INSTITUTE OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

FOI Working Paper

Comparing Parametric and Nonparametric Regression Methods for Panel Data: the Optimal Size of Polish Crop Farms

Tomasz Gerard Czekaj
Arne Henningsen

2012 / 12

FOI Working Paper 2012 / 12

Comparing Parametric and Nonparametric Regression Methods for Panel Data: the Optimal Size of Polish Crop Farms

Authors: Tomasz Gerard Czekaj, Arne Henningsen

Institute of Food and Resource Economics

University of Copenhagen

Rolighedsvej 25

DK 1958 Frederiksberg DENMARK

www.foi.life.ku.dk

Comparing Parametric and Nonparametric Regression Methods for Panel Data: the Optimal Size of Polish Crop Farms

TOMASZ GERARD CZEKAJ and ARNE HENNINGSEN

Department of Food and Resource Economics, University of Copenhagen,
Rolighedsvej 25, 1958 Frederiksberg C, Denmark
e-mail: tcz@foi.ku.dk, arne@foi.ku.dk

October 2012

Abstract

We investigate and compare the suitability of parametric and non-parametric stochastic regression methods for analysing production technologies and the optimal firm size. Our theoretical analysis shows that the most commonly used functional forms in empirical production analysis, Cobb-Douglas and Translog, are unsuitable for analysing the optimal firm size. We show that the Translog functional form implies an implausible linear relationship between the (logarithmic) firm size and the elasticity of scale, where the slope is artificially related to the substitutability between the inputs. The practical applicability of the parametric and non-parametric regression methods is scrutinised and compared by an empirical example: we analyse the production technology and investigate the optimal size of Polish crop farms based on a firm-level balanced panel data set. A nonparametric specification test rejects both the Cobb-Douglas and the Translog functional form, while a recently developed nonparametric kernel regression method with a fully nonparametric panel data specification delivers plausible results. On average, the nonparametric regression results are similar to results that are obtained from the parametric estimates, although many individual results differ considerably. Moreover, the results from the parametric estimations even lead to incorrect conclusions regarding the technology and the optimal firm size.

Keywords: production technology, nonparametric econometrics, panel data, Translog, firm size, Polish crop farms

JEL codes: C14, C23, D24, Q12

1. Introduction

The optimal farm structure and the optimal farm size have been the subject of a long-standing debate among agricultural economists and policymakers (Gorton and Davidova, 2004), because both of these issues have significant policy implications, especially in transition and developing economies. In order to avoid misleading policy recommendations, the investigation of the optimal size of a farm (or any other firm) requires the use of proper empirical methods. Two different approaches are predominantly used in empirical studies of the optimal firm size and the relationship between firm size and productivity, namely the nonparametric deterministic Data Envelopment Analysis (DEA) and the econometric estimation of a parametric production technology. However, both approaches have their shortcomings.

In the parametric approach to empirical production analysis, the most common parametric specifications of production technologies are the Cobb-Douglas and the Translog (transcendental logarithmic) functional forms. In the case of the Cobb-Douglas functional form, the elasticity of scale is constant, which limits the inference regarding the optimal size to the sample mean. Although it is possible to derive observation-specific elasticities of scale from the Translog functional form, this should not be applied to determine the optimal firm size, because this functional form is only *locally* flexible. Moreover, we show in this paper that the Translog functional form implies a linear relationship between the (logarithmic) firm size and the elasticity of scale, where the slope is artificially related to the substitutability between the inputs. Additionally, parametric approaches to empirical production analysis generally have the problem that the *a priori* chosen functional form may not be sufficiently similar to the “true” relationship between the inputs and the outputs and that this misspecification may lead to biased estimates.

The DEA method solves this problem by using a nonparametric specification of the production technology. However, the natural randomness of agricultural production calls for the use of stochastic methods. Therefore, the use of stochastic regression methods is generally preferred to the use of the deterministic DEA in empirical studies of agricultural production technologies.

In order to solve the problems that are inherent in the above-mentioned approaches, we propose the use of stochastic nonparametric kernel regression, because this technique does not require the specification of a functional form and at the same time can account for randomness in the production process. Hence, this nonparametric stochastic regression technique can be applied to determine production technologies in order to obtain consistent estimates of elasticities of scale, and thus, to determine the optimal firm size without artificial parametric constraints. Recent developments in nonparametric kernel regression even allow for fully non-parametric panel data specifications. We demonstrate the usefulness of our approach and its advantages over previously applied approaches by analysing the returns to scale and the optimal size of Polish crop farms.

The next section presents a literature review. Section three briefly introduces the parametric and nonparametric approaches used in econometric production analysis. Section four describes the data used in this study. Section five delivers the results of the conducted analyses and section six concludes.

2. Literature review

Although there has been a long-standing debate about economies of scale in agriculture and the optimal farm size, no general agreement has yet been reached (Kislev and Peterson, 1996). Empirical results depend on the farm type, the country, and the method applied. Initiated by the seminal work of Sen (1962), there is a vast literature, particularly in the field of development economics, that suggests an inverse relationship between farm size and productivity (e.g. Carter, 1984; Barrett, 1996; Barrett, Bellemare and Hou, 2010, just to mention a few). Nonetheless, there are many contrary findings (e.g. Rao and Chotigeat, 1981; Barbier, 1984; Dorward, 1999; Sarris, Savastano and Tritten, 2004, among others) and even Amartya Sen argued that the inverse relationship should not be generalised (Rudra and Sen, 1980). However, it is generally agreed that the optimal farm size can be determined by empirical studies and that this should be promoted (e.g. Munroe, 2001). In Central and Eastern European Countries (CEECs), especially during the transition period, the discussion regarding the optimal farm size focussed on the differences in efficiency and productivity between small-scale farms and large-scale farms. In most empirical works (e.g. Hughes, 2000; Mathijs and Swinnen, 2001; Curtiss, 2000; Lerman and Schreinemachers, 2002), individual farms are classified as small-scale farms whereas cooperatives and state-owned farms are classified as large-scale farms. Hence, the comparison of efficiencies and productivities between small-scale farms and large-scale farms is rather an investigation of different governance structures than of different sizes. Thus, the optimal size of individual farms in CEECs has not yet been thoroughly analysed. Table 1 presents summary results from empirical studies of different farm types in Poland and of crop farms in other European countries. There is only a single study of Polish crop farms (Latruffe et al., 2004) but the results of this study are very questionable: the monotonicity condition is violated for most input variables and the estimated elasticity of scale is only 0.08, which is unrealistically low. In contrast, studies of other farm types (including studies of all farm types) in Poland consistently report increasing returns to scale of around 1.10. Studies of crop farms in other European countries also report mostly increasing returns to scale. Only German crop farms were found to have decreasing returns to scale. However, even within the countries with increasing returns to scale, there is a large variation in the magnitude of the returns to scale with estimated elasticities of scales varying between 1.03 and 1.39. The elasticities of scale summarised in Table 1 are derived at the sample mean. Hence, the results of these studies indicate that the average farm in Poland and the average crop farm in most EU countries are below their optimal scale levels, i.e. productivity could be increased by increasing size. However, these studies do not provide a comprehensive analysis of the optimal farm size. Therefore, this paper fills this gap and analyses the optimal size of individual crop farms in Poland.

Most studies that investigate the optimal firm size use a parametric production function or distance function of the Translog functional form and evaluate the elasticity of scale at different firm sizes. For instance, Oh, Lee and Heshmati (2008) analysed South Korea's manufacturing industry in the years 1993–2003 and found that estimated elasticities of scale considerably depended on the plant size, and the authors concluded that larger manufacturing plants (with more than 300 employees) should expand their size while smaller plants (with less than 300 employees) should reduce their size to reach the technically optimal size level. Aw and Hwang (1995) analysed Taiwan's electronics industry based on cross-sectional data. They estimated returns to scale for different sectors and different firm sizes and found that in the Data Storage sector and the Processing Units sector, the elasticity of scale fell from about 1.04 to about 0.94 as size increased. Noam (1983) analysed scale economies of cable television companies in 1980 and found substantial increasing returns to scale which even increased with the firm size measured by the number of subscribers. In the field of agriculture, Vlastuin, Lawrence and Quiggin (1982) used a Translog production function to investigate the optimal size of Australian farms. The authors calculated elasticities of scale for different size groups of farms and found that in the years 1966/1967 elasticities of scale were close to one implying constant returns to scale for all farms, while in 1976/1977 the elasticities of scale increased with the farm size (although they were not significantly different from one).

