Semiparametric Stochastic Frontier Analysis: An Application to Polish Farms during Transition

Arne Henningsen[1] and Subal Kumbhakar[2]

[1] Institute of Food and Resource Economics, University of Copenhagen, Rolighedsvej 25, 1958 Frederiksberg C, Denmark, arne.henningsen@gmail.com

[2] Department of Economics, Binghamton University, State University of New York, PO Box 6000, Binghamton, NY 13902-6000, USA

Paper presented at the European Workshop on Efficiency and Productivity Analysis (EWEPA) in Pisa, Italy, June 24, 2009

Abstract

Almost twenty years after transition to a market economy and a few years after EU accession, the Polish farm sector is still largely underdeveloped and has an extremely low labor productivity. Studies about Polish farms using Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) report contracting results regarding efficiency levels. Hence, it is not clear whether inefficiencies are a major cause of the poor economic situation of many Polish farms. Since data from transition economies are relatively noisy and farms in transition countries generally apply rather different production technologies, neither the deterministic DEA nor the parametric SFA seems to be appropriate. As a solution, we apply a semiparametric stochastic frontier analysis, which has a stochastic component *and* models the production technology nonparametrically. Our preliminary results are compared with efficiency estimates, partial production elasticities, and elasticities of scale obtained from the conventional (parametric) SFA. The efficiency estimates of the semiparametric SFA are highly correlated with the estimates of the conventional SFA but they are – in contrast to the conventional SFA – rather high. Hence, the low profitability of many Polish farms cannot be explained by low technical efficiency.

Keywords: efficiency, farms, Poland, semiparametric

1 Introduction

Poland is the largest of twelve countries that have joined the EU in the course of its recent Eastern enlargements. The Polish farm sector is of particular importance because agricultural employment accounts for an extraordinarily high share of Poland's total labour force. However, almost twenty years after the start of the transition programme and a few years after EU accession, the Polish farm sector is still largely underdeveloped.

While the farm sectors of most Central and Eastern European Countries (CEECs) were dominated by large state-owned or cooperative farms during the period of socialism, small family farms were prevalent in Poland (Borzutzky and Kranidis, 2005, p. 628). They cultivated approximately 80% of Poland's utilised agricultural area (UAA) (Lerman and Schreinemachers, 2005, p. 682). Before transformation, the average size of private farms was only around 5 ha (Borzutzky and Kranidis, 2005, p. 629). It has increased only slightly during the transformation (European Commission, 2002, p. 8) and is currently approximately 7.6 ha (GUS, 2007, Table 41). Even EU accession did not have a significant impact on farm structure (Wilkin, 2007, p. 7).

Although 17% to 29% of total Polish workforce are employed in the farm sector[1], the agricultural sector contributes only 3% to the Polish gross domestic product (GDP) (e.g. Pacuszka, 2005; FAO, 2005). The large gap between the proportion of employment and the proportion of GDP indicates the low labour productivity of the Polish farm sector (e.g. Lerman and Schreinemachers, 2005, p. 678, USDS, 2006). Gross value added per agricultural worker in Poland is only 8.4% of the EU-15 average (Pouliquen, 2001, p. 35) and is the third lowest in the EU-27 (European Commission, 2007, Table 3.3.8.i). Davidova et al. (2002, p. 86, 2005, p. 669) compare Polish farms with farms in two regions in the EU-15 and find that the net value added per agricultural worker in Poland is only 8.6% of the value in South-East England and 7.0% of the value in the Navarra region of Spain. This necessarily leads to a low remuneration of agricultural labour, and hence to low incomes from agricultural production. The resulting lack of profitability is a major problem for Polish farms (Davidova et al., 2005, p. 665). Given their low profitability, the long-term survival of many farms is questionable (Davidova et al., 2002, p. 86).

The poor income situation of many farm households has been somewhat alleviated after EU accession, because the newly introduced direct payments considerably improve the financial situation of Polish farms (Wilkin, 2007, pp. 6–7). Since these payments are granted per hectare of cultivated land, operators of smaller farms benefited from them much less — in particular because smaller farms have a higher labour intensity (Lerman and Schreinemachers, 2005, p. 675) and thus receive less support per agricultural worker.

[1]Surveys on the proportion of agricultural employment in Poland present varying results. While GUS (2006, p. 15) and Góra et al. (2006, pp. 20–21) report a proportion of approximately 17% for the year 2005, Dries and Swinnen (2002, p. 457), Lerman and Schreinemachers (2005, p. 682), Pacuszka (2005, p. 6) and the USDS (2006) report proportions between 25% and 29%.

