Movers and stayers in the farming sector: Accounting for unobserved heterogeneity in structural change

Laurent Piet, Legrand Dunold Fils Saint-Cyr

Movers and stayers in the farming sector: Accounting for unobserved heterogeneity in structural change

Legrand D. F. SAINT-CYR[*,†], **Laurent PIET**[‡]

[*] *AGROCAMPUS OUEST, UMR1302 SMART, F-35000 Rennes, France*
[‡] *INRA, UMR1302 SMART, F-35000 Rennes, France*

Abstract

The Markov chain model (MCM) has become a popular tool in the agricultural economics literature to describe how farms experience structural change and to study the impact of various drivers of this process, including public support. Even though some studies have accounted for heterogeneity across farms by letting transition probabilities depend on covariates depicting farms/farmers' characteristics, only observed heterogeneity has been considered. Assuming that structural change may also relate to unobserved individual farms' characteristics, we applied a restricted mixed Markov chain model (M-MCM), namely the mover-stayer model (MSM), in the agricultural context to relax the assumption of homogeneity in the transition process which grounds the usual MCM. We consider a mixture of two types of farms, the 'stayers' who always remain in their initial size category, and the 'movers' who follow a first-order Markovian process. An empirical application to a panel of commercial French farms over 2000-2013 shows that the MSM is a better modeling framework to recover the underlying transition probability matrix as well as to perform long-run farm size distribution forecasts.

Keywords:
Farm, Unobserved heterogeneity, Markov chain, Mover-stayer model, EM algorithm

JEL classification: Q12, C15, D92

[†]Corresponding author: AGROCAMPUS OUEST, UMR1302 SMART, 4 allée Adolphe Bobierre, CS 61103, F-35011 Rennes cedex; lsaintcy@agrocampus-ouest.fr.

Acknowledgments: Legrand D. F. Saint-Cyr benefits from a research grant from Crédit Agricole en Bretagne in the framework of the chair "Enterprises and Agricultural Economics" created in partnership with Agrocampus-Ouest.

1. Introduction

The farming sector has faced important structural changes over the last decades. In most developed countries, particularly in Western Europe and United States, the total number of farms has been decreased significantly and their average size increasing continually, implying some changes in the distribution of farms by category of sizes. Such structure changes in the farming sector may have important consequences for equity within agriculture (regarding income distribution and competitiveness among farms), for productivity and efficiency of farming as well as on the demand for government services and infrastructure and the well-being of local communities (Weiss, 1999). Therefore, structural change has been the subject of considerable interest among agricultural economists and policy makers, in particular to understand the mechanisms underlying these changes in order to identify key drivers that influence the observed trends and to generate prospective scenarios either to reverse the situation or to draw appropriate support programs.

As Zimmermann et al. (2009) show, it has become quite common in the agricultural economic literature to study the way farms experience structural change thanks to the so-called Markov chain model (MCM). Basically, this model states that, as the size of farms changes according to some stochastic process, farms move from one size category to another over time. Methodologically, most of these studies have used 'aggregate' data, that is, cross-sectional observations of the distribution of a farm population into a finite number of size categories: such data are most often easier to obtain than individual-level data, and Lee et al. (1965) and Lee et al. (1977) have shown that robustly estimating a MCM from such aggregate data is possible. Since then, because estimating a MCM may well be an ill-posed problem as the number of parameters to be estimated is often larger than the number of observations (Karantininis, 2002), much effort has been dedicated to developing efficient ways to parameterize and estimate these models, ranging from a discrete multinomial logit formulation (MacRae, 1977; Zepeda, 1995), the maximization of a generalized cross-entropy model with instrumental variables (Karantininis, 2002; Huettel and Jongeneel, 2011; Zimmermann and Heckelei, 2012), a continuous re-parameterization (Piet, 2011), to the use of Bayesian inference (Storm et al., 2011).

Empirically, MCMs have been first used within a stationary and homogeneous approach, assuming that transition probabilities are invariant over time and that all agents in the population change categories according to the same unique stochastic process. Despite improvements in the specification and the estimation method of this basic model, the resulting estimated parameters generally lead to erroneous farm size distributions forecasts (Hallberg, 1969; Stavins and Stanton, 1980) because of this homogeneity assumption (McFarland, 1970). Since then, several studies have been therefore devoted to improve the Markov chain modeling framework. Two approaches have been particularly investigated; first, assuming that transition probabilities of farms may vary over time, non-stationary MCMs have been developed in order to investigate the effects of time-varying variables on farm structural change, including agricultural policies (see Zimmermann et al. (2009)); second, assuming that the transition process may be different according to some farms/farmers' characteristics (regional location, type of farms, legal status, age group, etc.), some studies have accounted for farms heterogeneity in modeling structural change.

In the agricultural economics literature, heterogeneity of farms has been mostly incorporated using MCMs either by letting the transition probabilities depend on a set of dummy variables (see Zimmermann and Heckelei (2012) for a recent example) or by fitting the usual MCM to sub-populations, partitioned ex-ante based on the some exoge-

nous variables (see Huettel and Jongeneel (2011) for example). To our knowledge, only observed heterogeneity has been considered in this studies, implying that all farms sharing same observed characteristics follow the same and unique stochastic process. However, as it has been found in some other strands of the economic literature, the factors driving the evolution of the structure of the farms at the individual level may also relate to unobserved farms/farmers' characteristics (Langeheine and Van de Pol (2002)). Therefore, as farm-level data become more widely available, we propose to use a more general modeling framework than the simple MCM, namely the mixed-MCM (M-MCM), which allows accounting for unobserved heterogeneity in the transition process. This extended modeling framework has been applied to study population mobility or structural change in some other strands of the economic literature. For example, applications have been made to study labor mobility (Blumen et al., 1955; Fougère and Kamionka, 2003), credit rating, income or firm size dynamics (Dutta et al., 2001; Frydman and Kadam, 2004; Frydman and Schuermann, 2008; Cipollini et al., 2012).

Since structural change in agriculture refers to a long-run process (it may take time for farms to make at least one transition), we assume that accounting for heterogeneity in the rate of movement of farms (unobserved heterogeneity) may allow recovering the data generating this process in a more efficient way than the simple MCM. Therefore, using the M-MCM should lead to estimate more accurately the transition probabilities of farms. This modeling framework should thus leads to better farm size distribution forecasts and to investigate more efficiently the effects of time-varying variables on farm structural change, including agricultural policies as well as individual farms/farmers' characteristics. As an illustration, we apply the simplest version of the M-MCM, the mover-stayer model (MSM), to estimate the transition probability matrices and to perform short-, medium- and long-run out-of sample forecasts of farm size distributions using an unbalanced panel of 14,298 commercial French farms observed over 2000-2013. The objective is to compare the performance of a such modeling framework with respect to the simple MCM: first, in predicting the transition probabilities of farms; then, in performing farm size distribution forecasts over time.

The originality of this article is threefold: first, we implemented a MSM to account for unobserved heterogeneity across farms in their transition process; second, we computed a new index to compare distance between transition probability matrices or between farm size distributions; third, we use bootstrap sampling method to insure robustness of our results. The article is structured as follows. First, we introduce how the traditional MCM can be generalized into the M-MCM and the how specific MSM is derived. Second, we develop the method used to estimate the MSM parameters and to assess the performance of the model. Third, reports our application to France, first describing the data used and then presenting the results. Finally, we conclude with some considerations on how to extend further the approach described here.

2. Modeling transition process using Markov chain framework

Consider a population of agents which is partitioned into a finite number J of categories or 'states of nature'. Assuming that agents move from one category to another during a certain period of time r according to a stochastic process leads to defining the number

$n_{j,t+r}$ of individuals in category j at time $t+r$ as given by:

$$n_{j,t+r} = \sum_{i=1}^{J} n_{i,t}\phi_{ij,t}^{(r)} \tag{1}$$

where $n_{i,t}$ is the number of individuals in category i at time t, and $\phi_{ij,t}^{(r)}$ is the probability of moving from category i to category j between t and $t+r$. As such, $\phi_{ij,t}^{(r)}$ is subject to the standard non-negativity and summing-up to unity constraints for probabilities:

$$\begin{array}{rl} \phi_{ij,t}^{(r)} & \geq 0, \quad \forall i,j,t \\ \sum_{j=1}^{J} \phi_{ij,t}^{(r)} & = 1, \quad \forall i,t. \end{array} \tag{2}$$

In the following, without loss of generality, we restrict our analysis to the stationary case where the r-step transition probability matrix (TPM), $\mathbb{P}_t^{(r)} = \{\phi_{ij,t}^{(r)}\}$, is time-invariant, *i.e.*, $\mathbb{P}_t^{(r)} = \mathbb{P}^{(r)}$ for all t. In matrix notation, equation (1) then rewrites:

$$\mathbf{N}_{t+r} = \mathbf{N}_t \times \mathbb{P}^{(r)} \tag{3}$$

where $\mathbf{N}_{t+r} = \{n_{j,t+r}\}$ and $\mathbf{N}_t = \{n_{j,t}\}$ are row vectors.

