The evolution of farm size distribution: revisiting the Markov chain model

Laurent PIET[1,2]
[1] **INRA, UMR1302 SMART, 35000 Rennes, France**
[2] **Agrocampus, UMR1302 SMART, 35000 Rennes, France**
Adresse de correspondance : INRA, UMR1302 SMART, 4 allée Adolphe Bobierre, CS 61103 – F-35011 Rennes cedex – laurent.piet@rennes.inra.fr

2èmes journées de recherches en sciences sociales

INRA SFER CIRAD

11 & 12 décembre 2008 – LILLE, France

The evolution of farm size distribution: revisiting the Markov chain model

Laurent PIET[1,2,*]

June 5, 2008

Abstract

In this paper, a continuous version of the Markov Chain Model (MCM) is proposed to project the number and the population structure of farms. It is then applied to the population of professional French farms. Rather than working directly with transition probabilities as in the traditional, discontinuous, MCM, this approach relies on the close but not identical concept of growth rate probabilities and exploits the Gibrat's law of proportionate effects which appears to be supported by the French data. It is shown that the proposed continuous MCM is a more general approach, since it enables to derive more in-depth detail on the distrubtion of the projected population and the traditional MCM transition probability matrix can be easily reconstructed from the estimated growth rate probabilities. Though the continuous MCM is presented in this paper in a stationary framework, it should be possible to develop a non-stationary version in a similar way traditional MCMs are now made non-stationary.

Keywords: *Distribution of farm size, Gibrat's law, Markov Chain Model, Continuous model*

JEL Classification: *Q12, C15, C53*

[1] INRA, UMR1302 SMART, F-35000 Rennes

[2] Agrocampus Rennes, UMR1302 SMART, F-35000 Rennes

* Corresponding address : INRA, UMR1302 Unité Mixte de Recherche SMART, 4 allée Adolphe Bobierre, CS 61103 – F-35011 Rennes cedex – tél. : 02.23.48.53.83 – fax. : 02.23.48.53.80 – laurent.piet@rennes.inra.fr

1. INTRODUCTION

The so-called Markov chain model (MCM) is becoming a more and more popular tool to predict the number and the distribution structure of a population of agricultural firms. Soon after the pioneering applications to iron and steel industries (Adelman, 1958), the MCM has been widely applied to food industries (Hallberg, 1969, Padberg, 1962) and to farm units (Colman, 1967, Harrison and Alexander, 1969, Judge and Swanson, 1961, Krenz, 1964, Stanton and Kettunen, 1967).

Early researches such as Lee et al. (1965) showed that it is possible, with the use of econometric estimation, to build a robust MCM from aggregate (cross-sectional) data only, alleviating the difficulty of individual (panel) data availability (see also Lee et al. (1977)). Since then, most, if not all, the available literature on the use of the MCM approach in agricultural economics employs aggregate data (see Table 1 in the next section). Though the issue of building a non-stationary[1] MCM was early addressed to take into account the dynamic nature of the underlying microeconomic processes (Hallberg, 1969), one has to wait for the works of Chavas and Magand (1988) and Disney et al. (1988) to see such a feature introduced in a MCM of farm size and number evolution. By this time, only market variables, especially prices or input/output price ratios, were used as explanatory variables of the underlying non-stationary model parameters. Though not really non-stationary *per se*, the work by Keane (1991) introduced the influence of a policy variable, namely the introduction of milk quotas in the Irish dairy sector, by building a MCM for each of the two periods, before and after quotas. In the last decades, following von Massow et al. (1992), Zepeda (1995a, 1995b) or Rahelizatovo and Gillespie (1999), an increasing number of market (prices), macroeconomic (interest rates), policy (price support, direct payments, diversion or termination programs…) and individual farm financial (debt-equity ratio) or technical (productivity indices) variables have been introduced into explicitly non-stationary models.

With years and experience, we can see from above that the models used in the MCM literature applied to agricultural economics have become more sophisticated. Still, there is one aspect that, to our knowledge, modellers have not addressed so far. As will be shown in the next section, the underlying transition matrix of a MCM is built by discretizing the whole population of farms into a (limited) number

[1] See section 2 for a definition of this term.

of classes on the basis of some particular size criterion. In other words, a traditional MCM implements a "histogram" approach rather than a truly "distribution" point of view. We intend to tackle here this issue of continuity. Our results show that the continuous approach we propose is much more informative, as far as the evolution of the structural distribution of farms is concerned, than the traditional, discontinuous, approach.

The rest of the paper is structured as follows. In the next section, we describe the traditional MCM approach and its most common features, showing why its discrete nature may constitute an issue. Section 3 presents the continuous version of the MCM we propose, and section 4 applies it to the case of French professional farms. Section 5 discusses some key elements and concludes.

2. THE TRADITIONAL MCM APPROACH AND THE "HISTOGRAM" ISSUE

In the traditional MCM, the studied population, a population of farms in our case, is broken down into a finite number J of classes, the so-called "states-of-nature". Denoting the number of individuals (farms) in the $j-th$ state (with $j=\{1..J\}$) at time t by $n_{j,t}$, the demography of the population follows a Markov chain process of degree 1 if, between two dates t and $t+1$, the following relation holds:

$$n_{j,t+1}=\sum_{k=1}^{J}p_{kj}n_{k,t} \qquad (1)$$

where p_{kj} is the probability for a farm to move from state k to state j in one time-period τ, with $p_{kj}\geq 0$ and $\sum_{j=1}^{J}p_{kj}=1$. In the MCM approach, p_{kj} are the model's parameter to be determined. In most economic and social sciences studies, the states-of-nature of the MCM correspond to groupings on the basis of some size variable, whatever the definition of size according to the type of population studied.