3. Parametric and Nonparametric Econometric Production Analysis

One of the most common approaches in applied economic production analysis is the econometric estimation of production functions. The idea of an algebraic relationship between inputs and output was developed in the eighteenth century in the works of Turgot and Malthus and in the nineteenth century in the works of Ricardo and von Thünen (Humphrey, 1997)[1]. Finally, Wicksteed (1894) was the first to explicitly use the concept of an algebraic production function while Cobb and Douglas (1928) were the first to use econometric techniques to estimate a production function.

[1] The production function proposed by von Thünen in 1840 is in fact the same, albeit in indirect form, as probably the most famous production function, the so-called Cobb-Douglas function (Humphrey, 1997).

Table 1: Recent empirical studies of farm productivity in Poland and other European countries

	Country	Time period	Farm type	Method	Partial output / distance elasticities of					EoS
					labour	land	capital	int. inputs	other	
Hockmann, Pieniadz and Goraj (2007)	PL	1994-2001	all types	TL ODF	0.19	0.16	0.08	0.66	.	1.09
Renner et al. (2009)	PL	1994-2001	all types	TL PF	0.18	0.20	0.10	0.62	.	1.10
Brümmer, Glauben and Thijssen (2002)	PL	1991-1994	dairy	TL ODF	0.16	0.21	0.02	0.71	.	1.10
Lerman (2002)	PL	2000	all types	CD PF	0.09	0.42	0.19	0.29	0.15	1.14
Latruffe et al. (2004)	PL	2000	crop	CD SFPF	0.17	-0.02	-0.17	-0.10	.	0.08
Latruffe et al. (2004)	PL	2000	livestock	CD SFPF	0.14	0.18	-0.11	0.86	.	1.07
Bakucs et al. (2010)	HU	2001-2005	all types	TL SFPF	0.32	0.18	0.12	0.41	.	1.03
Latruffe and Sauer (2010)	FR	1980-2006	crop	GL TF	0.17	0.20	0.16	0.47	0.50	1.07
Latruffe and Sauer (2010)	UK	1980-2006	crop	GL TF	0.19	0.21	0.18	0.33	0.36	1.03
Kumbhakar and Lien (2010)	NO	1991-2006	crop	TL SFPF	0.11	0.95	0.21	.	.	1.28
Heshmati and Kumbhakar (1997)	SE	1976-1988	crop	TL SFPF	0.30	0.20	0.18	0.71	.	1.25
Zhu and Lansink (2010)	DE	1995-2004	crop	TL ODF	-0.01	0.61	0.04	0.12	0.16	0.92
Zhu and Lansink (2010)	NL	1995-2004	crop	TL ODF	0.04	0.65	0.15	0.10	0.27	1.21
Zhu and Lansink (2010)	SE	1995-2004	crop	TL ODF	0.03	0.43	0.17	0.18	0.25	1.06
Rasmussen (2010)	DK	1985-2006	crop	TL IDF	0.31	0.21	0.09	0.24	.	1.39

Note that:

CD - Cobb-Douglas, TL - Translog (transcendental logarithmic), GL - Generalized Linear, PF - production function, TF - transformation function, SFPF - stochastic frontier production function, IDF - input distance function, ODF - output distance function,

EoS - Elasticity of scale.

The purpose of regression analysis is to evaluate the effects of one or more explanatory variables (e.g. inputs) on a single dependent variable (e.g. output). This is done by evaluating the conditional expectation of the dependent variable given the explanatory variables, which can be expressed as:

$$y_i = f(X_i) + \varepsilon_i, \tag{1}$$

where y_i is the observed dependent variable, X_i is a set of explanatory variables, $f(.)$ is the unknown regression function that returns the conditional expectation of the dependent variable $E[y_i|X_i]$, ε_i is a random error term, and i indicates the observation.

3.1. Parametric Approach

The specification of the functional form for $f(.)$ is one of the most crucial decisions in the parametric approach to econometric production analysis. Initially, the most widely used functional form in applied production economics was the Cobb-Douglas function:

$$\ln y_i = \alpha_0 + \sum_j \alpha_j \ln x_{ij}, \tag{2}$$

where x_{ij} indicates the quantity of the jth input and the αs are the coefficients.

Because of the strong assumptions that the Cobb-Douglas production function imposes on the underlying technology, Christensen, Jorgenson and Lau (1971) proposed a more flexible generalisation of the Cobb-Douglas function, the Translog (transcendental logarithmic) production function:

$$\ln y_i = \alpha_0 + \sum_j \alpha_j \ln x_{ij} + \frac{1}{2} \sum_j \sum_k \alpha_{jk} \ln x_{ij} \ln x_{ik} \tag{3}$$

with $\alpha_{jk} = \alpha_{kj}$.

These two functional forms have played a predominant role in applied production and efficiency analysis for the past 40 years. One important reason for this is that both functions are (after logarithmic transformation) linear in parameters and hence, can be estimated by simple linear regression techniques. In the case of cross-sectional data (several firms observed at a single period of time), these functional forms can be estimated by the ordinary least squares (OLS) method. If several observations are available for each firm, the usual panel data methods such as the fixed effects (FE) estimator or the random effects (RE) estimator can be suitable, as they can account for individual or time specific heterogeneity. The Hausman test (Hausman, 1978) can be used to determine whether the RE estimator is consistent, and if it is, it should be used as it is more efficient than the FE estimator.

Because the selection of the functional form for modelling the relationship between the inputs and the output is rather arbitrary, the *a priori* specification of a functional form runs the risk of choosing a functional form that is not similar to the "true" relationship between the inputs and the output. This misspecification can result in biased parameter estimates and hence, also in biased measures which are derived from the parameters, such as marginal products, partial output elasticities, and elasticities of scale. Moreover, various statistical tests become incorrect. This problem has been addressed by introducing flexible functional forms. However, most flexible functional forms such as the Translog functional form are only *locally* flexible, i.e. they are only flexible at a single point, the approximation point, which is usually the sample mean.

Although it is possible to obtain individual elasticities of scale for each observation (firm) from *locally* flexible functional forms, we argue that this approach should not be used to investigate the optimal firm size, because *locally* flexible functional forms may be too inflexible to appropriately model the production technology of firms that are larger or smaller than the sample mean and thus, measures such as elasticities are only valid at the sample mean.

The Translog functional form possesses a further shortcoming in that if all inputs are changed proportionally by the scale factor $s > 0$, the elasticity of scale is linearly related to the logarithm of the firm size, where the slope is equal to the sum of the second-order coefficients:

$$\frac{\partial \epsilon}{\partial \ln s} = \sum_j \sum_k \alpha_{jk} \tag{4}$$

Thus, the elasticity of scale changes linearly with the logarithm of the firm's scale.[2]

As the second-order coefficients of the Translog production function determine, among other things, the substitutability between the inputs, the application of the Translog functional form imposes an artificial relationship between input substitutability and the effect of the firm size on the elasticity of scale. Table 2 presents the slopes of the elasticities of scale with respect to the (logarithmic) farm size that we derived from the results of some previously mentioned studies that used the Translog functional form. In two thirds of the studies, this slope is positive, which implies that the derived elasticities of scale increase with farm size.

[2] In case of a Translog output distance function, the elasticity of scale also changes linearly with the logarithm of the firm's scale, where the slope is $\partial\epsilon/\partial \ln s = -\sum_j \sum_k \alpha_{jk}$ and the α_{jk}s denote the second-order coefficients of the input quantities. In the case of a Translog input distance function, the *inverse* elasticity of scale changes linearly with the logarithm of the firm's scale, where the slope is $\partial\epsilon^{-1}/\partial \ln s = -\sum_j \sum_k \beta_{jk}$ and the β_{jk}s denote the second-order coefficients of the output quantities. This means that the elasticity of scale of the Translog input distance function changes non-linearly with the logarithm of the firm's scale, where the (non-constant) slope is $\partial\epsilon/\partial \ln s = \epsilon^2 \sum_j \sum_k \beta_{jk}$. The proofs are presented in appendix B and substantially draw on Ray (1998) who showed that the elasticities of scale derived from a Translog production function are constant over all firms, if the second-order coefficients sum to zero, i.e. the Translog production function is homothetic.

