



HAL
open science

Prise en compte de l'hétérogénéité inobservée des exploitations agricoles dans la modélisation du changement structurel : illustration dans le cas de la France.

Legrand Dunold Fils Saint-Cyr

► To cite this version:

Legrand Dunold Fils Saint-Cyr. Prise en compte de l'hétérogénéité inobservée des exploitations agricoles dans la modélisation du changement structurel : illustration dans le cas de la France.. Economies et finances. Agrocampus Ouest, 2016. Français. NNT : 2016NSARE043 . tel-01649738

HAL Id: tel-01649738

<https://theses.hal.science/tel-01649738v1>

Submitted on 27 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AGRO
CAMPUS
OUEST

Legrand Dunold Fils SAINT-CYR • 12 décembre 2016

Thèse AGROCAMPUS OUEST
sous le label de l'Université Bretagne Loire
pour obtenir le grade de
DOCTEUR D'AGROCAMPUS OUEST
Spécialité Sciences Économiques et de Gestion

ÉCOLE DOCTORALE • Sciences de l'homme, des
organisations et de la Société (SHOS)

LABORATOIRE D'ACCUEIL • Laboratoire d'études et de
recherches en économie sur les structures et marchés
agricoles, ressources et territoires (SMART-LERECO)

**Prise en compte de
l'hétérogénéité inobservée
des exploitations agricoles
dans la modélisation
du changement structurel :
illustration dans le cas
de la France**

Cécile DETANG-DESSENDRE

Directrice de Recherche, INRA / *présidente*

Thierry KAMIONKA

Directeur de recherche, CNRS / *rapporteur*

Roel JONGENEEL

Senior Researcher, Wageningen UR / *rapporteur*

Alain CARPENTIER

Directeur de recherche, INRA / *examineur*

Ani KATCHOVA

Associate Professor, Ohio State University / *examinatrice*

Jérôme PAYOUX

Directeur de Clientèle, CA 35/ *examineur*

Laure LATRUFFE

Chargée de recherche, INRA Rennes / *directrice de thèse*

Laurent PIET

Ingénieur de recherche, INRA Rennes / *co-directrice de thèse*



N° d'ordre : 2016-39

N° de série : E-43

THESE / AGROCAMPUS OUEST

Sous le label de l'Université Européenne de Bretagne

Pour obtenir le grade de

DOCTEUR D'AGROCAMPUS OUEST

Spécialité: **Sciences Economiques et de Gestion**

Ecole doctorale: **Sciences de l'Homme des Organisations et de la Société (SHOS)**

Présentée par

Legrand Dunold Fils SAINT-CYR

Accounting for Unobserved Farm Heterogeneity in Modelling Structural Change: Evidence from France

Soutenue le 12 décembre 2016 devant la Commission d'Examen

Composition du Jury:

Cécile Détang-Dessendre	Directrice de Recherche, INRA	Présidente
Thierry Kamionka	Directeur de Recherche, CNRS	Rapporteur
Roel Jongeneel	Senior Researcher, Wageningen UR	Rapporteur
Alain Carpentier	Directeur de Recherche, INRA	Examineur
Ani Katchova	Associate Professor, Ohio State University	Examinatrice
Jérôme Payoux	Directeur de Clientèle, CA 35	Examineur
Laure Latruffe	Chargée de Recherche, INRA	Directrice de thèse
Laurent Piet	Ingénieur de Recherche, INRA	Co-encadrant



Abstract

The farming sector has faced important structural changes over the last decades, particularly in Western Europe. Such changes may have important consequences for the viability of farming and for other related economic sectors. It is thus crucial to understand the mechanisms underlying this process, which results from individual farmers' decisions, in order to identify its main drivers and to generate prospective scenarios. The thesis contributes to this objective by analysing the impact of farm heterogeneity on structural change in farming.

The complexity resulting principally from the multidimensionality of structural change, has led to the development and application of a variety of modelling approaches in the literature. Among others, the Markov chain modelling framework is particularly suitable to model farm size change over time. While some efforts have been devoted to improve the models' specifications as well as the estimation techniques, less attention has been paid so far to taking into account farm heterogeneity, especially using this modelling approach. However, as heterogeneity is obviously present in farming, studies which assume homogeneity may be misleading. Hence, the thesis aims at controlling for both observed and unobserved heterogeneity in modelling farm size change under a Markovian-like process. To do this, a mixture Markov framework is applied for the first time in agricultural economics, to evidence heterogeneity in the transition process of farms. The performance of this approach is first investigated using a restrictive form of the model, and its general form is then applied to investigate the impact of agricultural policies on structural change.

While Markovian approaches enable mirroring farm interdependency by considering at the same time several dimensions of structural change (entry/exit and growth process), they are unable to identify specific types of interactions between farms. However, understanding the relationships that may exist between farms may give a better insight into the ongoing structural change in farming. Another objective of this thesis is thus to provide a modelling approach that allows identifying specific types of relationships existing between farms and their neighbours in specific farming contexts. To address this question, a mixture modelling framework is applied to investigate heterogeneity in spatial interdependence between farms, considering farm survival and farm growth.

The main conclusions of this thesis are threefold. Firstly, ignoring farm heterogeneity when studying structural change, in particular farm size dynamics, could lead to inaccurate

projection of the future distribution of farm sizes. Secondly, the impacts of agricultural policy measures and other time-varying variables on farm size dynamics depend on specific types of farms that are not observable. Lastly, the spatial interdependence that exists between farms could be a starting point to understand structural change observed at an aggregate or regional level.

Keywords: Agricultural policy; Expectation-Maximisation (EM) algorithm; farms; Markovian process; mixture models; spatial interdependence; structural change; unobserved heterogeneity.

Résumé

Le secteur agricole a fait face à d'importants changements structurels depuis les dernières décennies, en particulier en Europe de l'Ouest. Ces changements peuvent avoir des conséquences non seulement pour la viabilité du secteur mais aussi pour certains autres secteurs de l'économie qui lui sont liés. Il est donc crucial de comprendre les mécanismes sous-jacents à ce processus, qui résulte en particulier des décisions des agriculteurs au niveau individuel, afin d'identifier les principaux déterminants et ainsi de pouvoir dégager des scénarios prospectifs. La thèse s'inscrit dans cette optique en analysant l'impact de l'hétérogénéité des comportements des agriculteurs sur le changement structurel.

La complexité due au caractère multidimensionnel du changement structurel a conduit au développement et à l'application de différentes approches de modélisation dans la littérature. Entre autres, les modèles de chaînes de Markov sont particulièrement adaptés à la modélisation du changement de taille des exploitations agricoles. Alors que beaucoup d'efforts ont été consacrés à l'amélioration de la spécification des modèles ainsi qu'aux techniques d'estimation, moins d'attention a été accordée à ce jour à la prise en compte de l'hétérogénéité des exploitations agricoles, en particulier dans le cadre des modèles Markoviens. Cependant, l'hétérogénéité étant certainement présente en agriculture, les études faisant l'hypothèse d'homogénéité peuvent aboutir à des résultats erronés. De ce fait, la thèse vise à prendre en compte à la fois l'hétérogénéité observée et l'hétérogénéité non observée dans l'analyse du changement de taille des exploitations, tout en considérant une approche Markovienne. Pour ce faire, une approche par les modèles de mélange de chaînes de Markov est appliquée pour la première fois en économie agricole pour mettre en évidence l'hétérogénéité dans le processus de transition des exploitations agricoles. La performance de cette approche est d'abord étudiée, en utilisant une forme restrictive du modèle, puis l'approche est appliquée sous sa forme générale pour étudier l'impact de mesures de politique agricole sur le changement structurel.

Alors que les approches Markoviennes permettent de considérer l'aspect spatial et l'interdépendance pouvant exister entre les exploitations agricoles, en prenant en compte plusieurs dimensions du changement structurel (installation, agrandissement, et cessation d'activités), elles ne permettent pas d'identifier des types spécifiques d'interactions entre les exploitations. Cependant, comprendre l'interdépendance spatiale entre les exploitations peut apporter une meilleure compréhension de l'évolution structurelle dans l'agriculture. Un autre objectif de cette thèse est donc de proposer une approche de modélisation à partir de laquelle il est possible

d'identifier les canaux d'interdépendance entre les exploitations agricoles dans des contextes spécifiques. Pour répondre à cette question, l'approche par les modèles de mélange a été appliquée pour étudier l'hétérogénéité et l'interdépendance entre exploitations voisines, dans le cadre la survie et du processus d'agrandissement dans le secteur agricole.

Trois principales conclusions découlent de cette thèse. Tout d'abord, ignorer l'hétérogénéité dans le comportement des agriculteurs, en particulier dans le cadre de l'analyse de l'évolution de la taille des exploitations, pourrait conduire à des résultats erronés en termes de prédiction et projection de distribution future des tailles d'exploitation. Deuxièmement, l'impact des politiques agricoles et de certaines autres variables qui varient dans le temps, sur l'évolution de la taille d'exploitation dépend des types spécifiques d'exploitations, généralement non observables. Enfin, l'interdépendance spatiale qui peut exister entre les exploitations agricoles, pourrait être le point de départ pour bien comprendre le processus de changement structurel observé à un niveau global ou régional.

Keywords: Politiques agricoles; algorithme Espérance-Maximisation (EM); exploitations agricoles; processus Markovien; modèles de mélange; interdépendance spatiale; changement structurel; hétérogénéité inobservée.

Acknowledgements

I would like to thank my supervisors Laure Latruffe and Laurent Piet, both from INRA SMART-LERECO in Rennes, France, for their support and for giving me the opportunity to carry out my PhD as well as for their dexterous guidance. Besides my supervisors, I am very grateful to:

- The members of the PhD examining committee: Thierry Kamionka from CNRS, France; Roel Jongeneel from Wageningen UR, The Netherlands; Alain Carpentier from INRA SMART-LERECO in Rennes, France; Ani Katchova from the Ohio State University, USA; Cécile Détang-Dessendre, from INRA CESAER in Dijon, France; and Jérôme Payoux from ‘Crédit Agricole en Bretagne’, France, for their comments and scientific questions to improve the manuscript;
- Thomas Heckelei from the University of Bonn, Germany, and Muriel Roger from the Université Paris 1, France, for their invaluable advice as members of the PhD committee meeting;
- Hugo Storm from the University of Bonn, Germany, for his collaboration on a joint paper;
- Colleagues from the Research Unit SMART-LERECO (INRA and AGROCAMPUS OUEST) in Rennes, France, for their advice and for the valuable time we spent discussing;

I also thank:

- Participants to the 14th Congress of the European Association of Agricultural Economists (EAAE) in Ljubljana (Slovenia), the 89th annual Conference of the Agricultural Economics Society in Warwick (England), the 8th Conference for Research in Social Sciences (‘Journées de Recherche en Sciences Sociales’) in Grenoble (France), the 6th EAAE PhD Workshop in Rome (Italy), the 2016 Agricultural and Applied Economics Association Annual Meeting in Boston (USA), the 7th International Doctoral Meeting of Montpellier in Economics, Management and Finance in Montpellier (France), and the 149th EAAE Seminar ‘Structural change in agri-food chains: new relations between farm sector, food industry and retail sector’ in Rennes (France), for their valuable discussions and advice;

- The Chair EEA ('Entreprises et Economie Agricole') between 'Crédit Agricole en Bretagne' and AGROCAMPUS OUEST, for providing the financial grant support;
- My relatives, colleagues and close friends for their support and friendliness;
- And last but not least, my wife Rose-Sandra for her support, her kindness and encouragements; and my son Daniel who has been my great motivation to finish my doctoral works.

Contents

Abstract	ii
Acknowledgements	vi
Contents	viii
List of Figures	xi
List of Tables	xii
1 General introduction	1
1.1 Introduction	1
1.2 Trends and patterns of farm structural change in France	3
1.3 On modelling structural change in agricultural economics	9
1.4 Objectives and research questions	15
1.5 Main contributions	17
1.6 Outline of the thesis	18
Bibliography	20
Appendix	28
2 Movers and stayers in the farming sector: accounting for unobserved heterogeneity in structural change	31
2.1 Introduction	31
2.2 Modelling farm transition process using the Markov chain framework	34
2.2.1 The simple Markov chain model (MCM)	35
2.2.2 Accounting for unobserved heterogeneity: the mixed Markov chain model (M-MCM)	36
2.2.3 A simple implementation of the M-MCM: the Mover-Stayer model (MSM)	37
2.3 Estimation method	37
2.3.1 Likelihood maximisation under complete information	37
2.3.2 The expectation-maximisation (EM) algorithm under incomplete information	39
2.4 Model comparison	41
2.4.1 Likelihood ratio (LR) test	41
2.4.2 Average Marginal Error (AME)	42
2.5 Data	43
2.6 Results	47

2.6.1	In-sample estimation results	47
2.6.2	In-sample model comparison	48
2.6.3	Out-of-sample projections	50
2.7	Robustness tests	53
2.7.1	Sample size	53
2.7.2	Definition of farm size categories	55
2.8	Concluding remarks	57
Bibliography		59
Appendices		62
3	Unobserved farm heterogeneity and agricultural policy impacts on size dynamics	67
3.1	Introduction	67
3.2	The non-stationary mixed Markov chain model (M-MCM)	70
3.2.1	Model specification	72
3.2.2	Estimation procedure	74
3.3	Farm types and elasticities of explanatory variables	77
3.3.1	Choosing the optimal number of farm types	77
3.3.2	Probability elasticities	78
3.3.3	Farm structure elasticities	78
3.4	Empirical application	79
3.4.1	Data	79
3.4.2	Explanatory variables	81
3.5	Results	84
3.5.1	Type membership and transition probability matrices	84
3.5.2	Impact of explanatory variables	86
3.6	Concluding remarks	91
Bibliography		94
Appendix		100
4	Heterogeneity and spatial interdependence in farm survival and growth: evidence from France	105
4.1	Introduction	106
4.2	On heterogeneity and spatial interdependence between farms	108
4.3	Modelling approach	109
4.4	Estimation of the models	111
4.4.1	Implementing the Expectation-Maximisation (EM) algorithm	112
4.4.2	Choosing optimal number of farm types	113
4.5	Empirical application	114
4.5.1	Data used	114
4.5.2	Dependent and explanatory variables	115
4.6	Results	118
4.6.1	Farm survival model	118
4.6.2	Farm growth model	123

4.7 Concluding remarks	127
Bibliography	129
5 General conclusion	133
5.1 Overview and main findings	133
5.2 Recommendations	137
5.3 Discussion and directions for further research	139
Bibliography	143

List of Figures

- 1.1 Distribution of farms according to their production specialisation in France in 1988, 2000 and 2010. 6
- 1.2 Evolution of farm investment and debt ratio in France from 1990 to 2010. 9
- 2.1 Extrapolated population shares by farm size categories and average economic size (ES). 46
- 2.2 Average marginal error ($AME_{FSD}(t+r)$) between the out-of-sample projections of farm size distributions and the actually observed ones for both models. 53
- 2.3 Average marginal error (AME_{TPM}) between the predicted 10-year ($\hat{\mathbf{I}}^{(10)}$ and $\hat{\mathbf{\Phi}}^{(10)}$) and the observed ($\mathbf{P}^{(10)}$) transition probability matrices (TPMs), by sub-samples. 54
- 2.4 Distribution of movers by category of sizes. 55
- 3.1 Comparison of model-fit statistics for different numbers of types. 84
- 3.2 Yearly structure elasticities of farm subsidies from Pillar One of the CAP both for homogeneous and mixed Markov chain models. 91
- 3.3 Yearly structure elasticities of farm subsidies from Pillar Two of the CAP both for homogeneous and mixed Markov chain models. 92
- 4.1 Probability to survive for varying municipality-level average farm sizes by unobserved farm types (predicted margins with 95% confidence intervals). 123
- 4.2 Total land used for varying average farm size in municipality by type of farms (predicted margins with 95% confidence intervals). 126

List of Tables

- 1.1 Evolution of the number of farms in France by category of sizes from 1955 to 2010. 4
- 1.2 Evolution of the number of farms in France by type of legal status between 1970 and 2010. 5
- 1.3 Salaried and non-salaried labour in French farming from 1970 to 2010. 7
- 1.4 Some empirical applications of the Markov chain models using individual-level data 28
- 1.5 Some empirical applications of the Markov chain models using aggregate data . . 29
- 1.6 Some empirical applications of the Markov chain models using aggregate data (continued) 30

- 2.1 Distribution by economic size (ES) and average ES for the studied sample, 2000-2013. 45
- 2.2 Number of individual farms, observations and transitions by sub-samples 46
- 2.3 Estimated 1-year transition probability matrix under Markov chain model ($\hat{\mathbf{H}}$). . 48
- 2.4 Estimated stayer shares (\hat{s}_i) and 1-year transition probability matrices (TPMs) for movers ($\hat{\mathbf{M}}$) and for the overall population ($\hat{\mathbf{\Phi}}$). 49
- 2.5 Observed and predicted 10-year transition probability matrices (TPMs) for both models. 51
- 2.6 Average marginal error (AME_{TPM}) between the predicted 10-year ($\hat{\mathbf{H}}^{(10)}$ and $\hat{\mathbf{\Phi}}^{(10)}$) and the observed ($\mathbf{P}^{(10)}$) transition probability matrices (TPMs). 52
- 2.7 Average marginal error (AME_{TPM}) between the predicted 10-year ($\hat{\mathbf{H}}^{(10)}$ and $\hat{\mathbf{\Phi}}^{(10)}$) and the observed ($\mathbf{P}^{(10)}$) transition probability matrices (TPMs), for different number of size categories. 56

- 3.1 Observed 1-year transitions across size categories, 2000-2013. 80
- 3.2 Descriptive statistics and definition of explanatory variables (n=92,626). 83
- 3.3 Estimated farm type shares and 1-year transition probability matrices (TPMs). . 85
- 3.4 Descriptive statistics for observed farm characteristics by type membership. . . . 86
- 3.5 Overall population 1-year transition probability matrix (TPM). 86
- 3.6 Yearly probability elasticities of farm growth and decline both for homogeneous and mixed Markov chain models. 88
- 3.7 Yearly probability elasticities of farms to remain in the same category of sizes both for homogeneous and mixed Markov chain models. 89
- 3.8 Estimated parameters of farm probabilities to grow between two consecutive years for the homogeneous and the mixed Markov chain models. 100
- 3.9 Z-scores for testing the equality of the farm types' odds ratios coefficients for farm growth. 101
- 3.10 Estimated parameters of farm probabilities to decline between two consecutive years for the homogeneous and the mixed Markov chain models. 102

3.11	Z-scores for testing the equality of the farm types' odds ratios coefficients for farm decline.	103
4.1	Definition and descriptive statistics of explanatory variables (n=344,617).	118
4.2	Estimated parameters for both the pooled and the mixture probit model for farm survival.	119
4.3	Z-scores for testing the equality of the estimated farm types' coefficients from the mixture probit model for farm survival.	120
4.4	Main characteristics of each type of farms identified with the mixture probit model for farm survival.	122
4.5	Estimated parameters for both the pooled and the mixture linear regression model for farm growth.	124
4.6	Z-scores for testing the equality of the estimated farm types' coefficients from the mixture simple linear regression for farm growth.	125
4.7	Main characteristics of types from the mixture of normal regression for farm growth.	127

Chapter 1

General introduction

1.1 Introduction

The farming sector has faced significant changes over recent decades, predominately in developed countries such as Western Europe and the United States (Eastwood *et al.*, 2010). More specifically, the number of farms has been decreasing significantly and their average size has been increasing continually, implying changes in the distribution of farm sizes. The increase in farm size has led to a reduction of the total labour force in agriculture (Vidal, 2001). The human labour force has been progressively replaced by new technology, generally adopted by large farms in order to improve their efficiency (Carlin and Saupe, 1990; Albrecht, 1997). The need for qualified labour force in agriculture to manage new technologies implied an increase in the share of hired labour with respect to total labour. Capital has substituted for labour on farms, increasing thus labour productivity over time (Bollman *et al.*, 1995). At the same time, the market structure has also evolved with, for instance, the development of futures trading or forward contracts to cope with market instability (Roussillon-Montfort, 2008). As a consequence, production systems tend, on average, to specialise in order to satisfy the market requirements while remaining competitive (Chavas, 2001). Farming has become more capital intensive with an expansion of corporate forms to the expense of individual structures (Pritchard *et al.*, 2007). Such changes in the structure of the farming sector are typically identified by the term ‘farm structural change’ and, has long been the subject of considerable interests among agricultural economists, policy makers and other stakeholders in farming.

However, structural change is a wide concept and it remains difficult to define. The ‘New Palgrave’ Dictionary of Economics defines structural change as «*a change in the relative weight of significant components of the aggregative indicators of the economy, such as product and expenditure, exports and imports, and population and labour force*» (Goddard *et al.*, 1993). While this global definition rather refers to macroeconomic systems, it can be more restrictively applied

to specific economic sectors for the analysis of changes in, for instance, firm characteristics or structures. The starting point in the analysis of farm structural change is thus the definition of farm structure, which is as complex a concept as structural change (Stanton, 1991). Indeed, the term ‘*structure*’ in farming may refer to a variety of indicators such as farm size distribution, technology used, production systems and labour force characteristics, ownership and financing patterns as well as inter- and intra-sector linkages (Stanton, 1991; Boehlje, 1992). Structural change can be therefore viewed as a multi-dimensional process and can be analysed through various indicators related to the structure of a specific economic sector. Structural change is also a space-time process, which is relatively stable in the short term and varies slowly in the long term (Akimowicz *et al.*, 2013). This raises the difficulty to find a globally accepted definition for structural change in agricultural economics and complicates its analysis (Stanton, 1991). According to (Goddard *et al.*, 1993), structural change can be defined as changes in the essential characteristics of the productive activities. Any analysis of this process therefore requires to focus on certain aspects of this process because accounting for all its facets at the same time is merely impossible. This generally leads to consider only one or a limited set of indicators to depict the process of structural change in farming.

Despite the absence of a globally accepted definition of farm structural change, all studies agree that this process may have important consequences for the sector. Though affecting each farm individually, farm structural change may impact the viability of the sector as a whole and have social, economic and environmental consequences at a local, regional or country level (Flaten, 2002). This process thus affects equity within agriculture in term of income distribution between farmers, farms’ productivity and efficiency as well as the well-being of rural communities (Weiss, 1999; Breustedt and Glaubien, 2007). Structural change in farming may also impact other sectors of the economy because of their interconnections. Changes in farm structures may have direct effects on the upstream sector, that is to say on the providers of inputs and services to farmers, as well as on the downstream sector, that is to say the agri-food industry (Jones *et al.*, 1995). According to the European Economic and Social Committee, between four and five jobs in these two sectors are dependent upon each job in the farming sector (EESC, 2004). It thus becomes very important to understand the mechanisms underlying changes in farm structures in order to identify key drivers that influence the observed trends. This could help to generate prospective scenarios.

This introductory chapter presents a global view of structural change in farming. The first section describes the overall trends of farm structural change in France in the recent decades. The second section reviews the modelling approaches used to analyse this process, specially focusing on Markov chain models because this approach is adopted in a large part of this thesis. Section three presents the objectives of the thesis and the main research questions addressed. Section four underlines the main contributions. The last section explains the thesis’ outline.

1.2 Trends and patterns of farm structural change in France

The structure of the farming sector has been changing rapidly in France since the 1950s (Desriers, 2007). As previously mentioned, changes in farm structures have been observed considering different indicators (farm size, farm number, production systems, labour force, market structure, etc.). Early trends are still observed nowadays and have even accelerated in recent years. Hereafter, the observed trends over recent decades are briefly presented, focusing on some specific indicators that are mostly reported in the literature.

- **Concentration, professionalisation and specialisation**

Farm number and average size are two indicators, related to each other, that are generally used to highlight the degree of structural change in farming. As in most developed countries, farm number and average size have evolved in two opposite directions in France. In 1955, farms in France were more than 2.3 million. This number has progressively decreased over the years. After observing a relative diminution between 1970 and 1988, the rate of decline in farm number has been increasing (see Table 1.1). The total number of farms has fallen progressively to about 1 million in 1988 and less than 500,000 in 2010. The French farming sector has thus lost almost 80% of its farm population in less than 60 years and more than half in only 20 years. Smaller farms, in terms of total utilised agricultural area (UAA) or production potential, have been most affected by this farm number decline.¹ The number of farms operating less than 20 hectares in terms of UAA has decreased rapidly between 1955 and 2010, while the number of those operating more than 100 hectares has increased. More precisely, between 1988 and 2010 the number of farms operating less than 20 hectares has decreased by more than 60%, falling from 556,000 to 211,400. In contrast, the number of farms with more than 100 hectares has increased from 45,000 to 93,220 over the same period. The same trend has been observed for the evolution of farm economic sizes in terms of total standard output (SO). The share of small farms (i.e., farms with less than 25,000 Euros of SO) has decreased from about 47% to 36%, while the share of farms with more than 100,000 Euros of SO has increased by almost 20%, going from 147,000 to 162,000 farms (AGRESTE, 2012a).

A farm number decline, specifically in smaller farm categories, contributes to an increase in average size. This is visible in France, where, between 1955 and 2010, the average farm size has been multiplied by almost four, going from 14 to 56 hectares. Farm size has increased most rapidly between 1988 and 2010, going from 28 to 56 hectares (see Table 1.1). In 2010, 19% of

¹The production potential of a farm is measured through its economic size. In accordance with the European Union (EU) regulation (CE) No 1242/2008, a farm's economic size is measured as the total standard output (SO) of the holding expressed in Euros. SO is the average monetary value of the agricultural output at farm-gate price of each agricultural product (crop or livestock) in a given region. *This calculation notably allows the various types of production to be ranked relatively to each other.*

TABLE 1.1: Evolution of the number of farms in France by category of sizes from 1955 to 2010.

		1955	1970	1979	1988	2000	2010
Size categories (in hectares)	< 20	1,822	1,097	767	556	324	211
	20-50	392	370	347	288	139	88
	50-100	92	93	114	128	122	98
	≥ 100	1	28	35	45	79	93
All farms		2,307	1,588	1,263	1,017	664	490
Average size (in hectares)		14	19	23	28	43	56

Note: farm numbers in thousands.

Source: Agreste – Agricultural Censuses 1955, 1970, 1979, 1988, 2000 and 2010

farms operated 58% of the total UAA of the country (AGRESTE, 2012a). The average size of large farms in terms of economic size has increased from 68 to 108 hectares over the same period (AGRESTE, 2012b). This observed concentration in French farming particularly results from an increase in the rate of farm exit in smaller category of sizes. In general, it is more difficult for small farms to find a successor than their medium and large counterparts (Butault and Delame, 2005). In such cases, farms are mostly purchased or rented by large farms, which become more profitable, encouraging the enlargement process in farming (Rattin, 2009; Mann *et al.*, 2013). Non-succession may also be the consequence of the development of other economic sectors, giving some alternatives to potential farm holders' successors (Fennell, 1981; Kennedy, 1991).

Concentration may be also related to the development of corporate forms of farming. Indeed, corporate holdings have been expanding sharply in France to the expense of individual farms in recent decades. In the 50s, almost all farms were operated under the individual legal status, where farmers undertook their activity independently, and each farm thus corresponded to one manager (Schutz-Stephan and Savy, 2010). Since 1962, corporate companies have been developing continually, fostered by agricultural laws of 1960 and 1962 which aimed at integrating the agricultural sector in the wider economy (Desriers, 2007; Schutz-Stephan and Savy, 2010). In 1970, corporate farms represented less than 1% of the farm population and operated about 2% of the total UAA. Their proportion as well as their types have evolved from simple partnerships (French 'Groupement Agricole d'Exploitation en Commun' - GAEC) to more structured commercial groups such as 'Exploitation Agricole À Responsabilité Limitée' (EARL) 'Société Civile d'Exploitation Agricole' (SCEA), 'Société Anonyme' (SA) or 'Société etc.' (SARL). Between 1988 and 2010, the share of corporate farms in the total farm number has increased from 6% to

TABLE 1.2: Evolution of the number of farms in France by type of legal status between 1970 and 2010.

	1970	1979	1988	2000	2010
Individual operators	1,567.4	1,214.8	946.1	537.6	339.9
Total for all corporate forms and others	20.2	47.9	70.7	126.3	150.1
Partnerships (French ‘GAEC’)	3.9	15.3	37.7	41.5	37.2
Limited liability agricultural firm (French ‘EARL’)	-	-	1.5	55.9	78.6
Other corporate forms (French ‘SCEA’, ‘SA’, ‘SARL’, others)	5.1	26.1	26.3	26.3	30.8
Other types of status (natural person or legal entity)	11.2	6.5	5.2	2.6	3.5
All farms (excluding overseas territories)	1,587.6	1,262.7	1,016.8	663.8	490.0

Note: farm numbers in thousands.

Source: Agreste – Agricultural Censuses 1970, 1979, 1988, 2000 and 2010

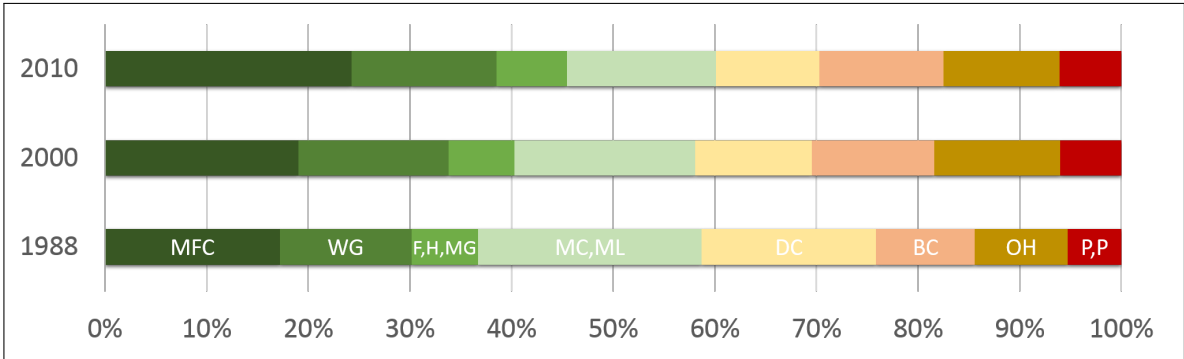
30% while the share of individual farms has fallen from 93% to 69% (see Table 1.2). Introduced for the first time in 1985, the limited liability status (EARL) developed more rapidly than other corporate forms, going from less than 1% in 1988 to about 16% in 2010. Conversely, the proportion of GAEC tends to decrease in recent years. While GAEC status is mostly used to facilitate succession in farming, EARL holdings are less risky because of the distinction allowed between private and corporate assets (Schutz-Stephan and Savy, 2010). Limited liability status can be therefore considered as more convenient for the farming sector nowadays than group share operation status found in partnership farms.

In general, corporate holdings are larger than farms operated under the individual legal status both in terms of UAA and total SO. In 2010, corporate farms represented 57% of the total UAA and 64% of the total SO of the French farming sector (AGRESTE, 2014). The average size of corporate farms is now 101 hectares whereas it is only 35 hectares for individual farms (AGRESTE, 2012b). In particular, corporate statuses allow farmers to share workforce and/or integrate non-agricultural capital investors (Bathelemy and Dussol, 2000; Delame, 2002; Schutz-Stephan and Savy, 2010). Often, corporate farms, especially under GAEC legal status, result from the merger of two pre-existing individual ones. On average, single-person EARL farms were about 1.6 times larger in terms of UAA than farms operated under individual legal status in 2010 (AGRESTE, 2014).

The expansion of large corporate holdings and the tendency to farm specialisation were concomitant in France (Rattin, 2009). Farm specialisation in specific types of production is in general associated to specific legal statuses (Barthélémy and Dussol, 2002). For example, individual legal status as well as limited liability farms (EARL) are more adequate for farms where farmers can work alone (*e.g.*, major field crops, cattle farms). By contrast, the GAEC legal status is more represented in production specialisations that are labour intensive such as dairy farms. This type of status was important for milk quotas allocation and since livestock premium rights were related to the number of farmers. Today, specific measures of the French implementation of the Common Agricultural Policy (CAP) are still more beneficial to GAECs because of this ‘transparency’. Other corporate forms such as SCEA, SA and SARL are more appropriate for farms that are capital intensive. These types of farms are more represented in wine growing farming, market gardening and horticulture. Farm specialisation has emerged as a solution to deal market requirements and improve productivity (Chavas, 2001).

In 1970, dairy farms were the most numerous, accounting for 20% of the total farm number in France (Rattin, 2009). Since then, their proportion has decreased significantly. The trend has not changed despite the introduction of milk quotas of in 1984 by the CAP. Between 1988 and 2010, the total number of dairy farms decreased by almost 30%, going from about 175,000 to 50,000. In contrast, the proportion of major field crop farms has increased (*e.g.*, from about 17% in 1988 to almost 25% in 2010), this type of farm becoming the largest farm type in 2010 despite a diminution of the number of these farms (see Figure 1.1). At the same time, beef cattle farms shares have also increased. One reason is the conversion to beef production of dairy farms following the milk quota implementation (Rattin, 2009). Nevertheless, the number of traditional mixed farms has decreased more rapidly than that of specialised farms. Mixed crop, mixed livestock (including mixed cattle) and mixed crop-livestock farms have lost about 30% of their total number, falling from 224,000 in 1988 to 72,000 in 2010.

FIGURE 1.1: Distribution of farms according to their production specialisation in France in 1988, 2000 and 2010.



Note: : MFC=major field crops; WM=wine-growing; F,H,MG=fruits, horticulture and market gardening; MC,ML: mixed crops and mixed livestock; DC=dairy cattle; BC= beef cattle; OH=other grazing livestock; P,P=Pigs and poultry; .

Source: Agreste – Agricultural Censuses 1988, 2000 and 2010

- **Modernisation and capital intensification**

The ongoing process of farm concentration, professionalisation and specialisation in the French agriculture has led to significant changes in labour characteristics. Overall, the labour force has declined sharply both in terms of the total number of employed workers and the total number of Annual Working Units (AWU) used in recent decades. The total number of permanent employees and of AWU used respectively decreased by about 75% and 70% between 1970 and 2010 (see Table 1.3). This decline has been observed both in the family and the non-family labour force. However, the total number of family workers has decreased more rapidly than non-family employees in farming, going from more 3.5 million permanent workers in 1970 to about 811,000 in 2010. This implied a progressive increase in the share of non-family labour both in terms of total number of workers and total AWU over years. Over the same period, the share of non-family agricultural workers indeed doubled, going from about 8% to 16%. Likewise, the total number of family AWU used decreased by about 73% between 1970 and 2010 while the reduction was only around 57% for non-family workers (permanent salaried, seasonal workers, agricultural external workers -ETA, see below- and shared equipment cooperatives -CUMA, see below).

TABLE 1.3: Salaried and non-salaried labour in French farming from 1970 to 2010.

	1970	1979	1988	2000	2010
Permanent workers	3,846.6	2,687.3	2,038.2	1,319.2	966.3
Family	3,511.7	2,454.5	1,876.9	1,155	811.4
Non-family salaried	334.9	232.8	161.3	164.2	154.9
Total AWU	2,373.7	1,872.3	1,445.6	957.4	751.4
Family	1,992.0	1,581.3	1,214.9	718.1	532.9
Non-family permanent salaried	280.1	198.4	142.7	137.5	127.9
Seasonal work	96.7	84.7	81.5	93.8	78.9
Contractual work (ETA, CUMA)	4.9	7.9	6.5	8.0	11.7

Note: numbers in thousands; an Annual Work Unit (AWU) is the equivalent of a full-time worker over an entire year; ETA stands for external contractors; CUMA stand for shared machinery cooperatives.

Source: Agreste – Agricultural Censuses 1970, 1979, 1988, 2000 and 2010.

In recent years, farms have tended to use relatively more seasonal workers and more contractual work, in the form of labour from service companies (the French ‘Entreprises de Travaux Agricoles’, ETA) or shared equipment cooperatives (the French ‘Coopératives d’Utilisation de Matériel Agricole’, CUMA). The share of seasonal labour has increased with the recourse to

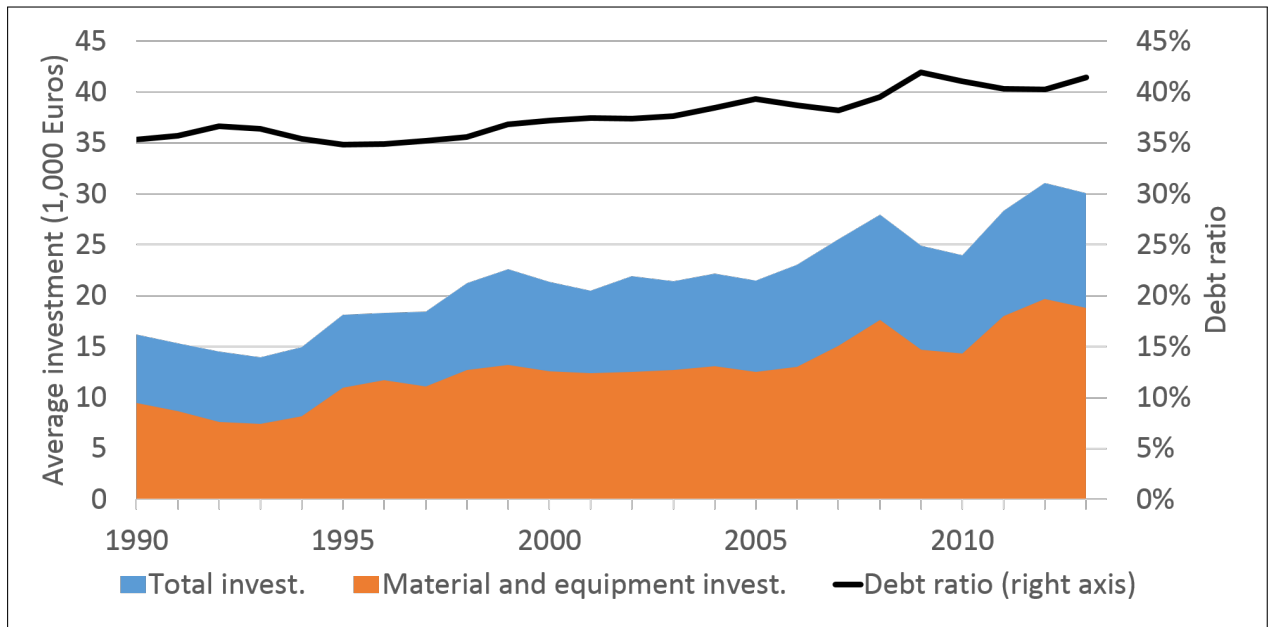
foreign seasonal workers (Darpeix *et al.*, 2014). Their proportion has increased gradually over years, going from less than 5% of the total AWU used in 1955 to about 10% in 2010. Despite their limited share in the total AWU, the development of the recourse to contractual work (ETA, CUMA, and others service companies) facilitates agricultural tasks by outsourcing labour (Darpeix *et al.*, 2014; Cahuzac and Détang-Dessendre, 2011). In 2013, 60% of the French farms used this type of labour (Lerbourg and Dedieu, 2016). In particular, resorting to CUMA enables reducing cost by allowing farmers to share agricultural machinery, which may be very expensive.