2 Efficiencies of Polish Farms

The low productivity of Polish farms can be caused by inefficient agricultural production (Kumbhakar and Lovell, 2000, p. 2). During socialism, many private and state farms were rather inefficient (Brada and King, 1993, 1994) and gains in efficiency would have been needed to achieve competitiveness and stabilise farm incomes in a market economy (Brooks et al., 1991, p. 153). However, most Polish farms are still in a poor economic situation. Zegar and Floriańczyk (2003, p. 14) claim that one of the main reasons is low efficiency.

A few studies on the efficiency of Polish farms can be found in the literature:[2] van Zyl, Miller Jr. and Parker (1996), Lerman (2002) and Latruffe et al. (2005) analyse efficiency with the Data Envelopment Analysis (DEA); Munroe (2000, 2001) and Brümmer, Glauben and Thijssen (2002) apply a Stochastic Frontier Analysis (SFA). While Munroe (2000, 2001) uses a Cobb-Douglas production function, Brümmer, Glauben and Thijssen (2002) apply a Translog distance function.

Table 1: Efficiency of Polish farms

	time period	method	technical efficiency	scale efficiency	allocative efficiency	total efficiency
Brümmer, Glauben and Thijssen (2002)	1991–1994	SFA	76%	—	—	—
van Zyl, Miller Jr. and Parker (1996)	1996	DEA	98%	98%	77%	73%
Munroe (2000, 2001)	1996	SFA	57%	—	—	—
Latruffe et al. (2005)	1996	DEA	76%	94%	—	—
Lerman (2002)	2000	DEA	25%	—	—	—
Latruffe et al. (2005)	2000	DEA	71%	92%	—	—

Note: The efficiency measures of Brümmer, Glauben and Thijssen (2002), van Zyl, Miller Jr. and Parker (1996), and Latruffe et al. (2005) have been calculated as unweighted or weighted mean of the published values.

Source: see first column

These studies report varying results (see Table 1). For instance, values for the average technical efficiency range between 98% (van Zyl, Miller Jr. and Parker, 1996) and 25% (Lerman, 2002). Neither a clear improvement nor a clear decrease in technical efficiency can be observed over time. Scale efficiency has been analysed only by van Zyl, Miller Jr. and Parker (1996) and Latruffe et al. (2005), who report average values of 98% and around 93% respectively. Allocative efficiency has been examined solely by van Zyl, Miller Jr. and Parker (1996), who obtained an average value of 77%.

The results on the relationship between farm size and efficiency vary.[3] Munroe (2000, 2001) reports that farms larger than 15 ha are less technically efficient. van Zyl, Miller Jr. and Parker (1996, p. 34) also show that technical efficiency is higher for smaller farms (≤ 15 ha) than for larger farms (> 15 ha), while allocative efficiency and scale efficiency do not depend on the size of the farm. However, Latruffe et al. (2005, p. 287) find that most crop farms operate under increasing returns to scale, which means that these farms are too small. The results of Lerman (2002, p. 8) are

[2] In the following, we do not consider efficiency analyses of the pre-reform time (e.g. Brada and King, 1993, 1994) because these studies are not relevant to the current state of the Polish farm sector.

[3] These findings must be interpreted with care because analyses of the relationship between size and efficiency are plagued by empirical as well as conceptual problems (Kislev and Peterson, 1996).

somehow in-between; they indicate that the smallest farms (≤2 ha) and the largest farms (>30 ha) achieve relatively high technical efficiency, while mid-sized farms are characterised by low technical efficiency. However, he also shows that 83% of the farms with up to 5 ha have increasing returns to scale (Lerman, 2002, p. 8), indicating that smaller farms are less scale-efficient.

There are also some results regarding the relationship between efficiency and other factors. Latruffe et al. (2005, p. 287) show that crop farms are not as technically and scale-efficient as livestock farms. Moreover, Munroe (2000, 2001) finds a positive impact of the farmer's experience (measured as his age) and the modernisation level of the farm (measured as electricity and gas heating use) on technical efficiency.

