Working Papers SMART - LERECO

Trends in the French commercial farm population

Madior FALL, Laurent PIET, Muriel ROGER

Working Paper SMART ï LERECO N°10-04

June 2010

UMR INRA-Agrocampus Ouest **SMART** (Structures et Marchés Agricoles, Ressources et Territoires)
UR INRA **LERECO** (Laboratoires dÉtudes et de Recherches Economiques)

Trends in the French commercial farm population

Madior FALL

INRA, UR1043, Laboratoire d'Economie Appliquée, F-75000, Paris, France.

Laurent PIET

INRA, UMR1302 SMART, F-35000 Rennes, France
Agrocampus Ouest, UMR1302 SMART, F-35000 Rennes, France

Muriel ROGER

INRA, UR1043, Laboratoire d'Economie Appliquée, F-75000, Paris, France.

Acknowledgements

We thank Karine Daniel, Nejla Ben Arfa for helpful comments. Usual disclaimers apply.

Auteur pour la correspondance / Corresponding author

Laurent PIET
INRA, UMR SMART
4 allée Adolphe Bobierre, CS 61103
35011 Rennes cedex, France
Email: Laurent.Piet@rennes.inra.fr
Téléphone / Phone: +33 (0)2 23 48 53 83
Fax: +33 (0)2 23 48 53 80

Trends in the French commercial farm population

Abstract

Knowledge and projection of farm numbers and the structure of their population is an important issue for agricultural economists and policy makers. Although Markov chain models have enjoyed decades of popularity in forecasting total farm numbers, they generally fail to provide a detailed insight of the farm population's structure; to overcome this caveat we estimate a parametric distribution of the utilized agricultural area of French commercial farms. Our method provides detailed information on the structure of the population and accounts for the specificity of off-land farming. We also model the influence of variables such as the farm's legal status, type of farming and farm holder's age. The estimation leads to a relevant description of the entire population of professional farm. When compared with the 2005 Farm Structure Survey data, our simulations based on FADN data display a close match across a number of key variables.

Keywords: farm structures, farm size distribution, maximum likelihood, simulation

JEL classifications: Q12, C13, C15

Lôévolution de la population des exploitations agricoles professionnelles françaises

Résumé

Il est important, pour lôéconomiste agricole et le décideur public, de connaître et de prévoir le nombre des exploitations agricoles et la structure de leur population. Très populaire depuis plusieurs décennies, lôutilisation des chaînes de Markov permet de projeter à un horizon donné lôeffectif total dôune population mais nôoffre pas, en général, une analyse fine de la structure de celle-ci ; pour pallier cette limite, nous nous fondons sur une estimation paramétrique de la distribution des surfaces agricoles utiles des exploitations professionnelles françaises. Cette méthode permet une analyse fine de celle-ci et tient compte de la spécificité des exploitations hors-sol. Elle permet également de caractériser lôinfluence de variables telles que le statut juridique de lôexploitation, lôorientation productive ou encore lôâge du chef sur cette distribution. Nous obtenons ainsi une représentation pertinente de lôensemble des exploitations. Confrontées aux données de lôenquête « Structures » 2005, les simulations réalisées à partir des données du RICA montrent une très bonne adéquation de nos résultats pour plusieurs variables dôintérêt.

Mots-clefs : structures agricoles, taille des exploitations agricoles, maximum de vraisemblance, simulation

Classifications JEL : Q12, C13, C15

Trends in the French commercial farm population

1. Introduction

As shown by Chavas (2001), the organisation of agricultural production and the way in which it has developed are central issues for analysis in agricultural economics. In fact, the same level of production and satisfaction of food needs can be provided by any number and configuration of farming structures: from a large number of very small farms, as is often the case in developing countries, to a small number of very large farms, as in certain Eastern European, South American and Australasian countries, and the more common combination of the two. It is largely this organisation that defines what is generally called a country or regionôs ñagricultural modelò.

Yet each type of organisation has its own repercussions in terms of family/paid labour breakdown, land use and rural vitality, landscape structure and the environment. It also affects food availability, variety, quality and prices. This is why policymakers are generally keen to know the impact of the regulatory instruments they put in place or reform on the number and structure of farms,[1] and to even set goals in this area.[2] Some policies are therefore designed to explicitly encourage a certain type of ñmodelò such as the structural policy introduced in France in the 1960s. In the European Union, in particular, the recent Common Agricultural Policy shift towards increasingly direct farm income support instruments, i.e. decoupled from market variables (prices and quantities), produce and factors of production, has prompted questions about the distribution of the aid at individual level and the equity of this distribution.

On the basis of early work by Judge and Swanson (1961), Krenz (1964) and Hallberg (1969), it has become ñtraditionalò in agricultural economics literature to address this question of change in the number and structure of farms using Markov chain theory models. In this approach, the population studied is first divided into a small number of categories, generally

---

[1] This paper does not discuss the substance of public agricultural policies. It merely notes their existence.

[2] For example, the policy paper published by the French Ministry for Agriculture and Fisheriesô High Council for the Co-ordination and Steering of the Agricultural and Food Economy (CSO) states, ñThe existence of a dense network of farms, food industries and local craft and food trade firms is part and parcel of the European identity. It calls for farming and the many farmers to be upheld as the economic bedrock of a huge number of territoriesò (MAP, 2007, p. 4, our translation).

less than a dozen,[3] based on a size criterion (utilised agricultural area, the number of livestock or a measurement of the economic size of the farms). The probabilities of moving from one size class to another over the time period considered is then deduced from observing the distribution of the farms across the different categories at two different dates. Once allowance has been made for the "entries" and "exits" categories, these transition probabilities can be used to project the total number of farms and their distribution across the different categories at a later time step. Recent extensions of these models are non-stationary, i.e. transition probabilities can change over time (Zepeda, 1995a and b; Karantininis, 2002; Ben Arfa *et al.*, 2006; Stokes, 2006).

The Markov chain approach is interesting in that it can be used to forecast the total number of farms at a given time in the future. Yet we believe that it suffers from three major shortcomings. Firstly, although it can be used to simulate both the total number of farms and their breakdown into each of the size classes considered, it cannot furnish a detailed picture of the population's structure. Given that the distribution of farms in each category is unknown, since their numbers are often very small, as already mentioned, and the intervals they define rarely correspond to precise statistical scales such as deciles, it is strictly speaking impossible to calculate an indicator as simple as average farm size at the simulation date. Secondly, as we have seen, the categories are defined on the basis of a single criterion, which rules out any differentiation of the projected population's structure by other key variables such as individual or incorporated farm status and farm holder's age. Lastly, due precisely to these categories being defined by the choice of a single size criterion, which has to be relevant for the entire population studied, the studies generally only look at one type of farm (cereal crops *or* dairy *or* pork producers, etc.) and cannot paint a picture of the entire agricultural sector.

