MAGYAR TUDOMÁNYOS AKADÉMIA
Közgazdaság- és Regionális Tudományi Kutatóközpont

MTA KRTK KTI

Centre for Economic and Regional Studies
HUNGARIAN ACADEMY OF SCIENCES

MŰHELYTANULMÁNYOK DISCUSSION PAPERS

MT-DP – 2016/36

Technological differences, theoretically consistent frontiers and technical efficiency: a Random parameter application in the Hungarian crop producing farms

LAJOS BARÁTH – HEINRICH HOCKMANN

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES,
HUNGARIAN ACADEMY OF SCIENCES - BUDAPEST, 2016

Discussion papers
MT-DP – 2016/36

Institute of Economics, Centre for Economic and Regional Studies,
Hungarian Academy of Sciences

Technological differences, theoretically consistent frontiers and technical efficiency: a Random parameter application in the Hungarian crop producing farms

Authors:

Lajos Baráth
research fellow
Institute of Economics, Centre for Economic and Regional Studies,
Hungarian Academy of Sciences
E-mail: barath.lajos@krtk.mta.hu

Heinrich Hockmann
Deputy Head of Department Agricultural Markets
Leibniz Institute of Agricultural Development in Transition Economies (IAMO)
E-mail: hockmann@iamo.de

December 2016

ISBN 978-615-5594-77-9
ISSN 1785 377X

Technological differences, theoretically consistent frontiers and technical efficiency: a Random parameter application in the Hungarian crop producing farms

Lajos Baráth – Heinrich Hockmann

Abstract

Existing studies of agricultural production largely neglect technology heterogeneity. However, the assumption of homogeneous production may result in inadequate policy implications. There is a growing literature on this issue. In this paper we contribute to this literature by modelling the effect of heterogeneous technologies and its impact on technological parameters and technical efficiency using a reformulated Random parameter Model. Our approach is based on the model developed by Alvarez et al. (2004). However, the original version of this model faces one crucial econometric problem: the assumption of independence of technical inefficiency and input variables does not, hence the estimated results are not necessarily consistent. Therefore we reformulate the model to allow for a more consistent estimation. Additionally, we examine the importance of the fulfilment of theoretical consistency: monotonicity and quasi-concavity. In order to fulfil these criteria we apply constrained maximum likelihood estimation, more specifically we build linear and non-linear constraints into the model and force it to yield theoretically consistent results, not only in the mean but also in different approximation points. For the empirical analysis we use farm level data from the Hungarian FADN Database. The results showed that considering technological differences is important. According to model selection criteria the modified Alvarez model with constraints was the preferred specification. Additionally, the results imply that the consideration of the effect of heterogeneous technologies on production potential and efficiency crucial in order to get adequate policy implication.

JEL: C5, D24, Q12

Keywords:
technical efficiency, technological heterogeneity, Random Parameter Model, theoretical consistency, monotonicity, quasi-concavity, Hungarian agriculture

Technológia különbségek, elméleti konzisztencia és technikai hatékonyság: a magyar növénytermesztő üzemek vizsgálata random paraméter modellel

Baráth Lajos – Heinrich Hockmann

Összefoglaló

A mezőgazdasági termelés hatékonyságát modellező tanulmányok nagy része nem veszi figyelembe az üzemek közötti technológiai különbségeket. A homogén technológia feltételezése azonban hibás agrárpolitikai javaslatokhoz vezethet. Egyre növekszik azon tanulmányok száma, amelyek erre a problémára hívják fel a figyelmet. Cikkünkben egy módosított random paraméter modell segítségével a különböző technológiák figyelembe vételének hatását vizsgáljuk a becsült technológiai paraméterekre és a technikai hatékonyságra. A cikkben használt modell az Alvarez és társai által 2004-ben javasolt modellen alapul. Az eredeti modellben azonban a technikai hatékonyság és az input változók függetlenségének kritériuma nem feltétlenül teljesül, így az eredmények torzítottak lehetnek. Az eredeti modellt ezért módosítjuk és bemutatjuk, hogy a módosított verzió esetében nem jelentkezik ez a probléma. Második célkitűzésünk az elméleti konzisztenciának (monotonitás, kvázi konkávitás) való megfelelés hatásának vizsgálata. Lineáris és nem lineáris korlátokat építünk a modellbe és biztosítjuk, hogy az átlagtól távolabbi megfigyelési pontokon is konzisztens eredményt kapjunk. Az empirikus elemzéshez a Tesztüzemi Rendszer (FADN) adatait használtuk. Modellszelekciós kritériumok alapján arra következtethetünk, hogy a módosított, lineáris és nem lineáris korlátokkal becsült modell illeszkedik legjobban a vizsgált adatokra. Az eredmények alátámasztják, hogy a technológiai különbségek figyelembe vétele és az elméleti konzisztenciának való megfelelés döntő jelentőségű a megfelelő agrárpolitikai javaslatok kidolgozásában.

JEL: C5, D24, Q12

Tárgyszavak:
technikai hatékonyság, technológiai különbségek, random parameter modell, elméleti konzisztencia, montonitás, kvázi konkávitás, magyar mezőgazdaság

INTRODUCTION

Analysis of farm efficiency using frontier methods can deliver significant insights into the competitiveness of farms and their potential for increasing productivity and improving resource use. Policy makers are particularly interested in the potential impact of their decisions on performance of firms. Thus findings from the study of technical efficiency (TE), far-reaching policy implications (Abdulai-Tietje, 2007; Bauer et al., 1998).

There exist numerous papers concerning technical efficiency of the agricultural sector, but in the majority of technical efficiency literature homogenous technology is assumed for every farms1. However, farms may adopt different technologies or face different natural resource and economic conditions for a variety of reasons. Without considering this possible heterogeneity, the efficiency and productivity estimates of farms can be over-estimated.

There is a growing body of macro agricultural productivity literature which emphasise the importance of modelling these possible technologies. In a recent paper Eberhardt-Teal (2013) conducted an extensive comparison among different linear parametric models and revealed that the assumption of a homogeneous production function in the farm sector may mask or distort important insights into development and demonstrates that failure to account for technology heterogeneity leads to misspecified empirical models with serious implications for any TFP estimates obtained (Eberhardt-Teal, 2013).

In contrast, in Stochastic Frontier context the number of agricultural efficiency/productivity paper accounting for technological differences is limited. Concerning a related issue, the separation of unobserved heterogeneity from efficiency estimates, using different mainly variable intercept models, extensive investigations has already been conducted, but there is a lack of systematic investigation of a more generally formulated models, which are able to distinguish efficiency from technological differences across farms.

Two classical methods have been developed in the frontier context which allow to model different technologies: the random parameters model and the latent class models. Random parameters formulation models consider firm heterogeneity in the form of continuous parameter variation. The latent class model, on the other hand, can be viewed as an approximation to this since variation of the parameters are treated as generated by a discrete distribution instead (Greene, 2005).

Although, Latent Class models was used and compared with traditional SFA models in some paper (e.g. Alvarez and del Corral, 2010; Alvarez et al., 2012; Emvalomatis, 2007; Sauer and Morrison Paul, 2013), only a few author examined efficiency with random parameter

[1] Technology, in this context, refers to the shape of the production function.

model (e.g Cechura et al., 2014; Wang et al., 2012) and extensive comparison of the results with other model hasn't been conducted yet. In addition, the results of these paper might be biased, as they used a special type of RPM, namely the model that was originally proposed by Alvarez et al, 2004. However, it might not yield consistent estimates, because the fundamental assumption of the independence of inefficiency (u) and explanatory variables (x), does not necessarily hold.

Our first contribution to the literature is that we propose a reformulated version of original Alvarez et al (2004) model that provide more consistent estimates.

In our second contribution we focus on the changes in the parameter estimates, efficiency and productivity scores in agricultural context when we move from variable intercept SF model to RPM. More precisely, among the available variable intercept models for purposes of comparison, our starting point is the True Random Effect (TRE) model which was found in many paper the most suitable to separate unobserved heterogeneity from efficiency estimates (e.g, Abdulai-Tietje, 2007; Kuenzle, 2005; Farsi et al., 2005) and compare the results of these models with the original Alvarez et al., 2004 and the reformulated Alvarez et al model. Allowing for heterogeneous technologies is much closer to real world, therefore such comparison have important policy implication. As public policies designed to improve agricultural productivity can be targeted at the different components of productivity. Efficient policy requires careful and realistic estimates.

In our third contribution we examine, how does the fulfilment of the requirement of theoretical consistency influence the estimates? Efficiency estimates are often used without a critical assessment with respect to the literature on theoretical consistency, specifically on monotonicity and quasi-concavity. The robustness of policy suggestions based on inferences from efficiency measures nevertheless crucially depends on theoretically well-founded estimates (Sauer-Hockmann, 2005). In order to fulfil the criteria of monotonicity and quasi-concavity we build linear and non-linear constraints into the model and force the model to yield theoretically consistent results.