3 Parametric, Nonparametric, and Semiparametric Efficiency Analysis

It is highly questionable if the non-parametric deterministic Data Envelopment Analysis (DEA) and the parametric Stochastic Frontier Analysis (SFA) are appropriate for analyzing farm efficiency in transition economies at all. Data from transition economies are relatively noisy (Gorton and Davidova, 2004, p. 6) so that a deterministic approach (such as the DEA) seems to be inappropriate as it does not allow for noise in the data. On the other hand, small subsistence farms, market-oriented family farms, and large commercial farms in transition countries generally apply rather different production technologies so that even flexible functional forms such as the translog cannot model their production technologies adequately. Hence, a parametric approach (such as the SFA) seems to be inappropriate (Gorton and Davidova, 2004, pp. 6–7) because it requires the specification of a functional form and selecting a wrong functional form may lead to severely biased estimation results.

Given that the data are noisy *and* production units might have rather different technologies (in parametric sense), a stochastic *and* nonparametric approach is required. In cases like this, the semiparametric approach proposed by Fan, Li and Weersink (1996) could be more appropriate, because it allows for statistical "noise" and does not require the specification of a functional form for production technologies. Fan, Li and Weersink (1996) first estimated a nonparametric (average) production function

$$y = f(x) + \epsilon, \tag{1}$$

where y is the output quantity, x is a vector of input quantities, $f(.)$ is a function of an unknown functional form, and ϵ is a disturbance term.[4] In a second step, they decomposed the residuals from the first step ($\hat{\epsilon} = y - \widehat{y}$) into a constant ($\mu$), a non-negative technical inefficiency term ($u \geq 0$), and a statistical noise term (v):

$$\hat{\epsilon} = \mu + v - u. \tag{2}$$

The constant μ must be included, because the expected values of the error term (ϵ) and the statistical noise term (v) are both zero so that including (a positive) μ allows for a non-zero expected value of $v - u$, which further allows for positive values of the inefficiency term u (Fan, Li

[4]Please note that in our specification, we define the average production function $f(x) = E[y|x]$ and the corresponding error term $\epsilon = y - f(x) = y - E[y|x]$ with $E[\epsilon|x] = 0$, while Fan, Li and Weersink (1996) define the frontier production function $g(x) = f(x) + \mu$, and the corresponding error term $\varepsilon = y - g(x) = \epsilon - \mu$ with $E[\varepsilon|x] \neq 0$.

and Weersink, 1996, p. 462). Hence, the entire semi-parametric frontier model becomes

$$y = \mu + f(x) + v - u. \tag{3}$$

Although this approach seems to be more appropriate than the SFA and DEA in many empirical applications, it has not been used much in applied studies, perhaps because nonavailability of softwares. However, this has changed in recent years. In the first step, we use the powerful and feature-rich "np" package (Hayfield and Racine, 2008) to estimate the nonparametric local-linear production function. In contrast to Fan, Li and Weersink (1996), we use logarithmic output and input quantities in the nonparametric regression:

$$\ln y = f(\ln x) + \epsilon. \tag{4}$$

This has three advantages: (i) the gradients of the nonparametric model can be interpreted as partial output elasticities and their sum as elasticity of scale; (ii) the observations on the logarithmic output and input quantities are more equally distributed, which is advantageous when using constant bandwidths, and (iii) this allows us to use the usual specification of a stochastic frontier function, where the dependent variable is logarithmic so that the predicted dependent variables cannot become negative. The bandwidths of the regressors in the non-parametric regression are selected according to the expected Kullback-Leibler cross-validation criterion (Hurvich, Simonoff and Tsai, 1998). The Epanechnikov kernel is used for the continuous regressors and the kernel proposed by Aitchison and Aitken (1976, p. 29) for categorical variables. From this non-parametric estimation, we retrieve the residuals ($\hat{\epsilon} = \ln y - \widehat{\ln y}$). In the second step, we decompose the residuals as shown in equation (2) above. We use the "frontier" package (Coelli and Henningsen, 2008) to do this decomposition, where the residuals from the non-parametric estimation ($\hat{\epsilon}$) are used as dependent variable and the only regressor is a constant (see equation 2). Hence, our entire semi-parametric frontier model becomes

$$\ln y = \mu + f(\ln x) + v - u. \tag{5}$$

Alternatively, we use a specification that allows for a more flexible adjustment of the frontier in the second step:

$$\ln y = \mu + \lambda\, f(\ln x) + v - u. \tag{6}$$

In this case, we retrieve the predicted (fitted) logarithmic output quantities from the (first-step) non-parametric estimation ($\widehat{\ln y}$) and estimate the following stochastic frontier model in the second step:

$$\ln y = \mu + \lambda\, \widehat{\ln y} + v - u, \tag{7}$$

i.e. we use the observed logarithmic output quantity ($\ln y$) as dependent variable and a constant as well as the non-parametrically fitted logarithmic output quantity ($\widehat{\ln y}$) as regressors. However, in our empirical application, the estimated coefficient λ turns out to be virtually one so that the extended model (6) becomes virtually identical to the basic model (5) and both procedures return almost identical parameter estimates and efficiency estimates.