Equation (1) expresses the fact that the structure of the population (the number of farms lying in each category) at one date only depends on the structure of the (whole) population at the previous date.[2]

[2] Higher degree MCMs allow for the structure of the population at one date to depend on the structure of the population at several previous dates (Berchtold, 1998).

In order to ensure that $\sum_{j=1}^{J} p_{kj} = 1$, an “exit” state-of-nature is usually added, stating that farms may disappear between to dates; similarly, an “entry” category usually allows to account for new comers. All together, the set of probabilities p_{kj} define a square matrix $\mathbf{P} = \left(p_{kj}\right)$ which is called the transition probability matrix (TPM) so that equation (1) may be re-written in a matricial form:

$$\mathbf{N}'\left(t+1\right) = \mathbf{P}' \times \mathbf{N}\left(t\right) \qquad (2)$$

where $\mathbf{N}\left(t\right)$ is the vector representing the structure of the population at time t.

[insert Table 1]

Table 1 presents in a synthetic way some of the key features characterizing MCMs applied to populations of farms from a (non-exhaustive) literature review. It first shows that the MCM approach has been used in several national or sub-national contexts and for studying various types of farms (of which the relative frequency of dairy farms is worth noticing). Second, it is striking that the vast majority of studies use aggregate data to determine the transition probabilities: though it is relatively straightforward to compute such probabilities from the survey of a panel of individual farms, such data are usually not or too costly available, therefore precluding the direct calculation of the TPM. As it was mentioned in the introduction of the paper, Lee et al. (1965) and later Lee et al. (1977) made a seminal contribution showing how to retrieve the elements of the TPM from aggregate data only. Third, the models used are more and more non-stationary: in the stationary MCM, the transition probabilities p_{kj} are assumed to be constant parameters over time; if this can seem a fairly reasonable first approximation, there are many good reasons for which it should not be so, and recent works have tried to show some evidence of the impact of various variables such as input and/or output prices, technical efficiency, and more and more political variables. In the non-stationary framework, transition probabilities are no longer supposed fixed but vary over time (either because of time itself or because explanatory variables vary themselves).

The last feature listed in Table 1 that is worth shedding light on, is the number of classes used, i.e., the number of states-of-nature considered, as it is related to the motivation for this paper.[3] Table 1 shows that this number is usually limited: it rarely exceeds 10, the 19 classes used by Karantininis (2002) appearing to be an unusual maximum, the average being 7. Apart from keeping the model tractable, the stability issue of the TPM is an important (mathematical) reason for keeping such a small number of classes (Berchtold, 1998). Nevertheless, this characteristic of the traditional MCM presents the main drawback that the resulting TPM is strongly diagonal: the size intervals defining the states-of-nature are so wide that the highest probability for a farm is by large to stay in the same class; complementarily, the probability for a farm to actually experience a change in size corresponding to one or more categories rapidly falls to zero. Of course, the TPM is less diagonal when the transition period over which it is estimated is longer.

[insert Table 2*]*

Table 2 is a typical example of such a diagonal TPM, corresponding to annual transitions, taken from the recent study by Stokes (2006): for a farm lying in the size interval $[50;99]$, staying in the same class means experiencing, in one year, a relative change in size (both increasing or decreasing) as large as by a factor of two; as for moving two categories upward, that is to say joining the size class $[200;499]$, it means, again in one year, at least doubling, at most growing by a factor of about 10! Figuring explicitly those underlying relative changes makes the structure of the transition probabilities no longer surprising.

Adopting such a limited discretization of the population structure must not be viewed as a problem *per se*, since the primary objective of a MCM is to predict the total number of farms in the population. But by doing so, this model fails to give useful information on the fine structure of the projected

[3] In particular, though an interesting issue, we shall not discuss here the question of the variable used to define the size of a farm.

population or a precise indication of, among other interesting indicators, the future average size of farms.[4] Also, such a discretization may lead to the spurious conclusion that the distribution of the farm population is, or is becoming, bimodal, a feature that may only represent an artifact due to the definition of the size interval bounds (a classic issue when dealing with histograms). To avoid these problems, we propose in the next section a continuous version of the MCM, then and apply it to the population of professional French farms in section 4.

3. A CONTINUOUS MARKOV CHAIN MODEL

Table 3 shows that the cumulated distributions with respect to the utilised agricultural area (UAA), used as the size variable, of the population of professional farms represented in the French sample of the Farm Accounting Data Network (FADN)[5] may be adjusted, each year from 1980 to 2005, by a lognormal density function of the form:[6]

$$n(h)_t = \frac{N_t}{h\sqrt{2\pi}\sigma_t} e^{\frac{1}{2}\frac{(\ln(h)-\mu_t)^2}{\sigma_t^2}} \quad (3)$$

for $h > 0$ and where μ_t and σ_t are the log-normal distribution parameters and N_t is the total number of farms in the population; in the rest of the paper, we will denote this distribution as:

$$n(h)_t = \frac{N_t}{h}\phi\left(\frac{\ln(h)-\mu_t}{\sigma_t}\right) \quad (3')$$

with ϕ the probability density function of the standard normal distribution.

