Modelling the Dairy Farm Size Distribution in Poland Using an Instrumental Variable Generalized Cross Entropy Markov Approach

Axel Tonini[*], *Roel Jongeneel*[**]

[*] European Commission, Joint Research Centre (JRC), Institute for Prospective and Technological Studies (IPTS), Agriculture and Life Sciences in the Economy (AGRILIFE Unit), Seville, Spain. E-mail: axel.tonini@ec.europa.eu .

[**] Agricultural Economics and Rural Policy Group, Wageningen University, Wageningen and Agricultural Economics Institute, The Hague, both in The Netherlands. E-mail: roel.jongeneel@wur.nl .

Paper prepared for presentation at the joint IAAE - 104th EAAE Seminar

Agricultural Economics and Transition:

„What was expected, what we observed,

the lessons learned."

Corvinus University of Budapest (CUB)

Budapest, Hungary. September 6-8, 2007

ABSTRACT

The aim of this paper is to analyse the evolution of the dairy farm structure of Poland during the post-socialist period. First the paper focuses on how the farm structure has changed over time and what path it is likely to follow in the coming decade. Second, it is tested whether the evolution of farm size is explained by non-stationary effects. Finally, several statistical indicators are computed on farm mobility and on which farms are likely to survive. An instrumental variable generalised cross entropy Markov chain approach which incorporates prior information is applied for estimation. Prior information included general and plausible information on farm mobility and structural adjustments based on independent literature. The projections show that dairy farm numbers will continue to decline, although accompanied by an increase in the number of medium-sized and large farms. Subsistence dairy farms are expected to slowly leave the sector in the coming decade.

Keywords: dairy, farm size, Poland, Markov chain, generalised cross entropy.

1 INTRODUCTION

The aim of this paper is to analyse the evolution of the dairy farm structure of Poland during the post-socialist period. This analysis is of interest for policy makers in providing insight into how the farm structure is likely to evolve over time. A relevant issue is what will happen to the subsistence and semi-subsistence farms in the restructuring process. Last but not least, the analysis is of interest also for the upstream and downstream industries that have to decide on investments in dairy processing capacity, milk collection schemes, and providing farm input supplies. The objectives are threefold. First the paper focuses on how the farm structure has changed over time and to what path it is likely to follow in the coming decade by making several projections. Second, it is tested whether the evolution of farm size is explained by non-stationary effects. Finally, several statistical indicators are computed on farm mobility and on which farms are likely to survive.

This study employs a Markov probability model (LEE et al. 1970) of farm size distribution which is able to analyse movements of individuals between different states when only aggregate data on finite size categories are available for a given time period. A generalised cross entropy (GCE) estimator is used (see GOLAN et al. 1996 and MITTELHAMMER et al. 2000). Entropy estimators are particularly suited when dealing with limited data which is often the case for empirical applications on Central Eastern European Countries (CEECs). This paper further extends the approaches of GOLAN and VOGEL (2000), COURCHANE et al. (2000), KARANTININIS (2002) and JONGENEEL et al. (2005) by allowing for a heteroscedastic version of the set of Markov equations and for seemingly unrelated regressions (SUR) estimation. Assuming a common and constant variance matrix across the different Markov states, as it is done for example in KARANTININIS (2002) and JONGENEEL et al. (2005) could easily create bias on the estimated Markov transition probabilities affecting related indicators as well as projections.

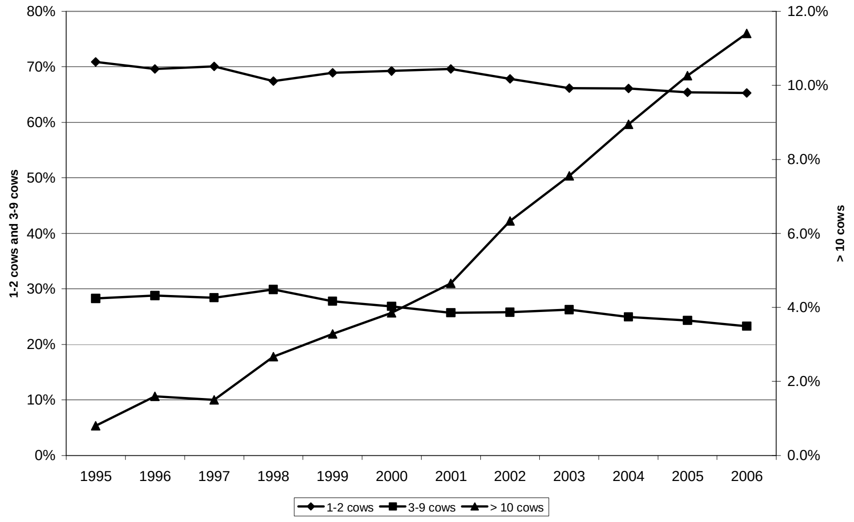
The remainder of this paper is organized as follows. Section 2 describes the farm structure of Poland, with a focus on dairy farming. Section 3 specifies the Markov chain entropy formalism. Section 4 discusses the sample data as well as prior information. Section 5 discusses results. In Section 6 the conclusions are presented.