4 Data

The empirical analysis is conducted using three separate cross-sectional data sets (1994, 2000, and 2006) of Polish farms. We have aggregated the input and outputs to one composite output and four inputs: land, labour, capital, and intermediate inputs (seed, fertiliser, pesticides, fuel, purchased feed, ...). Furthermore, a categorical variable specifying the region where the farm is located is added to account for differences in soil quality and climate.

5 Results

5.1 Semiparametric Efficiency Analysis

Currently, we have only preliminary results for the year 1994. The capital input did not have a significant effect on the output quantity and its marginal products were even mostly negative. This supports the argument of Latruffe et al. (2005, p. 293) that many Polish farmers made poor investment decisions in the beginning of the transition period and therefore had stocks of obsolete capital. As negative marginal products may lead to perverse efficiency estimates (Henningsen and Henning, 2008), capital was not used as input in this analysis. Figure 1 shows that partial production elasticities of land and labour are rather low and nearly constant. In contrast, the partial production elasticity of intermediate inputs is rather high and varies with the input quantity. The high partial production elasticity of intermediate inputs compared to labour can be explained by the scarcity of intermediate inputs and the abundance of labour in the early transition period. The efficiency estimates are rather high. They have a mean of 0.966 and range between 0.951 and 0.976. Hence, the low profitability of many Polish farms cannot be explained by low technical efficiency. However, the low marginal product of labour, which is most likely caused by the abundance of agricultural workers, seems to be the main cause of the low profitability of Polish farms.

5.2 Comparison of Parametric and Semiparametric Efficiency Analysis

Figures 2, 3, and 4 compare the partial production elasticities of intermediate inputs, labour, and land, respectively estimated by the semiparametric SFA and the (parametric) SFA with Cobb-Douglas and Translog functional form. These approaches lead to partial production elasticities that are generally in a similar range but are only slightly correlated. Similarly, the scale elasticities are in a similar range but also only slightly correlated (Figure 5). Figure 6 demonstrates that the efficiency estimates of the semiparametric approach are generally highly correlated with estimates of a parametric SFA estimation using the Translog functional form. However, the semiparametric approach returns clearly higher efficiency levels than the parametric approach with the Translog functional form. No inefficiency can be identified in the parametric SFA with Cobb-Douglas functional form.