---

[4] The study by Butault and Delame (2005) appears as a worth noticing exception: using large scale panel data, the authors worked with a large number of states-of-nature which are not only defined upon the size in hectares but also the region, the type of farming, the economic size, the legal status of the farm or the age of the operator; thought they thus obtained a fine picture of the structure of the population regarding these politically interesting variables, neither can they directly recover indicators such as the projected average size of farms.

[5] These FADN cumulative distributions are obtained by taking into account the extrapolation factor attached to each farm in the sample.

[6] Actually, these distributions are best fitted by a 3-parameter log-normal density function, rather than the (standard) 2-parameter log-normal density function used here. At the time of writing, calculations with such 3-parameter distributions are still in progress.

[insert Table 3*]*

It can be shown that the lognormal nature of a size distribution might be related to the so-called Gibrat's law of proportionate effects (Kleiber and Kotz, 2003). Schematically, this empirical "law" states that the probability for, say, firms to experience a certain relative growth between two dates is independent of the initial size of the firm and is the same for all firms exhibiting the same initial size. Some authors have shown that this law does not hold in the agricultural context they studied, e.g. Canada for Shapiro et al. (1987) or upper Austria for Weiss (1999); inversely, the analysis reported by Butault and Delame (2005) for France does not lead to a full rejection of the Gibrat's law hypothesis.

Here, the persistence of the lognormal nature of the size distributions of the FADN farm population across time inclines to assume the Gibrat's law assumption as sufficiently plausible. Then, it can be written that the probability for farm i, with initial size $h_{i,t}$ at time t, to exhibit a size of $h_{i,t+\tau} = (1+\delta h) h_{i,t}$ at time $t+\tau$ (i.e., to experience a relative growth at rate δh with $-1 < \partial h < +\infty$) is constant, whatever the initial size:

$$P\left(h_{i,t+\tau} = (1+\delta h) h_{i,t}\right) = p(\delta h)_{\tau} = \text{constant} \tag{4}$$

Under the further assumption that the growth of a particular farm between two dates is independent of the growth of other farms[7], the total number of farms that will have a size h at time $t+\tau$, $n(h)_{t+\tau}$, is given by the following convolution over δh formula:

$$n(h)_{t+\tau} = \int_{-1}^{\infty} p(\delta h)_{\tau} . n\left(\frac{h}{1+\delta h}\right)_{t} d(\delta h) \tag{5}$$

where $n\left(\frac{h}{1+\delta h}\right)_{t}$ is the number of farms that exhibited a size $\frac{h}{1+\delta h}$ a time t.

[7] This is obviously not true at the individual scale: the actual growth possibilities of one particular farm will depend on its growing opportunities, that is, on the quantity of land made available by the reduction in size or exit of neighbouring farms and on its position in the competition for such land. But, at a macro, aggregate scale, this assumption may be thought reasonable, in the sense that the growth of a particular farm in one region is independent of the growth of farms in other (remote) regions.

Equation (5) may be regarded as a continuous Markov model which gives the population at time $t+\tau$ from the observed population at time t, and the probability $p(\delta h)_\tau$ for farms to grow at rate δh between these two dates. This continuous Markov model is not directly equivalent to the traditional MCM though, since the probabilities $p(\delta h)_\tau$ do not exactly correspond to the transition probabilities of the traditional MCM. One can however easily turn back to a traditional MCM, since the probability of transition to state-of-nature j, defined by the size interval $[X,Y]$ (with $0<X<Y$), over a period τ for farms initially in state-of-nature k, defined by the size interval $[x,y]$ (with $0<x<y$), is given by:

$$p_{k=[x,y]\to j=[X,Y],\tau} = \frac{\int_x^y \left(\int_{\frac{X-h}{h}}^{\frac{Y-h}{h}} p(\delta h)_\tau . n(h)_t \, d(\delta h) \right) dh}{\int_x^y n(h)_t \, dh} \qquad (6)$$

with the notations of equation (5).

The interesting point is that, while the traditional MCM only derives a limited number of transition probabilities, the continuous approach that we propose allows to calculate any of these probabilities from equation (6), especially by making size intervals $[x,y]$ and/or $[X,Y]$ as small as desired.

In other words, the continuous MCM proposed here outperforms by far the traditional MCM in terms of the richness of the structural information it allows to bring into light. As in the traditional MCM, it remains to determine, for a given transition period τ, the value of the probability $p(\delta h)_\tau$ for any growth rate δh. We examine this issue in the next section on the example of French professional farms.

4. AN APPLICATION TO THE DISTRIBUTION OF FRENCH FARMS

In this section, we present the results of estimating the probabilities $p(\delta h)_\tau$ from French FADN data for transition periods τ ranging from 1 year to 15 years, using data over the 1980-2005 period.[8] Estimations

[8] When dealing with a particular τ year transition, some of the combinations of the available years could not be used in the estimation process; the reason lies in the fact that the extrapolation coefficients attached to farms in the FADN sample are not re-evaluated each year but only every two or three years, when the results of the most recent survey on farm structures are made available.

were obtained by using the lognormal parameters displayed in Table 3 and the following equation derived from equations (3') and (5):

$$\phi\left(\frac{\ln(h)-\mu_{t+\tau}}{\sigma_{t+\tau}}\right)= \frac{N_t}{N_{t+\tau}}\int_{-1}^{+\infty} p(\delta h)_{\tau}.(1+\delta h).\phi\left(\frac{\ln\left(h/(1+\delta h)\right)-\mu_t}{\sigma_t}\right)d(\delta h)+\varepsilon_h \quad (7)$$

where ε_h are error terms.