Table 2: Effect of farm size on the elasticity of scale derived from selected studies listed in Table 1

	Country	$\partial\epsilon/\partial \ln s$
Hockmann, Pieniadz and Goraj (2007)	PL	0.01
Renner et al. (2009)	PL	0.05
Brümmer, Glauben and Thijssen (2002)	PL	-0.20
Bakucs et al. (2010)	HU	0.04
Kumbhakar and Lien (2010)	NO	0.25
Zhu and Lansink (2010)	DE	0.02
Zhu and Lansink (2010)	NL	-0.27
Zhu and Lansink (2010)	SE	0.02
Rasmussen (2010)	DK	-0.30

Note: Zhu and Lansink (2010) used the quadratic terms $(\ln x_j)^2$ as regressors rather than $1/2(\ln x_j)^2$ as given in the Translog specification. Therefore, we had to multiply the coefficients of the quadratic terms by two in order to obtain the α_{jj} coefficients of the Translog function.

Globally flexible functional forms such as the Fourier flexible functional form (Gallant, 1982) or methods employing neural networks (Michaelides, Vouldis and Tsionas, 2010) solve the problems that we described for non-flexible or only *locally* flexible functional forms. However, in this paper, we use a different approach: nonparametric regression.

3.2. Nonparametric Approach

Non-parametric approaches do not require a parametric specification of the functional relationship between the explanatory variables and the dependent variable. Hence, they avoid a possible misspecification of the functional form. The predominant nonparametric method in applied production analysis is the deterministic Data Envelopment Analysis (DEA) introduced by Charnes, Cooper and Rhodes (1978), while stochastic nonparametric methods such as non-parametric regression analysis are very rarely used. Stochastic methods are particularly advantageous in stochastic environments such as agricultural crop production, where random weather conditions have an important influence on the outcome. Therefore, we apply a stochastic and nonparametric local-linear kernel estimator in this study. One can think of this estimator as a set of weighted linear regressions, where a weighted linear regression is performed at each observation and the weights of the other observations decrease with the distance from the respective observation. The weights are determined by a kernel function and a set of bandwidths, where a bandwidth for each explanatory variable must be specified. The smaller the bandwidth, the faster the weight decreases with the distance from the respective observation. In our study, we make the frequently used assumption that the bandwidths can differ between regressors but are constant over the domain of each regressor. While the bandwidths were initially determined by using a rule of thumb, nowadays increased computing power allows us to select the optimal bandwidths for a given model and data set according to the expected

Kullback-Leibler cross-validation criterion (Hurvich, Simonoff and Tsai, 1998). Hence, in nonparametric kernel regression, the overall shape of the relationship between the inputs and the output is determined by the data and the (marginal) effects of the explanatory variables can differ between observations without being restricted by an arbitrarily chosen functional form.

The first applications of stochastic nonparametric kernel methods in production analysis, mainly in the stochastic frontier framework, are Vinod and Ullah (1988), Kneip and Simar (1996) and Fan, Li and Weersink (1996). More recent applications can be found in Henderson and Simar (2005) and Kumbhakar et al. (2007).

In agricultural economics, the most recent studies which have employed stochastic nonparametric regression methods are Kumbhakar and Tsionas (2009) who estimated the production function of Norwegian salmon farms, Livanis, Salois and Moss (2009) who analysed the crop response production function for corn production in Illinois and Indiana, and Verschelde et al. (forthcoming) who investigated the relationship between farm productivity and farm size in Burundi.

Although parametric regression methods still constitute the standard for estimating panel data models, the popularity of nonparametric regression methods for panel data sets has recently increased (e.g. Porter, 1996; Lin and Carroll, 2000; Wang, 2003; Henderson and Ullah, 2005; Henderson, Carroll and Li, 2008). Initially, nonparametric kernel regression methods usually used the Epanechnikov kernel or the Gaussian kernel and could only include continuous explanatory variables. In order to account for the panel structure, these nonparametric studies usually applied data transformation methods that were developed for parametric panel data regression such as the "within" or "between" transformations or first differencing (e.g. Henderson, Carroll and Li, 2008). However, these data transformations are not valid in the nonparametric framework unless one assumes additive separability of the individual and time effects, which would make the use of nonparametric regression questionable in many applications (Racine, 2008). Furthermore, these data transformation methods result in a considerable loss of observations (degrees of freedom), particularly when the time dimension of the panel data set is rather short. This is a substantial drawback when using nonparametric regression, because these methods demand a large number of observations.

Therefore, we follow Henderson and Simar (2005) and Racine (2008) who estimate a fully nonparametric two-ways panel data model by applying the nonparametric regression method proposed by Li and Racine (2004) and Racine and Li (2004) that can handle both continuous and categorical explanatory variables. The authors use time and firm id as (ordered and unordered) categorical explanatory variables so that not only the level of the dependent variable ("intercept") but also the effects of the explanatory variables on the dependent variable ("slopes") may differ between time periods and individual firms. This advantage makes the method very useful in applied production analysis, because it allows us to estimate observation-specific measures such as elasticities and marginal products, etc. which are not artificially constrained by an arbitrarily chosen functional form.

The statistical significance levels of the explanatory variables can be obtained by bootstrapping using the methods proposed by Racine (1997) and Racine, Hart and Li (2006).

Furthermore, nonparametric specification tests can be used to scrutinise the functional form of parametrically estimated (production) models. In most applications of kernel based regression models, commonly used functional forms (i.e. Cobb-Douglas and Translog) were rejected in favour of a nonparametric specification (e.g. Verschelde et al., forthcoming; Livanis, Salois and Moss, 2009) and the results (such as elasticities) obtained by nonparametric methods were economically more intuitive (e.g. Kumbhakar and Tsionas, 2009).

4. Data

In this study, we use balanced panel data from the Polish Farm Accountancy Data Network (FADN) which consists of 342 farms specialising in crop production[3] in the period 2004 to 2010. Hence, our data set includes 2394 observations in total.

The dependent variable of the production function is the farms' output which is measured as the value of the total agricultural production. Four inputs are used in the regression analyses: labour, land, intermediate inputs, and capital. Labour is measured by Annual Work Units (AWU), where 1 AWU equals 2200 hours of work. Total utilised agricultural area in hectares is used as a measure of land input. Intermediate inputs are measured as the sum of total farming overheads (e.g. maintenance, energy, services, other direct inputs) and specific costs (e.g. fertilisers, pesticides, seeds). Capital input is measured as the value of total fixed assets excluding the value of land. Since data on total agricultural production, intermediate inputs, and capital are expressed monetarily in current Polish Złoty (PLN), these data were deflated by national price indices published by the Central Statistical Office of Poland (GUS, 2012b).[4] Descriptive statistics of the regression variables are presented in Table 3. It is worth stressing that the crop farms in our data set are individual farms that mainly use family labour. The farm size varies from 7.3 ha to 632.0 ha with an average of 82.9 ha, while the average size of individual

[3] We selected all farms that are classified as specialised producers of cereals, oilseed and protein crops (according to the FADN's methodology) for at least 5 out of the 7 years covered in the data set. As the FADN typology classifies farms according to the expected gross margin ("standard gross margin") of the different types of production rather than the actual production, we only included farms that have a share of actual (observed) crop production in total agricultural output higher than 50% in each year of the sample.

[4] The value of agricultural production is deflated by the price index of agricultural production; the value of variable inputs is deflated by the price index of purchased goods and services for current agricultural production; and the value of the capital stock is deflated by the price index of purchased goods and services for investment.

farms (with an area exceeding 1 ha of arable land) increased from 7.5 ha in 2004 to 8.6 ha in 2010 (GUS, 2012a).