2 FARM STRUCTURE IN POLAND, WITH A FOCUS ON DAIRY FARMING

Poland is one of the most important dairy producers in the European Union (EU). In 2005 it accounted for about 8 percent of the total EU-27 cow milk production, being the fourth EU milk producers after Germany, France and United Kingdom. In the last five years, dairy cow numbers have declined by 9.4 per cent and milk yields have improved by 15.1 per cent (FAOSTAT 2006). Since the socialist regime, the Polish dairy sector has presented a highly fragmented dairy farm structure, with a large number of small private family farms, just as in other sectors of agriculture. In 1987, about 67 per cent of the dairy farms had only 1-2 cows and these accounted for 41 per cent of the national herd. The number of private dairy farms had already shrunk greatly before transition by about 25 per cent from 1981 to 1987. Dairy cow numbers declined concomitantly. At the beginning of transition, about 80 per cent of the national milk production was produced from farms with 10 cows or less (SZNAJDER 2002, pp. 242-244).

In Poland, dairy producers after the transition reform can be classified in three main categories: farmers with 1-2 cows, who produce milk mostly for the farm household (i.e. subsistence dairy farms); farmers with more than 3-4 cows, who produce milk for sale in local markets and for their own needs (i.e. semi-subsistence dairy farms); and farmers with more than 10 cows, who produce almost exclusively for the dairy industry (SZNAJDER 2002, p. 248). In 1996, about one quarter of Polish milk was produced by almost 1 million of individual farms holding 1 to 3 cows, while half was produced by farms with 3 to 9 cows (EUROPEAN-COMMISSION 1998, p. 36). This underscores the great fragmentation of Polish milk production even after transition. In 2005 there were about 700 000 dairy farms: a decline of about 51 per cent as compared with the number of farms in 1995. In the same year, about 65 per cent of the farms with dairy cows belonged to subsistence farms with 1-2 cows (Figure 1) and about 53 per cent of the dairy cow stock was concentrated in farms with 1-9 cows. The Polish Ministry of Agriculture forecasts a 76 per cent decline in the number of total farms from 1996 to 2010 (AGRAEUROPE 2000, pp. 18-19). A first inspection of Figure 1 suggests that the evolution of Polish dairy farms proceeded without being affected by the EU milk quota system which was announced in 2004 and effectively introduced in 2006. In addition it appears that the size class with 3 to 9 cows constitutes the switch size class: farms with smaller herd sizes (i.e. dairy farms with 1-2 cows) show a tendency to decline, whereas for farms with larger herd sizes (i.e. dairy farms with more than 10 cows) the opposite holds. This suggests that part of the dairy farms in the size class with 3 to 9 cows will go out of business, scale down and scale up to large farm size classes.

Figure 1: Dairy farms in Poland, 1995-2006



Note: Percentages are expressed relatively to the total number of active dairy farms.
Source: Own calculations based on KRAWIECKA (2006).

3 AN INSTRUMENTAL VARIABLE GENERALISED CROSS ENTROPY MARKOV CHAIN

The Markov chain approach is very suitable when the only data available are count data in the form of observable proportions or aggregates rather than data at the level of micro units. Movements from state to state are represented by a stochastic process and are typically modelled by estimating the so-called Markov transition probabilities. It is often the case that the proportions/count data are only available for the total aggregate and not for the net shifts, so that the number of unknowns in terms of transition probabilities to be estimated might exceed the number of available data points. In this context, the maximum entropy (ME) algorithm developed in GOLAN et al. (1996), FOMBY and CARTER HILL (1997) and MITTELHAMMER et al. (2000) is a suitable candidate for extracting the maximal signal from an initial 'out-of-focus' problem.

This paper is based on a GCE formalism which is founded on the directed divergence or minimal discriminability principles of KULLBACK (1959) and GOOD (1963). GCE is suitable when some 'educated' guesstimates based on previous data, experiments or economic theory are available. GCE selects out of all feasible solutions the one that minimizes the divergence between the data and the priors, the final solution being the closest to the data and priors. Considering the dynamic farm growth process in a Markov problem, it is possible to envisage that farm growth can be explained by non-stationary effects. Several economic variables are then expected to affect the unknown transition probabilities[1]. Applying the formulation as developed in GOLAN and VOGEL (2000) and COURCHANE et al. (2000)[2], it is possible to assess the impact of key variables on the Markov transition probabilities therewith potentially improving the explanatory power of the model. In formalizing the problem, the non-stationary GCE Markov problem can be formulated as follows:

$$\min I\left(p_{lk}, q_{lk}, w_{tkh}, u_{tkh}\right) = \sum_{l}\sum_{k} p_{lk} \ln\left(p_{lk} / q_{lk}\right) + \sum_{t}\sum_{k}\sum_{h} w_{tkh} \ln\left(w_{tkh} / u_{tkh}\right) \qquad (1)$$

---

1 For example, a literature review suggests that out of all possible covariates the following appear to be likely to affect the transition probabilities of dairy farms: technological shift, milk price, feed price, dairy cow stock price (see GODDARD et al. 1993, ZEPEDA 1995b and KARANTININIS 2002).