It appears that a parametric approach may be used to derive the $p(\delta h)_{\tau}$ probabilities since assuming that they follow a 3-parameter log-normal distribution leads to satisfying results, especially in maintaining the log-normal distribution of the projected farm population distribution. Thus, we suppose that the $p(\delta h)_{\tau}$ probabilities that we want to estimate are of the form:

$$p(\delta h)_{\tau}=\frac{\alpha_{\tau}}{(\delta h+1)}\phi\left(\frac{\ln(\delta h+1)-\eta_{\tau}}{\nu_{\tau}}\right) \quad (8)$$

for $-1<\delta h<+\infty$ and where α_{τ}, η_{τ} and ν_{τ} are the parameters that we finally want to estimate from equation (7). These estimations were obtained using the nonlinear least-squares estimation procedure available in the Stata 10.0 software.

[insert Figure 1*]*

Figure 1 shows the estimated log-normal probability density functions of a relative change in area for the twelve 10-year transitions considered, both when they are studied individually and when they are used altogether. Figure 2 represents the evolution of this probability density function for transitions ranging from 1 year to 15 years; the corresponding parameters are given in Table 4. As could be expected, the curve shifts to the right and flattens when the transition period increases, meaning that the probability of larger growth rates (both negative and positive) also increases.

[insert Figure 2*]*

[insert Table 4*]*

We have then used these results to project the population of professional French farms at the horizon of 2015. To do so, we used the 10-year probability density function estimated above (with $\alpha_{10} = 0.5717$, $\eta_{10} = 0.1442$ and $\nu_{10} = 0.4209$, see Table 4) and used the 2005 population as a starting point. This leads to a projected 2015 population of $N_{2015} = 249{,}812$ farms, with $\mu_{2015} = 4.3826$ and $\sigma_{2015} = 0.9470$. The average farm size in 2015 is then easily obtained as $\tilde{h}_{2015} = e^{\mu_{2015} + \frac{\sigma_{2015}^{2}}{2}} = 125.34 \text{ ha}$.

It can be shown that, thanks to the log-normal distribution of the projected population, other interesting indicators such as quantiles can also easily be computed by:

$$F^{-1}(u) = e^{\mu + \sigma . \Phi^{-1}(u)} \quad (9)$$

where u is the desired quantile (with $0 < u < 1$) and Φ is the cumulative density function of the standard normal distribution (Kleiber and Kotz, 2003). In our case, with the figures above, we can then compute that:

- 10% of the population will operate less than 23.78 ha (1st decile);
- 50% of the population will operate less than 80.05 ha (median);
- 10% of the population will operate more than 269.41 ha (10th decile).

[insert Table 5*]*

Finally, as indicated in the previous section, a traditional transition probability matrix can be reconstructed from these estimations, according to equation (6). Table 5 reproduces such a TPM obtained starting from year 2005 and using the 10-year probability density function defined by the parameters of Table 4. Here, the bounds defining the size classes have been chosen arbitrarily but, once again, our approach enables us to set them at whatever value we want (except zero). Globally, this "reconstructed" TPM reproduces both the diagonal nature of the traditional MCM matrix and the fact that farms rarely move by more than one or two categories (upward or downward). Here, the probability of reaching neighbouring categories is however quite high compared to what was shown in Table 2; this is just because the transition period is longer (10 years instead of 1 year). It would be easy to show that the transition probabilities of such a matrix intrinsically depend on the bounds used to defined the size intervals: a supporting evidence of this in the example of Table 5 is that the diagonal element of the matrix becomes smaller as the relative width of the size interval with respect to the central value reduces (compare diagonal elements for classes $[50;99]$, $[100;149]$ and $[150;199]$). We can notice the poor information that would be derived from such a TPM on the evolution of the sub-population of larger farms in a traditional MCM setting, as the last category ($[200;+]$) definitely acts as an "absorbing" class. On the other hand, Table 5 also demonstrate that our approach has the major drawback that the exit probability is constant whatever the size class, as it is simply given by:

$$p(-1)_{\tau} = 1 - \frac{N_{t+\tau}}{N_t} \qquad (10)$$

This issue will be discussed in the next section.

5. DISCUSSION AND CONCLUDING REMARKS

In this paper, we presented a continuous version of the traditional MCM. Rather than working directly with transition probabilities, this approach relies on the close but not identical concept of growth rate probabilities. We have shown that this is a more general approach since it brings more in-depth detail on the distribution of the projected population and the traditional MCM can be easily reconstructed from the estimated growth rate probabilities.

Just like any other model, the traditional MCM relies on a set of assumptions: these were strong and quite arbitrary in the early implementations of the MCM to agriculture (like in Krenz (1964)); recent approaches, especially those using a Generalised Cross Entropy method to estimate the model's parameters (e.g. Jongeneel (2002), Karantininis (2002), Zepeda (1995a, 1995b)), are more flexible and insist on the use of prior information rather than rigid a priori assumptions. Here, the model is only based on two assumptions, the strongest of which being the Gibrat's law of proportionate effects[9]. Nonetheless, it appears to be quite well supported by data in the French case and can be seen as a plausible first approximation.