2.1. The simple Markov chain model (MCM)

The simple (stationary) MCM approach consists in approximating the r-step TPM ($\mathbb{P}^{(r)}$) by the 1-step transition matrix $\mathbb{P}^{(1)}$ raised to the power r. Under first-order Markov assumption (*i.e.*, the situation at any future period depends only on the situation at the preceding period) and using individual level data, Anderson and Goodman (1957) have shown that the true (observed) r-step transition probabilities ($\phi_{ij}^{(r)}$) which can be computed from a contingency table, are the maximum likelihood of the MCM parameters ($\mathbf{\Pi}^{(r)} = \{\pi_{ij}^{(r)}\}$) given by:

$$\hat{\pi}_{ij}^{(r)} = \frac{n_{ij}^{(r)}}{\sum_j n_{ij}^{(r)}} \tag{4}$$

where $n_{ij}^{(r)}$ is the total number of r-step transitions from category i to category j during the period of observation and $\sum_j n_{ij}^{(r)}$ the total number of r-step transitions out of category i. Thus, $\phi_{ij}^{(1)} = \hat{\pi}_{ij}^{(1)}$. Then, under population homogeneity and stationary assumptions, state $\hat{\mathbf{\Pi}}^{(r)} = (\hat{\mathbf{\Pi}})^r$ to approximate $\mathbb{P}^{(r)}$.

In doing so, the MCM approach assumes that the agents in the population are homogeneous, *i.e.*, they all move according to the same stochastic process described by $\hat{\mathbf{\Pi}}$. However, in general, $\hat{\mathbf{\Pi}}^{(r)}$ proves to be a poor estimate of $\mathbb{P}^{(r)}$ (Blumen et al., 1955; Spilerman, 1972). In particular, the main diagonal elements of $\hat{\mathbf{\Pi}}^{(r)}$ largely underestimate those of $\mathbb{P}^{(r)}$. This means that, in general, $\hat{\pi}_{ii}^{(r)} \ll \phi_{ii}^{(r)}$. In the farming context, this mean that the simple MCM tends to overestimate mobility of farm because of the homogeneity assumption.

2.2. Accounting for unobserved heterogeneity: the mixed Markov chain model (M-MCM)

One way to obtain a 1-step TPM which leads to a more consistent r-step estimate consists in relaxing the assumption of homogeneity in the transition process which underlies the

MCM approach. This leads to considering a mixture of Markov chains which may captures population heterogeneity for example in the rate of movement among states (Frydman, 2005).

More precisely, consider a population partitioned in a discrete number G of homogeneous types of agents instead of just one and each agent follows one of these types which describe elementary Markov processes, the general form of the M-MCM consists in decomposing the 1-step transition matrix as:

$$\mathbf{P} = \{p_{ij}\} = \sum_{g=1}^{G} \mathbf{S}_g \mathbf{M}_g \tag{5}$$

where $\mathbf{M}_g = \{m_{ij,g}\}$ is the TPM defining the 1-step Markov process followed by type-g agents, and $\mathbf{S}_g = \text{diag}(s_{i,g})$ is a diagonal matrix which gathers the shares of type-g agents in each category. Since every agent in the population has to belong to one and only one type g, the constraint that $\sum_{g=1}^{G} \mathbf{S}_g = \mathbf{I}$ must hold, where $\mathbf{I}$ is the $J \times J$ identity matrix.

Because we consider here only the stationary case, it is assumed that neither $\mathbf{M}_g$ nor $\mathbf{S}_g$ vary over time.

Then, the r-step TPM for any future time period r can be approximated as:

$$\mathbf{P}^{(r)} = \sum_{g=1}^{G} \mathbf{S}_g (\mathbf{M}_g)^r. \tag{6}$$

With the so-defined MCM and M-MCM modeling frameworks, it should be noted that: (i) the M-MCM reduces to the MCM if $G = 1$, that is, the homogeneity assumption holds and; (ii) the aggregate overall M-MCM process described by $\mathbf{P}^{(r)}$ may no longer be Markovian even if each agent type follows a specific Markov process.[1]

2.3. The Mover-Stayer model (MSM)

In this article, we stick to the simplest version of the M-MCM, namely the mover-stayer model (MSM) first proposed by Blumen et al. (1955). In this restricted approach, only two types of homogeneous agents are considered, those who always remain in the same category (the 'stayers') and those who follow a first-order Markovian process (the 'movers'). Formally, this leads rewriting equation (5) in a simpler form as:

$$\mathbf{P} = \mathbf{S} + (\mathbf{I} - \mathbf{S})\,\mathbf{M}. \tag{7}$$

With respect to the general formulation (5), this corresponds to setting $G = 2$ and defining $\mathbf{S}_1 = \mathbf{S}$ and $\mathbf{M}_1 = \mathbf{I}$ for stayers, and $\mathbf{S}_2 = (\mathbf{I} - \mathbf{S})$ and $\mathbf{M}_2 = \mathbf{M}$ for movers. According to the Frydman (2005)'s specification of M-MCM presented in Section A.1 in appendix, the mover-stayer model is equivalent to imposing the rate of movement for stayers to be zero. Thus, the overall population r-step TPM can be approximated as:

$$\mathbf{P}^{(r)} = \mathbf{S} + (\mathbf{I} - \mathbf{S})\,\mathbf{M}^r. \tag{8}$$

[1] According to equation (6), the situation at future periods not only depends on the situation at one or some previous periods but also depends on the initial agent distribution.

3. Estimating the mover-stayer model (MSM)

At first, Blumen et al. (1955) have used a simple calibration method base on the maximum likelihood to estimate the MSM parameters. Then, since Goodman (1961) has shown that Blumen et al. (1955) estimators are biased, alternative methods have been developed to obtain consistent ones using, for example, minimum chi-square (Morgan et al., 1983), maximum likelihood (Frydman, 1984; Frydman and Kadam, 2004) or Bayesian inference (Fougère and Kamionka, 2003). Frydman (2005) is the first who has developed a maximum likelihood estimation method for the general M-MCM from which can be easily derived estimators for the MSM. We report this strategy simplifying for the MSM, using our own notations introduced above.

3.1. The maximum likelihood under complete information

Within the MSM framework where only two types of agents are considered ('S' standing for stayers and 'M' for movers) and under complete information, that is, stayers and movers (identified by indicators $Y_{k,S}$ and $Y_{k,M} = 1 - Y_{k,S}$, respectively) are perfectly known, the log-likelihood of the MSM for the whole population writes:

$$\log L_{MSM} = \sum_{k=1}^{n} Y_{k,S} \log l_{k,S} + \sum_{k=1}^{n} (1 - Y_{k,S}) \log l_{k,M} \tag{9}$$

where the first sum on the right hand-side is the log-likelihood of stayers and the second one is the log-likelihood of movers. Conditional on knowing that k was initially in category i, the likelihood that k is a stayer, $l_{k,S}$, is equivalent to $s_{i_k,S}$, the proportion of agents who never move out of category i during the whole period of observation (see section A.2 in appendix). And, the likelihood that agent k is a mover writes (Frydman and Kadam, 2004):

$$l_{k,M} = s_{i_k,M} \prod_{i \neq j} (m_{ij})^{n_{ij,k}} \prod_{i} (m_{ii,M})^{n_{ii,k}} \tag{10}$$

where $s_{i_k,M} = 1 - s_{i_k,S}$ is the share of movers initially in category i_k, $n_{ij,k}$ is the number of transitions from category i to category j made by agent k, $n_{ii,k}$ is the total times that agent k stay in category i. Therefore, on the right hand-side of equation (10), the first product is the probability to move out of category i while the second one is the probability to stay in this category from one period to the next even if agent k is a mover.

Substituting $l_{k,S}$ and $l_{k,M}$ in equation (9), the log-likelihood of the MSM for the whole population rewrites:

$$\log L = \sum_{i} b_i \log(1 - s_i) + \sum_{i} b_{i,S} \log[s_i/(1 - s_i)] + \sum_{i \neq j} n_{ij} \log(m_{ij}) + \sum_{i} n_{ii,M} \log(m_{ii}) \tag{11}$$

where b_i and $b_{i,S}$ are the total number of agents and the total number of stayers, respectively, who were initially in category i, s_i is category-i share of stayers, $n_{ij} = \sum_{k=1}^{n} n_{ij,k}$ is the total number of transitions from category i to category j, m_{ij} and m_{ii} are the elements of the generator matrix ($\mathbf{M}$) of movers and $n_{ii,M}$ is the total number of times that movers stay in category i.

Then, maximizing equation (11) with respect to the unknown parameters s_i and m_{ij} leads to the optimal values of the MSM parameters. Therefore, solving $\partial \log L_{MSM}/\partial s_i = 0$ gives the optimal share of stayers in each category i:

$$\hat{s}_i = \frac{b_{i,S}}{b_i}. \tag{12}$$

Likewise, solving $\partial \log L_{MSM}/\partial m_{ij} = 0$, gives:

$$m_{ij} = \frac{n_{ij} m_{ii}}{n_{ii,M}} \quad \forall i \neq j \tag{13}$$

Then, setting $\sum_{i \neq j}^{J} m_{ij} = 1 - m_{ii}$, the maximum likelihood of m_{ii} (*i.e.*, the probability for movers to remain in their starting category i) is obtained by:

$$\hat{m}_{ii} = \frac{n_{ii,M}}{n_i + n_{ii,M}} \tag{14}$$

where n_i is the total number of transitions out of category i, and $n_{ii,M} = n_{ii} - n_{ii,S}$ with n_{ii} and $n_{ii,S}$ are the total number of times that all agents and stayers remain in state i, respectively. Finally, substituting equation (14) into (13), the maximum likelihood of m_{ij} (*i.e.*, the probability for movers to make a transition from the category i to the category j) is given by:

$$\hat{m}_{ij} = \frac{n_{ij}}{n_i}(1 - \hat{m}_{ii}) \quad \forall i \neq j, \quad i, j = 1, \cdots, J \tag{15}$$

3.2. The expectation-maximization (EM) algorithm under incomplete information

Some authors have shown that equation (11) is actually difficult to use directly because it is unlikely that one knows beforehand which agents are stayers and which are movers (Frydman, 1984; Fuchs and Greenhouse, 1988; Swensen, 1996). Indeed, because the transition process is assumed to be a stochastic process, even movers may remain for a long time period in their initial category before moving, so that they may not appear as movers on the observed period. Alternatively, Fuchs and Greenhouse (1988) and van de Pol and Langeheine (1989) suggest that the MSM parameters can be estimated using the EM algorithm developed by Dempster et al. (1977). Concretely, the EM algorithm allows estimating the probability to be a stayer for each agent in the population given its initial category.