The approach we propose here responds in part to these criticisms, its main limitation at present being that it cannot project the total number of farms. However, if this number is given, it can forecast the population's structure in detail based on different key criteria and cover all the commercial farms working in all types of farming. As we will see, the estimation method is simple and applies to usual and easily available data. It is based on the specification, maximum likelihood estimation and projection of the distributions of utilised

3 Butault and Delame (2005) are a notable exception to this. Using panel data rather than aggregate data like most of the other Markov chain studies, these authors consider a large number of different categories defined by the farm's physical size in hectares, economic size, region of establishment, the farm holder's age, individual or incorporated status, and type of farming.

agricultural area (UAA) as an indicator of physical farm size. Yet the method could easily be applied to other size criteria such as standard gross margin (SGM) as an indicator of economic size.

The data used are taken from the French strand of the Farm Accountancy Data Network (FADN) for the years in common with the Farm Structure Survey (FSS) for the 1990-2000 period. For each of the resulting five dates, we have all the individual data we need (repeated cross-sections) to be able to work simultaneously on the individual and time elements of the information. The characteristics chosen to study surface area trends are respectively type of farming (TF), legal status and farm holder's age. The list is not exhaustive and could be extended to other variables obtainable from the data, such as the region of the farm's registered office.

Following a brief overview of the data in the second section, the third section presents the econometric model and its estimation. The fourth section sets out to validate our results by comparing the projections made using the proposed method with the data observed in the 2004 FADN and the 2005 FSS. These comparisons find that the projected data closely match the observed data. The last section concludes with a discussion of the presented method's limitations and suggested ways to improve and develop it.

2. Data used

2.1. The FADN

The Farm Accountancy Data Network (FADN) is a European survey for evaluating the income of agricultural holdings. It provides, on an annual basis, accountancy and technico-economic data for a sample of agricultural holdings in the European Union. The survey covers only farms which, due to their size, are considered as commercial ones.[4] We are interested solely in the survey's French strand. Yet the fact that the data are harmonised across all the countries means that the method could easily be extended to the European Union as a whole.

The FADN sample is stratified using three criteria: region, type of farming (TF) and economic size (ES). Within each stratum, a set of individuals is drawn pseudo-randomly from all the corresponding farms. Each of the sample's farms is then assigned an extrapolation coefficient

[4] For more information: http://ec.europa.eu/agriculture/index.cfm

based on its representativeness within the stratum, in keeping with what is known as the ñcalibrationò method (Deville *et al.*, 1993).

Calibration entails knowing the total number of farms present (N) every year. Yet this total number is only available for certain years: whereas it is an accurate figure when drawn from the French Farm Censuses (the most recent being in 1988 and 2000), it is determined on the basis of the Farm Structure Surveys for a certain number of years between two censuses (the most recent being 1990, 1993, 1995, 1997, 2003 and 2005). The weighting allocated to each of the FADN individuals is therefore calculated such that the total number of farms extrapolated from the entire sample is consistent with the findings of the above mentioned surveys for the corresponding years. However, the coefficients are not ñreallyò updated for the years not covered by one of these two surveys: they are determined such that the total extrapolated number of farms does not differ too much from the number found by the most recent survey, all the while checking that certain aggregate economic variables (e.g. total output value) are consistent with the data in the National Agricultural Accounts for the year in question.[5] We therefore do not include these ñintermediateò years in our analysis and use only the data from 1990, 1993, 1995, 1997 and 2000.[6]

Moreover, note that the FADN observations are not panel data since farms can ñenterò and ñexitò the FADN sample every year for reasons that are not purely demographic. Here, we are interested in all the farms present in each of the five separate years considered, not just the constant sub-sample (often called ñcylinderò) of those present on all five dates.

---

[5] In other words, the total number of commercial farms determined from the FADN changes ñin plateauxò rather than continuously, with ñjumpsò in the years corresponding to the FSS. For example, we observe that $N^{FADN}_{1990} = 521{,}644$, $N^{FADN}_{1991} = 526{,}123$, $N^{FADN}_{1992} = 526{,}521$ and that $N^{FADN}_{1993} = 452{,}125$, $N^{FADN}_{1994} = 461{,}241$. Note also that the calibration can sometimes give rise to a (slight) increase in the total number of commercial farms found by the FADN between two Farm Structure Surveys, whereas the general consensus is of a steady downward trend in the number of farms.

[6] For technical reasons, the FADN coefficients were not updated in 2003 to reflect the 2003 FSS. Hence we have not used this particular year in our analysis.

2.2. The Farm Structure Surveys

The Farm Structure Survey (FSS) is conducted regularly by the French Ministry for Agricultureös Statistics and Forecasting Service (SSP).[7] The FSS system tracks a panel of farms drawn from the population identified by the most recent farm census. The sampleös longevity is guaranteed by detailed affiliation rules defining the farms to be interviewed in the case of an event such as closure, spin-off or merger. Since 1962, interviews have been conducted face to face on the farm itself. The data concerning crop year t/t+1 are collected at the end of year t+1. The survey provides information on the structure of the farms, the farmös manpower and any changes to this manpower, and the agricultural factors of production (surface areas, livestock and certain machinery).

Aside from the fact that we base our analysis solely on the years when the FADN sampleös weighting coefficients are determined on the basis of either the Farm Censuses (1988 and 2000) or the Farm Structure Surveys (1990, 1993, 1995 and 1997),[8] we do not otherwise explicitly use the data from these surveys for our projections. The 2005 FSS is nonetheless used in the third part to test the validity of our forecasting method.