2 One limitation of this approach is that the type of covariates cannot differ across the different Markov states.

subject to the following constraints:

$$\sum_t z_{tn} y_{tk} = \sum_t \sum_l z_{tn} x_{tl} p_{lk} + \sum_t z_{tn} e_{tk} \ , \quad \forall n = 1, \ldots, N, \text{ and } \forall k = 1, \ldots, K \tag{2}$$

with

$$e_{tk} = \sum_h V_{tkh} w_{tkh} \tag{3}$$

and

$$\sum_k p_{lk} = 1 \tag{4}$$

$$\sum_h w_{tkh} = 1 \tag{5}$$

Equation (1) represents the GCE criterion which minimizes the divergence between the data in the form of posterior transition probabilities p_{lk} and the transition priors q_{lk}[3]; p_{lk} denotes the probability a farm in size class *l* at time *t* moves to size class *k* at time t+1. Probabilities p_{lk} are elements of a $L \times K$ squared matrix of transition probabilities where *l, k* =1,…, *K* and q_{lk} are the counterpart prior elements; w_{tkh} are the elements of a $TKH \times 1$ vector of error posterior probabilities and u_{tkh} are the counterpart prior elements. Equation (2) represents the Markov data consistency constraints, where y_{tk} are the elements of a $TK \times 1$ vector of known proportions falling in the *k*-th Markov states in time (t+1), x_{tl} are the elements of a $TL \times 1$ vector of known proportions falling in the *l*-th Markov states in time (t). The covariates z_{tn}, which operate like instrumental variables, are forming a $T \times N$ matrix, explaining the non-stationarity effects[4].

The error term e_{tk}, included in equation (2), is reparameterised as given by equation (3) following the classical maximum entropy formalism (GOLAN et al. 1996, pp. 107-110), where $\mathbf{V}_{tk}$ is an *H*-dimensional vector of support points and $\mathbf{w}_{tk}$ is an *H*-dimensional vector of proper probabilities with $H \geq 2$[5]. Given that each Markov state can be characterized by a different variance as such a specific definition of support bounds for each Markov size class is desired. Specification of a common and constant variance for each Markov states in such a case can lead to specify relatively large support bounds for size classes where the variance is relatively small. As a consequence of this the final Markov probabilities estimates for these size classes are likely to converge to the prior estimates and underutilize the information present in the sample data. To avoid this, size class-specific variances are specified, following

---

[3] By analogy, the GCE criterion minimizes also the divergence between the error in the form of posterior probabilities w_{tkh} and the priors u_{tkh} where u_{tkh} are taken to be uniform since no prior information is available on the error term.

[4] The alternative simpler Markov stationary problem can be obtained by simply withdrawing the covariates z_{tn} from equation (2).

[5] In defining the $\mathbf{V}_{tk}$ vector, several choices can be followed. One possibility would be to set $\mathbf{V}_{tk} = [-1, \ldots, 0, \ldots, 1]$ given that the Markov states are expressed in terms of proportions/shares and y_{tk} and x_{tl} follow in a range between zero and one. A second possibility would be to set $\mathbf{V}_{tk} = \left[-1/K\sqrt{T}, \ldots, 0, \ldots, 1/K\sqrt{T}\right]$ where K is the number of states and T number of years as suggested in GOLAN and VOGEL (2000), COURCHANE et al. (1991), and KARANTININIS (2002). Both choices although empirically plausible assume a common and constant variance matrix across the different Markov states.

the statistical model presented in GOLAN et al. (1996, pp. 182-185). In so doing different error support bounds are specified for each Markov states relying on the 3σ rule of PUKELSHEIM (1994) based on the empirical standard deviation of y_k. Equation (4) represents the set of additivity constraints for the required Markov row constraint, while equation (5) does so for the proper probabilities of the reparameterised error. All proper probabilities of signal and noise are required to be non-negative $(\mathbf{p},\mathbf{w}) >> 0$. The minimization of (1) subject to (2) - (5) yields the solutions for the estimated values of $\tilde{p}_{lk}$ and $\tilde{w}_{tkh}$ (GOLAN and VOGEL 2000, pp. 458-459). The estimation procedure allowed for the possibility of non-zero covariances following the one-step GCE-SUR as described by GOLAN et al. (1996, p. 186).

The relative information content of the estimated parameters can be evaluated through the normalized entropy measure described in GOLAN et al. (1996, p.93). The measure is defined for values between zero and one, with values approaching zero in the case of perfect information (i.e. perfectly degenerated distribution) and values approaching one in the case of perfect uncertainty (i.e. uniform distribution). Additional entropy statistics used in the paper are the so-called: entropy-ratio and an analogous entropy Chi-square measure both described in GOLAN and VOGEL (2000, pp. 454-455). In an instrumental variable GCE (IV GCE) Markov approach, non-stationary effects can be determined by the following elasticity that determines the cumulative effects of a unit change in each covariate z_{tn} on y_{tk}, the vector of proportion falling in the *k*-th Markov state in time (t+1), as given by KARANTININIS (2002, p. 10):

$$\eta^{y}_{kn} = \frac{\partial y_{kt}}{\partial z_{tn}} \frac{\bar{z}_{tn}}{\bar{y}_k} = \frac{\bar{z}_n}{\bar{y}_k} \sum_l \left[\tilde{p}_{lk} \bar{x}_l^2 \left(\tilde{\lambda}_{nk} - \sum_k \tilde{p}_{lk} \tilde{\lambda}_{nk} \right) \right] \qquad (6)$$

Following the Markov formalism based on the Markov equilibrium distribution and absorbing states notions (JUDGE and SWANSON 1962, pp. 58-59), it is possible to compute several indicators such as the mean number of years it costs a farm being in a certain Markov state before absorption in a final state, as well as the probability that a non-absorbing Markov state will end up in a particular absorbing state. The projections of farm numbers were obtained following two steps. In the first step the Markov transition probability matrix was multiplied by itself *n* times in order to recover the transition probability matrix during *n* time periods. In the second step individual elements of the transition probability matrix were multiplied by the farm number in their respective size class in the base year used for projections.