Table 3: Descriptive statistics of regression variables

Variable	Min	Median	Mean	Max	Std. dev.
Output (Y) [in k PLN]	6.34	125.40	188.40	1563.00	188.61
Labour (L) [in AWU]	0.19	1.55	1.66	6.81	0.77
Land (A) [in ha]	7.31	59.30	82.86	632.00	79.25
Intermediate inputs (V) [in k PLN]	4.42	75.94	113.20	1167.00	112.99
Capital stock (K) [in k PLN]	5.52	233.40	354.00	2160.00	334.37

5. Results

All estimations and calculations were conducted within the statistical software environment "R" (R Development Core Team, 2012) using the add-on packages "plm" (Croissant and Millo, 2008) for panel data estimations and the add-on package "np" (Hayfield and Racine, 2008) for nonparametric regression and specification tests.

5.1. Parametric Approach

The Cobb-Douglas and Translog production functions were both estimated with three different estimators: fixed effects (FE), random effects (RE), and pooled OLS (i.e. ignoring the panel structure of the data). Diagnostic tests for both functional forms are presented in Table 4. For both functional forms, a Hausman test shows that the RE model is inconsistent. F-tests indicate that both individual effects and time specific effects are statistically significant so that we can reject the pooled OLS models, the one-way individual FE models, and the one-way time FE models in favour of the two-ways FE models. A Wald test which was used to compare the two functional forms clearly rejects the Cobb-Douglas FE model in favour of the Translog FE model. The regression error specification test (RESET) proposed by Ramsey (1969) rejects the linearity (in logs) of the Cobb-Douglas model but accepts the functional form of the Translog model (P-value = 0.112). Furthermore, we apply the nonparametric specification test described in Hsiao, Li and Racine (2007) to check whether the functional forms used in the two parametric models (Cobb-Douglas and Translog) are consistent with the "true" relationship between the inputs and the output in our data set, i.e. whether the dependent variables are indeed linear in all regressors. This test shows that neither the Cobb-Douglas FE specification nor the Translog FE specification is consistent with the data. Hence, while the Cobb-Douglas functional form is clearly rejected by the Wald test, Ramsey's RESET test, and the nonparametric specification test, the Translog functional form is accepted by

Table 4: Results of diagnostic tests for Cobb-Douglas and Translog production functions

Test	Function	Statistics	Decision
Poolabillity test (FE vs. Pooled)	CD	$F(347, 2042) = 6.538$, $p < 0.001$	Pooled model rejected
	TL	$F(347, 2032) = 6.451$, $p < 0.001$	Pooled model rejected
2-way FE vs. Individual FE	CD	$F(6, 2042) = 39.470$, $p < 0.001$	Signif. time effects
	TL	$F(6, 2032) = 5.988$, $p < 0.001$	Signif. time effects
2-way FE vs. Time FE	CD	$F(341, 2042) = 6.089$, $p < 0.001$	Signif. indiv. effects
	TL	$F(341, 2032) = 5.988$, $p < 0.001$	Signif. indiv. effects
Hausman test (RE vs. FE)	CD	$\chi^2(4) = 121.808$, $p < 0.001$	RE model rejected
	TL	$\chi^2(14) = 124.391$, $p < 0.001$	RE model rejected
Wald test (TL FE vs. CD FE)	TL vs. CD	$\chi^2(10) = 35.561$, $p < 0.001$	CD FE rejected
RESET test (linear specification)	CD	$RESET(8, 2381) = 3.687$, $p < 0.001$	CD FE rejected
	TL	$RESET(8, 2381) = 3.687$, $p = 0.112$	TL FE accepted
Nonparametric model specification test	CD	$Jn = 1.017$, $p = 0.015$	CD FE rejected
	TL	$Jn = 0.882$, $p = 0.018$	TL FE rejected

Ramsey's parametric RESET test, but rejected by the nonparametric specification test. The estimation results of the most suitable parametric model, the two-ways FE Translog production function, are presented in Table 5.[5] The sum over all elements of the matrix of second-order coefficients (see equation 4) is 0.014.[6] This indicates that the elasticity of scale approximately increases by 0.012 units if all input quantities increase by 1%.

5.2. Nonparametric Approach

In our nonparametric model, we use the logarithm of the output quantity as the dependent variable and the logarithms of the input quantities as continuous explanatory variables. Using the logarithms of these variables has two advantages. First, given the right-skewed distribution of the output and input quantities, the logarithms of these variables are more evenly distributed, which is desirable when using fixed (constant) bandwidths.[7] Second, we can interpret the marginal effects (gradients) of the logarithmic input quantities on the logarithmic output quantity as partial output elasticities and the sum of these gradients as elasticity of scale.

[5] The results of the other parametric models are presented in appendix tables A1 and A2.

[6] Please note that the coefficients of the interaction terms that are presented in table 5 have to be added twice, because they correspond to the off-diagonal elements of the symmetric coefficient matrix and hence, appear twice in this matrix.

[7] Without taking the logarithm, there would be many observations within the bandwidths for small values of the explanatory variables (small farms), but only very few observations within the bandwidths for large values of the explanatory variables (large farms) so that the regression function would risk becoming over-smoothed for small farms and/or under-smoothed for large farms.

Table 5: Results of parametric fixed-effects estimation of Translog production function

Regressor	Estimate	Std. Error	t value	P-Value
$\ln(L)$	-0.101	0.168	-0.600	0.548
$\ln(A)$	0.227	0.198	1.150	0.250
$\ln(V)$	0.519	0.139	3.734	0.000
$\ln(K)$	0.067	0.108	0.620	0.536
$1/2 \ln(L)^2$	-0.107	0.070	-1.534	0.125
$\ln(L)\ln(A)$	0.066	0.061	1.070	0.285
$\ln(L)\ln(V)$	-0.068	0.054	-1.248	0.212
$\ln(L)\ln(K)$	0.040	0.030	1.328	0.184
$1/2 \ln(A)^2$	-0.146	0.076	-1.911	0.056
$\ln(A)\ln(V)$	0.254	0.058	4.385	0.000
$\ln(A)\ln(K)$	-0.044	0.032	-1.381	0.167
$1/2 \ln(V)^2$	-0.289	0.063	-4.554	0.000
$\ln(V)\ln(K)$	0.025	0.030	0.835	0.404
$1/2 \ln(K)^2$	0.008	0.021	0.386	0.699
$R^2 = 0.406$				

As we chose the two-ways fixed effects model for the parametric estimation, we used the IDs of the individual farms and the time (year) as additional (categorical) explanatory variables in the nonparametric regression so that both modelling approaches are comparable. While the ID of the individual farm is clearly an unordered categorical variable, we estimated the nonparametric model both with the year as an unordered categorical variable (assuming that the effects of successive years are unrelated due to random weather events) and with the year as an ordered categorical variable (assuming that there is a clear trend over time).

In order to find the most suitable kernels for the continuous and categorical variables and the most suitable specification of the time variable, we estimated our nonparametric model with all 16 combinations of the different kernels and the different specifications of the time variable. We chose the nonparametric model, where the year is modelled as an ordered categorical variable, the second-order Epanechnikov kernel is used for the continuous regressors (i.e. the four input variables), the kernel proposed by Wang and van Ryzin (1981) is used for the ordered categorical explanatory variable (year) and the kernel proposed by Li and Racine (2003) is used for the unordered categorical variable (farm ID) because in this specification all explanatory variables have a notable and statistically significant effect. The estimation results of the chosen nonparametric model are summarised in Table 6. The bandwidths of the continuous explanatory variables are very large, which indicates that the relationship between each logarithmic input quantity and the logarithmic output quantity—holding the other explanatory variables constant—is approximately linear. Similarly, the bandwidth of the time variable is virtually at its upper bound, which indicates that—everything else being constant—the effect of time on the output is very limited. However, even if the bandwidths are very large, non-parametric local-linear kernel regression allows for interaction effects between the regressors so that the model is not necessarily linear (Cobb-Douglas in our application) and independent of time, because the effect of one regressor on the dependent variable may still depend on the values of the other regressors. Therefore, our estimated partial output elasticities (i.e. the gradients of the continuous variables) and the rate of technological change (i.e. the gradients of the time variable) vary considerably between farms (see table 6). On average, slight technological regress of 0.3% per year has occurred but there is considerable variation between observations with almost half of the observations (45%) exhibiting technological progress. The productivity particularly increased in the years 2007 and 2010. The large variation in the gradients of the ID variable indicates considerable productivity differences between farms.