3. Projection model

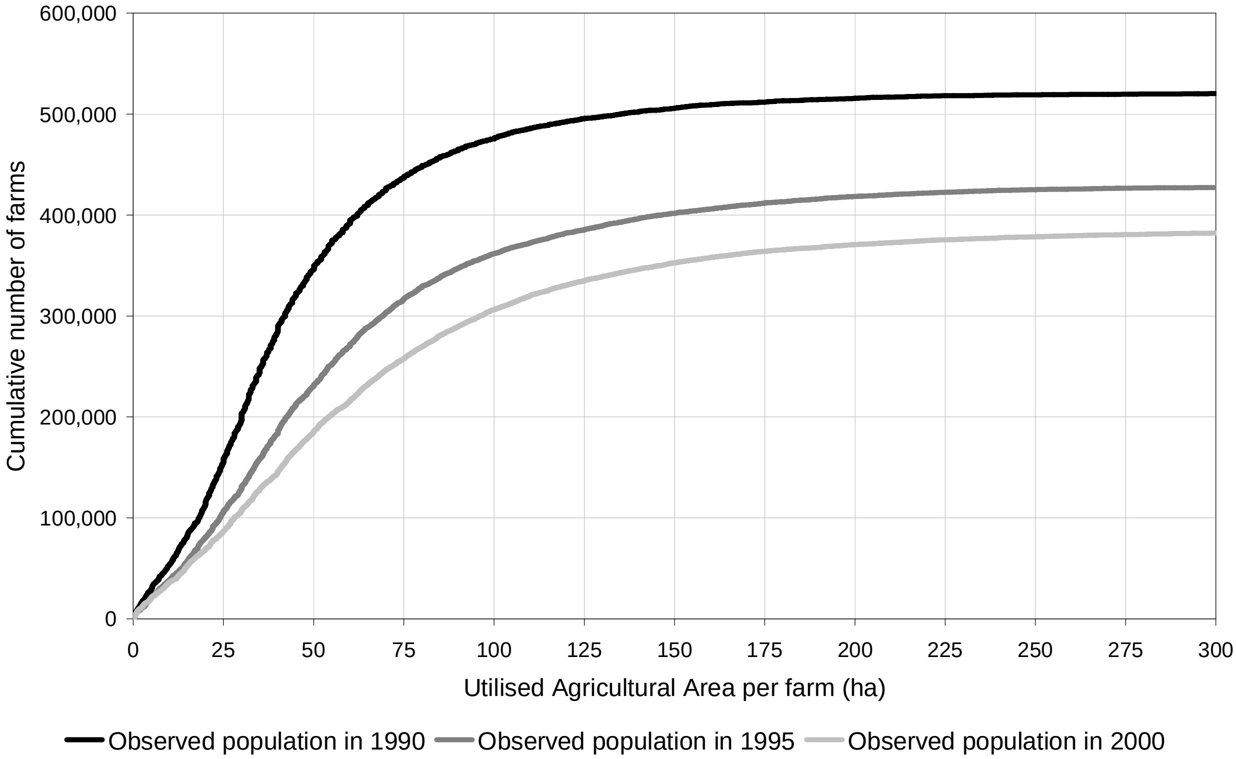
3.1. Econometric model

For each year, we have first constructed the cumulative distribution of farms based on their utilised agricultural area (UAA), obviously taking into account the weighting coefficients affecting each of them. We obtain, for example, the curves presented in Figure 1.

[7] In the summer of 2008, the SSP took over all the tasks previously assigned to the Central Service for Statistical Surveys and Studies (SCEES).

[8] We do not use the 2003 FSS data, see Note 6.

Figure 1: Cumulative distribution of commercial farms by Utilised Agricultural Area (UAA)



Source: Farm Accounting Data Network (FADN) 1990, 1995 & 2000

In addition to the fact that the total number of commercial farms decreases, the curves veer off to the right indicating an increase in their average surface areas. As Fansten (1969) has already observed, all these curves are similar to log-normal cumulative distribution functions. We therefore assume that the distribution S^t of surface areas on date t follows a log-normal distribution whose parameters μ^t and σ^t are defined by:

$$\mu^t = \exp(X^t \alpha)$$
$$\sigma^t = \exp(Z^t \beta)$$

where X^t and Z^t are the matrices of explanatory variables at each corresponding date, with the exponential form being included to guarantee that μ^t and σ^t are positive, and α and β are the coefficients.

We need to take account of the farms in the data whose UAA is zero. We hence define the probability $1/\left(1+p^t\right)$ of having a zero surface area (the probability of having a strictly positive surface area being $p^t/\left(1+p^t\right)$) using a logistic regression where :

$$p^t = \exp\left(W^t\delta\right)$$

where W^t is the matrix of explanatory variables at each date t.[9]

The likelihood is specified as the weighted joint distribution of the two endogenous variables p^t and UAA^t and is maximized in one step. The optimization model is thus given by:

$$\underset{p,\sigma,\mu}{Max} \quad \sum_t \sum_i w_i^t . \ln\left(s_i^t\right)$$

$$\text{where} \quad s_i^t = \begin{cases} 1/\left(1+p_i^t\right) & \text{if} \quad UAA_i^t = 0 \\ \dfrac{p_i^t}{\left(1+p_i^t\right)} \times \dfrac{1}{UAA_i^t\sqrt{2\pi}\sigma_i^t} \times \exp^{-\frac{\left(\ln\left(UAA_i^t\right)-\mu_i^t\right)^2}{2\sigma_i^{t\,2}}} & \text{if} \quad UAA_i^t > 0 \end{cases} \qquad (1)$$

where w_i^t represents the extrapolation coefficient (weight) for individual i in the FADN sample for year t and where parameters μ, σ and p have to be replaced by their expressions.

3.2. Results of the estimations

In practice, the characteristics chosen for the explanatory variables X^t and Z^t are the same, *i.e.*, a constant, the log of time $\ln(t)$, type of farming (denoted $\mathbf{TOF}_i$), legal status (denoted

[9] Here too, we use the exponential form to guarantee that p_i^t is positive.

$\mathbf{LST}_i$), and the farm holder's age (denoted $\mathbf{AGE}_i$); these are all qualitative variables whose categories are given in Table 1.

Table 1: Explanatory variable categories for parameters μ^t and σ^t

Type of farming (TOF)	
1	Field crops (TF 13+14+60 in the FADN classification)
2	Market garden cropping and horticulture (TF 28+29)
3	Wine (TF 37+38)
4	Fruit and other permanent crops (TF 39)
5	Dairying and cattle dairying, rearing and fattening combined (TF 41+43)
6	Cattle rearing and fattening (TF 42)
7	Sheep, goats and other grazing livestock (TF 44)
8	Pigs/poultry (TF 50+72)
9	Other crops and livestock combined (TF 71+81+82+90)

Legal status (LST)	
1	Individual farms
2	Other (incorporated)

Age of farm holder (AGE)	
1	≤ 35 years
2	> 35 years and ≤ 50 years
3	> 50 years and ≤ 65 years
4	> 65 years

We therefore have:

$$
\begin{aligned}
\mu_i^t &= \exp\left(\alpha_0 + \alpha_1 .\ln(t) + \boldsymbol{\gamma}_2.\mathbf{TOF}_i + \boldsymbol{\gamma}_3.\mathbf{LST}_i + \boldsymbol{\gamma}_4.\mathbf{AGE}_i\right) \\
\sigma_i^t &= \exp\left(\beta_0 + \beta_1 .\ln(t) + \boldsymbol{\delta}_2.\mathbf{TOF}_i + \boldsymbol{\delta}_3.\mathbf{LST}_i + \boldsymbol{\delta}_4.\mathbf{AGE}_i\right)
\end{aligned}
\qquad (2)
$$

where α_i and β_i are the coefficients ($\alpha_0, \alpha_1, \beta_0$ and β_1 are scalars and the others vectors). The reference categories correspond to individual farms specialised in "field crops" whose farm holder is aged 35 to 50 years ($TOF = 1$, $LST = 1$ and $AGE = 2$). Farms with a UAA of zero are mainly off-land farms.