The evolution of the different types of agricultural labour depends on farm characteristics. For example, permanent employment has significantly increased in intensive livestock farms, horticulture farms and in farms specialised in quality wine-making, but it has decreased in other wine-making farms and those specialised in orchards (Rattin, 2003). In general, corporate farms use more salaried labour than individual farms, particularly in one-person EARL and other commercial legal statuses such as SCEA, SA and SARL. In 2010, non-family salaried labour represented 42% of the total AWU in a one-person EARL and almost 70% in farms with other commercial legal statuses, while this proportion was about 15% in individual farms as well as in GAEC (AGRESTE, 2014). Non-family permanent employees are particularly numerous in farms operated under commercial legal statuses. They represented 47% of the workforce in 2010, whereas their proportion remained under 25% in all other types of farms. Seasonal labour has increased particularly in off-land farms (Rattin, 2003). As farms become larger and given the ongoing technological progress, human labour has been progressively substituted by more powerful agricultural machines (Chavas, 2001; Blanc *et al.*, 2008). In 1979, 94% of the tractors used had less than 80 horsepower, but the figure fell under 50% in 2013 (Lerbourg and Dedieu, 2016). As a consequence, the need for qualified labour to manage the new technologies used has increased, as well as the apparent labour productivity.

Farm enlargement and modernisation strongly influenced investment in farming. French farms have become capital intensive and therefore more and more indebted. The investment in medium and large farms almost doubled in only 20 years, ranging from 16,200 Euros per farm on average in 1990, to 30,100 Euros in 2010 (see Figure 1.2). The evolution of farm investment is particularly related to material and equipment purchases. The share of such capital expenditures increased by about 5 percentage points between 1990 and 2010. To acquire new agricultural machinery, which are very expensive, farmers resort more and more to credit. As a consequence, the debt ratio of French farms has increased by almost 6 percentage points in only 20 years to reach 41% in 2010.

Changes are also observed in the market structure for of agricultural products. Different systems of commercialisation have developed in the recent years. Farmers tend to use more and more production contracts and participate in futures and options markets. This type of commercialisation helps reducing risk in the context of price instability (Roussillon-Montfort,

FIGURE 1.2: Evolution of farm investment and debt ratio in France from 1990 to 2010.



Note: only medium and large size farms (>25,000 Euros of standard output) in France excluding overseas territories.

Source: Agreste – Agricultural accounts information network (RICA)

2008) and in the presence of transaction costs, and/or favours better coordination (Chavas, 2001; Vavra, 2009). Farmers also developed strategies to sale directly to the consumer in order to increase their profit from farming activities. In 2010, more than 20% of farm sale their products directly to the consumers (Bary, 2012).

It is clear that the French farming sector has faced important structural over recent decades. The multiplicity of the indicators and potential correlations that may exist between them showed the complexity of this process. Analysis of farm structural change in order to identify its main drivers and to generate prospective scenarios thus requires special care.

1.3 On modelling structural change in agricultural economics

Many theoretical studies argued that structural change in agriculture is the observable result of many social, economic and environmental factors. For example, Boehlje (1992) presents such factors according to five models of structural change in farms and agribusiness firms: the technology model, the human capital model, the financial model, the institutional model, and the sociological model. The author argues that even these models should be integrated because of their interrelation, each providing useful information on the key determinants of the changing structure of the agricultural sector. Goddard *et al.* (1993) and Harrington and Reinsel (1995) review different models proposed to explain structural changes in farming in order to identify

the main drivers of this process. The former authors differentiate various attributes and causal factors of farm structural change. They highlight major factors that can influence structural change in farming and discuss the interrelation between these factors. [Harrington and Reinsel \(1995\)](#) analyse different views to reduce to common ground the issues about the factors affecting structural change in farming. The literature provides thus a large number of factors that may affect farm structural change. These factors have been divided into several categories, but it is generally supposed that farm structural change may be mainly influenced by population and economic growth, technology improvement, the market structure for inputs as well as for outputs, human capital and agricultural policies and other non-agricultural public programmes especially for rural development and market regulation (see [Goddard *et al.* \(1993\)](#); [Harrington and Reinsel \(1995\)](#); [Chavas \(2001\)](#); [Eastwood *et al.* \(2010\)](#) for detailed discussions). Possible interrelations between these factors make it difficult in general to determine the main drivers of structural change in a specific farming context. Furthermore, the impacts of specific causal factors such as agricultural policies are not unambiguous since they may depend on a variety of other factors which are specific to the farming sector considered ([Goddard *et al.*, 1993](#); [Harrington and Reinsel, 1995](#)). This thus prompted the necessity to empirically investigate the main drivers of observed structural change in some specific farming contexts.

- **Markov chain analysis**

In agricultural economics, several empirical studies have been devoted to analysing structural change and to identifying the main drivers of this process in various farming contexts. The complexity of the process led to the development and application of a variety of modelling approaches such as cohort analysis, econometric models or multi-agent systems (see [Zimmermann *et al.* \(2009\)](#) for a review). As usual and as [Zimmermann *et al.* \(2009\)](#) discuss, each of these modelling approaches has pros and cons for the analysis of farm structural change depending on the objective pursued. Among these alternatives, the Markov chain model (MCM) has become a widely used approach to describe the way farm experience structural change. The Markov chain approach is particularly suitable for the analysis of structural change when this process is defined as changes in farm distribution across farm types or size categories. While other econometric models focus on one dimension of structural change only (entry, or exit, or farm growth, etc.), MCMs provide a comprehensive and consistent approach by considering different aspects of farm structural change at the same time, thus mirroring farm interdependence [Huettel and Margarian \(2009\)](#). In comparison to more structured modelling approaches such as multi-agent systems, Markov chain models also present the advantage to be less demanding in programming time and require less data ([Zimmermann *et al.*, 2009](#)).

The (relative) mathematical simplicity and stability properties of MCMs made the approach attractive for modelling dynamic random phenomena in economics ([Meyn and Tweedie,](#)

2009). In agricultural economics, several studies early proved the applicability of MCMs to analyse structural change in farming (*e.g.*, [Judge and Swanson \(1962\)](#); [Bostwick \(1962\)](#); [Krenz \(1964\)](#); [Colman \(1967\)](#)). Basically, the Markov model states that a farm type or size at a given date is the result of a probabilistic process which only depends on its type or size in the previous period since, in general, a first-order Markov process is assumed. While the MCM framework is essentially a statistical modelling approach, [Stokes \(2006\)](#) showed that a Markovian transition process may be derived from a structural model of inter-temporal profit maximisation, giving theoretical grounds to using the MCM. The author argues that if farmers choose inputs consistently with the objective of maximising their profit and face a Markovian-like uncertainty, then the development in farm size is also Markovian.

MCMs were first applied in their stationary version, *i.e.*, assuming that transition probabilities are invariant over time. Studies using this approach aimed in particular at recovering transition process of farms to perform projections over time. Since the stationarity assumption proved to be restrictive and inconsistent with the evolution of farm sizes ([Hallberg, 1969](#)), non-stationary versions of MCMs have been considered to allow transition probabilities to vary over time. This approach was particularly used to study the impact of specific factors on farm size change. Methodologically, most studies have used ‘aggregate’ data, that is, cross-sectional observations of the distribution of a farm population into a finite number of size categories because, first, such data are often easier to obtain than individual-level data, and second, [Lee *et al.* \(1965\)](#), [Lee *et al.* \(1970\)](#) and [MacRae \(1977\)](#) demonstrated that robustly estimating an MCM from such aggregate data is possible. More recently, because estimating an MCM using this type of data may well be an ill-posed problem (*i.e.*, 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 parameterise and estimate these models. Improvements range from the application of a discrete multinomial logit formulation (*e.g.*, [MacRae \(1977\)](#); [Zepeda \(1995b\)](#)), the maximisation of a generalised cross-entropy model with instrumental variables (*e.g.*, [Karantininis \(2002\)](#); [Huettel and Jongeneel \(2011\)](#); [Zimmermann and Heckelei \(2012\)](#)), a continuous re-parameterisation ([Piet, 2011](#)), to Bayesian inference ([Storm *et al.*, 2016](#)).

MCMs have been applied to investigate farm structural change in various agricultural contexts. Many studies focused on specific farms types. For example, several studies analysed structural change in dairy farming in different regions of the United States of America (USA) (*e.g.*, [Padberg \(1962\)](#); [Hallberg \(1969\)](#); [Stavins and Stanton \(1980\)](#); [Chavas and Magand \(1988\)](#); [Zepeda \(1995a\)](#); [Stokes \(2006\)](#)), in England and Wales (*e.g.*, [Colman \(1967\)](#); [Colman and Leech \(1970\)](#)) in Ireland (*e.g.*, [O’Dwyer \(1968\)](#); [Keane \(1976, 1991\)](#); [Gaffney \(1997\)](#); [McInerney and Garvey \(2004\)](#)), in Scotland (*e.g.*, [Buckwell *et al.* \(1983\)](#)), in Poland (*e.g.*, [Tonini and Jongeneel \(2009\)](#)), or in France (*e.g.*, [Butault and Delame \(2005\)](#); [Ben Arfa *et al.* \(2015\)](#)). Other analyses focus on other types of farming such as pig farms (*e.g.*, [Judge and Swanson](#)

(1962); Disney *et al.* (1988); Gillespie and Fulton (2001); Karantininis (2002)), grain farms (*e.g.*, Bostwick (1962); Garcia *et al.* (1987)), cotton farms (*e.g.*, Salkin *et al.* (1976); Ethridge *et al.* (1985)). Finally, some studies considered all farms at the same time (*e.g.*, Harrison and Alexander (1969); MacMillan *et al.* (1974); Edwards *et al.* (1985); Piet (2011)). Following Rahelizatovo and Gillespie (1999), Gillespie and Fulton (2001) and Zimmermann and Heckelei, 2012 conducted a cross-regional analysis to compare structural change in different regions in the USA and in the Western European Union (EU-15), respectively. Huettel and Jongeneel (2011) compared structural change in the Netherlands and West Germany by estimating two sub-models separately. All these studies found significant differences between the considered regions.

MCM applications described farm size changes using various size variables such as farm area, herd size, production potential or actual market sales. Different numbers of size categories have been also considered, ranging from three (*e.g.*, Chavas and Magand (1988); Zepeda (1995a)) up to 19 (*e.g.*, Karantininis (2002)). Recently, Piet (2011) developed a continuous approach which allows the reconstruction of transition probabilities for any number of size categories. When considered, entries and exits in farming are generally account for through an additional category representing net entries or exits (*i.e.*, total entries minus total exits), but some studies represented net new entries probabilities separately (*e.g.*, Chavas and Magand (1988); Zepeda (1995a)). In general, yearly transition probabilities are estimated, but larger time intervals have been also used. For example, Huettel and Jongeneel (2011) estimated farm transitions in Germany considering 22 years for both pre-quota period (1961-1983) and quota period (1985-2007)

Whatever the farm type, the region considered or the technique used, studies mostly predict a decrease in farm number for small and medium sized farms to the benefit of larger entities. Markov analyses also evidenced the impacts of several factors on structural change in farming. Among others, input and output prices, interest rate (proxy of cost of capital), technological change (using various proxies) and agricultural policies were found to play an important role in this process whatever the farm type or the region considered. In general, the input-output price ratio is found to increase net entry (entries minus exits) and to favour farm growth (*e.g.*, Disney *et al.* (1988); Massow *et al.* (1992); Karantininis (2002); Stokes (2006)). Technological change has been found to increase farm exits and encourage farm growth (*e.g.*, Ethridge *et al.* (1985); Zepeda (1995b); Zimmermann and Heckelei (2012)). In contrast, the impacts of specific agricultural policy measures are rather ambiguous. For example, Rahelizatovo and Gillespie (1999) and Stokes (2006) find a negative impact of the dairy termination programme on farm exit from dairy in Pennsylvania and Louisiana (USA), respectively, while Zepeda (1995a) finds an opposite effect for Wisconsin (USA). Likewise, Huettel and Jongeneel (2011) showed that EU milk quotas tend to increase farm growth Germany and Netherlands while Ben Arfa *et al.*

(2015) found a negative impact of the same policy on farm growth in France.

- **Some limits of the usual MCMs: farm heterogeneity and spatial interdependence**

Despite the usefulness of the MCMs presented above for the analysis of structural change in farming, applications so far face certain limits for a better understanding of the process. A first limitation of the MCM that should be tackled is the homogeneity assumption of the transition process of farms. Indeed, while much effort has been devoted to improve the specification and estimation techniques of MCM transition probabilities, much less attention has been dedicated to take into account farm heterogeneity in the analysis of farm structural change. Two reasons may explain why there has been so few studies that explicitly account for farm heterogeneity so far. Firstly, most previous studies focus on specific farm types such as dairy farming or hog production. Authors may therefore have considered that such groups are homogeneous enough to justify discarding heterogeneity issues. Secondly, studies that investigate the potential drivers of structural change in farming emphasise the role of economic factors such as market or policy variables. Authors have therefore plainly left heterogeneity issues in the background, considering it as a control variable rather than one of key interest. Furthermore, previous studies of farm structural change often faced data limitations. As mentioned earlier, structural change in farming is generally investigated using aggregate data because such data are easier to obtain than individual level data. Even when individual level data are used, yearly transition probabilities are estimated first for the overall population, then the impacts of causative factors are investigated on the resulting probabilities in a second step (*e.g.*, [Hallberg \(1969\)](#); [Ethridge et al. \(1985\)](#); [Smith \(1988\)](#); [Rahelizatovo and Gillespie \(1999\)](#)). It seems difficult if not impossible to take into account farm heterogeneity within such approaches.

However, it is clear that the process observed in general at a macro level first originates from farmers' individual decisions ([Bollman et al., 1995](#); [Jackson-Smith, 1999](#)). As farmers' behaviour may depend on specific farm and/or farmer characteristics, we argue that more attention should be paid to farm heterogeneity. While it is clear that farms may behave differently according to a variety of individual characteristics, controlling for all sources of heterogeneity in the analysis of farm structural change is challenging for two main reasons.

Firstly, heterogeneity in farming may originate from several sources. Farms display a huge diversity in terms of their physical and human capital and natural resource endowments, goals, and opportunity costs ([Babb, 1979](#)). In agricultural economics, [Zimmermann and Heckelei \(2012\)](#) is the only MCM application that explicitly attempted to take farm heterogeneity into account, to the best of our knowledge. Their application is however restrictive since the authors only considered the age of farmers to control for farm heterogeneity. Nonetheless, it seems

practically impossible to control for all sources of farm heterogeneity within such a methodology. In fact, this would considerably increase the total number of parameters to be estimated, thus decreasing the total number of degrees of freedom, which is by essence limited when using aggregate data.

Secondly, not all sources of heterogeneity can be observed or some of them cannot be fully linked to specific observed farm and/or farmer characteristics. Unobserved heterogeneity may originate from several sources that relate to, for example, the imperfect functioning of land and/or credit markets, or a differentiated accessibility to new technology and credit, to risk aversion or risk perception (Bowman and Zilberman, 2013; Conradt *et al.*, 2014; Trujillo-Barrera *et al.*, 2016). Depending on the characteristics that dominate, farmers' response to specific stimuli such as market price or public support programmes may not be the same (Key and Roberts, 2009; Mzoughi, 2011). For example, when considering size as measured in terms of the operated area, farmers may not all have the same opportunities to enlarge their farm simply because they may not be faced with the same land supply: because land is mostly released upon retirement, a farmer surrounded only by young colleagues will not find plots to buy or rent. Even if land offers exist, farmers may have different abilities to negotiate with land owners, may face unequal conditions to access credit or simply may not be willing to extend their business for personal reasons. Indeed, while farmers are generally supposed to maximise the expected profit from their farming activities, it has been evidenced that not all farmers give priority to economic motives (Maybery *et al.*, 2005; Howley *et al.*, 2014). For example, some non-competitive farms may keep their activity because of less or even non-financial/pecuniary motives (Harrington and Reinsel, 1995) perhaps enjoying the farming lifestyle values (Hallam, 1991). Some farms may also keep up production at sub-optimal levels because of the non-pecuniary benefits that they derive from their farming activities (Howley, 2015). This may be the case of some environmentally oriented farms (Willock *et al.*, 1999) or hobby farms (Daniels, 1986; Holloway *et al.*, 2002).

A second limitation of the usual MCMs is that they are unable to identify specific types of interactions between neighbouring entities despite their considering several dimensions of farm structural change at the same time. Indeed, understanding interactions between farms is very important in the analysis of structural change in farming particularly because of the fixity and immobility characteristics of land in the short to medium term (Storm *et al.*, 2015). However, it may exist different types of interactions between farms. For example, the existing literature distinguishes two types of effects of neighbouring farm size, which particularly originate from technology adoption. On the one hand, neighbours can be viewed as competitors for plots given land market regulations (Weiss, 1999). In this case, farmers surrounded by larger farms may be constrained to close their operation since larger farms are more likely to adopt new technologies earlier given their potentially greater access to information and better financial capacity (Goddard *et al.*, 1993). On the other hand, neighbours can be considered as a source of

motivation to adopt new technologies, which may lead to a positive impact of neighbouring farm size (Case, 1992; Holloway *et al.*, 2002). In addition to these two types of farms' relationships, one could postulate that neighbouring farm characteristics may not influence, or may have a limited impact on, decisions of some farmers since not all of them give priority to the commercial aspect of the farming activities, as mentioned above. For example, in the case of some environmentally oriented and hobby farms, new technology may be evaluated for different aspects other than financial viability before adoption. Then, the technology will be adopted only if it is considered as conformed to their expectations or interests (Mzoughi, 2011).

In any case, these types of rationality may not be fully linked to observable variables related to human capital such as education, to land market such as land prices, to credit market such as the interest rate, etc. As a result, even if they share the same observed characteristics, farms with an equal initial size may not experience structural change at the same speed or to the same extent. Ignoring these sources of heterogeneity using any modelling approach may lead to inconsistent parameters and unrealistic impact of structural change driving factors when aggregated at the population level (Pennings and Garcia, 2004; Holloway *et al.*, 2007). Therefore, the analysis of structural change in farming under the assumption of a homogeneous population (i.e., considering that all farmers will behave alike given a specific stimulus) may be inconsistent and may lead to unrealistic results or may be insufficient to explain the process at hand. The resulting conclusions and tentative recommendations may be therefore misleading. This situation raises the necessity to develop and apply modelling frameworks that allow capturing both observed and unobserved heterogeneity in the analysis of structural change in farming. As farm-level data become more widely available, allowing for the observation of individual transitions across time, controlling for unobserved heterogeneity both in transition processes and in relationships between farms, may leads to a better understanding of farm structural change and therefore to evidence more efficiently the impacts of causal factors.

1.4 Objectives and research questions

The main purpose of this thesis is to improve the understanding of the observed structural change in farming over the last decades. As it is in general the case in agricultural economics, this thesis concentrates on the variation of the total number of farms and the evolution of farm average size as indicators of farm structural change. More precisely, the thesis aims at: i) providing a discussion on the importance of accounting for farm heterogeneity in the analysis of structural change in agriculture; ii) showing that controlling for both observed and unobserved farm heterogeneity may lead to a better understanding and representation of structural change in farming; iii) identifying the main drivers of the observed trends of this process, with a focus on France.

The thesis adopts a mixture modelling framework to capture farm unobserved heterogeneity in modelling structural change. Finite mixture models present advantages over other modelling frameworks that can be used to capture unobserved heterogeneity, such as fixed and random effect models, individual or random parameter models, that make it a widely used approach (Becker *et al.*, 2015; Compiani and Kitamura, 2016). A review of this modelling approach and its application will be presented in the different chapters of the thesis in due places. Briefly, these models can be viewed as a special case of individual parameter models, where the parameters of the model are supposed to differ according to specific types (Greene, 2004). Finite mixture models, also called latent class models, are generally used to partition a population into homogeneous types to evidence heterogeneous behaviours. In particular, the mixture modelling framework is attractive because of its simplicity and intuitiveness (McLachlan and Peel, 2004). One of the main advantages of the mixture modelling approach is that it allows the data itself to sample select and gather observations into the most appropriate designations corresponding to the relationship between dependent and independent variables (Holloway *et al.*, 2007). The mixture modelling approach could therefore help separate farms with similar behaviours based on their main characteristics.

The Markov modelling framework has been chosen as the corner-piece approach of this thesis, and it is used in two out of the three chapters. As regards to the first and second specific objectives, the thesis thus intends to apply a mixture modelling approach that enables capturing unobserved heterogeneity under Markovian-like transition process of farms. To do this, a mixture Markov framework that captures heterogeneity in the transition process of farms is used to model farm size change over time. The objective is to: firstly, evidence a heterogeneous transition process in farming; and, secondly, investigate whether or not controlling for such unobserved heterogeneity using mixed-Markov chain models (M-MCMs) could help representing structural change in farming more efficiently than the homogeneous MCMs usually applied in agricultural economics. Therefore, the first research question is formulated as follows:

Question 1: How important is it to control for farm heterogeneity in the analysis of structural change in agriculture, and does a mixture modelling approach enable capturing unobserved heterogeneity under Markovian-like transition process of farms?

As regards to the second objective, the M-MCM is applied to investigate the impacts of causal factors on transition probabilities of farms. In doing so, the thesis intends to determine whether or not impacts of structural change factors depend on the specific unobserved types that farms belong to. A further objective is to quantify the impacts of recent agricultural policies on structural change, considering the presence of different unobservable types in the farm population. Then the second research question addressed by the thesis is formulated as follows:

Question 2: Do the impacts of causal factors of structural change in farming vary according to some unobservable farm types?

Since the Markov analysis is unable to identify or measure specific types of relationship between farms, the thesis also intends to apply a non-Markov mixture model to evidence heterogeneity in the spatial interdependence between farms. To understand structural change in a specific farming context, it seems important to identify the type of interactions that coexist and which farms are more likely to belong to these specific types. The third research question is thus:

Question 3: Do mixture models enable evidencing specific types of unobservable relationships or spatial interdependence between farms?

1.5 Main contributions

The contribution of this thesis is primarily empirical. However, its main contributions to the existing agricultural economics literature are twofold.

Firstly, from a methodological point of view, the thesis imports into agricultural economics a modelling approach which has been already used in other strands of the economic literature, so as to specifically account for potential unobserved heterogeneity in the process of farm structural change. The thesis thus highlights some well-known limitations of the traditional MCMs generally applied to analyse structural change in farming. In particular, because an assumption of homogeneity in the transition process of farms underlies the MCMs used so far, the results from such models could be misleading because this assumption is unrealistic in practice. As a more general framework, the mixture modelling approach applied in the thesis allows identifying potential different groups of similar behaviours among the farm population. The M-MCM model is suitable to capture unobserved heterogeneity in the transition process of farms without losing the flexibility of the usual MCM. It thus enables recovering the data-generating process more efficiently than the homogeneous MCM and therefore deriving more accurately the long-run transition matrix and forecasting future farm size distributions. Likewise, the mixture of probit and mixture of regression models applied to represent farm survival and growth allow evidencing unobserved heterogeneity in the neighbouring relationships between farms. The resulting impacts of structural change drivers are found to be significantly different from the mean effect obtained under the homogeneous assumption when aggregated at the overall population level.

Secondly, from an empirical point of view, the thesis shows that there exist two main types of farms in the French agriculture: farms which tend to remain in the same category of size so that they can be identified as ‘likely stayers’; and farms which are more likely to change category

of sizes, namely ‘likely movers’. As it has been shown in other strands of economics, ignoring such a difference in the process of farm size transition generally leads to overestimate farms’ mobility and therefore to inaccurate projections of farm distribution across size categories. Additionally, the thesis results show that the impact of factors affecting farm structural change, including agricultural policies, depends on unobserved characteristics of farms and farmers that shape their behaviours. This finding is especially relevant for agricultural policy assessment, suggesting that considering an ‘average farmer’ to design policy measures may be inadequate given potential sources of farm heterogeneity. The thesis also identifies specific types of relationships between farms such as competition for land and positive spill overs of new technology adoption, which have different impacts on their probability to remain in the sector and change their operated size. The thesis also evidences neutral relationships between neighbouring farms. This shows the importance to take into account such types of relationships between farms in analysing factors that affect structural change.

The empirical findings in this thesis show that accounting for unobserved farm heterogeneity is important when analysing farm structural change, and that a mixture modelling framework is relevant to do so. However, while we could put forward some theoretical explanations for the presence of unobserved heterogeneity, to our knowledge, there is, as yet, no formal framework in agricultural economics that could help identifying the potential sources of such heterogeneity and guide the empirical measurement of how they may affect the transition process of farms. We hope that the results presented in the thesis will contribute to encouraging the development of such theoretical frameworks for a better understanding of structural change under the assumption of heterogeneous farm populations.

1.6 Outline of the thesis

The thesis is organised into five chapters including this general introduction and a concluding chapter. It builds on three research articles which have been prepared during the course of the PhD and which have been extended to constitute the thesis central chapters. In doing so, chapters may overlap to some extent but all of them are consistent with the main objective of the thesis to incorporate potential farm heterogeneity in the analysis of structural change in farming.

Chapter 2 addresses the first research question of the thesis. It presents a mixture modelling approach that allows evidencing unobserved heterogeneity in the process of farm size transition. A restricted version of the mixed-Markov chain model (M-MCM), the so-called ‘mover-stayer’ model (MSM), is applied to analyse structural change in the farming context. The model is applied to a sample of French farms from the ‘Réseau d’Information Comptable Agricole’ (RICA) surveyed from 2000 to 2013. A comparison with the usual homogeneous

Markov chain model (MCM) is performed to demonstrate the usefulness of the applied mixture modelling approach to capture unobserved farm heterogeneity.

Chapter 3 investigates the second research question of the thesis. In this chapter, the generalised M-MCM is applied to identify the main drivers of farm structural change. From a methodological point of view, this chapter therefore generalises chapter 2 and adds the analysis of potential explanatory variables for the transition probabilities. A special attention is paid to the impacts of agricultural policy measures. RICA data from 2000 to 2013 are also used for the empirical application. The results from the M-MCM are compared to those obtained under the assumption of a homogenous farm population.

Chapter 4 addresses the third research question. This chapter deals with both farm heterogeneity and the spatial interdependency between neighbouring farms. From a methodological point of view, it takes advantage of the M-MCM methods developed in the previous chapters to extend an econometric model which also accounts for multi-level spatial variables. An application to a panel of French farms located in Brittany surveyed by the ‘Mutualité Sociale Agricole’ (MSA) is performed to evidence different relationships between farms and how this may affect farm survival and growth in this specific region of France.

The last chapter draws some concluding comments along with a summary of the main findings of the thesis. It also underlines the limits of the proposed analyses and provides some suggestions for further research.

Bibliography

- AGRESTE (2012a). *A Handbook of Agricultural Statistics*. Ministry of Agriculture, Agrifood and Forests. Department of Statistics and Foresight Analysis (SSP), 2012th ed.
- AGRESTE (2012b). Structures et moyens de production agricoles. *GraphAgri* 2: 17–24.
- AGRESTE (2014). Le statut juridique des exploitations agricoles évolutions 1970-2010. *Agreste Les Dossiers. No 20* .
- Akimowicz, M., Magrini, M.-B., Ridier, A., Bergez, J.-E. and Requier-Desjardins, D. (2013). What Influences Farm Size Growth? An Illustration in Southwestern France. *Applied Economic Perspectives and Policy* 35: 242–269.
- Albrecht, D. E. (1997). The changing structure of US agriculture: dualism out, industrialism in. *Rural Sociology* 62: 474.
- Babb, E. (1979). Some causes of structural change in us agriculture. *US Department of Agriculture, Economics, Statistics, and Cooperative Service, Structure Issues of American Agriculture. Report 438*: 51–60.
- Barthélémy, D. and Dussol, A.-M. (2002). Sociétés agricoles: entre modernité et tradition. *Agreste Cahiers* 2: 31–36.
- Bary, C. (2012). Commercialisation des produits agricoles. un producteur sur cinq vend en circuit court. *Agreste Primeur. No 275* .
- Becker, J.-M., Ringle, C. M., Sarstedt, M. and Völckner, F. (2015). How collinearity affects mixture regression results. *Marketing Letters* 26: 643–659.
- Ben Arfa, N., Daniel, K., Jacquet, F. and Karantininis, K. (2015). Agricultural Policies and Structural Change in French Dairy Farms: A Nonstationary Markov Model. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroéconomie* 63: 19–42.
- Blanc, M., Cahuzac, E., Elyakime, B. and Tahar, G. (2008). Demand for on-farm permanent hired labour on family holdings. *European Review of Agricultural Economics* 35: 493–518.
- Boehlje, M. (1992). Alternative models of structural change in agriculture and related industries. *Agribusiness* 8: 219–231.
- Bollman, R. D., Whitener, L. A. and Tung, F. L. (1995). Trends and patterns of agricultural structural change: A Canada–US comparison. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroéconomie* 43: 15–28.

- Bostwick, D. (1962). *Yield probabilities as a Markov process*. Marketing Economics Division, Economic Research Service, US Department of Agriculture.
- Bowman, M. S. and Zilberman, D. (2013). Economic factors affecting diversified farming systems. *Ecology and Society* 18: 33.
- Breustedt, G. and Glauben, T. (2007). Driving forces behind exiting from farming in Western Europe. *Journal of Agricultural Economics* 58: 115–127.
- Buckwell, A. E., Shucksmith, D. and Young, D. (1983). Structural projections of the Scottish dairy industry using micro and macro Markov transition matrices. *Journal of Agricultural Economics* 34: 57–68.
- Butault, J.-P. and Delame, N. (2005). Concentration de la production agricole et croissance des exploitations. *Economie et statistique* 390: 47–64.
- Cahuzac, É. and Détang-Dessendre, C. (2011). Le salariat agricole. une part croissante dans l'emploi des exploitations mais une précarité des statuts. *Economie rurale* : 82–92.
- Carlin, T. A. and Saupe, W. E. (1990). *Structural change in farming and its relationship to rural communities*. University of Wisconsin, Department of Agricultural Economics.
- Case, A. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics* 22: 491–508.
- Chavas, J.-P. (2001). *Chapter 5 Structural change in agricultural production: Economics, technology and policy*. Elsevier, *Volume 1, Part A*. 263–285.
- Chavas, J.-P. and Magand, G. (1988). A dynamic analysis of the size distribution of firms: The case of the US dairy industry. *Agribusiness* 4: 315–329.
- Colman, D. (1967). The Application of Markov Chain Analysis to Structural Change in the North West Dairy Industry. *Journal of Agricultural Economics* 18: 351–363.
- Colman, D. and Leech, D. (1970). A forecast of milk supply in England and Wales. *Journal of Agricultural Economics* 21: 253–265.
- Compiani, G. and Kitamura, Y. (2016). Using mixtures in econometric models: A brief review and some new results. *The Econometrics Journal* Forthcoming.
- Conradt, S., Bokusheva, R., Finger, R., Kussaiynov, T. *et al.* (2014). Yield trend estimation in the presence of farm heterogeneity and non-linear technological change. *Quarterly Journal of International Agriculture* 53: 121–140.
- Daniels, T. L. (1986). Hobby farming in America: Rural development or threat to commercial agriculture? *Journal of Rural Studies* 2: 31–40.

- Darpeix, A., Bignebat, C. and Perrier-Cornet, P. (2014). Demand for Seasonal Wage Labour in Agriculture: What Does Family Farming Hide? *Journal of Agricultural Economics* 65: 257–272.
- Desriers, M. (2007). L'agriculture française depuis cinquante ans: des petites exploitations familiales aux droits à paiement unique. *Agreste cahiers* 2: 3–14.
- Disney, W. T., Duffy, P. A. and Hardy, W. (1988). A Markov chain analysis of pork farm size distributions in the South. *Southern Journal of Agricultural Economics* 20: 57–64.
- Eastwood, R., Lipton, M. and Newell, A. (2010). Farm size. *Handbook of agricultural economics* 4: 3323–3397.
- Edwards, C., Smith, M. G. and Peterson, R. N. (1985). The changing distribution of farms by size: A Markov analysis. *Agricultural Economics Research* 37: 1–16.
- EESC (2004). Opinion of the European Economic and Social Committee on ‘the agricultural employment situation in the EU and the accession countries: options for action for 2010’. *Official Journal of the European Union* .
- Ethridge, D. E., Roy, S. K. and Myers, D. W. (1985). A Markov chain analysis of structural changes in the Texas high plains cotton ginning industry. *Southern Journal of Agricultural Economics* 17: 11–20.
- Fennell, R. (1981). Farm succession in the European Community. *Sociologia Ruralis* 21: 19–42.
- Flaten, O. (2002). Alternative rates of structural change in Norwegian dairy farming: impacts on costs of production and rural employment. *Journal of Rural Studies* 18: 429 – 441.
- Gaffney, P. (1997). A projection of Irish agricultural structure using Markov chain analysis. Tech. rep., CAPRI Working Paper 97-10.
- Garcia, P., Offutt, S. E. and Sonka, S. T. (1987). Size distribution and growth in a sample of Illinois cash grain farms. *American Journal of Agricultural Economics* 69: 471–476.
- Gillespie, J. M. and Fulton, J. R. (2001). A Markov chain analysis of the size of hog production firms in the United States. *Agribusiness* 17: 557–570.
- Goddard, E., Weersink, A., Chen, K. and Turvey, C. G. (1993). Economics of Structural Change in Agriculture. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 41: 475–489.
- Greene, W. (2004). Convenient estimators for the panel probit model: Further results. *Empirical Economics* 29: 21–47.

- Hallam, A. (1991). Economies of size and scale in agriculture: An interpretive review of empirical measurement. *Review of Agricultural Economics* 13: 155–172.
- Hallberg, M. C. (1969). Projecting the size distribution of agricultural firms-an application of a Markov process with non-stationary transition probabilities. *American Journal of Agricultural Economics* 51: 289–302.
- Harrington, D. H. and Reinsel, R. D. (1995). A synthesis of forces driving structural change. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 43: 3–14.
- Harrison, H. and Alexander, D. (1969). Projection of farm numbers and farm-size structure with Markov chains. *Irish Journal of Agricultural Economics and Rural Sociology* 2: 105–116.
- Holloway, G., Lacombe, D. and LeSage, J. P. (2007). Spatial econometric issues for bio-economic and land-use modelling. *Journal of Agricultural Economics* 58: 549–588.
- Holloway, G., Shankar, B. and Rahmanb, S. (2002). Bayesian spatial probit estimation: a primer and an application to hyv rice adoption. *Agricultural Economics* 27: 383–402.
- Howley, P. (2015). The Happy Farmer: The Effect of Nonpecuniary Benefits on Behavior. *American Journal of Agricultural Economics* 97: 1072–1086.
- Howley, P., Dillon, E. and Hennessy, T. (2014). It s not all about the money: understanding farmers labor allocation choices. *Agriculture and Human Values* 31: 261–271.
- Huettel, S. and Jongeneel, R. (2011). How has the EU milk quota affected patterns of herd-size change? *European Review of Agricultural Economics* 38: 497–527.
- Huettel, S. and Margarian, A. (2009). Structural change in the West German agricultural sector. *Agricultural Economics* 40: 759–772.
- Jackson-Smith, D. B. (1999). Understanding the Microdynamics of Farm Structural Change: Entry, Exit, and Restructuring among Wisconsin Family Farmers in the 1980s. *Rural Sociology* 64: 66–91.
- Jones, W., Freshwater, D. and Fiarchuk, R. (1995). Why the structure of agriculture still matters to national governments. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 43: 223–240.
- Judge, G. G. and Swanson, E. R. (1962). Markov chains: basic concepts and suggested uses in agricultural economics. *Australian Journal of Agricultural Economics* 6: 49–61.
- Karantininis, K. (2002). Information-based estimators for the non-stationary transition probability matrix: An application to the Danish pork industry. *Journal of Econometrics* 107: 275–290.