Following Frydman and Kadam (2004), the four steps of the EM algorithm are defined in the case of the MSM as follows:

(i) Initialization: Arbitrarily choose initial values s_i^0 for the share of stayers and m_{ii}^0 for the main diagonal entries of the generator matrix ($\mathbf{M}$) of movers.

(ii) Expectation: At iteration p of the algorithm, compute the probability of observing agent k as generated by a stayer, $E^p(Y_{k,S})$. If at least one transition is observed for agent k, then set $E^p(Y_{k,S}) = 0$, otherwise set it to:

$$E^p(Y_{k,S}) = \frac{s_i^p}{s_i^p + (1 - s_i^p)(m_{ii}^p)^{n_{ii,k}}}. \tag{16.i}$$

Using the resulting $E^p(Y_{k,S})$, then compute:

- the expected value of the total number of stayers in category i, $E^p(b_{i,S})$, as:

$$E^p(b_{i,S}) = \sum_{k=1}^{n} E^p(Y_{k,S}) \tag{16.ii}$$

- the expected value of the total number of times that stayers remain in category i, $E^p(n_{ii,S})$, as:

$$E^p(n_{ii,S}) = \sum_{k=1}^{n} E^p(Y_{k,S}) n_{ii,k} \tag{16.iii}$$

- and the expected value of the total number of times that movers remain in category i, $E^p(n_{ii,M})$, as:

$$E^p(n_{ii,M}) = n_{ii} - E^p(n_{ii,S}) \tag{16.iv}$$

(iii) Maximization: Update s_i^p and m_{ii}^p as follows:

$$s_i^{p+1} = \frac{E^p(b_{i,S})}{b_i} \quad \text{and} \quad m_{ii}^{p+1} = \frac{E^p(n_{ii,M})}{n_i + E^p(n_{ii,M})}. \tag{16.v}$$

(iv) Iteration: Return to step (ii) using s_i^{p+1} and m_{ii}^{p+1} and iterate until convergence.

When convergence is reached, $\hat{s}_i^*$ and $\hat{m}_{ii}^*$ so obtained are considered as the optimal estimators. Then, $\hat{m}_{ij}^*$ derives from $\hat{m}_{ii}^*$ as in equation (15).

According to Frydman (2005), the standard errors for the MSM parameters were computed directly from the EM equations using the method proposed by Louis (1982) presented in the section A.3 in appendix. Then, the standard errors on the 1-year TPM were derived using the Delta method according to equation (7). As equation (8) is relatively more complicated, bootstrap sampling method was used to compute standard deviations on r-step TPMs (Efron, 1979, 1981).

4. Assessing the MSM performance

To assess the performance of the MSM with respect to the simple MCM, two types of measure were used.

4.1. Likelihood test ratio

To test the goodness-of-fit of the MSM with respect to a simple MCM, a likelihood test ratio was performed. This statistical test allows comparing the performance of the two models in recovering the data generating the process under study. According to general Frydman and Kadam (2004), the likelihood ratio statistic for the MSM is given by:

$$\Gamma = \frac{L_{MCM}(\hat{\mathbf{\Pi}})}{L_{MSM}(\hat{\mathbf{S}}, \hat{\mathbf{M}})} \tag{17}$$

where L_{MCM} and L_{MSM} are the estimated maximum likelihood for MCM and MSM, respectively. Theoretically, the asymptotic distribution of $-2\log\Gamma$, under H_0, is chi-square with $(G-1) \times J$ degrees of freedom. In the case of the MSM, the likelihood ratio tests the hypothesis that the process involves according to a simple MCM ($H_0 : \mathbf{S} = 0$) against the hypothesis that it is a mixture of movers and stayers ($H_1 : \mathbf{S} \neq 0$). The log-likelihood for both models can be derived from equation (11), where $s_i = 0$ and $n_{ii,M} = n_{ii}$ ($\forall i \in J$) for the simple MCM.

4.2. TPMs and distributions distance

In order to test the usefulness of the MSM and to compare its merits relative to the MCM, the parameters from both models were estimated. The estimated parameters were then used to compute r-step TPMs (where $\hat{\mathbf{\Pi}}^{(r)} = (\hat{\mathbf{\Pi}})^r$ for the MCM and $\hat{\mathbf{P}}^{*(r)} = \hat{\mathbf{S}} + (\mathbf{I} - \hat{\mathbf{S}})(\hat{\mathbf{M}}^*)^r$ for the MSM) which were used to perform out-of-sample short-, medium- and long-run forecasts of farm distributions across categories according to equation (1). The resulting TPMs and distributions from both models were compared to the observed ones based on the average of marginal errors (AME) given by:

$$AME = \frac{1}{W} \sum_{i,j} \sqrt{\left(\frac{\hat{\phi}_{ij}^{(r)} - \phi_{ij}^{(r)}}{\phi_{ij}^{(r)}} \right)^2} \tag{18}$$

where $W = J^2$ is the total number of elements; $\hat{\phi}_{ij}^{(r)}$ and $\phi_{ij}^{(r)}$ are predicted and observed values, respectively. $\hat{\phi}_{ij}^{(r)}$ are the elements of the r-step TPM predicted using either the MCM ($\hat{\phi}_{ij}^{(r)} = \hat{\pi}_{ij}^{(r)}$) or the MSM ($\hat{\phi}_{ij}^{(r)} = \hat{p}_{ij}^{*(r)}$) estimates as described above, respectively, and $\phi_{ij}^{(r)}$ are elements of the observed r-step TPM ($\mathbb{P}^{(r)}$) given by (Anderson and Goodman, 1957):

$$\phi_{ij}^{(r)} = \frac{n_{ij}^{(r)}}{\sum_j n_{ij}^{(r)}}, \tag{19}$$

where $n_{ij}^{(r)}$ is the total number of r-step transitions from category i to category j and $\sum_j n_{ij}^{(r)}$ the total number of r-step transitions from category i. When compared distributions the AMEs is computed over the resulting row vector, that is, $W = J$

Contrary to some indexes of dissimilarity (see Jafry and Schuermann (2004)) or a residual matrix (see Frydman et al. (1985) for example), the AME provides a global distance between the predicted TPM or the distribution across categories and the observed ones. The AME can be interpreted as the average percentage of deviations on predicting the observed overall population TPM or the distribution across categories. The higher is the AME, the more the resulting TPM or distribution is different from the observed ones. Therefore, the best model is the one which gives the lowest AME.

5. Empirical application

5.1. Data used

For our empirical application, we used data from the "Réseau d'Information Comptable Agricole" (RICA) for France. The RICA (or Farm Accountancy Data Network (FADN) in English) is defined at the European Union level and consists of an annual survey carried out by the Member States of the European Union. In France, the RICA focuses on 'commercial' farms, that is, farms whose standard output (SO) is greater than or equal to 25,000 Euros. The information collected on farms refer to physical and structural characteristics, on the one hand, economic and financial characteristics, on the other hand. To comply with French accounting standards, the specific questionnaire defined at the European Union level has been adapted at national level and referenced as RICA France. The RICA France is produced and disseminated by the statistical and foresight service of the French ministry for agriculture. Individual farm level data were available

from 2000 to 2013 for the full sample surveyed, *i.e.*, around 7,000 farms each year. To each farm in the dataset is assigned has a weighted factor which allows insuring the representativeness of the sample. The weighted factor also allows extrapolating the total number of commercial French farms by years based on the Farm Structure Survey (FSS) organised by Eurostat.[2]

As we considered all farms in the sample whatever their type of farming, we chose to concentrate on size as defined from an economic perspective. In accordance with the European regulation (CE) Nº1242/2008, RICA France farms are classified into 14 economic size (ES) categories, evaluated in terms of total SO expressed in Euros.[3] This corresponds to ES category 6 and above. To simplify and because some categories of farms are less represented in the sample than others, we aggregated the 9 size categories available in the RICA France sample into 5: less than 50,000 Euros of SO and below (ES6); between 50,000 and 100,000 Euros of SO (ES7); between 100,000 and 150,000 Euros of SO and between 150,000 and 250,000 Euros of SO (both in ES8); more than 250,000 Euros of SO (ES9 and above).

Since the RICA France database is a rotating panel, farms which enter (respectively, leave) the sample a given year cannot be considered as representing actual entries into (exits from) the agricultural sector (around 10% of the French FADN sample is renewed each year). Because we cannot identify entries into nor exits from the sector, we cannot thus work directly on farm numbers. Alternatively, we chose to work on size change of on-going farms, *i.e.*, without considering entries nor exits. We concentrated on farm sizes in terms of shares of farms by size categories, which is another way to analyze farm size distribution. We used the weighted factors to compute the actual farm size distributions for the whole population of commercial French farms for each available year. Table 1 and Figure 1 present the evolution over the whole studied period of sample farm numbers by ES categories and average ES in thousand of Euros of SO for the studied panel.