The specification used for the logistic regression is given by:

$$p_i^t = \exp\left(\delta_0 + \delta_1.\ln(t) + \delta_2.PPF_i + \delta_3.TOF_i^8\right) \quad (3)$$

where δ_i are the parameters to be estimated, PPF_i is a continuous variable calculated as the ratio of purchased concentrated and coarse pig and poultry feed to gross farm income, and TOF_i^8 is a dichotomous variable indicating whether farm i belongs to the "pigs/poultry" type of farming or not. With these last two variables, we believe we adequately capture the farm's greater or lesser specialisation in off-land production and therefore its probability of having a zero surface area.

The results obtained are presented in Table 2. The estimation is highly satisfactory. Only a few parameters are not significant at the 1% level: these are, firstly, δ_1 and β_4^1 (which are, however, significant at the 10% and 5% levels respectively) and, secondly, β_2^4, β_3^2, β_4^3, β_4^4 and α_2^6 (which are not significant, even at the 10% level). Moreover, the signs of the parameters are as we expected, in that:

- The higher the level of pig and poultry feed purchases to gross farm income, the closer p_i^t is to zero (since δ_2 is negative), and therefore the closer the probability of having a zero surface area, $1/(1+p_i^t)$, is to 1; likewise if the farm belongs to the "pigs/poultry" specialisation;
- The average farm surface area, expressed by $\exp^{\mu_i^t + \sigma_i^{t^2}/2}$, tends to rise over time since both α_1 and β_1 are strictly positive;
- The hierarchy of types of farming in terms of average surface area is respected: for example, with a highly negative α_2 ($\alpha_2^2 = -1.67$), even though β_2^2 is positive ($\beta_2^2 = 0.59$), the farms specialised in market garden cropping have a much lower average surface area than the others;
- The incorporated farms generally have a larger surface area (α_3^2 and β_3^2 are both positive);
- Farm surface area initially grows with age and then decreases from 50 years old onwards. Farmers aged 35 to 50 years have the largest surface areas on average whereas those

aged over 65 years farm the smallest surface areas. This result can be an age effect but could also reflect a cohort one since the average size of farms in the 1950s was smaller than nowadays (Desriers, 2007).

Table 2: Results of the estimation of the model

Estimation of p_i^t		**Estimation of** μ_i^t		**Estimation of** σ_i^t	
δ_0	7.8554 (0.4618)***	α_0	1.3666 *(0.0041)****	β_0	-0.3235 *(0.0206)****
δ_1	-0.3823 (0.2043)*	α_1	0.0226 *(0.0016)****	β_1	0.0260 *(0.0087)****
δ_2	-3.2382 (1.2323)***	α_2^2	-1.6699 *(0.0786)****	β_2^2	0.5915 *(0.0303)****
δ_3	-2.6246 (0.6284)***	α_2^3	-0.4760 *(0.0078)****	β_2^3	0.1618 *(0.0203)****
		α_2^4	-0.3967 *(0.0125)****	β_2^4	-0.0064 *(0.0320)*
		α_2^5	-0.0699 *(0.0036)****	β_2^5	-0.4639 *(0.0255)****
		α_2^6	-0.0052 *(0.0049)*	β_2^6	-0.2629 *(0.0469)****
		α_2^7	-0.0399 *(0.0065)****	β_2^7	-0.1324 *(0.0331)****
		α_2^8	-0.2465 *(0.0115)****	β_2^8	0.3295 *(0.0480)****
		α_2^9	-0.0262 *(0.0042)****	β_2^9	-0.2314 *(0.0215)****
		α_3^2	0.1597 *(0.0033)****	β_3^2	0.0196 *(0.0332)*
		α_4^1	-0.0173 *(0.0031)****	β_4^1	-0.0454 *(0.0182)***
		α_4^3	-0.0412 *(0.0036)****	β_4^3	0.0150 *(0.0206)*
		α_4^4	-0.0631 *(0.0181)****	β_4^4	0.0235 *(0.0700)*

: significant at the 10% level; **: significant at the 5% level; *: significant at the 1% level*

3.3. Projection for horizon "t"

The method for projecting the population from an observed date t^0 to another date t consists in computing the extrapolation coefficient w_i^t that each individual i observed in t^0 should have if it were included in the t FADN sample. To do so, we alter the observed extrapolation coefficient $w_i^{t^0}$, which affects individual i in the t^0 sample, according to the econometric model.

Once Model (1) has been estimated, equations (2) and (3) allow to calculate μ_i^t, σ_i^t and p_i^t at any time t for each individual i observed in t^0, holding other variables constant. Then, we can calculate the probability:

$$s_i^t = \begin{cases} 1/\left(1+p_i^t\right) & \text{if} \quad UAA_i^t = 0 \\ \dfrac{p_i^t}{\left(1+p_i^t\right)} \times \dfrac{1}{UAA_i^t\sqrt{2\pi}\sigma_i^t} \times \exp^{-\frac{\left(\ln\left(UAA_i^t\right)-\mu_i^t\right)^2}{2{\sigma_i^t}^2}} & \text{if} \quad UAA_i^t > 0 \end{cases} \tag{4}$$

where the surface area of individual i is also kept constant, *i.e.*, $UAA_i^t = UAA_i^{t^0}$.

