	[95% Conf.	Interval]
lnMat	0.5366338	0.00847	63.36	0.000	0.5200177	0.5532498
lnLab	0.185674	0.0122416	15.17	0.000	0.1616591	0.2096889
lnExt	0.087042	0.0055375	15.72	0.000	0.0761789	0.0979051
lnCap	0.0642858	0.0094572	6.80	0.000	0.0457333	0.0828384
lnOth	0.1049571	0.0075441	13.91	0.000	0.0901575	0.1197566
_cons	1.646474	0.0711816	23.13	0.000	1.506834	1.786114

F test that all $u_i = 0$: $F(235, 1332) = 39.60$ Prob $> F = 0.0000$

Table 1 indicates that all the coefficients are significant at all conventional levels and that the regression is fitted good.

Further, technical efficiency scores are calculated using equation (4). The summary statistics of obtained technical efficiencies scores that are presented in Table 2, indicate low level of technical efficiency. Specifically, industries could have produced, on average, the same level of output using only 65% of inputs they actually used.

Table 2. – Sample Statistics: Efficiency Measure

variable	# obs	mean	st. d.	min	max
TE	236	.6517072	.0674251	.3502355	1

On the following figure histogram and *Kernel*[8] density estimates of technical efficiencies scores of 236 industries are shown:

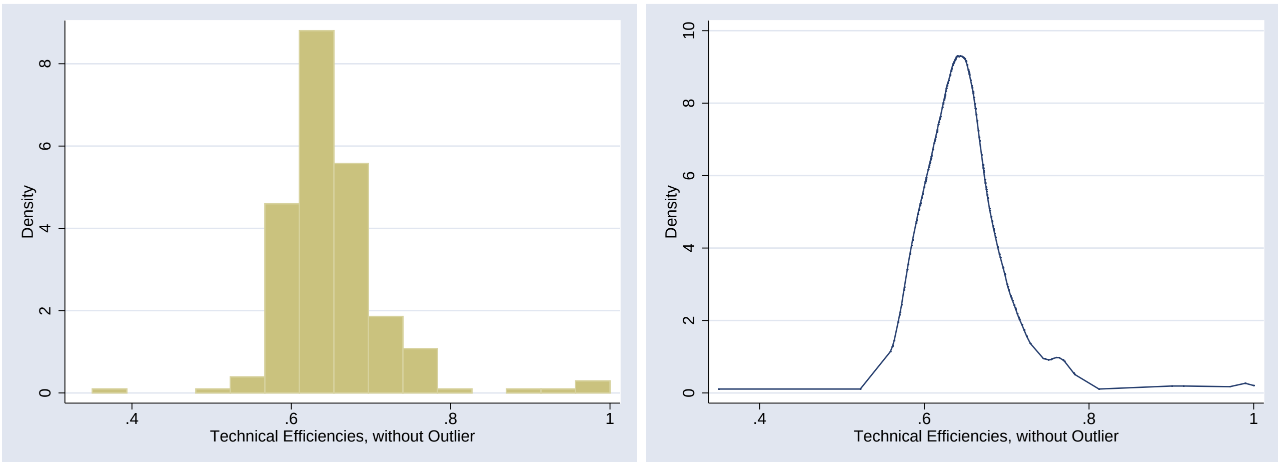

Figure 2: Histogram and *Kernel* density estimate for technical efficiency produced using equation (4)

The distribution of TE (Figure 2) is only slightly positively skewed, contrary to the rationale for using a one-sided distribution for the efficiencies. This problem has been noticed by other researchers (e.g. Carree, 2002), and so far the only proposed solution involves changing the assumed distribution

[8]We choose Gaussian normal kernel, with optimal bandwidth by Silverman (1986).

of the u_i's in equation (1). However, since LSDV estimation does not assume a particular distribution for the firm level inefficiencies, our purified-of-outliers scores of technical efficiencies can be trusted and used as endogenous variable in further analysis.

Correlates of Technical Efficiencies

In this subsection regression analysis is used to determine whether the efficiency scores are related to characteristics of industries, defined and described in section Data and Variable Definition, in Additional Information description.

The output of the regression is presented in Table 3. The independent variables do not account for the significant share in the variation of technical efficiency measure.

Table 3. – OLS regression

$$TE_i = \beta_{0i} + X\beta + \epsilon_i$$

Number of obs = 236
$F(8, 227) = 4.99$, $Prob > F = 0.0000$
Adj. $R^2 = .12$

Tech. Eff.	Coef.	Std. Err.	t	p-value	[95% Conf.	Interval]
herfindahl	-.0958505	.0354733	-2.70	0.007	-.1657496	-.0259515
sales growth	.1514117	.1284603	1.18	*0.240*	-.1017154	.4045389
new firm	5.267175	1.570813	3.35	0.001	2.171936	8.362413
r&d	-.2346449	.2878304	-0.82	*0.416*	-.8018058	.3325161
cap	-.1976795	.1808955	-1.09	*0.276*	-.5541286	.1587695
hum cap	.2381159	.0791692	3.01	0.003	.0821154	.3941164
east	.0003808	.0004095	0.93	*0.353*	-.0004261	.0011877
size	-.0045804	.006763	-0.68	*0.499*	-.0179066	.0087458
_cons	.633303	.0518139	12.22	0.000	.5312053	.7354007

However, Herfindahl Concentration Index is negatively related, while new firm formation and human capital are positively related to the technical efficiency; and significantly so. Additionally, F-statistic indicates that multiple regression equation is statistically significant.