Furthermore, the average economic size of farms has increased over the period of observation as well as the standard deviations which is a common observed feature for the whole population of farms in France (see Butault and Delame (2005) and Agreste Primeur (2011)). Table 1 shows that the observed shares of farms by category of sizes for the overall population as well as the average ES are different from for the RICA France sample ones. This is because farms in smaller categories have higher individual weights than those in larger ones. However, the same trend is observed, that is, the shares of farms for smaller categories (less than 100,000 Euros of SO) have decreased from 2000 to 2013 while the ones for larger categories (150,000 Euros of SO and above) have increased. Likewise, the average ES has also decreased while the standard deviation has increase during the same period. Figure 1 shows that the shares of farms in the category of sizes less than 100,000 Euros of SO tend to decrease from 2000 to 2013 while the ones for categories with more than 150,000 Euros of SO tend to increase.

Despite the limitations mentioned above, the RICA France farm sample provides an opportunity to use the approach developed in this article because: first, individual farm history is available; second, some farms remain a relative long time period in the sample. In order to observe at least one transition for each agent, we kept only farms

[2]To learn more about RICA France, see `http://www.agreste.agriculture.gouv.fr/`. To learn more about FADN in general, see `http://ec.europa.eu/agriculture/rica/index.cfm`.

[3]SO is being used as the measure of economic size since 2010. Before this date, economic size was measured in terms of standard gross margin (SGM). However, SO calculations have been retropolated for 2000 to 2013, allowing for consistent time series analysis (European Commission, 2010).

Table 1

Distribution by economic size (ES) class and average ES for the studied sample[a]

Years	Number of farms by ES class					Total	Average ES	
	0-50	50-100	100-150	150-250	≥ 250		(std. dev.)	
2000	790	2,234	1,629	1,762	1,342	7,757	168.88	(179.01)
	87,924	129,691	59,857	67,367	41,457	386,296	134.46	(151.72)
2001	746	2,231	1,625	1,817	1,382	7,801	170.98	(180.88)
	84,442	123,900	57,583	67,741	41,890	375,556	136.75	(155.04)
2002	713	2,128	1,663	1,818	1,443	7,765	177.57	(198.12)
	81,228	118,571	58,104	65,448	42,344	365,695	140.99	(184.50)
2003	690	1,975	1,562	1,693	1,393	7,313	176.27	(193.55)
	78,249	113,662	56,961	64,946	42,859	356,677	141.08	(176.08)
2004	707	1,940	1,538	1,707	1,437	7,329	177.67	(188.47)
	75,481	109,118	56,118	64,252	43,419	348,388	142.63	(169.30)
2005	741	1,927	1,516	1,711	1,467	7,362	178.03	(181.95)
	72,896	104,906	54,811	64,112	44,007	340,732	144.55	(161.46)
2006	756	1,922	1,488	1,688	1,491	7,345	181.21	(209.21)
	70,516	101,035	54,202	63,443	44,740	333,936	146.99	(171.49)
2007	774	1,845	1,552	1,694	1,511	7,376	182.27	(191.10)
	68,286	97,435	54,032	62,390	45,491	327,634	150.08	(172.33)
2008	780	1,866	1,511	1,721	1,587	7,465	185.49	(200.25)
	66,201	94,098	52,412	62,889	46,338	321,938	153.47	(185.00)
2009	778	1,816	1,517	1,734	1,624	7,469	188.43	(205.95)
	64,243	90,970	51,137	63,151	47,278	316,779	156.14	(186.03)
2010	652	1,885	1,537	1,770	1,608	7,452	190.53	(199.03)
	62,429	88,104	51,320	62,062	48,267	312,182	157.88	(174.96)
2011	638	1,856	1,468	1,791	1,658	7,411	194.89	(207.58)
	60,743	85,444	49,285	63,292	49,381	308,145	162.11	(189.23)
2012	651	1,797	1,396	1,794	1,679	7,317	200.28	(249.45)
	59,152	82,943	47,911	63,953	50,626	304,585	166.69	(227.41)
2013	658	1,769	1,361	1,804	1,701	7,293	202.41	(240.15)
	57,668	80,638	46,821	64,414	51,939	301,480	169.49	(225.53)

[a] ES in 1000 Euros of standard output

Notes: For each year, the first row is the total farm numbers by ES class and the average ES observed in the RICA France full sample; the second row is the total number of farms by ES class and the average ES extrapolated using the weighted factor.

Source: Agreste, RICA France 2000-2013 – authors' calculations

present during at least two consecutive years in the database. Our unbalanced panel thus counted 14,831 individual farms, out of the 17,285 farms in the original database. This led to observe 85,196 individual 1-year transitions from 2000 to 2013 (100,027 observations for the selected panel).

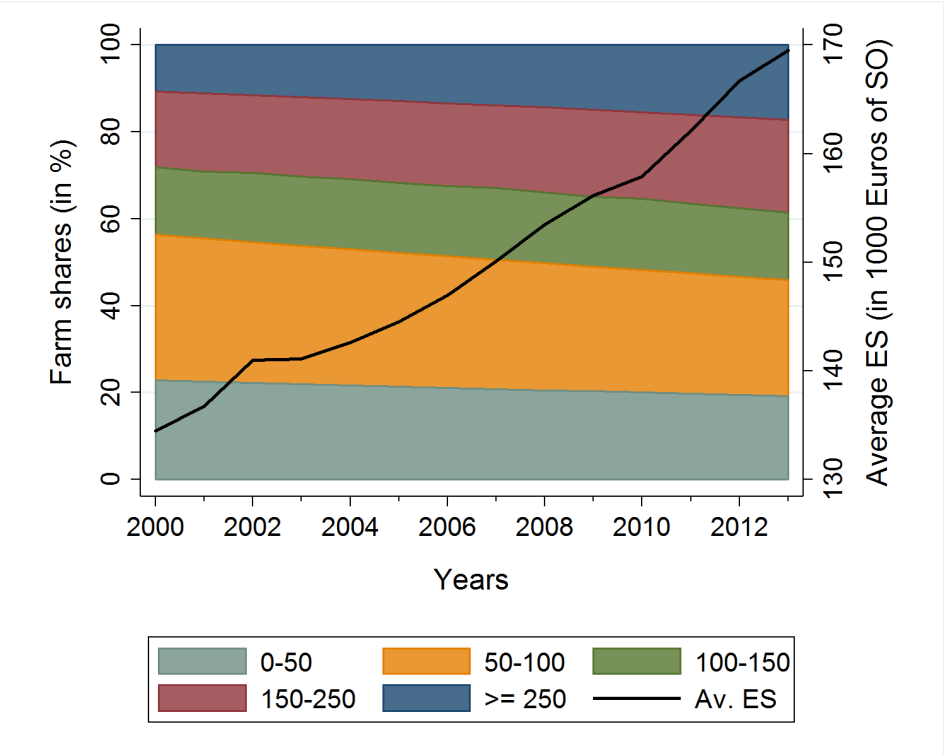

Figure 1: Extrapolated farm shares by category of sizes and average Economic Size (ES)

Source: Agreste, RICA France 2000-2013 – authors' calculations

Before proceeding with the results of our analysis, it should be noted that because we chose to work with a subset of the full sample, the transition probabilities reported in the next section should be viewed as size change probabilities *conditional* on having been observed over a specific number of consecutive years during the whole period under study, and should not be considered as representative for the whole population of commercial French farms.

5.2. Results

We estimated the MSM and the MCM on the RICA France data. As the goal of this study is to develop a robust modeling approach to describe the observed transition process of farms which may lead to better farm size distributions forecasts, the quality of the model was evaluated on out-of-sample forecasts. Therefore, we split the RICA France sample into two parts: (i) observations from 2000 to 2010 were used to estimate parameters for both models; (ii) the resulting parameters were then used to forecast farm size distributions in 2011, 2012 and 2013 in order to assess the out-of-sample forecasting power of each model on a short-, medium- and long-run perspectives. We assumed that 11 years is long enough to observe at least one transition for farms therefore to provide consistent parameter estimation for both models. We used three different years for out-of-sample forecasts in order to insure robustness of the results.

Table 2

Farm numbers and total observations from 2000 to 2010, by subsamples

	Subsamples									
	1	2	3	4	5	6	7	8	9	10
Farms	13,123	11,291	9,322	7,680	6,257	5,107	4,112	3,448	2,801	2,170
Observations	78,434	74,770	68,863	62,295	55,180	48,280	41,315	36,003	30,180	23,870

Notes: subsample 1 corresponds to the subset of farms remaining present in the database at least 2 consecutive years; subsample 2 to those remaining at least 3 consecutive years; and so forth.

Source: Agreste, RICA France 2000-2010 – authors' calculations

5.2.1. In-sample estimation

For the estimation phase, ten subsamples could be constructed according to the minimum number of consecutive years a farm remains present in the database, from two to eleven. Thus, subsample 1 corresponds to the subset of farms remaining present in the database at least 2 consecutive years; subsample 2 to those remaining at least 3 consecutive years; and so forth . Table 2 shows that from subsample 1 to subsample 10 we lost around 83% of the total number of farms and about 70% of the total number of observation meaning that several farms remain relative short time period in the sample. Then, using each subsample the MCM and MSM parameters were estimated and used to predict the 10-year (or 10-step) TPMs (where $\hat{\mathbf{\Pi}}^{(10)} = (\hat{\mathbf{\Pi}})^{10}$ for the MCM and $\hat{\mathbf{P}}^{*(10)} = \hat{\mathbf{S}} + (\mathbf{I} - \hat{\mathbf{S}})(\hat{\mathbf{M}}^{*})^{10}$ for the MSM). The resulting 10-year TPMs were compared to the observed one (*i.e*, the TPM describing movements of farms across ES classes from 2000 to 2010) based on the average of marginal errors (AME) as previously described .