In reality, s_i^t is the probability of observing individual i in the sub-sample of individuals sharing the same characteristics, *i.e.*, individuals with the same characteristics as the model's explanatory variables. Denoting this sub-sample using index k, formula (4) gives us the probability $s_{i\in k}^t$. Following the sampling method used in FADN, this probability may also be expressed as:[10]

$$s_{i\in k}^t \cong \frac{w_i^t}{N_k^t}$$

[10] The transition from a law of continuous distribution (log normal) to discontinuous sampling along with the very nature of the FADN's stratified sampling mean that we do not rigorously have $s_{i\in k}^t \equiv w_i^t / N_k^t$.

where w_i^t is the extrapolation coefficient of individual i in the t sample and $N_k^t = \sum_{i \in k} w_i^t$ is the total number of individuals in sub-sample k.[11] With this approximation, which holds regardless of t, we easily find that:

$$\frac{s_{i\in k}^t}{s_{i\in k}^{t^0}} = \frac{w_i^t}{w_i^{t^0}} \frac{N_k^{t^0}}{N_k^t}$$

which gives, after rearrangement:

$$w_i^t = w_i^{t^0} \frac{s_{i\in k}^t}{s_{i\in k}^{t^0}} \frac{N_k^t}{N_k^{t^0}} \quad (5)$$

Among the five terms in the right-hand side of equation (5), only N_k^t is still missing; if it is known somehow, the extrapolation coefficient w_i^t we were looking for can be easily computed.

4. Simulations and validity of the projections

4.1. Simulation for 2004

We measure our model's predictive capacity by simulating the distribution of commercial farms in 2004 and comparing it with the distribution actually observed in the FADN for the same year.[12] This simulation is conducted with different starting years in order to study this parameter's influence on the quality of the projection.

We therefore apply the formula in equation (5) as follows:

$$w_i^{2004} = w_i^{t^0} \frac{s_{i\in k}^{2004}}{s_{i\in k}^{t^0}} \frac{N_k^{2004}}{N_k^{t^0}}$$

---

[11] We obviously make the assumption that no empty sub-sample is defined, *i.e.*, that $N_k^t > 0$ whatever k.

[12] The idea here is not to "forecast" the numbers for 2004, since we use those given by the FADN (despite the fact that the extrapolation coefficients were not updated in 2003-2004 to bring them in line with the FSS 2003, see Note 10), but to compare the simulated and observed distributions.

where the $s_{i \in k}^{2004}$ are estimated from the econometric model as described above, the numbers N_k^{2004} are those observed in the FADN 2004 and where t^0 successively takes the values 1990, 1993, 1995, 1997, 2000 and 2003.

The analysis then consists in comparing the distribution of farms derived from the estimated w_i^{2004} with the distribution actually observed in the FADN 2004 data. At this stage, bear in mind that 2004 (like 2003) was not used to estimate the econometric model: the analysis conducted is therefore a pertinent test of the model's predictive power.

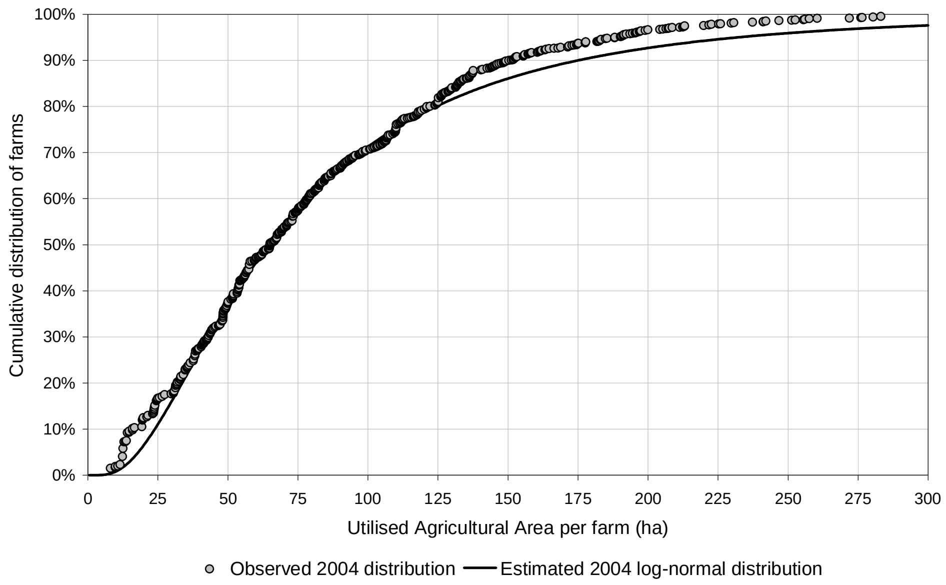
Figure 2 presents the observed and estimated cumulative distributions for the sub-population of individual farms specialised in field crops and whose holder's age lies between 35 and 50 years (*i.e.*, farms characterised by $TOF = 1$, $LST = 1$ and $AGE = 2$). It shows that, for this particular sub-population, the log-normal assumption: i) is quite satisfactory on a wide range of UAA (between 25 ha and 125 ha) where the observed farms are most numerous; but ii) gives a poorer picture of the population at either ends of the distribution.

Figure 3 presents the relative deviation between the 2004 observed cumulative distribution and the simulated ones when the various starting years are used for the projection, for all farms. We observe that, regardless of the year used as the base year, the deviation between the two distributions is less than 5% in absolute value across the entire range of UAAs, a result that is in itself highly satisfactory. However, the projection appears to be better on the whole the closer the starting year is to the simulated year: the maximum deviation in absolute value narrows the closer the base year is to 2004, and even dips below 1% with 2003. Yet this seems logical as, the more time goes by, the more "things happen" that divert the real distribution from what it would be when following the trend found by our method.

In addition to these positive and reassuring findings, it is also interesting to note that the different curves in Figure 3 are all fairly similar. Even if the UAA bounds vary from one year to the next, our model tends to:

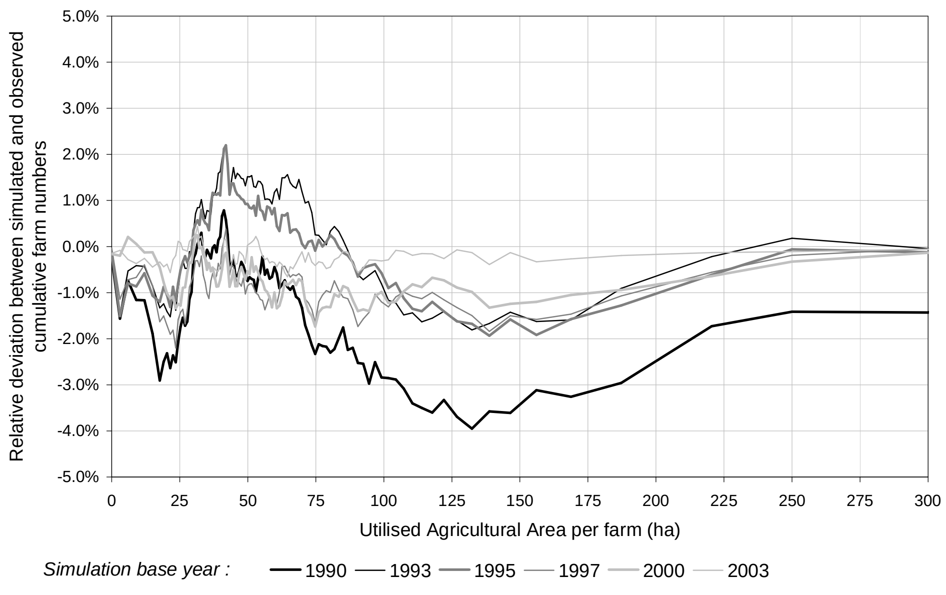
- Underestimate the numbers of farms with a small UAA;
- Overestimate them for the "average" UAAs;
- Underestimate them again for the largest farms.

Figure 2: Observed and estimated log-normal distributions for the 2004 sub-population of individual farms (LST = 1) specialised in field crops (TOF = 1) and whose holder's age lies between 35 and 50 years (AGE = 2)



Source: estimations and Farm Accounting Data Network (FADN) 2004

Figure 3: Deviation between simulated and observed distributions in 2004 by farm size and simulation base year (all farms)



Source: simulations and Farm Accounting Data Network (FADN) 2004

Although we cannot be sure purely on the basis of these results that these "biases" are a persistent characteristic of our model, it would be interesting to see their cause.[13] We could then endeavour to remedy them either *ex ante*, by altering the model, or *ex post*, by adjusting its results.

4.2. Simulation for 2005

The above subsection looks at the similarity between observed and simulated distributions in terms of overall farm numbers; it says nothing regarding the quality of our projection with respect to other key variables of interest which are present in the FADN database but which are not used as covariates in our model (cropping patterns, quantities produced, yields, labour used, economic variables, etc.).

In order to assess whether our model is satisfactory or not on this chapter also, we computed the distributions of several such variables with the FADN 2005 database, but in place of using the true 2005 extrapolation coefficients – the ones available in the database – we used our simulated 2005 coefficients with year 2000 as the base year. Then, we compared the obtained distributions with the ones actually observed in the 2005 FSS.

The variables which were available in the 2005 FSS and could be used for this purpose are the following:

- distribution of UAA and Annual Work Units (AWUs) by type of farming (TF);
- distribution of farm numbers, UAA and AWUs by farm surface area classes;
- distribution of UAA and AWUs by legal status;
- distribution of farm numbers and standard gross margin (SGM) by economic size classes measured in European Size Units (ESUs)

While this is not an exhaustive study, it is enough to provide some elements of evaluation. The results of these comparisons are given in Figures 4 to 7. In relative terms, i.e. in terms of distribution, the similarity between our projection and the figures actually observed in the 2005 FSS is entirely satisfactory.

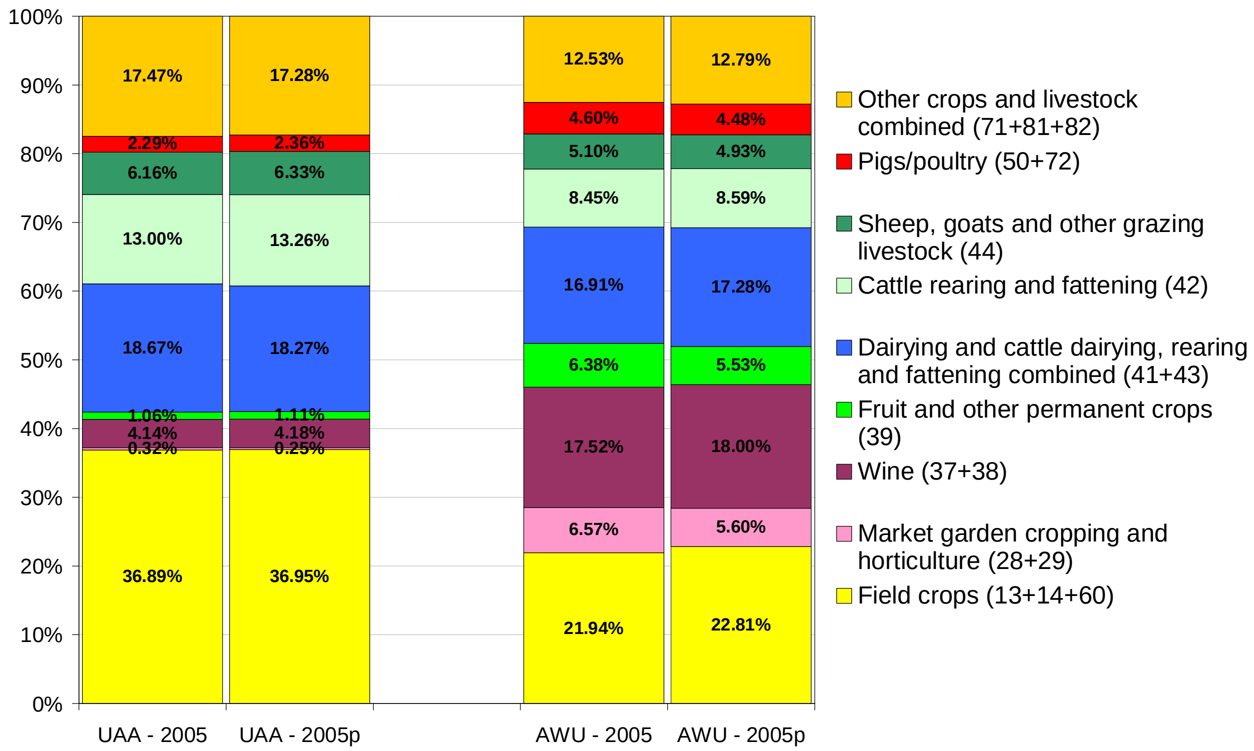
[13] They could be due, for example, to the fact that the log-normal approximation can be but imperfect, especially at either ends of the distributions, as noticed in Figure 2.

The observed deviations rarely exceed 1% and never 3%. In absolute terms, our method renders:

- a total commercial UAA of 25.4 million hectares when the 2005 FSS estimates it at 25.3 million hectares, representing a deviation of less than 0.5%;
- a total number of 681,552 AWUs when the 2005 FSS counts 751,155 AWUs, representing a deviation of less than 10%;
- a total SGM of 27.1 million ESUs when the 2005 FSS estimates it at 26.7 million ESUs, representing a deviation of just over 1%.