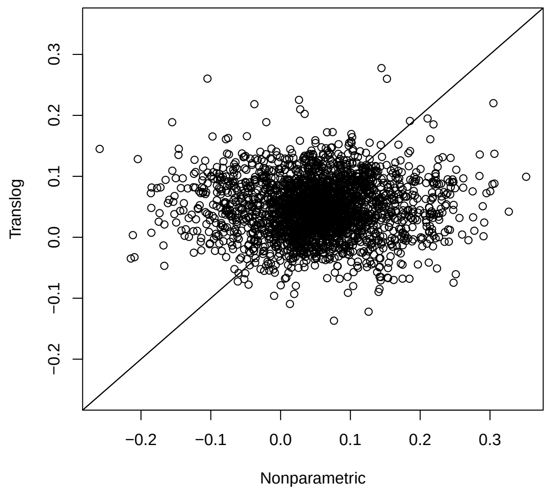
Table 6: Results of the nonparametric local-linear kernel regression

Regressor	Bandwidth	Gradients		P-Value
		Mean	Std. Dev	
$\ln(L)$	1047168	0.054	0.079	<0.001
$\ln(A)$	969597	0.147	0.087	<0.001
$\ln(V)$	366895	0.785	0.094	<0.001
$\ln(K)$	399621	0.084	0.048	<0.001
year (ordered)	1.000	-0.003	0.032	0.003
ID (unordered)	0.006	—	0.129	<0.001
$R^2 = 0.956$				

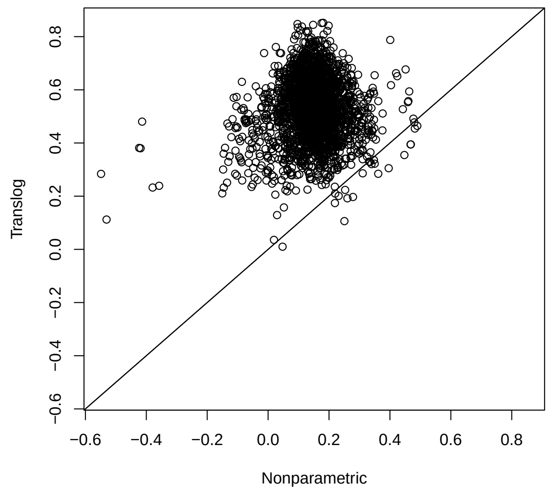
5.3. Comparison of Parametric and Nonparametric Results

In order to collate the results of the parametric Translog model and the chosen nonparametric model, we compare the partial production elasticities of each of the four inputs as well as the elasticities of scale in figures 1 and 2, respectively. The elasticities of the individual observations differ considerably between the parametric and the nonparametric model. Interestingly, there is not even any considerable correlation between the elasticities based on the Translog function and the elasticities based on the nonparametric production function. The monotonicity condition is partly violated in both parametric and nonparametric models.

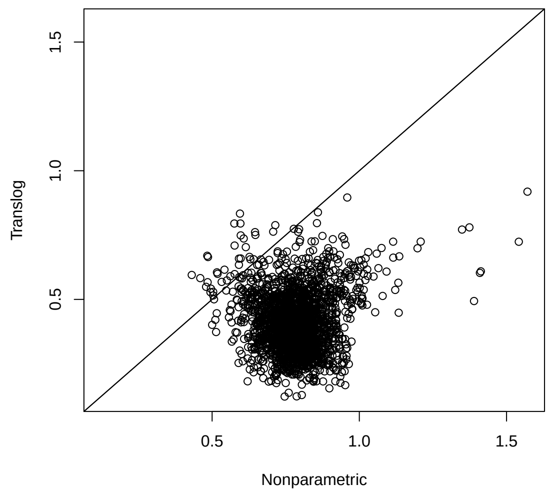
We aim to identify the optimal farm size by investigating the relationship between farm size and the elasticity of scale. Figure 3 presents the relationship between farm size (in



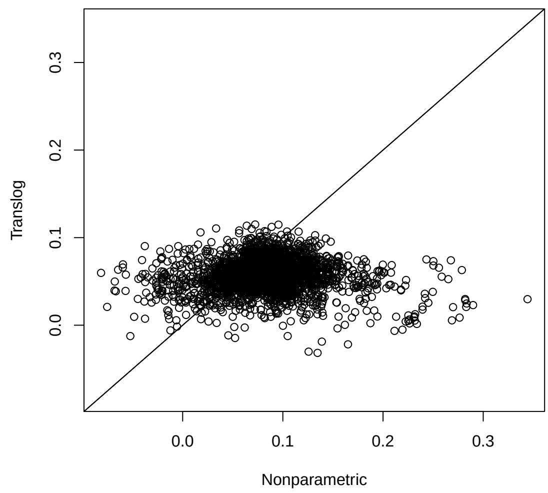
(a) Labour



(b) Land



(c) Intermediate inputs



(d) Capital

Figure 1: Partial output elasticities of labour (1a), land (1b), intermediate inputs (1c) and capital (1d) based on the FE Translog model and the nonparametric kernel regression

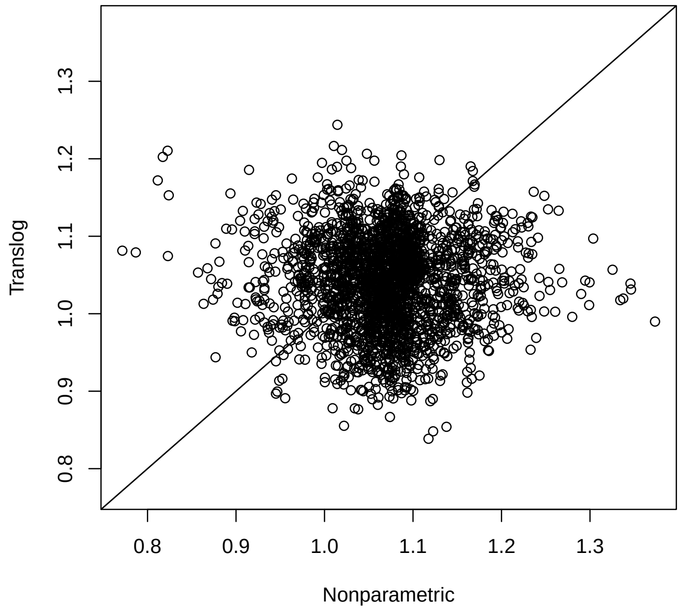
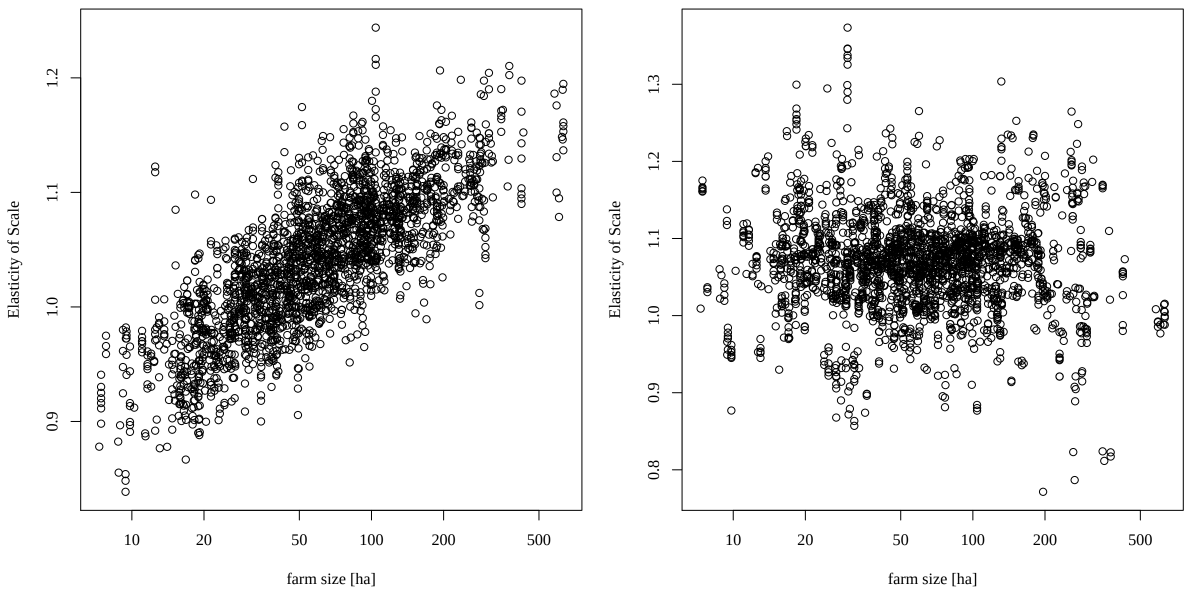