Following, the original Alvarez et al., 2004 model we formulate the production frontier in translog form. However, the translog specification fulfils these criteria only locally (Diewert-Wales, 1988). The usual practice is checking the fulfilment of theoretically consistency at the mean of the data. However, the lack of global consistency yields that production far apart from this point can't be consistently interpreted. As our aim is to investigate farms individual technologies, we need consistent estimates not only in the mean of the data. We overcome this problem as follows. We identify several approximation points and force the model – through the linear and non-linear constraint – to fulfil the required criteria at every approximation point. We then estimate the model with and without constraints thus we are able to check how the fulfilment of these restrictions changes the estimation results and whether it leads to different policy suggestion.

For empirical analysis we use Hungarian FADN Data. Considering the characteristics of the Hungarian agriculture we assume that there exist at least two significant background factors which imply that the technology parameters are different across farms and therefore the estimation of a homogeneous production function might lead to inadequate policy implication. First, the Hungarian agriculture has a typical dual structure with a large number of small-scale farms on one side and a small number of large-scale farms on the other and within these categories there are further significant differences among farms (i.e. the standard deviation of output and input variables are high). Second, the ecological conditions are very diverse – according to an agro-economic potential survey, 35 ecological regions has been distinguished (Láng et al., 1983). Because of these we think that an RPM is an adequate approach to model the production structure and efficiency in such a diverse production environment.

The remainder of the paper is organised as follows. We begin by briefly examining previous studies concerning heterogeneity and then we outline the methods used in the analysis (Sec.2 and 3). In the fourth Section we present the data used in our analysis. The following section (Section 5) discusses our results. Here special weight is put to the estimated efficiencies and the interpretation of unobserved heterogeneity. Finally, Sec. 6 summarizes our findings and draws conclusion about future modelling.

THEORETICAL BACKGROUND: THE BASIC EFFICIENCY MODEL

We start with the basic stochastic frontier model. Following Aigner et al. (1977) and Meeusen and van den Broeck (1977) the stochastic production frontier model in a general form might be written as follows:

$$lny_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i - u_i,$$

where $y_i$represents the output of the i-th farm, xi, is a vector of inputs, □□is a vector of unknown parameters, v_i is a symmetric random error, which accounts for statistical noise and u_i is a non-negative random variable associated with technical inefficiency.

Several techniques can be used to estimate the unknown parameters in the model. Aigner, Lovell and Schmidt (ALS) (1977) obtained ML estimates under the assumptions (Coelli et al., 2005):

$v_i \sim iid\, N(0, \sigma_v^2)$

$u_i \sim iid\, N^+(0, \sigma_u^2)$

The assumptions says that the v_i s are independently and identically distributed normal random variables with zero means and variances σ_v^2 and the u_i s are independently and identically distributed half-normal variables with scale parameter σ_u^2 (Coelli et al., 2005). ALS (1977) parameterised the log-likelihood function for this so-called normal; half-normal model in terms of $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\lambda = \sigma_v^2/\sigma_u^2$. Using this parameterisation, the model relies of the likelihood of the compound error term $\varepsilon_i = v_i - u_i$:

$f(\varepsilon) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left(\frac{\lambda\varepsilon}{\sigma}\right)$, where

ϕ and Φ represent the density and distribution of the standard normal. Using this relation, the log-likelihood function can be determined in the usual way (ALS, 1997).

Following the Model of ALS (1977), stochastic frontier models have been subject of a great body of literature resulting in a large number of econometric models. An extensive review can be found in Kumbhakar and Lovell (2000), Coelli et al. (2005) and Fried et al. (2008).

One of the most important issues in these models is adjusting for the unobserved heterogeneity among firms functioning in different production environments (Farsi et al., 2005). Early work on heterogeneity focused at the inclusion of environmental variables in the production models (Agrell and Brea-Solis, 2015). However, in many cases it may not be feasible to include such kind of variables because of shortages of degree of freedom or multicollinearity or they may simple not be observable (Hsiao, 2014). Therefore, later works, explored various statistical techniques to determine omitted variables (i.e. unobserved heterogeneity) (Agrell and Brea-Solis, 2015).

Conventional panel data models such as fixed-effects or random-effects models can be employed to account for unobserved heterogeneity (Pitt and Lee, 1981; Schmidt and Sickles, 1984). However, there are two major limitations of these models: (i) the treatment of the inefficiency term as time-invariant, which raises a fundamental identification problem and (ii) they fail to distinguish between cross individual heterogeneity and inefficiency (Abdulai-Tietje, 2007; Greene, 2005).

To account for these limitations, Greene (2005) proposed two stochastic frontier models that are time-variant and that distinguish unobserved heterogeneity from the inefficiency component. These models are termed the 'true' fixed-effects (TFE) and 'true' random-effects (TRE) models (Abdulai-Tietje, 2007). However, as pointed out by Greene (2005), the TFE model might produce biased individual effects and efficiency estimates, because the presence of the individual effects creates an incidental parameter problem. In contrast, TRE models produce unbiased inefficiency estimates, therefore we focus here on the description of TRE model, only.

The TRE model might be written as follows:

$$y_{it} = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_{it} + w_i + v_{it} - u_{it},$$

where w_i is a random firm specific effect. Other variables as defined earlier2.

Although TRE is able to separate TE and heterogeneity, it is insufficient to capture underlying variation in agricultural technology. However, as Hsiao, 2014 argues: in farm production it is likely that unobserved heterogeneity could also impact the marginal productivity of inputs used such as soil characteristics (e.g., slope, soil fertility, water reserve, etc.) or climatic conditions. In this context, random parameter models appear to be more capable of capturing the unobserved heterogeneity than a model with only individual- and/or time-specific effects (variable-intercept models) (Hsiao, 2014). Despite of this fact, variable-coefficient models have not gained as wide acceptance in empirical work as has the variable-intercept models, may be because of computational complexity.

This paper attempts to close this gap: (i) by applying an RPM in agricultural context with farm level panel data and (ii) by modification of the RPM, which was originally proposed by Alvarez et al., 2004.

MODELLING TECHNOLOGY WITH UNOBSERVED HETEROGENEITY

In order to model technological differences we use a modified version of the fixed-management model, originally proposed by Alvarez et al. (2004):

$$y_{it} = f(\boldsymbol{x}_{it}, t, m_i^*) + v_{it} - u_{it},$$

where mi* is a firm specific latent variable, other variables as defined earlier.

The key feature of the model is the interactions of “mi*” with the input variables. This model allows not only the constant to change, but also the structural parameters. Such specification therefore can be used to model the heterogeneity of the production structure. Without this interaction the model doesn’t differ from standard variable intercept models.

Alvarez et al., (2004) interpret the latent variable (mi*) as an effect that accounts for management differences between firms. However, there doesn't exist any underlying theory or empirical justification which would state why this term should capture only the effect of management. It might capture various differences among firms. Recent empirical papers (Wang et al., 2012; Belyaeva et al., 2014) argue that this term might capture the effect of various sources of unobserved heterogeneity such as differences in regional/farm characteristics, input quality, environmental conditions or socio-economic characteristics etc. We follow this interpretation.

[2] The estimation procedure is similar to the one of the models discussed below. In order to avoid duplications we omit a detailed description of the estimation procedure here.

The expectation of the model is that the production is monotonically increasing in mi*. The maximal output for given $\boldsymbol{x}_{it}$ (i.e the frontier output, y_{it}^*) is achieved with the maximal level of this component, i.e with m_i^*.

Assuming a translog function the production frontier of the original Alvarez et al., 2004 model may be written as:

$$lny_{it}^* = \alpha_0 + \beta_t t + \frac{1}{2}\,\beta_{tt}t^2 + \boldsymbol{\beta}_k' ln\boldsymbol{x}_{it} + \boldsymbol{\beta}_{kt}' ln\boldsymbol{x}_{it} * t + \frac{1}{2} ln\boldsymbol{x}_{it}' \boldsymbol{B}_{kk}\, ln\boldsymbol{x}_{it}$$
$$+\beta_m m_i^* + \beta_{mt} m_i^* * t + \frac{1}{2}\beta_{mm} m_i^{*2} + \boldsymbol{\beta}_{mk} ln\boldsymbol{x}_{it} * m_i^* + v_{it},$$

where y_{it}^* is the optimal output. $\boldsymbol{\beta}_k'$ and $\boldsymbol{x}$ vectors of parameter estimates and exogenous inputs, respectively. The k×k matrix Bkk contains the second order parameters. Additionally a time trend (t) and its square (tt) are introduced to capture non-monotonic technical change. The variable m_i^* enters all first order terms: the constant, the time trend and the first order parameters of the exogenous variables with

(2) $\dfrac{\partial \ln y^*}{\partial m^*} \geq 0$,

e. g. the unobserved heterogeneity influences production positively.

