Applying linear mixed model to farm panel data set

Alina Sinisalo[1]

[1]MTT Agrifood Research Finland, e-mail: alina.sinisalo@mtt.fi

August 2014

Abstract

The development of production costs were studied in Finnish farms in 2000-2011 with a linear mixed model taking into account farm-level information (production type, economic size, arable land) and time effect. Interindividual differences in intraindividual changes over time were analyzed. The production cost increased over time and as the size of arable land and standard output increased.
Keywords: linear mixed model, panel data, farm, cost, Finland

1 Introduction

The structure of agriculture has changed. In 2000-2011 the overall number of farms has dropped by 24% (pig farms 59%, dairy farms 52%, cattle farms 45%, horticulture outdoor 43%, poultry farms 39%, horticulture indoor 38%, mixed production 17%, cereal farms 9%). On the contrary the number of sheep and goat husbandry farm increased by 9% and other crop farms by 13% (Tike, 2014). During the same period the average farm size has grown, i.e. farms have more animals and more arable land.

This study focuses on the development of production costs in 2000–2011 on Finnish farms by production type. The prices of goods and services currently consumed in agriculture (+47%) and goods and services contributing to agricultural investments (+42%) have increased faster than the Consumer Price Index that is used to measure general inflation rate (+21%) (OSF, 2014a,b).

In almost all production types the production costs have increased: sheep and goat husbandry 220%, pig farms 111%, cattle farms 107%, dairy farms 92%, poultry farms 69%, cereal farms 64%, mixed production 55%, horticulture outdoor 43%, and horticulture indoor 28%. On the contrary the production costs on other crop farms have decreased by 26% (MTT, 2014).

Quite often production cost studies are based on average results (MTT, 2014), one year observations (Riepponen, 2003; Rantala, 1997) or different kind of models (Ala-Mantila, 1998). There are few studies in which the panel characteristics of dataset are implemented. The goal of this study was to study how the production costs in farms have developed in 2000s taking into account farm-level information (production type, economic size, arable land) and time effect by observing the same farms for several years, and to explore possible differences in the development between farms.

2 Materials

Farms participating in MTT profitability bookkeeping were studied for the years 2000–2011. The data set was formed as panel. Each farm was repeatedly measured in one year intervals. There were 11115 observations from 1564 different farms and on average 926 different farms every year in the data. Hence, the data set was unbalanced. This is due to the fact that it is voluntary to participate in MTT bookkeeping activities and, on the other hand, some farms had exited the business.

In this study the total production costs (continuous dependent variable) were studied. The total is sum of following components: material, livestock, machinery, building, wages and interest costs. Only costs targeted to agriculture were included; costs targeted to forestry and private operations were not studied. The production costs were deflated by using Consumer price indices (OSF, 2014a) year to 2011 prices (2000=100).

The farm-level data were weighted with weight factors calculated individually for each farm taking into account the type of operations, economic size and location by support areas. Weights were calculated for each farm by stratum indicators separately for every year. Furthermore weights were calibrated taking into account the total arable land in Finland. Thus, after weighting, the used data can be used to describe the results of all Finnish farms.

In this study 10 different production types were included: cereal farms, other crop farms, horticulture indoor, horticulture outdoor, dairy farms, cattle farms, sheep and goats, pig farms, poultry farms and mixed production. Cereal farms were compared to the other production types. Farm size is often described by arable land, and it was included in the model as continuous explanatory variable describing the average amount of arable land for each year.

The standard output (SO) is the average monetary value of the agricultural output at farm-gate price per unit. There is a regional standard output coefficient for each product (average value over 5 year period). The sum of all standard outputs per unit multiplied by hectares or headcount forms the overall economic size, expressed in euro (Eurostat, 2014). We classified the standard output variable into three classes. We chose the smallest class (SO less than 50000 €) as basis for testing so that the medium class (SO 50000–100000 €) and the largest class (SO more than 100000 €) were compared to the smallest class.

3 Methods

A linear mixed model includes both fixed and random effects. A fixed effect model or a random effect model can be considered as a special case of a mixed effects model.