In the following, because it would take too much space to present the results obtained from all subsamples, estimated matrices and detailed results are reported for subsample 10 only, that is, when considering the subset of farms which remain at least 11 consecutive years in the database which is a balanced panel for 2,170 individual farms, leading to 23,870 observations over the 11 years (see table 2). However, it should be noted that the derived matrices and results, hence conclusions, remain very similar to those reported here when considering any of the other subsamples.

The corresponding observed 1-year TPM computed from subsample 10 (which is the maximum likelihood of the MCM) is reported in table 3. As it usually found in the literature, we observe that this TPM is strongly diagonal, meaning that its main diagonal elements exhibit by far the largest values and that probabilities rapidly decrease as we move away from the main diagonal. This means that, overall, farms are more likely to remain in their initial size category from one year to the next (see Piet (2011)). This does not mean no size change at all but, at least, no sufficient change to move to another category as we defined them. Table 3 also shows that the probability for commercial French farms to remain in their starting ES class from one year to the next is lower for farms in intermediate categories. This means that farms from these categories are more likely to change category of sizes which is also a common feature in the agricultural economics literature (see Piet (2011); Ben Arfa et al. (2014) for example).

Table 3

Computed observed 1-year TPM ($\mathbb{P}^{(1)}$) (subsample 10)

		ES class					Total
		0-50	50-100	100-150	150-250	≥ 250	transitions
ES class	0-50	0.917	0.079	0.002	0.002	0.001	1,651
	50-100	0.030	0.898	0.065	0.005	0.002	5,083
	100-150	0.002	0.062	0.854	0.080	0.002	4,514
	150-250	0.001	0.004	0.054	0.886	0.055	5,685
	≥ 250	0.000	0.001	0.003	0.048	0.948	4,767

Notes: the observed 1-year TPM computed from the contingency table is the the maximum likelihood of the MCM; the computed log-likelihood is $\log L_{MCM} = -8,689.36$.

Source: Agreste, RICA France 2000-2010 – authors' calculations

In order to estimate the stayer shares $\mathbf{S}$, and the generator matrix of movers $\mathbf{M}$, which both define the MSM, we implemented the EM algorithm estimation method as previously developed. Table 4 reports the corresponding shares of stayers by size category and generator matrix of movers. Firstly, the estimated stayer shares show that the probability to be a stayer is closer or above 0.30 whatever the category considered. In other words, for every category, 30% of the farms are likely to remain in their initial category; this share even goes beyond 60% for farms over 250,000 Euros of SO and is almost 50% for farms below 50,000 Euros of SO. meaning that farms of these categories are more likely to remained in their starting category than those in the intermediate categories. This result could be explained by the fact that farms below 50,000 Euros of SO may face some economic constraints while farms above 250,000 Euros of SO may reach an optimal economic size. Secondly, the generator matrix reveals that, conditional on having been observed eleven times, movers remain between more than four years and a half (for intermediate ES class) and almost eight years (for farms above 250,000 Euros of SO) in their initial category before leaving it, recalling that the average time spent by movers in a particular category is given by $1/(1-m_{ii})$ (see section A.1 in appendix). This result thus confirms that farms which remain in a particular category for a long time, even during the whole observation period, are not necessarily stayers. Altogether, these two results are in agreement with the strong diagonality found for the observed 1-year TPM (see Table 3). This is reflected in 1-year MSM TPM ($\hat{\mathbf{P}}^{*}$) which is also strongly diagonal.

Table 4

Estimated stayer shares ($\hat{s}_i^*$), mover generator matrix ($\hat{\mathbf{M}}^*$) and overall population 1-year TPM ($\hat{\mathbf{P}}^*$) (subsample 10)

		Stayers shares	**Movers generator matrix** ($\hat{\mathbf{M}}^*$)					**Overall population TPM** ($\hat{\mathbf{P}}^*$)				
		($\hat{s}_i^*$)	0-50	50-100	100-150	150-250	≥ 250	0-50	50-100	100-150	150-250	≥ 250
ES class	0-50	**0.494**	**0.837**	**0.154**	**0.004**	**0.004**	**0.001**	**0.917**	**0.078**	**0.002**	**0.002**	**0.001**
		(0.036)	(0.041)	(0.012)	(0.002)	(0.002)	(0.001)	(0.019)	(0.011)	(0.001)	(0.001)	(.)
	50-100	**0.422**	**0.055**	**0.815**	**0.118**	**0.009**	**0.003**	**0.032**	**0.893**	**0.068**	**0.005**	**0.002**
		(0.021)	(0.004)	(0.022)	(0.006)	(0.002)	(0.001)	(0.003)	(0.012)	(0.005)	(0.001)	(0.001)
	100-150	**0.291**	**0.002**	**0.089**	**0.793**	**0.113**	**0.003**	**0.002**	**0.062**	**0.854**	**0.080**	**0.002**
		(0.016)	(0.001)	(0.005)	(0.020)	(0.005)	(0.001)	(0.001)	(0.004)	(0.014)	(0.005)	(0.001)
	150-250	**0.371**	**0.002**	**0.007**	**0.087**	**0.816**	**0.088**	**0.001**	**0.005**	**0.055**	**0.884**	**0.055**
		(0.017)	(0.001)	(0.001)	(0.004)	(0.020)	(0.004)	(0.000)	(0.001)	(0.004)	(0.012)	(0.004)
	≥ 250	**0.650**	**0.001**	**0.003**	**0.007**	**0.114**	**0.875**	**0.000**	**0.001**	**0.003**	**0.040**	**0.956**
		(0.021)	(0.001)	(0.001)	(0.002)	(0.007)	(0.027)	(0.000)	(0.001)	(0.001)	(0.005)	(0.009)

Notes: estimated parameters in bold font; bootstrap standard errors in parenthesis (1000 replications); the computed log-likelihood is $\log L_{MSM} = -7,783.90$.

Source: Agreste, RICA France 2000-2010 – authors' calculations

The 1-year MSM TPM for the overall population, $\hat{\mathbf{P}}^*$, derives from the estimated parameters $\hat{\mathbf{S}}^* = \{\hat{s}_i^*\}$ and $\hat{\mathbf{M}}^*$ according to equation (7). At first glance, the resulting matrix, which is reported in Table 4, may look different from the observed one reported in Table 3, which also defines the MCM. However, a close examination of the standard errors associated with the elements of $\hat{\mathbf{P}}^*$ reveals that every observed probabilities fall within the 95% confidence interval of their estimated counterpart p_{ij}. In other words, this means that the MSM leads to estimate a matrix which is not statistically different from the true (observed) underlying transition process. Moreover, the fit of the models measured by the likelihood ratio shows that the MSM better fit the data than the MCM. The likelihood ratio test is $-2\log\Gamma = 1{,}810.99$ which is highly significant (the critical value is $\chi^2_{0.001}(5) = 20.52$). This means that the MSM allows recovering the data generating the transition process of commercial French farms in a more efficient way than the simple MCM. Therefore, the MSM should lead to better approximation of farm transition probabilities of farms over time,, that is, the r-step TPM.

Table 5 reports both the 10-year MCM TPM, $\hat{\mathbf{\Pi}}^{(10)} = (\hat{\mathbf{\Pi}})^{10}$, and the 10-year MSM TPM, $\hat{\mathbf{P}}^{*(10)}$ obtained from $\hat{\mathbf{S}}^*$, $\hat{\mathbf{M}}^*$ given equation (8). Table 5 shows that the TPMs predicted by both models are obviously different from the observed one. Most of the predicted transition probabilities fall out of the 5%-95% percentile confidence interval for both models. Overall, the MSM matrix however appears to be a better approximation than the MCM matrix when compared to the actually observed 10-year TPM ($\mathbb{P}^{(10)}$). In particular, we find as expected that $\hat{\pi}_{ii}^{(10)} \ll \phi_{ii}^{(10)}$ while $\hat{p}_{ii}^{*(10)}$ is much closer to $\phi_{ii}^{(10)}$. This means that the MCM tends to largely overestimate mobility of farms particularly on the long-run, with respect to the MSM. Furthermore, a close examination of all the transition probabilities individually shows that overall the MSM leads to a better approximation than the MCM in 16 out of the 25 predicted probabilities and in general the transition probabilities are more robustly predicted using the MSM.