Farms generally do not exploit their full production capacities (y_{it}^*), e. g. their actual output (y_{it}) can lie below the optimal level. The observed output (y_{it}) can be modelled using the same structure as y_{it}^* (see eq.1). The only difference is: instead of the maximum level of the unobserved components (m_i^*), the production model is defined with the actual level of this component (m_i), where ($m_i < m_i^*$).

Having defined the optimal and actual level of production we can define technical efficiency (TEit or u_{it}) – which is the ratio of observed to potential output:

(3) $TE_{it} = -u_{it} = lny_{it}\text{-}lny_{it}^*$

For estimation, a critical assumption is the absence of correlation between u_{it} and the input levels ($\boldsymbol{x}_{it}$); therefore Alvarez et al., 2004 highlight that it is important to show explicitly the definition of u_{it} in the model:"

(4) $$TE_{it} = (\beta_m + b_{mt}t + \boldsymbol{\beta}_{mk}' ln\boldsymbol{x}_{it})(m_i - m_i^*) + \frac{1}{2}\,\beta_{mm}(m_i^2 - m_i^{*2}) \geq 0$$

Although $ln\boldsymbol{x}_{it}$ appear in u_{it}, Alvarez et al., 2004 assume that it does not influence $(m_i - m_i^*)$ and u_{it} can be calculated using only the first summand in Eq. 3. However, because of the inclusion of the second summand [$\frac{1}{2}\,\beta_{mm}(m_i^2 - m_i^{*2})$], it is not possible to separate properly u_{it} and the other input variables. Hence, u_{it} and $ln\boldsymbol{x}_{it}$ might correlate.

In order to avoid this possible correlation, we reformulate the model to provide more consistent estimations. As there doesn't exists any economic theory that would suggest the inclusion of the second order term of term $\frac{1}{2}\beta_{mm}m_i^{*2}$, we estimate the model without this term. This allows us to properly separate $(m_i - m_i^*)$ (the efficiency effect) from the input level. As a result we can avoid the possible correlation between the input level and inefficiency, and the reformulated model ensures more consistent estimates.

Following this modification the model takes the form:

(5) $lny = lny^* + TE + v = f'(\boldsymbol{x}, t, m^*) - h(\boldsymbol{x}, t)u + v$

(6) $TE = -h(\boldsymbol{x}, t)u = -(a_m + b_{mt}t + b_{xm}ln\boldsymbol{x})(m - m^*)$

Equation 6 clearly show that in this formulation TE takes the form suggested in (Eq. 5).

Although (Eq.4) involves an unobservable variable (m_i^*) it is possible to translate the model, similarly as in the case of the original model, into an empirically estimable form. We continue to employ the translog function, so the model might be written as follows:

(7) $lny_{it}^* = \alpha_0 + \beta_t t + \frac{1}{2}\beta_{tt}t^2 + \boldsymbol{\beta}_k' ln\boldsymbol{x}_{it} + \boldsymbol{\beta}_{kt}' ln\boldsymbol{x}_{it} * t + \frac{1}{2} ln\boldsymbol{x}_{it}' \boldsymbol{B}_{kk} ln\boldsymbol{x}_{it}$

$+\beta_m m_i^* + \beta_{mt} m_i^* * t + \boldsymbol{\beta}_{mk} ln\boldsymbol{x}_{it} * m_i^* - u_{it} + v_{it}$, with

$u_{it} = h(\boldsymbol{x}_{it}, t)(m_i^* - m_i) = h(\boldsymbol{x}_{it}, t)u_i$

$v_{it} \sim iid\, N(0, \sigma_v)$ $u_i = m_i^* - m_i \sim iid\, N^+(0, \sigma_u)$ $m^* \sim N(0,1)$,

In this form the model takes the appearance of an RPM. It differs from more familiar random coefficients models in two respects. First, only the constant and first order terms are randomly distributed. Second, the random component m_i^* of each random parameter is the same for all exogenous variables.

The likelihood function will have the same structure as in conventional stochastic frontier analysis. However, two important modifications have to be mentioned: the compound variance of the error term and the measure of the extent of inefficiency. The variance of the compound error term in the original model is given by:

$$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$$

Since in our model σ_u^2 is determined through the relation $h(\boldsymbol{x}_{it}, t)u_i$ the standard calculation changes to

$$\sigma = \sqrt{h(\boldsymbol{x}, t)^2\sigma_u^2 + \sigma_v^2}.$$

Similarly, the original measure for the extent of efficiency $\lambda = \frac{\sigma_u}{\sigma_v}$ changes to

$$\lambda = \frac{h(x,t)\sigma_u}{\sigma_v}.$$

These modifications result in a similar likelihood function as in the Meeusen and van den Broeck (1977) case.

$f(\varepsilon) = \frac{2}{\sigma}\phi\left(\frac{\epsilon}{\sigma}\right)\Phi\left(\frac{\lambda\varepsilon}{\sigma}\right)$, with σ and λ defined above.

However, it remains how to treat the unobserved heterogeneity. The parameters of the model can be estimated by maximum simulated likelihood technique (Greene, 2005; Alvarez et al., 2004), where m^* is assumed to be standard normally distributed with zero mean and variance 1. We can then take several draws from this distribution, plug in the values of m_i^* into the likelihood function and construct the simulated maximum likelihood function as follows:

$$L = ln\left\{\prod_i \frac{1}{R}\sum_r \prod_t \left[\frac{2}{\sigma}\phi\left(\frac{\epsilon}{\sigma}\right)\Phi\left(\frac{\lambda\varepsilon}{\sigma}\right)\right]\right\}$$

$$\lambda = \frac{h(\boldsymbol{x},t)\sigma_u}{\sigma_v} \sigma^2 = h(\boldsymbol{x},t)^2\sigma_u^2 + \sigma_v, \ \varepsilon = v - h(\boldsymbol{x},t)u = y - f'(\boldsymbol{x},t,m^*)$$

Here R represents the number of replications in the simulation process. According to Greene, 2005: „in order to achieve a reasonable approximation to the true likelihood function, a reasonably large number of random draws are required. The process can be greatly accelerated by using 'intelligent' draws, such as Halton sequences (see Bhat (1999) or Train (1999) for discussion)." Following this suggestion we used Halton sequences for the simulation. We decided to use 1000 replication in the final model since we experienced that with lower number of replication the m_i^* do not seem to be give stable results for m_i^*. Moreover, the number of replications was also used by Alvarez et al. 2004 in their paper which applied this PRM model.

However, after the likelihood is optimized the m_i^* can be estimated via (Alvarez et al 2004, Greene 1986-2007):

$$E\left[m_i^* \mid y,\mathbf{x},t\right] = \frac{1/R\sum_{r=1}^{R} m_{i,r}^* \hat{f}\left(y \mid m_{i,r}^*,\mathbf{x},t\right)}{1/R\sum_{r=1}^{R} \hat{f}\left(y \mid m_{i,r}^*,\mathbf{x},t\right)},$$

where the function $\hat{f}\left(y \mid m_{i,r}^*,\mathbf{x},t\right)$ denotes the likelihood function for farm i evaluated at the parameter estimates and the current draw of mi*.

Given the estimates of the model and the expectation of the m_i^* the efficiency score can be estimated according to the conventional Jondrow et al. (1982) procedure (Alvarez et al. 2004):

$$E\left[u_{it} \mid m_i^*, \varepsilon\right] = \frac{\sigma\lambda}{\left(1+\lambda^2\right)}\left[\frac{\phi\left(-\frac{\left(\varepsilon_{it} \mid m_i^*\right)\lambda}{\sigma}\right)}{\Phi\left(-\frac{\left(\varepsilon_{it} \mid m_i^*\right)\lambda}{\sigma}\right)} - \frac{\left(\varepsilon_{it} \mid m_i^*\right)\lambda}{\sigma}\right]$$

In the estimation special attention was paid to theoretically consistency. Thus, we force the model to fulfil the requirements of theoretical consistency from an economic point of view, i.e. monotonicity and curvature properties (Coelli et al. 2005, Fuss and McFadden 1978).

"The translog specification fulfils these requirements only locally (Diewert and Wales 1988). In many cases the desired properties are not checked at all, or checked only at the mean of the dataset. The lack of global consistency yields that production far apart from the approximation points may not be consistently interpreted. We overcome this problem by forcing the estimation to provide theoretically consistent results for a number of approximation points by applying corresponding linear and nonlinear inequality restrictions. The approximation points were calculated as follows: for each variable the one sigma deviation from the mean was calculated. All observations inside of the one sigma deviation were excluded. For the resulted data sets the mean of each variable was computed; these means in the constructed new data set were used as new approximation points. Since we have 4 exogenous variables we constructed 8 approximation points. For each point we have 4 linear monotonicity restriction and one sign restriction for the impact of unobserved heterogeneity (Equation 2). In addition there are 4 nonlinear curvature restrictions for each approximation point. This procedure gives us consistent results from an economic point of view for a wide range of observations."