Figure 2: Elasticities of scale based on the FE Translog model and the nonparametric kernel regression



(a) Translog (b) Nonparametric

Figure 3: Elasticities of scale for different farm sizes derived from the FE Translog model (3a) and the nonparametric kernel regression (3b)

ha, on a logarithmic scale) and the elasticities of scale obtained from our parametric and nonparametric estimations. Both models show that most farms operate under increasing returns to scale. The elasticities of scale obtained from the nonparametric model do not significantly depend on the farm size. Both small farms as well as large farms usually exhibit increasing returns to scale and hence, could increase their productivity by increasing their size. The largest farms in our sample (i.e. farms with more than 500 ha) have approximately constant returns to scale. Thus, the results of our nonparametric regression suggest that the optimal farm size is at least as large as the largest farms in our sample. On the contrary, the results obtained with the Translog functional form are rather surprising. According to the results of the Translog model, most farms with less than 50 ha would become less productive if they increased their size, while most farms with more than 50 ha would become more productive if they increased their size. Furthermore, the Translog model implies that the largest farm in the sample still has large economies of scale, which would indicate that the optimal farm size is much larger than the largest farms in our sample. The positive and nearly linear relationship between the (logarithmic) farm size and the elasticity of scale seems to be implausible and is an artefact of our estimated Translog model rather than the truth, as we showed that the Translog functional form implies a linear relationship between the scale of the farm and the elasticity of scale. Therefore we found that the use of this functional form to investigate the optimal firm size is rather limited. In contrast, the nonparametric model does not impose this artificial relationship and gives plausible results.

6. Conclusion

We propose to increase the use of nonparametric econometric methods in empirical production analysis. First, these methods can be applied to verify the functional form used in parametric estimations of production functions. Second, they can be directly used to estimate production functions without the specification of a functional form and hence, they can avoid possible misspecification errors.

We found that the two functional forms, which are most widely used in empirical production analysis, i.e. the Cobb-Douglas and the Translog functional form, are both inconsistent with the "true" relationship between the inputs and the output in our specific data set. Furthermore, we showed that the Translog production function implies an implausible relationship between farm size and the elasticity of scale, where the slope is artificially related to the substitutability between the inputs. Thus, the Translog functional form seems to be unsuitable for analysing the optimal firm size.

We solved this problem by using nonparametric regression. This approach delivered reasonable results, which were on average not too dissimilar from the results of the parametric estimations, although many individual results were rather different. This indicates that parametric regression methods are suitable for investigating the average properties of the production technology but are too restrictive for investigating the properties at individual observations.

We applied the nonparametric regression method proposed by Li and Racine (2004) and Racine and Li (2004) that can handle both continuous and categorical explanatory variables. This approach allowed us to include the time and the farm ID as categorical explanatory variables to estimate a two-ways nonparametric panel data model that accounts for the panel structure of our data set in a fully nonparametric way without any data transformation or loss of observations.

Based on the results from the nonparametric approach, we conclude that the optimal size of Polish individual crop farms is at least as large as the largest farms in our sample, i.e. around 630 ha.

Acknowledgements

We would like to thank Jeff Racine for his many very helpful comments and suggestions regarding nonparametric regression, Subal Kumbhakar for encouraging us to use nonparametric kernel regression in econometric production analysis, Mette Asmild and the participants of the Asia-Pacific Productivity Conference in Bangkok 2012 for helpful suggestions on an earlier draft of this paper, and Lech Goraj as well as the Agricultural Accountancy Department of IAFE-NRI (IERiGŻ-PIB) in Warsaw for providing access to the data. We are also grateful to the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) for financially supporting this research. Of course, all remaining errors are the sole responsibility of the authors.

Appendix

A. Tables with detailed results

Table A1: Results of parametric regressions with Cobb-Douglas functional form

	Pooled	Random Effects	Fixed Effects
$\ln(L)$	0.052**	0.068**	0.064*
	(0.018)	(0.022)	(0.026)
$\ln(A)$	0.121***	0.388***	0.507***
	(0.016)	(0.023)	(0.032)
$\ln(V)$	0.815***	0.548***	0.448***
	(0.017)	(0.019)	(0.022)
$\ln(K)$	0.080***	0.076***	0.047**
	(0.009)	(0.013)	(0.016)
(Intercept)	0.337***	0.427***	–
	(0.040)	(0.076)	–
R^2	0.905	0.700	0.396

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table A2: Results of parametric regressions with Translog functional form

	Pooled	Random Effects	Fixed Effects
$\ln(L)$	0.088	−0.036	−0.101
	(0.129)	(0.149)	(0.168)
$\ln(A)$	0.113	0.330*	0.227
	(0.121)	(0.150)	(0.198)
$\ln(V)$	0.981***	0.643***	0.519***
	(0.119)	(0.127)	(0.139)
$\ln(K)$	0.102	0.077	0.067
	(0.076)	(0.092)	(0.108)
$1/2 \ln(L)^2$	0.032	−0.084	−0.107
	(0.065)	(0.065)	(0.070)
$\ln(L)\ln(A)$	0.038	0.069	0.066
	(0.051)	(0.055)	(0.061)
$\ln(L)\ln(V)$	−0.103†	−0.083	−0.068
	(0.053)	(0.051)	(0.054)
$\ln(L)\ln(K)$	0.044	0.038	0.040
	(0.028)	(0.028)	(0.030)
$1/2 \ln(A)^2$	−0.319***	−0.226***	−0.146†
	(0.064)	(0.069)	(0.076)
$\ln(A)\ln(V)$	0.282***	0.269***	0.254***
	(0.056)	(0.055)	(0.058)
$\ln(A)\ln(K)$	0.018	−0.035	−0.044
	(0.025)	(0.029)	(0.032)
$1/2 \ln(V)^2$	−0.236***	−0.285***	−0.289***
	(0.064)	(0.060)	(0.063)
$\ln(V)\ln(K)$	−0.052†	0.007	0.025
	(0.027)	(0.028)	(0.030)
$1/2 \ln(K)^2$	0.022	0.019	0.008
	(0.017)	(0.019)	(0.021)
(Intercept)	−0.050	0.371	–
	(0.220)	(0.326)	–
R^2	0.907	0.701	0.406

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

B. Elasticity of scale and firm's scale in Translog production technologies

B.1. Translog production function

For the Translog production function defined in (3), we get the following partial output elasticities (ignoring the subscript i that indicates the observation):

$$\epsilon_j = \frac{\partial \ln f(x)}{\partial \ln x_j} = \alpha_j + \sum_k \alpha_{jk} \ln x_k \tag{5}$$

and the following elasticity of scale:

$$\epsilon = \sum_j \epsilon_j = \sum_j \alpha_j + \sum_j \sum_k \alpha_{jk} \ln x_k \tag{6}$$

Scaling all input quantities in the Translog production function (3) by a factor $s > 0$, the elasticity of scale of the "scaled firm" is:

$$\epsilon = \sum_j \alpha_j + \sum_j \sum_k \alpha_{jk} \ln(s\ x_k) = \sum_j \alpha_k + (\ln s) \sum_j \sum_j \alpha_{jk} + \sum_j \sum_k \alpha_{jk} \ln x_k \tag{7}$$

Hence, the effect of the (logarithmic) scale on the elasticity of scale is:

$$\frac{\partial \epsilon}{\partial \ln s} = \sum_j \sum_k \alpha_{jk} \tag{8}$$

Thus, the elasticity of scale changes linearly with the logarithm of the firm's scale.