4 DATA AND PRIOR INFORMATION

Aggregate data on the size distribution of dairy farms in Poland are used. Holdings were classified according to their herd size classes. The data cover the period from 1995 to 2006 and allow the recovery of the number dairy farms belonging to eight[6] farm size classes: 1 cow, 2 cows, 3-9 cows, 10-29 cows, 30-49 cows, 50-99 cows, 100-199 cows, > 200 cows (KRAWIECKA 2006). In order to account for exit and entry an additional size class was defined which contains the 'inactive farms' and 'potential entrants' ($l,k = 0$). Data were normalized by a common scalar equal to the maximum number of farms contained in the aggregate transition counts. In order to capture potential non-stationary effects on the Markov transition probabilities only a trend variable z_{t1} was introduced during estimation. Due to the limited

[6] Nine farm size classes considering the artificial entry and exit class size.

number of observations (i.e. number of transitions) the inclusion of other potentially relevant policy variable was not considered a feasible option.

The researcher may follow several principles in order to best approximate the farm size growth and to guess or estimate the probability of a farm to be in a given size class. In order to avoid data mining and ensure efficiency in estimation, the prior information should be derived from sources independent from the sample data as much as possible. In this study an extensive investigation of previous research was done and the lessons (general patterns) drawn from this formed the basis of the used prior information (see Table 1)[7]. The prior information on Markov transition probability estimates may concern three types of information: the probability of a farm to persist in the same farm size class (i.e. persistency), the probability a farm enters and/or exits the sector (i.e. entry/exit), and the probability to move to another farm size class (i.e. net shifts).

Persistency:

- Table 1 provides an overview of the estimated persistency's probabilities encountered in the literature, both for dairy studies and other studies. Although the studies found in the literature are not directly comparable (different countries, different sectors, different size class-width definitions used, and different time span) it appears that on average about 82.5 percent of dairy farms persist in the same size class from one period to another. When analyzing the aforementioned studies in further detail it also appears that persistency is generally lower for small farm size classes as compared to large farm size classes. Based on these findings in the literature, the priors on the diagonal transitional probabilities were set, moving from the top left corner to the low right corner of the transition probability matrix from 0.80 to 0.90 (i.e. $p_{lk} = 0.80 \quad l = k$ for $l,k = 2,3,4$ and $p_{lk} = 0.90 \quad l = k$ for $l,k = 5,\ldots,8$).

Table 1: Transition probability estimates: Literature overview

Authors	Year	Average Estimates	Smallest Class Estimates	Largest Class Estimates	Number of Classes	Transition
			Dairy Studies			
Padberg	1962	0.691	0.733	0.960	4	5 years
Hallberg	1969	0.879	0.768	0.961	5	annual
Keane	1991	0.756	0.360	0.945	7	6 years
Zepeda	1995	0.901	0.877	0.944	3	annual
Stokes	2006	0.898	0.805	0.999	6	annual
			Other Studies			
Judge and Swanson	1962	0.511	0.412	0.672	6	annual
Krenz	1964	0.862	0.804	1.000	6	5 years
Lee et al.	1965	0.650	0.473	0.572	4	annual
Ethridge et al.	1985	0.957	0.919	0.986	5	annual
Edwards et al.	1985	0.687	0.781	0.813	8	4 years
Garcia et al.	1987	0.836	0.930	0.929	11	annual
Disney et al.	1988	0.605	0.400	0.732	4	5 years
Karantininis	2002	0.531	0.386	0.768	18	annual

Note: Estimates may reflect different transition period lengths as indicated by the last column.
Source: Own calculations based on estimates from the literature.

Entry/Exit:

[7] A recent example neglecting this independence-requirement is STOKES (2006). For this reason the results he obtained are likely to over fit the sample data.

- As regards exit the literature shows two basic results. Small farms are more likely to exit than large farms (see also remark made before). Moreover, the smaller a farm, the higher the probability of exit is. Combining this with the already specified priors on persistence (which was set to 0.8 for small farms) the priors on the exit probabilities p_{10}, p_{20} and p_{30} were set to 0.20, 0.15 and 0.10 respectively.

- With respect to entry in all the studies observed the total number of businesses shows a clear tendency to decline over time. Generally very little information was known about entering firms, let alone about the probabilities of entrance in different size classes. Given this finding and the character of our data, which required us to focus on net-transitions (net entry), it was decided to specify no positive priors on any entry probabilities $(p_{0k} = 0, \forall k \neq 0)$. Since by definition $\sum_k p_{0k} = 1$ these priors on entry also imply that once a farm is out of business it will stay out of business (see previous remark about the Entry/Exit size class as an absorbing state and the prior estimate $p_{00} = 1$).

Net Shifts:

- As regarding the net shifts one pattern observed from the literature is that farms show a tendency to gradually develop. This implies that the probability a farm moves from its current size class to an adjacent size class is generally higher than the probability to move to more distant size classes. A second finding is that usually there is a switch-size class, below which farms show a tendency to decline and ultimately go out of business, whereas above this size class farms expand their business. This finding is likely to be partly related to the dominant family-business character of farming. As a consequence of this farm succession is tied to the family cycle (e.g. in case of no succession farmers getting older are likely to gradually downsize their business). Another explanatory factor might be that farms need a certain critical scale in order to be considered as 'viable', i.e. being able to finance expansion relying on generated internal savings and to the possibilities for attracting external credit (see SWINNEN and MATHIJS 1997, TONINI and JONGENEEL 2002). Reviewing previous studies it appeared that the location of the turning point size class is generally country and case specific (depending for example also on the specified number and width of size classes). Our prior estimate of the switch size class is therefore based on the particular sample considered and set equal to the size class with 3 to 9 cows (see also Figure 1). As regards the farms in this size class our prior is that they have a fifty-fifty probability to move up or down ($p_{32} = p_{34} = 0.05$, i.e. uninformative priors). Farms in larger size classes are assumed to move up to the adjacent size class with a probability of 0.10, whereas farms in lower size classes are assumed to move down to the next size class with the same probability (conditional on prior assumptions previously made about exit for the lower size classes). The prior assumptions made so far imply that most of the lower and upper off-diagonal elements of the transition probability matrix have prior expectations equal to zero (see DISNEY et al. (1988), ZEPEDA (1995) for a similar approach).