These findings suggest that improving competition (that is, reducing concentration) has a potential to increase technical efficiency. By the same token, foundation of new firms and increasing the number of employees with university degree would facilitate technical efficiency improvement.

Summary and Conclusions

German industries in the sample during 1995-2001 time span are characterized by quite low level of technical efficiency. The scores of technical efficiency are negatively related to concentration indices and positively related to new firm formation and human capital proxies.

Performed analysis reveals that (i) R&D expenditures, (i) capital intensity, (iii) proportion of east German firms and (iv) size of the firm do not have influence on technical efficiency. This does not seem to be plausible in the real world. However, this analysis is based on the aggregated data, and the aggregation might have disclosed important properties of the data. That is why, the prospect for the future research would be the same analysis, but with firm-level data.

References

[1] Aly, Hassan Y.; Grabowski, Richard; Pasurka, Carl; Rangan, Nanda "Technical, Scale, and Allocative Efficiencies in U.S. Banking: An Empirical Investigation. " *The Review of Economics and Statistics*, Vol. 72, No. 2. (May, 1990), pp. 211-218.

[2] Carree, Martin A. "Technological Inefficiency and the Skewness of the Error Component in Stochastic Frontier Analysis. " *Economic Letters*, 77, 2002, pp. 101-107.

[3] Kumbhakar, Subal C.; Lovell, C.A.Knox. Stochastic Frontier Analysis, Cambridge University Press, 2000

[4] Schmidt, P., and Sickles R.C. "Production Frontiers and Panel Data " *Journal of Business and Economic Statistics*, 2:4 October 1984, pp. 367-74.

[5] Silverman B.W. Density Estimation: for Statistics and Data Analysis, Chapman and Hall, London, 1986.

Appendix

Table A1 *Fixed-Effects regression*

$$\ln Prod_{it} = \beta_{0i} + \ln Mat_{it} + \ln Lab_{it} + \ln Ext_{it} + \ln Cap_{it} + \ln Oth_{it} + v_{it}$$

Number of groups = 237
Number of obs = 1580
Overall R^2 = .9931

lnProd	Coef.	Std. Err.	t	p-value	[95%Conf.	Interval]
lnMat	.5338688	.0085244	62.63	0.000	.5171462	.5505914
lnLab	.1882965	.0123227	15.28	0.000	.1641227	.2124703
lnExt	.0858638	.0055798	15.39	0.000	.0749178	.0968099
lnCap	.0644981	.0095362	6.76	0.000	.0457906	.0832056
lnOth	.1056842	.0075993	13.91	0.000	.0907763	.1205921
cons	1.657035	.0717926	23.08	0.000	1.516196	1.797873

F test that all $u_i = 0$: $F(236, 1338) = 58.82$ Prob $> F = 0.0000$

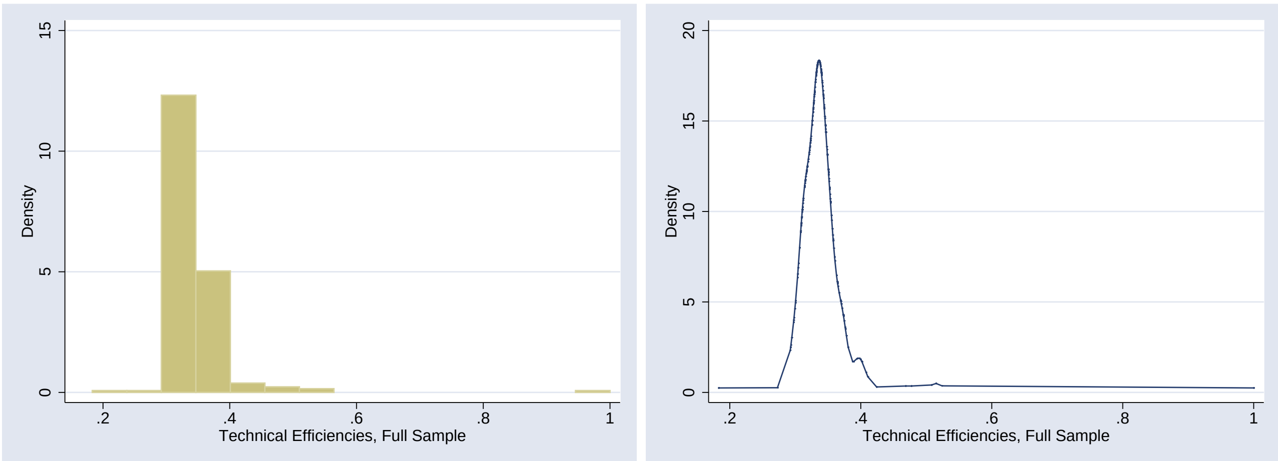

Figure 3: Histogram and Kernel density estimate for technical efficiency produced using equation (4)