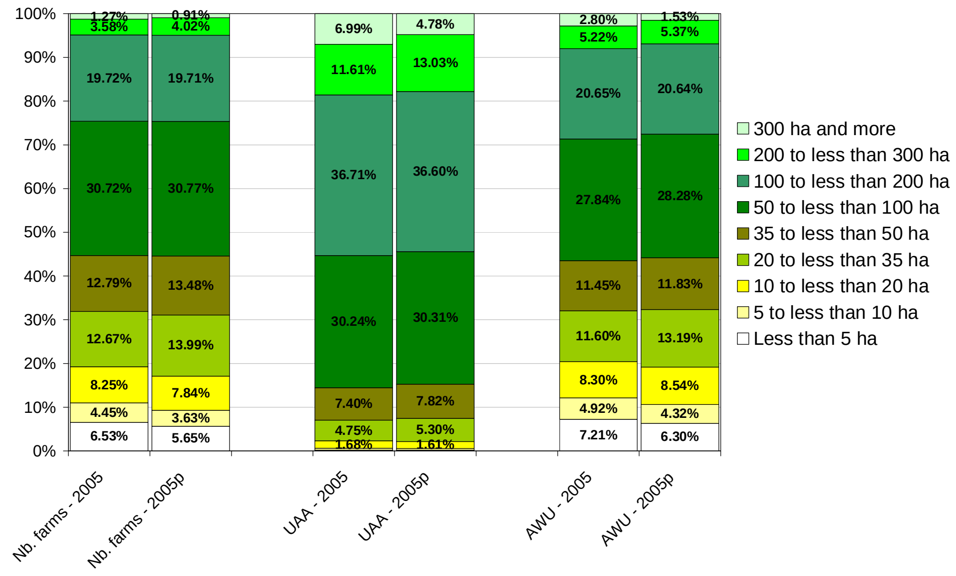
Here again, the similarity is highly satisfactory, including when accounting for variables (AWUs and SGM) not used in our modelôs estimations.

Figure 4: Comparison of the distribution of Utilised Agricultural Area (UAA) and Annual Work Units (AWU) by Type of Farming (TF) between the 2005 Farm Structure Survey (ô2005ô columns) and our simulations (ô2005pô columns)



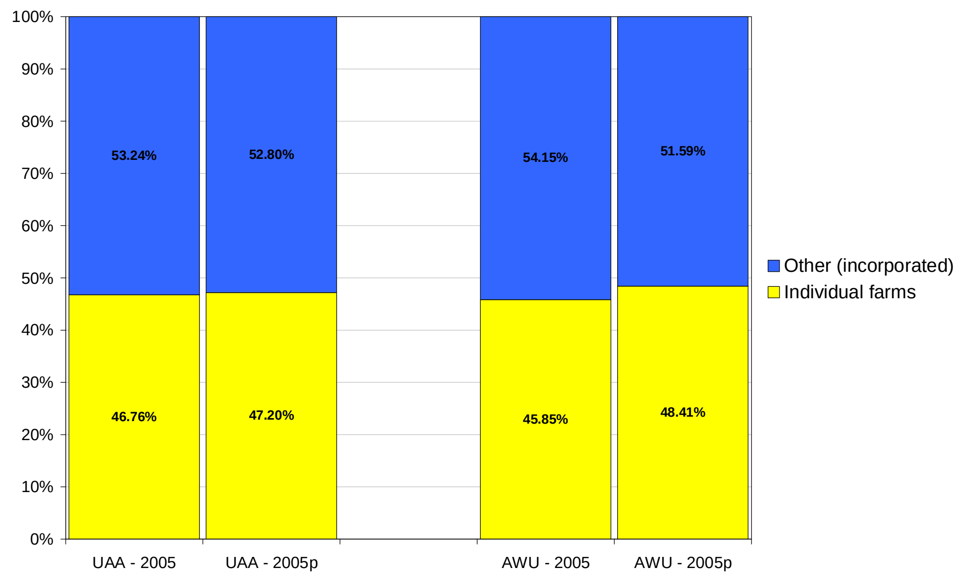
Source: simulations and Farm Structure Survey (FSS) 2005

Figure 5: Comparison of the distribution of farm numbers, Utilised Agricultural Area (UAA) and Annual Work Units (AWU) by farm surface area classes between the 2005 Farm Structure Survey (ó2005ô columns) and our simulations (ó2005pô columns)



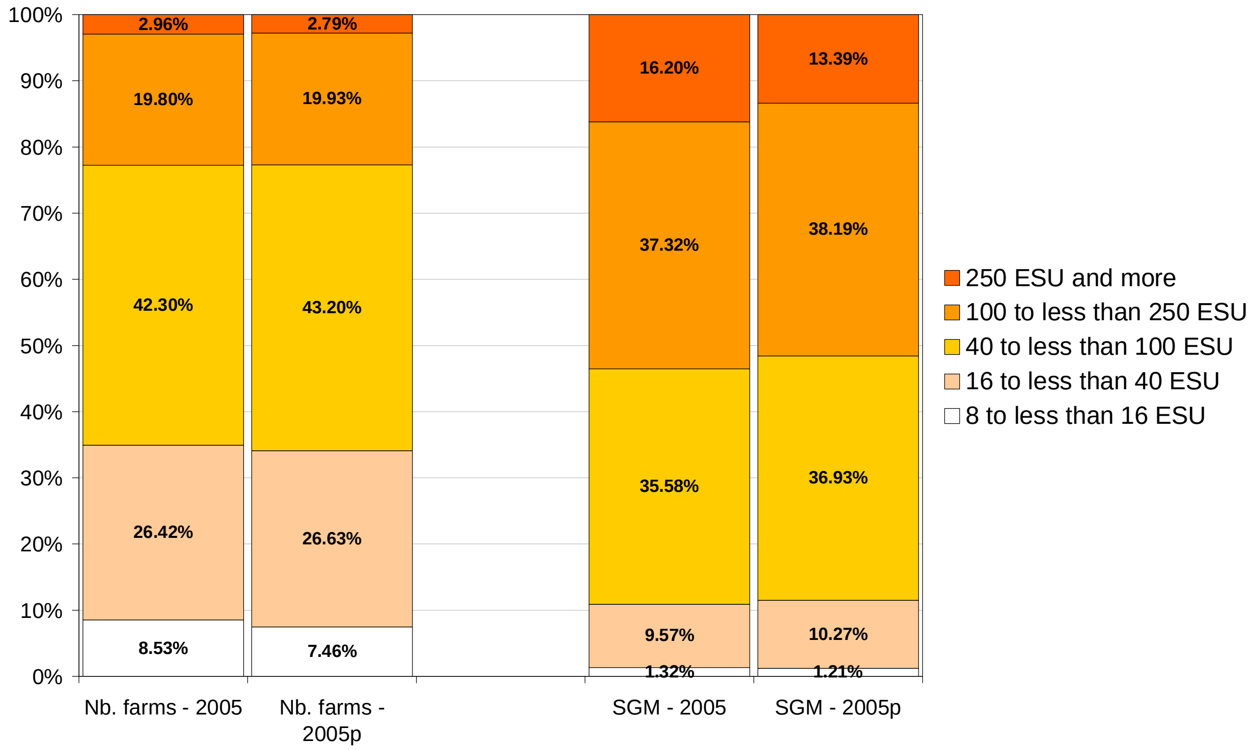
Source: simulations and Farm Structure Survey (FSS) 2005