This assumption may be seen as quite strong, though, in the general case, especially as far as exit is concerned: it is usually accepted that the probability of exiting the agricultural sector is higher far smaller farms, especially because exit is often preceded by a "decapitalization" phase (Butault and Delame, 2005)[10]. It should be noted though that the model developed here does not deal with entry: "exit" should thus be seen as net exit. While this might not totally compensate, it could happen that entries are also more frequent at smaller sizes, that is at the size where farms are made available for takeover by exits. Anyway, it is an on-going research to study if and how this continuous version of the MCM could be refined by releasing the Gibrat's law assumption for some or all sizes and/or growth rates and how to better account for both entries and exits.

Finally, we would like to stress that the discrete approach developed in the traditional MCM does not constitute a "problem" *per se* as far as predicting the total number of farms is concerned: this is a powerful and relatively easy to implement tool to do so. But it gives little information on the fine structure of the projected population. The continuous MCM developed here overcomes this lack of information and is no less efficient in forecasting the total number of farms. We presented here a stationary version, but there is no doubt it could be made non-stationary, in the same line as the one used in the recent developments made to the traditional MCM; it could then bring into light valuable information regarding the impact of policy instruments and other market and/or technical variables. This is also a possible direction for further research.

---

[9] Recall that the second assumption is that the growth of a particular farm between two dates is independent of the growth of other farms (see supporting elements in footnote 7).

[10] In fact, this might be true especially in field crops, dairy and cattle production; farms specialised in horticulture are usually relatively small in terms of operated area and still they are viable; in their case, an economic indicator would be better adapted to represent their size. Again, the discussion regarding the definition of size in agriculture is beyond the scope of this paper.

6. REFERENCES

Adelman, I. G. "A Stochastic Analysis of the Size Distribution of Firms." *Journal of the American Statistical Association* 53, no. 284(1958): 893-904.

Benarfa, N., and K. Daniel (2007) Agricultural policies and structural change of French dairy farms: Application of a non-stationary Markov model. European Association of Agricultural Economists Ph-D Workshop, Rennes (France), 21 pp.

Benarfa, N., F. Jacquet, and G. Flichman (2006) Decoupling of the farm income support and structural change : a Markov chain model application to drop farms of Midi-Pyrénées. 93rd Seminar of the European Association of Agricultural Economists, Prague (Czech Republic), 17 pp.

Berchtold, A. (1998) *Chaînes de Markov et modèles de transition : application aux sciences sociales*. Paris (France): Editions Hermès.

Buckwell, A. E., D. M. Shucksmith, and D. A. Young. "Structural Projections of the Scottish Dairy Industry Using Micro and Macro Markov Transition Matrices." *Journal of Agricultural Economics* 34, no. 1(1983): 57-69.

Butault, J.-P., and N. Delame. "Concentration de la production agricole et croissance des exploitations." *Economie et Statistique*, no. 390(2005): 47-64.

Chavas, J.-P., and G. Magand. "A dynamic analysis of the size distribution of firms: The case of the US dairy industry." *Agribusiness* 4, no. 4(1988): 315 - 329.

Colman, D. R. "The Application of Markov Chain Analysis to Structural Change in the North West Dairy Industry." *journal of Agricultural Economics* 18, no. 3(1967): 351-363.

Disney, W. T., P.-A. Duffy, and W.-E. Hardy. "A Markov Chain Analysis of Pork Farm Size Distributions in the South." *Southern Journal of Agricultural Economics* 20, no. 2(1988): 57-64.

Edwards, C., M.-G. Smith, and R. N. Peterson. "The Changing Distribution of Farms by Size: A Markov Analysis." *Agricultural-Economics-Research* 37, no. 4(1985): 1-16.

Gaffney, P. (1997) "A Projection of Irish Agricultural Structure Using Markov Chain Analysis." CAPRI Working Papers. Dept of Economics, NUI Galway, 21 pp.

Hallberg, M. C. "Projecting the Size Distribution of Agricultural Firms. An Application of a Markov Process with Non-Stationary Transition Probabilities." *American Journal of Agricultural Economics* 51, no. 2(1969): 289-302.

Hammond, J. W. (1994) *Trends In The Size Distribution Of Dairy Farms In Minnesota And Wisconsin*: Staff Papers 94-27, University of Minnesota, Department of Applied Economics, 18 pp.

Harrison, H. "Trends in Agricultural Land, Labour and Farming Units in Northern Ireland from the present time until 2000 AD." *Irish Journal of Agricultural Economics and Rural Sociology* 6, no. 1(1976): 75-87.

Harrison, H., and D. J. Alexander. "Projection of Farm Numbers and Farm-Size Structure with Markov Chains." *Irish Journal of Agricultural Economics and Rural Sociology* 2, no. 1(1969): 105-116.

Jongeneel, R. A. (2002) An analysis of the impact of alternative EU dairy policies on the size distribution of Dutch dairy farms: an information based approach to the non-stationary Markov chain model. Xth Congress of the European Association of Agricultural Economists, Zaragoza (Spain).

Jongeneel, R. A., N. B. P. Polman, and L. H. G. Slangen (2005) Explaining the changing institutional organisation of Dutch farms: the role of farmer's attitudes, advisory network and structural factors. XIth Congress of the European Association of Agricultural Economists, Copenhagen (Denmark).

Judge, G. G., and E. R. Swanson. "Markov chains: Basic concepts and suggested uses in agricultural economics." *Australian Journal of Agricultural Economics* 6, no. 2(1961): 49-61.

Karantininis, K. "Information-based estimators for the non-stationary transition probability matrix: an application to the Danish pork industry." *Journal of Econometrics* 107, no. 1-2(2002): 275-290.