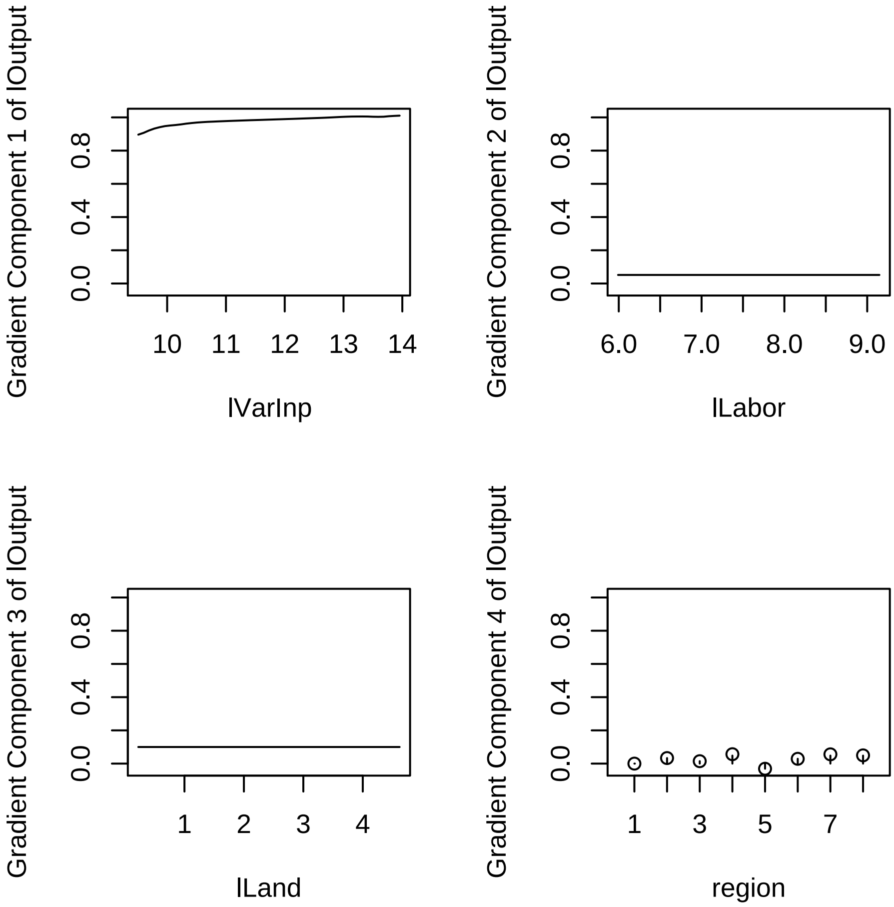

Figure 1: Partial production elasticities estimated with the semiparametric approach

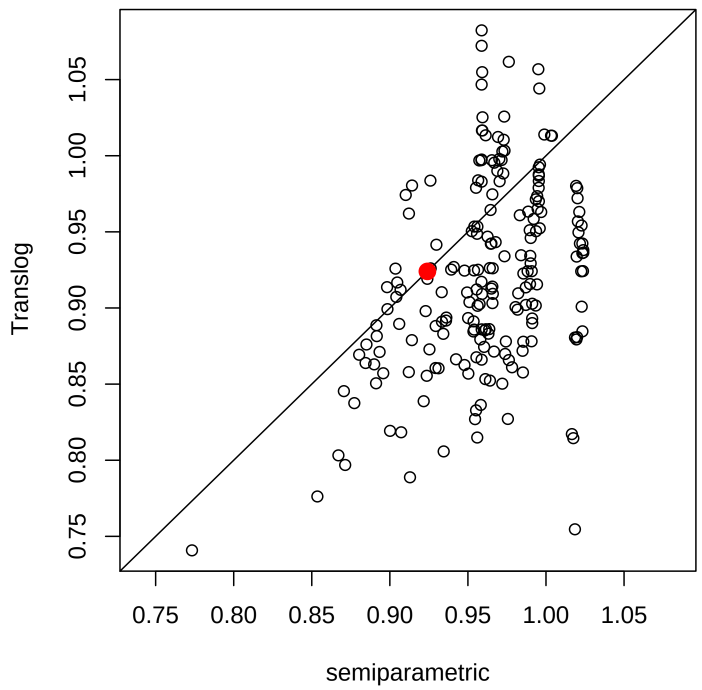

Figure 2: Partial production elasticities of intermediate inputs based on the semiparametric SFA and on the (parametric) SFA with Cobb-Douglas (red dot) and Translog functional form

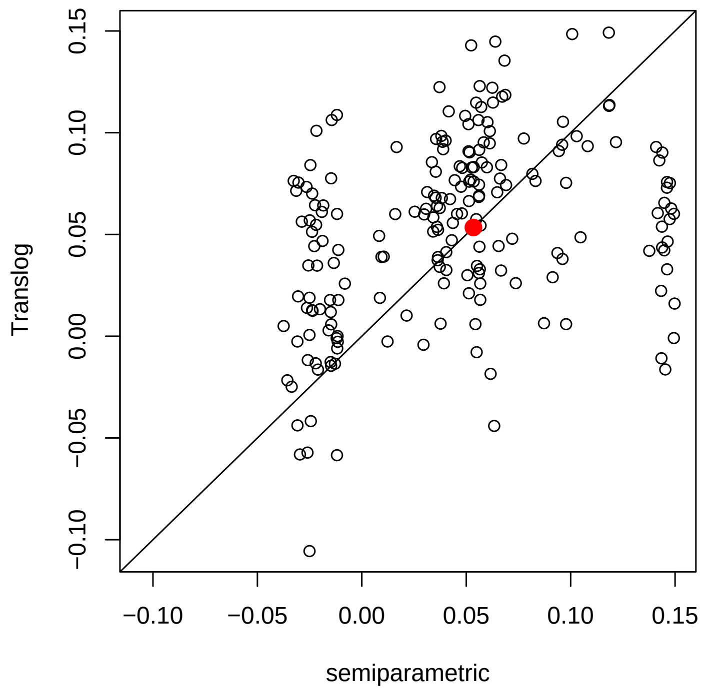

Figure 3: Partial production elasticities of labor based on the semiparametric SFA and on the (parametric) SFA with Cobb-Douglas (red dot) and Translog functional form