5.2.2. In-sample assessment

The AMEs obtained for each model prove that the MSM leads to a better approximation of the observed 10-year TPM than the MCM whatever the subsample considered (see Figure 2). For all subsamples, the AME is never higher than 0.85 for the MSM while it is always over 0.95 for the MCM meaning that, with respect to the MCM, the 10-year TPM obtained using the MSM is always closer to the observed one in term of percentage of deviations. Figure 2 also shows that the accuracy of the 10-year TPM prediction increases for both models when increasing the number of consecutive years farms remain in the database. This could be explained by the fact that farms remaining short time periods in the sample could be noise for the model parameter estimation. The shorter is the time period that farms remain in the database the more incomplete is the information about them. Therefore, a balanced panel should provide better approximation of the underlying transition probabilities for both models. From Figure 2, it should be noted also that the resulting AMEs for the MSM seem more stable than those for the MCM, suggesting that farms remaining short time periods are more noise for the MCM than for the MSM. Nevertheless, the computed standard errors show that the accuracy of both models decreases when increasing the number of consecutive years farms remain in the database, probably because a decrease of the number of observations when increase the

Table 5

Observed 10-year TPM and the predicted ones for both models (subsample 10)

		ES class				
		0-50	50-100	100-150	150-250	≥ 250
ES class	0-50	0.715	0.235	0.029	0.014	0.007
	50-100	0.107	0.641	0.199	0.038	0.015
	100-150	0.020	0.146	0.536	0.268	0.030
	150-250	0.010	0.032	0.096	0.630	0.232
	≥ 250	0.005	0.021	0.020	0.124	0.830

a) Observed 10-year TPM ($\mathbb{P}^{(10)}$)

		ES class				
		0-50	50-100	100-150	150-250	≥ 250
ES class	0-50	**0.476**	**0.361**	**0.106**	**0.043**	**0.014**
		(0.028)	(0.020)	(0.008)	(0.006)	(0.004)
	50-100	**0.141**	**0.467**	**0.240**	**0.116**	**0.036**
		(0.011)	(0.015)	(0.011)	(0.007)	(0.004)
	100-150	**0.044**	**0.234**	**0.338**	**0.281**	**0.103**
		(0.004)	(0.011)	(0.013)	(0.011)	(0.007)
	150-250	**0.015**	**0.082**	**0.193**	**0.428**	**0.282**
		(0.002)	(0.005)	(0.010)	(0.013)	(0.013)
	≥ 250	**0.005**	**0.026**	**0.068**	**0.245**	**0.656**
		(0.001)	(0.003)	(0.005)	(0.013)	(0.018)

b) Predicted MCM 10-year TPM ($\hat{\mathbf{P}}^{(10)}$)

		ES class				
		0-50	50-100	100-150	150-250	≥ 250
ES class	0-50	**0.690**	**0.140**	**0.097**	**0.053**	**0.020**
		(0.017)	(0.010)	(0.007)	(0.005)	(0.003)
	50-100	**0.060**	**0.684**	**0.126**	**0.090**	**0.040**
		(0.007)	(0.010)	(0.007)	(0.005)	(0.003)
	100-150	**0.041**	**0.119**	**0.586**	**0.164**	**0.090**
		(0.004)	(0.007)	(0.012)	(0.009)	(0.006)
	150-250	**0.018**	**0.062**	**0.117**	**0.676**	**0.127**
		(0.002)	(0.004)	(0.006)	(0.010)	(0.009)
	≥ 250	**0.005**	**0.021**	**0.048**	**0.093**	**0.833**
		(0.001)	(0.002)	(0.003)	(0.005)	(0.008)

c) Predicted MSM 10-year TPM ($\hat{\mathbf{\Pi}}^{(10)}$)

Notes: estimated parameters in bold font; bootstrap standard deviations in parenthesis (1000 replications); the observed TPM ($\mathbb{P}^{(10)}$) was computed directly from data (see text).

Source: Agreste, RICA France 2000-2010 – authors' calculations

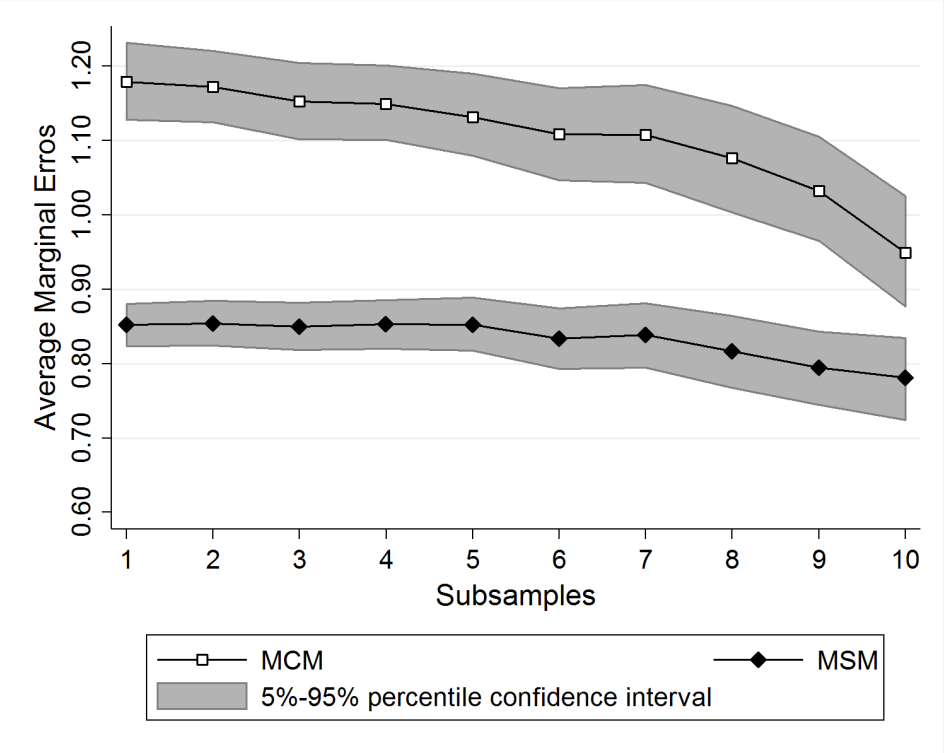

Figure 2: Computed average of marginal errors (AME) between the observed and predicted 10-year TPMs for both models, by subsamples

Notes: subsample 1 corresponds to the subset of farms remaining present in the database at least 2 consecutive years; subsample 2 to those remaining at least 3 consecutive years; and so forth. The confidence intervals are based on the percentiles, that is, the lower and upper bounds are the percentiles 5% and 95%, respectively.

Source: Agreste, FADN France 2000-2010 – authors' calculations

number of years that farms remain in the database (see Table 2).[4]

Considering subsample 10, the AME on the overall predicted 10-year TPM is around 0.95 for the MCM while it is about 0.78 for the MSM meaning that the predicted 10-year TPM for the MSM is about 17% closer the observed TPM one than for the MCM (see Table 6). However, the MSM improvement mainly comes from the main diagonal elements: when only considering those, the MSM does about five times better than the MCM (0.292/0.057=5.12), while both models almost compare for off-diagonal elements. This result confirms that accounting for heterogeneity in the rate of movement avoid overestimating the mobility of farms. Thus, the transition process of farms is more accurately estimated particularly for the main diagonal elements. Since transitions of farms may be relatively slow, such a modeling framework (the MSM) should lead to more accurate forecasts of farm sizes distribution on long-run, with respect to the MCM.

[4]Since we used bootstrap to compute standard deviations, we used the percentile method to construct the confidence interval for the AMEs as well as for the r-step transition probabilities. This method is relevant for censored data and when the bootstrapped distribution for the estimated parameter is approximatively normal (Efron, 1981).

Table 6

Average of marginal errors (AME) between observed 10-year TPM ($\mathbb{P}^{(10)}$) and predicted ones ($\hat{\mathbf{\Pi}}^{(10)}$ and $\hat{\mathbf{P}}^{*(10)}$) (subsample 10)

TPM	Overall	Main diagonal	Off-diagonal
MCM	0.949	0.292	0.657
	(0.044)	(0.010)	(0.036)
MSM	0.781	0.057	0.724
	(0.034)	(0.007)	(0.035)

Notes: bootstrap standard deviations in parenthesis (1000 replications).

Source: Agreste, RICA France 2000-2010 – authors' calculations

5.2.3. Out-of-sample forecasting

Given the transition process described by the models within the resulting TPMs and observed farm distributions, one might want to know how these distributions look like some years after and also how the models recover them. Therefore, out-of-sample short-, medium- and long-run forecasts were performed using both models as follows. For a short-run perspective, the estimated 1-year TPMs ($\hat{\mathbf{\Pi}}$ for the MCM and $\hat{\mathbf{P}}^*$ for the MSM) and the observed distributions in 2010, 2011 and 2012 were used to forecast the distributions in 2011, 2012 and 2013, respectively. For medium-run forecasts, the predicted 5-year TPMs (where $\hat{\mathbf{\Pi}}^{(5)} = (\hat{\mathbf{\Pi}})^5$ for the MCM and $\hat{\mathbf{P}}^{*(5)} = \hat{\mathbf{S}} + (\mathbf{I} - \hat{\mathbf{S}})\hat{\mathbf{M}}^{*5}$ for the MSM) and the observed distributions in 2006, 2007 and 2008 were used while for long-run forecasts the predicted 11-year TPMs (where $\hat{\mathbf{\Pi}}^{(11)} = (\hat{\mathbf{\Pi}})^{11}$ for the MCM and $\hat{\mathbf{P}}^{*(11)} = \hat{\mathbf{S}} + (\mathbf{I} - \hat{\mathbf{S}})\hat{\mathbf{M}}^{*11}$ for the MSM) and the observed distributions in 2000, 2001 and 2002 were used. Then, the resulting short-, medium- and long-run distributions from both models were compared to the observed distributions in 2011, 2012 and 2013 based on the AMEs. Figure 3 presents the AMEs computed for both model using 1000 bootstrap replications.

Figure 3 shows that as expected the accuracy of both models decreases when increasing the forecast horizon time. The computed AMEs are smaller for short-run forecasts than for medium- and long-run ones for both models. However, while both models almost compare short-run farm distribution forecasts, the MSM performs better than the MCM for medium- and long-run forecasts. The resulting AMEs are almost equivalent for short-run forecasting while for the long-run, for example, the MSM does almost one time and a half better than the MCM. This means that overall the MSM leads to a closer approximation of the observed farm size distributions than the MCM particularly on the long-run. Figure **??** also shows that the accuracy on forecasting farm size distributions as well as the robustness of the results decrease more rapidly for the MCM than for the MSM when increased the horizon time of projections. Considering only medium- and long-run forecasts, the AME increases about 27% on average over the three years for the MCM (0.088/0.121=0.273) while it increases only about 17% for the MSM (0.068/0.082=0.171).