Keane, M. "Changes in the Size Structure of Irish Dairy Farms." *Irish Journal of Agricultural Economics and Rural Sociology* 14, no. 1(1991): 67-74.

Kleiber, C., and S. Kotz (2003) *Statistical size distributions in economics and actuarial sciences*, ed. D. J. Balding, et al. Hoboken (New Jersey), John Wiley and Sons, Inc., 332 pp.

Krenz, R. D. "Projection of Farm Numbers for North Dakota with Markov Chains." *Agricultural. Economics Research* 16(1964): 77-83.

Lee, T. C., G. Judge, and A. Zellner (1977) *Estimating the parameters of the Markov probability model from aggregate time series data*. Amsterdam North Holland, 254 pp.
Lee, T. C., G. G. Judge, and T. Takayama. "On Estimating the Transition Probabilities of a Markov Process." *Journal of Farm Economics* 47, no. 3(1965): 742-762.
MacMillan, J. A., F. L. Tung, and J. R. Tulloch. "Migration Analysis and Farm Number Projection Models: A Synthesis." *American Journal of Agricultural Economics* 56, no. 2(1974): 292-299.
McInerney, N., and E. Garvey. (2004) "Farm Structure and Agricultural Labour." CAPRI Working Papers. Dept of Economics, NUI Galway, 18 pp.
Padberg, D. I. "The Use of Markov Processes in Measuring Changes in Market Structure." *Journal of Farm Economics* 44, no. 1(1962): 189-199.
Rahelizatovo, N., and J. Gillespie. "Dairy Farm Size, Entry and Exit in a Declining Production Region." *Journal of Agricultural and Applied Economics* 31, no. 2(1999): 333-348.
Shapiro, D., R. D. Bollman, and P. Ehrensaft. "Farm Size and Growth in Canada." *American Journal of Agricultural Economics* 69, no. 2(1987): 477-483.
Stanton, B. F., and L. Kettunen. "Potential Entrants and Projections in Markov Process Analysis." *Journal of Farm Economics* 49, no. 3(1967): 633-642.
Stokes, J. R. "Entry, Exit, and Structural Change in Pennsylvania's Dairy Sector." *Agricultural and Resource Economics Review* 35, no. 2(2006): 357-373.
Tonini, A., and R. Jongeneel (2007) Modelling the Dairy Farm Size Distribution in Poland Using an Instrumental Variable Generalized Cross Entropy Markov Approach. 104th Seminar of the European Association of Agricultural Economists, Budapest (Hungary).
von Massow, M., A. Weersink, and C. G. Turvey. "Dynamics of Structural Change in the Ontario Hog Industry." *Canadian Journal of Agricultural Economics* 40(1992): 93-107.
Weiss, C. R. "Farm Growth and Survival: Econometric Evidence for Individual Farms in Upper Austria." *American Journal of Agricultural Economics* 81, no. 1(1999): 103-116.
Zepeda, L. "Asymmetry and nonstationarity in the farm size distribution of Wisconsin milk producers: an aggregate analysis." *American Journal of Agricultural Economics* 77(1995): 837-852.
Zepeda, L. "Technical change and the structure of production: a non-strationary Markov analysis." *European Review of Agricultural Economics* 22(1995): 41-60.

ACKNOWLEDGEMENT

The author would like to thank Fabrice Levert and Christine Valade, both from the UMR INRA-Agrocampus SMART, for their support, respectively in data preparation and bibliographic material collection.

Table 1. Synthetic review of the literature on the use of the Markov chain model

Reference	Country/State	Type of farms	Type of data	Type of model	Number of classes [a]	Size variable(s)	Transition
Judge and Swanson (1961)	Illinois (USA)	Hog	Individual	Stationary	7	Litters of hogs produced	Annual
Krenz (1964)	North Dakota (USA)	All	Aggregate	Stationary	7	Acres	5 years
Colman (1967)	Northwest region of England	Dairy	Individual	Stationary	6	Herd size in heads	Annual
Stanton and Kettunen (1967)	New York State (USA)	Dairy	Individual	Stationary	4	Herd size in heads	4 years
Harrison and Alexander (1969)	Ireland	All	Aggregate	Stationary	6	Standard man-days	6 years
MacMillan et al. (1974)	Canadian Prairie Provinces	All	Aggregate	Stationary	7	Gross receipts	5 years
Harrison (1976)	North Ireland	All	Aggregate	Stationary	5	Acres	Annual
Buckwell et al. (1983)	Scotland	Dairy	Both	Stationary	7	Herd size in heads	3 years
Edwards et al. (1985)	USA	All	Individual	Stationary	9	Acres	4 years
Chavas and Magand (1988)	Five regions of the USA	Dairy	Aggregate	Non-stationary	4	Herd size in heads	Annual
Disney et al. (1988)	South Atlantic Census division of the USA	Hog	Aggregate	Non-stationary	5	Number of hogs marketed	4-5 years
Keane (1991)	Ireland	Dairy	Aggregate	Stationary	8	Herd size in heads	6 years
von Massow et al. (1992)	Ontario (USA)	Hog	Aggregate	Both	6	Number of hogs marketed	Annual
Hammond (1994)	Minnesota and Wisconsin (USA)	Dairy	Aggregate	Stationary	6	Herd size in heads	5 years
Zepeda (1995a)	Wisconsin (USA)	Dairy	Aggregate	Non-stationary	5	Herd size in heads	Annual
Zepeda (1995b)	Wisconsin (USA)	Dairy	Aggregate	Non-stationary	4	Herd size in heads	Annual
Gaffney (1997)	Ireland	Dairy, cattle, hogs, sugar beet and cereals	Aggregate	Stationary	5 to 7	Herd size in heads, hectares	12 years
Rahelizatovo and Gillespie (1999)	Louisiana (USA)	Dairy	Individual	Non-stationary	5	Number of lbs/day produced	7 years
Karantininis (2002)	Denmark	Hog	Aggregate	Non-stationary	19	Number of hogs marketed	Annual
McInerney and Garvey (2004)	Ireland	Dairy	Aggregate	Stationary	7	Herd size in heads	12 years
Butault and Delame (2005)	France	All	Individual	Stationary	na [b]	Standard Gross Margin	9 years
Jongeneel et al. (2005)	The Netherlands, Germany, Poland and Hungary	Dairy	Aggregate	Non-stationary	8	Herd size in heads	Annual
Benarfa et al. (2006)	Midi-Pyrénées (France)	Cash crops	Aggregate	Non-stationary	7	Hectares	2-3 years
Stokes (2006)	Pennsylvania (USA)	Dairy	Aggregate	Non-stationary	7	Herd size in heads	Annual
Benarfa and Daniel (2007)	France	Dairy	Aggregate	Non-stationary	8	Herd size in heads	Annual
Tonini and Jongeneel (2007)	Poland	Dairy	Aggregate	Non-stationary	9	Herd size in heads	Annual