- Keane, M. (1976). Projections of change in creamery supplier structure. *Irish Journal of Agricultural Economics and Rural Sociology* : 177–184.
- Keane, M. (1991). Changes in the size structure of Irish dairy farms. *Irish Journal of Agricultural Economics and Rural Sociology* : 67–74.
- Kennedy, L. (1991). Farm succession in modern Ireland: elements of a theory of inheritance. *The Economic History Review* 44: 477–499.
- Key, N. and Roberts, M. J. (2009). Nonpecuniary benefits to farming: Implications for supply response to decoupled payments. *American Journal of Agricultural Economics* 91: 1–18.
- Krenz, R. D. (1964). Projection of farm numbers for North Dakota with Markov chains. *Agricultural Economics Research* 16: 77–83.
- Lee, T., Judge, G. and Zellner, A. (1970). *Estimating the parameters of the Markov probability model from aggregate time series data*. Amsterdam: North Holland.
- Lee, T. C., Judge, G. G. and Takayama, T. (1965). On estimating the transition probabilities of a Markov process. *Journal of Farm Economics* 47: 742–762.
- Lerbourg, J. and Dedieu, M.-S. (2016). L'équipement des exploitations agricoles: Un recours à la propriété moins marqué pour les machines spécialisées. *Agreste Primeur* 334.
- MacMillan, J. A., Tung, F. and Tulloch, J. R. (1974). Migration analysis and farm number projection models: a synthesis. *American Journal of Agricultural Economics* 56: 292–299.
- MacRae, E. C. (1977). Estimation of time-varying Markov processes with aggregate data. *Econometrica* 45: 183–198.
- Mann, S., Mittenzwei, K. and Hasselmann, F. (2013). The importance of succession on business growth: A case study of family farms in Switzerland and Norway. *Yearbook of Socioeconomics in Agriculture* 12: 109–38.
- Massow, M., Weersink, A. and Turvey, C. G. (1992). Dynamics of structural change in the ontario hog industry. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 40: 93–107.
- Maybery, D., Crase, L. and Gullifer, C. (2005). Categorising farming values as economic, conservation and lifestyle. *Journal of Economic Psychology* 26: 59 – 72.
- McInerney, N. and Garvey, E. (2004). Farm structure and agricultural labour. Tech. rep., CAPRI Working Papers. Dept of Economics, NUI Galway.
- McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

- Meyn, S. P. and Tweedie, R. L. (2009). *Markov chains and stochastic stability*. Cambridge University Press.
- Mzoughi, N. (2011). Farmers adoption of integrated crop protection and organic farming: Do moral and social concerns matter? *Ecological Economics* 70: 1536–1545.
- O’Dwyer, T. (1968). Structural changes in milk supply in a cooperative creamery area. *Irish Journal of Agricultural Economics and Rural Sociology* 1: 207–219.
- Padberg, D. I. (1962). The use of Markov processes in measuring changes in market structure. *Journal of Farm Economics*. 44: 189–199.
- Pennings, J. M. and Garcia, P. (2004). Hedging behavior in small and medium-sized enterprises: The role of unobserved heterogeneity. *Journal of Banking & Finance* 28: 951–978.
- 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).
- Pritchard, B., Burch, D. and Lawrence, G. (2007). Neither ‘family’ nor ‘corporate’ farming: Australian tomato growers as farm family entrepreneurs. *Journal of rural studies* 23: 75–87.
- Rahelizatovo, N. C. and Gillespie, J. M. (1999). Dairy farm size, entry, and declining production region. *Journal of Agricultural and Applied Economics* 31: 333–347.
- Rattin, S. (2003). Recrudescence du temps partiel chez les salariés permanents non familiaux. *Agreste* 3: 9–15.
- Rattin, S. (2009). Evolution des structures: l’exploitation française est devenue une entreprise. *Demeter-Economie et stratégies agricoles* : 197–221.
- Roussillon-Montfort, M.-A. (2008). Les marchés à terme agricoles en europe et en france. *Notes et Études Économiques* 30: 99–124.
- Salkin, M., Just, R. and Cleveland, J., O.A. (1976). Estimation of nonstationary transition probabilities for agricultural firm size projection. *The Annals of Regional Science* 10: 71–82.
- Schutz-Stephan, N. and Savy, H. (2010). La diversification des statuts juridiques des exploitations comme outil de politique agricole. *Pour* 1: 75–84.
- Smith, M. G. (1988). A Conditional Approach to Projecting Farm Structure. Tech. rep., Economic Research Service (DOA), Washington, DC. Agricultural and Rural Economics Division.
- Stanton, B. F. (1991). *Farm Structure: Concept and Definition*. Department of Agricultural Economics, Cornell University Agricultural Experiment Station, New York State College of Agriculture and Life Sciences, Cornell University.

- Stavins, R. N. and Stanton, B. F. (1980). *Alternative Procedures for Estimating the Size Distribution of Farms*. Department of Agricultural Economics, New York State College of Agriculture and Life Sciences.
- Stokes, J. R. (2006). Entry, exit, and structural change in Pennsylvania's dairy sector. *Agricultural and Resource Economics Review* 35: 357–373.
- Storm, H., Heckelei, T. and Mittelhammer, R. C. (2016). Bayesian estimation of non-stationary Markov models combining micro and macro data. *European Review of Agricultural Economics* 43: 303–329.
- Storm, H., Mittenzwei, K. and Heckelei, T. (2015). Direct Payments, Spatial Competition, and Farm Survival in Norway. *American Journal of Agricultural Economics* 97: 1192–1205.
- Tonini, A. and Jongeneel, R. (2009). The distribution of dairy farm size in poland: a Markov approach based on information theory. *Applied Economics* 41: 55–69.
- Trujillo-Barrera, A., Pennings, J. M. E. and Hofenk, D. (2016). Understanding producers' motives for adopting sustainable practices: the role of expected rewards, risk perception and risk tolerance. *European Review of Agricultural Economics* 43: 359–382.
- Vavra, P. (2009). Role, usage and motivation for contracting in agriculture. *OECD Food, Agriculture and Fisheries Papers* OECD Publishing.
- Vidal, C. (2001). Thirty years of agriculture in Europe: Changes in agricultural employment. Report.
- Weiss, C. R. (1999). Farm Growth and Survival: Econometric Evidence for Individual Farms in Upper Austria. *American Journal of Agricultural Economics* 81: 103–116.
- Willock, J., Deary, I. J., Edwards-Jones, G., Gibson, G. J., McGregor, M. J., Sutherland, A., Dent, J. B., Morgan, O. and Grieve, R. (1999). The Role of Attitudes and Objectives in Farmer Decision Making: Business and Environmentally-Oriented Behaviour in Scotland. *Journal of Agricultural Economics* 50: 286–303.
- Zepeda, L. (1995a). Asymmetry and nonstationarity in the farm size distribution of Wisconsin milk producers: An aggregate analysis. *American Journal of Agricultural Economics* 77: 837–852.
- Zepeda, L. (1995b). Technical change and the structure of production: a non-stationary Markov analysis. *European Review of Agricultural Economics* 22: 41–60.
- Zimmermann, A. and Heckelei, T. (2012). Structural change of European dairy farms: A cross-regional analysis. *Journal of Agricultural Economics* 63: 576–603.

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

Appendix

TABLE 1.4: Some empirical applications of the Markov chain models using individual-level data

Year	Author	Region	Field of study	modelling approach	Estimation method	Size variable	Categories	Explanatory variables	Transition
1961	Judge and Swanson	Illinois, USA	Pig farms	Stationary	Maximum likelihood	Number of litters of hogs	6 + entry/exit		Annual
1962	Bostwick	Hjort, USA	Wheat	Stationary	Maximum likelihood	yield to cover cash costs	5	–	Annual
1962	Padberg	California, USA	Milk industry	Stationary	Maximum likelihood	Market share	3 + entry/exit	–	5 years
1967	Colman	Northwest, England	Dairy	Stationary		Herd size in heads	6	–	Annual
1967	Stanton and Ketunen	New York State (USA)	Dairy	Stationary	Maximum likelihood	Herd size in heads	4	–	4 years
1969	Hallberg	Pennsylvania, USA	Milk products plants	Stationary, non-stationary	Maximum likelihood + least squares (2-step)	Sales volume	4 + entry1/exit	Wages, population, per capita income, farm-gate price for milk, retail price	Annual
1976	Salkin <i>et al.</i>	Oklahoma, USA	Cotton warehouses	Stationary, non-stationary	Least squares + geometric model (2-step)	warehouse capacity	5 + entry1/exit	Time	Annual
1980	Stavins and Stanton	New York, USA	Dairy farms	Stationary, 2-step non-stationary	Maximum likelihood + multinomial logit (2-step)	Milk supply	9 + entry1/exit	Milk-feed price ratio	S: 9 years, N-S: annual
1983	Buckwell <i>et al.</i>	Scotland	Dairy	Stationary		Herd size in heads	7	3 years	
1985	Ethridge <i>et al.</i>	West Texas, USA	Cotton gin firms	Stationary, non-stationary	Maximum likelihood + least squares (2-step)	Gin capacity	12 + new entrants, dead gin firms, inactive class	Wages, energy costs, plant capacity, technical change	Annual
1985	Edwards <i>et al.</i>	USA	All types	Stationary	Maximum likelihood	Acres	8+entry/exit	–	4 years
1987	Garcia <i>et al.</i>	Illinois USA	Cash grain farms	Stationary	Maximum likelihood	gross value of farm product/tillable acres	11	–	9

TABLE 1.5: Some empirical applications of the Markov chain models using aggregate data

Year	Author	Region	Farms	Modelling approach	Estimation method	Size variable	Categories	Explanatory variables	Transition
1964	Krenz	North Dakota, USA	All types	Stationary	Maximum likelihood	Acres	6 + entry/exit	–	5 years
1969	Harrison and Alexander	Ireland	All types	Stationary	Maximum likelihood	Standard man-days	5 + exits	–	6 years
1974	MacMillan <i>et al.</i>	Canadian Prairie Provinces	All types	Stationary	quadratic programming	Gross receipts	6 + exits	–	5 years
1976	Keane	South of Ireland	Dairy farms	Stationary	Maximum likelihood	Milk supply	6 + entry/exit	–	Annual
1983	Buckwell <i>et al.</i>	Scotland	Dairy	Stationary	Maximum likelihood	Herd size	6 + entry/exit	–	3 years
1988	Disney <i>et al.</i>	Southern states, USA	Pig farms	Stationary, non-stationary	Minimum absolute deviation	Sales market hogs/year	4 + entry/exit	Hog-corn price ratio	4-5 years
1988	Chavas and Magand	Different regions, USA	Dairy farms	Non-stationary	Multinomial logit	Herd size	4 Net entry ¹	Economies of size, sunk costs, market prices	Annual
1991	Keane	Ireland	Dairy farms	Stationary	Maximum likelihood	Milk supply	7 + entry/exit	–	6 years
1992	Von Massow <i>et al.</i>	Ontario, Canada	Pig farms	Stationary, non-stationary	Minimization of median absolute deviation	number of hogs marketed	5 + entry/exit	Hog-corn price ratio, interest rate, labour capital price ratio	Annual
1995a	Zepeda	Wisconsin, USA	Dairy farms	Non-stationary	Multinomial logit	Herd size	3 + entry/exit	Milk-feed price ratio, interest rate, dairy termination program	Annual
1995b	Zepeda	Wisconsin, USA	Dairy farms	Non-stationary	Multinomial logit	Herd size	4	Milk production per cow	Annual
1997	Gaffney	Ireland	Dairy, cattle, hogs, sugar beet and cereals	Stationary	Maximum likelihood	Herd size in heads, hectares	5 to 7	–	12 years
2001	Gillespie and Fulton	17 US states	Pig farms	Non-stationary	Multinomial logit	Number of hogs	3 + net entry	Regional dummies, hog-corn price ratio, interest rate, corporate farm laws, meat processing capacity, percentage of land in farms	Annual
2002	Karantininis	Denmark	Pig farms	Stationary, non-stationary	GCE, IV-GCE	Number of hogs	18 + entry/exit	Input and output prices of pork and other livestock, fertilizer prices, interest rate	Annual

TABLE 1.6: Some empirical applications of the Markov chain models using aggregate data (continued)

Year	Author	Region	Farms	Modelling approach	Estimation method	Size variable	Categories	Explanatory variables	Transition
2002	Jongeneel	Netherlands	Dairy farms	Non-stationary	GCE + linear explanation function	herd size	3 + inactive	Milk output, milk price, policy dummy, trend	Annual
2004	McInerney and Garvey	Ireland	Dairy	Stationary	Maximum likelihood	Herd size	6+ net exits	–	12 years
2005	Jongeneel <i>et al.</i>	Netherlands, Germany, Poland, Hungary	Dairy farms	Non-stationary	IV-GCE	Herd size	NL, E-DE, H: 7 + entry/exit, W-DE: 6 + entry/exit, PL: 4 + entry/exit	Trend, milk output, milk price, quota auction dummy	Annual
2006	Stokes	Pennsylvania, USA	Dairy farms	Non-stationary	GCE + SUR (2-step)	Herd size	6 + entry/exit	Milk price, milk price volatility, productivity, interest rates, land values, policy dummy for exit probabilities	Annual
2008	Tonini and Jongeneel	Poland	Dairy farms	Non-stationary	IV-GCE	Herd size	8 + entry/exit	Trend	Annual
2008	Piet	France	All types	Stationary	Non-linear least-squares	Utilised Agricultural Area	Continuous	–	1-2-5-7-10-15
2011	Huettel and Jongeneel	Germany, Netherlands	Dairy farms	Stationary	GCE	Herd size	W-G: 3+ inactive, NL: 4 + inactive	–	WG: 22 NL: 11-22
2011	Piet	EU-15, USA	All types	Stationary	Continuous re-parametrisation	UE: E SU, USA: gross sales;	Continuous	–	Annual
2014	Ben Arfa et al.		Dairy farms	Non-stationary	GCE	Herd size	7 + entry/exit	Price of cow milk, Yield of cow milk, Direct payments, Milk quota, young farmers financial aid	Annual

Chapter 2

Movers and stayers in the farming sector: accounting for unobserved heterogeneity in structural change¹

Abstract

This chapter investigates whether accounting for unobserved heterogeneity in farms' size-transition process improves the representation of structural change in agriculture. Considering a mixture of two types of farms, the mover-stayer model (MSM) is applied for the first time in an agricultural economics context. The maximum-likelihood method and the expectation-maximisation (EM) algorithm are used to estimate the model's parameters. An empirical application to a panel of French farms from 2000 to 2013 shows that the MSM outperforms the homogeneous Markov chain model (MCM) in recovering the transition process and predicting the future distribution of farm sizes.

2.1 Introduction

In most developed countries, the number of farms has decreased sharply and their average size has increased continually over recent decades. Structural change has long been a subject of considerable interest to agricultural economists and policy makers. Previous studies aimed in particular to understand the mechanisms underlying this process in order to identify the key drivers that influence the observed trends, and to generate prospective scenarios. This study

¹This chapter has been written in collaboration with Laurent Piet (INRA, SMART-LERECO, France). A shorter version is published in the Journal of the Royal Statistical Society, Series 'Applied Statistics'.

aims at making a methodological contribution to the literature. More specifically, the aim of the research was to investigate whether a mixture modelling approach could account for unobserved heterogeneity in the size-transition process of farms, and in doing so, improve the estimation and prediction of structural change in agriculture.

As [Zimmermann *et al.* \(2009\)](#) show in their review, it has become quite common in agricultural economics to study the way farms experience size change by applying the so-called Markov chain model (MCM) to aggregate series of farm numbers by size categories (see [Huettel and Jongeneel \(2011\)](#), [Zimmermann and Heckelei \(2012\)](#) or [Ben Arfa *et al.* \(2015\)](#) for recent examples). While the MCM framework is essentially a statistical modelling approach, [Stokes \(2006\)](#) showed that a Markovian transition process may be derived from a structural model of inter-temporal profit maximisation, giving theoretical grounds to using the MCM. Basically, this model states that a farm's size at a given date is the result of a probabilistic process which only depends on its size in the previous period since, in general, a first-order Markov process is assumed. In its simplest version, the MCM framework assumes stationarity and homogeneity, *i.e.*, it assumes that transition probabilities are invariant over time and that all farms move from one category to another according to the same unique stochastic process, conditional on their initial size. While MCMs were first developed in this simplest form, two directions have been investigated to improve the modelling framework when applied to farming. Firstly, assuming that farms' transition probabilities may vary over time, non-stationary MCMs have been developed to investigate the effects of time-varying variables such as agricultural policies on farm structural changes (see [Zimmermann *et al.* \(2009\)](#) for a review). Secondly, assuming that the transition process may differ depending on farm and/or farmer characteristics, some studies have accounted for farm heterogeneity in modelling structural change. However, to date, to the best of our knowledge, only observed heterogeneity (such as regional location, type of farming, legal status or age group) has been considered, implying that all farms sharing the same observed characteristics follow the same stochastic process (see [Zimmermann and Heckelei \(2012\)](#) for a recent example).

We argue that structural change in agriculture may also pertain to individual characteristics or behaviours of farms and/or farmers which are not observable in practice. From a theoretical point of view, this unobserved heterogeneity may originate from several sources that relate to, for example, the imperfect functioning of land and/or credit markets, or farmer preferences. For example, when considering size as measured in terms of the operated area, farmers may not all have the same opportunities to enlarge their farm simply because they may not be faced with the same land supply: because land is mostly released upon retirement, a farmer surrounded only by young colleagues will not find plots to buy or rent. Even if land offers exist, farmers may have different abilities to negotiate with land owners, may face unequal conditions to access credit, or simply may not be willing to extend their business for personal

reasons. In any case, such types of rationale of these types may not be fully linked to observable variables related to human capital such as education, to land market such as land prices, to credit market such as the interest rate, etc. As a result, even if they share the same observed characteristics, farms with an equal initial size may not experience structural change at the same speed or to the same extent. Therefore, as farm-level data have become more widely available, accounting for unobserved heterogeneity in farms' size-transition process may allow the underlying data-generating process to be recovered in a more efficient way than with a simple homogeneous MCM. In this study, we investigate whether or not this is the case by comparing two alternative modelling frameworks, namely the homogeneous MCM, which does not account for heterogeneity, and the mover-stayer model (MSM), which accounts for unobserved heterogeneity.

The MSM extends the MCM framework by considering two types of agents following two different transition processes: the 'movers' who follow a first-order Markovian process and the 'stayers' who never experience transition. Because it derives from the MCM, the MSM exhibits the same mathematical simplicity and stability properties that have made Markovian approaches attractive for modelling dynamic random phenomena in economics (Meyn and Tweedie, 2009). However since it has been shown that economic transition processes are often not 'purely' Markovian (Bickenbach and Bode, 2001), the MSM enables relaxing the homogeneity assumption underlying the MCM. The MSM has thus allowed evidencing the presence of unobserved heterogeneity in several economic issues such as social (Singer and Spilerman, 1974, 1976), labour (Blumen *et al.*, 1955; Fougere and Kamionka, 2003, 2008), or income mobility (Shorrocks, 1976; Dutta *et al.*, 2001), financial rating (Frydman and Kadam, 2004), or firm-size dynamics (Cipollini *et al.*, 2012).

There are two reasons why there have been so few studies that explicitly account for farm heterogeneity. Firstly, most previous studies focus on specific farm types such as dairy farming or hog production. Authors may therefore have assumed that such groups are homogeneous enough to justify discarding heterogeneity issues. Secondly, studies that investigate the potential drivers of structural change in farming emphasise the role of economic factors such as market or policy variables. Authors have therefore plainly left heterogeneity issues in the background, considering it as a control variable rather than one of key interest. While our implementation of the MSM compares with similar studies in other strands of economics, the primary contribution of our work is to demonstrate that unobserved heterogeneity is also present in the process of farm structural change, and that the MSM also constitutes a relevant model in this specific sector. In our case, the studied agents are farms and we investigate how their size, measured in economic terms, evolves over time. 'Mover'-farms change size according to a first-order Markovian process, while 'stayer'-farms always remain in their initial size category. In our view, considering this 'stayer' type of farmers is justified because, since structural change in agriculture is a long-term process and because of the potential sources of unobserved heterogeneity listed above, some

farms may experience no size change at all or no sufficiently large size changes to move to another size category for a long enough period of time to qualify as ‘stayers’. To establish our result, we compare the respective strength of both the MCM and the MSM using a panel of 17,285 commercial French farms observed during the period 2000-2013. We estimate transition probability matrices and perform short- to long-term out-of-sample projections of farm-size distribution for both models, which allows us to compare their respective performance, firstly, in predicting farms’ size-transition probabilities and, secondly, in performing farm-size distribution forecasts over time.

This chapter is structured as follows: in the first section we introduce the way in which the homogeneous MCM can be generalised into the M-MCM and how the specific MSM is derived. In the next two sections, we describe the method used to estimate MSM parameters and the two measures, namely the likelihood ratio (LR) and the average marginal error (AME), used to compare the respective performances of the MCM and the MSM. We then report our application to France, starting with a description of the data used and following with a presentation of the results. Finally, we conclude with some considerations on how further to extend the approach described here.

2.2 Modelling farm transition process using the Markov chain framework

By definition, a Markov chain is a stochastic process (*i.e.*, the outcomes may be subjected to probabilistic analysis) characterised by the Markov property and where the outcome takes values on a finite state space (Lee *et al.*, 1970; Berchtold, 1998; Meyn and Tweedie, 2009). A process is Markovian if it is memoryless, that is, future states are independent from the past given the present (Meyn and Tweedie, 2009). A finite Markov process can evolve through discrete time steps or continuously along a certain time interval. This leads to two different stochastic processes so-called discrete-time and continuous-time Markov chain. Agricultural economics in general deals with discrete-time Markov chain since available data are generally recorded at discrete time periods.

Consider a population of N farms that is partitioned into a finite number J of categories or ‘states of nature’. Assuming that farms move from one category to another during a certain period of time r according to a stochastic process, we define the number $n_{j,t+r}$ of farms in category j at time $t + r$ as:

$$n_{j,t+r} = \sum_{i=1}^J n_{i,t} P_{ij,t}^{(r)} \quad (2.1)$$

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

$$\begin{aligned} p_{ij,t}^{(r)} &\geq 0, \quad \forall i, j, t \\ \sum_{j=1}^J p_{ij,t}^{(r)} &= 1, \quad \forall i, t. \end{aligned} \quad (2.2)$$

In matrix notation, equation (2.1) can then be rewritten as:

$$\mathbf{n}_{t+r} = \mathbf{n}_t \times \mathbf{P}^{(r)} \quad (2.3)$$

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

Using individual-level data, the observed r -step transition probabilities can then be computed from a contingency table as:

$$p_{ij}^{(r)} = \frac{\nu_{ij}^{(r)}}{\sum_j \nu_{ij}^{(r)}} \quad (2.4)$$

where $\nu_{ij}^{(r)}$ is the total number of r -step transitions from category i to category j during the period of observation and $\sum_j \nu_{ij}^{(r)}$ the total number of r -step transitions out of category i .

2.2.1 The simple Markov chain model (MCM)

Assuming that the transition process of agents is Markovian first-order, [Anderson and Goodman \(1957\)](#) showed that the maximum-likelihood estimator of the 1-step TPM corresponds to the observed transition matrix, that is, $\hat{\mathbf{\Pi}} = \mathbf{P}^{(1)}$. Then, since we focus on the issue of accounting for heterogeneity, similar to work done by others ([Major, 2008](#); [Cipollini et al., 2012](#)), we restrict our analysis to the stationary case where the r -step transition probability matrix (TPM). Under the assumption of stationarity, the r -step TPM ($\hat{\mathbf{\Pi}}^{(r)}$) is then obtained by raising the 1-step transition matrix to the power r (see appendix A):

$$\hat{\mathbf{\Pi}}^{(r)} = (\hat{\mathbf{\Pi}})^r. \quad (2.5)$$

The MCM approach thus assumes that 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, while $\hat{\mathbf{\Pi}}$ is a consistent estimate of $\mathbf{P}^{(1)}$, $\hat{\mathbf{\Pi}}^{(r)}$ proves to be a poor estimate of $\mathbf{P}^{(r)}$ ([Blumen et al., 1955](#)). In particular, the main diagonal elements of $\hat{\mathbf{\Pi}}^{(r)}$ largely underestimate those of $\mathbf{P}^{(r)}$. This means that, in general, $\hat{\pi}_{ii}^{(r)} \ll p_{ii}^{(r)}$. In other words, the simple MCM tends to overestimate agents' mobility because of the homogeneity assumption.

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

One way to obtain a 1-step TPM that leads to a more consistent r -step estimate consists in relaxing the assumption of homogeneity in the transition process underlying the MCM approach.

Following Poulsen (1983), Frydman (2005) proposed grounding the source of population heterogeneity on the rate of agents' movement; farms may move across categories at various speeds, each according to one of several types of transition process. This usually constitutes unobserved heterogeneity because observing the set of transitions that an farm actually experienced does not, in general, unambiguously reveal which stochastic process generated this specific sequence, hence the farm's type.

Implementing this idea leads to considering a mixture of Markov chains to capture the population heterogeneity. It is considered that the population is divided into a discrete number G of homogeneous types or groups of farms instead of just one, each farm belonging to one and only one of these types. Assuming that each farm type is characterised by its own elementary Markov process, the general form of the M-MCM then consists in decomposing the 1-step transition matrix as:

$$\Phi = \{\phi_{ij}\} = \sum_{g=1}^G \mathbf{S}_g \mathbf{M}_g \quad (2.6)$$

where $\mathbf{M}_g = \{m_{ij,g}\}$ is the TPM defining the 1-step Markov process followed by type- g farms, and $\mathbf{S}_g = \text{diag}(s_{i,g})$ is a diagonal matrix that gathers the shares of type- g farms in each category. Since every farm 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.

Since we consider only the stationary case here, it is assumed that neither \mathbf{M}_g nor \mathbf{S}_g varies over time. Then, the r -step TPM for any future time period r can be defined as the linear combination of the r -step processes:

$$\Phi^{(r)} = \sum_{g=1}^G \mathbf{S}_g (\mathbf{M}_g)^r. \quad (2.7)$$

With the MCM and M-MCM modelling frameworks as defined above, it should be noted that: (i) the M-MCM reduces to the MCM if $G = 1$, *i.e.*, only one type of farms is considered or, equivalently, the homogeneity assumption holds; and (ii) the aggregate overall M-MCM process described by $\Phi^{(r)}$ as defined by equation (2.7) may no longer be Markovian even if each farm type follows a specific Markov process.

Since the number of parameters to estimate increases rapidly with the number of homogeneous farm types (G), the estimation of matrix Φ as defined in the general case by equation

(2.6) can quickly become an ill-posed problem, in particular when the number of parameters to estimate is large and the number of observations is relatively small. As an alternative, the M-MCM parametrisation proposed by [Frydman \(2005\)](#) can help to decrease the number of parameters to be estimated (see appendix B).

2.2.3 A simple implementation of the M-MCM: the Mover-Stayer model (MSM)

In this analysis, we stick to the simplest version of the M-MCM, namely the MSM first proposed by [Blumen *et al.* \(1955\)](#). In this restricted approach, only two types of homogeneous farms are considered, those who always remain in their initial category (the ‘stayers’) and those who follow a first-order Markovian process (the ‘movers’). Formally, this leads to rewriting equation (2.6) in a simpler form as:

$$\Phi = \mathbf{S} + (\mathbf{I} - \mathbf{S}) \mathbf{M}. \quad (2.8)$$

With regard to the general formulation (2.6), 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. Thus, following equation (2.7), the MSM overall population r -step TPM can be expressed as:

$$\Phi^{(r)} = \mathbf{S} + (\mathbf{I} - \mathbf{S}) (\mathbf{M})^r. \quad (2.9)$$

2.3 Estimation method

Various methods have been used to estimate the mover-stayer model (MSM) parameters ranging from a simple calibration method ([Goodman, 1961](#)), minimum chi-square ([Morgan *et al.*, 1983](#)), maximum likelihood ([Frydman, 1984](#)) to Bayesian inference ([Fougere and Kamionka, 2003](#)). [Frydman \(2005\)](#) was the first to develop a maximum-likelihood estimation method for the general M-MCM. In the following section, we present the corresponding strategy in the simplified case of the MSM, which consists of two steps: firstly, under complete information, *i.e.*, as if the population heterogeneity were perfectly observable; secondly, under incomplete information, *i.e.*, accounting for the fact that population heterogeneity is not actually observed.

2.3.1 Likelihood maximisation under complete information

Under complete information the status of each farm k , either stayer (denoted ‘ S ’) or mover (denoted ‘ M ’), is perfectly known *ex-ante* and can be recorded through a dummy variable $Y_{k,S}$

where $Y_{k,S} = 1$ if farm k is a stayer and $Y_{k,S} = 0$ if farm k is a mover.

$$\log L = \sum_{k=1}^N Y_{k,S} \log l_{k,S} + \sum_{k=1}^N (1 - Y_{k,S}) \log l_{k,M} \quad (2.10)$$

where the first sum on the right hand side is the overall log-likelihood associated with stayers and the second sum is the overall log-likelihood associated with movers.

At the individual level, conditional on knowing that k was initially in size category i :

1. the likelihood that farm k is a stayer, $l_{k,S}$, is given by s_i the share of farms who never move out of category i during the whole period of observation (see appendix C);
2. the likelihood that farm k is a mover is (Frydman and Kadam, 2004):

$$l_{k,M} = (1 - s_i) \prod_{i \neq j} (m_{ij})^{\nu_{ij,k}} \prod_i (m_{ii})^{\nu_{ii,k}} \quad (2.11)$$

where $\nu_{ij,k}$ is the number of transitions from category i to category j made by farm k and $\nu_{ii,k}$ is the total number of times farm k stayed in category i . On the right hand side of equation (2.11), the first product is thus the probability of farm k moving out of category i , while the second product is the probability of farm k staying in category i from one period to the next, even though k is a mover.

Substituting $l_{k,S}$ and $l_{k,M}$ in equation (2.10), the log-likelihood of the MSM for the whole population can be expressed as:

$$\log L = \sum_i n_i \log(1 - s_i) + \sum_i n_{i,S} \log \{s_i / (1 - s_i)\} + \sum_{i \neq j} \nu_{ij} \log(m_{ij}) + \sum_i \nu_{ii,M} \log(m_{ii}) \quad (2.12)$$

where n_i and $n_{i,S}$ are, respectively, the numbers of farms and stayers who were initially in category i , $\nu_{ij} = \sum_k \nu_{ij,k}$ is the total number of transitions from category i to category j , $\nu_{ii,M} = \sum_k (1 - Y_{k,S}) \nu_{ii,k}$ is the total number of times movers stayed in category i , and m_{ij} and m_{ii} are the elements of the generator matrix (\mathbf{M}) of movers.

Then, maximising equation (2.12) with respect to the unknown parameters s_i and m_{ij} leads to the optimal values of the MSM parameters:

1. solving for $\partial \log L / \partial s_i = 0$ yields the optimal share of stayers in each category i :

$$\hat{s}_i = \frac{n_{i,S}}{n_i} \quad (2.13)$$

2. solving for $\partial \log L / \partial m_{ij} = 0$ for $i \neq j$ and noting that, by definition of \mathbf{M} as a stochastic matrix, $m_{ii} = 1 - \sum_{i \neq j} m_{ij}$ yields:

$$\hat{m}_{ii} = \frac{\nu_{ii,M}}{\nu_i + \nu_{ii,M}} \quad (2.14)$$

and

$$\hat{m}_{ij} = \frac{\nu_{ij}}{\nu_i} (1 - \hat{m}_{ii}) \quad \forall i \neq j \quad (2.15)$$

where $\nu_i = \sum_{j \neq i} \nu_{ij}$ is the total number of transitions out of category i .

2.3.2 The expectation-maximisation (EM) algorithm under incomplete information

Since, as already mentioned, it is unlikely in practice that one knows beforehand which farms actually behave as stayers among those who remain in their starting category during the period of observation. Therefore, equation (2.12) cannot be used directly to estimate the MSM parameters in the farming context. Indeed, because the transition process is assumed to be stochastic, even movers may remain for a long time in their initial category before moving, so that they may not appear as movers but as stayers over the observed period. In such a case, [Fuchs and Greenhouse \(1988\)](#) suggested that the MSM parameters could be estimated using the EM algorithm developed by [Dempster *et al.* \(1977\)](#): rather than observing the dummy variable $Y_{k,S}$, the EM algorithm allows its expected value $E\{Y_{k,S} | X_k(0) = i\}$ to be estimated, *i.e.*, the probability for each farm k to be a stayer, given farm k 's initial category i and observed transition sequence.

Following [Frydman and Kadam \(2004\)](#), the four steps of the EM algorithm are defined in the case of the MSM as follows.

(i) Initialisation: Arbitrarily choose initial values s_i^0 for the shares of stayers and m_{ii}^0 for the main diagonal entries of the TPM of movers (\mathbf{M}).

(ii) Expectation: At iteration p of the algorithm, compute the probability of farm k being a stayer, $E^p\{Y_{k,S} | X_k(0) = i, \boldsymbol{\theta}^p\}$, where $\boldsymbol{\theta}^p = (s_i^p, m_{ii}^p)$ is the vector of iteration- p values of the parameters to be estimated. If at least one transition is observed for farm k then set $E^p\{Y_{k,S} | X_k(0) = i, \boldsymbol{\theta}^p\} = 0$, otherwise this probability is given by, according to Bayes' law:

$$E^p\{Y_{k,S} | X_k(0) = i, \boldsymbol{\theta}^p\} = \frac{l_{k,S}}{l_{k,S} + l_{k,M}}. \quad (2.16)$$

Replacing $l_{k,S}$ and $l_{k,M}$ in the above equation, we obtain:

$$E^p \{Y_{k,S} | X_k(0) = i, \boldsymbol{\theta}^p\} = \frac{s_i^p}{s_i^p + (1 - s_i^p)(m_{ii}^p)^{\nu_{ii,k}}}. \quad (2.17.i)$$

Then compute:

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

$$E^p(n_{i,S}) = \sum_k^N E^p \{Y_{k,S} | X_k(0) = i, \boldsymbol{\theta}^p\} \quad (2.17.ii)$$

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

$$E^p(\nu_{ii,M}) = \sum_k^N [1 - E^p \{Y_{k,S} | X_k(0) = i, \boldsymbol{\theta}^p\}] \nu_{ii,k}. \quad (2.17.iii)$$

(iii) Maximisation: Update s_i^p and m_{ii}^p by maximising equation (2.12) using the expected values of $n_{i,S}$ and $\nu_{ii,M}$ obtained at iteration p . The new parameters are thus given by:

$$s_i^{p+1} = \frac{E^p(n_{i,S})}{n_i} \quad \text{and} \quad m_{ii}^{p+1} = \frac{E^p(\nu_{ii,M})}{\nu_i + E^p(\nu_{ii,M})}. \quad (2.17.iv)$$

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

When convergence is reached, the optimal values s_i^* and m_{ii}^* are considered to be the searched for estimators \hat{s}_i and \hat{m}_{ii} . Then, \hat{m}_{ij} derives from \hat{m}_{ii} as in equation (2.15).

Following [Frydman \(2005\)](#), the standard errors attached to the MSM parameters can be computed directly from the EM equations using the method proposed by [Louis \(1982\)](#). Since the log-likelihood function given by equation (2.12) is twice differentiable with respect to the model parameters, both the Fisher information matrix and the missing information matrix can be straightforward derived (see [McLachlan and Krishnan \(2007\)](#) for more details).

Two components are required to compute standard errors from the EM algorithm equations ([Louis, 1982](#)): (i) the observed information matrix given by the negative of the Hessian matrix of the log-likelihood function and; (ii) the missing information matrix obtained from the gradient vector, that is, the vector of score statistics based on complete information. Since the log-likelihood function given by equation (2.12) is twice differentiable with respect to the model parameters, the standard errors can be computed as follows.

Let $\boldsymbol{\Omega}_c(\mathbf{z}; \hat{s}_i, \hat{m}_{ij})$ and $\boldsymbol{\Omega}_m(\mathbf{z}; \hat{s}_i, \hat{m}_{ij})$ ($i, j = 1, \dots, J$) be the observed ($d \times d$) information matrices in terms of complete and missing information, respectively, where $Z_k = (X_k, Y_k)$ gathers the transition sequence X_k of farm k and the unobserved farm's type dummy variable Y_k , and d is the number of estimated parameters. The observed information matrix in terms of incomplete information can then be derived as:

$$\boldsymbol{\Omega}(\mathbf{x}; \hat{s}_i, \hat{m}_{ij}) = \boldsymbol{\Omega}_c(\mathbf{z}; \hat{s}_i, \hat{m}_{ij}) - \boldsymbol{\Omega}_m(\mathbf{z}; \hat{s}_i, \hat{m}_{ij}) \quad (2.18)$$

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

$$\boldsymbol{\Omega}_m(\mathbf{z}; \hat{s}_i, \hat{m}_{ij}) = E[\mathbf{s}_c(\mathbf{z}; \hat{s}_i, \hat{m}_{ij}) \times \mathbf{s}_c(\mathbf{z}; \hat{s}_i, \hat{m}_{ij})'] \quad (2.19)$$

where $\mathbf{s}_c(\mathbf{z}; \hat{s}_i, \hat{m}_{ij})$ is the vector of score statistics in terms of complete information.

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

$$\mathbf{se} = \{\sqrt{\psi_{ll'}}\} \quad (2.20)$$

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

The standard errors attached to the overall 1-year TPM $\boldsymbol{\Phi}$ can then be derived applying the standard Delta method to equation (2.8) (Oehlert, 1992). Finally, because it is more complicated to apply the Delta method to equation (2.9) as it involves powers of matrices, we used a bootstrap sampling method to compute standard errors attached to the r -step TPMs $\boldsymbol{M}^{(r)}$ and $\boldsymbol{\Phi}^{(r)}$ (Efron, 1979; Efron and Tibshirani, 1986).

2.4 Model comparison

Two types of analysis were conducted to assess whether or not the MSM outperforms the MCM.

2.4.1 Likelihood ratio (LR) test

The likelihood ratio test allows comparison of the in-sample performance of the two models in recovering the data-generating process to be compared. As stated by Frydman and Kadam

(2004), the likelihood ratio statistic for the MSM is given by:

$$LR = \frac{L_{MCM}(\hat{\boldsymbol{\Pi}})}{L_{MSM}(\hat{\boldsymbol{S}}, \hat{\boldsymbol{M}})} \quad (2.21)$$

where L_{MCM} and L_{MSM} are the estimated maximum likelihoods for the MCM and the MSM, respectively. The asymptotic distribution of $-2\log(LR)$, 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 follows an MCM ($H_0 : \hat{\boldsymbol{S}} = \mathbf{0}$) against the hypothesis that it is a mixture of movers and stayers ($H_1 : \hat{\boldsymbol{S}} \neq \mathbf{0}$). The observed log-likelihood for both models can be derived from equation (2.10), by imposing $Y_{k,S} = 0$ for all farms k for the MCM and by replacing $Y_{k,S}$ by its optimal expected value ($E^*(Y_{k,S})$) for the MSM.