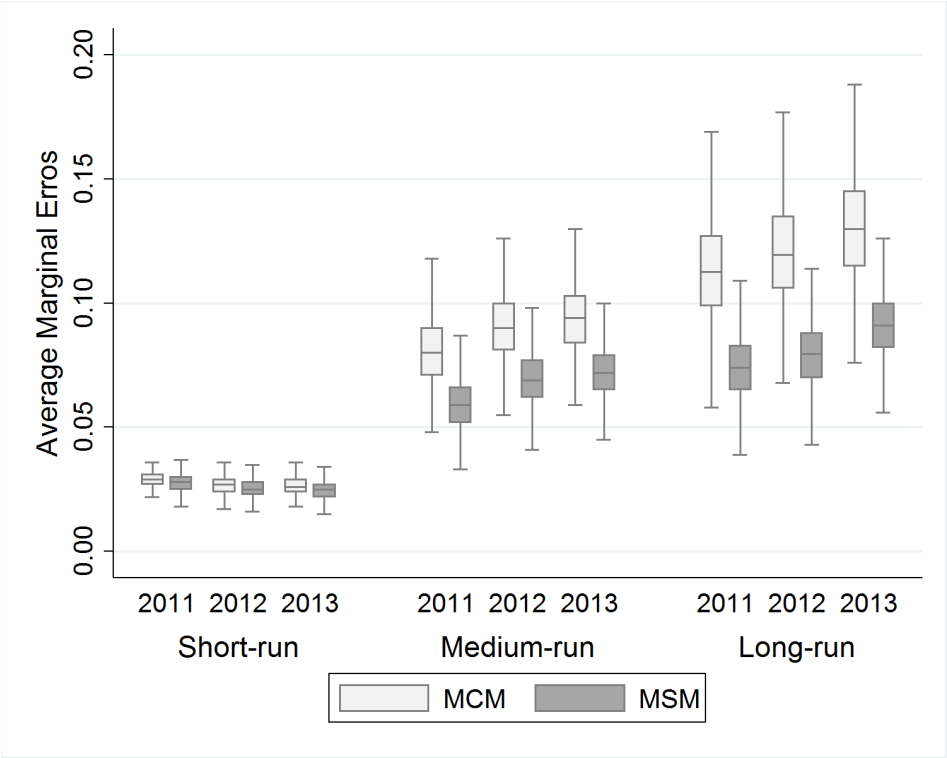

Figure 3: Average of marginal errors (AME) between observed and forecasting farm size distributions for both models.

Notes: see text for an explanation on how short-, medium- and long-run forecasts were obtained.

Source: Agreste, RICA France 2000-2013 – authors' calculations

Concluding remarks

As has already been found in other strands of the economic literature, the empirical analysis provided in this article reveals that accounting for unobserved heterogeneity to relax the assumption of homogeneity in the transition process which grounds the simple Markov chain model (MCM), leads to a better representation of the underlying structural change process also in the farming sector. Using a more general framework to accounting for heterogeneity in the rate of movement of farms, the 1-year transition probability matrix is decomposed into, on the one hand, a fraction of 'stayers' who remain in their initial size category and, on the other hand, a fraction of 'movers' who follow a standard Markovian process. Accounting for such unobserved farm heterogeneity using a mover-stayer model (MSM)allows deriving a closer estimate of the observed long-run transition matrix as well as of the distribution of farms across size categories.

The results also show that the MSM leads to more accurate and robust estimates for both the transition probabilities and the farm distribution forecasts than the simple MCM whatever the number of time that farms remained in the database. This suggest that the MSM is a more consistent modeling framework to describe farm transition process and to perform farm size distribution forecasts than the simple MCM. Therefore, we conclude that such a modeling framework allows recovering the data generating the process of farm

structural change in a more efficient way than the simple MCM, and is a better choice to describe this process as well as to perform farm size distribution forecasts over time.

Still, the MSM, as proposed by Blumen et al. (1955) and implemented here, is quite a restricted and simplified version of the more general model which was presented at the beginning of this article. Even though we improved Blumen et al. (1955)'s calibration method by using the elaborate expectation-maximization (EM algorithm) estimation method of Frydman (2005), extending the MSM framework could lead to even more economically sound, as well as statistically more accurate, models for the farming sector. We briefly mention some of such extensions which we think are promising. Firstly, more heterogeneity across farms could be incorporated by allowing for more than two types of farms. Considering for example movers at different rate of movement may lead to a better representation of the structural change process in the farming sector. Secondly, the quite strong assumption of a 'pure stayer' type could be relaxed because it may look unlikely that some farms 'never move', *i.e*, won't change size category over their entire lifespan.

Finally, the last direction we recommend to extend this modeling framework consists in accounting for entries and exits and developing a non-stationary version of the M-MCM model. Indeed, we think that such a generalized version of the MSM approach could certainly prove very insightful for analyzing structural change in the farming sector, in particular to get a better understanding of the impact of some explanatory variables, including agricultural policies, on the development of farm numbers and sizes because it should allow recovering the transition process in a more efficient way.

References

Agreste Primeur (2011, December). Production is concentrated in specialised farms. In *Agricultural census 2010: Farm structure*, Number number 272, pp. 4. Agreste: la statistique agricole.

Anderson, T. W. and L. A. Goodman (1957, 03). Statistical inference about markov chains. *Annals of Mathematical Statistics 28*(1), 89–110.

Ben Arfa, N., K. Daniel, F. Jacquet, and K. Karantininis (2014). Agricultural policies and structural change in french dairy farms: A nonstationary markov model. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*.

Blumen, I., M. Kogan, and P. J. McCarthy (1955). *The industrial mobility of labor as a probability process*, Volume VI. Cornell Studies in Industrial and Labor Relations.

Butault, J.-P. and N. Delame (2005). Concentration de la production agricole et croissance des exploitations. *Economie et statistique 390*(1), 47–64.

Cipollini, F., C. Ferretti, and P. Ganugi (2012). Firm size dynamics in an industrial district: The mover-stayer model in action. In A. Di Ciaccio, M. Coli, and J. M. Angulo Ibanez (Eds.), *Advanced Statistical Methods for the Analysis of Large Data-Sets*, Studies in Theoretical and Applied Statistics, pp. 443–452. Springer Berlin Heidelberg.

Dempster, A. P., N. M. Laird, D. B. Rubin, et al. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal statistical Society 39*(1), 1–38.

Dutta, J., J. A. Sefton, and M. R. Weale (2001). Income distribution and income dynamics in the united kingdom. *Journal of Applied Econometrics 16*(5), 599–617.

Efron, B. (1979, 01). Bootstrap methods: Another look at the jackknife. *Ann. Statist. 7*(1), 1–26.

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association 76*(374), 312–319.

European Commission (2010, 11). *Farm Accounting Data Network. An A to Z of methodology*. Brussels (Belgium): DG Agri.

Fougère, D. and T. Kamionka (2003). Bayesian inference for the mover-stayer model in continuous time with an application to labour market transition data. *Journal of Applied Econometrics 18*(6), 697–723.

Frydman, H. (1984). Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association 79*(387), 632–638.

Frydman, H. (2005). Estimation in the mixture of Markov chains moving with different speeds. *Journal of the American Statistical Association 100*(471), 1046–1053.

Frydman, H. and A. Kadam (2004). Estimation in the continuous time mover-stayer model with an application to bond ratings migration. *Applied Stochastic Models in Business and Industry 20*(2), 155–170.

Frydman, H., J. G. Kallberg, and D.-L. Kao (1985). Testing the adequacy of Markov chain and mover-stayer models as representations of credit behavior. *Operations Research 33*(6), 1203–1214.

Frydman, H. and T. Schuermann (2008, June). Credit rating dynamics and Markov mixture models. *Journal of Banking and Finance 32*(6), 1062–1075.

Fuchs, C. and J. B. Greenhouse (1988). The EM algorithm for maximum likelihood estimation in the mover-stayer model. *Biometrics 44*, 605–613.

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association 56*(296), 841–868.

Hallberg, M. C. (1969). Projecting the size distribution of agricultural firms-an application of a markov process with non-stationary transition probabilities. *Amer. J. Agr. Econ. 51*(2), 289–302.

Huettel, S. and R. Jongeneel (2011, January). How has the EU milk quota affected patterns of herd-size change? *European Review of Agricultural Economics 38*(4), 497–527.

Jafry, Y. and T. Schuermann (2004). Measurement, estimation and comparison of credit migration matrices. *Journal of Banking & Finance 28*(11), 2603 – 2639. Recent Research on Credit Ratings.

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

Langeheine, R. and F. Van de Pol (2002). *Applied latent class analysis*, Chapter Latent markov chains, pp. 304–341. Cambridge Univ Pr.

Lee, T., G. Judge, and A. Zellner (1977). *Estimating the parameters of the Markov probability model from aggregate time series data*. Amsterdam: North Holland.

Lee, T. C., G. G. Judge, and T. Takayama (1965). On estimating the transition probabilities of a Markov process. *Journal of Farm Economics 47*(3), 742–762.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 226–233.

MacRae, E. C. (1977). Estimation of time-varying Markov processes with aggregate data. *Econometrica 45*(1), 183–198.

McFarland, D. D. (1970). Intragenerational social mobility as a markov process: Including a time-stationary markovian model that explains observed declines in mobility rates over time. *Am. Sociol. Rev.*, 463–476.