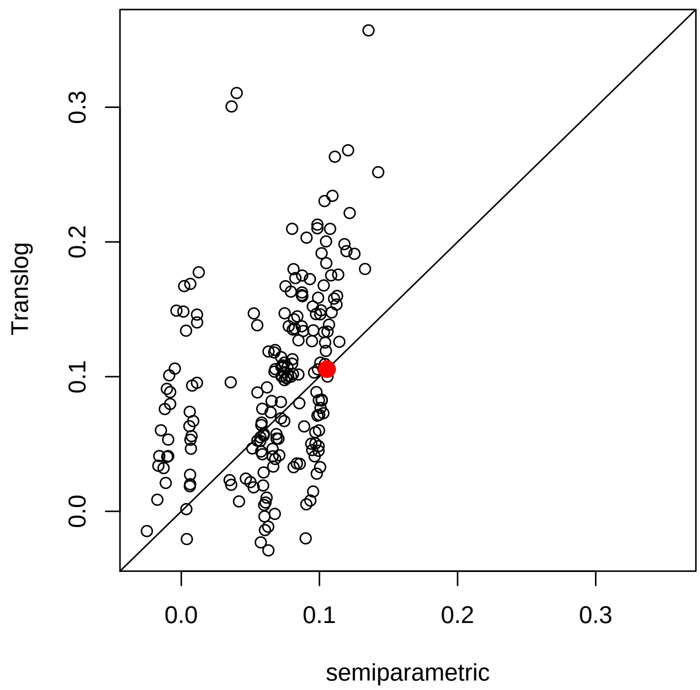

Figure 4: Partial production elasticities of land based on the semiparametric SFA and on the (parametric) SFA with Cobb-Douglas (red dot) and Translog functional form

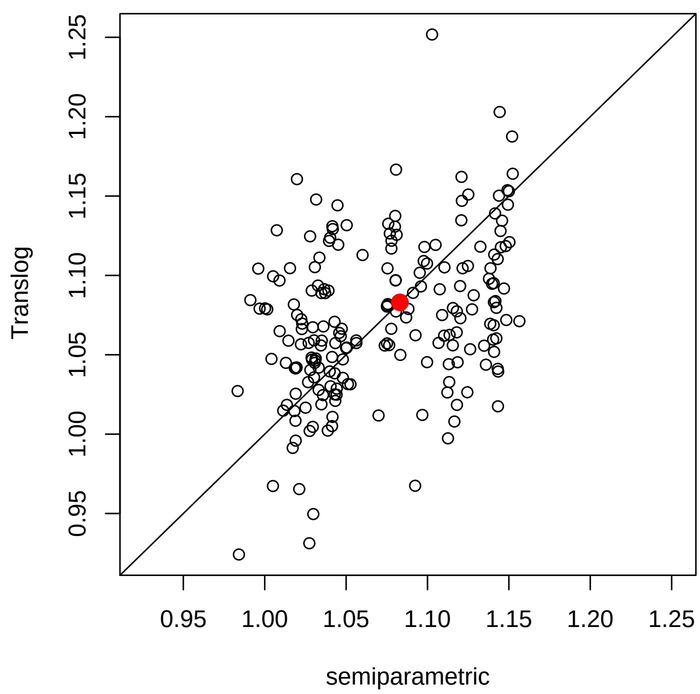

Figure 5: Elasticities of scale based on the semiparametric SFA and on the (parametric) SFA with Cobb-Douglas (red dot) and Translog functional form