2.4.2 Average Marginal Error (AME)

The estimated parameters were used to compute the corresponding r -step TPMs, *i.e.*, $\hat{\boldsymbol{\Pi}}^{(r)} = (\hat{\boldsymbol{\Pi}})^r$ for the MCM and $\hat{\boldsymbol{\Phi}}^{(r)} = \hat{\boldsymbol{S}} + (\mathbf{I} - \hat{\boldsymbol{S}})(\hat{\boldsymbol{M}})^r$ for the MSM. These r -step TPMs were then used to perform out-of-sample short- to long-term projections of farm distributions across size categories according to equation (2.1) (see below).

On the one hand, TPMs from both models were compared to the observed one, providing a second in-sample assessment in addition to the likelihood ratio test. Following [Frydman *et al.* \(1985\)](#) who compared percentage differences between predicted and observed transition probabilities, we compute an average marginal error (AME) for each model based on [Cipollini *et al.* \(2012\)](#)'s formulation of the expected marginal distribution. The AMEs for the transition probability matrices are thus obtained as:

$$AME_{TPM} = \frac{1}{J \times J} \sum_{i,j} \sqrt{\left(\frac{\hat{p}_{ij}^{(r)} - p_{ij}^{(r)}}{p_{ij}^{(r)}} \right)^2} \quad (2.22)$$

where $\hat{p}_{ij}^{(r)}$ and $p_{ij}^{(r)}$ are the predicted and observed TPM entries, respectively:

- $\hat{p}_{ij}^{(r)} \equiv \hat{\pi}_{ij}^{(r)}$ under the MCM while $\hat{p}_{ij}^{(r)} \equiv \hat{\phi}_{ij}^{(r)}$ under the MSM
- $p_{ij}^{(r)}$ derives from equation (2.4).

On the other hand, AMEs were similarly computed for both the MCM and MSM projections of farm-size distributions with respect to those actually observed, providing an out-of-sample comparison of the models. The AMEs for the projected farm-size distributions at time

$t + r$ are thus obtained as:

$$AME_{FSD}(t + r) = \frac{1}{J} \sum_j \sqrt{\left(\frac{\hat{n}_{j,t+r} - n_{j,t+r}}{n_{j,t+r}}\right)^2} \quad (2.23)$$

where $\hat{n}_{j,t+r}$ are the predicted proportions of farms in size category j at time $t + r$ using either $\hat{\pi}_{ij}^{(r)}$ for the MCM or $\hat{\phi}_{ij}^{(r)}$ for the MSM, and $n_{j,t+r}$ are the observed counterparts.

In contrast to some dissimilarity indexes reported by [Jafry and Schuermann \(2004\)](#) or the matrix of residuals computed by [Frydman *et al.* \(1985\)](#), the AME provides a global view of the distance between the predicted TPM or population distribution across size categories and the observed ones. It can be interpreted as the average percentage of deviations on predicting the observed TPM or population distribution across size categories. Thus, the higher the AME, the greater the difference between the computed TPM or distribution and the observed one. The better model is therefore the one that yields the lowest AME.

2.5 Data

In our empirical application, we used the 2000-2013 “Réseau d’Information Comptable Agricole” (RICA) data, the French strand of the Farm Accountancy Data Network (FADN). FADN is an annual survey defined at European Union (EU) level and carried out in each member state. The information collected at individual level relates to both the physical and structural characteristics of farms and their economic and financial characteristics. Note that, to comply with accounting standards that may differ from one country to the next, the EU-level questionnaire was adapted for France, but the results were unchanged whether the original or adapted version of the questionnaire was used.

In France, RICA is produced and disseminated by the Ministry for Agriculture’s statistical and foresight office. It focuses on ‘medium and large’ farms (see below) and constitutes a stratified and rotating panel of approximately 7,000 farms surveyed each year. Some 10% of the sample is renewed every year so that, on average, farms are observed during 5 consecutive years. However, some farms may be observed only once, and others several but not consecutive times. Some farms remained in the database over the entire period studied, *i.e.*, fourteen consecutive years. Each farm in the dataset is assigned year-specific weighting factors that reflect its annual stratified sampling probabilities, allowing for extrapolation at population level when the database is used in its cross-sectional dimension. However, when the database is used as a panel, as is the case here, these extrapolation weights should not be used because they are not relevant in the inter-temporal dimension.

Because we considered all 17,285 farms present in the 2000-2013 sample, whatever the type of farming, we chose to concentrate on size as defined in economic terms. In accordance with EU regulation (CE) N°1242/2008, European farms are classified into fourteen economic-size (ES) categories, evaluated in terms of total standard output (SO) expressed in Euros. As mentioned above, in France, RICA focuses on ‘medium and large’ farms, those whose SO is greater than or equal to 25,000 Euros; this corresponds to ES category 6 and above. Since size categories are not equally represented in the sample, we aggregated the nine ES categories available in RICA into five categories: strictly less than 50,000 Euros of SO (ES6); from 50,000 to less than 100,000 Euros of SO (ES7); from 100,000 to less than 150,000 Euros of SO (lower part of ES8); from 150,000 to less than 250,000 Euros of SO (upper part of ES8); 250,000 Euros of SO and more (ES9 to ES14).

RICA being a rotating panel, farms that either enter or leave the sample in a given year cannot be considered as actual entries into or exits from the agricultural sector. Thus, we could not work directly on the evolution of farm numbers but rather on the evolution of population shares by size categories, *i.e.*, the size distribution in the population. Table 2.1 presents the year-on-year evolution by size categories of farm numbers in the sample as well as for the extrapolated population. It also reports the average ES in thousands of Euros of SO both at the sample- and extrapolated-population levels.

Figure 2.1 shows that the share of smaller farms (below 100,000 Euros of SO) decreased from 56% to 46% between 2000 and 2013 while the share of larger farms (above 150,000 Euros of SO) increased from 28% to 38%, and the share of intermediate farms (100,000 to less than 150,000 Euros of SO) remained stable at 16%. As a consequence, as can be seen from Table 2.1 and Figure 2.1, the average economic size of French farms was multiplied by more than 1.25 over this period. Note that Table 2.1 also reveals that the size distribution became more heterogeneous since the standard deviation of economic size was multiplied by almost 1.5, a feature that has already been observed for the population of French farms as a whole and in other periods (Butault and Delame, 2005; Desriers, 2011). Lastly, Table 2.1 reveals that these observations also apply at sample level, even though the latter is skewed towards larger sizes with regard to the population as a whole.

To assess which model performed better, we compared the MCM and the MSM on the basis of both in-sample estimation and out-of-sample size-distribution forecasts. To do so, we split the RICA database into two parts: (i) observations from 2000 to 2010 were used to estimate the parameters of both models; (ii) observations from 2011 to 2013 were used to compare actual farm-size distributions with their predicted counterparts for both models. Note that in doing so we assumed that, in the case of the MSM, eleven years is a sufficiently long interval for a robust estimation of the movers’ transition process.

TABLE 2.1: Distribution by economic size (ES) and average ES for the studied sample, 2000-2013.

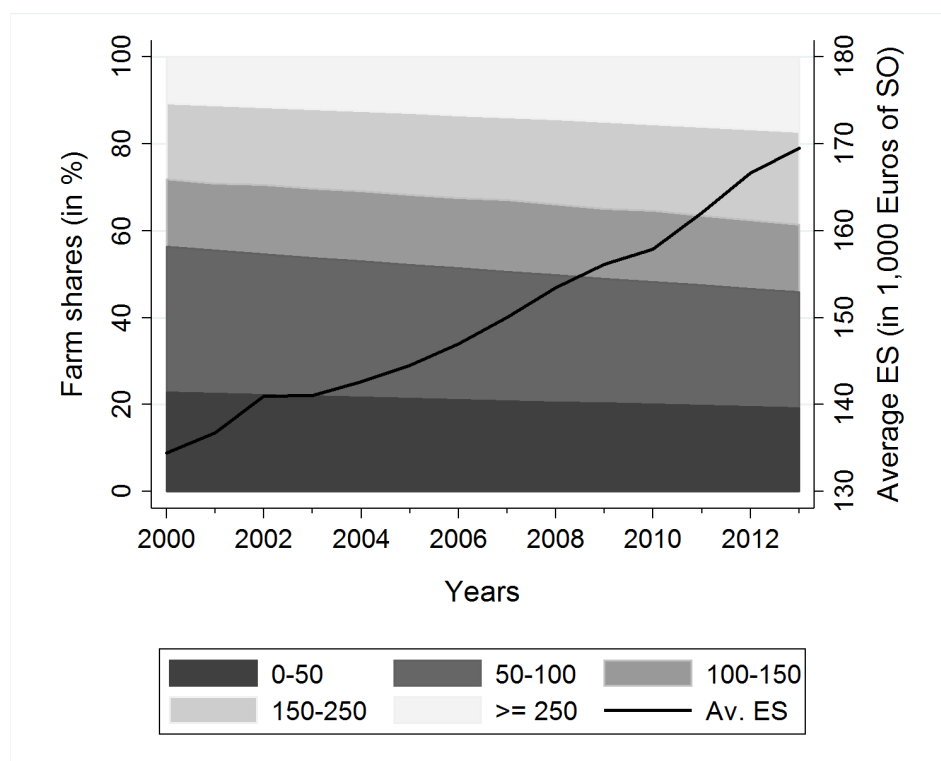
Year	Number of farms by ES category					Total	Average ES S	
	0-50	50-100	100-150	150-250	≥ 250		(std. dev.)	
2000	790	2,234	1,629	1,762	1,342	7,757	168.88	(179.01)
	87,924	129,691	59,857	67,367	41,457	386,296	134.46	(151.72)
2001	746	2,231	1,625	1,817	1,382	7,801	170.98	(180.88)
	84,442	123,900	57,583	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)

Note: ES in 1,000 Euros of standard output (SO). For each year, the first row reports figures at the sample level and the second row reports figures for the extrapolated population using weighted factors.

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

For in-sample estimations, we also had to restrict the sub-sample to farms which were present in the database for at least two consecutive years in order to observe potential transitions. The corresponding unbalanced panel then comprised 13,123 farms out of the 17,285 farms in the original database (76%), leading to 78,434 farm×year observations and 65,311 individual 1-year transitions (including staying in the same category) from 2000 to 2010. Furthermore, ten sub-samples could be constructed from this unbalanced panel, according to the minimum number of consecutive years a farm remained present in the database, from two to eleven: sub-sample #1 thus corresponds to the subset of farms which remained for at least two consecutive

FIGURE 2.1: Extrapolated population shares by farm size categories and average economic size (ES).



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

TABLE 2.2: Number of individual farms, observations and transitions by sub-samples

	Sub-samples									
	#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
1-year transitions										
$\forall(i, j)$	65,311	63,479	59,541	54,615	48,923	43,173	37,203	32,555	27,379	21,700
$\forall j \neq i$	7,374	7,138	6,672	6,110	5,426	4,700	4,045	3,478	2,889	2,212

Note: sub-sample #1 corresponds to the subset of farms which remained present in the database for at least two consecutive years; sub-sample #2 to those which remained for at least three consecutive years; and so forth; the third row gives the number of observed 1-year transitions from category i to category j , including remaining in the initial category; and the fourth row gives the number of observed 1-year transitions from category i to category j excluding remaining in the initial category.

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

years (full 2000-2010 unbalanced panel); sub-sample #2 to those which remained for at least three consecutive years; and so forth, up to sub-sample #10 which corresponds to the 2000-2010 balanced panel. Table 2.2 reports the corresponding numbers of individual farms and observed transitions for each sub-sample.

Before proceeding with the results, it should also be noted that because we worked with

panel subsets of the full sample and therefore could not account for extrapolation weights in the estimations as explained above, the transition probabilities reported in the next section can only be viewed as *conditional* on having been observed over a specific number of consecutive years during the entire period under study. Projected population distributions obtained from these transition probabilities should therefore be interpreted with due care from an empirical point of view. They were performed here to compare the forecasting power of the studied models, for which the above caveat applies to the same extent.

2.6 Results

In this section, we first report the results of in-sample estimations, *i.e.*, the estimated 1-year TPMs for the MCM and the MSM. We then compare the models based on in-sample results in order to assess which one performs better in recovering the underlying transition process both from a short-term and a long-term perspective. Lastly, we compare the models' ability to forecast future farm-size distributions based on out-of-sample observations.

2.6.1 In-sample estimation results

The MCM and MSM parameters were estimated using ten sub-samples according to the minimum length of time farms remain in the database. To save space, we only report here those for the balanced sub-sample, *i.e.*, using the 2000-2010 balanced panel consisting of 2,170 individual farms and 21,700 observed transitions over the 11 years. However, when any of the unbalanced panels were considered, the results, and hence conclusions, remained similar to those reported here.

Table 2.3 reports the estimated 1-year TPM under the MCM assumption. As is usually the case, the matrix 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. Note that this does not mean no size change at all but, at least, no change sufficient to move to another category as we defined them.

Table 2.4 reports the estimated shares of stayers and transition probability matrix of movers along with the corresponding 1-year TPM for the entire population under the MSM assumption. Two main results can be seen from this table. Firstly, the estimated stayer shares (Panel a of Table 2.4) show that the probability of being a stayer is close to or above 30% whatever the category considered; it reaches almost 50% for farms below 50,000 Euros of SO and is even higher than 60% for farms over 250,000 Euros of SO. This means that, according to

TABLE 2.3: Estimated 1-year transition probability matrix under Markov chain model ($\hat{\Pi}$).

		ES class				
		0-50	50-100	100-150	150-250	≥ 250
ES class	0-50	0.917 (0.024)	0.079 (0.007)	0.002 (0.001)	0.002 (0.001)	0.001 (0.001)
	50-100	0.030 (0.002)	0.898 (0.013)	0.065 (0.004)	0.005 (0.001)	0.002 (0.001)
	100-150	0.002 (0.001)	0.062 (0.004)	0.854 (0.014)	0.080 (0.004)	0.002 (0.001)
	150-250	0.001 (0.000)	0.004 (0.001)	0.054 (0.003)	0.886 (0.012)	0.055 (0.003)
	≥ 250	0.000 (0.000)	0.001 (0.001)	0.003 (0.001)	0.048 (0.003)	0.948 (0.014)

Log-likelihood: $\log L_{MCM} = -8,689.36$

Note: estimated parameters in bold font, standard errors in parentheses.

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

the MSM and depending on the size category, at least 30% of the farms are likely to remain in their initial category for at least another 10 years. Secondly, the transition probability matrix (Panel b of Table 2.4) reveals that even though movers are by definition expected to move from one category to another in the next ten years, nevertheless, the highest probability for them is to remain in the same category from one year to the next. Since the average time spent by movers in a particular category is given by $1/(1 - m_{ii})$, it can be seen from Table 2.4 that movers in the intermediate ES class [$1/(1 - 0.793) = 4.8$] were likely to remain in the same category for almost five years, while those above 250,000 Euros of SO [$1/(1 - 0.875) = 8$] were likely to remain for eight years before moving. In other words, farms that remained in a particular category for quite a long time (theoretically even over the entire observation period) were not necessarily stayers but could well be movers who had not yet moved. Together, these two results yield a 1-year TPM for the whole population that is also highly diagonal (Panel c of Table 2.4).

2.6.2 In-sample model comparison

Both models show that, at overall population level, farms were more likely to remain in their initial size category from year to year, confirming that structural change in the agricultural sector is a slow process which would be worth investigating in the long-term. In this respect, while the 1-year TPMs look very similar across both models, the resulting long-term transition model differs between the MCM, given by equation (2.5), and the MSM, given by equation (2.9). It is therefore important to assess which model performs better in recovering the true underlying transition process.

TABLE 2.4: Estimated stayer shares (\hat{s}_i) and 1-year transition probability matrices (TPMs) for movers (\hat{M}) and for the overall population ($\hat{\Phi}$).

ES class	0-50	0.494 (0.036)
	50-100	0.422 (0.021)
	100-150	0.291 (0.016)
	150-250	0.371 (0.017)
	≥ 250	0.650 (0.021)

a) Stayer shares (\hat{s}_i)

		ES class				
		0-50	50-100	100-150	150-250	≥ 250
ES class	0-50	0.837 (0.041)	0.154 (0.012)	0.004 (0.002)	0.004 (0.002)	0.001 (0.001)
	50-100	0.055 (0.004)	0.815 (0.022)	0.118 (0.006)	0.009 (0.002)	0.003 (0.001)
	100-150	0.002 (0.001)	0.089 (0.005)	0.793 (0.020)	0.113 (0.005)	0.003 (0.001)
	150-250	0.002 (0.001)	0.007 (0.001)	0.087 (0.004)	0.816 (0.020)	0.088 (0.004)
	≥ 250	0.001 (0.001)	0.003 (0.001)	0.007 (0.002)	0.114 (0.007)	0.875 (0.027)

b) Mover TPM (\hat{M})

		ES class				
		0-50	50-100	100-150	150-250	≥ 250
ES class	0-50	0.917 (0.019)	0.078 (0.011)	0.002 (0.001)	0.002 (0.001)	0.001 (0.001)
	50-100	0.032 (0.003)	0.893 (0.012)	0.068 (0.005)	0.005 (0.001)	0.002 (0.001)
	100-150	0.002 (0.001)	0.062 (0.004)	0.854 (0.014)	0.080 (0.005)	0.002 (0.001)
	150-250	0.001 (0.000)	0.005 (0.001)	0.055 (0.004)	0.884 (0.012)	0.055 (0.004)
	≥ 250	0.000 (0.000)	0.001 (0.001)	0.003 (0.001)	0.040 (0.005)	0.956 (0.009)

c) Overall population TPM ($\hat{\Phi}$)

Log-likelihood: $\log L_{MSM} = -8,384.08$

Note: estimated parameters in bold font, standard errors in parentheses.

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

The first assessment method used, *i.e.*, the likelihood ratio test, reveals that the MSM yields a better fit than the MCM: the value of the test statistic as defined by equation (2.21) is $-2\log(LR) = -2 \times (-8,689.36 + 8,384.08) = 610.56$, which is highly significant since the critical value of a chi-square distribution with $(G - 1) \times J = 5$ degrees of freedom and the 1% significance level is $\chi_{0.99}^2(5) = 15.09$. This leads to the rejection of the H_0 assumption that the stayer shares are all zero, and thus to the conclusion that the MSM allows the data-generating process to be recovered more efficiently than the MCM. The MSM should therefore also lead to a better approximation of transition probabilities in the long-term.

The second assessment method used allows this very point to be evaluated. Estimated parameters for both models were used to derive the corresponding 10-year TPMs, namely $\hat{\mathbf{\Pi}}^{(10)} = (\hat{\mathbf{\Pi}})^{10}$ for the MCM and $\hat{\mathbf{\Phi}}^{(10)} = \hat{\mathbf{S}} + (\mathbf{I} - \hat{\mathbf{S}})(\hat{\mathbf{M}})^{10}$ for the MSM. These estimated long-term matrices were then compared with the observed one, $\mathbf{P}^{(10)}$, which is derived from the contingency table using equation (2.4). It appears from the visual inspection of the three corresponding panels of Table 2.5 that the MSM 10-year matrix more closely resembles the one actually observed than the MCM 10-year matrix. In particular, we find as expected that the diagonal elements of $\hat{\mathbf{\Pi}}^{(10)}$ largely underestimate those of $\mathbf{P}^{(10)}$ while those of $\hat{\mathbf{\Phi}}^{(10)}$ are far closer. This means that the MCM largely tends overestimate the farms' mobility in the long-term, with regard to the MSM. The AMEs reported in Table 2.6 confirm the MSM's superiority over the MCM in modelling the long-term transition process. The AME for the overall predicted 10-year TPM is around 0.95 for the MCM while it is about 0.78 for the MSM. This means that the MSM is about 17 percentage points closer to the observed TPM than the MCM in the long-term. However, the AMEs also confirm that the improvement arises principally from the main diagonal elements: when only these are considered, the MSM performs about five times better than the MCM ($0.292/0.057 = 5.1$), while both models are almost comparable for off-diagonal elements, with the MCM this time performing slightly better ($0.657/0.724 = 0.9$).

2.6.3 Out-of-sample projections

In-sample estimation results lead us to conclude that accounting for unobserved heterogeneity in the rate of farms' movement, as the MSM does, avoids overestimating their mobility across size categories. The MSM should therefore also lead to a more accurate prediction of the farms' size distribution in the long-term without hampering its short-term prediction.

To validate this point, we performed out-of-sample short- to long-term projections of farm-size distribution using the parameters for both models as estimated with sub-sample #10. To do this, farm-size distributions in 2011, 2012 and 2013 were predicted from a short- to long-term perspective, by applying the estimated r -step TPMs (for $1 \leq r \leq 11$) to the corresponding observed distributions from 2000 to 2012. In other words, distributions in 2011, 2012 and 2013

TABLE 2.5: Observed and predicted 10-year transition probability matrices (TPMs) for both models.

		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 ($P^{(10)}$)

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

b) MCM predicted 10-year TPM ($\hat{I}^{(10)}$)

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

c) MSM predicted 10-year TPM ($\hat{\Phi}^{(10)}$)

Note: estimated parameters in bold font, bootstrap standard errors in parentheses (1,000 replications).

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

TABLE 2.6: Average marginal error (AME_{TPM}) between the predicted 10-year ($\hat{\mathbf{I}}^{(10)}$ and $\hat{\mathbf{F}}^{(10)}$) and the observed ($\mathbf{P}^{(10)}$) transition probability matrices (TPMs).

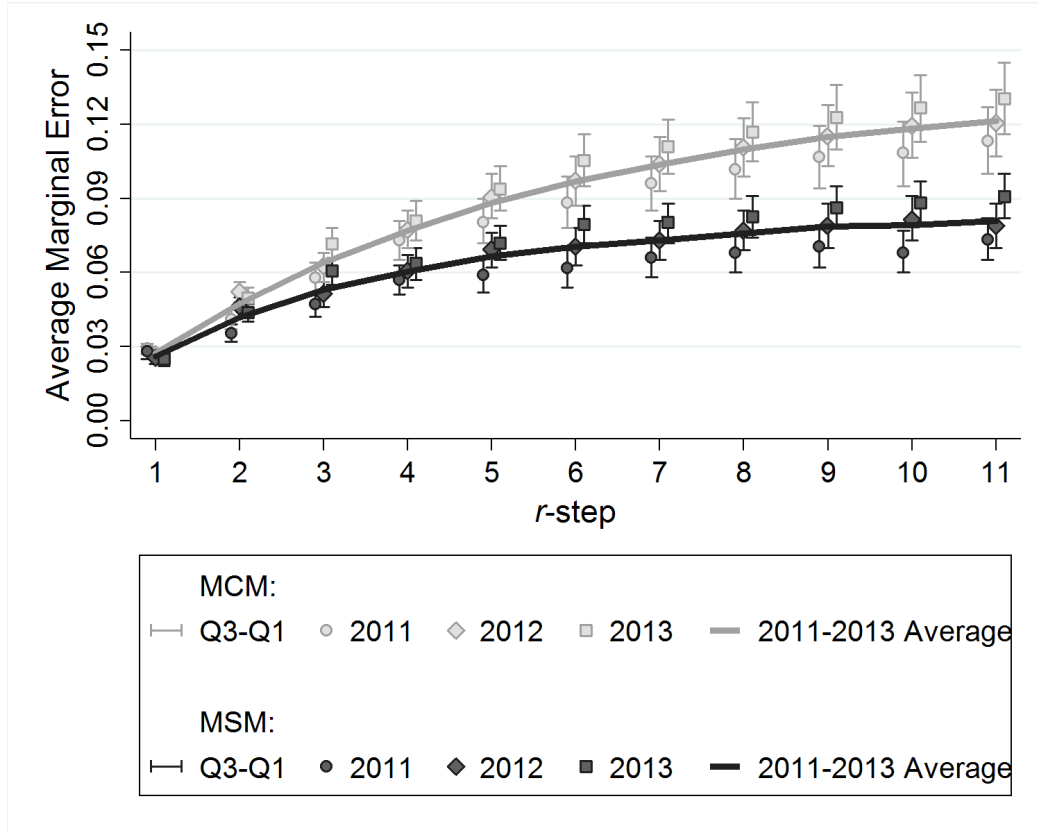
Model	Overall matrix	Main diagonal elements	Off-diagonal elements
MCM	0.949 (0.044)	0.292 (0.010)	0.657 (0.036)
MSM	0.781 (0.034)	0.057 (0.007)	0.724 (0.035)

Note: bootstrap standard errors in parentheses (1,000 replications).

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

were predicted: by applying the estimated 1-year TPMs ($\hat{\mathbf{I}}$ for the MCM and $\hat{\mathbf{F}}$ for the MSM) to the observed distributions in 2010, 2011 and 2012, respectively; by applying the estimated 2-year TPMs ($\hat{\mathbf{I}}^{(2)} = (\hat{\mathbf{I}})^2$ for the MCM and $\hat{\mathbf{F}}^{(2)} = \hat{\mathbf{S}} + (\mathbf{I} - \hat{\mathbf{S}})(\hat{\mathbf{M}})^2$ for the MSM) to the observed distributions in 2009, 2010 and 2011, respectively; and so forth. This process was continued by applying the estimated r -step TPMs ($\hat{\mathbf{I}}^{(r)} = (\hat{\mathbf{I}})^r$ for the MCM and $\hat{\mathbf{F}}^{(r)} = \hat{\mathbf{S}} + (\mathbf{I} - \hat{\mathbf{S}})(\hat{\mathbf{M}})^r$ for the MSM) to the observed distributions in $(2011 - r)$, $(2012 - r)$ and $(2013 - r)$ and varying r up to eleven. Then, the resulting distributions for both models were compared to distributions actually observed in 2011, 2012 and 2013 (see Table 2.1). The corresponding AMEs reported in Figure 2.2 summarise the results obtained for the 1,000 bootstrap replications. Four conclusions can be drawn from Figure 2.2. Firstly, as expected, the accuracy of both models decreases when increasing the time horizon of projections: the computed AMEs are significantly smaller in the short-term than they are in the medium and long-term for both models. Secondly, both models are almost comparable in the short run, confirming that adopting the MSM modelling framework does not degrade year-on-year forecasts with regard to the MCM. Thirdly, the MSM performs significantly better than the MCM in both the medium- and long-term. For example, the average AME for the MSM (0.068) is 1.3 times lower than that of the MCM (0.088) for 5-year interval projections, and 1.5 times lower for 11-year interval projections (0.082 for the MSM compared with 0.121 for the MCM). Fourthly, Figure 2.2 also shows that the accuracy and robustness of farm-size distribution predictions decrease more rapidly for the MCM than for the MSM when increasing the time horizon of projections since the AMEs as well as the interquartile ranges of the 1,000 bootstrap replications increase more rapidly for the MCM than for the MSM.

FIGURE 2.2: Average marginal error ($AME_{FSD}(t+r)$) between the out-of-sample projections of farm size distributions and the actually observed ones for both models.



Note: see text for an explanation on how short- to long-term projections were obtained; interquartile ranges (Q3-Q1) obtained from the 1,000 bootstrap replications.

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

2.7 Robustness tests

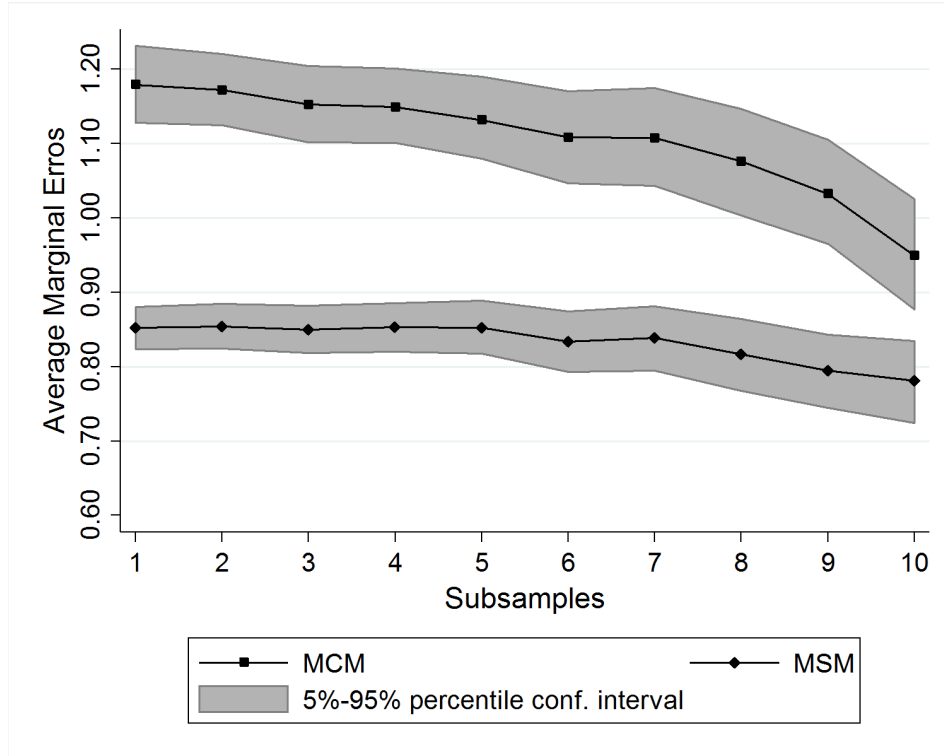
2.7.1 Sample size

As mentioned before, ten sub-samples can be constructed according to the length of time observations remain in the 2000-2010 RICA panel data. The balanced panel used for the result presented above consists of less observations than the other nine unbalanced panels (see Table 2.2). As farm transition across size categories is rather a the long-run process, we check the robustness of the results with regard to the other sub-samples.

The AMEs obtained for each sub-sample prove that the MSM leads to a better prediction of the observed 10-year TPM than the MCM whatever the sub-sample considered (see Figure 2.3)². For all sub-samples, the AME is never higher than 0.85 for the MSM while it

²Since 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

FIGURE 2.3: Average marginal error (AME_{TPM}) between the predicted 10-year ($\hat{\Pi}^{(10)}$ and $\hat{\Phi}^{(10)}$) and the observed ($P^{(10)}$) transition probability matrices (TPMs), by sub-samples.



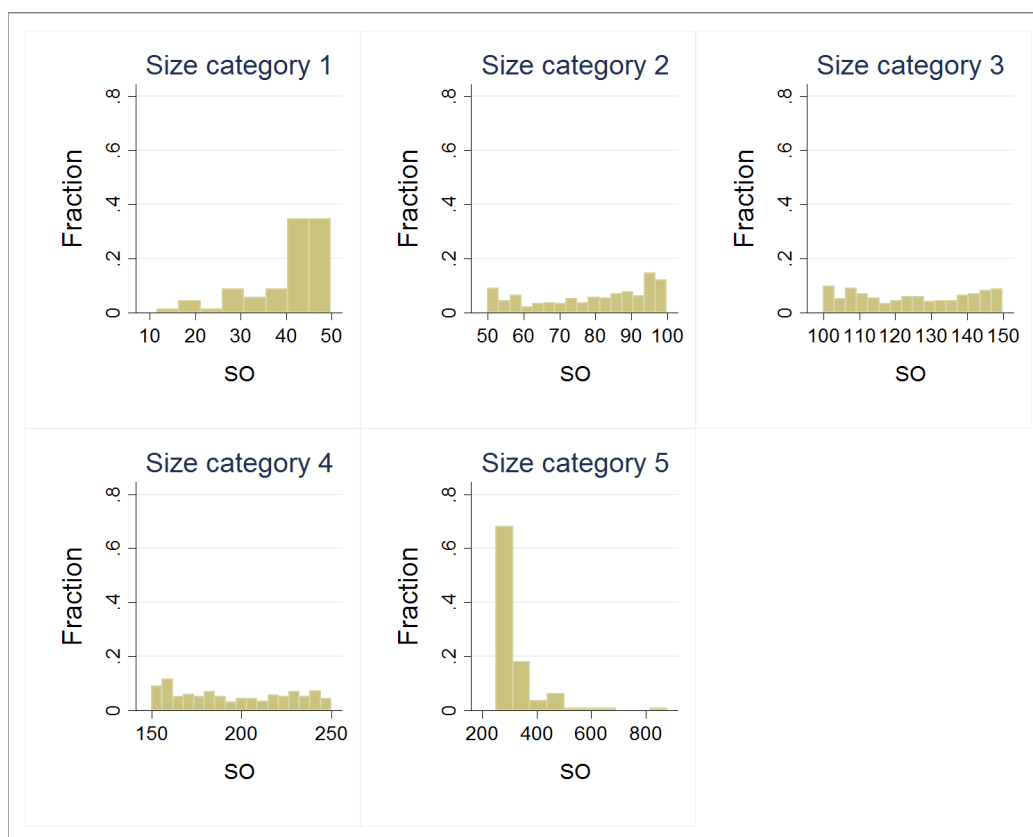
Notes: sub-sample 1 corresponds to the subset of farms which remained present in the database for at least two consecutive years; sub-sample 2 to those which remained for at least three consecutive years; and so forth; the confidence intervals are based on the percentiles, that is, the lower and upper bounds are the percentiles 5% and 95%, respectively.

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

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 TPM in term of percentage of deviation. Figure 2.3 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 for a short time period in the sample in fact add noise for the model parameter estimation. Indeed, the shorter farms remain in the database, the more incomplete the information about them for the algorithm to decide whether they are movers or stayers. Therefore, a balanced panel should provide a better prediction of the underlying transition probabilities for both models. Also, the relative accuracy loss seems larger for the MCM, suggesting that farms remaining a short time period add more noise to the MCM than for the MSM.

censored data and when the bootstrapped distribution for the estimated parameter is approximatively normal (Efron and Tibshirani, 1986).

FIGURE 2.4: Distribution of movers by category of sizes.



Note: Standard output (SO) in 1,000 Euros.

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

2.7.2 Definition of farm size categories

As size category of farms were defined using a continuous variables, one could argued that the results may depend one the chosen category bounds, because farms which are close to the bounds need less size variation to change category. Therefore, these farms may be more likely to behave as movers. We thus performed some robustness checks with regard to the definition of the size categories to validate the findings of the study.

Firstly, we present the distribution of mover farms by size category. Figure 2.4 shows that farms which are near the intervals' bounds for the two extreme categories (less than 50,000 and more than 250,000 Euros of SO) are indeed more likely to behave as movers. Conversely, farms are rather equally distributed in the other size categories. This suggests that some of the quantitative results of the model such as the proportion of stayers may vary according to the definition of size categories to in some extent, but qualitative results should remain the same.

Secondly, we tested the robustness with respect to the number of size categories. We

TABLE 2.7: Average marginal error (AME_{TPM}) between the predicted 10-year ($\hat{\mathbf{I}}^{(10)}$ and $\hat{\Phi}^{(10)}$) and the observed ($\mathbf{P}^{(10)}$) transition probability matrices (TPMs), for different number of size categories.

Model	Overall matrix	Main diagonal elements	Off-diagonal elements
MCM	0.836 (0.041)	0.230 (0.009)	0.607 (0.034)
MSM	0.675 (0.036)	0.049 (0.007)	0.627 (0.036)

a) Four ES class: < 50; 50-100; 100-250; 250 and more

Model	Overall matrix	Main diagonal elements	Off-diagonal elements
MCM	0.854 (0.048)	0.299 (0.010)	0.556 (0.040)
MSM	0.677 (0.033)	0.083 (0.014)	0.594 (0.035)

b) Five ES class: < 100; 100-200; 200-300; 300-400; 400 and more

Model	Overall matrix	Main diagonal elements	Off-diagonal elements
MCM	0.845 (0.034)	0.328 (0.007)	0.517 (0.028)
MSM	0.658 (0.022)	0.044 (0.009)	0.614 (0.023)

c) Six ES class: < 50; 50-100; 100-150; 150-200; 200-250; 250 and more

Note: Economic size (ES) in thousands euros of standard output; bootstrap standard errors in parentheses (1,000 replications).

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

thus estimated the models using different numbers of size categories from four to six. The four size categories, in thousands Euros of SO were defined as : < 50; 50-100; 100-250; 250 and more. For five the five size categories, we changed in this section the interval definition with respect to those used in previous sections. Here, the five size categories were therefore defined as: < 100; 100-200; 200-300; 300-400; 400 and more. Finally, the six size categories were defined as: < 50; 50-100; 100-150; 150-200; 200-250; 250 and more. In the three panels of Table 2.7, which are similar to Table 2.6, we report the AMEs between the predicted 10-year TPMs (MCM and MSM) and the observed TPM using these different numbers of farm size categories or a different definition of size intervals. The results show that despite the different values obtained for the computed AMEs, it is clear that the MSM always leads to closer TPMs than the usual MCM.

2.8 Concluding remarks

Accounting for farm heterogeneity is crucial in the analysis of structural change in farming. Existing literature however has so far only accounted for observed heterogeneity. This study investigate whether accounting for unobserved heterogeneity in farms' size-transition process improves the representation of structural change. Considering a mixture of two types of farms, the so-called mover-stayer model (MSM) is applied for the first time in an agricultural economics context. Our results show that, with regard to the MCM and even using the simple MSM framework, accounting for unobserved farm heterogeneity allows for closer estimates of both the observed transition matrix and farms' distribution across size categories in the long-term to be derived, without degrading any short-term analysis. This result is consistent with findings in other strands of economic literature, *i.e.*, by relaxing the assumption of homogeneity in the transition process, which is the basis of the MCM, the MSM gives a better representation of the underlying structural-change process. Moreover, robustness checks show that these conclusions remain valid when using different boundaries for size interval definitions and/or using different numbers of size categories.