McLachlan, G. and T. Krishnan (2007). *The EM algorithm and extensions*, Volume 382. John Wiley & Sons.

Morgan, T. M., C. S. Aneshensel, and V. A. Clark (1983). Parameter estimation for mover-stayer models analyzing depression over time. *Sociological Methods & Research 11*(3), 345–366.

Piet, L. (2011). Assessing structural change in agriculture with a parametric Markov chain model. illustrative applications to EU-15 and the USA. XIIIth Congress of the European Association of Agricultural Economists, Zurich (Switzerland).

Spilerman, S. (1972). The analysis of mobility processes by the introduction of independent variables into a Markov chain. *American Sociological Review 37*(3), 277–294.

Stavins, R. N. and B. F. Stanton (1980). *Alternative Procedures for Estimating the Size Distribution of Farms*. Department of Agricultural Economics, New York State College of Agriculture and Life Sciences.

Storm, H., T. Heckelei, and R. C. Mittelhammer (2011). Bayesian estimation of non-stationary Markov models combining micro and macro data. Discussion Paper 2011:2, University of Bonn, Institute for Food and Resource Economics, Bonn (Germany).

Swensen, A. R. (1996). On maximum likelihood estimation in the mover-stayer model. *Communications in statistics. Theory and methods A 25*(8), 1717–1728.

van de Pol, F. and R. Langeheine (1989). *Multiway data analysis*. Amsterdam, The Netherlands: North-Holland Publishing Co.

Weiss, C. R. (1999). Farm growth and survival: Econometric evidence for individual farms in upper austria. *Amer. J. Agr. Econ. 81*(1), 103–116.

Zepeda, L. (1995). Asymmetry and nonstationarity in the farm size distribution of Wisconsin milk producers: An aggregate analysis. *American Journal of Agricultural Economics 77*(4), 837–852.

Zimmermann, A. and T. Heckelei (2012, September). Structural change of European dairy farms: A cross-regional analysis. *Journal of Agricultural Economics 63*(3), 576–603.

Zimmermann, A., T. Heckelei, and I. P. Dominguez (2009). Modelling farm structural change for integrated ex-ante assessment: Review of methods and determinants. *Environmental Science and Policy 12*(5), 601–618.

A. Appendix

A.1. Frydman (2005)'s specification of the mixed Markov chain model (M-MCM)

As the number of parameters to estimate increases with the number of homogeneous agent types (G), the estimation of equation (5) could be an ill-posed problem. As an issue, Frydman (2005) proposed a parameterization of the M-MCM to decrease the number of parameters to be estimated. Considering heterogeneity in the rate of movement of agents and assuming that all types-g TPMs are related to a specific one namely the generator matrix ($\mathbf{M}$), only the elements of $\mathbf{S}_g$ (*i.e.*, the shares of type-g agents in each category) and the generator matrix elements (m_{ij}) are estimated. All the others possible TPMs ($\mathbf{M}_g$, $\forall\ 1 \leq g \leq G-1$) are then derived from the generator matrix using the relative rate of movement of agent type-g.

Assuming that all type-g transition probability matrices (TPMs) are related to a specific one, chosen arbitrarily as that of the last agent type, the TPM of any agent type-g is writes as:

$$\mathbf{M}_g = \mathbf{I} - \mathbf{\Lambda}_g + \mathbf{\Lambda}_g\mathbf{M} \quad \text{for} \quad 1 \leq g \leq G-1 \tag{20}$$

where $\mathbf{\Lambda}_g = \text{diag}(\lambda_{i,g})$, $\mathbf{M} = \mathbf{M}_G$ (*i.e.*, $\mathbf{\Lambda}_G = \mathbf{I}$) and $0 \leq \lambda_{i,g} \leq \frac{1}{1-m_{ii}}$ $(\forall i \in J)$ with $0 \leq m_{ii} \leq 1$ the main diagonal elements of matrix $\mathbf{M}$, that is, the probability to remain in the starting category from one period to the next for movers.

The $\lambda_{i,g}$ parameters inform about heterogeneity in the rates of movement across homogeneous agent types: $\lambda_{i,g} = 0$ if type-g agents starting in state i never move out of i; $0 < \lambda_{i,g} < 1$ if they move at a lower rate than the generator matrix $\mathbf{M}$; $\lambda_{i,g} > 1$ if they move at a higher rate than the generator matrix $\mathbf{M}$ and; the expected time spent in state i of observations generated by $\mathbf{M}_g$ is given by $1/[\lambda_{i,g}(1-m_{ii})]$ $(\forall \lambda_{i,g} > 0)$.

A.2. Maximum likelihood of the mixed Markov chain model (M-MCM)

Consider a population of n agents k, each agent k is observed at some discrete time points on time interval $[0, T_k]$ with $T_k \leq T$, where T is the time horizon of all observations. According to Anderson and Goodman (1957), the likelihood that the transition history of agent k (X_k) was generated by the specific Markov chain (*i.e.*, that k belongs to type g), conditional on knowing that k was initially in state i_k, is given by:

$$l_{k,g} = s_{i_k,g} \prod_{i \neq j} (m_{ij,g})^{n_{ij,k}} \prod_{i} (m_{ii,g})^{n_{ii,k}} \tag{21}$$

where $s_{i_k,g}$ is the share of type-g agents initially in category i_k, $n_{ij,k}$ is the number of transitions from i to j made by agent k, with $j \neq i$, $n_{ii,k}$ is the total time spent by k in category i and $m_{ii,g}$ and $m_{ij,g}$ are elements of matrix $\mathbf{M}_g$.

Under Frydman (2005)'s specification of the M-MCM as defined by equation (20), that is, all specific matrices $\mathbf{M}_g$ are related to a generator matrix $\mathbf{M}$, the likelihood rewrites:

$$l_{k,g} = s_{i_k,g} \prod_{i \neq j} (\lambda_{i,g} m_{ij})^{n_{ij,k}} \prod_{i} (m_{ii,g})^{n_{ii,k}} \tag{22}$$

where $\lambda_{i,g}$ is the relative rate of movement of agent type-g and therefore $m_{ii,g} = 1 - \lambda_{i,g} + \lambda_{i,g} m_{ii}$ as a consequence of the relation established in equation (20).

Then, the log-likelihood function for the whole population writes:

$$\log L = \sum_{k=1}^{n} \sum_{g=1}^{G} (Y_{k,g} \log l_{k,g}) \tag{23}$$

where $Y_{k,g}$ is an indicator variable which equals 1 if agent k belongs to type g and 0 otherwise. The likelihood of the MSM can be easily derived by stating G=2 and Λ_1=0 according to the relation established in equation (20).

A.3. Computing standard errors from EM algorithm equations

To compute standard errors from EM algorithm equations, two components are required (Louis, 1982): the observed information matrix given by the negative of the Hessian matrix of the log-likelihood function and the missing information matrix obtained from the gradient vector, that is, the vector of score statistic based on complete information. Since the log-likelihood function given by equation (11) is twice differentiable with respect to the model parameters, the standard errors can be thus computed as follows.

Let $\mathbf{\Omega}_c(\mathbf{Z}; \hat{s}_i^*, \hat{m}_{ij}^*)$ and $\mathbf{\Omega}_m(\mathbf{Z}; \hat{s}_i^*, \hat{m}_{ij}^*)$ $(i, j = 1, ..., J)$ be the observed $(d \times d)$ information matrices in terms of complete and missing information, respectively, where $Z_k = (X_k, Y_k)$, (with Y_k unobserved information) and d is the number of estimated parameters. The observed information matrix in terms of incomplete information can then be derived as:

$$\mathbf{\Omega}(\mathbf{X}; \hat{s}_i^*, \hat{m}_{ij}^*) = \mathbf{\Omega}_c(\mathbf{Z}; \hat{s}_i^*, \hat{m}_{ij}^*) - \mathbf{\Omega}_m(\mathbf{Z}; \hat{s}_i^*, \hat{m}_{ij}^*), \tag{24}$$

where $\mathbf{\Omega}_m(\mathbf{Z}; \hat{s}_i^*, \hat{m}_{ij}^*)$ is given by:

$$\mathbf{\Omega}_m(\mathbf{Z}; \hat{s}_i^*, \hat{m}_{ij}^*) = \mathbf{E}[\mathbf{S_c}(\mathbf{Z}; \hat{s}_i^*, \hat{m}_{ij}^*) \times \mathbf{S_c}(\mathbf{Z}; \hat{s}_i^*, \hat{m}_{ij}^*)'], \tag{25}$$

where $\mathbf{S_c}(\mathbf{Z}; \hat{s}_i^*, \hat{m}_{ij}^*)$ is the vector of score statistic in terms of complete information.

Therefore, if the observed information matrix in terms of incomplete information just described $\mathbf{\Omega}(\mathbf{X}; \hat{s}_i^*, \hat{m}_{ij}^*)$ is invertible, the standard errors are given by:

$$\mathbf{SE} = \{\psi_{ll'}^{1/2}\}, \tag{26}$$

where $\mathbf{SE}$ is the $1 \times d$ vector of standard errors, $\mathbf{\Psi} = \{\psi_{ll'}\} = \mathbf{\Omega}^{-1}(\mathbf{X}; \hat{s}_i^*, \hat{m}_{ij}^*)$ is defined as the asymptotic covariance matrix of the maximum likelihood estimators $\hat{s}_i^*$ and $\hat{m}_{ij}^*$ $(\forall i, j = 1, ..., J)$ under incomplete information and $l, l' = 1, ..., d$ (McLachlan and Krishnan, 2007).