In order to fulfil the monotonicity criteria we build linear constraints into the model as follow:

$$S_{x_i} = \frac{\partial\, lny^*}{\partial\, lnx_i} \geq 0, \text{ for } i \leq 4 \text{ and}$$

$$S_{m^*} = \frac{\partial \ln y^*}{\partial m^*} \geq 0,$$

and force the model to fulfil this criteria at every defined approximation points.

Similarly, in order to fulfil the criteria of quasi-concavity we build nonlinear constraints into the model. We guaranteed that the bordered Hessian at the different approximation points is negative semidefinite. For this we apply the Cholesky decomposition (Lau 1978): Since every bordered Hessian Matrix H can be written in in the following form:3

3 Here the matric A refers to the second derivatives of the production function with respect to inputs and a represents the vector of first derivatives of the production function. The Cholesky factorization of

$$x'Hx \Leftrightarrow x'[\mathrm{B_{kk}} + aa' - a]x + x'ax \Leftrightarrow x'LDL'x + x'ax < 0.$$

Using $L'x = y$ the above expression changes to

$$x'Hx = y'Dy + x'ax < 0$$

$$\Leftrightarrow d_{11}y_1^2 + \sum_{i=2}^{n} d_{ii}y_i^2 + x'ax < 0$$

This results in the following curvature restrictions

(i) $d_{11}y_1^2 + x'ax \leq 0$ and

(ii) $d_{ii} \leq 0 \;\; \forall\, 2 \geq i \;\; \geq 4$

DATA

For purposes of empirical analysis we use Hungarian FADN Data. We used data on specialised COP farms over the 2004–2009 periods. Agricultural farms can join and leave the Hungarian FADN system, and to maintain representativeness, farms that leave the system are replaced by similarly characterised farms (Keszthelyi and Pesti, 2009). Our primary goal is to examine technological differences between farms. This question can be better examined if in every year the same farms are in the sample, therefore we used balanced panel. Our sample contains 3984 observations, 664 for each year. The data were provided by the Research Institute for Agricultural Economics.

Table 1

Descriptive Statistics

	Symbol	Mean	Standard Deviation	Minimum	Maximum
Output (EUR)	Y	40097.8	84487.8	128.51	931774.
Labour (Awu)	A	3.73	8.30	0.01	100.09
Land (ha)	L	237.41	428.57	3.68	3787.
Capital (EUR)	K	17309.6	42077.1	5.53	339055.
Variable Inputs (EUR)	V	28224.6	60186.5	323.26	657902.

Source: Authors' calculations based on Hungarian FADN data.

We estimated the model with one output (Y – total agricultural production in constant EUR) and four inputs: (1) labour in Annual Work Units (A), (2) utilised agricultural area (UAA) in hectares (L), (3) capital input (as a sum of depreciation and services) in constant EUR (K) and (4) variable input (intermediate consumption) in constant EUR (V). All of the variables expressed in nominal prices were deflated to 2005 prices with the use of the

---

$[\mathrm{B_{kk}} + aa' - a]$ is L'DL (Lau 1978). The matrix D represents the Cholesky factors dii for i =1,4 and L is a lower triangular matrix with element lij, for i, j =1,4 and lii = 1.

appropriate deflators; precisely, the output (Y) was deflated by the agricultural output price index, the total specific costs (V) by the price index of purchased goods and services and the corresponding values of total fixed assets (K) by the price index of agricultural investments. Some descriptive statistics are presented in Table 1.

The high variance of the individual variables is apparent; e.g. the labour input had a minimum value of 0.01 AWU and a maximum of 100 AWU, and the values for (UAA) ranged from 8.5to 3787 hectares. These high differences suggest that heterogeneity plays an important role in Hungarian agriculture. The huge differences between the minimum and maximum values also imply that the marginal products of these inputs are different among farms thus it seems to be reasonable to assume that farms with such a heterogeneous input endowment use different technologies and it is important to account for these differences in the production model.

RESULTS

We start with the discussion of parameter coefficients. Table 1 reports the parameter estimates of the models estimated: (1) TRE, (2) original Alvarez et al., 2004 Model and (3) the reformulated Alvarez Model.

All variables were divided by their geometric mean, thus the first order coefficients can be interpreted as output elasticities evaluated at the geometric mean of the sample. We interpret and compare the results among the different models from 5 points of view: (1) the characteristic of technology, (2) the effect of technological change, (3) the impact of the unobserved heterogeneity, (4) the importance of inefficiency in comparison to statistical noise and (5) return to scale.

First, table 2 clearly shows that the first order coefficients are different among the estimated models. The estimates of TRE and original Alvarez Model are rather similar, however there is a marked difference between these and the results of the modified Alvarez Model, especially the estimates of material inputs are round 10% lower in the modified Alvarez model. The reason for that might be that financial constraints are important in the Hungarian agriculture and many farms are not able to afford enough fertilisers and crop protection materials, therefore their technologies is less intensive. The lower estimates of the modified Alvarez Model, which are supposed to more consistently estimate technological differences, might reflect to this fact.

Second, all model suggest that technological progress occurred and the growth rate was increasing in the Hungarian agriculture over the analysed period. However, the modified Alvarez Model suggests that the technological progress was much higher. There are no differences among the models in terms of the characteristic of technological change: all models suggest that it was capital using.

Third, the effect of the unobserved latent variable, assumed to capture the effect of heterogeneity, is significant in all model versions. The interaction of this variable is significant not only with the constant (the assumption of TRE) but also with observed inputs, which suggest that there exist significant technological differences among farms and it is important to model these differences in order to get unbiased results.

Fourth, σ_u and σ_v is significant in all model and the estimates of σ_u is much higher in the first two models, suggesting that inefficiency variation has a significantly larger impact on output variation than statistical noise, confirming that technical inefficiency is an important phenomenon in Hungarian agriculture. The modified Alvarez Model differs from the first two models, since σ_u is much higher in this model. However, when calculating λ for the modified model the definition $\lambda = \frac{h(x,t)\sigma_u}{\sigma_v}$ has to be taken into account. Computing this expression at the sample mean gives $\lambda = 2.322$. Thus, all models give similar results for the significance of inefficiency.

Table 2

Parameter Estimates- Model Comparison

	Constant	**TRE**		**O_Alvarez**		**M_Alvarez**	
		0.273	***	0.267	***	0.241	***
Neutr TF	**T**	0.006	**	0.005	*	0.028	***
	TT	0.009	***	0.009	**	0.018	***
inputs- first order	**A**	0.054	***	0.040	***	0.074	***
	L	0.158	***	0.167	***	0.177	***
	K	0.132	***	0.125	***	0.142	***
	V	0.664	***	0.672	***	0.571	***
biased technical change	**A*T**	-0.001		-0.001		0.000	
	L*T	-0.006		-0.001		0.000	
	K*T	0.008	**	0.007	**	0.007	*
	V*T	-0.004		-0.007		-0.009	
second order	**AA**	0.038	**	0.018		0.010	
	LL	0.052		0.041		0.126	**
	KK	0.067	***	0.062	***	0.071	***
	VV	0.023		-0.004		0.023	
	AL	-0.107	***	-0.135	***	-0.077	***
	AK	0.017	*	-0.002		0.010	
	AV	0.046	**	0.094	**	0.051	**
	LK	0.015		0.014		-0.008	
	LV	0.002		-0.020		-0.057	
	KV	-0.082	***	-0.096	***	-0.051	**
ve d he te	**AM**	0.181	***	0.174	***	0.179	***

	AM_T	-		0.010	***	0.018	***
	AM_A	-		0.044	***	0.027	***
	AM_L	-		-0.002		0.015	**
	AM_K	-		0.002		-0.001	
	AM_V	-		-0.081	***	-0.070	***
	AMM	-		0.021	***	-	
auxiliary parameters	**SV**	0.167	***	0.163		0.167	***
	SU	0.395	***	0.394		2.083	***
	λ	2.368		2.414		2.232	
	RTS	1.008		1.004		0.964	
model selection	**Log L**	-1011.391		-961.402		-853.328	
	AIC	2070.782		1982.804		1764.656	
	BIC	2109.190		2030.814		1811.065	

Note: TRE true random effect, **O_Alvarez** =original Alvarez model, **M_Alvarez** = modified Alvarez model, all models were estimated without constraints