However, this modelling framework remains a somewhat restricted and simplified version of the more general, mixed-Markov chain model (M-MCM), that we presented as an introduction to the MSM. Extending the MSM further could therefore lead to even more economically-sound and, statistically more accurate models for the farming sector. We briefly mention here some of the extensions we believe are promising. Firstly, more heterogeneity across farms could be incorporated by allowing for more than two unobserved types. For example, accounting for different types of movers could provide a better representation of the structural-change process in the farming sector by allowing farms that mainly tend to enlarge to be separated from farms that mainly tend to shrink. Secondly, the relatively strong assumption of a 'pure stayer' type could be relaxed because it appears unlikely that some farms 'never move at all', *i.e.*, will not change size category over their entire lifespan even if they do not do so for a relatively high number of consecutive years. In other words, an 'almost stayer' type could be hypothesised rather than a 'pure stayer' type. In this respect, the robustness of the MSM to the number of years during which farms are observed could be investigated. Indeed, by considering eleven years to perform in-sample estimations, our results show that movers may stay five to eight years in a given category before experiencing a sufficiently large (positive or negative) size change to reach another category. We can then infer that the shorter the observation period, the higher the number of farms that would be inappropriately considered as 'pure stayers'. From a methodological point of view, this leads to the conclusion that sufficiently long panel data needs to be available if the MSM is to be empirically implemented. This could be seen as a shortcoming of the MSM with regard to the MCM but it is in fact consistent with farm structural change being a long-term and slow process.

The empirical application in this study showed that accounting for unobserved farm heterogeneity is important when analysing of structural change in farming, and that a mixture-modelling framework such as the MSM is relevant to do this. However, while we could put forward some theoretical explanations for the presence of unobserved heterogeneity, to our knowledge, there is, as yet, no formal framework in agricultural economics that could help identify the potential sources of this heterogeneity and guide the empirical measurement of how they may affect farms' transition process. We hope that the results presented here and the limitations mentioned above will contribute to encouraging the development of such theoretical frameworks for a better understanding of structural change under the assumption of heterogeneous farm populations.

Our final recommendations are that, when implemented for thorough empirical studies, the proposed modelling framework should be extended to account for entries and exits, that factors driving homogeneous group membership – including observed heterogeneity variables – should be investigated, and that a non-stationary version of the M-MCM model should be developed. Since this should allow the transition process to be recovered more efficiently, it would surely prove to be very useful for analysing the factors that drive structural change in the farming sector, including agricultural policies, not only from a size-distribution perspective, but also with regard to the evolution of farm numbers.

Bibliography

- Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about markov chains. *Annals of Mathematical Statistics* 28: 89–110.
- Ben Arfa, N., Daniel, K., Jacquet, F. and Karantininis, K. (2015). Agricultural policies and structural change in french dairy farms: A nonstationary markov model. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 63: 19–42.
- Berchtold, A. (1998). *Chaînes de Markov et modèles de transition: applications aux sciences sociales*. Hermès.
- Bickenbach, F. and Bode, E. (2001). Markov or not Markov-This should be a question. Tech. rep., Kieler Arbeitspapiere.
- Blumen, I., Kogan, M. and McCarthy, P. J. (1955). *The industrial mobility of labor as a probability process, VI*. Cornell Studies in Industrial and Labor Relations.
- Butault, J.-P. and Delame, N. (2005). Concentration de la production agricole et croissance des exploitations. *Economie et statistique* 390: 47–64.
- Cipollini, F., Ferretti, C. and Ganugi, P. (2012). Firm size dynamics in an industrial district: The mover-stayer model in action. In Di Ciaccio, A., Coli, M. and Angulo Ibanez, J. M. (eds), *Advanced Statistical Methods for the Analysis of Large Data-Sets*. Springer Berlin Heidelberg, 443–452.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39: 1–38.
- Desriers, M. (2011). Farm structure. agricultural census 2010. production is concentrated in specialised farms. *Agreste Primeur* 272.
- Dutta, J., Sefton, J. A. and Weale, M. R. (2001). Income distribution and income dynamics in the united kingdom. *Journal of Applied Econometrics* 16: 599–617.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7: 1–26.
- Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1: 54–75.
- Fougere, D. and Kamionka, T. (2003). Bayesian inference for the mover-stayer model in continuous time with an application to labour market transition data. *Journal of Applied Econometrics* 18: 697–723.

- Fougere, D. and Kamionka, T. (2008). *Econometrics of individual labor market transitions*. Springer.
- Frydman, H. (1984). Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association* 79: 632–638.
- Frydman, H. (2005). Estimation in the mixture of Markov chains moving with different speeds. *Journal of the American Statistical Association* 100: 1046–1053.
- Frydman, H. and Kadam, A. (2004). Estimation in the continuous time mover-stayer model with an application to bond ratings migration. *Applied Stochastic Models in Business and Industry* 20: 155–170.
- Frydman, H., Kallberg, J. G. and Kao, D.-L. (1985). Testing the adequacy of Markov chain and mover-stayer models as representations of credit behavior. *Operations Research* 33: 1203–1214.
- Fuchs, C. and Greenhouse, J. B. (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: 841–868.
- Huettel, S. and Jongeneel, R. (2011). How has the EU milk quota affected patterns of herd-size change? *European Review of Agricultural Economics* 38: 497–527.
- Jafry, Y. and Schuermann, T. (2004). Measurement, estimation and comparison of credit migration matrices. *Journal of Banking & Finance* 28: 2603–2639.
- Lee, T., Judge, G. and Zellner, A. (1970). *Estimating the parameters of the Markov probability model from aggregate time series data*. Amsterdam: North Holland.
- Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 44: 226–233.
- Major, K. (2008). Income disparities among Hungarian micro-regions: The mover-stayer model. *Acta Oeconomica* 58: 127–156.
- McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*, 382. John Wiley & Sons.
- Meyn, S. P. and Tweedie, R. L. (2009). *Markov chains and stochastic stability*. Cambridge University Press.
- Morgan, T. M., Aneshensel, C. S. and Clark, V. A. (1983). Parameter estimation for mover-stayer models analyzing depression over time. *Sociological Methods & Research* 11: 345–366.

- Oehlert, G. W. (1992). A note on the delta method. *The American Statistician* 46: 27–29.
- Poulsen, C. S. (1983). Latent structure analysis with choice modeling applications. Ph.D. thesis, Wharton School, University of Pennsylvania.
- Shorrocks, A. F. (1976). Income mobility and the markov assumption. *The Economic Journal* 86: 566–578.
- Singer, B. and Spilerman, S. (1974). Social mobility models for heterogeneous populations. *Sociological Methodology* 5: 356–401.
- Singer, B. and Spilerman, S. (1976). The representation of social processes by Markov models. *American Journal of Sociology* 82: 1–54.
- Stokes, J. R. (2006). Entry, exit, and structural change in Pennsylvania’s dairy sector. *Agricultural and Resource Economics Review* 35: 357–373.
- Zimmermann, A. and Heckelei, T. (2012). Structural change of European dairy farms: A cross-regional analysis. *Journal of Agricultural Economics* 63: 576–603.
- Zimmermann, A., Heckelei, T. and Dominguez, I. P. (2009). Modelling farm structural change for integrated ex-ante assessment: Review of methods and determinants. *Environmental Science and Policy* 12: 601–618.

Appendices

A: Estimating a first-order Markov chain model

Let consider a stochastic process $\{X(t), t = 1, \dots, T\}$ where $X(t)$ are categories or state occupations by agents. Assuming a first-order Markovian process, transitions between categories can thus be described by the probability function (Lee *et al.*, 1970):

$$\text{prob}[X(t+1) = j | X(t) = i] = \pi_{ij} \quad (2.24)$$

where π_{ij} is interpreted as the probability to move from category i to category j in one occurrence, constrains to $\pi_{ij} \geq 0$ and:

$$\sum_{j=1}^J \pi_{ij} = 1, \quad \forall i. \quad (2.25)$$

Let ν_{ij} denote the number of transition from category i to category j over the period of observation. Assuming independence between state occupations implies that the random variables $\sum_j \nu_{ij}$ have multinomial distribution with multinomial probabilities π_{ij} and the likelihood function of the MCM thus writes (Anderson and Goodman, 1957):

$$L_{MCM} = \prod_{i,j} (\pi_{ij})^{\nu_{ij}}. \quad (2.26)$$

The Lagrangian form of the log-likelihood given the restrictions in equation (2.25), writes (Lee *et al.*, 1970):

$$\mathcal{L}(\log L_{MCM}, \lambda_i) = \sum_{i,j} \nu_{ij} \log \pi_{ij} - \lambda_i \left(\sum_j \pi_{ij} - 1 \right) \quad (2.27)$$

Maximising the log-likelihood with respect to the parameters π_{ij} , then give:

$$\frac{\partial \mathcal{L}(\log L_{MCM}, \lambda_i)}{\partial \pi_{ij}} = \frac{\nu_{ij}}{\pi_{ij}} - \lambda_i = 0 \quad (2.28)$$

implying that:

$$\lambda_i = \frac{\nu_{ij}}{\pi_{ij}} \quad (2.29)$$

And, maximising this log-likelihood with respect to the parameters λ_i , we obtain:

$$\frac{\partial \mathcal{L}(\log L_{MCM}, \lambda_j)}{\partial \lambda_j} = \sum_j \pi_{ij} - 1 = 0 \quad (2.30)$$

From equation (2.29), we have $\sum_j \nu_{ij} = \lambda_i \sum_j \pi_{ij}$ and equation (2.30) implies that:

$$\lambda_i = \sum_j \nu_{ij} \quad (2.31)$$

And, by replacing λ_i in equation (2.29), we obtain:

$$\hat{\pi}_{ij} = \frac{\nu_{ij}}{\sum_j \nu_{ij}} \quad (2.32)$$

Under stationarity assumption, transition probabilities are obtained as:

$$\hat{\pi}_{ij} = \frac{\sum_{t=1}^T \nu_{ij,t}}{\sum_{t=1}^T \sum_j \nu_{ij,t}} \quad (\forall i, j \in J) \quad (2.33)$$

where $\sum_{t=1}^T \nu_{ij,t}$ is the total number of transition from category i to category j at time t .

B: [Frydman \(2005\)](#)'s specification of the mixed Markov chain model (M-MCM)

Recall that the general form of the M-MCM is given by equation (2.6):

$$\Phi = \{\phi_{ij}\} = \sum_{g=1}^G \mathbf{S}_g \mathbf{M}_g \quad (2.34)$$

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.

Assuming that all type g TPMs are related to a specific one, arbitrarily chosen as that of the last type G , [Frydman \(2005\)](#) gives the TPM of any type g as:

$$\mathbf{M} =_g \mathbf{I} - \mathbf{A}_g + \mathbf{A}_g \mathbf{M} \quad \text{for } 1 \leq g \leq G - 1 \quad (2.35)$$

where \mathbf{I} is the $J \times J$ identity matrix, $\mathbf{A}_g = \text{diag}(\lambda_{i,g})$ and $\mathbf{M} = \mathbf{M}_G$ (*i.e.*, $\mathbf{A}_G = \mathbf{I}$), subject to $0 \leq \lambda_{i,g} \leq \frac{1}{1-m_{ii}}$ ($\forall i \in J$) and $0 \leq m_{ii} \leq 1$, where m_{ii} are the main diagonal elements of matrix \mathbf{M} .

Within this specification, the $\lambda_{i,g}$ parameters give information about heterogeneity in the rates of movement across homogeneous agent types:

- $\lambda_{i,g} = 0$ if type- g agents originally in category i never move out of i ;
- $0 < \lambda_{i,g} < 1$ if type- g agents originally in category i move at a lower rate than the generator matrix \mathbf{M} ;

- $\lambda_{i,g} > 1$ if type- g agents originally in category i move at a higher rate than the generator matrix M .

Then the expected time spent in category i by type- g agents is given by $\frac{1}{(\lambda_{i,g}(1-m_{ii}))}$ ($\forall \lambda_{i,g} > 0$).

Formally, an ill-posed problem arises when the number of parameters to be estimated is greater than the number of available observations. With the general formulation of the mixed Markov chain model (M-MCM) stated in equation (2.34), the number of parameters to be estimated is given by $L^G = (G-1) \times J + G \times J \times (J-1)$, where G is the number of homogeneous types and J is the number of size categories. On the right hand side of this expression, the first term of the sum corresponds to the number of homogeneous type shares and the second term of the sum corresponds to the total number of transition probabilities. This number increases rapidly with the number of homogeneous groups and the number of size categories. The larger the number of parameters to estimate, the harder the estimation. Though she does not state it explicitly, the parametrisation introduced by [Frydman \(2005\)](#) in equation (2.35) can help reduce the number of parameters to be estimated. Indeed, under Frydman's formulation of the M-MCM, the number of parameters to be estimated is $L^F = (G-1) \times J + (G-1) \times J + J \times (J-1)$, where the first term of the sum corresponds again to the number of homogeneous type shares, the second term of the sum corresponds to the number of 'rate-of-movement' vector entries \mathbf{A}_g , and the third term of the sum corresponds to the number of transition probabilities in matrix M . Under this parametrisation, $L^F = 30$ with $G = 2$ and $J = 5$, $L^F = 170$ with $G = 5$ and $J = 10$, and $L^F = 740$ with $G = 10$ and $J = 20$. As we can see, the number of parameters to be estimated is sharply decreased: it is reduced by one third in our case, by two thirds in the second example, and is divided by more than five with the very large, untypical, values for G and J .

C: Maximum likelihood estimation of the mixed Markov chain model (M-MCM)

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

$$l_{k,g} = s_{i,g} \prod_{i \neq j} (m_{ij,g})^{\nu_{ij,k}} \prod_i (m_{ii,g})^{\nu_{ii,k}} \quad (2.36)$$

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

Under [Frydman \(2005\)](#)'s specification of the M-MCM as defined by equation (2.35) the above likelihood can be rewritten as:

$$l_{k,g} = s_{i,g} \prod_{i \neq j} (\lambda_{i,g} m_{ij})^{\nu_{ij,k}} \prod_i (1 - \lambda_{i,g} + \lambda_{i,g} m_{ii})^{\nu_{ii,k}} \quad (2.37)$$

where $\lambda_{i,g}$ is the relative rate of movement of type- g agents.

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

$$\log L = \sum_{k=1}^N \sum_{g=1}^G (Y_{k,g} \log l_{k,g}) \quad (2.38)$$

where $Y_{k,g}$ is an indicator variable which equals 1 if agent k belongs to type g and 0 otherwise. Note that the log-likelihoods of the MCM and MSM can easily be derived from equation (2.38) by stating, respectively, $G=1$ and $\lambda_{i,1} = 1$ for the MCM and $G=2$, $\lambda_{i,1} = 0$ and $\lambda_{i,2} = 1$ for the MSM.

The maximum likelihood estimators, $\hat{s}_{i,g}$, $\hat{\lambda}_{i,g}$, \hat{m}_{ii} and \hat{m}_{ij} , are obtained using the EM algorithm in a similar way to that described in the main text.

Chapter 3

Unobserved farm heterogeneity and agricultural policy impacts on size dynamics

Abstract

This chapter aims at investigating the impact of agricultural policies on structural change in farming. As not all farms may behave alike, a mixed Markov chain modelling approach is applied to capture unobserved heterogeneity in transition process of farms. A multinomial logit specification is used for transition probabilities and the parameters are estimated by maximum likelihood method and the expectation-maximisation (EM) algorithm. An empirical application to an unbalanced panel from 2000 to 2013 shows that French farms mainly consist of a mixture Two farm types characterised by specific transition processes. The main finding is that the impact of farm subsidies from both pillars of the Common Agricultural policy (CAP) highly depends on the type membership of farms. This result suggests that more attention should be paid to both observed and unobserved farm heterogeneity in assessing agricultural policy impact on structural change in farming.

3.1 Introduction

The farming sector has faced important structural change over recent decades. In particular, the number of farms has been decreasing sharply and their average size has been increasing continually in most developed countries, implying changes in farm-size distribution. Such changes may

have important consequences for equity among farmers, productivity and efficiency of farming (Weiss, 1999; Breustedt and Glauben, 2007). Structural change has been therefore the subject of considerable interest to agricultural economists and policy makers. Many theoretical studies pointed out potential impacts of agricultural policies on farm-size changes, but these impacts remain ambiguous (Goddard *et al.*, 1993; Harrington and Reinsel, 1995). Since then, several studies empirically investigated the impact of some recent policy programmes on structural change in various farming contexts.

It has become common in agricultural economics to study structural change in farming and the impact of time-varying variables including agricultural policies by using the so-called Markov chain model (MCM). Focusing on farm number and size evolution, the Markovian framework has proved to be a convenient modelling approach to represent the transition process of farms across categories (Bostwick, 1962; Padberg, 1962; Krenz, 1964). Basically, this model states that farm size at a given time period is the result of a probabilistic process where future farm size only depends on size in the immediately previous period since, in general, a first order process is assumed. Stokes (2006) showed that a Markovian transition process may derive from a structural model of inter-temporal profit maximisation, giving theoretical grounds to using the MCM.

This chapter aims at investigating the impacts of farm subsidies from the Common agricultural Policy (CAP) on farm structural change, especially in France. It adds to the existing literature in two ways. Firstly, a mixture modelling framework is applied to take into account potential unobserved farm heterogeneity in the analysis of structural change in farming. Secondly, transition probabilities are specified at the individual farm level within a discrete choice modelling approach to analyse transition process at farm level.

In agricultural economics, the heterogeneity issue has been so far mostly left in the background of farm structural change analysis, maybe because previous studies generally focused on specific farm types and regions. Two reasons may explain why there has been so few studies that explicitly account for farm heterogeneity. Firstly, most previous studies focus on specific farm types. For example, several studies analysed structural change in dairy farming, especially in some regions of United States of America (Chavas and Magand, 1988; Zepeda, 1995a; Stokes, 2006), in Ireland (Keane, 1991; Gaffney, 1997; McInerney and Garvey, 2004), in France (Ben Arfa *et al.*, 2015). Other analysis focused on other types of farming such as pig farms (Judge and Swanson, 1962; Disney *et al.*, 1988; Gillespie and Fulton, 2001; Karantinis, 2002), cash grain farms (Bostwick, 1962; Garcia *et al.*, 1987). Authors may therefore have assumed that such groups are homogeneous enough to justify discarding heterogeneity issues. Secondly, studies that investigate the potential drivers of structural change in farming emphasise the role of economic factors such as market or policy variables. Authors have therefore plainly left heterogeneity issues in the background, considering it as a control variable rather than one of key interest.

Furthermore, previous studies of farm structural change often faced data limitations. Indeed, structural change in farming has generally been investigated using aggregate data, *i.e.*, cross-sectional observations of the farm distribution across a finite number of size categories, because such data are most often available to the researchers. It could be complicated even impossible to control for farm heterogeneity using such data.

However, accounting for heterogeneity may be crucial to fully understand structural change in farming because this results from individual farmers' decisions (Freshwater and Reimer, 1995; Bollman *et al.*, 1995; Jackson-Smith, 1999). Farm heterogeneity may originate from several sources. One of the most important sources of farm heterogeneity is farmers' motivation. While the standard economic theory of production states that farmers are normally expected to maximise their total profit from farming activities, it has been shown that not all farmers actually give a priority to profit maximisation (Maybery *et al.*, 2005; Mzoughi, 2011; Howley *et al.*, 2014). It is the case of environmentally oriented farms (Willock *et al.*, 1999) or some hobby farms (Daniels, 1986; Holloway, 2002). The existence of non-financial/pecuniary motives or potential farming lifestyle values may shape farmers' behaviours (Hallam, 1991; Harrington and Reinsel, 1995; Howley, 2015). The ability to change operated farm size may also depend on some other factors such as accessibility to inputs (land, new technology), managerial capacity, risk perception, risk tolerance of farmers, etc. (Bowman and Zilberman, 2013; Conradt *et al.*, 2014; Trujillo-Barrera *et al.*, 2016). Therefore, the implicit homogeneity assumption of the usual MCM, *i.e.*, that all farms have the same probability to change category of sizes given their initial size, is likely to not hold.

Previous studies which controlled for farm heterogeneity in modelling structural change using the MCM approach only accounted for observed sources of heterogeneity so far (see Zimmermann and Heckelei (2012) for a recent example). This appears to be restricted since not all sources of farm heterogeneity are observable or can be linked to observed farm and farmer characteristics. The mixed Markov chain model (M-MCM) applied here captures unobserved heterogeneity in the transition process of farms. The existence of different transition processes in farming may reflect heterogeneity in farmer behaviours, which may relate both to observed and unobserved farm and farmer characteristics.

Mixture models have a long story in the literature and have proved to have several advantages over alternative techniques also used in order to account for unobserved heterogeneity (Greene and Hensher, 2003; Garver *et al.*, 2008). While the M-MCM has been the subject of considerable applications in some other strands of economics, we are the first to evidence heterogeneity in farm transition process, to the best of our knowledge. Chapter 2 of this thesis showed that even a restricted mover-stayer model represents more efficiently farm size dynamics than the usual MCM, as has been evidenced in other strands of economics (Blumen *et al.*, 1955; Frydman *et al.*, 1985; Major, 2008; Cipollini *et al.*, 2012). The present chapter extends this

approach in two directions. Firstly, it allows for more than two types of farms and also relaxes the ‘pure stayer’ assumption. Secondly, it develops a non-stationary approach to study the impact of agricultural policies on farm size dynamics. This approach leads to separate the farm population into meaningful clusters with different transition patterns (Vermunt, 2010). As agricultural policy impacts may depend on some observed and unobserved farm and farmer characteristics, a mixture approach may therefore provide more informative results in the analysis of structural change.

Since Lee *et al.* (1965), Lee *et al.* (1970) and MacRae (1977) showed that robustly estimating an MCM from aggregate data is possible, structural change has been investigated mostly using aggregate data because such data are most often easier to obtain than individual-level data. Structural change determinants are thus usually investigated at a macro level. Focusing on agricultural policies, the impacts of public supports on transition probabilities of farms across category of sizes are generally investigated using transition probability matrices estimated for the overall population of farms (Huettel and Jongeneel, 2011; Zimmermann and Heckelei, 2012; Ben Arfa *et al.*, 2015). Even when individual level data are used, yearly transition probability matrices are first computed for the overall population of farms, effects of exogenous variables being estimated on transition probabilities in a second step (Hallberg, 1969; Ethridge *et al.*, 1985; Rahelizatovo and Gillespie, 1999). In doing so, individual effects of farms cannot be taken into account. The discrete choice approach adopted here to model farm transitions enables to more easily incorporate individual farm effects on structural change.

This chapter is structured as follows. Section 3.2 presents the proposed model, the specification method for transition probabilities and the estimation procedure. In Section 3.3, methods for assessing and interpreting the model’s result are presented. The two next sections report the application to a panel of French farms, starting with a description of the data used and explanatory variables investigated in Section 3.4, followed by a presentation of the results in Section 3.5. Finally, concluding remarks are provided in Section 3.6 with some considerations on possible improvements of this study for further research.

3.2 The non-stationary mixed Markov chain model (M-MCM)

Let N be the total number of farms in the population and K the total number of farm size categories or choice alternatives. As farm size categories are observed at discrete times, generally 1-year intervals, a discrete-time process is assumed. Denote by y_{it} the category of size of a specific farm i ($i \in N$) at time t ($1 \leq t \leq T$). The indicator variable y_{i1} is equal to j ($\forall j = 0, 1, 2, \dots, K$) if farm i is in category j at time $t = 1$. The category $j = 0$ may indicate entry in or exit from farming. Since farms may enter or leave the farming sector at different dates, the length of

time T_i a farm is observed may vary across farms (*i.e.*, $T_i \leq T$). Over time period T_i , the size evolution of a specific farm i can therefore be represented by the vector $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT_i})$.

In agricultural economics, farm movements across size categories are generally supposed to follow a first order Markov process (Zimmermann *et al.*, 2009). It is thus assumed that the size category of farm i at any time t (y_{it}) only depends on its immediately preceding state, *i.e.*, its size category at time $t-1$ (y_{it-1}). The Markov assumption implies that the observed random variables ($y_{i1}, y_{i2}, \dots, y_{iT_i}$) are not independent from each other (Dias and Willekens, 2005). Then, the probability function describing farm movements across size categories can be derived as:¹

$$f(\mathbf{y}_i) = \prod_{t=1}^{T_i} P(y_{it}|y_{it-1}) \quad (3.1)$$

where $P(y_{it}|y_{it-1})$ is the transition probability, that is, the probability that farm i chooses a specific size category at time t given its state at time $t-1$.

Suppose now that the observed sample of farms is divided into G homogeneous types instead of just one, each type gathering farms with a similar transition process. The density function of \mathbf{y}_i , as a discrete mixing distribution with G support points, can be rewritten (McLachlan and Peel, 2001):

$$f(\mathbf{y}_i) = \sum_{g=1}^G \pi_g f_g(\mathbf{y}_i) \quad (3.2)$$

where $f_g(\mathbf{y}_i)$ is the probability function describing farm size dynamics in type g as specified in equation (3.1); and π_g , the mixing proportions, are non-negative and sum up to one. In statistics, π_g is called the mixing distribution and $f_g(\mathbf{y}_i)$ is called the mixed function (Lindsay and Lesperance, 1995). As we defined a finite number of farm types, the mixed model can be also called a ‘latent class model’ with G latent transition processes (Garver *et al.*, 2008; Train, 2009). The density function of \mathbf{y}_i is thus conditional on the mixing distribution and we can represent farm size dynamics as:

$$f(\mathbf{y}_i) = \sum_{g=1}^G P(g_i = g) \left[\prod_{t=1}^{T_i} P(y_{it} = k | y_{it-1} = j, g_i = g) \right] \quad (3.3)$$

From equation (3.3), it can be seen that under the M-MCM, farm size evolution thus depends on two sets of probabilities. The first term represents the probabilities that farm i belongs to a specific farm type g while the second term represents the probabilities of making transitions across size categories given that farm i belongs to type g .

¹This formulation does not account for the entry and exit processes in farming because of data limitations underlying in Section 3.4.1.

3.2.1 Model specification

Both parts of equation (3.3) can be specified as a function of exogenous variables (Vermunt, 2010). In this case, the model is fully parametric. However, it has been shown that the resulting estimates could depend on the specification of the mixing distribution (Heckman and Singer, 1984); and the models' parameters may be biased when mis-specification occurred (Wedel, 2002). Furthermore, a parametric approach is convenient when performing discriminant analysis, that is, if we know a priori the total number of types in the population which depends on socio-economic characteristics (Garver *et al.*, 2008). Because we have no information about farm types, a semi-parametric approach is applied here. Only transition probabilities are specified as a function of exogenous variables.

As size categories of farms are finite, exhaustive and mutually exclusive, a discrete-choice approach is used to specify transition probabilities. The discrete choice approach assumes that farmers' choice of the initial size category as well as to making some transitions across categories can be represented by a random utility model (Train, 2009). Farmer's utility may represent in our case the net benefit that arises from choosing (*i.e.*, moving to) a specific size category given the preceding state.

Denote by U_{ijkt} the utility of farm i arising upon moving from category j to category k between time periods t and $t + 1$. Under the standard behavioural assumption, it is supposed that farmers choose the category which maximises their utility. Therefore, a move from j to k (*i.e.*, $y_{it} = k | y_{it-1} = j$) will be observed if and only if $U_{ijkt} \geq U_{ijlt}$ ($\forall j, k, l \in K$). Farms staying in the same category two consecutive dates are considered as making a transition from j to j . Under a mixture assumption, the utility level of a farm is conditional on its specific type g . A farm belonging to type g will thus make a transition from a specific category j to category k if and only if $U_{ijkt|g} \geq U_{ijlt|g}$ ($\forall g \in G$), where $U_{ijkt|g}$ is the utility of farms conditional on belonging to type g .

As the utility of farmers may vary over time because of changes in their social and economic environment, transition probabilities are specified as a function of a set of causal factors \mathbf{x} . Under the mixture assumption, the utility that would accrue to farm i upon moving from category j to category k at time t given type g can be expressed as:

$$P(y_{it} = k | y_{it-1} = j, g_i = g, \mathbf{x}_{it-1}) = P(U_{ijkt|g} \geq U_{ijlt|g}) \quad (3.4)$$

$$p_{ijkt|g} = f(\mathbf{x}_{it-1}, \beta_g, \epsilon_{ijkt|g}), \quad \forall t \in T_i \quad j, k, l \in K \quad g \in G$$

where \mathbf{x}_{it-1} are the considered explanatory variables, β_g are the parameters to be estimates and $\epsilon_{ijkt|g}$ an *iid* random error term specific to farm type g . Explanatory variables are lagged 1-year since farmers' decisions for entering or the farming sector as well as for expansion or contraction

are likely to depend on information available during the previous period (Zepeda, 1995b). According to equation (3.4), the utility function of farmers thus consist in an observed part ($V_{ijkt|g}$) which depends on some observable individual farm level exogenous variables and a stochastic component ($\epsilon_{ijkt|g}$) which is the difference between the utility level that farm i experiments and the representation obtained from the chosen explanatory variables (Train, 2009).

Because farmers face multiple choices at each occasion, it is econometrically convenient to use a multinomial specification (Greene, 2006). This specification assumes that the error terms $\epsilon_{ijkt|g}$ are randomly drawn from a Gumbel distribution.² The density for each unobserved (stochastic) component of utility is given by (Train, 2009):

$$f(\epsilon_{ijkt|g}) = e^{-\epsilon_{ijkt|g}} e^{-e^{-\epsilon_{ijkt|g}}}$$

with cumulative distribution

$$F(\epsilon_{ijkt|g}) = e^{-e^{-\epsilon_{ijkt|g}}}$$

The difference between two extreme values thus follows the logistic distribution (Train, 2009):

$$F(\epsilon_{ijkt|g} - \epsilon_{g,ijlt}) = \frac{e^{(\epsilon_{ijkt|g} - \epsilon_{g,ijlt})}}{1 + e^{(\epsilon_{ijkt|g} - \epsilon_{g,ijlt})}}$$

According to equation (3.4) and following McFadden (1974), the probability that farm i makes a transition from category j to category k at time t given type g can be written as:

$$p_{ijkt|g} = P(\epsilon_{ijkt|g} \geq \epsilon_{ijlt|g} + V_{ijkt|g} - V_{ijlt|g} \quad \forall k \neq l).$$

Since the $\epsilon_{ijkt|g}$ are independent, the choice probability can be derived as:

$$p_{ijkt|g} = \int (p_{ijkt|g} | \epsilon_{ijkt|g}) f(\epsilon_{ijkt|g}) d\epsilon_{ijkt|g}$$

with

$$p_{ijkt|g} | \epsilon_{ijkt|g} = \prod_{k \neq l} F(\epsilon_{ijkt|g})$$

the cumulative distribution over all $j \neq k$. Some mathematical manipulation of the above equation leads to the logit multinomial expression for the conditional probability of making a transition from category j to category k at date t :

$$P(y_{it} = k | y_{it-1} = j, g_i = g, \mathbf{x}_{it-1}) = \frac{\exp(\beta'_{jk|g} \mathbf{x}_{it-1})}{\sum_{l=1}^K \exp(\beta'_{jl|g} \mathbf{x}_{it-1})} \quad (\forall j, k = 1, 2, \dots, K) \quad (3.5)$$

²It should be noted that, the multinomial logit model is often used in the agricultural economics literature to specify transition probabilities even when using aggregate data (see Zimmermann and Heckelei (2012) for a recent example). Storm *et al.* (2016) also used this specification method to generate individual transition probabilities for farms in a Monte Carlo simulation.

where $\beta_{jk|g}$ is a vector of parameters specific to each farm type g and each jk transition. Choosing that staying in the same category two consecutive years is the reference leads to set $\beta_{jj|g} = \mathbf{0} \quad \forall g = 1, 2, \dots, G$ and $\forall j = 1, 2, \dots, K$ for identification purposes. This constrains to compute $K - 1$ logarithmic odds ratios for each initial size category j as:

$$\ln \left[\frac{P(y_{it} = k | y_{it-1} = j, g_i = g, \mathbf{x}_{it-1})}{P(y_{it} = j | y_{it-1} = j, g_i = g, \mathbf{x}_{it-1})} \right] = (\beta'_{jk|g} - \beta'_{jj|g}) \mathbf{x}_{it-1} \quad (3.6)$$

3.2.2 Estimation procedure

The parameters of the model are estimated using the maximum likelihood estimation method. Let $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ be the observed random sample of farms obtained from the mixture density, where the vector $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT_i})$ gathers farm i states over time period $T_i \leq T$. Let denote the conditional transition probability of moving from category j to category k by:

$$P(y_{it} = k | y_{it-1} = j, g_i = g, \mathbf{x}_{it-1}) = p_{ijk|g} = P(\mathbf{x}_{it-1}; \beta_{jk|g})$$

Under a mixture assumption, the log-likelihood (LL) function for the parameters (β) of the model, conditional on observing \mathbf{y} , writes:

$$LL(\beta) = \sum_{i=1}^N \ln \left\{ \sum_{g=1}^G \pi_g \prod_{t=1}^{T_i} \prod_{j,k} [P(\mathbf{x}_{it-1}; \beta_{jk|g})]^{d_{ijk|g}} \right\} \quad (3.7)$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_G)$ is a matrix of parameters with $\beta_g = \{\beta_{jk|g}\} \quad \forall g \in G$ and $j, k = 1, 2, \dots, K$; the indicator $d_{ijk|g}$ is equal to 1 if farm i moves from category j to category k at time t (*i.e.*, $y_{it} = k | y_{it-1} = j$) and zero otherwise. Since farm type is unknown beforehand and given some numerical difficulties associated to the maximisation of the above expression, the expectation-maximisation (EM) algorithm is generally used to estimate the parameters of such models (McLachlan and Krishnan, 2007).³ The EM algorithm developed by Dempster *et al.* (1977) simplifies the complex log-likelihood in equation (3.7) into a set of easily solvable log-likelihood functions by introducing a so-called ‘missing variable’.

Let v_{ig} be a discrete unobserved variable indicating the type-membership of each farm. The random vector $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{iG})$ is thus g -dimensional with $v_{ig} = 1$ if farm i belongs to type g and zero otherwise. Assuming that v_{ig} is unconditionally multinomial distributed with probability π_g , the complete likelihood for (β, π) , conditional on observing $\mathbf{y}_c = (\mathbf{y}, \mathbf{v})$, therefore

³The likelihood does not yield to an explicit solution for the model parameters. The EM algorithm transforms the maximisation of a log of sums into a recursive maximisation of the sum of logs (McLachlan and Krishnan, 2007).

writes:⁴

$$L_c(\boldsymbol{\beta}, \boldsymbol{\pi}) = \prod_{i=1}^N \prod_{g=1}^G \left\{ \pi_g \prod_{t=1}^{T_i} \prod_{j,k}^K \left[P(\mathbf{x}_{it-1}; \boldsymbol{\beta}_{jk|g}) \right]^{d_{ijkt}} \right\}^{v_{ig}} \quad (3.8)$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_G)$ is the vector gathering shares of farm types which has also to be estimated. The complete log-likelihood is thus obtained as:

$$LL_c(\boldsymbol{\beta}, \boldsymbol{\pi}) = \sum_{i=1}^N \sum_{g=1}^G v_{ig} \ln \left\{ \pi_g \prod_{t=1}^{T_i} \prod_{j,k}^K \left[P(\mathbf{x}_{it-1}; \boldsymbol{\beta}_{jk|g}) \right]^{d_{ijkt}} \right\} \quad (3.9)$$

In this case, v_{ig} is called the ‘posterior’ probability that farm i belongs to type g given that \mathbf{y}_i has been observed, that is $P(v_{ig} = 1 | \mathbf{y}_i)$, while π_g is the ‘prior’ probability of the mixture (McLachlan and Peel, 2001). This log-likelihood can then be split into two components:

$$\begin{aligned} LL_1 &= \sum_{i=1}^N \sum_{g=1}^G v_{ig} \ln \pi_g \\ LL_2 &= \sum_{i=1}^N \sum_{g=1}^G v_{ig} \sum_{t=1}^{T_i} \sum_{j,k}^K d_{ijkt} \ln \left\{ P(\mathbf{x}_{it-1}; \boldsymbol{\beta}_{jk|g}) \right\} \end{aligned} \quad (3.10)$$

As the farm type is not observed, the posterior probability that farm i belongs to type g has to be estimated from the observations. The EM algorithm therefore consists of the four following steps:

(i) Initialisation: Arbitrarily choose initial values $\boldsymbol{\Psi}^0 = (\boldsymbol{\Phi}_1^0, \boldsymbol{\Phi}_2^0, \dots, \boldsymbol{\Phi}_G^0)$ where $\boldsymbol{\Phi}_g^0 = (\pi_g^0, \boldsymbol{\beta}_{jk|g}^0)$ $\forall j, k = 1, 2, \dots, K$ and $\forall g = 1, 2, \dots, G$ for the parameters of the model, with $\boldsymbol{\beta}_{jj|g}^0$ set to zero for identification as previously mentioned in Section 3.2.1.

(ii) Expectation: At iteration $p + 1$ of the algorithm, compute the expected probability that farm i belongs to a specific type g while observing \mathbf{y}_i and given parameters $\boldsymbol{\Psi}^p$. This conditional expectation probability, that is, the posterior probability $v_{ig}^{(p+1)} = v_{ig}(\mathbf{y}_i; \boldsymbol{\Psi}^p)$, can be obtained according to the Bayes’ law:

$$v_{ig}^{(p+1)} = \frac{\pi_g^p \prod_{t=1}^{T_i} \prod_{j,k}^K \left[P(\mathbf{x}_{it-1}; \boldsymbol{\beta}_{jk|g}^p) \right]^{d_{ijkt}}}{\sum_{h=1}^G \pi_h^p \prod_{t=1}^{T_i} \prod_{j,k}^K \left[P(\mathbf{x}_{it-1}; \boldsymbol{\beta}_{jk|h}^p) \right]^{d_{ijkt}}} \quad (3.11)$$

Replacing v_{gi} in equation (3.9) by its expected value $v_{ig}^{(p+1)}$ leads to the conditional expectation of the complete data log-likelihood.