5 ESTIMATION RESULTS AND DISCUSSION

The IV GCE Markov model was estimated including a trend capturing for structural change. The normalized signal entropy $S(\tilde{\mathbf{p}})$ for the system was 0.663 whereas the normalized noise entropy $S(\tilde{\mathbf{w}})$ for the system was 0.971. The information index $I(\tilde{\mathbf{p}})$ or pseudo-R^2 for the signal was 0.337. The estimated $\chi^2_{\sim\,(K\text{-}1)}$ statistic was 0.416, indicating that the estimated transition probabilities did not statistically differ from the priors at five per cent significance level. A similar result was obtained computing the signal entropy ratio (i.e. only considering the signal distribution) which was equal to 2.324. At five percent significance level the hypothesis of normal errors could not be rejected relying on the Jarque-Bera test (VERBEEK 2004, p. 185). Statistical testing, at least for the signal part, was done under negative degrees of freedom given that Kx(K-1) independent[8] transition probabilities had to be estimated only having K total aggregate data of finite size categories for T transitions. However estimates were rather robust to changes in the prior magnitude[9].

Even though the power of statistical tests can be relatively low when there are negative degrees of freedom several stylized facts can be drawn from the above results. The computed statistics suggest that the data did not push the final estimates too much away from the prior, indicating either a relatively poor data signal or data-conforming prior estimate. This finding also is likely to be related to the negative number of degrees of freedom. Table 2 presents the estimated IV GCE Markov model (i.e. non-stationary model).

The estimated transition probability matrix itself already provides insight into the dynamic adjustment of dairy farms. For example, during the period considered there is a strong tendency for farms to persist in the same size class from one year to the next (see transition probabilities on the diagonal containing elements p_{kk}). The off-diagonal elements of the transition matrix provide information on the extent dairy farms are going to scale up or down. For example, from one period to another about 2 per cent of all farms with 10-29 cows will probably grow into a dairy farm with 30-49 cows. In Table 2 the cumulative effects of the trend z_{t1} on the number of dairy farms y_{tk} in terms of elasticity is presented in the last row. The trend impact found implies that over time there is a contraction in the farms with 1-9 cows and increase in the remaining farms. The trend also has a positive impact on number of farms in the inactive size class (Exit). Our results fits in with SZNAJDER (2002, p. 253) who shows that in order to have full return from the engaged capital, including rent of the land, Polish dairy farms need to have a herd of at least 10-15 dairy cows. This suggests that the minimum efficient size of dairy farms, minimizing the per unit costs, or the minimum locus on the long-run average costs level for farms is at a herd size of 10 cows or more.

Table 2: IV GCE-SUR Markov transition probabilities and non-stationary effects

Class	Exit	1	2	3-9	10-29	30-49	50-99	100-199	> 200	$S(\mathbf{p}_i)$
Entry	1.000									1.000
1	0.118	0.882								0.727
2	0.116	0.054	0.829							0.919
3-9	0.063		0.044	0.872	0.021					0.722
10-29					0.980	0.020				0.302
30-49						0.919	0.081			0.862
50-99							0.984	0.016		0.254
100-199								0.989	0.011	0.183
> 200									1.000	1.000
z_{t1}	0.011	-0.007	-0.002	-0.007	0.011	0.047	0.003	0.132	2.524	

Source: Own estimates.

Table 3 reports the estimated mean number of years in each transient state for each non-absorbing states (i.e. transient periods) as well as the probabilities of absorption for each non-absorbing states into the two absorbing states (i.e. absorption probabilities). These estimates

---

8 This is obtained subtracting from the KxK transition probability matrix the K row adding-up condition in equation (4).

9 For a given prior configuration we carried several estimations by only changing the prior magnitude by one digit each time. This did not bring remarkable changes on the final estimates. To save space results are not reported here but they are available upon request from the authors.

provide an additional indicator on the rate of change in the number of dairy farms by herd size class. Thus for a dairy farm with 10-29 dairy cows the mean number of years before absorption is about 50 years whereas for a dairy farms with 2 cows the mean number of years before absorption is about 6 years. This suggests a larger rate of change for the small dairy farms as compared to the medium and large dairy farms. From the last two columns of Table 3 it also appears that in equilibrium the majority of the dairy farms with 1 and 9 cows will leave the sector, whereas the dairy farms belonging to the remaining size states will continue in dairying. More precisely, only 16 per cent of the dairy farms with 3-9 cows will persist in the dairy sector, whereas 84 per cent are expected to leave the sector.