B.2. Translog output distance function

Given the Translog output distance production with N inputs and M outputs:

$$\begin{aligned} \ln D = \alpha_0 &+ \sum_{j=1}^{N} \alpha_j \ln x_j + \frac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \alpha_{jk} \ln x_j \ln x_k \\ &+ \sum_{j=1}^{M} \beta_j \ln y_j + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \beta_{jk} \ln y_j \ln y_k \\ &+ \sum_{j=1}^{N} \sum_{k=1}^{M} \delta_{jk} \ln x_j \ln y_k \end{aligned} \tag{9}$$

with $\alpha_{jk} = \alpha_{jk}$, $\beta_{jk} = \beta_{kj}$, $\sum_{j=1}^{M} \beta_j = 1$, $\sum_{j=1}^{M} \beta_{jk} = 0 \ \forall \ k$, and $\sum_{k=1}^{M} \delta_{jk} = 0 \ \forall \ j$, we get the following distance elasticities of the inputs:

$$\epsilon_j = \frac{\partial \ln D(x, y)}{\partial \ln x_j} = \alpha_j + \sum_{k=1}^{N} \alpha_{jk} \ln x_k + \sum_{k=1}^{M} \delta_{jk} \ln y_k \tag{10}$$

and the following elasticity of scale:

$$\epsilon = -\sum_{j=1}^{N} \epsilon_j = -\sum_{j=1}^{N} \alpha_j - \sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_{jk} \ln x_k - \sum_{j=1}^{N}\sum_{k=1}^{M} \delta_{jk} \ln y_k \tag{11}$$

Scaling all input quantities in the Translog output distance function (9) by a factor $s > 0$, we have to scale all outputs by a factor t with

$$\ln t = \epsilon \ \ln s - \frac{1}{2}(\ln s)^2 \sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_{jk} \tag{12}$$

in order to maintain a constant distance D. This scaling of the inputs and outputs results in the following elasticity of scale of the "scaled firm":

$$\epsilon = -\sum_{j=1}^{N} \alpha_j - \sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_{jk} \ln(s\ x_k) - \sum_{j=1}^{N}\sum_{k=1}^{M} \delta_{jk} \ln(t\ y_k) \tag{13}$$

$$= -\sum_{j=1}^{N} \alpha_j - (\ln s)\sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_{jk} - \sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_{jk} \ln x_k - (\ln t)\sum_{j=1}^{N}\sum_{k=1}^{M} \delta_{jk} - \sum_{j=1}^{N}\sum_{k=1}^{M} \delta_{jk} \ln y_k \tag{14}$$

$$= -\sum_{j=1}^{N} \alpha_j - (\ln s)\sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_{jk} - \sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_{jk} \ln x_k - \sum_{j=1}^{N}\sum_{k=1}^{M} \delta_{jk} \ln y_k \tag{15}$$

Hence, the effect of the (logarithmic) scale on the elasticity of scale is:

$$\frac{\partial \epsilon}{\partial \ln s} = -\sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_{jk} \tag{16}$$

Thus, the elasticity of scale changes linearly with the logarithm of the firm's scale.

B.3. Translog input distance function

Given the Translog input distance production with N inputs and M outputs:

$$\begin{aligned}\ln D = \alpha_0 &+ \sum_{j=1}^{N} \alpha_j \ln x_j + \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} \alpha_{jk} \ln x_j \ln x_k \\ &+ \sum_{j=1}^{M} \beta_j \ln y_j + \frac{1}{2}\sum_{j=1}^{M}\sum_{k=1}^{M} \beta_{jk} \ln y_j \ln y_k \\ &+ \sum_{j=1}^{N}\sum_{k=1}^{M} \delta_{jk} \ln x_j \ln y_k\end{aligned} \tag{17}$$

with $\alpha_{jk} = \alpha_{jk}$, $\beta_{jk} = \beta_{kj}$, $\sum_{j=1}^{N} \alpha_j = 1$, $\sum_{j=1}^{N} \alpha_{jk} = 0 \; \forall \; k$, and $\sum_{j=1}^{N} \delta_{jk} = 0 \; \forall \; k$, we get the following distance elasticities of the outputs:

$$\epsilon_j = \frac{\partial \ln D(x,y)}{\partial \ln y_j} = \beta_j + \sum_{k=1}^{M} \beta_{jk} \ln y_k + \sum_{k=1}^{N} \delta_{kj} \ln x_k \tag{18}$$

and the following *inverse* elasticity of scale:

$$\epsilon^{-1} = -\sum_{j=1}^{M} \epsilon_j = -\sum_{j=1}^{M} \beta_j - \sum_{j=1}^{M} \sum_{k=1}^{N} \beta_{jk} \ln y_k - \sum_{j=1}^{M} \sum_{k=1}^{N} \delta_{kj} \ln x_k \tag{19}$$

Scaling all input quantities in the Translog input distance function (17) by a factor $s > 0$, we have to scale all outputs by a factor t with

$$\ln t = \epsilon^{-1} \; \ln s - \frac{1}{2} (\ln s)^2 \sum_{j=1}^{M} \sum_{k=1}^{M} \beta_{jk} \tag{20}$$

in order to maintain a constant distance D. This scaling of the inputs and outputs results in the following inverse elasticity of scale of the "scaled firm":

$$\epsilon^{-1} = -\sum_{j=1}^{M} \beta_j - \sum_{j=1}^{M} \sum_{k=1}^{M} \beta_{jk} \ln(s \; y_k) - \sum_{j=1}^{M} \sum_{k=1}^{N} \delta_{kj} \ln(t \; x_k) \tag{21}$$

$$= -\sum_{j=1}^{M} \beta_j - (\ln s) \sum_{j=1}^{M} \sum_{k=1}^{M} \beta_{jk} - \sum_{j=1}^{M} \sum_{k=1}^{M} \beta_{jk} \ln y_k - (\ln t) \sum_{j=1}^{M} \sum_{k=1}^{N} \delta_{kj} - \sum_{j=1}^{M} \sum_{k=1}^{N} \delta_{kj} \ln x_k \tag{22}$$

$$= -\sum_{j=1}^{M} \beta_j - (\ln s) \sum_{j=1}^{M} \sum_{k=1}^{M} \beta_{jk} - \sum_{j=1}^{M} \sum_{k=1}^{M} \beta_{jk} \ln y_k - \sum_{j=1}^{M} \sum_{k=1}^{N} \delta_{kj} \ln x_k \tag{23}$$

Hence, the effect of the (logarithmic) scale on the inverse elasticity of scale is:

$$\frac{\partial \epsilon^{-1}}{\partial \ln s} = -\sum_{j=1}^{M} \sum_{k=1}^{M} \beta_{jk} \tag{24}$$

Thus, the inverse elasticity of scale changes linearly with the logarithm of the firm's scale. This means that the elasticity of scale changes non-linearly with the logarithm of the firm's scale with following (non-constant) slope:

$$\frac{\partial \epsilon}{\partial \ln s} = \frac{\partial \epsilon}{\partial \epsilon^{-1}} \; \frac{\partial \epsilon^{-1}}{\partial \ln s} = \epsilon^2 \; \sum_{j=1}^{M} \sum_{k=1}^{M} \beta_{jk} \tag{25}$$

References

Aw, B. and Hwang, A. (1995). Productivity and the export market: A firm-level analysis. *Journal of Development Economics* 47: 313–332.

Bakucs, L. Z., Latruffe, L., o, I. F. and Fogarasi, J. (2010). The impact of EU accession on farms' technical efficiency in Hungary. *Post-Communist Economies* 22: 165–175.

Barbier, P. (1984). Inverse relationship between farm size and land productivity: A product of science or imagination? *Economic and Political Weekly* 19: A189–A191+A193–A198.

Barrett, C. B. (1996). On price risk and the inverse farm size-productivity relationship. *Journal of Development Economics* 51: 193–215.

Barrett, C. B., Bellemare, M. F. and Hou, J. Y. (2010). Reconsidering conventional explanations of the inverse productivity-size relationship. *World Development* 38: 88–97.

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.

Carter, M. R. (1984). Identification of the inverse relationship between farm size and productivity: An empirical analysis of peasant agricultural production. *Oxford Economic Papers* 36: 131–45.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978). Measuring the efficiency of desicion making units. *European Journal of Operational Research Policy* 2: 429–444.

Christensen, L. R., Jorgenson, D. W. and Lau, L. J. (1971). Conjugate duality and the transcendental logarithmic functions. *Econometrica* 39: 255–256.