⁴This assumption means that the distribution of the complete-data vector (\mathbf{y}_c) implies the appropriate distribution for the incomplete-data vector (\mathbf{y}) (McLachlan and Peel, 2001).

(iii) Maximisation: Update Ψ^p by maximising the complete log-likelihood conditional on the observations. The model parameters are thus updated as:

$$\beta^{(p+1)} = \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^N \sum_{g=1}^G v_{ig}^{(p+1)} \sum_{t=1}^{T_i} \sum_{j,k}^K d_{ijkt} \ln \left[P(\mathbf{x}_{it-1}; \beta_{jk|g}) \right] \quad (3.12)$$

The maximisation process of the above equation is straightforward. The transition probability parameters (β^p) are updated considering $v_{gi}(\mathbf{y}_i; \Psi^p)$ as a weighting factor for each observation (Pacífico and Yoo, 2012). Given equation (3.5), the log-likelihood function for the transition probabilities rewrites:

$$\begin{aligned} LL_2(\beta) &= \sum_{i=1}^N \sum_{g=1}^G v_{ig}^{(p+1)} \sum_{t=1}^{T_i} \sum_{j,k}^K d_{ijkt} \ln \left[\frac{\exp(\beta'_{jk|g} \mathbf{x}_{it-1})}{\sum_{l=1}^K \exp(\beta'_{jl|g} \mathbf{x}_{it-1})} \right] \\ &= \sum_{i=1}^N \sum_{g=1}^G v_{ig}^{(p+1)} \sum_{t=1}^{T_i} \sum_{j,k}^K d_{ijkt} \left\{ (\beta'_{jk|g} \mathbf{x}_{it-1}) - \ln \left[\sum_{l=1}^K \exp(\beta'_{jl|g} \mathbf{x}_{it-1}) \right] \right\} \end{aligned} \quad (3.13)$$

The first-order derivative of this log-likelihood function then becomes:⁵

$$\begin{aligned} \frac{\partial LL_2(\beta)}{\partial \beta_{jk|g}} &= \sum_{i=1}^N \sum_{g=1}^G v_{ig}^{(p+1)} \sum_{t=1}^{T_i} \sum_j^K d_{ijkt} \mathbf{x}_{it-1} - \sum_{i=1}^N \sum_{g=1}^G v_{ig}^{(p+1)} \sum_{t=1}^{T_i} \sum_{j,k}^K d_{ijkt} \left[P(\mathbf{x}_{it-1}; \beta_{jk|g}) \mathbf{x}_{it-1} \right] \\ &= \sum_{i=1}^N \sum_{g=1}^G v_{ig}^{(p+1)} \sum_{t=1}^{T_i} \sum_{j=1}^K d_{ijkt} \mathbf{x}_{it-1} - \sum_{i=1}^N \sum_{g=1}^G v_{ig}^{(p+1)} \sum_{t=1}^{T_i} \sum_j^K \left[P(\mathbf{x}_{it-1}; \beta_{jk|g}) \mathbf{x}_{it-1} \right] \\ &= \sum_{i=1}^N \sum_{g=1}^G v_{ig}^{(p+1)} \sum_{t=1}^{T_i} \sum_{j=1}^K \left[d_{ijkt} - P(\mathbf{x}_{it-1}; \beta_{jk|g}) \right] \mathbf{x}_{it-1} \end{aligned} \quad (3.14)$$

The model's parameters are obtained by solving the equation $\partial LL_2(\beta)/\partial \beta = 0$.

Then, the probabilities of belonging to farm type g are updated as follows :

$$\pi_g^{(p+1)} = \frac{\sum_{i=1}^N v_{ig}^{(p+1)}}{\sum_{i=1}^N \sum_{h=1}^G v_{ih}^{(p+1)}}, \quad \forall g \in G \quad (3.15)$$

(iv) Iteration: Return to expectation step (ii) using $\pi^{(p+1)}$ and $\beta^{(p+1)}$ and iterate until convergence of the observed log-likelihood given by equation (3.7). At convergence, the resulting parameters are considered as the optimal estimators ($\hat{\Psi}$).

A problem which often occurs in a mixture analysis with several components is that some solutions may be suboptimal. Indeed, the non-concavity of the log-likelihood function in

⁵Note that $\sum_k^K d_{ijkt} = 1$ since farm i can be in only one category at time t .

equation (3.7) does not allow the identification of a global maximum in the mixture model, even for discrete mixtures of multinomial logit (Hess *et al.*, 2006). Given the potential presence of a number of local maxima, the EM solutions may be highly dependent on the initial values chosen for Ψ^0 . Various techniques are used in the literature to avoid suboptimal solutions (see Baudry and Celeux (2015) for a short review). In the present study, the EM algorithm was run several times with various randomly chosen initial values and those providing the largest likelihood at convergence were chosen as the best ones.

3.3 Farm types and elasticities of explanatory variables

3.3.1 Choosing the optimal number of farm types

The total number of components for a mixture model can be chosen either thank to an ‘*a priori*’ assumption or via information criteria. In the latter case, selection criteria are generally based on the value of $-2LL_G(\mathbf{y}; \hat{\Psi})$ of the corresponding model, where matrix $\hat{\Psi}$ represents the maximum likelihood estimates adjusted for the number of free parameters in the model with a total of G homogeneous types. The basic principle under these information criteria is parsimony, that is, all other things being the same, the model with fewer parameters is preferred (Andrews and Currim, 2003). The selection criteria derive from the following formula:

$$C_G = -2 \left\{ LL_G(\mathbf{y}; \hat{\Psi}) \right\} + \kappa N_G \quad (3.16)$$

where $LL_G(\mathbf{y}; \hat{\Psi})$ is the overall population log-likelihood value computed with the resulting estimated parameters for the model specified with G types; N_G is the total number of free parameters in the model and κ is a penalty constant. Different values of κ lead to the two well known information criteria: the Akaike Information Criterion (AIC) with $\kappa = 2$ and the Bayesian Information Criterion (BIC) with $\kappa = \log N$ with N the total number of observations. Other information criteria can also be derived such as the Consistent Akaike Information Criterion (CAIC) with $\kappa = \log N + 1$ and the modified AIC (AIC3) with $\kappa = 3$.

There is no consensus in the literature about the specific type of criteria to choose an optimal number of components for a mixture model. However, some studies suggest that the CAIC and AIC3 may be more useful in the context of mixture models since these criteria more severely penalise the addition of parameters (Andrews and Currim, 2003; Dias and Willekens, 2005; Sarstedt, 2008).⁶

⁶It should be noted that based on Monte Carlo simulations the AIC3 is the best segment-retention criterion for a large variety of multinomial data configurations (Andrews and Currim, 2003) as well as using Markov modelling approaches (Dias and Willekens, 2005). The choice of a good criterion were based on the stability of the resulting estimated parameters and the standard errors variation.

3.3.2 Probability elasticities

The model tests whether the investigated exogenous variables have significant impacts on farm transition probabilities. As the estimated coefficients indicate marginal effects on the log-odds ratios of transition probabilities, the impacts of the explanatory variables are difficult to interpret directly (Greene, 2006). In this case, the impacts of explanatory variables are usually evaluated in terms of elasticities. The ‘probability elasticities’ measure the effect of a 1% change in the i th explanatory variable on the transition probabilities (Zepeda, 1995a). In the mixture model these probability elasticities may depend on farm type. Yearly transition-probability elasticities for farms belonging to a specific type g are obtained as:

$$\delta_{jkt|g} = \frac{\partial p_{jkt|g}}{\partial \mathbf{x}_{t-1}} \times \frac{\mathbf{x}_{t-1}}{p_{jkt|g}}, \quad \forall j, k \in K \quad \forall g \in G \quad (3.17)$$

where $\delta_{jkt|g}$ is a vector gathering elasticities at the means of the explanatory variables in vector \mathbf{x}_{t-1} ; and $p_{jkt|g}$ is the probability to move from category j to category k at time period t conditional on belonging to type g . The first term of equation (3.17) thus represents the marginal effects of explanatory variables and is given by (Greene, 2006):

$$\frac{\partial p_{jkt|g}}{\partial \mathbf{x}_{t-1}} = p_{jkt|g} \left(\beta_{jk|g} - \sum_{l=1}^K \beta_{jl|g} p_{jlt|g} \right) \quad (3.18)$$

where $\beta_{jk|g}$ is the vector of estimated parameters.

Replacing the marginal effects in equation (3.17) leads to expressing the transition-probability elasticities as:

$$\delta_{jkt|g} = \left(\beta_{jk|g} - \sum_{l=1}^K \beta_{jl|g} p_{jlt|g} \right) \mathbf{x}_{t-1} \quad (3.19)$$

Given the constraint that $\beta_{jj|g} = 0$ for identification, the probability elasticities for the reference pair jj is obtained as:

$$\delta_{jkt|g} = \left(- \sum_{l=1}^K \beta_{jl|g} p_{jlt|g} \right) \mathbf{x}_{t-1}, \quad \forall g \in G \quad (3.20)$$

3.3.3 Farm structure elasticities

Yearly ‘structure elasticities’ are also derived to measure the impacts of exogenous variable on the distribution of farms across size categories. Farm structure elasticities measure the percentage change in the total number of farms in a specific category j at time t for a 1% change in the investigated explanatory variable (Zepeda, 1995a).

Under the mixture modelling framework, the total number of farms in a specific category k at time t can be obtained as:

$$n_{kt} = \sum_{j=1}^G \pi_g \sum_j^K n_{jt-1} p_{jkt|g}, \quad \forall k \in K, \quad \forall t \in T \quad (3.21)$$

where π_g is the probability of belonging to type g ; n_{jt-1} is the total number of farms in category j at time $t - 1$; and $p_{jkt|g}$ is the probability for farm i to make a transition from category j to category k at time t . Farm structure elasticities are then defined as:

$$\boldsymbol{\eta}_{kt} = \frac{\partial n_{kt}}{\partial \mathbf{x}_{t-1}} \times \frac{\mathbf{x}_{t-1}}{n_{kt}} \quad (3.22)$$

In equation (3.21), only transition probabilities depend on exogenous variables (\mathbf{x}_{t-1}). Farm structure elasticities can be therefore obtained using the corresponding probability marginal effects in equation (3.18). At any specific time t , farm structure elasticities are thus derived as:

$$\boldsymbol{\eta}_{kt} = \left(\sum_{g=1}^G \pi_g \sum_{j=1}^K n_{jt-1} \frac{\partial p_{jkt|g}}{\partial \mathbf{x}_{t-1}} \right) \frac{\mathbf{x}_{t-1}}{n_{kt}} \quad (3.23)$$

where the marginal effects at the means of the corresponding explanatory variables are replaced by their values derived in the previous subsection.

3.4 Empirical application

3.4.1 Data

For the empirical application, an unbalanced panel from the ‘Réseau d’Information Comptable Agricole’ (RICA) database is used. RICA is the French implementation of the Farm Accountancy Data Network (FADN) and data are available from 2000 to 2013. FADN is an annual survey which is defined at the European Union (EU) level and is carried out in each member state. The information collected at the individual level relates to both the physical and structural characteristics of farms and their economic and financial characteristics. In particular, this database provides information about the total subsidies received by farms from the Common Agricultural Policy (CAP) (see <http://ec.europa.eu/agriculture/rica/index.cfm> to learn more about FADN). In France, RICA is produced and disseminated by the statistical and foresight office of the French ministry for agriculture. It focuses on ‘medium and large’ farms and constitutes a stratified and rotating panel of approximately 7,000 farms surveyed each year. Some 10% of the sample is renewed every year so that, on average, farms are observed during 5 consecutive years. However, some farms may be observed only once, and others several, yet not consecutive, years. Some farms remained in the database over the whole studied period,

i.e., fourteen consecutive years. Each farm in the dataset is assigned a weighting factor which reflects its stratified sampling probability, allowing for extrapolation at the population level (see <http://www.agreste.agriculture.gouv.fr/> to learn more about RICA France).

This empirical study concentrates on farm size as defined in economic terms in order to consider all farms in the sample whatever their type of production. In accordance with the EU regulation (CE) N°1242/2008, European farms are classified into fourteen economic size (ES) categories, evaluated in terms of total standard output (SO) expressed in Euros. As mentioned before, in France, RICA focuses on ‘medium and large’ farms, those whose SO is greater than or equal to 25,000 Euros; this corresponds to ES category 6 and above. The nine ES categories available in RICA are aggregated into three categories: strictly less than 100,000 Euros of SO (ES6); from 100,000 to less than 250,000 Euros of SO (ES7); 250,000 Euros of SO and more (ES8 to ES14). It should be noted that, according to the EU regulation (CE) N°1242/2008, the two last categories correspond to ‘large farms’. Here, these farms are separated in two classes so that, in the following, they are referred to as, respectively, ‘large’ and ‘very large’ farms. RICA being a rotating panel, farms which either enter or leave the sample in a given year cannot be considered as actual entries into nor exits from the agricultural sector. Therefore, constant population is assumed and only transitions between size categories as defined above will be investigated.

For estimation purposes, the sample is restricted to farms which were present in the database for at least two consecutive years, so as to observe at least one transition. The corresponding unbalanced panel then comprised 13,786 farms out of the 16,217 farms in the original database (85.01%), leading to 92,626 (farm×year) observations and 78,840 individual 1-year transitions (including staying in the same category) from 2000 to 2013. Table 3.1 shows that farms are more likely to remain in their initial category of sizes two consecutive years. More than 90% of the farms remain in their initial size category whatever the category considered. Finally, it should be noted that remaining in the initial category does not mean that farms do not increase or decrease their size, but that the change is not sufficient to end into a different category as they were defined in this study.

TABLE 3.1: Observed 1-year transitions across size categories, 2000-2013.

		ES class			Total transitions
		25-100	100-250	≥ 250	
ES class	25-100	27,121	1,335	31	28,487
	100-250	1,280	32,822	1,089	35,191
	≥ 250	39	872	14,251	15,162

Note: Economic Size (ES) in 1,000 Euros of standard output (SO).

Source: Agreste, RICA France 2000-2013 – author’s calculations

To insure convergence of the EM algorithm especially when a high number of homogeneous farm type is considered, it is further assumed that farms do not move from medium to very large size and vice versa. because only a few movements between these two size extreme categories are observed (see Table 3.1). This is a common procedure in the agricultural economics when using Markov chain model to study structural change in farming (see [Ben Arfa *et al.* \(2015\)](#) for a recent example).

3.4.2 Explanatory variables

The literature provides a large number of factors that may affect farm structural change. These factors have been divided into several categories, but it is generally supposed that farm structural change may be mainly influenced by population and economic growth, technology improvement, the market structure for inputs as well as for outputs, human capital and some agricultural policies or public programmes (see [Goddard *et al.* \(1993\)](#); [Harrington and Reinsel \(1995\)](#); [Chavas \(2001\)](#); [Eastwood *et al.* \(2010\)](#) for detailed discussions). Given the multiplicity of these factors and the potential interactions existing between some of them, empirical studies generally focus on determining the impact of specific drivers of structural change. This study aims at investigating the impacts of agricultural policies on farm structural change in France. Subsidies received by farms are thus used to explain transition probabilities. Considering all farms together, this study analyses the impacts of public support programmes originating from the Common Agricultural Policy (CAP) of the European Union (EU).

Theoretically, no consensus has been found on the effects of such public programmes on structural change in farming ([Zimmermann and Heckelei, 2012](#)). The impact of such public support may depend on policy objectives and their implementation scheme ([Goddard *et al.*, 1993](#)). Public farm supports from the CAP are divided into two main components called ‘pillars’. While Pillar One aims at improving farmers’ competitiveness through, in particular, income support, Pillar Two is rather dedicated to promote rural development with broader environmental goals ([Gay *et al.*, 2005](#); [Jaime *et al.*, 2016](#)). Despite some reforms aimed at promoting more sustainable agriculture, Pillar One has remained the dominant part of the CAP, accounting for 89% of the CAP budget in 2003 ([Gay *et al.*, 2005](#)). Still targeting better performance of public supports, the 2013 reform of the CAP introduced some new regulations such as allowing for the transfer up to 15% from Pillar One to Pillar Two ([European Union, 2013](#)). In the following, we analyse the impacts of farm supports from the CAP focusing on some specific types of subsidies from both pillars.

For Pillar One, direct payments to farms, which are the most important element of EU’s agricultural policy, are considered. Direct payments are divided into coupled subsidies and decoupled subsidies (single farm payments) and account for 72% of CAP expenditure in the

2012 budget (Matthews, 2011). Both kinds of farm subsidies are supposed to encourage farm growth since they may relax liquidity and credit constraints (Roberts and Key, 2008; O’Toole and Hennessy, 2015). Despite the fact that single farm payments were introduced to reduce the distorting impacts of public supports, it has been proved that this kind of farm support may have *ex-post* coupled effects (Latruffe *et al.*, 2010). Single farm payments may influence farmers’ decisions through various channels. In particular, the authors argued that decoupled subsidies influence farmers’ decisions through expectations about future payments since such subsidies are often based on historical behaviour (Bhaskar and Beghin, 2009). Focusing on direct payments gathering coupled and decoupled farm subsidies, Pillar One is thus expected to positively affect farm growth.

Considering Pillar Two, we gathered compensatory allowances for less favoured areas (LFA) and for agri-environmental measures (AEM), and subsidies dedicated to farm investment. While supports for farm investment and agricultural infrastructure can help shift farming practices away from more damaging or wasteful forms of production and towards greater resource efficiency, agri-environmental measures are supposed to facilitate change in farming systems towards more resilient production types, better able to cope with future climate-related stress (Dwyer, 2013). These types of support programmes are more likely to inhibit farm growth. Focusing on these three types of farm support, Pillar Two is expected to have a positive impact on the probability of farms to remain in the same size category over time.

Several factors other than the agricultural policy have been shown to affect structural change in farming (see Goddard *et al.* (1993); Zimmermann *et al.* (2009)). In this study, only factors related to farm path dependency and the economic environment of farms are considered. The reason for only considering these factors is mainly because of data limitations. Indeed, proxies for other factors that may play an important role on farm size change, such as market condition, technical change, etc., are not available at the individual farm level in the RICA database. Nevertheless, as factors that affect transition process of farms may relate to each other (Goddard *et al.*, 1993; Harrington and Reinsel, 1995), proxies used may also control for the impacts of such other causal factors of farm size change.

Following Zimmermann and Heckelei (2012), the total initial stock (products of the farm and supplies in store) is used as a proxy of path dependency. Initial stocks is supposed to negatively affect farm size decline since high initial stock are assumed to result from former investment. Gross Operating Surplus (GOS) minus the total amount of subsidies received and the debt rate of farms are used to reflect the economic environment of farms. The GOS divided by the total number of non-salaried Annual Working Units (AWU) measures the financial capacity of a farm and as such is a very important indicator to obtain credit from a bank for new investments. It could also relate to the self-financing capacity of farms. A positive effect of the GOS is thus expected on the probability for a farm to grow. Debt rate is also expected to have

a positive impact on farm growth since credit generally gives to firms the resources necessary to expand (Goddard *et al.*, 1993).

Finally, some farm and farmer characteristics such as age, managerial capacity of farmers, legal status, localisation, type of farming are also used as control variables in the specification of transition probabilities. Such farm and farmer characteristics may allow controlling for observed heterogeneity and are introduced in the model specification using dummy variables. Table 3.2 presents the description and summary statistics for all the chosen explanatory variables.

Because of potential sources of unobserved heterogeneity as discussed in Section 3.1, we further expect that impacts of explanatory variables, especially farm subsidies, may differ from type to type. The results are presented and discussed in the following section.

TABLE 3.2: Descriptive statistics and definition of explanatory variables (n=92,626).

Variable	Unit	Mean	Std. Dev.	Min.	Max.
<i>pillar1</i> Subsidies from Pillar One	1,000 Euros	29.66	29.00	0.00	315.40
<i>pillar2</i> Subsidies from Pillar Two	1,000 Euros	12.05	21.85	0.00	443.98
<i>in_stock</i> Total initial stock	1,000 Euros	99.07	166.28	0.00	3,624.47
<i>GOS/ns-awu</i> GOS per non-salaried AWU	1,000 Euros	27.43	45.90	-1,052.02	1,279.47
<i>debt_rate</i> Total debt/liabilities	ratio	39.89	28.62	0.00	578.80
<i>age_55</i> 1 if farmer is over 55 years old ^a	dummy	0.19	0.39	0.00	1.00
<i>crops</i> 1 if farm is specialised in crops	dummy	0.45	0.50	0.00	1.00
<i>individual</i> 1 if farm has individual legal status	dummy	0.51	0.50	0.00	1.00
<i>LFA</i> 1 if farm is located in LFA	dummy	0.40	0.49	0.00	1.00
<i>agri_skills</i> 1 if farmer has agricultural education	dummy	0.93	0.25	0.00	1.00

Less Favoured Area (LFA); Gross Output Surplus (GOS)

^aIn order to take into account farmers' anticipation, 55 years old is used instead of 60 which is the minimum age for retirement in France.

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

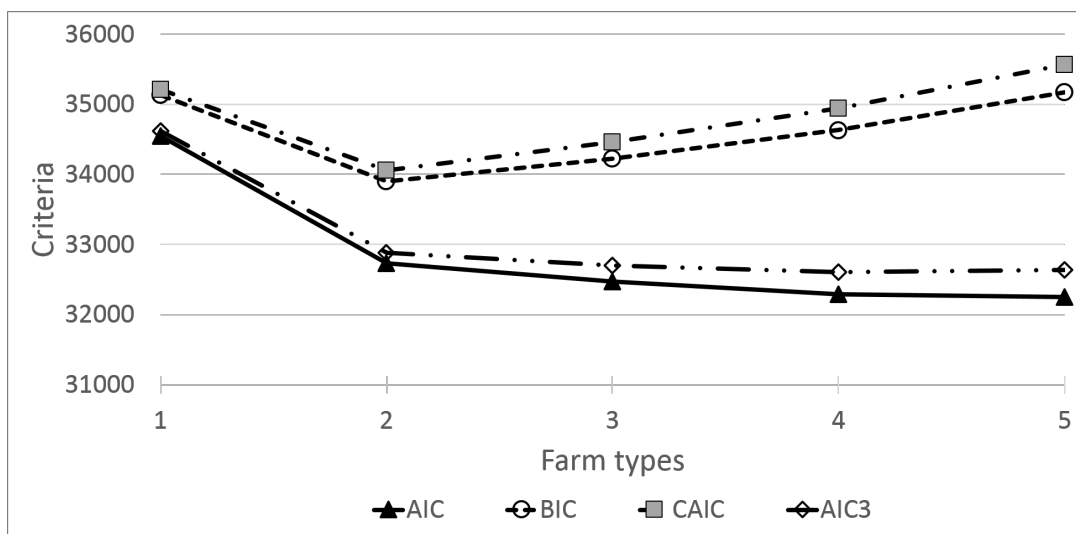
3.5 Results

3.5.1 Type membership and transition probability matrices

According to the information criteria presented in Section 3.3, a mixture of two types of farms seems to be the most appropriate data generating process. Two types are chosen as the optimal number specifically for two reasons: first, the BIC and CAIC criteria indicate that two types is the optimal number of farm types; second, even if the AIC and AIC3 criteria indicate a higher optimal number of farm types, the results show that the improvement in these criteria is relatively small when specifying more than two types of farms (see Figure 3.1). This means that increasing the total number of homogeneous types in the population increases much more the total number of parameters to be estimated than the representativeness of the data generating process. For the sake of parsimony, a mixture of two types of farms is thus preferable to represent farm size dynamics in France.

Table 3.3 reports the estimated type shares and the resulting transition probability matrices (TPMs) for the two farm types. As expected, the resulting TPMs are quite different from each other. The average posterior probabilities of belonging to a specific type indicate that about 68% of the sample consists of farms which tend to remain in their initial size category almost indefinitely, that is, at least during the entire period of observation. We therefore shall name this type the ‘almost stayers’. The probabilities to remain in the same category of sizes two consecutive years for those farms are close or over 0.99 (see Table 3.3.a). This result means that these farms have about 99% of chance to remain in the same category during a long time

FIGURE 3.1: Comparison of model-fit statistics for different numbers of types.



Source: Agreste, RICA France 2000-2013 – author’s calculations

TABLE 3.3: Estimated farm type shares and 1-year transition probability matrices (TPMs).

		Shares	ES class		
		68.75%	25-100	100-250	≥ 250
ES class	25-100	40.68%	0.990(0.019)	0.010(0.019)	-
	100-250	43.39%	0.07(0.017)	0.986(0.018)	0.007(0.007)
	≥ 250	15.93%	-	0.011(0.019)	0.989(0.019)

a) ‘Almost stayers’ TPM

		Shares	ES class		
		31.25%	25-100	100-250	≥ 250
ES class	25-100	39.03%	0.872(0.090)	0.128(0.090)	-
	100-250	44.78%	0.083(0.076)	0.822(0.071)	0.081(0.070)
	≥ 250	16.19%	-	0.159(0.097)	0.841(0.097)

b) ‘Likely movers’ TPM

Note: Economic size (ES) in 1,000 Euros of standard output (SO); standard deviations in parenthesis; transition between extreme size categories (25-100 and ≥ 250) are constrained to be zero.

Source: Agreste, RICA France 2000-2013 – author’s calculations

period. Conversely, farms belonging to the second type, about 32% of the sample, are more likely to change category two consecutive years than farms in the first type. We shall then name this type the ‘likely movers’. The transition probability matrix for this farm type is also strongly diagonal meaning that even farms which are likely to change category also have a high probability to remain in their initial category two consecutive years (Table 3.3.b). However, the probability to remain in the same category for these farms is around 0.85 meaning that they are about 15% more likely to change their size categories than those in the ‘almost stayer’ type.

Summary statistics for various farm and farmer characteristics for both the ‘almost stayers’ and ‘likely movers’ farm types were computed in order to identify the profile of farms in each type. Table 3.4 shows that the probability of belonging to a specific type does not correlate to the farm and farmer characteristics considered in the model specification. There is no significant difference in the distribution of these observed characteristics between the two types of farms. This result means that the unobserved heterogeneity cannot be sufficiently controlled by the observed farms and/or farmer characteristics considered and accounting for both kinds of heterogeneity may therefore lead to more efficiently estimate the impacts of explanatory variables, including agricultural policies.

The transition probability matrix for the overall population of farms can be easily derived by summing the two types of TPMs weighted by their respective shares in the population. The resulting 1-year TPM for the overall population is reported in Table 3.5. The transition

TABLE 3.4: Descriptive statistics for observed farm characteristics by type membership.

Variable	‘Almost stayers’		‘Likely movers’	
<i>age_55</i>	0.190	(0.392)	0.189	(0.391)
<i>crops</i>	0.438	(0.496)	0.433	(0.495)
<i>individual</i>	0.534	(0.499)	0.517	(0.500)
<i>LFA</i>	0.407	(0.491)	0.408	(0.491)
<i>agri_skills</i>	0.932	(0.252)	0.937	(0.244)

Note: Less Favoured Area (LFA); standard deviations in parentheses.

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

TABLE 3.5: Overall population 1-year transition probability matrix (TPM).

		ES class		
		25-100	100-250	≥ 250
ES class	25-100	0.954	0.046	-
	100-250	0.035	0.934	0.031
	≥ 250	-	0.058	0.942

Note: Economic size (ES) in 1,000 Euros of standard output (SO); transition between extreme size categories (25-100 and ≥ 250) are constrained to be zero.

Source: Agreste, RICA France 2000-2013 – author’s calculations

probabilities show that, overall, farms are more likely to remain in their initial category two consecutive years which is a common feature in agricultural economics (see [Hallberg \(1969\)](#); [Stokes \(2006\)](#) for examples). Indeed, the overall population TPM is strongly diagonal. The probability to remain in the same category of size two consecutive years varies from 0.93 to 0.95. This high probability to remain in the initial category is due to the high proportion of the ‘almost stayer’ type in the population. As a consequence, considering an homogeneous population to describe farm size dynamics as well as to investigate the impact of some explanatory variables on transition probabilities may be not sufficiently informative. Analysis under a mixture approach may be more informative by separating the impacts of explanatory variables, including agricultural policies, on different farm patterns.

3.5.2 Impact of explanatory variables

We estimated both the mixture model and an homogeneous model, *i.e.*, without accounting for unobserved farm heterogeneity. The results confirm our expectation that the impacts of most explanatory variables depend on the type of farms considered: even if the same sign is most often observed, the coefficient values are generally significantly different. As the values of the odds-ratios coefficients are difficult to interpret, we present here only transition probability

elasticities for the main explanatory variables.⁷ The estimated coefficients for the odds ratios of transition probabilities for both models as well as z-score tests for the comparison of the estimated parameters are reported in appendix. The values of the statistic lead to a rejection of the null hypothesis with a p-value of 5% in most cases.

Tables 3.6 and 3.7 respectively report the probability elasticities for farms to grow or decline and to remain in the same size category two consecutive years for both the homogeneous model and the mixture model. The first two columns of the tables present the results obtained using the homogeneous model. The four following columns present the results for the mixture model considering respectively the ‘almost stayers’ type and the ‘likely movers’ type. The results confirm our expectation for almost all explanatory variables. The results from both models show that, on overall, subsidies from the two pillars of the CAP, initial stocks, total GOS and debt rate of farms have a positive impact on the probability to grow and a negative impact on the probability to shrink.⁸ These results are also consistent with the literature. Indeed, several studies showed that farm subsidies from support programmes of the CAP are likely to favour farm growth (see [Ben Arfa *et al.* \(2015\)](#) for a recent example). [Zimmermann and Heckelei \(2012\)](#) found that initial stock positively affects farm growth. Contrary to the latter authors, the present results also support that initial stocks decrease the probability of farms to decline.

The results from both models also show that, overall, farms are more likely to decrease size when farmers are over 55 years old. Individual legal status as well as specialisation in crop productions have the same impacts on farm size. These results can be explained in three ways. Firstly, older farmers may be less motivated to increase their capacity of production when they are close to retirement because they may be more interested by the farm succession ([Potter and Lobley, 1992](#)) and younger farmers may be more likely to seek to increase agricultural activity, as they would be less financially secure than their older counterparts ([Howley *et al.*, 2014](#)). Secondly, individual legal status farms may face more economic constraints (capital, access to credit, etc.) than corporate farms, which may hamper new investments, or such farms may just have less financial motives than corporate farms ([Boehlje, 1992](#)). Thirdly, in France, it may be more difficult for farms specialised in crop productions to increase their economic size because of the regulation of the land market, since increasing the production capacity for crop farms would generally mean increasing the operated area contrary to livestock breeding farms which may increase the number of animal raised. Conversely, the probability to grow increases if most of the farm is located in an area without natural constraints or if the farmer has a minimum

⁷Here, an odds ratio is the ratio of the probability that a farm chooses a specific size category over the reference choice of remaining in the same category two consecutive years (see Section 3.2.2).

⁸Here, ‘grow’ and ‘shrink’ mean that farms move from one size category to another one. Nevertheless, farms remaining in the same category two consecutive years does not mean that they do not increase or decrease operated size at all but, at least, such size changes are not sufficient to move to another size category as defined in this study.

TABLE 3.6: Yearly probability elasticities of farm growth and decline both for homogeneous and mixed Markov chain models.

Probability	Variables	Homogeneous		‘Almost stayers’		‘Likely movers’	
Medium/Large p(12)	<i>pillar1</i>	0.514***	(0.042)	0.841***	(0.066)	0.344***	(0.038)
	<i>pillar2</i>	0.068**	(0.022)	0.135***	(0.033)	0.044*	(0.019)
	<i>in_stock</i>	0.143***	(0.028)	0.178***	(0.021)	0.188***	(0.034)
	<i>GOS/ns-awu</i>	0.172***	(0.017)	0.183***	(0.026)	0.147***	(0.019)
	<i>debt_rate</i>	0.356***	(0.031)	0.512***	(0.046)	0.340***	(0.027)
	<i>age_55</i>	-0.033*	(0.016)	-0.126***	(0.031)	-0.025	(0.015)
	<i>crops</i>	-0.169***	(0.032)	-0.156**	(0.051)	-0.160***	(0.028)
	<i>individual</i>	-0.470***	(0.058)	-0.586***	(0.091)	-0.421***	(0.049)
	<i>LFA</i>	-0.278***	(0.044)	-0.406***	(0.074)	-0.244***	(0.037)
	<i>agri_skills</i>	0.217	(0.127)	0.456*	(0.228)	0.199	(0.107)
Large/V. large p(23)	<i>pillar1</i>	0.290***	(0.052)	-0.792***	(0.084)	0.475***	(0.044)
	<i>pillar2</i>	0.089***	(0.021)	0.139***	(0.032)	0.032	(0.021)
	<i>in_stock</i>	0.259***	(0.032)	0.069	(0.047)	0.313***	(0.032)
	<i>GOS/ns-awu</i>	0.268***	(0.028)	0.123**	(0.040)	0.343***	(0.028)
	<i>debt_rate</i>	0.611***	(0.054)	0.371***	(0.082)	0.657***	(0.053)
	<i>age_55</i>	-0.017	(0.016)	-0.003	(0.027)	-0.024	(0.017)
	<i>crops</i>	-0.265***	(0.041)	0.377***	(0.058)	-0.450***	(0.040)
	<i>individual</i>	-0.172***	(0.034)	-0.086	(0.051)	-0.249***	(0.035)
	<i>LFA</i>	-0.095**	(0.030)	-0.162**	(0.054)	-0.084**	(0.029)
	<i>agri_skills</i>	0.275	(0.171)	0.319	(0.266)	0.304	(0.161)
Large/Medium p(21)	<i>pillar1</i>	-0.648***	(0.051)	-0.546***	(0.110)	-0.570***	(0.044)
	<i>pillar2</i>	-0.161***	(0.030)	-0.770***	(0.092)	-0.134***	(0.027)
	<i>in_stock</i>	-0.681***	(0.078)	-2.618***	(0.211)	-0.523***	(0.063)
	<i>GOS/ns-awu</i>	-0.498***	(0.035)	-0.804***	(0.072)	-0.496***	(0.037)
	<i>debt_rate</i>	-0.236***	(0.057)	0.209*	(0.102)	-0.368***	(0.053)
	<i>age_55</i>	0.049***	(0.013)	0.23***	(0.023)	0.020	(0.013)
	<i>crops</i>	-0.025	(0.034)	-0.577***	(0.069)	0.066*	(0.030)
	<i>individual</i>	0.253***	(0.029)	0.226***	(0.051)	0.231***	(0.024)
	<i>LFA</i>	0.188***	(0.024)	0.083	(0.046)	0.192***	(0.021)
	<i>agri_skills</i>	-0.056	(0.137)	0.882**	(0.277)	-0.329**	(0.122)
V. Large/Large p(32)	<i>pillar1</i>	-0.069	(0.049)	-0.014	(0.081)	-0.307***	(0.053)
	<i>pillar2</i>	-0.087*	(0.036)	-0.109	(0.064)	-0.043	(0.033)
	<i>in_stock</i>	-0.737***	(0.093)	-0.385**	(0.127)	-0.752***	(0.077)
	<i>GOS/ns-awu</i>	-0.335***	(0.052)	-0.421***	(0.080)	-0.320***	(0.045)
	<i>debt_rate</i>	-0.32***	(0.080)	-0.691***	(0.143)	-0.354***	(0.070)
	<i>age_55</i>	0.014	(0.017)	0.148***	(0.029)	-0.001	(0.016)
	<i>crops</i>	-0.100*	(0.040)	-0.048	(0.072)	0.017	(0.030)
	<i>individual</i>	0.073***	(0.021)	0.303***	(0.034)	0.023	(0.017)
	<i>LFA</i>	0.065**	(0.020)	-0.142**	(0.044)	0.052**	(0.019)
	<i>agri_skills</i>	0.137	(0.172)	-0.297	(0.247)	-0.006	(0.142)

Note: Less Favoured Area (LFA); Gross Output Surplus per non-salaried AWU (GOS/ns-awu); standard errors in parentheses; ***, ** and * indicate significance at 0.1%, 1% and 5%, respectively.

Source: Agreste, RICA France 2000-2013 – author’s calculations

TABLE 3.7: Yearly probability elasticities of farms to remain in the same category of sizes both for homogeneous and mixed Markov chain models.