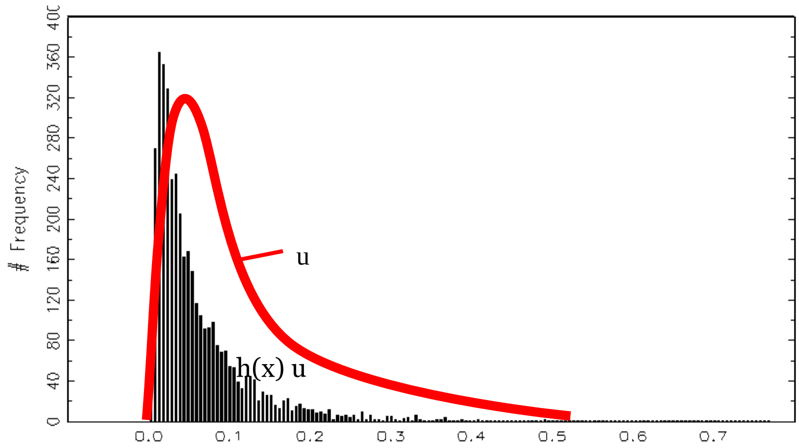
Source: Own estimation

The consequence of the different definitions can be seen in Figure 1. The red line represents the distribution of u for all three models4. The differences in the distributions were only marginal so that only one line is drawn. However, this distribution of u represent the also the distribution of technical efficiency in the first two models. The distribution of technical efficiency in the third model is presented by the histogram in Figure 1. It can be seen at the first glance that efficiencies of the first model are much lower than in the other models. From our point of view this result is very convincing. While in the first two model a large share of the farm have efficiency lower than 25% (e-u, u > .3) but the shares of firm with low efficiency reduces to a marginal fraction in the third model. Given the farms are operating in a homogeneous institutional environment the wide span of efficiencies are difficult to justify. Moreover, the parameter value for intermediate inputs in the modified Alvarez is much closer to the actual value of intermediate input share on total revenues provided by the data. Thus we conclude that the third model depicts the inefficiency much more appropriately than the first two models.

4 The line represents only the rough characterization of the distribution of u. Since it is used for illustrative purpose only this procedure appears reasonable.

Figure 1

Distribution of inefficiencies by model



Source: own calculation

Fifth, returns to scale are also different among the models. TRE and the original Alvarez Model suggest slightly increasing return to scale, whereas the modified Alvarez Model suggests slightly decreasing RTS.

Until now, we have seen that the estimated models yield different results. An obvious question, which suit better to these dataset? In order to select the most appropriate specification, we used model selection criteria. Specifically, we used the Akaike Information criteria (AIC) and Bayesian information criterion (BIC)[5]. The preferred model is that for which the value of these statistics are lowest. Table 1 clearly shows that both the values of AIC and BIC is getting lower, when we move from TRE to the modified Alvarez Model, i.e. the modified Alvarez model is the preferred model to these data. Two important conclusion can be drawn from these information: (i) the fact that AIC and BIC is smaller in the case of the original Alvarez Model compared to the TRE model confirm that a model which account for technological differences suit better to these dataset; (ii) the fact that AIC and BIC is even smaller in the case of the reformulated Alvarez model compared to the original Alvarez model confirm that it is also important to consider the possible correlation between TE and the input variables. As the model selection criteria suggest that the modified Alvarez Model is the preferred specification, the following examinations will be based only on this model.

In the next step of our analysis we examine the effect of constraints. Before discussing the different parameter estimates with and without constraints we check how many percent of the observation are consistent in the case of the models with and without constraints. We also check the number of binding restrictions. The results are presented in Table 3.

---

5 Also known as Schwarz criterion (SBC or SBIC).

Table 3

Check for theoretical consistency (% of observations, number)

	Mono-tonicity	Quasi-concavity	consistent	binding restrictions	
				linear	nonlinear
MA_without constraints	88%	75%	73%	-	
MA_ with constraints	97%	93%	92%	7	3

Note: MA refers to the modified Alvarez model.
Source: Own estimation

Table 3 reveals that the percentage of consistent estimates increased substantially. Despite the fact that we estimated the model with a great number of restrictions, only 10 restrictions in total were binding.

Table 4

Parameter Estimates- The effect of constraints

	MA_without constraints		MA_with constraints		Difference (%)+	
Constant	0.2412	***	0.2513	***	4.0%	***
T	0.0283	***	0.0288	***	1.7%	***
TT	0.0176	***	0.0176	***	0.0%	
A	0.0735	***	0.0750	***	2.0%	***
L	0.1768	***	0.1748	***	-1.1%	***
K	0.1423	***	0.1397	***	-1.9%	***
V	0.5711	***	0.5716	***	0.1%	
A*T	0.0001		0.0007		85.7%	***
L*T	-0.0003		0.0021		114.3%	***
K*T	0.0073	*	0.0070	**	-4.3%	***
V*T	-0.0092		-0.0116	**	20.7%	***
AA	0.0102		0.0087		-17.2%	***
LL	0.1263	**	0.0882	**	-43.2%	***
KK	0.0713	***	0.0481	***	-48.2%	***
VV	0.023		-0.0147		256.5%	***
AL	-0.0771	***	-0.0592	***	-30.2%	***
AK	0.0104		0.0082		-26.8%	***
AV	0.0512	**	0.0416	**	-23.1%	***
LK	-0.0075		-0.0063		-19.0%	***
LV	-0.0569		-0.0324		-75.6%	***
KV	-0.0507	**	-0.0300	***	-69.0%	***
AM	0.179	***	0.1746	***	-2.5%	***
AM_T	0.0178	***	0.0173	***	-2.9%	***

AM_A	0.0267	***	0.0240	***	-11.3%	***
AM_L	0.0148	**	0.0155	***	4.5%	***
AM_K	0.0006		0.0025		76.0%	***
AM_V	-0.0697	***	-0.0714	***	2.4%	***
SV	0.1671	***	0.1681	***	0.6%	***
SU	2.0828	***	2.1287	***	2.2%	***
λ	2.232		2.211		0.9%	

Note: The significance of the difference was tested with a Welch t-test (Ruxton 2006).
Source: Own estimation

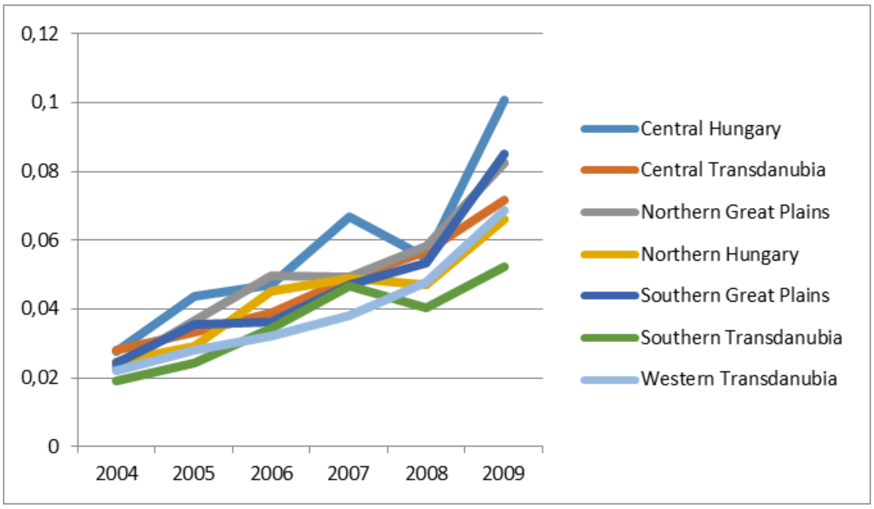
Including the restrictions in the estimation changes all parameter values. This is astonishing since only a few constraints were found binding. Our interpretation is that the binding constraint forms an envelope of all parameter constraints that forces the variation in the optimal parameters. The changes in in the first order parameters are very marginal compared to the changes in the second order parameters. The change of the second order parameter often exceeds often 30 % in one case even more 250 %. This confirms that is mainly the relation of the values in the matrix Bkk that determines whether the results are theoretically consistent or not. This concerns not only the pure absolute value of the parameter estimated but in many cases also the sign.

Moreover, the value of λ in the modified Alvarez model with constraints is the more or less the same than in the other models presented here. Together with a stable value of AM_T this implies that the distribution of inefficiency is similar to the histogram in Figure 1.

In the following we restrict our attention to the model with constraints. First, we move to the regional distribution of estimated efficiencies (Figure 2). In general the inefficiencies are very low, e.g. much of the variation of output is explained by technology and by the impact of unobserved heterogeneity. There are hardly seen regional differences in the data. The only pronounced effects in inefficiency are the pronounced reduction in efficiency in the year 2007 and 2009. These years we observed severe weather conditions which affected all regions in Hungary similarly. Moreover, there is a negative trend in efficiency over time. This can be seen in connection with the high impact of technical change in Hungarian grain production. The frontier is determined by the farms which apply the most modern technology. Farmers who hesitate do adopt these are falling more and more behind or show greater inefficiencies.