[a]: when appropriate, "entry" and/or "exit" states-of-nature are included in the counting.

[b]: the exact number of classes is not given since the transition matrix is built as a multidimensional one, crossing such variables as hectares, region, type of farming, economic size measured in ESU (European Size Unit), legal status of the farm (individual or corporate), etc.

Table 2. A typical "traditional" MCM transition probability matrix [a] (adapted from Stokes (2006))

t \ $t+1$	1 to 29	30 to 49	50 to 99	100 to 199	200 to 499	500+	Exit
1 to 29	0.8051	0.0478	0.0314	0.0051	0.0014	0.0001	0.1092
30 to 49	0.0295	0.8312	0.0735	0.0020	0.0008	0.0002	0.0628
50 to 99	0.0000	0.0593	0.8696	0.0254	0.0006	0.0003	0.0449
100 to 199	0.0000	0.0000	0.0612	0.8975	0.0068	0.0000	0.0344
200 to 499	0.0000	0.0000	0.0000	0.0146	0.9853	0.0000	0.0000
500+	0.0000	0.0000	0.0000	0.0000	0.0000	0.9999	0.0000
Entry	0.2459	0.1712	0.0601	0.0003	0.0001	0.0000	0.5224

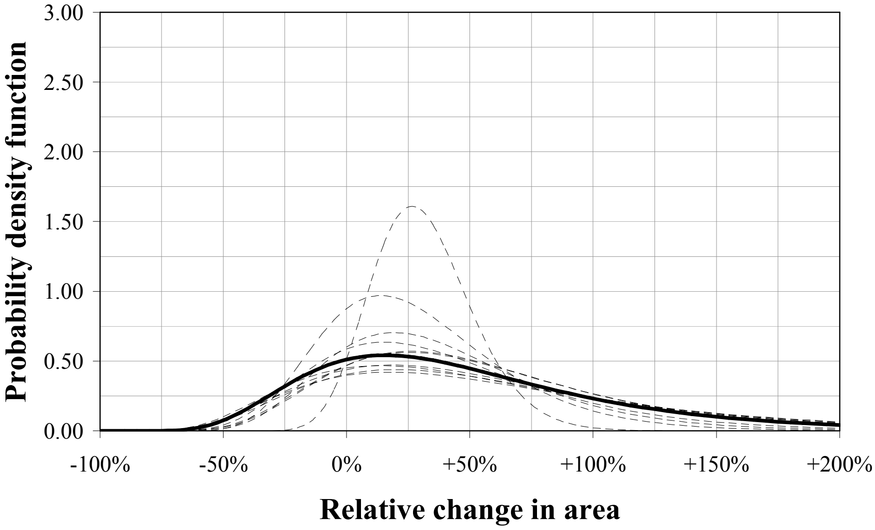
[a]*: intervals are for herd size in number of cow heads; the probabilities correspond to annual transitions and have been estimated over the 1980-2002 period; only the shaded cells significantly differ from zero.*

Table 3. Lognormal parameter estimations on cumulative FADN distributions for France [a]