Table 3: Estimated transient periods and absorption probabilities

Class	1	2	3-9	10-29	30-49	50-99	100-199	0	> 200
1	8.447							1.000	0.000
2	2.689	5.865						1.000	0.000
3-9	0.919	2.005	7.825	8.182	2.030	10.164	15.240	0.836	0.164
10-29				49.980	12.402	62.087	93.091	0.001	0.999
30-49					12.403	62.089	93.094	0.001	0.999
50-99						62.089	93.094	0.001	0.999
100-199							93.098	0.001	0.999

Note: The last two columns of the table report the absorption probabilities.
Source: Own estimates.

Finally, the estimated Markov transition probability matrixes were used to make several projections of the number of dairy farms in the coming decade. In order to assess the predictive power of the estimated Markov models, projected values and actual values were first compared for the most recent available year (i.e. 2006). We compared two types of models: the IV GCE Markov model estimated with SUR, hereinafter called IV GCE-SUR (i.e. non-stationary model) and the similar model without the inclusion of the trend (i.e. stationary model). In addition for each type of model we compared the model with the priors as defined in Section 4 with a model estimated using uniform (i.e. non-informative) priors. In terms of projections the best performance was obtained for the IV GCE-SUR model with non uniform priors. In addition from our results it appears useful to impose some sort of prior information on the estimated Markov transition probabilities given the relatively low projection power of the models estimated with uniform priors.

Table 4: Dairy farm size distribution: projected versus actual numbers for 2006

1	2	3-9	10-29	30-49	50-99	100-199	> 200	Total
				IV GCE-SUR				
286690	124949	148573	68203	5591	1155	140	42	635343
2.47	*-5.37*	*1.15*	*5.99*	*-6.43*	*3.34*	*-7.19*	*21.05*	*0.74*
				IV GCE-SUR (Uniform Prior)				
183155	111209	120992	37372	4275	1184	253	69	458508
-34.54	*-15.77*	*-17.63*	*-41.92*	*-28.46*	*-15.88*	*51.34*	*82.05*	*-27.30*
				GCE-SUR				
292110	126837	153170	67985	5564	1146	127	41	646979
-4.40	*-3.94*	*4.28*	*5.65*	*-6.88*	*-18.63*	*-24.15*	*8.85*	*2.59*
				GCE-SUR (Uniform Prior)				
252441	154765	167159	22858	1779	1286	105	22	600415
-9.78	*17.21*	*13.80*	*-64.48*	*-70.23*	*-8.67*	*-37.21*	*-41.48*	*-4.79*
				Actual 2006				
279791	132037	146887	64350	5975	1408	167	38	630653

Note: Percentage deviations are reported in italics.
Source: Own estimates.

The estimated IV GCE-SUR model predicts reasonably well the total aggregate number of dairy farms, although the model has the tendency to overestimate the number of farms in most of the size classes an exception made for the farms with 2, 30-49 and 100-199 cows where the model underestimates the total number of farms. This is mainly attributable to the effect plaid by net shifts from one size class to the adjacent size class. Table 5 provides the projections associated with the IV GCE-SUR model. As can be seen it is predicted that in 2013 about 47 percent of the number of active dairy farms in 2007 will leave the sector (*ceteris paribus*).

Table 5: Projected dairy farm size distribution (IV GCE-SUR)

Year	1	2	3-9	10-29	30-49	50-99	100-199	> 200	Total
2007	253833	115943	128116	66135	6781	1867	188	40	572902
2008	230074	101772	111744	67492	7557	2384	216	42	521281
2009	208359	89303	97464	68480	8298	2955	252	44	475155
2010	188538	78335	85009	69149	8999	3576	297	47	433950
2011	170468	68693	74146	69544	9657	4244	351	50	397153
2012	154015	60221	64671	69703	10270	4955	415	54	364303
2013	139049	52779	56406	69662	10837	5703	489	58	334982
Average Growth Rates (%)									
	-10.3	-12.3	-13.8	0.3	9.5	22.8	17.9	4.4	-9.1

Source: Own estimates.

6 CONCLUSION

The projections showed that the number of dairy farms will continue to decline in the coming decade, although with an increase in the number of farms of medium and large size. The increase will be in farms with more than 30 cows. Therefore a consolidation process is expected, where small dairy farms (i.e. semi-subsistence farms) will continue to exit from the sector although their relative share on the total number of dairy farms will rather persist. The estimated mean number of years before the small subsistence dairy farms with 1-2 cows leave the dairy sector is approximately 7 years. In addition, only dairy farms with at least 10-29 cows and about 16 percent of the dairy farms with 3-9 cows are expected to survive at the Markov equilibrium.

Overall, these findings suggest that Poland will be characterized by a polarized dairy farm structure with on one side a persistent fringe of subsistence and semi-subsistence self-employed small dairy farms and on the other side a growing fringe of business oriented dairy farms. However, the predicted transition from a subsistence farming style to a more modern and specialized farm structure is also subject to a number of other influencing factors, not directly included in our model.