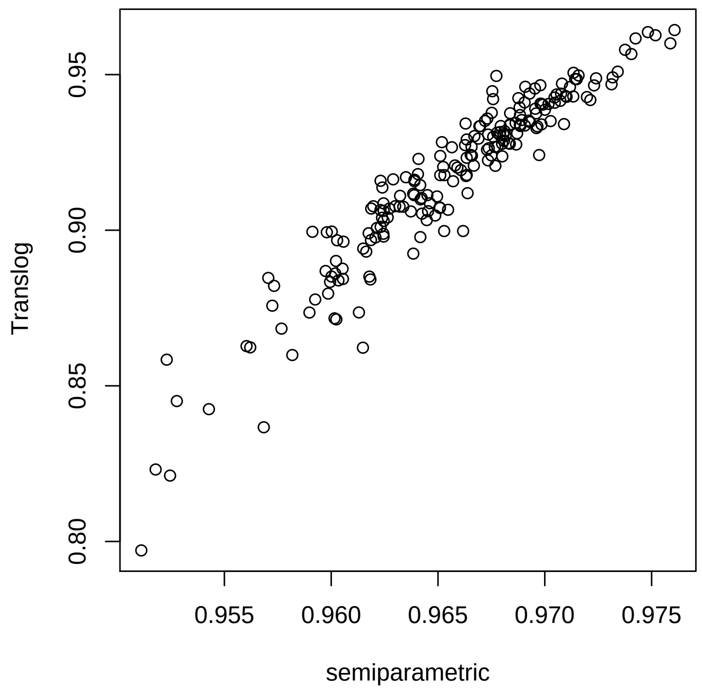

Figure 6: Efficiency estimates based on the semiparametric SFA and on the SFA with Translog functional form

6 Conclusions

Most results of the semiparametric SFA and the standard (parametric) SFA are rather similar. The efficiency estimates are even very highly correlated so that the ranking of the firms with respect to their efficiencies does not depend on the methodology. In contrast, the actual values of the efficiency estimates clearly depend on the methodology, where the estimates of the semiparametric SFA are in between the estimates of the parametric SFAs with Cobb-Douglas and translog functional form.

The efficiency estimates of Polish farms seem to be rather high. Therefore, inefficiencies cannot be the primary cause of the poor situation of many Polish farms. Instead, it seems that their low profit is caused by low remuneration of farm work, which is in turn caused by the extremely high labour intensity.

Acknowledgements

Arne Henningsen is grateful to the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) for financially supporting this research.

References

Aitchison, J. and Aitken, C. G. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika* 63: 413–420.

Borzutzky, S. and Kranidis, E. (2005). A struggle for survival: The Polish agricultural sector from communism to EU accession. *East European Politics and Societies* 19: 614–654.

Brada, J. C. and King, A. E. (1993). Is private farming more efficient than socialized agriculture? *Economica* 60: 41–56.

Brada, J. C. and King, A. E. (1994). Differences in the technical and allocative efficiency of private and socialized agricultural units in pre-transformation Poland. *Economic Systems* 18: 363–376.

Brooks, K., Guasch, J. L., Braverman, A. and Csaki, C. (1991). Agriculture and the transition to the market. *Journal of Economic Perspectives* 5: 149–161.

Brümmer, B., Glauben, T. and Thijssen, G. (2002). Decomposition of productivity growth using distance functions: The case of dairy farms in three European countries. *American Journal of Agricultural Economics* 84: 628–644.

Coelli, T. and Henningsen, A. (2008). frontier: Stochastic Frontier Analysis. R package version 0.9, `http://CRAN.R-project.org`.

Davidova, S., Gorton, M., Ratinger, T., Zawalinska, K. and Iraizoz, B. (2005). Farm productivity and profitability: A comparative analysis of selected new and existing EU member states. *Comparative Economic Studies* 47: 652–674.

Davidova, S., Gorton, M., Ratinger, T., Zawalinska, K., Iraizoz, B., Kovács, B. and Mizo, T. (2002). An Analysis of Competitiveness at the Farm Level in the CEECs. IDARA Working Paper 2/11, IDARA, available on `http://www.agp.uni-bonn.de/agpo/rsrch/idara/idara_e.htm`.

Dries, L. and Swinnen, J. F. M. (2002). Institutional reform and labor reallocation during transition: Theory evidence from Polish agriculture. *World Development* 30: 457–474.

European Commission (2002). Agricultural situation in the candidate countries – country report on Poland. Directorate-General for Agriculture, `http://ec.europa.eu/agriculture/external/enlarge/publi/countryrep/poland.pdf` [accessed 14-December-2006].

European Commission (2007). Rural development in the european union - statistical and economic information - report 2007. European Union, Directorate-General for Agriculture and Rural Development.

Fan, Y., Li, Q. and Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics* 14: 460–68.