Figure 2

Regional development of inefficiencies, 2004-2009

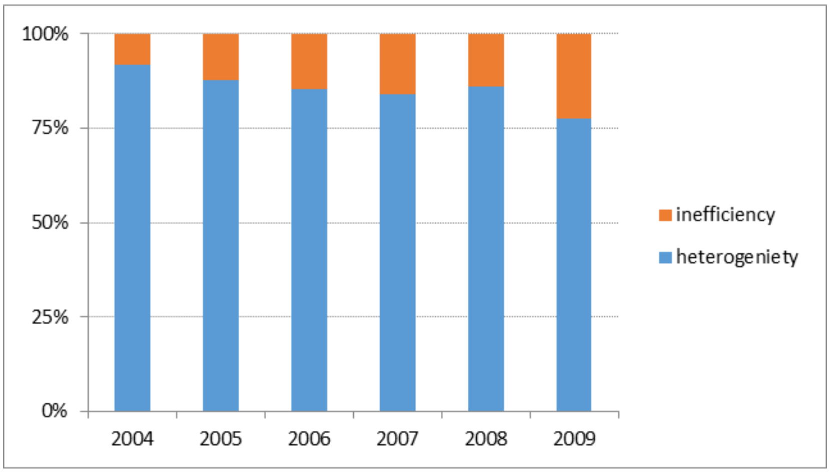


Source: Own estimation

It was already mentioned in the theoretical part of the paper that heterogeneity might have a greater impact on production than inefficiency. In order to check this conjecture the share of heterogeneity and inefficiency on the variation of output has been calculated (Figure 3). The results confirm that heterogeneity is much more important than inefficiency. The latter is only responsible for 5 – 20 % of the total variance of these sources. Not considering heterogeneity in the estimation can lead to totally wrong policy recommendation. Moreover, because of the expected overestimation of efficiency when neglecting heterogeneity, the policies may not only direct in the wrong direction, the impact on increasing efficiency measure may be overestimated as well. The increased impact of inefficiency at the end of the investigation period can be seen, as explained in the last paragraph, as a consequence of the high rate of technical change in the sector (see Table 4) and/or the severe weather condition at the end of the investigation period.

Figure 3

Joint impact of the variance of heterogeneity and efficiency

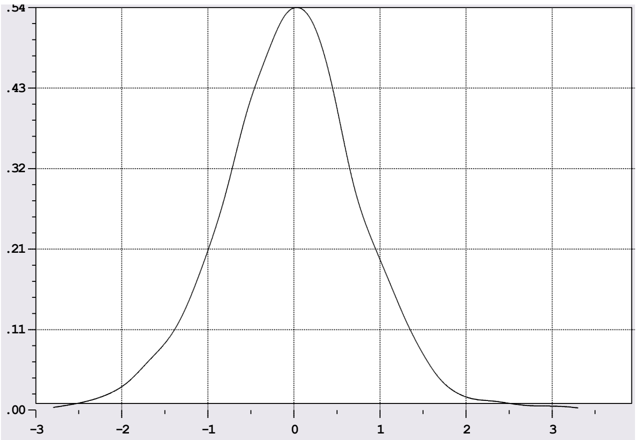


Source: own estimation

Having demonstrated the importance of heterogeneity for the variation of output it remains to discuss its impact on production and its sources in more detail. First, Figure 4 given the distribution of estimated m* in the dataset. According to our assumption (see Section 3) the mean of m* is about zero and ranges from -3 to +3. The kernel density estimation has not exact the form of a standard normal distribution, however, given that it is the results of a simulation process, the deviation appear acceptable

Figure 4

Density of m* in the sample



Kernel density estimate of m*

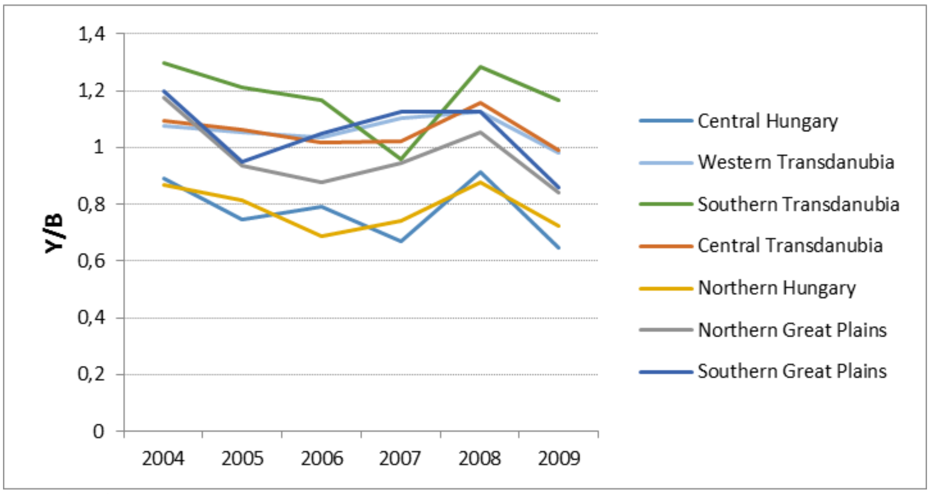
Source: own estimation

In the following we will investigate the source of heterogeneity in more detail. For this, we will refer to two partial productivities which can be viewed as indicator of the natural and economic factor of location.

Figure 5 shows the development of land productivity in the Hungarian regions (NUTS 2 level) over the investigation period. In this figure land productivity in the regions is measured relative to the average land productivity in the sample. Two regions shows below average land productivities: Northern Hungary and Central Hungary. The other regions have above average productivities with the highest in Southern Transdanubia. In this regions land productivity was about 20% higher than average. One more thing has to be considered: The ordering of the regions according to land productivity remains relatively constant over the years. This suggests that our conjecture that we could assume a constant term for the impact of heterogeneity is supported by the data.

Figure 5

Development of relative land productivity by region, 2004-2009



Source: own calculations

Figure 6 provides more details of the technologies used in Hungarian grain production. It relates the three indicators labour productivity (Y/A), land productivity (Y/B) and land-man ratio (B/A) for all regions over time according to (Herlemann and Stamer 1958, Hayami and Ruttan 1971):

$$\frac{Y}{A} = \frac{Y}{B} * \frac{B}{A}$$

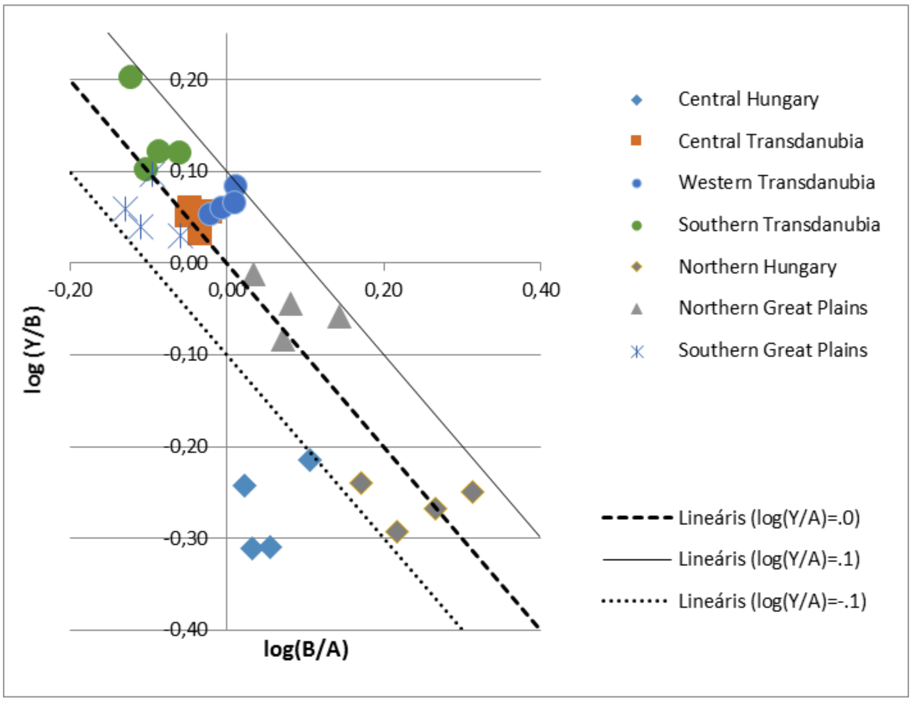
or in log terms:

$$\log\left(\frac{Y}{A}\right) = \log\left(\frac{Y}{B}\right) + \log\left(\frac{B}{A}\right)$$

Land productivity can be seen as an indicator for the natural condition of location. This indicator reflects basically soil conditions, sufficient water and sunshine etc. However, land productivity is also influenced by the level of economic development through the availability of sufficient production enhancing inputs like fertilizer or pesticide, however, given the homogeneous institutional conditions in Hungary it can be assumed that this impact is of minor importance for land productivity differences across regions. On the other hand differences in labour productivity can be more regarded as a consequence of the economic conditions of location. The opportunity cost of labour determines how much labour will be devoted to this sector. Moreover, in combination with factor prices the land man ratio determines the capital to labour ratio in the sector which not at least depend on the economic infrastructure in the region (Herlemann and Stamer 1958). However, we admit that the relation between economic condition of location and labour productivity is much weaker than between natural condition and land productivity. Differences in labour productivity can be also the consequence of various farm structures: usually large scale agriculture is less labour intensive than small scale agriculture but much more mechanized. The dual structure agricultural structure in Hungary provides some support for this interpretation. Given the lack of data unfortunately we are not able to dig deeper into this problem. However, independently whether the high land man ratio is the results of economic forces or whether the farm structures are due to political (institutional) decision, a higher labour productivity can be viewed as in indicator of better performance.