Although the Markov chain approach appears to be flexible to handle a wide scope of dynamic factors, the predicted evolution of the Polish dairy sector might be also affected by other factors which are not explicitly included or not sufficiently accounted for in the model. To mention some important ones:

- Most of the time, exiting is not an option for farmers in CEECs, simply because the industrial or service sectors are not able to absorb the redundant unskilled labourers, given the difficult economic environment (PETRICK and WEINGARTEN 2004, p. 6). According to the last Agricultural Census in 2002, about 1 million of individual farmers have failed to find a job, thus fuelling the so-called 'hidden unemployment'[10]. In addition, from 1 May 2006, Polish farmers have been entitled to receive direct payments following a simplified framework

[10] Note that in Poland, the owners and holders of farms with an area equal to or exceeding 2 hectares cannot be registered as unemployed (ZMIJA and TYRAN 2004, p. 75).

which allocates the premiums per hectare of land. Direct payments consist of a per hectare Single Area Payment System (SAPS) and supplemental eligible crop area payments. The eligibility criteria for the SAPS[11] require that farmers own over 1 hectare of arable land, provided that the arable 'plot' is no smaller than 0.1 hectare (USDA 2005, p. 5). The impact of this is ambiguous. On the hand it creates an incentive for small farms to stay in farming, whereas on the other hand facilitates faster expansion and modernization of medium and larger sized farms (relaxing liquidity constraints);

- As LYSON and WELSH (1992), DUPUIS (1993), and LYSON and GILLESPIE (1995) have also observed, the size structure of dairy farming is related to changes in the milk market. The entry of large-scale foreign investors with mass production dairy-processing facilities, for example, is usually accompanied by a decline in the number of small units unable to comply with the quality requirements imposed, and by an increase in the number of large-scale producers;

- The concentration of land in fewer but more efficient farms depends on the mediating role of a well-defined and functioning land market. When lacking, this not only hampers efficient land allocation, but also limits the access to capital (land credit, mortgage) and hence investments. The increased land price after the EU accession is also likely to affect land allocation towards large dairy farms;

- The recent access to the EU implies that the milk quota regime is imposed now on the Polish dairy sector. This constraint might affect the sectoral evolution. Although there are expectations that this will fix the sectoral structure, there are also signals that the impact might be limited or go the other way around. The value of the quota might also act as an exit-payment inducing some farmers to leave the sector even earlier than initially planned;

Since these factors were or could not be taken explicitly into account in the present analysis the actual evolution could be different from projected one, in particular for the subsistence sector.

ACKNOWLEDGEMENTS

We would like to thank Francisco J. André for the invitation and fruitful discussions at the seminar on Economic Analysis at the Department of Economics of Universidad Pablo de Olavide, Seville, Spain on May 7th, 2007. We would like also to acknowledge Dr. Lucyna Krawiecka, Head of Animal Production Section, Agriculture and Environment Division, Central Statistical Office, Warsaw, Poland for providing us the data.

REFERENCES

AGRAEUROPE (2000): Milk production to post further decline in 2000, *AgraFood East Europe, Vol. April*, pp. 18-19.

COURCHANE, M., GOLAN, A., NICKERSON, D. (2000): Estimation and evaluation of loan discrimination: An informational approach, *Journal of Housing Research, Vol. 11*, pp. 67-90.

---

[11] In the Malopolska region, where the average farm size is about 2.10 hectares and about 45.5 per cent of the farming population receive income from pension schemes, disability benefits and other social security, the distribution of direct payments may act as an additional social support, keeping subsistence and semi-subsistence farming in business.

DISNEY, W.T., DUFFY, P.A., HARDY JR, W.E. (1988): A Markov chain analysis of pork farm size distributions in the South, *Southern Journal of Agricultural Economics, Vol. 20*, pp. 57-64.

DUPUIS, M. (1993): Subnational state institutions and the organization of agricultural resource use: The case of the dairy industry, *Rural Sociology, Vol. 58*, pp. 440-460.

EDWARDS, C., SMITH, M.G, PETERSON, R.N. (1985): The changing distribution of farms by size: A Markov analysis, *Agricultural Economics Research, Vol. 37*, pp. 1-16.

ETHRIDGE, D.E., ROY, S.K, MYERS, D.W. (1985): A Markov chain analysis of structural changes in the Texas high plans cotton ginning industry, *Southern Journal of Agricultural Economics, Vol. 17*, pp. 11-20.

EUROPEAN COMMISSION (1998): Agricultural Situation and Prospects in the Central European Countries-Poland, *Working Document*, Brussels, Directorate General for Agriculture (DG VI), <http://ec.europa.eu/agriculture/publi/peco/poland/pol_en.pdf>.

FAOSTAT (2006): FAOSTAT Agriculture, <http://faostat.fao.org/faostat/collections?version=ext&hasbulk=0&subset=agriculture>.

FOMBY, T.B., CARTER HILL, R. (1997): Applying Maximum Entropy to Econometric Problems, Advances in Econometrics, London, JAI Press.

GARCIA, P., OFFUTT, S.E., SONKA, S.T. (1987): Size distribution and growth in a sample of Illinois cash grain farms, *American Journal of Agricultural Economics, Vol. 69*, pp. 471-476.

GODDARD, E., WEERSINK, A., CHEN, K., TURVEY, C.G. (1993): Economics of structural change in agriculture, *Canadian Journal of Agricultural Economics, Vol. 41*, pp. 475-489.

GOLAN, A., JUDGE, G., MILLER, D. (1996): Maximum Entropy Econometrics: Robust Estimation with Limited Data, Chichester [etc.], John Wiley & Sons.

GOLAN, A., VOGEL, S.J. (2000): Estimation of non-stationary social accounting matrix coefficients with supply-side information, *Economic Systems Research, Vol. 12*, pp. 447-471.

GOOD, I.J. (1963): Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables, *Annals of Mathematical Statistics, Vol. 34*, pp. 911-934.

HALLBERG, M.C. (1969): Projecting the size distribution of agricultural firms – An application of a Markov process with non-stationary transition probabilities, *American Journal of Agricultural Economics, Vol. 51*, pp. 289-302.