FAO (2005). Compendium of food and agriculture indicators 2005. Food and Agriculture Organization of the United Nations, `http://www.fao.org/es/ess/compendium_2005/` [accessed 15-December-2006].

Góra, M., Radziwi, A., Sowa, A. and Walewski, M. (2006). Tax Wedge and Skills: Case of Poland in International Perspective. CASE Reports 64/2006, Center for Social and Economic Research, Warsaw.

Gorton, M. and Davidova, S. (2004). Farm productivity and efficiency in the CEE applicant countries: A synthesis of results. *Agricultural Economics* 30: 1–16.

GUS (2006). Employment in national economy in 2005. Glowny Urzad Statystyczny / Central Statistical Office of Poland, `http://www.stat.gov.pl/english/dane_spol-gosp/praca_ludnosc/pracujacy/2005/employment_2005.pdf` [accessed 14-December-2006].

GUS (2007). Agriculture in 2006. Glowny Urzad Statystyczny / Central Statistical Office of Poland, `http://www.stat.gov.pl/cps/rde/xbcr/gus/PUBL_agriculture_2006.pdf` [accessed 4-October-2007].

Hayfield, T. and Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software* 27: 1–32.

Henningsen, A. and Henning, C. H. C. A. (2008). Estimation of theoretically consistent stochastic frontier functions with a simple three-step procedure. Paper presented at the Fifth North American Productivity Workshop (NAPWV) in New York City, USA.

Hurvich, C. M., Simonoff, J. S. and Tsai, C. L. (1998). Smooting parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society Series B* 60: 271–293.

Kislev, Y. and Peterson, W. (1996). Economies of scale in agriculture: A reexamination of the evidence. In Antle, J. M. and Sumner, D. A. (eds), *The Economics of Agriculture. Papers in Honor of D. Gale Johnson*. University of Chicago Press, *2*, 156–170.

Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Latruffe, L., Balcombe, K., Davidova, S. and Zawalinska, K. (2005). Technical and scale efficiency of crop and livestock farms in Poland: Does specialistion matter? *Agricultural Economics* 32: 281–296.

Lerman, Z. (2002). Productivity and efficiency of individual farms in Poland: A case for land consolidation. Paper presented at the Annual Meeting of the American Agricultural Economics Association, Long Beach, July 28-31, 2001.

Lerman, Z. and Schreinemachers, P. (2005). Individual farming as a labour sink: Evidence from Poland and Russia. *Comparative Economic Studies* 47: 675–695.

Munroe, D. (2001). Economic efficiency in Polish peasant farming: An international perspective. *Regional Studies* 35: 461 – 471.

Munroe, D. K. (2000). Regional Variations in Polish Peasant Farm Efficiency: Composed Error, Spatial Econometric and Spatial Interaction Techniques. Ph.D. thesis, University of Illinois at Urbana-Champaign.

Pacuszka, R. (2005). Entwicklungsperspektiven polnischer Familienbetriebe unter Einfluss des EU-Beitritts: Eine Modellanalyse für typische Betriebe. Ph.D. thesis, University of Bonn.

Pouliquen, A. (2001). Competitiveness and Farm Incomes in the CEEC Agri-Food Sectors: Implications Before and After Accession for EU Markets and Policies. Tech. rep., European Commission, DG Agriculture and Rural Development, `http://ec.europa.eu/agriculture/publi/reports/ceeccomp/fullrep_en.pdf` [accessed 9-January-2007].

USDS (2006). Background note: Poland. U.S. Department of State, `http://www.state.gov/r/pa/ei/bgn/2875.htm` [accessed 12-December-2006].

Wilkin, J. (2007). Agriculture in new Member States - expectations and lessons learned. Plenary paper prepared for presentation at the 104th EAAE Seminar: Agricultural Economics and Transition "What was expected, what we observed, the lessons learned", Corvinus University of Budapest, Budapest, Hungary, September 6-8, 2007.

Zegar, J. S. and Floriańczyk, Z. (2003). Main problems of agriculture and rural areas in Poland in the period of transformation and integration with European Union. In Bański, J. and Owsiński, J. (eds), *Alternatives for European Rural Areas*. Warsaw: Institute of Agricultural and Food Economics & Institute of Geography and Spatial Organization, Polish Academy of Sciences, 7–18.

Zyl, J. van, Miller Jr., B. and Parker, A. (1996). The Agrarian Structure in Poland: The Myth of Large-Farm Superiority. Policy Research Working Paper 1596, The World Bank.