Figure 6

Partial productivities and man land ratio in Hungarian grain production 2005-2008 by regions



Note: Y/A land productivity; Y/B land productivity; B/A land man ratio
all variables are normalized by the geometric average of the total sample

Source: own calculations

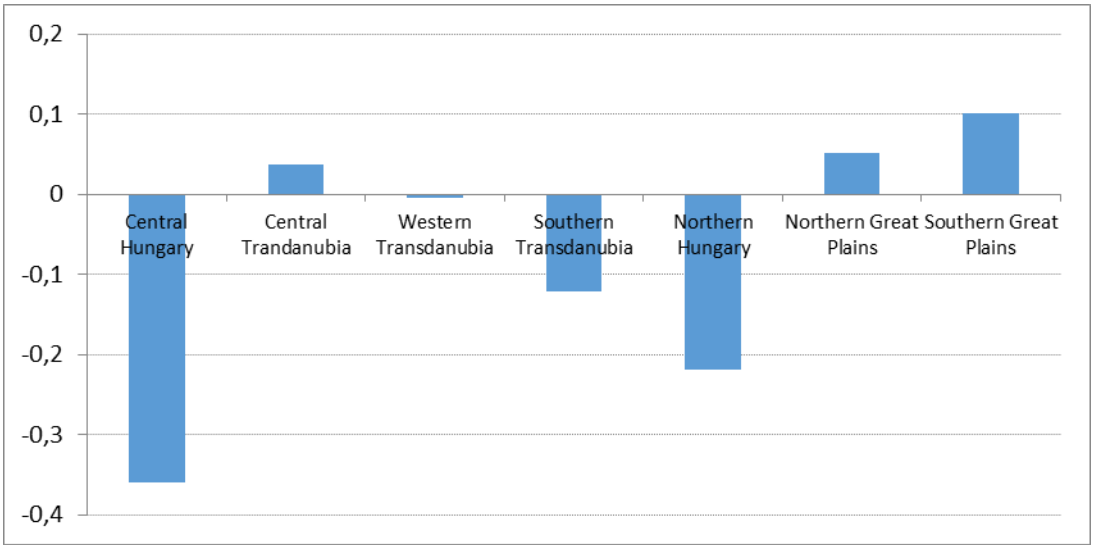
The numbers in Figure 6 represent 3 year averages. This calculation was conducted to eliminate outliers. Like in Figure 5 all number are in relation to the sample average. Thus on the horizontal axis, numbers larger than zero indicate above average land man ratios. A similar interpretation holds for the vertical axes for land productivity. The dashed line represents average labour productivity. Above this line are regions with higher labour productivities6. The figure confirms what has been deduced form the previous figure. Land productivity in Central Hungary and Northern Hungary is poorer than in the other regions. However, in Central Hungary the land man ratio is lower than in Northern Hungary. As a result, labour productivity in this region is the lowest of all Hungarian regions. A poor labour productivity is also observed in the Southern Great Plains. However, together with a low land ratio this allows nevertheless an above average land productivity. The region with a relatively high land and labour productivity and a high land man ratio is the Northern Great Plains. The high land man ratio suggests that agriculture or grain production is much more labour incentive than in other regions, especially the Southern Great Plains. The same holds for grain production in Northern Hungary.

In Figure 7 we investigate how the heterogeneity affected Hungarian agriculture. Central Hungary is the region where land and labour productivity were the lowest in our sample. This finds its expression in the highest negative m* values. North Hungary is similar. However, the labour productivity is higher than in Central Hungary. Accordingly, the m* value in this region are also negative, but also a bit higher than in Central Hungary. The highest m* values were found for the two Great Plains regions. In the northern part we have a high labour productivity, however, a low land productivity. The partial productivities in the southern part are opposite to the northern part: high land productivity and low labour productivity. In this region we found the highest value for m*. Another region with above average of m* value in Central Transdanubia. This region is characterized by above average land productivity and average labour productivity. Western Transdanubia doesn't fit into this pattern. Labour and land productivity are both relatively high but the average value of m* is only at an average level.

[6] The solid line represents labour productivity about 25% higher than average labour productivity. The dotted line correspondingly a labour productivity about 25% lower than the average.

Figure 7

Distribution of heterogeneity over Hungarian regions



Source: Own estimation

This analysis provide strong support for the view that unobserved heterogeneity can be regarded as an indicator the favourability of the regions in terms of their natural und economic conditions of production (Sec. 3). However, some part of the variation of m* is not explained this interpretation. In the model it was assumed that m* has a mean of zero and a variance of 1. Mean of zero is quite well captured since the estimated value of m* is around zero. However the variance is widely underrepresented and by far not homogeneous among the regions. Despite the differences are large enough that many pairwise comparison provided significant results. (see Appendix).

To sum up, land and labour productivities may provide some explanation of the variance of heterogeneity. The explanatory power of these two indicators appears to be relatively high but not sufficient. The actual value of m* depend on important other factor which might not be covered by natural and economic conditions. One candidate is that m* captures the distribution of management abilities of farmers (Alvarez et al. 2004). However, what are the reasons for different m* among farms cannot be uniquely decided form our analysis. More analyses are necessary to distinguish among the various sources of m*.

CONCLUSION

The goal of this paper is to model technological differences among farms and check how the consideration of technological heterogeneity affects structural parameters and technical efficiency. In order to model technological differences we used an RPM, which was originally proposed by Alvarez et al., 2014. However, the original version of the model might give biased result, because of the possible correlation between input variables and technical efficiency. Therefore, our first aim was to reformulate the model in such a form which avoids this possible correlation. Our second aim was to compare the results of the reformulated model with the original Alvarez model and with the TRE model, which can be seen as a restricted version of an RPM, where heterogeneity affects only the intercept, i.e. it is able to separate heterogeneity and technical efficiency, but insufficient to capture underlying variation in technology. Our third aim was to examine how the fulfilment of theoretical consistency influence the results. Our examination has a number of interesting methodological and agricultural policy implications.

The results revealed that in addition to the separation of unobserved heterogeneity and TE, the consideration of technological heterogeneity is also important and has significant effect on parameter estimates and TE. Our results showed that the interaction of unobserved heterogeneity was significant not only with the constant, but also with observed inputs and technological change, which reveal that heterogeneity influence also the applied technology and technological change. This suggest that a model which is able to estimate these different technologies and changes might lead to more adequate policy implication.

The comparison of the (1) modified Alvarez model with the (2) original Alvarez and (3) TRE model showed that TRE and the original Alvarez model give similar results which imply that the original Alvarez model might not capture well the underlying technological differences. In contrast, the modified Alvarez model yields substantially different results, suggesting that this model might estimate it better. Model selection criteria also suggest that the modified Alvarez Model is the preferred specification among the estimated models.

The modified Alvarez model yields reasonable results also in terms of technical (in)efficiency. The technical inefficiency estimates are much lower in this model. From our point of view this results is very convincing, While in the other two model a large share of the farm have efficiency lower than 25% (e-u, u > .3) but the shares of firm with low efficiency reduces to a marginal fraction in this model. Given the farms are operating in a homogeneous institutional environment the wide span of inefficiencies are difficult to justify. Moreover, the parameter value for intermediate inputs in the modified Alvarez is much closer to the actual value of intermediate input share on total revenues provided by the data. Thus we conclude that the modified Alvarez model depicts the inefficiency much more appropriately than the two other models.

The comparison of the modified Alvarez model with and without constraints showed that in the model with constraints the percentage of consistent estimates increased substantially, suggesting that these results can be interpreted even more adequately. The inclusion of the restrictions changes all parameter estimates. However, the changes in the first order parameters are very marginal compared to the changes in the second order parameters. The change of the second order parameters often exceed 30% in one case even more, 205%. This confirms that is mainly the relation of the values in the matrix Bkk that determines whether the results are theoretically consistent or not.

As the model with restrictions give theoretically more consistent result, we focused only on these results to examine some empirical question with important policy implication, especially we focused on regional differences and potential factors which determine unobserved heterogeneity.