Probability	Variables	Homogeneous		‘Almost stayers’		‘Likely movers’	
Medium/Medium p(11)	<i>pillar1</i>	-0.034***	(0.004)	-0.013***	(0.002)	-0.069***	(0.009)
	<i>pillar2</i>	-0.004**	(0.001)	-0.002**	(0.001)	-0.008*	(0.004)
	<i>in_stock</i>	-0.010***	(0.002)	-0.003***	(0.000)	-0.040***	(0.009)
	<i>GOS/ns-awu</i>	-0.013***	(0.002)	-0.003***	(0.001)	-0.034***	(0.006)
	<i>debt_rate</i>	-0.023***	(0.003)	-0.007***	(0.001)	-0.070***	(0.007)
	<i>age_55</i>	0.001*	(0.000)	0.001***	(0.000)	0.003	(0.002)
	<i>crops</i>	0.008***	(0.001)	0.002***	(0.000)	0.024***	(0.003)
	<i>individual</i>	0.018***	(0.002)	0.004***	(0.000)	0.052***	(0.005)
	<i>LFA</i>	0.012***	(0.002)	0.003***	(0.000)	0.033***	(0.004)
	<i>agri_skills</i>	-0.011	(0.006)	-0.005	(0.002)	-0.032	(0.018)
Large/Large p(22)	<i>pillar1</i>	0.007**	(0.002)	0.005***	(0.000)	-0.009	(0.007)
	<i>pillar2</i>	0.001	(0.001)	0.000	(0.000)	0.009**	(0.003)
	<i>in_stock</i>	0.006**	(0.002)	0.006***	(0.001)	0.003	(0.006)
	<i>GOS/ns-awu</i>	0.000	(0.002)	0.000	(0.001)	-0.005	(0.004)
	<i>debt_rate</i>	-0.015***	(0.003)	-0.004***	(0.001)	-0.036***	(0.008)
	<i>age_55</i>	-0.002*	(0.001)	-0.003***	(0.001)	-0.001	(0.002)
	<i>crops</i>	0.008***	(0.001)	-0.002	(0.001)	0.024***	(0.003)
	<i>individual</i>	-0.011***	(0.002)	-0.002*	(0.001)	-0.027***	(0.005)
	<i>LFA</i>	-0.007***	(0.002)	0.000	(0.000)	-0.026***	(0.005)
	<i>agri_skills</i>	-0.007	(0.008)	-0.008**	(0.003)	0.010	(0.018)
V.Large/V.Large p(33)	<i>pillar1</i>	0.004	(0.003)	0.000	(0.001)	0.049***	(0.007)
	<i>pillar2</i>	0.005**	(0.002)	0.001*	(0.000)	0.007	(0.005)
	<i>in_stock</i>	0.030***	(0.003)	0.003***	(0.001)	0.085***	(0.007)
	<i>GOS/ns-awu</i>	0.015***	(0.002)	0.003***	(0.000)	0.042***	(0.005)
	<i>debt_rate</i>	0.021***	(0.005)	0.007***	(0.001)	0.065***	(0.012)
	<i>age_55</i>	-0.001	(0.001)	-0.003**	(0.001)	0.000	(0.003)
	<i>crops</i>	0.005**	(0.002)	0.001	(0.001)	-0.003	(0.005)
	<i>individual</i>	-0.007**	(0.002)	-0.008***	(0.002)	-0.006	(0.004)
	<i>LFA</i>	-0.006**	(0.002)	0.001***	(0.000)	-0.012*	(0.005)
	<i>agri_skills</i>	-0.009	(0.012)	0.003	(0.003)	0.001	(0.027)

Note: Less Favoured Area (LFA); Gross Output Surplus per non-salaried AWU (GOS/ns-awu); standard errors in parentheses; ***, ** and * indicate significance at 0.1%, 1% and 5%, respectively.

Source: Agreste, RICA France 2000-2013 – author’s calculations

agricultural education. The latter result confirms the positive impact of the managerial capacity on farm size growth (Boehlje, 1992; Goddard *et al.*, 1993).

As mentioned above, the resulting elasticities are different considering the homogeneous model and a mixture of ‘almost stayers’ and ‘likely movers’ farms. The impacts of explanatory variables are larger on transition probabilities are also different for ‘likely movers’ farms and ‘almost stayers’. Overall, the impacts of farm subsidies from both pillar are in most cases higher for ‘almost stayers’ than for ‘likely movers’. In particular, the impact of subsidies two pillars of the CAP on probability of farms to grow is higher for farms belonging to the ‘almost stayer’

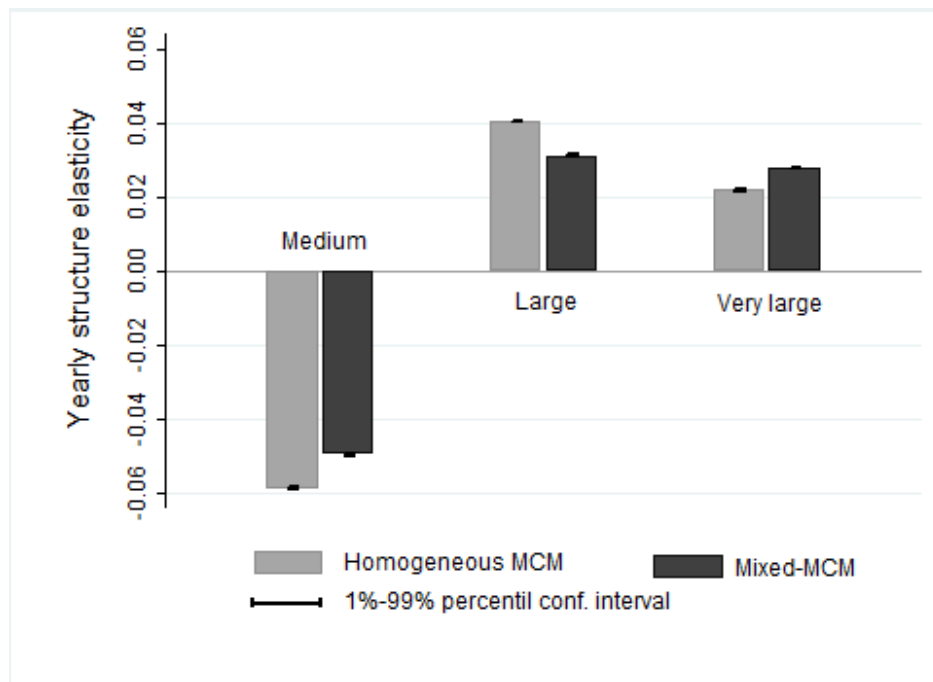
type than for farms in the ‘likely movers’ one. For example, a 1% increase in the amount of Pillar One subsidies is expected to increase the probability of farms to move from the medium to the large size by about 0.84% for the ‘almost stayers’ type but only 0.34% for the ‘likely movers’ type. The impact of coupled and decoupled subsidies is thus more than 2 times higher for ‘almost stayers’ than for ‘likely movers’. This result may be explained by the fact that the ‘almost stayers’ type may gather farms with liquidity constraints and support from the CAP may therefore have an income multiplier effect for such farms (Latruffe *et al.*, 2010).

The results from both the homogeneous model and the mixture of ‘almost stayers’ and ‘likely movers’ also show that Pillar One has higher effects on farm size change than Pillar Two supports. Considering the ‘almost stayers’ type of farms, the impact of Pillar One is about 6 times higher than that of Pillar Two. Furthermore, subsidies from Pillar Two are more likely to increase farm probability to grow while Pillar Two supports are more likely to decrease the probability to shrink. Such subsidies generally decrease the probability to decline. This result confirms the expectation and previous findings of the literature that subsidies from Pillar One of the CAP are more likely to encourage farm growth than subsidies from Pillar Two (Piet *et al.*, 2012).

Structure elasticities were then computed at the mean point for farm subsidies from both pillars of the CAP in order to assess their impact on the distribution of farm size. As entries and exits could not be considered because of data limitations, the structure elasticities were computed on farm proportions by size category. The structure elasticities thus represent the variation (in percent) of the share of farms in a specific category that would occur if the total amount of subsidies received by farmers were to increase by 1%. Figures 3.2 and 3.3 present the resulting structure elasticities for both the homogeneous MCM and the mixed-MCM.

Figures 3.2 and 3.3 show that the resulting structure elasticities are quite different from one model to the other. Overall, the homogeneous model tends to overestimate the impacts of subsidies from both pillars. For example, the homogeneous model predicts that a 1% increase of Pillar One subsidies will decrease the proportion of medium sized farms by about 0.06% while the mixture model predicts a decrease by about 0.05% (see Figure 3.2). Likewise, the homogeneous model predicts a 0.01% increase in the proportion of very large farm if Pillar Two subsidies were to increase by 1% while this proportion is about 0.005% for the mixture model (see Figure 3.3). These results confirm that ignoring unobserved heterogeneity in modelling farm size dynamics can lead to misleading results. The results also confirm that the impact of subsidies from Pillar One of the CAP on the variation of farm proportion in category of sizes is higher than the impact of supports from Pillar Two. The resulting structure elasticities are about 6 times higher for Pillar One than for Pillar Two whatever the size category considered.

FIGURE 3.2: Yearly structure elasticities of farm subsidies from Pillar One of the CAP both for homogeneous and mixed Markov chain models.



Note: ‘Medium’ sized farms are less than 100,000 Euros of SO; ‘Large’ farms are 100,000-250,000 Euros of SO; ‘Very large’ farms are 250,000 Euros of SO and more.

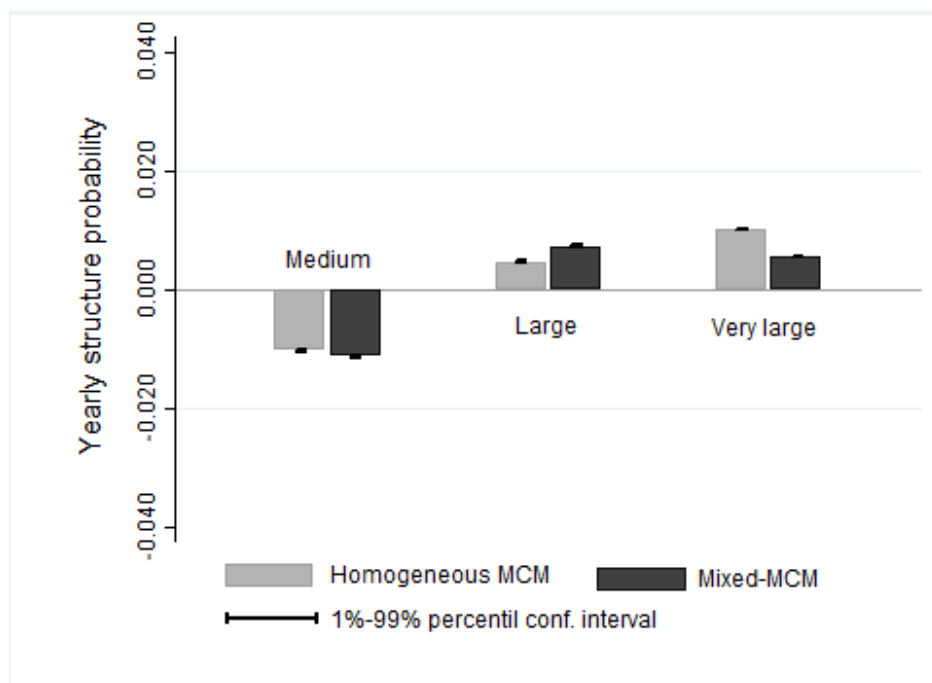
Source: Agreste, RICA France 2000-2013 – author’s calculations

3.6 Concluding remarks

This chapter aims at investigating the impacts of a selected set of causal factors likely to influence structural change in farming. From a methodological point of view, it extends the findings of the chapter of the thesis by applying a non-stationary mixed-Markov chain model (M-MCM) for the first time in agricultural economics. Using individual level data, a mixture of homogeneous farm types was considered in order to incorporate unobserved heterogeneity in the transition process of farms. A discrete choice modelling approach was used to describe farm size category choices of farmers. A multinomial logit specification for the transition probabilities was implemented and the expectation-maximisation (EM) algorithm allowed estimating the parameters of the mixed Markov chain model. Transition probability elasticities as well as structure elasticities were derived to analyse the impacts of a set of exogenous variables on farm size change in the French farming sector, especially focusing on the impact of specific measures of the Common Agricultural Policy (CAP) of the European Union (UE).

Using a sample of farms from the ‘Réseau d’Information Comptable Agricole’ (RICA) from 2000 to 2013, the results show that French farms can be divided into two unobserved types:

FIGURE 3.3: Yearly structure elasticities of farm subsidies from Pillar Two of the CAP both for homogeneous and mixed Markov chain models.



Note: ‘Medium’ sized farms are less than 100,000 Euros of SO; ‘Large’ farms are 100,000-250,000 Euros of SO; ‘Very large’ farms are 250,000 Euros of SO and more.

Source: Agreste, RICA France 2000-2013 – author’s calculations

‘almost stayers’ which are more likely to remain almost indefinitely in their initial category of sizes, and ‘likely movers’ which change size category more frequently. The results also show that the French farm population consists of a higher proportion of farms that behave like ‘almost stayers’ leading to a strongly diagonal transition probability matrix for the overall population. Descriptive statistics for both types of farms on a set of observed variables show that the probability to belong to the ‘almost stayer’ type or the ‘likely movers’ type are not significantly correlated with the chosen set of observed farm and/or farmer characteristics, meaning that unobserved heterogeneity cannot be fully controlled by observed heterogeneity.

The results also show that the impacts of farm subsidies from Pillar One and Pillar Two of the CAP depend on the type that farms belong to. Aggregated at the population level, structure elasticities show that the two models lead to different results, with the homogeneous model tending to either overestimate or underestimate the impacts of subsidies from both Pillars. These results are relevant for policy assessment since they confirm that ignoring potential unobserved heterogeneity in farmer behaviour may lead to inaccurate parameters and therefore to misleading policy recommendations.

Yet, this study has some limitations that may motivate further research. In the current

chapter, the estimations were performed without taking into account entries into nor exits from the farming sector because of the RICA data limitations at the individual level. Accounting for entries into and exits from farming would constitute an obvious way to analyse the impact of agricultural policy on the total number of farms by size category. The results could be also improved considering other explanatory variables than those used in this study since it has been proved that several other factors may play an important role on structural change in farming. Lastly, a possible improvement of this work is to introduce explanatory variables in the specification of farm types selection. This could lead to more insightful analysis of the impact of farm and/or farmer characteristics on the probability to belong to a specific farm type.

Bibliography

- Andrews, R. L. and Currim, I. S. (2003). Retention of latent segments in regression-based marketing models. *International Journal of Research in Marketing* 20: 315–321.
- Baudry, J.-P. and Celeux, G. (2015). Em for mixtures. *Statistics and Computing* 25: 713–726.
- Ben Arfa, N., Daniel, K., Jacquet, F. and Karantininis, K. (2015). Agricultural policies and structural change in French dairy farms: A nonstationary Markov model. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroéconomie* 63: 19–42.
- Bhaskar, A. and Beghin, J. C. (2009). How coupled are decoupled farm payments? a review of the evidence. *Journal of Agricultural and Resource Economics* : 130–153.
- Blumen, I., Kogan, M. and McCarthy, P. J. (1955). *The industrial mobility of labor as a probability process, VI*. Cornell Studies in Industrial and Labor Relations.
- Boehlje, M. (1992). Alternative models of structural change in agriculture and related industries. *Agribusiness* 8: 219–231.
- Bollman, R. D., Whitener, L. A. and Tung, F. L. (1995). Trends and patterns of agricultural structural change: A Canada–US comparison. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroéconomie* 43: 15–28.
- Bostwick, D. (1962). *Yield probabilities as a Markov process*. Marketing Economics Division, Economic Research Service, US Department of Agriculture.
- Bowman, M. S. and Zilberman, D. (2013). Economic factors affecting diversified farming systems. *Ecology and Society* 18: 33.
- Breustedt, G. and Glauben, T. (2007). Driving forces behind exiting from farming in western europe. *J. Agr. Econ.* 58: 115–127.
- Chavas, J.-P. (2001). *Chapter 5 Structural change in agricultural production: Economics, technology and policy*. Elsevier, *Volume 1, Part A*. 263–285.
- Chavas, J.-P. and Magand, G. (1988). A dynamic analysis of the size distribution of firms: The case of the us dairy industry. *Agribusiness* 4: 315–329.
- Cipollini, F., Ferretti, C. and Ganugi, P. (2012). Firm size dynamics in an industrial district: The mover-stayer model in action. In Di Ciaccio, A., Coli, M. and Angulo Ibanez, J. M. (eds), *Advanced Statistical Methods for the Analysis of Large Data-Sets*. Springer Berlin Heidelberg, 443–452.

- Conradt, S., Bokusheva, R., Finger, R., Kussaiynov, T. *et al.* (2014). Yield trend estimation in the presence of farm heterogeneity and non-linear technological change. *Quarterly Journal of International Agriculture* 53: 121–140.
- Daniels, T. L. (1986). Hobby farming in america: Rural development or threat to commercial agriculture? *Journal of Rural Studies* 2: 31 – 40.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39: 1–38.
- Dias, J. G. and Willekens, F. (2005). Model-based clustering of sequential data with an application to contraceptive use dynamics. *Mathematical population studies* 12: 135–157.
- Disney, W. T., Duffy, P. A. and Hardy, W. (1988). A Markov chain analysis of pork farm size distributions in the South. *Southern Journal of Agricultural Economics* 20: 57–64.
- Dwyer, J. (2013). Transformation for sustainable agriculture: what role for the second pillar of cap? *Bio-based and Applied Economics* 2: 29–47.
- Eastwood, R., Lipton, M. and Newell, A. (2010). Farm size. *Handbook of agricultural economics* 4: 3323–3397.
- Ethridge, D. E., Roy, S. K. and Myers, D. W. (1985). A markov chain analysis of structural changes in the Texas high plains cotton ginning industry. *Southern Journal of Agricultural Economics* 17: 11–20.
- European Union (2013). Overview of CAP Reform 2014-2020. Agricultural Policy Perspectives Brief 5.
- Freshwater, D. and Reimer, B. (1995). Socio-economic policies as causal forces for the structure of agriculture. *Canadian journal of agricultural economics/Revue canadienne d'agroeconomie* 43: 209–222.
- Frydman, H., Kallberg, J. G. and Kao, D.-L. (1985). Testing the adequacy of Markov chain and mover-stayer models as representations of credit behavior. *Operations Research* 33: 1203–1214.
- Gaffney, P. (1997). A projection of Irish agricultural structure using Markov chain analysis. Tech. rep., CAPRI Working Paper 97-10.
- Garcia, P., Offutt, S. E. and Sonka, S. T. (1987). Size distribution and growth in a sample of Illinois cash grain farms. *American Journal of Agricultural Economics* 69: 471–476.
- Garver, M. S., Williams, Z. and Taylor, G. S. (2008). Employing latent class regression analysis to examine logistics theory: An application of truck driver retention. *Journal of Business Logistics* 29: 233–257.

- Gay, S., Osterburg, B., Baldock, D. and Zdanowicz, A. (2005). Recent evolution of the eu common agricultural policy (cap): state of play and environmental potential. *MEACAP WP6, D4b, Federal Agricultural Research Centre, Braunschweig and Institute for European Environmental Policy, London* .
- Gillespie, J. M. and Fulton, J. R. (2001). A markov chain analysis of the size of hog production firms in the united states. *Agribusiness* 17: 557–570.
- Goddard, E., Weersink, A., Chen, K. and Turvey, C. G. (1993). Economics of structural change in agriculture. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 41: 475–489.
- Greene, W. H. (2006). *Econometric analysis*. Pearson Education India, 6th ed.
- Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological* 37: 681 – 698.
- Hallam, A. (1991). Economies of size and scale in agriculture: An interpretive review of empirical measurement. *Review of Agricultural Economics* 13: 155–172.
- Hallberg, M. C. (1969). Projecting the size distribution of agricultural firms. an application of a markov process with non-stationary transition probabilities. *American Journal of Agricultural Economics* 51: 289–302.
- Harrington, D. H. and Reinsel, R. D. (1995). A synthesis of forces driving structural change. *Canadian Journal of Agricultural Economics* 43: 3–14.
- Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52: 271–320.
- Hess, S., Bierlaire, M. and Polak, J. (2006). A systematic comparison of continuous and discrete mixture models. Tech. rep.
- Holloway, L. (2002). Smallholding, hobby-farming, and commercial farming: Ethical identities and the production of farming spaces. *Environment and Planning A* 34: 2055–2070.
- Howley, P. (2015). The Happy Farmer: The Effect of Nonpecuniary Benefits on Behavior. *American Journal of Agricultural Economics* 97: 1072 –1086.
- Howley, P., Dillon, E. and Hennessy, T. (2014). It s not all about the money: understanding farmers labor allocation choices. *Agriculture and Human Values* 31: 261–271.
- Huettel, S. and Jongeneel, R. (2011). How has the EU milk quota affected patterns of herd-size change? *European Review of Agricultural Economics* 38: 497–527.

- Jackson-Smith, D. B. (1999). Understanding the microdynamics of farm structural change: Entry, exit, and restructuring among wisconsin family farmers in the 1980s1. *Rural Sociology* 64: 66–91.
- Jaime, M. M., Coria, J. and Liu, X. (2016). Interactions between CAP agricultural and agri-environmental subsidies and their effects on the uptake of organic farming. *American Journal of Agricultural Economics* 98: 1114–1145.
- Judge, G. G. and Swanson, E. R. (1962). Markov chains: basic concepts and suggested uses in agricultural economics. *Australian Journal of Agricultural Economics* 6: 49–61.
- Karantininis, K. (2002). Information-based estimators for the non-stationary transition probability matrix: An application to the Danish pork industry. *Journal of Econometrics* 107: 275–290.
- Keane, M. (1991). Changes in the size structure of Irish dairy farms. *Irish Journal of Agricultural Economics and Rural Sociology* 14: 67–74.
- Krenz, R. D. (1964). Projection of farm numbers for North Dakota with Markov chains. *Agricultural Economics Research* 16: 77–83.
- Latruffe, L., Davidova, S., Douarin, E. and Gorton, M. (2010). Farm expansion in Lithuania after accession to the EU: The role of CAP payments in alleviating potential credit constraints. *Europe-Asia Studies* 62: 351–365.
- Lee, T., Judge, G. and Zellner, A. (1970). *Estimating the parameters of the Markov probability model from aggregate time series data*. Amsterdam: North Holland.
- Lee, T. C., Judge, G. G. and Takayama, T. (1965). On estimating the transition probabilities of a Markov process. *Journal of Farm Economics* 47: 742–762.
- Lindsay, B. G. and Lesperance, M. L. (1995). A review of semiparametric mixture models. *Journal of statistical planning and inference* 47: 29–39.
- MacRae, E. C. (1977). Estimation of time-varying Markov processes with aggregate data. *Econometrica* 45: 183–198.
- Major, K. (2008). Income disparities among Hungarian micro-regions: The mover-stayer model. *Acta Oeconomica* 58: 127–156.
- Matthews, A. (2011). Post-2013 eu common agricultural policy, trade and development. a review of legislative proposals. international centre for programme on agricultural trade and sustainable development (ictsd), issue paper no 39 .
- Maybery, D., Crase, L. and Gullifer, C. (2005). Categorising farming values as economic, conservation and lifestyle. *Journal of Economic Psychology* 26: 59 – 72.

- McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics* 3: 303 – 328.
- McInerney, N. and Garvey, E. (2004). Farm structure and agricultural labour. Tech. rep., CAPRI Working Papers. Dept of Economics, NUI Galway.
- McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*, 382. John Wiley & Sons.
- McLachlan, G. and Peel, D. (2001). *Finite mixture models*. John Wiley & Sons.
- Mzoughi, N. (2011). Farmers adoption of integrated crop protection and organic farming: Do moral and social concerns matter? *Ecological Economics* 70: 1536 – 1545.
- O’Toole, C. and Hennessy, T. (2015). Do decoupled payments affect investment financing constraints? evidence from Irish agriculture. *Food Policy* 56: 67–75.
- Pacifico, D. and Yoo, H. I. (2012). lcglogit: a stata module for estimating latent class conditional logit models via the expectation-maximization algorithm. *UNSW Australian School of Business Research Paper* .
- Padberg, D. I. (1962). The use of Markov processes in measuring changes in market structure. *J. Farm Econ.* 44: 189–199.
- Paternoster, R., Brame, R., Mazerolle, P. and Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology* 36: 859–866.
- Piet, L., Latruffe, L., Le Mouel, C. and Desjeux, Y. (2012). How do agricultural policies influence farm size inequality? the example of France. *European Review of Agricultural Economics* 39: 5–28.
- Potter, C. and Lobley, M. (1992). Ageing and succession on family farms: The impact on decision-making and land use. *Sociologia Ruralis* 32: 317–334.
- Rahelizatovo, N. C. and Gillespie, J. M. (1999). Dairy farm size, entry, and declining production region. *J. Agr. Appl. Econ.* 31: 333–347.
- Roberts, M. J. and Key, N. (2008). Agricultural payments and land concentration: a semiparametric spatial regression analysis. *American Journal of Agricultural Economics* 90: 627–643.
- Sarstedt, M. (2008). Market segmentation with mixture regression models: Understanding measures that guide model selection. *Journal of Targeting, Measurement and Analysis for Marketing* 16: 228–246.
- Stokes, J. R. (2006). Entry, exit, and structural change in pennsylvania’s dairy sector. *Agr. Resource Econ. Rev.* 35: 357.

- Storm, H., Heckelei, T. and Mittelhammer, R. C. (2016). Bayesian estimation of non-stationary Markov models combining micro and macro data. *European Review of Agricultural Economics* 43: 303–329.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- Trujillo-Barrera, A., Pennings, J. M. E. and Hofenk, D. (2016). Understanding producers' motives for adopting sustainable practices: the role of expected rewards, risk perception and risk tolerance. *European Review of Agricultural Economics* 43: 359–382.
- Vermunt, J. (2010). Longitudinal research using mixture models. In Montfort, K. van, Oud, J. H. and Satorra, A. (eds), *Longitudinal Research with Latent Variables*. Springer Berlin Heidelberg, 119–152.
- Wedel, M. (2002). Concomitant variables in finite mixture models. *Statistica Neerlandica* 56: 362–375.
- Weiss, C. R. (1999). Farm Growth and Survival: Econometric Evidence for Individual Farms in Upper Austria. *American Journal of Agricultural Economics* 81: 103–116.
- Willock, J., Deary, I. J., Edwards-Jones, G., Gibson, G. J., McGregor, M. J., Sutherland, A., Dent, J. B., Morgan, O. and Grieve, R. (1999). The role of attitudes and objectives in farmer decision making: Business and environmentally-oriented behaviour in scotland. *Journal of Agricultural Economics* 50: 286–303.
- Zepeda, L. (1995a). Asymmetry and nonstationarity in the farm size distribution of Wisconsin milk producers: An aggregate analysis. *American Journal of Agricultural Economics* 77: 837–852.
- Zepeda, L. (1995b). Technical change and the structure of production: a non-stationary markov analysis. *European Review of Agricultural Economics* 22: 41–60.
- Zimmermann, A. and Heckelei, T. (2012). Structural change of European dairy farms: A cross-regional analysis. *Journal of Agricultural Economics* 63: 576–603.
- Zimmermann, A., Heckelei, T. and Dominguez, I. P. (2009). Modelling farm structural change for integrated ex-ante assessment: Review of methods and determinants. *Environmental Science and Policy* 12: 601–618.

Appendix

TABLE 3.8: Estimated parameters of farm probabilities to grow between two consecutive years for the homogeneous and the mixed Markov chain models.

Odds ratios	Variables	Homogeneous		‘Almost stayers’		‘Likely movers’	
p12/p11	<i>intercept</i>	-3.872***	(0.184)	-6.352***	(0.330)	-2.602***	(0.172)
	<i>pillar1</i>	0.030***	(0.003)	0.047***	(0.004)	0.022***	(0.002)
	<i>pillar2</i>	0.008***	(0.003)	0.015***	(0.004)	0.006**	(0.002)
	<i>in_stock</i>	0.004***	(0.001)	0.005***	(0.001)	0.006***	(0.001)
	<i>GOS/ns-awu</i>	0.017***	(0.002)	0.017***	(0.003)	0.015***	(0.002)
	<i>debt_rate</i>	0.013***	(0.001)	0.017***	(0.002)	0.013***	(0.001)
	<i>age_55</i>	-0.200**	(0.094)	-0.738***	(0.180)	-0.161*	(0.094)
	<i>crops</i>	-0.454***	(0.085)	-0.410***	(0.135)	-0.452***	(0.077)
	<i>individual</i>	-0.615***	(0.076)	-0.734***	(0.114)	-0.614***	(0.070)
	<i>LFA</i>	-0.504***	(0.080)	-0.708***	(0.129)	-0.491***	(0.072)
	<i>agri_skills</i>	0.249*	(0.145)	0.503**	(0.252)	0.252*	(0.136)
p23/22	<i>intercept</i>	-4.916***	(0.221)	-5.668***	(0.345)	-4.050***	(0.226)
	<i>pillar1</i>	0.008***	(0.002)	-0.023***	(0.002)	0.015***	(0.002)
	<i>pillar2</i>	0.008***	(0.002)	0.012***	(0.003)	0.002	(0.002)
	<i>in_stock</i>	0.003***	(0.000)	0.001	(0.001)	0.004***	(0.000)
	<i>GOS/ns-awu</i>	0.010***	(0.001)	0.005***	(0.001)	0.013***	(0.001)
	<i>debt_rate</i>	0.015***	(0.001)	0.009***	(0.002)	0.016***	(0.001)
	<i>age_55</i>	-0.089	(0.097)	0.000	(0.155)	-0.134	(0.100)
	<i>crops</i>	-0.584***	(0.089)	0.804***	(0.125)	-1.025***	(0.089)
	<i>individual</i>	-0.387***	(0.082)	-0.207	(0.126)	-0.515***	(0.082)
	<i>LFA</i>	-0.253***	(0.088)	-0.479***	(0.160)	-0.164*	(0.084)
	<i>agri_skills</i>	0.298	(0.186)	0.346	(0.284)	0.310*	(0.183)

Note: Less Favoured Area (LFA); Gross Output Surplus per non-salaried AWU (GOS/ns-awu); standard errors in parentheses; ***, ** and * indicate significance at 0.1%, 1% and 5%, respectively.

Source: Agreste, RICA France 2000-2013 – author’s calculations

TABLE 3.9: Z-scores for testing the equality of the farm types' odds ratios coefficients for farm growth.

Odds ratios	Variables	'AS'-'HM'		'LM'-'HM'		'AS'-'LM'	
		score	p_value	score	p_value	score	p_value
p12/p11	<i>intercept</i>	3.871	0.000	-2.305	0.021	-5.699	0.000
	<i>pillar1</i>	1.597	0.110	-0.656	0.512	- 2.144	0.032
	<i>pillar2</i>	0.866	0.386	1.262	0.207	0.687	0.492
	<i>in_stock</i>	0.240	0.810	-0.602	0.547	- 0.725	0.468
	<i>GOS/ns-awu</i>	2.521	0.012	0.501	0.616	-2.096	0.036
	<i>debt_rate</i>	-2.648	0.008	0.288	0.773	2.839	0.005
	<i>age_55</i>	0.272	0.786	0.011	0.991	- 0.272	0.786
	<i>crops</i>	-0.873	0.383	0.008	0.994	0.901	0.368
	<i>individual</i>	-1.342	0.180	0.124	0.901	1.466	0.143
	<i>LFA</i>	0.872	0.383	0.015	0.988	-0.875	0.382
	<i>agri_skills</i>	-6.562	0.000	5.038	0.000	-10.079	0.000
p23/p22	<i>intercept</i>	-10.772	0.000	2.959	0.003	13.131	0.000
	<i>pillar1</i>	1.343	0.179	-2.159	0.031	-3.026	0.002
	<i>pillar2</i>	-3.302	0.001	1.296	0.195	4.272	0.000
	<i>in_stock</i>	-2.930	0.003	1.933	0.053	4.466	0.000
	<i>GOS/ns-awu</i>	-2.474	0.013	0.744	0.457	3.039	0.002
	<i>debt_rate</i>	0.487	0.626	-0.319	0.750	-0.724	0.469
	<i>age_55</i>	9.035	0.000	-3.504	0.000	-11.898	0.000
	<i>crops</i>	1.195	0.232	-1.094	0.274	-2.040	0.041
	<i>individual</i>	-1.235	0.217	0.739	0.46	1.745	0.081
	<i>LFA</i>	0.144	0.886	0.048	0.962	- 0.108	0.914
	<i>agri_skills</i>	-1.839	0.066	2.737	0.006	3.925	0.000

Note: 'Almost stayers' (AS), Homogeneous model (HM) and 'Likely movers' (LM); Least Favoured Area (LFA); Gross Output Surplus per non-salaried AWU (GOS/ns-awu); z-scores for testing the null hypothesis that coefficients of two different types are equal (Paternoster *et al.*, 1998).

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

TABLE 3.10: Estimated parameters of farm probabilities to decline between two consecutive years for the homogeneous and the mixed Markov chain models.

Odds ratios	Variables	Homogeneous		‘Almost stayers’		‘Likely movers’	
p21/p22	<i>intercept</i>	-1.957***	(0.195)	-3.593***	(0.373)	-0.633***	(0.184)
	<i>pillar1</i>	-0.019***	(0.001)	-0.016***	(0.003)	-0.017***	(0.001)
	<i>pillar2</i>	-0.014***	(0.003)	-0.068***	(0.008)	-0.012***	(0.002)
	<i>in_stock</i>	-0.008***	(0.001)	-0.029***	(0.002)	-0.006***	(0.001)
	<i>GOS/ns-awu</i>	-0.018***	(0.001)	-0.030***	(0.003)	-0.018***	(0.001)
	<i>debt_rate</i>	-0.005***	(0.001)	0.005**	(0.002)	-0.008***	(0.001)
	<i>age_55</i>	0.292***	(0.081)	1.348***	(0.136)	0.119	(0.086)
	<i>crops</i>	-0.070	(0.076)	-1.222***	(0.146)	0.091	(0.071)
	<i>individual</i>	0.640***	(0.074)	0.562***	(0.128)	0.598***	(0.068)
	<i>LFA</i>	0.567***	(0.074)	0.245*	(0.136)	0.612***	(0.070)
	<i>agri_skills</i>	-0.052	(0.150)	0.942***	(0.296)	-0.358**	(0.140)
	p32/p33	<i>intercept</i>	-1.552***	(0.240)	-3.296***	(0.351)	0.024
<i>pillar1</i>		-0.002	(0.001)	-0.000	(0.002)	-0.008***	(0.001)
<i>pillar2</i>		-0.006**	(0.002)	-0.007*	(0.004)	-0.003	(0.003)
<i>in_stock</i>		-0.004***	(0.001)	-0.002***	(0.001)	-0.005***	(0.000)
<i>GOS/ns-awu</i>		-0.007***	(0.001)	-0.008***	(0.002)	-0.008***	(0.001)
<i>debt_rate</i>		-0.007***	(0.002)	-0.014***	(0.003)	-0.008***	(0.002)
<i>age_55</i>		0.081	(0.102)	0.823***	(0.163)	-0.008	(0.108)
<i>crops</i>		-0.231**	(0.093)	-0.102	(0.153)	0.048	(0.083)
<i>individual</i>		0.342***	(0.098)	1.318***	(0.152)	0.125	(0.093)
<i>LFA</i>		0.323***	(0.104)	-0.709***	(0.221)	0.257***	(0.097)
<i>agri_skills</i>		0.154	(0.195)	-0.319	(0.266)	-0.007	(0.176)

Note: Less Favoured Area (LFA); Gross Output Surplus per non-salaried AWU (GOS/ns-awu); standard errors in parentheses; ***, ** and * indicate significance at 0.1%, 1% and 5%, respectively.

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

TABLE 3.11: Z-scores for testing the equality of the farm types' odds ratios coefficients for farm decline.

Odds ratios	Variables	'AS'-'HM'		'LM'-'HM'		'AS'-'LM'	
		score	p_value	score	p_value	score	p_value
p21/p22	<i>intercept</i>	0.952	0.341	0.978	0.328	-0.387	0.699
	<i>pillar1</i>	-6.326	0.000	0.600	0.549	6.646	0.000
	<i>pillar2</i>	-8.541	0.000	1.419	0.156	9.378	0.000
	<i>in_stock</i>	-3.768	0.000	0.015	0.988	3.715	0.000
	<i>GOS/ns-awu</i>	3.657	0.000	-1.337	0.181	-4.640	0.000
	<i>debt_rate</i>	6.655	0.000	-1.460	0.144	-7.630	0.000
	<i>age_55</i>	-7.008	0.000	1.559	0.119	8.104	0.000
	<i>crops</i>	-0.529	0.597	-0.418	0.676	0.251	0.802
	<i>individual</i>	-2.091	0.037	0.441	0.659	2.408	0.016
	<i>LFA</i>	3.000	0.003	-1.491	0.136	-3.975	0.000
	<i>agri_skills</i>	-3.885	0.000	4.940	0.000	7.116	0.000
p32/p33	<i>intercept</i>	0.577	0.564	-3.298	0.001	-2.987	0.003
	<i>pillar1</i>	-0.243	0.808	0.755	0.450	0.788	0.431
	<i>pillar2</i>	2.453	0.014	-0.802	0.423	-3.221	0.001
	<i>in_stock</i>	-0.642	0.521	-0.492	0.623	0.249	0.803
	<i>GOS/ns-awu</i>	-2.158	0.031	-0.603	0.547	1.766	0.077
	<i>debt_rate</i>	3.860	0.000	-0.597	0.551	-4.251	0.000
	<i>age_55</i>	0.719	0.472	2.242	0.025	0.863	0.388
	<i>crops</i>	5.380	0.000	-1.611	0.107	-6.693	0.000
	<i>individual</i>	-4.230	0.000	-0.464	0.643	4.008	0.000
	<i>LFA</i>	-1.438	0.150	-0.615	0.539	0.979	0.328
	<i>agri_skills</i>	-4.102	0.000	4.734	0.000	7.906	0.000

Note: 'Almost stayers' (AS), Homogeneous model (HM) and 'Likely movers' (LM); Least Favoured Area (LFA); Gross Output Suplus per non-salaried AWU (GOS/ns-awu); z-scores for testing the null hypothesis that coefficients of two different types are equal (Paternoster *et al.*, 1998).

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

Chapter 4

Heterogeneity and spatial interdependence in farm survival and growth: evidence from France¹

Abstract

Accounting for spatial interdependence is relevant for the analysis of structural change in farming and for the assessment of policy changes because of potential interactions between farms. However, neighbouring spatial interactions may vary according to farmers' motivations. To identify specific farms' relationships, a mixture modelling framework that enables capturing heterogeneity in spatial interdependence between farms is developed and applied to analyse farm survival and growth. An application to a panel of farms in Brittany in France from 2004 to 2014 shows that relationships between farms are more in terms of competition for land than positive spill overs of new technology adoption. This leads to a negative impact of neighbouring farms' size on the probability to survive for a majority of farms. The results also show that neighbouring farms' size has a positive effect on farm growth, but the impact varies in magnitude according to farm types. These results suggest that more attention should be paid to heterogeneity in spatial interdependence between farms for a better understanding of farm size dynamics.

¹This chapter has been written in collaboration with Thomas Heckelei (University of Bonn, Germany), Hugo Storm (University of Bonn, Germany) and Laurent Piet (INRA, SMART-LERECO, France).