Cobb, C. W. and Douglas, P. H. (1928). A theory of production. *The American Economic Review* 18: 139–165.

Croissant, Y. and Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software* 27: 1–43.

Curtiss, J. (2000). Technical Efficiency and Competitiveness of the Czech Agricultural Sector in Late Transition-the Case of Crop Production. In *proceedings KATO Symposium*.

Dorward, A. (1999). Farm size and productivity in Malawian smallholder agriculture. *The Journal of Development Studies* 35: 141–161.

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

Gallant, A. R. (1982). Unbiased determination of production technologies. *Journal of Econometrics* 20: 285–323.

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 (2012a). Bank Danych Lokalnych / Local Data Bank. Główny Urząd Statystyczny / Central Statistical Office of Poland, http://www.stat.gov.pl/bdlen/ [accessed 09-October-2012].

GUS (2012b). Prices in the national economy - january 2012. Główny Urząd Statystyczny / Central Statistical Office of Poland, http://www.stat.gov.pl/cps/rde/xbcr/gus/PUBL_pt_prices_national_economy_01m_2012.zip [accessed 05-March-2012].

Hausman, J. A. (1978). Specification test in econometrics. *Econometrica* 46: 1251–1272.

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

Henderson, D., Carroll, R. and Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics* 144: 257–275.

Henderson, D. and Ullah, A. (2005). A nonparametric random effects estimator. *Economics Letters* 88: 403–407.

Henderson, D. J. and Simar, L. (2005). A Fully Nonparametric Stochastic Frontier Model for Panel Data. Working Paper 0519, Departmentment of Economics, State University of New York at Binghamton.

Heshmati, A. and Kumbhakar, S. C. (1997). Estimation of technical efficiency in swedish crop farms: A pseudo panel data approach. *Journal of Agricultural Economics* 48: 22–37.

Hockmann, H., Pieniadz, A. and Goraj, L. (2007). Modeling Heterogeneity in Production Models: Empirical Evidence from Individual Farming in Poland. Discussion Paper 109, Leibniz Institute of Agricultural Development in Central and Eastern Europe, Halle (Saale), Germany.

Hsiao, C., Li, Q. and Racine, J. (2007). A consistent model specification test with mixed discrete and continuous data. *Journal of Econometrics* 140: 802–826.

Hughes, G. (2000). Total productivity of emergent farm structures in Central and Eastern Europe. In Stefan Tangermann, M. B. (ed.), *Central and Eastern European Agriculture in an Expanding European Union*. CABI Pub.

Humphrey, T. M. (1997). Algebraic production functions and their uses before Cobb-Douglas. *Federal Reserve Bank of Richmond Economic Quarterly* 83: 51–83.

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.

Kneip, A. and Simar, L. (1996). A general framework for frontier estimation with panel data. *Journal of Productivity Analysis* 7: 187–212.

Kumbhakar, S. C. and Lien, G. (2010). Impact of subsidies on farm productivity and efficiency. In Ball, V. E., Fanfani, R. and Gutierrez, L. (eds), *The Economic Impact of Public Support to Agriculture: An International Perspective*. Springer Verlag, *7*, 109–124.

Kumbhakar, S. C., Park, B. U., Simar, L. and Tsionas, E. G. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics* 137: 1–27.

Kumbhakar, S. C. and Tsionas, E. G. (2009). Nonparametric estimation of production risk and risk preference functions. In *Nonparametric Econometric Methods (Advances in Econometrics, Volume 25)*. Emerald Group Publishing Limited, *25*, 223–260.

Latruffe, L., Balcombe, K., Davidova, S. and Zawalinska, K. (2004). Determinants of technical efficiency of crop and livestock farms in Poland. *Applied Economics* 36: 1255–1263.

Latruffe, L. and Sauer, J. (2010). Subsidies, Production Structure and Technical Change: A Cross-Country Comparison. In *2010 Annual Meeting, July 25-27, 2010, Denver, Colorado*. Agricultural and Applied Economics Association.

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. (2002). Individual farming as a labor sink: Evidence from Poland and Russia. Paper presented at the Annual Meeting of the American Agricultural Economics Association Long Beach, July 28-31, 2001.

Li, Q. and Racine, J. S. (2003). Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis* 86: 266–292.

Li, Q. and Racine, J. S. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica* 14: 485–512.

Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* 95.

Livanis, G. T., Salois, M. J. and Moss, C. B. (2009). A nonparametric kernel representation of the agricultural production function: Implications for economic measures of technology. Paper presented at 83rd Annual Conference of Agricultural Economics Society, Dublin, Ireland.

Mathijs, E. and Swinnen, J. F. M. (2001). Production organization and efficiency during transition: An empirical analysis of East German agriculture. *The Review of Economics and Statistics* 83: 100–107.

Michaelides, P. G., Vouldis, A. T. and Tsionas, E. G. (2010). Globally flexible functional forms: The neural distance function. *European Journal of Operational Research* 206: 456–469.

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

Noam, E. M. (1983). Is cable television a natural monopoly? *Communications. The European Journal of Communication Research* 9: 241–259.

Oh, I., Lee, J. and Heshmati, A. (2008). Total factor productivity in Korean manufacturing industries. *Global Economic Review: Perspectives on East Asian Economies and Industries* 37: 23–50.

Porter, J. R. (1996). Essay in Econometrics. Ph.D. thesis, Massachusetts Institute of Technology. Dept. of Economics.

R Development Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Racine, J. S. (1997). Consistent significance testing for nonparametric regression. *Journal of Business and Economic Statistics* 15: 369–379.

Racine, J. S. (2008). Nonparametric econometrics: A primer. *Foundations and Trends in Econometrics* 3: 1–88.

Racine, J. S., Hart, J. and Li, Q. (2006). Testing the significance of categorical predictor variables in nonparametric regression models. *Econometric Reviews* 25: 523–544.

Racine, J. S. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* 119: 99–130.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* 31: 350–371.

Rao, V. and Chotigeat, T. (1981). The inverse relationship between size of land holdings and agricultural productivity. *American Journal of Agricultural Economics* 63: 571–574.

Rasmussen, S. (2010). Scale efficiency in Danish agriculture: An input distance–function approach. *European Review of Agricultural Economics* 37: 335–367.

Ray, S. C. (1998). Measuring scale efficiency from a translog production function. *Journal of Productivity Analysis* 11: 183–194.

Renner, S., Hockmann, H., Pieniadz, A. and Glauben, T. (2009). On flexibility in the Polish farming sector. Paper presented at the 111th EAAE-IAAE Seminar 'Small Farms: decline or persistence', University of Kent, Canterbury, UK.

Rudra, A. and Sen, A. (1980). Farm size and labour use: Analysis and policy. *Economic and Political Weekly* 15: 391+393–394.

Sarris, A., Savastano, S. and Tritten, C. (2004). Factor market imperfections and polarization of agrarian structures in Central and Eastern Europe. In Martin, P. and Peter, W. (eds), *The Role of Agriculture in Central and Eastern European Rural Development: Engine of Change or Social Buffer*. IAMO, *25*, 84–100.

Sen, A. (1962). An aspect of indian agriculture. *Economic Weekly* Annual Number 14.

Verschelde, M., D'Haese, M., Rayp, G. and Vandamme, E. (forthcoming). Challenging small-scale farming, a non-parametric analysis of the (inverse) relationship between farm productivity and farm size in burundi. *Journal of Agricultural Economics* .

Vinod, H. and Ullah, A. (1988). Flexible production function estimation by nonparametric kernel estimators. *Advances in Econometrics: Nonparametric and Robust Inference* 7: 139–160.

Vlastuin, C., Lawrence, D. and Quiggin, J. C. (1982). Size economies in Australian agriculture. *Review of Agricultural Economics* 50: 265–273.

Wang, M.-C. and van Ryzin, J. (1981). A class of smooth estimators for discrete distributions. *Biometrika* 68: 301–309.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* 90: 43–52.

Wicksteed, P. H. (1894). *Essay on the Coordination of the Laws of Distribution*. London: LSE, 1932nd ed.

Zhu, X. and Lansink, A. O. (2010). Impact of CAP subsidies on technical efficiency of crop farms in Germany, the Netherlands and Sweden. *Journal of Agricultural Economics* 61: 545–564.