The linear mixed model for an individual farm, i, was defined as followed:

$$\begin{cases} \mathbf{y}_i = \underbrace{\mathbf{X}_i\boldsymbol{\beta}}_{\text{fixed}} + \underbrace{\mathbf{Z}_i\mathbf{b}_i}_{\text{random}} + \underbrace{\boldsymbol{\epsilon}_i}_{\text{random}}, \\ \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D}), \\ \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i), \\ \mathbf{b}_1, \dots, \mathbf{b}_n, \boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n \text{ independent,} \end{cases} \tag{1}$$

where $\mathbf{y}_i$ is $n_i \times 1$ response vector for farm i, $\boldsymbol{\beta}$ denotes a $p \times 1$ vector of unknown population parameters and $\mathbf{X}_i$ denotes a known $n_i \times p$ design matrix linking $\boldsymbol{\beta}$ to $\mathbf{y}_i$. The $\mathbf{b}_i$ denotes a $k \times 1$ vector of unknown individual effects and $\mathbf{Z}_i$ a known $n_i \times k$ design matrix linking $\mathbf{b}_i$ to $\mathbf{y}_i$. The $\mathbf{b}_i$ are distributed as $\sim N(\mathbf{0}, \mathbf{D})$ (normal with mean 0 and covariance matrix $\mathbf{D}$), independently of each other and of the $\boldsymbol{\epsilon}_i$ that are distributed as $\sim N(\mathbf{0}, \mathbf{R}_i)$. The $\mathbf{D}$ is a $k \times k$ and $\mathbf{R}_i$ is an $n_i \times n_i$ positive-definite covariance matrix. Parameters $\boldsymbol{\beta}$ are treated as fixed effects and $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ as random effects. In our unbalanced panel data set the n_i varied between 1 to 12.

Unstructured (UN) covariance structure was chosen for random effects in the model since it is suitable for longitudinal data. Random effects were defined over farm register number (observation unit i). For residual random effects first-order autoregressive (AR1) covariance structure was chosen because it is suitable for data containing sequential observations and correlations declining exponentially with time.

Six fixed effects, p, and two random effects, k were included in the model. Intercept β_0, time β_1, land β_2, standard output $\beta_3 ... \beta_{12}$, standard output $\beta_{13} ... \beta_{15}$ and weight β_{16} were treated as fixed effects. Intercept b_0 and time b_1 were included in the model also as random effects. The unstructured 2×2 covariance matrix for the random effects is denoted as followed:

$$\mathbf{D} = \mathrm{Var}(\mathbf{b}_i) = \begin{pmatrix} \sigma^2_{b_0} & \sigma_{b_0,b_1} \\ \sigma_{b_0,b_1} & \sigma^2_{b_1} \end{pmatrix}, \tag{2}$$

where three parameters, b_0 variance, b_1 variance, b_0 and b_1 covariance, are denoted as UN(1,1), UN(2,2) and UN(2,1), respectively. The first-order autoregressive covariance matrix for residual is denoted

$$\mathbf{R}_i = \mathrm{Var}(\boldsymbol{\epsilon}_i) = \begin{pmatrix} \sigma^2 & \sigma^2\rho & \dots & \sigma^2\rho^{n_i-1} \\ \sigma^2\rho & \sigma^2 & \dots & \sigma^2\rho^{n_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho^{n_i-1} & \sigma^2\rho^{n_i-2} & \dots & \sigma^2 \end{pmatrix}. \tag{3}$$

The total number of different farms in our data set was $N = 1564$. N vectors of form (1) can be stacked

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_N \end{pmatrix}$$

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_N \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{pmatrix},$$

and presented as followed:

$$\begin{cases} \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \\ \mathbf{b} \sim N(\mathbf{0}, \mathbf{G}), \\ \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R}), \\ \mathrm{Cov}[\mathbf{b}, \boldsymbol{\epsilon}] = \mathbf{0} \end{cases} \tag{4}$$

observations between farms, i, being independent.