JONGENEEL, R., LONGWORTH, N., HUETTEL, S. (2005): Dairy farm size distribution in East and West: Evolution and sensitivity to structural and policy variables, paper presented at the 11th EAAE Congress, "The Future of Rural Europe in the Global Agri-Food System" held in Copenhagen, Denmark, 24.-27.08.05.

LEE, T.C., JUDGE, G.G., TAKAYAMA, T. (1965): On estimating the transition probabilities of a Markov process, *Journal of Farm Economics, Vol. 47*, pp. 742-762.

LEE, T.C., JUDGE, G.G., ZELLNER, A. (1970): Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, Amsterdam, North-Holland Publishing Company.

LYSON, T.A., GILLESPIE, G.W. (1995): Producing more milk on fewer farms: Neoclassical and neostructural explanations of changes in dairy farming, *Rural Sociology, Vol. 60*, pp. 493-504.

LYSON, T.A., WELSH, R. (1992): Toward a second agricultural devide: The restructuring of American agriculture, *Sociologia Ruralis, Vol. 32*, pp. 248-263.

MITTELHAMMER, R.C., JUDGE, G.G., DOUGLAS, J.M. (2000): Econometric Foundations, Cambridge, Cambridge University Press.

JUDGE, G. G., SWANSON, E. R. (1962): Markov chains: Basic concepts and suggested uses in agricultural economics, *Australian Journal of Agricultural Economics, Vol. 6*, pp. 49-61.

KARANTININS, K. (2002): Information-based estimators for the non-stationary transition probability matrix: An application to the Danish pork industry, *Journal of Econometrics, Vol. 107,* pp. 275-290.

KEANE, M. (1991): Changes in the size structure of Irish dairy farms, *Irish Journal of Agricultural Economics and Rural Sociology, Vol. 14*, pp. 67-74.

KRAWIECKA, L. (2006): Personal Communication on Data on Private Dairy Farms in Poland, Warsaw, National Statistical Office.

KRENZ, R.D. (1964): Projection of farm numbers for North Dakota with Markov chains, *Agricultural Economics Research, Vol. 16*, pp. 77-83.

KULLBACK, J. (1959): Information Theory and Statistics, New York, John Wiley.

PADBERG, D.I. (1962): The use of Markov processes in measuring changes in market structure, *Journal of Farm Economics, Vol. 44*, pp. 189-199.

PETRICK, M., WEINGARTEN, P. (2004): The role of agriculture in Central and Eastern European rural development: An overview, in: PETRICK, M., WEINGARTEN, P. (eds.): The Role of Agriculture in Central and Eastern European Rural Development: Engine of Change or Social Buffer, *Studies on the Agricultural and Food Sector in Central and Eastern Europe, Vol. 25,* Kiel, Vauk Wissenschaftsverlag, pp. 1-21.

PUKELSHEIM, F. (1994): The three sigma rule, *The American Statistician, Vol. 48,* pp. 88-91.

STOKES, J.R. (2006): Entry, exit, and structural change in Pennsylvania's dairy sector, *Agricultural and Resource Economics Review, Vol. 35*, pp. 357-373.

SWINNEN, J.F.M., MATHIJS, E. (1997): Agricultural privatization, land reform and farm restructuring in Central and Eastern Europe: A comparative analysis, in: SWINNEN, J.F.M., BUCKWELL, A., and MATHIS, E. (eds.): Agricultural Privatization, Land Reform and Farm Restructuring in Central and Eastern Europe, Aldershot, Ashgate, pp. 161-188.

SZNAJDER, M. (2002): I cambiamenti economici e strutturali delle aziende da latte in Polonia, in: PIERI, R., RAMA, D. (eds.): La Struttura della Zootecnia da Latte in Italia e in Europa, *Studi di Economia Agro-alimentare, Vol. 15,* Milano, Franco Angeli, pp. 237-261.

TONINI, A., JONGENEEL, R. (2002): Dairy farm size restructuring in Poland and Hungary, in: HINNERS-TOBRÄGEL, L., HEINRICH, J. (eds.): Agricultural Enterprises in Transition: Parallels and Divergences in Eastern Germany, Poland and Hungary, *Studies on the Agricultural and Food Sector in Central and Eastern Europe, Vol. 15,* Kiel, Wissenschaftsverlag, pp. 317-339.

USDA (2005): Poland, Trade Policy Monitoring, Poland's CAP Implementation 2005, *United States Department of Agriculture Foreign Agricultural Service*, Washington D.C., USDA.

VERBEEK, M. (2004): A Guide to Modern Econometrics, Chichester, John Wiley.

ZEPEDA, L. (1995a): Asymmetry and nonstationarity in the farm size distribution of Wisconsin milk producers: An aggregate analysis, *American Journal of Agricultural Economics, Vol. 77*, pp. 837-852.

ZEPEDA, L. (1995b): Technical change and the structure of production: A non-stationary Markov analysis, *European Review of Agricultural Economics, Vol. 22,* pp. 41-60.

ZMIJA, J., TYRAN, E. (2004): Agriculture in Southeastern Poland - Main problems of the systemic transformation process, in: PETRICK, M., WEINGARTEN, P. (eds.): The Role of Agriculture in Central and Eastern European Rural Development: Engine of Change or Social Buffer, *Studies on the Agricultural and Food Sector in Central and Eastern Europe, Vol. 25,* Kiel, Vauk Wissenschaftsverlag, pp. 73-82.