In general the inefficiencies are very low, e.g. much of the variation of output is explained by technology and by the impact of unobserved heterogeneity. There are hardly seen regional differences in the data. Moreover, there is a negative trend in efficiency over time. This can be seen in connection with the high impact of technical change in Hungarian grain production. The frontier is determined by the farms which apply the most modern technology. Farmers who hesitate do adopt these are falling more and more behind or show greater inefficiencies. This means at the same time also that there is a divergence among Hungarian farms.

The results showed that heterogeneity has much more important effect on the variance of output than inefficiency. The latter is only responsible for 5 – 20 % of the total variance of these sources. This clearly confirm that not considering heterogeneity in the estimation can lead to totally wrong policy recommendation. Moreover, because of the expected overestimation of inefficiency when neglecting heterogeneity, the policies may not only direct in the wrong direction, the impact on increasing efficiency measure may be overestimated as well. For instance, the increased impact of inefficiency at the end of the investigation period can be seen as a consequence of the high rate of technical change in the sector and/or the severe weather condition at the end of the investigation period.

Our results confirms that unobserved heterogeneity can be regarded as an indicator the favourability of the regions in terms of their natural und economic conditions of production. In fact, a large part of the variation of m* can be seen to be explained this interpretation. Variations of land and labour productivities may provide a good explanation of the variance of heterogeneity. However, the explanatory power of these two indicators appears to be not sufficient. Consequently, the modelled heterogeneity effect (m*) capture not only the natural and economic conditions but important other factors too. One candidate is that m* captures the distribution of management abilities of farmers (Alvarez et al. 2004). However, the determinants of different m* among farms cannot be uniquely decided form our analysis. More analyses are necessary to distinguish among the various sources of m*.

Appendix 1

Significance of regional differences

	Regional mean of m*	Standard deviation of m* in the region	Number of enterprises in the region	Central Hungary	Central Transdanu.	Western Transdanu.	Southern Transdanubia.	Northern Hungary	Northern Great Plains
Central Hungary	-0.3596	0.7275	53						
Central Transdanubia	0.0373	0.6694	72	***					
Western Transdanubia	-0.0045	0.8554	82	***					
Southern Transdanubia	-0.1208	0.6292	102	**	*				
Northern Hungary	-0.2193	0.8582	62		**				
Northern Great Plains	0.0516	0.8707	117	***			*	*	
Southern Great Plains	0.1013	0.7767	176	***			***	***	

Source: own calculations

REFERENCES

Abdulai, A. and Tietje, H. (2007): Estimating technical efficiency under unobserved heterogeneity with stochastic frontier models: application to northern German dairy farms. European Review of Agricultural Economics. Vol. 34. (3) 393-416.

Agrell, P.J., Brea-Solís, H. (2015): Stationarity of Heterogeneity in Production Technology using Latent Class Modelling. CORE DISCUSSION PAPER 2015/47

Aigner, D.J., C.A.K. Lovell and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," Journal of Econometrics, 6, 21-37.

Alvarez, A. and del Corral, J., (2010): Identifying different technologies using a latent class model: extensive versus intensive dairy farms. Eur. Rev. Agric. Econ. 37 (2), 231–250.

Alvarez, A., Arias, C. and Greene, W. (2004): Accounting for unobservables in production models: management and inefficiency. Working Paper. Fundación Centro de Estudios Andaluces. Serie Economía, E2004/72. 1–18. o.

Alvarez, A., del Corral, J. and Tauer, L. W. (2012): Modelling unobserved heterogeneity in New York dairy farms: one-stage versus two-stage models. Agric. Resour. Econ.Rev. 41 (3), 275–285.

Bauer, P. W., Berger, A. N. and Ferrier, G. D. – Humphrey, D. B. (1998): Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Method. Journal of Economics and Business Vol. 50 (2) 85-114.

Belyaeva, M., Hockmann, H. and Koch, F., (2014): Impact of regional diversity on production potential: an example of Russia. 142th EAAE Seminar. May 29-30, 2014. Corvinus University of Budapest.

Bhat, C. (1999): Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Logit Model. Msp., Department of Civil Engineering, University of Texas, Austin

Cambridge University Press.

Checura, L., Grau, A., Hockmann, H., Levkovych, I. and Kroupova, Z. (2014): Catching up or falling behind in Eastern European agriculture – the case of milk production. 142nd EAAE

Coelli, T, J, D.S. Rao, D. S. P., Q'Donnell, C. J. and Battese G. E. (2005): An introduction to efficiency and productivity analysis (2nd edition). Boston, MA: Springer US.

Diewert, W.E., and T.J. Wales. 1988. "A Normalized Quadratic Semiflexible Functional Form." Journal of Econometrics 37: 327–342.

Eberhardt, M. and Teal, F. (2013): No Mangoes in the Tundra: Spatial Heterogeneity in Agricultural Productivity Analysis. Oxford Bulletin of Economics and Statistics, 6 (2013) 0305-9049

Emvalomatis, G. (2007): A thesis in Agricultural, Environmental and regional Economics August 2007. Distinguishing Technologies and Measuring Performance in Farm Production: the Case of Pennsylvania Dairy. The Pennsylvania State University, pp. 56 pp.

Farsi, M., Filippini, M., Kuenzle, M. (2005). Unobserved heterogeneity in stochastic cost frontier models: an application to Swiss nursing homes. Applied Economics 37, 2127–2141.

Fried, H. O, Lovell, C.A. K. and Schmidt, S. S. (2008): The measurement of productive efficiency and productivity growth. Oxford University Press

Fuss, M., and McFadden, D. (1978): Production Economics: A Dual Approach to Theory and Applications. North-Holland Publishing Company.

Greene W. H. (1986-2007): LIMDEP, Version 9.0 Econometric Modelling Guide Vol.2. Ecomometric Software, Inc. Plainview and Castle Hill.

Greene. W. (2005): Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126: 269-303

Hayami, Y. and V. W. Ruttan (1971): Agricultural Development. Baltimore and London.

Herlemann, H.-H. and Stamer, H. (1958): Produktionsgestaltung und Betriebsgröße in der Landwirtschaft unter dem Einfluss der wirtschaftlich-technischen Entwicklung. Kieler Studien. Nr. 44. Kiel.

Hsiao, C. (2014): Analysis of Panel data (3rd edition). New York, Cambridge University Press

Jondrow, James, C A Knox, I.S. Materov, and Peter Schmidt. 1982. "On the Estiamtion of Technical Inefficiency in the Stochastic Frontier Production Function Model." Journal of Econometrics 19: 233–238.

Keszthelyi, Sz., Pesti, Cs., 2009. A Tesztüzemi Információs rendszer 2008. évi ered-ményei (in Hungarian with English abstract). Agrárgazdasági Információk. 3. AKI, Budapest, pp. 45 pp.

Kumbhaker, S.C. and Knox Lovell, C.A. (2000). Stochastic Frontier Analysis. Cambridge:

Kunezle, M. (2005): Cost Efficiency in Network Industries: Application of Stochastic Frontier Analysis. A dissertation submitted to the Swiss Federal Institute of Technology Zurich. Diss. ETH No. 16117

Láng, I., L. Csete and Zs. Harnos, 1983. The agro-ecopotential of the Hungarian Agriculture at the turn of 2000. Mezőgazdasági Kiadó, Budapest, pp. 1-266. (in Hungarian).

Lau J. L. (1978): Testing and Imposing Monotonicity, Convexity and Quasi-Convexity Constraints. In Fuss, M. and McFadden, D.: Production Economics: A Dual Approach and Application. Volume 1. The Theory of Production. North Holland Publishing Company. Amsterdam, New York, Oxford

Meeusen, W. and J. van den Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," International Economic Review, 18, 435-444.

Pitt, M. and Lee, L. (1981). The measurement and sources of technical inefficiency in indonesian weaving industry. Journal of Development Economics 9: 43–64.

Ruxton, G. D. (2006): The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. Behavioral Ecology, Vol. 17, pp. 688–690.

Sauer, J., Morrison Paul, C.J., 2013. The empirical identification of heterogeneous technologies and technical change. Appl. Econ. 45 (11), 1461–1479.

Sauer. J., Hockmann, H. (2005): The need for theoretically consistent efficiency frontiers. Paper prepared for presentation at the XIth EAAE congress, Copenhagen, Denmark, August 24-27, 2005

Schmidt, P. and Sickles, R. (1984). Production frontiers with panel data. Journal of Business and Economic Statistics 2: 365–374.

Train, K. (1999) Halton Sequences for Mixed Logit. Manuscript, Department of Economics, University of California, Berkeley.

Wang, X., Hockmann, H. and Bai, J. (2012): Technical Efficiency and Producers' Individual Technology: Accounting for Within and Between Regional Farm Heterogeneity, *Canadian Journal of Agricultural Economics*, Vol. 60, No. 4, pp.561-576.