Model (4) parameters are included in the fixed effect vector $\boldsymbol{\beta}$ and unknown variables in covariance matrices **G** and **R**. Explanatory variables are included in **X** and **Z**. Vector **b** includes random effect variables.

4 Results

The results of linear mixed model explaining the production cost are presented in Table (1).

Livestock farms have higher production costs compared to cereal farms. Costs increase year-to-year ($time = 6073$). As arable land is increased, also costs increase ($land = 825$). If farms are categorized by their standard output, small farms (standard output <50000 €) have greater production costs than medium-sized farms (50000–100000 €) and large farms (>100000 €). The results show that year-to-year correlation is strong (0.900, $p<0.001$). Productions costs change at different pace between farms ($p<0.001$).

5 Discussion

Linear mixed model presented in this study fits well for the data set, Finnish farms. Values predicted with the model slightly overestimate the average values besides for poultry farms, horticulture indoors and outdoors.

Table 1: The estimates for model parameters.

Effect	Parameter	Estimate	S.E.	p-value
intercept	β_0	78275	4595	<0.001
time	β_1	6073	469	<0.001
land	β_2	825	46	<0.001
Production type				
Other crop farms	β_3	1766	2324	0.447
Sheep and goat husbandry	β_4	8316	12979	0.641
Horticulture outdoor	β_5	12641	6015	0.036
Mixed production	β_6	20974	2891	<0.001
Cattle farms	β_7	26952	4011	<0.001
Dairy farms	β_8	31561	3896	<0.001
Pig farms	β_9	39122	4676	<0.001
Poultry farms	β_{10}	48889	6798	<0.001
Horticulture intdoor	β_{11}	218666	10245	<0.001
Cereal farms	β_{12}	0	0	
Standard output				
50000–100000 €	β_{13}	10068	1866	<0.001
100000–	β_{14}	25727	2377	<0.001
0–50000 €	β_{15}	0	0	
Weight factor				
weight	β_{16}	-2.599	3.759	0.489
Covariance parameters				
UN(1,1)	σ_0^2	4.253×10^9	1.012×10^9	<0.001
UN(2,1)	$\sigma_0\sigma_1$	0.829×10^9	0.059×10^9	<0.001
UN(2,2)	σ_1^2	0.162×10^9	0.012×10^9	<0.001
Residual				
AR1 diagonal	σ^2	6.846×10^9	0.815×10^9	<0.001
AR1 rho	ρ	0.900	0.010	<0.001
Observations		11115		
-2 REML log-likelihood		269583		
AIC		269593		
BIC		269630		

Weight factors were calculated for each observation and weights were included in model as continuous explanatory variable that, however, was not significant. The inclusion of weights should be further studied.

The forecasting possibilities should be further studied.

References

Ala-Mantila, O. (1998). *The production costs of agricultural products according to farms models (in Finnish with abstract in English)*. Research reports 222. Helsinki: Agricultural Economics Research Institute, pp. 6–93.

Eurostat (2014). *Glossary: Standard output*. Available at http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Standard_output_(SO).

MTT (2014). *Agriculture and horticulture. Data: MTT profitability bookkeeping results*. Available at http://www.mtt.fi/economydoctor.

OSF (2014a). *Official statistics of Finland. Consumer price index*. Available at http://www.stat.fi/til/khi/index_en.html.

OSF (2014b). *Official Statistics of Finland. Index of purchase prices of the means of agricultural production*. Available at http://www.stat.fi/til/ttohi/index_en.html.

Rantala, J. (1997). *Costs of dairy production*. Reports and discussion papers 151. Helsinki: Pellervo Economic Research Institute.

Riepponen, L. (2003). *Production costs of milk and cereals on Finnish bookkeeping farms in 1998—2000 (In Finnish with English abstact)*. Maa- ja elintarviketalous 19. Helsinki: MTT Agrifood Research Finland. Available at http://urn.fi/URN:ISBN:951-729-736-x.

Tike (2014). *Structural data of agriculture*. Available at http://www.maataloustilastot.fi.