Year	**Parameter μ_t**		**Parameter σ_t**		**Model**	N_t
	Coef.	**Std. Err.**	**Coef.**	**Std. Err.**	**Adj. R²**	
1980	3.3153	0.00021	0.6931	0.00045	0.9998	752,583
1981	3.3233	0.00023	0.6765	0.00047	0.9998	745,775
1982	3.3668	0.00025	0.6947	0.00051	0.9997	660,105
1983	3.3734	0.00030	0.6734	0.00062	0.9996	659,024
1984	3.3605	0.00024	0.6870	0.00054	0.9997	660,644
1985	3.4064	0.00024	0.6953	0.00057	0.9996	633,959
1986	3.4146	0.00025	0.6879	0.00060	0.9996	634,388
1987	3.4424	0.00030	0.7153	0.00068	0.9995	584,772
1988	3.5089	0.00034	0.7461	0.00082	0.9994	559,420
1989	3.5239	0.00040	0.7436	0.00087	0.9993	563,657
1990	3.5900	0.00044	0.7607	0.00099	0.9991	521,644
1991	3.5965	0.00045	0.7662	0.00100	0.9991	526,123
1992	3.6118	0.00044	0.7702	0.00096	0.9992	526,521
1993	3.7226	0.00048	0.8215	0.00109	0.9991	461,250
1994	3.7329	0.00050	0.8218	0.00106	0.9992	461,241
1995	3.8013	0.00049	0.8338	0.00108	0.9992	428,844
1996	3.8129	0.00054	0.8413	0.00116	0.9991	429,093
1997	3.8774	0.00053	0.8449	0.00116	0.9991	405,632
1998	3.8810	0.00055	0.8374	0.00116	0.9991	404,651
1999	3.8948	0.00058	0.8311	0.00119	0.9990	404,203
2000	3.9143	0.00067	0.8690	0.00141	0.9988	384,728
2001	3.9219	0.00082	0.8731	0.00163	0.9985	383,675
2002	3.9540	0.00082	0.8570	0.00159	0.9985	371,248
2003	3.9375	0.00080	0.8648	0.00157	0.9987	382,942
2004	3.9460	0.00083	0.8707	0.00166	0.9986	383,069
2005	4.0613	0.00085	0.8483	0.00180	0.9978	346,219

[a]*: all the reported parameters are significant at the 1% level*

Figure 1. Estimated probability density function of a relative change in area in a decade for France [a]



[a]: *dashed lines represent the estimated probability density function for each of the twelve studied 10-year transitions when considered individually; the bold line represent the estimated probability density function for the twelve 10-year transitions considered altogether* [the two outlying distributions correspond to the most recent studied transitions, 1994-04 and 1995-05].

Figure 2. Evolution of the estimated probability density function of a relative change in area for France

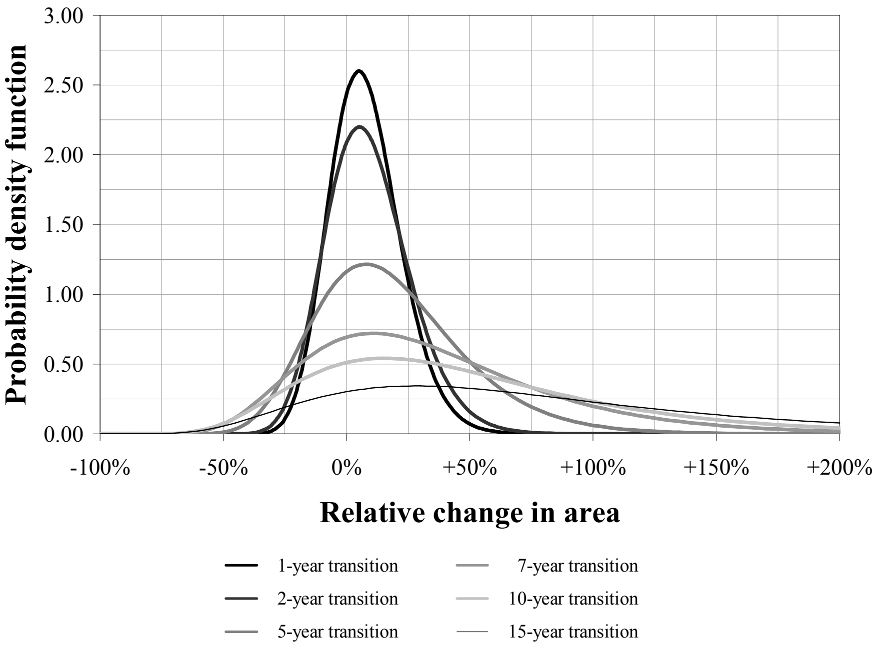


Table 4. Parameters of the estimated probability density functions depicted on Fig. 2 [a]

Transition	α_τ	η_τ	ν_τ	R^2	Nb. of transitions studied
1-year	0.8671	0.0482	0.1330	0.9979	11
2-year	0.8552	0.0503	0.1551	0.9972	11
5-year	0.7687	0.0755	0.2523	0.9939	15
7-year	0.6631	0.1026	0.3675	0.9953	12
10-year	0.5717	0.1442	0.4209	0.9926	12
15-year	0.4239	0.2477	0.4928	0.9961	9

[a]: *all the reported parameters are significant at the 1% level*

Table 5. An example of reconstructed transition probability matrix for the 10-year transition period for France [a]

t \ $t+10$	**< 9**	**10 to 24**	**25 to 49**	**50 to 99**	**100 to 149**	**150 to 199**	**> 200**	**Exit**
< 9	0.3406	0.3629	0.0180	0.0001	0.0000	0.0000	0.0000	0.2785
10 to 24	0.0229	0.3483	0.3028	0.0467	0.0008	0.0000	0.0000	0.2785
25 to 49	0.0001	0.0479	0.3131	0.3126	0.0418	0.0051	0.0009	0.2785
50 to 99	0.0000	0.0011	0.0554	0.3280	0.2142	0.0812	0.0416	0.2785
100 to 149	0.0000	0.0000	0.0020	0.0858	0.2069	0.1869	0.2400	0.2785
150 to 199	0.0000	0.0000	0.0001	0.0164	0.0914	0.1534	0.4602	0.2785
> 200	0.0000	0.0000	0.0000	0.0018	0.0198	0.0593	0.6407	0.2785

[a]*: with the starting year 2005 (t)and using the 10-year transition probability density function defined by the parameters shown in Table 4; intervals are for size in hectares of UAA.*