Figure 6: Comparison of the distribution of Utilised Agricultural Area (UAA) and Annual Work Units (AWU) by legal status between the 2005 Farm Structure Survey (ó2005ô columns) and our simulations (ó2005pô columns)



Source: simulations and Farm Structure Survey (FSS) 2005

Figure 7: Comparison of the distribution of farm numbers and Standard Gross Margin (SGM) by economic size classes between the 2005 Farm Structure Survey (ó2005ô columns) and our simulations (ó2005pô columns)



Source: simulations and Farm Structure Survey (FSS) 2005

5. Conclusion

The projection method developed appears to be relatively effective and robust. It can simulate the distribution of commercial farms based on the trends observed in recent years, bearing in mind that these trends do not reflect just ñpurelyò demographic effects, but also take in the impact of changes in the farmsô economic environment over the period considered and especially the impact of policies affecting the agricultural sector. This said, it may well not be surprising to find that the projection with 1990 as its starting year is not as good as the projection starting in 2003, since the Common Agricultural Policy (CAP) was extensively reformed twice in the intervening years with the MacSharry reform in 1992 and the Agenda 2000 reform in 1999. However, we do not consider the model to be sufficiently tried and tested to be able to purely and simply put the deviation between the two simulations down to the structural impact of these two reforms alone.

This proviso is all the greater since, such as it is presented here, our model assumes an identical growth rate for all the types of farms considered. The model could then be improved by introducing interactions between the qualitative explanatory variables and the time variables in order to show expansion and contraction differences (where such exist) between farm categories. In recent years, for example, incorporated farms have grown faster than the average (Ratin, 2007).

Another way to develop the model further would be to enrich it with new explanatory variables available in the FADN. These could be ñstructuralò variables, such as the farmôs region and the breakdown of labour between family labour and paid labour, or economic and financial variables, such as the level of net worth and liabilities.[14] With the growing decoupling of CAP subsidies, it would also be interesting to introduce the level of direct aid (whether the ñoldò aid of the kind introduced in 1992 or the more recent Single Payment Scheme support) for a direct study of its impact, in terms of farm structure, in line with what the non-stationary Markov chain models do today.

Yet the fact remains that, as mentioned in the introduction, our approach is limited in that it cannot simulate the number of farms and can only simulate their detailed distribution characteristics. An approach such as the one we propose, combined with an econometric estimation of farm survival rates, by duration model, for example, would develop a cohort of farms over time that could be used to gain a highly detailed picture of the most probable population of farms at any given moment in time.

[14] However, as with all regression models, caution is called for when studying the correlations between variables, as correlations presumably exist between the dominant type of labour, family or paid, and the farmôs legal status.

References

Ben Arfa, N., Jacquet, F., Flichman, G. (2006). *Decoupling of the farm income support and structural change: a Markov chain model application to crop farms of Midi-Pyrénées.* Paper presented at the 93rd seminar of the European Association of Agricultural Economists, « Impacts of decoupling and cross-compliance on agriculture in the enlarged EU », Prague, 22-23 September 2006.

Butault, J.P., Delame, N. (2005). Concentration de la production agricole et croissance des exploitations. *Economie et Statistiques*, 390: 47-64.

Chavas, J.P. (2001). Structural change in agricultural production. In: Gardner, B.L., Rausser, G.C. (eds.), *Handbook of Agricultural Economics*, Vol. 1A, Elsevier: 263-285.

Desriers, M. (2007). L'agriculture française depuis cinquante ans : des petites exploitations familiales aux droits à paiement unique. I*n: L'agriculture nouveaux défis*, coll. INSEE Références, 17-30.

Deville, J.C., Särndal, C.E., Sautory, O. (1993). Generalized raking procedures in survey sampling. J*ournal of the American Statistical Association*, 88(423): 1013-1020.

Fansten, M. (1969). Distribution des facteurs de production et comportements individuels dans l'activité agricole, *Série Etudes du Service Central des Enquêtes et Etudes Statistiqu*e 48, Paris, Ministère de l'Agriculture et de la Pêche, 178 p.

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*, 51(2): 289-302.

Judge, G.G., Swanson E.R. (1961). Markov chains: basic concepts and suggested uses in agricultural economics. *Australian Journal of Agricultural Economics*, 6(2): 49-61.

Karantininis, K (2002). Information-based estimators for the non-stationary transition probability matrix: an application to the Danish pork industry. *Journal of Econometric*, 107(1-2): 275-290.

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

MAP (2007). Assises de l'agriculture. Quels objectifs pour une politique agricole dans une perspective 2013 ?, *Document d'orientation du Conseil supérieur de coordination et* *d'orientation de l'économie agricole et agroalimentaire*, Paris, Ministère de l'Agriculture et de la Pêche, 6 p.

Ratin, S. (2007). Les sociétés agricoles en pleine croissance. *Agreste Cahiers, 2*: 49-56.

Stokes, J.R. (2006). Entry, exit and structural change in Pennsylvania's dairy sector. *Agricultural and Resource Economics Review*, 35(2): 339-347.

Zepeda, L. (1995a). Technical change and the structure of production: a non-stationary Markov analysis. *European Review of Agricultural Economics*, 22: 41-60.

Zepeda, L. (1995b). Asymmetry and nonstationarity on the farm size distribution of Wisconsin milk producers: an aggregate analysis. American Journal of Agricultural Economic, 77: 837-852.