4.1 Introduction

The farming sector faced considerable structural change over the last decades. In most developed countries, the total number of farms decreased significantly and their average size increased, implying changes in the distribution of farm sizes (Bollman *et al.*, 1995; Eastwood *et al.*, 2010). Understanding the factors that affect farmers' decisions to enter or exit farming as well as farm growth has been a concern of agricultural economists and policy makers for quite some time. For example, Shapiro *et al.* (1987) examined the relationship between farm size and growth in Canada over the period 1966-81; Sumner and Leiby (1987) analysed the effects of human capital on size and growth to a sample of Southern dairy farms in the United States; Weiss (1999) investigated farm survival and growth in Upper Austria; Breustedt and Glauben (2007) examined the exit process of Western European farmers; Dong *et al.* (2010) studied the exit decision of finisher hog producers in North America. In France, Aubert and Perrier-Cornet (2009) studied factors that influence survival and growth of small farms while Bakucs *et al.* (2013) investigated the relationship between size and farm growth. Among others, these studies identified important aspects of structural change in farming and showed that farm survival and growth processes may help understand farm dynamics in different farming contexts. More recently, Storm *et al.* (2015) empirically investigated the effects of direct payments on exit rates of Norwegian farms and showed that the spatial interdependence between farms is an important factor in farmers' decisions to maintain their production activities. The authors showed that accounting for spatial interdependence between farms may be highly relevant for an aggregate assessment of policy changes in agriculture.

The study presented in the present chapter adds to the existing literature especially in three ways. Firstly, we account for spatial interdependence between farms in both farmers' decision to continue their production activities over time and the process of farm growth. Indeed, neighbouring farms' characteristics may influence farm survival as well as farm growth since these two processes may be related to each other when they impact farms which are close to each other. Accounting for interdependence in both farmers' decision to stay in business and farm growth process may therefore improve the analysis of structural change in farming. Some studies have already investigated both farm growth and survival in different farming contexts (Weiss, 1999; Aubert and Perrier-Cornet, 2009). However, none of the existing studies have taken into account the potential spatial interdependence between farms, to the best of our knowledge.

Secondly, we extend the existing methods by using a mixture modelling approach to investigate heterogeneity in spatial interdependence between farms. Generally, studies in this strand of the literature estimate mean effects of neighbouring farms' characteristics on farmers' decision to exit farming or to increase their operated farm size (see Storm *et al.* (2015) for a recent example). The results from these studies are therefore based on the assumption that

all farms will behave alike given the investigated characteristics. However, some farms may be more or less sensitive to the characteristics of their neighbours (e.g., direct payment received, farm size, etc.), due to some specific individual characteristics. If all of these characteristics were observed, controlling for them would lead to a more efficient estimation of the impacts of neighbouring farms' characteristics. Otherwise, the resulting parameters may be biased and inconsistent due to unobserved farm heterogeneity (Kyriazidou, 1997; Pennings and Garcia, 2004). One way to tackle this issue is to use modelling frameworks that allow controlling for unobserved farm heterogeneity. Various modelling approaches such as fixed and random effect, random parameter and mixture models can be used to control for unobserved heterogeneity (Greene, 2012). Holloway *et al.* (2007) reported several strategies that could be adopted in order to control for unobserved heterogeneity in modelling spatial dependence. Among these strategies, these authors argued that the mixture modelling framework seems to be the most attractive since it is simple and intuitive. According to them, one of the main advantages of a mixture modelling approach is that it allows the data themselves to sample select and it gathers observations characterised by similar relations between dependent and independent variables. The mixture modelling approach can group farms with similar behaviours and therefore could help identify specific impacts of neighbouring farm's characteristics.

Thirdly, we develop the mixture modelling approach in order to handle panel data to capture potential dynamic effects in farmers' decisions. Some studies in the literature have already used panel data to study farmers' decisions to exit farming and farm growth process (see Bakucs *et al.* (2013) for a recent example). However, to the best of our knowledge, we are the first to investigate spatial farm interdependence both using panel data and controlling for unobserved heterogeneity. The fundamental advantage of a panel dataset over a cross section one is that the former allows greater flexibility in modelling differences in the behaviour of individuals (Greene, 2012). We can therefore expect that using a mixture modelling approach could group farms with similar behaviours and thus reveal different impacts of neighbouring farms' characteristics on farmers' decision to exit farming or to increase their production capacity.

This chapter extends the two preceding chapters of the thesis by investigating both farm heterogeneity and the spatial interdependency between neighbouring farms. Indeed, while Markovian approaches enable mirroring farm interdependency by considering at the same time several dimensions of structural change, it is unable to identify specific types of interactions between farms. However, understanding the relationships that may exist between farms may give a better insight into the ongoing structural change in farming. The methodology applied here takes advantage of the M-MCM methods developed in chapter 2 and 3 to extend an econometric model which also accounts for multi-level spatial variables.

This chapter is structured as follows. The next section provides theoretical arguments supporting the empirical application of this study. Sections 4.3 and 4.4 present the modelling

approach and the corresponding estimation procedure, respectively. The data used for our empirical application and explanatory variables for the model specification are presented in Section 4.5. Section 4.6 reports the main results. The last section concludes with some considerations on possible improvements of this study for further research.

4.2 On heterogeneity and spatial interdependence between farms

Neighbouring farms' characteristics may have important impacts on own farm size and/or on farmers' decision to exit farming. Indeed, farmer's decisions result from their interactions on the land market which is highly localized because of the immobility of land (Happe *et al.*, 2006). As farms compete for a fixed asset of land, transactions within the land market are driven by the relative difference between farms' willingness to pay (WTP) for land. Focusing on farm survival, a farm will survive if its WTP for land is greater than the WTP for land of its neighbours. As the WTP for land of a farm depends on the farm characteristics, the farmer's decision resulting from the difference in his/her WTP for land is therefore related to his/her neighbouring farms' characteristics. In this study, we argue that the impact of the characteristics of neighbouring farms also depends on the own characteristics of the farmer under consideration. Focusing on the neighbouring farms' specific characteristic that is size, we extend Storm *et al.* (2015)'s theoretical background providing some additional elements supporting this proposition.

The existing literature distinguishes two types of effects of neighbouring farms' size originating from technology adoption. On the one hand, neighbours can be viewed as competitors especially for the acquisition of plots (Weiss, 1999). In this case, a farmer surrounded by larger farms may be constrained to close his/her operation since larger farms are more likely to adopt new technologies earlier given their potentially greater access to information and better financial capacity (Goddard *et al.*, 1993). Larger neighbours therefore have a higher WTP for land, leading to a negative impact on the probability to survive and to enlarge, for the farm under consideration. On the other hand, neighbours can be considered as a source of motivation and example to adopt new technologies (Case, 1992; Holloway *et al.*, 2002). In this case, neighbouring farms' size positively influences the survival of the farm under consideration, because a farmer surrounded by larger farms is more likely to benefit from the innovation of larger neighbouring farms (Harrington and Reinsel, 1995). This may imply an increase in the WTP for land for those neighbouring farms since new technology adoption generally requires acquisition of land for an optimal use of the technology.

However, these interactions among neighbours may depend on the farm under consideration. Indeed, we expect that the effect of neighbouring farms' size is rather heterogeneous across farms under consideration, and crucially depends on the type and characteristics of the farm and the farmer considered. One of the most important sources of farm heterogeneity that may

shape farmers' behaviour is their motivation: neighbouring farms' size is more likely to have an impact (positive or negative) on farmers who are mainly motivated by profit maximisation. In the context of a free market competition, such business-oriented farms are constrained either to innovate or to exit, leaving resources to be acquired by more innovative competitors in the latter case (Harrington and Reinsel, 1995; Jackson-Smith, 1999). The persistence of commercial farms thus depends on their competitiveness, that is, on their capacity to innovate. However, this capacity differs across farms and depends on a variety of factors such as accessibility to technology and land, managerial capacity, risk perception, attitudes towards risk, etc. (Bowman and Zilberman, 2013; Conradt *et al.*, 2014; Trujillo-Barrera *et al.*, 2016).

Nevertheless, not all farmers give priority to the commercial aspect of farming activities (Maybery *et al.*, 2005; Howley *et al.*, 2014). Some non-competitive farmers may keep their activity because of prevailing non-pecuniary motives (Harrington and Reinsel, 1995), perhaps enjoying the farming lifestyle (Hallam, 1991), and may maintain production at sub-optimal levels (Howley, 2015). For example, it may be the case for some environmentally oriented farms (Willock *et al.*, 1999) or certain hobby farms (Daniels, 1986; Holloway, 2002). For such kinds of farms, the new technology is evaluated on aspects other than financial viability (Mzoughi, 2011). Then, technology is adopted only if it is considered as conform to some predefined criteria that are set by the farmer based on his/her non-pecuniary objectives. Overall, one can thus expect a lower or even no specific impact of neighbouring farms' size on the probability to survive or to increase the operated size, for those farmers characterised by prevailing non-pecuniary motives.

Based on these considerations, we hypothesise that there are at least two different types of farms that respond differently to neighbouring farms' size: a negative response because of competition for land, or a positive response resulting from positive spill overs of new technology adoption. Two questions arise from this: which farms are more likely to be in each of these specific types? And which type of relationship is the prevailing one in a specific farming context? Investigating both questions may help understand farm size dynamics in specific farming contexts. In the following, we apply a modelling framework that enables identifying specific types of relationships between farms and thus contributes to answer the second question. Even though, we identify the main characteristics of farms in each of these types in a second step, the first question remain out of the scope of this study and would need to be investigated in future works.

4.3 Modelling approach

Modelling farm survival and farm growth requires special care because the estimation of farm growth drivers may face the problem of the so-called selection bias due to the fact that growth only applies to surviving farms (Weiss, 1999; Aubert and Perrier-Cornet, 2009). As a solution,

the well-know Heckman sample selection model (Heckman, 1976, 1979) is a widely used approach to link two such models. However, the data we used for this analysis face limitations that impede using Heckman’s approach. Indeed, implementing the sample selection model would require exclusion restrictions, that is, variables that are good predictors of farm survival but that are not linked to farm growth. In the absence of exclusion restrictions, Heckman’s two-step estimator may be subject to collinearity problems (Little, 1985; Leung and Yu, 1996; Puhani, 2000; Lennox *et al.*, 2012). As will seen in section 4.5.1, such explanatory variables are not available in the database used for this study. We thus separately investigate the impacts of neighbouring farms’ size on farm survival on the one hand, and on the process of farm growth on the other hand, keeping the potential selection bias in mind.

Regarding farm survival, a probit model is applied. The probit model can be motivated by a latent utility model, where the latent variable represents the utility that is obtained from staying in or exiting the farming sector. Farmers’ utility may be affected by their own characteristics as well as their neighbours’ characteristics. The latent variable y_{it}^* underlying the probit model determines the outcome of the farmer’s decision to stay in business in two consecutive years. As yearly information about farmers’ decisions is available, we define:

$$\begin{aligned} y_{it} &= 1 & \text{if } y_{it}^* > 0, & \quad \forall t \in T_i \\ y_{it} &= 0 & \text{if } y_{it}^* \leq 0 \end{aligned} \tag{4.1}$$

where y_{it} is the observed outcome at time t for farm i which takes values: $y_{it} = 1$ if the farm survives two consecutive years, and $y_{it} = 0$ otherwise; T_i is the length of time that farm i is observed. The latent variable at time t is in turn given by (Greene, 2004):

$$y_{it}^* = \mathbf{x}_{it-1}\boldsymbol{\beta} + \epsilon_{it}, \quad t = 1, 2, \dots, T_i \leq T \tag{4.2}$$

where $\boldsymbol{\beta}$ are the parameters to be estimated, \mathbf{x}_{it-1} are own and neighbouring farm characteristics. The disturbances ϵ_{it} are T-variate, normally distributed with $T \times T$ positive definite covariance matrix $\boldsymbol{\Sigma}$.

Farm growth is represented as a simple linear regression model. The total land used at any specific time t is thus given by:

$$z_{it} = \mathbf{x}_{it-1}\boldsymbol{\theta} + u_{it}, \quad t = 1, 2, \dots, T_i \leq T \tag{4.3}$$

where $\boldsymbol{\theta}$ is the vector of parameters to be estimated, x_{it-1} are own and neighbouring farms’ characteristics, and u_{it} is an *iid* normally distributed error term.

As argued in the previous section, neighbouring farms’ size may influence farmers’ decisions in various ways. To capture the heterogeneity in farmers’ responses to their neighbouring

farms' characteristics, we apply a mixture modelling approach, which allows capturing unobserved heterogeneity. The mixture modelling approach supposes that the farm population is divided into more than one homogeneous group; each type of farms is characterised by a specific effect of the exogenous variables, including neighbouring farms' size, on farmers' decisions. Let $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)$ denote the observed random sample where \mathbf{y}_i is the sequence of choices or states of farm i over a certain period of time. Under a mixture approach, the density $f(\mathbf{y}_i)$ is written as (McLachlan and Peel, 2004):

$$f(\mathbf{y}_i) = \sum_{g=1}^G \pi_g f_g(\mathbf{y}_i) \quad (4.4)$$

where π_g is the proportion of farms belonging to type g with $g = 1, 2, \dots, G$, and f_g is type- g density as described by equations (4.2) and (4.3) for farm survival and growth process, respectively. Since the unobserved types have to be exhaustive and mutually excluding, the π_g proportions are non-negative and sum up to unity.

Under such a mixture approach, the conditional probability density for the observed data for farm i is:

$$f(\mathbf{y}_i | \mathbf{X}_i; \Psi) = \sum_{g=1}^G \pi_g f_g(\mathbf{y}_i | \mathbf{X}_i; \Phi_g) \quad (4.5)$$

where $\Psi = (\pi_1, \dots, \pi_G, \Phi_1, \dots, \Phi_G)$ are the parameters to be estimated with $\Phi_g = \beta_g$ for the probit model and $\Phi_g = \theta_g$ simple linear regression model; and f_g is the respective probability density function specific to farm type g , given by:

$$f_g(\mathbf{y}_i | \mathbf{X}_i; \Phi_g) = f(\mathbf{x}_{it-1}; \beta_g) = [F(\mathbf{x}_{it-1} \beta_g)]^{y_{it}} [1 - F(\mathbf{x}_{it-1} \beta_g)]^{(1-y_{it})} \quad (4.6)$$

for the probit model where $F(\mathbf{x}_{it-1}; \beta_g)$ is the cumulative density function for and farm type g and y_{it} is the observed outcome. For the simple linear regression model, the density function writes:

$$f_g(\mathbf{y}_i | \mathbf{X}_i; \Phi_g) = f(\mathbf{x}_{it-1}; \theta_g) = \frac{1}{\sqrt{2\pi\sigma_g^2}} \exp \left\{ -\frac{1}{2\sigma_g^2} (z_{it} - \mathbf{x}_{it-1} \theta_g)^2 \right\} \quad (4.7)$$

where σ_g^2 is the variance of the error term specific to farm type g , and z_{it} is the dependent variable.

4.4 Estimation of the models

The mixture models described in the previous section are estimated using the maximum likelihood method. Assuming that, for each model, observations are independent within farm types given \mathbf{x}_{it-1} , the log-likelihood (LL) function for the parameters Ψ of the model, conditional on

observing \mathbf{y}_i , is written as:

$$LL(\Psi) = \sum_{i=1}^N \ln \left\{ \sum_{g=1}^G \pi_g \prod_{t=1}^{T_i} f(\mathbf{x}_{it-1}; \Phi_g) \right\} \quad (4.8)$$

4.4.1 Implementing the Expectation-Maximisation (EM) algorithm

As the type of farms is unknown beforehand, the expectation-maximisation (EM) algorithm is used to estimate the parameters of the models. The EM algorithm simplifies the complex log-likelihood in equation (4.8) into a set easily solvable log-likelihood functions by treating the unobserved farm type as a missing information (McLachlan and Krishnan, 2007). Using a non-parametric mixing distribution approach, the complete log-likelihood is thus written as (Train, 2008):

$$LL_c(\Psi) = \sum_{i=1}^N \sum_{g=1}^G v_{ig} \ln \left\{ \pi_g \prod_{t=1}^{T_i} f(\mathbf{x}_{it-1}; \Phi_g) \right\} \quad (4.9)$$

In this case, v_{ig} is called the ‘posterior’ probability that farm i belongs to type g given \mathbf{y}_i , that is $P(v_{ig} = 1 | \mathbf{y}_i)$, while π_g is the ‘prior’ probability of the mixture (McLachlan and Peel, 2004). The expression in equation (4.9) can then be divided into two components:

$$\begin{aligned} LL_1 &= \sum_{i=1}^N \sum_{g=1}^G v_{gi} \ln \pi_g \\ LL_2 &= \sum_{i=1}^N \sum_{g=1}^G v_{gi} \sum_{t=1}^{T_i} \ln f(\mathbf{x}_{it-1}; \Phi_g) \end{aligned} \quad (4.10)$$

As the farm type is not observed, the posterior probability that farm i belongs to type g (*i.e.*, v_{ig}) has to be estimated from the observations. The EM algorithm therefore consists of the four following steps:

(i) Initialisation: Arbitrarily choose initial values $\Psi^0 = (\pi_1^0, \dots, \pi_G^0, \Phi_1^0, \dots, \Phi_1^0)$ for the parameters of the models.

(ii) Expectation: At iteration $p + 1$ of the algorithm, compute the expected probability that farm i belongs to a specific type g while observing \mathbf{y}_i and given the parameters $\Psi^{(p)}$. This conditional expectation probability, that is, the posterior probability $v_{ig}^{(p+1)} = v_{ig}(\mathbf{y}_i; \Psi^{(p)})$, can be obtained according to the Bayes’ law:

$$v_{ig}^{(p+1)} = \frac{\pi_g^{(p)} \prod_{t=1}^{T_i} f(\mathbf{x}_{it-1}; \Phi_g^{(p)})}{\sum_{h=1}^G \pi_h^{(p)} \prod_{t=1}^{T_i} f(\mathbf{x}_{it-1}; \Phi_h^{(p)})} \quad (4.11)$$

Replacing v_{ig} by its expected value in equation (4.9) leads to the conditional expectation of the complete data log-likelihood.

(iii) Maximisation: Update Φ^p by maximising the complete log-likelihood conditional on the observations. The models' parameters are thus updated as:

$$\Phi^{(p+1)} = \underset{\Phi}{\operatorname{argmax}} \sum_{i=1}^N \sum_{g=1}^G v_{ig}^{(p+1)} \sum_{t=1}^{T_i} \ln f(\mathbf{x}_{it-1}; \Phi_g) \quad (4.12)$$

The maximisation process of equation (4.12) is straightforward. The parameters of the model ($\hat{\Phi}^{(p)}$) are updated considering $v_{ig}(\mathbf{y}_i; \Psi^{(p)})$ as weighting factors for each observation. Then, the prior probability of belonging to type g is updated as:

$$\pi_g^{(p+1)} = \frac{\sum_{i=1}^N v_{ig}^{(p+1)}}{\sum_{i=1}^N \sum_{h=1}^G v_{ih}^{(p+1)}}, \quad \forall g \in G \quad (4.13)$$

(iv) Iteration: Return to expectation step (ii) using $\pi^{(p+1)}$ and $\Phi^{(p+1)}$ and iterate steps (ii) and (iii) until convergence of the observed log-likelihood given by equation (4.8).

At convergence, the resulting parameters ($\hat{\Psi}$) are considered as optimal. Because of the potential presence of a high number of local maxima (Hess *et al.*, 2006), the EM algorithm is run several times with various randomly chosen initial values, and those providing the largest likelihood at convergence are chosen as the best ones.

4.4.2 Choosing optimal number of farm types

Despite the intuition about relationships between farms as described in Section 4.2, we have no a priori information about the optimal number of homogeneous farm types that may exist in a specific farming context. The total number of components for the mixture of probit models as well as for the mixture of linear regression models are thus chosen based on information criteria. The selected criteria are derived on the resulting value of the log-likelihood of the corresponding model $LL_G(\mathbf{y}; \hat{\Psi})$ for a total of G homogeneous types. The basic principle under these information criteria is parsimony, that is, all other things being the same, the model with fewer parameters is preferred (Andrews and Currim, 2003). The selection criteria thus derive from the following formula:

$$C_G = -2 \left\{ LL_G(\mathbf{y}; \hat{\Psi}) \right\} + \kappa N_G \quad (4.14)$$

where $LL_G(\mathbf{y}; \hat{\Psi})$ is the overall population log-likelihood value computed with the resulting estimated parameters for the model specified with G types, N_G is the total number of free parameters in the model and κ is a penalty constant.

Depending on the value chosen for κ , we obtain the well-known Akaike Information Criterion (AIC) if $\kappa=2$, and Bayesian Information Criterion (BIC) if $\kappa = \log(N)$ where N is the total number of observations. However, based on Monte Carlo simulations, it has been proved that the modified AIC (AIC3) with $\kappa=3$ is preferable to the basic AIC and BIC since the two latter criteria more severely penalise the addition of parameters than AIC3 ([Andrews and Currim, 2003](#)). For all of these heuristic criteria, smaller values mean more parsimonious models.

4.5 Empirical application

4.5.1 Data used

For our empirical application, we used data provided by the ‘Mutualité Sociale Agricole’ (MSA), the French authority for farmers’ healthcare and social security. The MSA database contains information about all individuals who declare carrying out a non-salaried farming activity in France, and about their farm. Information is collected annually and is available for farmers who were active on January 1st of each year, from 2004 to 2014. The database can be considered as almost exhaustive for the French farm population², so we can assume that a farm: i) survives if it remains in the MSA database two consecutive years; ii) starts business if it enters the database after 2004; iii) quits farming if it leaves the database before 2014.

Using the MSA data requires several preliminary treatments to prepare the database before analysis. Here we only mention the most significant ones. First, the data have to be consolidated at the farm level because the MSA collects information at the affiliated physical person level, *i.e.*, the farm holder level and not the farm itself. This is made possible because, in the database, each individual person is assigned the farm number he/she belongs to. Second, the utilised agricultural area (UAA) has to be aggregated with care at the farm level because it is not simply the sum of the areas reported for each partner of the farm. Indeed, the recorded UAA at the individual partner level is calculated with respect to the proportion of the total social shares of the farm he/she holds. Then, whenever one or several partners are not affiliated to the MSA because they are external to the farming sector, social shares in the database do not sum up to unity, and hence the total UAA has to be computed taking into account that a part of it accrues to partners who are not observed in the database. Third, assumptions have

²The database is considered as ‘almost’ exhaustive because only it does not survey small farms which do not contribute to the MSA as well as corporate farms employing only salaried workforce.

to be made for other variables when consolidated at the farm level in the case where the farm is run by several partners. This is typically the case for some of the variables used in the analysis: age, farming specialisation and legal status. As for age, a choice has to be made on the 'age' of which person characterises the farm. Here, we arbitrarily chose to retain the median age of the farm's partners. As for the farm production specialisation, it is determined from the category of professional risks each partner is registered to. In the database, there exists 16 such risk categories such as 'cereals and industrial crops', 'dairy cattle breeding', 'pig farming', 'wine growing', etc. Then, when the farm includes several production units (*e.g.*, a 'crop' unit plus a 'livestock' unit in dairy farms), each partner may subscribe for only one of the corresponding professional risks depending on the unit he/she is specialised in. Therefore, several such risk categories may coexist on the same farm and an assumption has to be made on how to aggregate them to avoid classifying such farms as 'mixed farms'. Here, we chose to assign to the farm the risk category which represented the two thirds of partners or, whenever such majority did not exist, to classify the farm as mixed. In a last step, the MSA risk categories were translated into 13 'types of farming' (farm production specialisation) chosen in the nomenclature used by the French agricultural statistics office for the agricultural census and related surveys. Finally, as regards the legal status, while it should be recorded identically for the different partners of a specific farm, it appeared that this was not so in some cases. In such situations, we assigned the predominant legal status (*i.e.*, the mode of the observed values) or, when this was not possible because of several modes, we assigned the status corresponding to the higher degree of incorporated form of association.

In this empirical application, we restricted our investigation to farms located in Brittany (Western France), which is one of the largest agricultural regions in France.

4.5.2 Dependent and explanatory variables

For the analysis of farm survival, the dependent variable takes the value 1 if the farm survives and 0 otherwise. Since we consider all farms whatever their production specialisation or legal status, a farm is said to survive from one year to the next if it remains present in the database these two consecutive years. For the analysis of farm growth, the dependent variable consists in the change in the total UAA in hectares over the period of observation, with an increase representing a positive growth or enlargement, and a decrease representing a negative growth or decline.

The analysis of the spatial interdependence between neighbouring farms in the process of farm survival and growth requires special attention because the MSA database exhibits two main limitations for such a study.

Firstly, the MSA database contains only a few variables that can be used to explain farm survival and growth. We thus choose to concentrate on the possible impacts of the limited set of available variables. Other databases are merged with the MSA to provide additional information especially at different spatial scales. The most important farm characteristic that may play a role in the probability to survive is farm size in terms of total UAA (*area*) and farm total agricultural profit (*agri_profit*). Both explanatory variables are expected to positively influence the probability to survive and to increase the operated farm size because such farm characteristics may increase the farm's WTP for land. While the total land used is rather a proxy of path dependency, the total agricultural profit indicates whether or not farming is a profitable activity. The age of the farm holder (*age*), dummies indicating that the farm production specialisation is pig and/or poultry (*pig/poultry*), and a dummy indicating that the legal status of the farm is a corporate farm in opposition to partnerships or individual farms (*corporate*), are also included in the model specification. These variables are introduced to capture farm observed heterogeneity. Age square is used to capture non-linear effect of the farm holder age. Indeed, the age of farmers may be positively related to the probability of surviving and of increasing the operated size, since farmers' experience and skills may increase over years; by contrast, older farmers especially close to retirement time, may be less motivated to either compete for land or to adjust their operated size over time, leading to the opposite effect. As farm specialisation in pig and/or poultry generally requires less land, this type of farms may be less likely to compete for new plots. We thus expect a positive effect on farm survival and a negative impact on farm size change in term of total land used. Corporate farm legal status is also supposed to be positively related to farm survival and growth because these farms are generally in a better position to compete for land since they may have lower financial and credit constraints. As a farm's WTP for land may decrease at retirement time despite high agricultural profits, we control for the impact of retirement time by using an interaction term between farm agricultural profit and a dummy indicating that the farmer is close to retirement time (*agr_profit × retirement*). According to the MSA, the minimum age for retirement in France is 60 years old but farmers' behaviour may change earlier. Since some studies have indeed shown that farmers' succession is prepared between 5 and 10 years in advance, we choose to retain 55 years old and above as the indicator of retirement closeness (Gaté and Latruffe, 2016).

Secondly, the MSA database contains no information about the precise geographical location of the farmstead and farm plots. It is therefore impossible to determine the actual distance between farms. Only the municipality where the farmstead is located is available in the database. As municipalities in France are relatively small and given the dispersion of farm plots on French farms, farms may compete for land in their own municipality and even in neighbouring municipalities (Piet and Cariou, 2014; Latruffe and Piet, 2014). We thus use average farm characteristics at the municipality level to capture the effects of neighbouring farms' size on a farm's survival and growth. At a first spatial scale, we consider farms located in the

same municipality as the farm under consideration. The sample consists of 1,252 municipalities over the 1,270 in Brittany with an average area of 21 (with a standard deviation of 12) square km. The average number of farms by municipality in the sample is 64 (with a standard deviation of 38). From this, we calculate the average farm size by municipality (*average_mun_area*) and use it as a proxy for neighbouring farms' size. We also calculate, using the MSA database, the average age of farm holders (*average_mun_age*), the share of farms specialised in pig and/or poultry (*mun_pig/poultry_share*) and the share of corporate farms (*mun_corporate_share*) at the municipality level.

Following Storm and Heckeles (2016), we also include the same variables calculated at a larger spatial scale than the municipality. This allows distinguishing the effects of farm interactions that take place on a smaller spatial scale from spatial correlation arising from unobserved spatially correlated regional characteristics at a larger scale. Specifically, we calculate the average characteristics and shares for small agricultural regions (SAR), which is a geographical unit that usually groups several municipalities. The SAR level is a zoning that was specifically designed to define units with homogeneous conditions in terms of agricultural systems, soil and climate. The mean size of a French SAR is 22.4 (with a standard deviation of 13) square km (Teillard *et al.*, 2012). Based on the INSEE 2007 classification, there exist 25 SAR in Brittany. The average number of municipality by SAR in the sample is 50 (with a standard deviation of 42). The variables (farm area, age of farm holder, pig and/or poultry specialisation and corporate legal status) are defined here at the SAR level as: *average_sar_area*, *average_sar_age*, *sar_pig/poultry_share* and *sar_corporate_share*.³

Additionally, we use the rate of unemployment in employment regions (*unempl_rate*). The unemployment rate captures the opportunities for off-farm activities and is thus supposed to have a direct effect on the probability for farms to remain in farming, and only an indirect effect on farm growth through its impact on farm survival. A time trend is used in addition. It may capture potential effects of, for example, technical change in farming that may influence farm survival and growth.

We use the same set of explanatory variables to explain both farm survival and growth. In the second model describing farm growth, the logarithm of the total UAA is used instead of total UAA *per se* to facilitate convergence of the model. All explanatory variables are lagged one year because it is supposed that farmers take their decisions based on available information in the year just preceding. Descriptive statistics are reported in Table 4.1.

Before starting with the results, it should be finally noted that the mixture of probit and the mixture of linear regression models are estimated separately. As explanatory variables are lagged one year for both models, the estimations are thus performed using 317,177 and 294,288

³It should be noted that farm agricultural profit is not defined here neither at the municipality level nor at the SAR level since it is found highly non-significant at both spatial levels.

TABLE 4.1: Definition and descriptive statistics of explanatory variables (n=344,617).

Variable	Code	Mean	St.Dev.	Min	Max
<i>Farm level</i>					
Age of the farm holder (years)	<i>age</i>	48.45	9.12	18.50	99.00
Total UAA (ha)	<i>area</i>	48.82	41.20	0.00	580.30
Total agricultural profit (1,000 Euros)	<i>agri_profit</i>	10.78	12.72	-313.92	465.72
Pig/poultry specialisation dummy (1 if yes)	<i>pig/poultry</i>	0.18	0.38	0.00	1.00
Corporate farm dummy (1 if yes)	<i>corporate</i>	0.46	0.49	0.00	1.00
<i>Municipality level (mun)</i>					
Average farm holder age	<i>average_mun_age</i>	48.45	2.33	25.00	88.00
Average farm size	<i>average_mun_area</i>	48.82	13.60	0.00	227.29
Share of pig/poultry farms (%)	<i>mun_pig/poultry_share</i>	18.00	13.00	0.00	100.00
Share of corporate farms (%)	<i>mun_corporate_share</i>	46.00	14.00	0.00	100.
<i>Small agricultural region level (sar)</i>					
Average farm holder age	<i>average_sar_age</i>	48.45	1.12	44.30	51.28
Average farm size	<i>average_sar_area</i>	48.82	7.66	13.92	70.61
Share of pig/poultry farms	<i>sar_pig/poultry_share</i>	18.00	7.00	1.00	29.00
Share of corporate farms	<i>sar_corporate_share</i>	46.00	8.00	24.00	70.00
<i>Employment regional level</i>					
Unemployment rate (%)	<i>unempl_rate</i>	7.03	1.21	3.70	9.90

Source: MSA COTNS database, Bretagne 2004-2014 - authors' calculations

observations out of the 344,617 in the database, respectively for the models of farm survival and farm growth. In the following section, the results are thus presented for each model separately.

4.6 Results

4.6.1 Farm survival model

Table 4.2 reports the estimated parameters for both a pooled estimation where unobserved heterogeneity is not considered and the mixture probit model. Estimated parameters of the pooled probit thus constitute a mean effect of the considered farms' own characteristics and neighbouring farms' characteristics on the probability to survive from one year to the next, while the mixture model identifies impacts which are specific to the endogenously determined homogeneous farm types.

The results from the homogeneous model are consistent with our expectations. Overall, a positive impact is observed for the age of farm holders, the operated farm size (land), and the total agricultural profit. The results show a non-linear impact both for the age of farm holders and the total farm area. The negative impact of the square of age means that older farm holders are less likely to remain active over years. By contrast, the effect of the square of farm area is also positive but lower than the effect of area, which may suggest that very large farms may face some constraints that tend to decrease their probability to survive in comparison to smaller

TABLE 4.2: Estimated parameters for both the pooled and the mixture probit model for farm survival.

Variable code	Pooled	Mixture		
		type 1	type 2	type 3
<i>intercept</i>	0.0358 (0.3396)	1.3314** (0.4774)	-0.8903* (0.3875)	-60.2809*** (1.3082)
<i>time_trend</i>	0.0062** (0.0028)	-0.0222*** (0.0040)	-0.0286*** (0.0033)	0.1037*** (0.0092)
<i>age</i>	0.0104*** (0.0025)	-0.0258*** (0.0036)	0.0066* (0.0027)	3.2279*** (0.0328)
<i>age_square</i>	-0.0003*** (2.23e-05)	2.10e-05 (3.26e-05)	-0.0001*** (2.51e-06)	-0.040*** (0.0004)
<i>area</i>	0.0042*** (0.0002)	-0.0173*** (0.0004)	0.0092*** (0.0002)	0.0007 (0.0008)
<i>area_square</i>	9.64e-06*** (1.09e-06)	0.0001*** (2.29e-6)	2.40e-05*** (1.30e-06)	4.37e-05*** (6.01e-06)
<i>agri_profit</i>	0.0009** (0.0004)	-0.0499*** (0.0007)	0.004*** (0.0004)	0.0508*** (0.0045)
<i>agri_profit_x_retirement</i>	-0.0185*** (0.0006)	-0.0283*** (0.0008)	-0.0162*** (0.0006)	-0.0515*** (0.0046)
<i>pig/poultry</i>	0.0228** (0.0105)	0.2074*** (0.0167)	0.0281* (0.0119)	0.2085*** (0.0333)
<i>corporate</i>	0.3093*** (0.0091)	1.2208*** (0.0144)	0.2897 (0.0101)	-0.0157 (0.0319)
<i>average_mun_age</i>	0.0051*** (0.0018)	-0.0010 (0.0025)	0.0053* (0.0021)	0.0202*** (0.0061)
<i>average_mun_area</i>	-0.0003 (0.0004)	0.0049*** (0.0005)	-0.0013*** (0.0004)	-0.0001 (0.0012)
<i>mun_pig/poultry_share</i>	0.0104 (0.0354)	-0.0373 (0.0504)	0.0062 (0.0407)	-0.1104 (0.1149)
<i>mun_corporate_share</i>	-0.0545 (0.0358)	-0.4082*** (0.0497)	-0.0202 (0.0409)	-0.0499 (0.1140)
<i>average_sar_age</i>	0.0181** (0.0071)	0.0435*** (0.0099)	0.026*** (0.0081)	0.1549*** (0.0232)
<i>average_sar_area</i>	0.0018*** (0.0007)	0.0026** (0.0010)	0.0032*** (0.0008)	-0.0022 (0.0024)
<i>sar_pig/poultry_share</i>	0.1488** (0.0709)	1.0079*** (0.0993)	0.1223 (0.0816)	0.1305 (0.2306)
<i>sar_corporate_share</i>	0.1402** (0.0677)	0.4428*** (0.0962)	0.1994** (0.0770)	0.0997 (0.2210)
<i>unempl_rate</i>	0.0104*** (0.0033)	0.0311*** (0.0046)	0.0164*** (0.0038)	0.0797*** (0.0105)
Type shares		17.90%	54.20%	27.90%
Number of observations	317,177		317,177	
Correct predictions	92.73%		93.85%	
Log pseudo-likelihood	-76,323		-73,696	
AIC	152,684		147,470	
BIC	152,886%		147,886	
AIC3	152,703		147,509	

Note: *, ** and *** indicate significance at 10%, 5% and 1% levels, respectively; standard errors in parentheses.

Source: MSA COTNS database, Bretagne 2004-2014 - authors' calculations

farms. A positive effect is also observed for farm specialisation in pig and/or poultry and for farms operated under a corporate legal status. The first result may be linked to the second one. Indeed, corporate farms may be in a better place to compete because of lower financial and credit constraints. In general, farms specialised in pig and/or poultry are most often operated under corporate legal status in France particularly in Brittany. Therefore, these farms may have a higher probability to survive. Farm agricultural profit is also found to positively affect farm survival, but farm holders close to retirement time tend to leave the farming sector although this activity is profitable. Such a result could be explained by the fact that profitable farm may be in a better position to find a successor. The average farm size at the municipality level is not significant, meaning that the pooled probit does not identify any link between neighbouring farm size and the probability of farm to survive. However, the probability to survive is positively related to the average farm size at the small agricultural region level, which indicates unobserved spatial correlation between regional characteristics.

TABLE 4.3: Z-scores for testing the equality of the estimated farm types' coefficients from the mixture probit model for farm survival.

Variable code	z-score (1-2)	p-value	z-score (1-3)	p-value	z-score (2-3)	p-value
<i>intercept</i>	-3.61	0.00	-44.24	0.00	-43.53	0.00
<i>time trend</i>	-1.22	0.22	12.49	0.00	13.50	0.00
<i>age</i>	7.18	0.00	98.55	0.00	97.84	0.00
<i>age_square</i>	-2.49	0.01	-107.10	0.00	-106.99	0.00
<i>area</i>	55.92	0.00	19.68	0.00	-9.86	0.00
<i>area_square</i>	-26.93	0.00	-1.59	0.11	11.02	0.00
<i>agri_profit</i>	67.28	0.00	22.27	0.00	10.43	0.00
<i>agri_profit_x_retirement</i>	11.67	0.00	-4.94	0.00	-7.58	0.00
<i>pig/poultry</i>	-8.75	0.00	0.03	0.98	5.11	0.00
<i>corporate</i>	-53.11	0.00	-35.38	0.00	-9.14	0.00
<i>average_mun_age</i>	1.90	0.06	3.21	0.00	2.31	0.02
<i>average_mun_area</i>	-8.95	0.00	-3.81	0.00	0.91	0.36
<i>mun_pig/poultry_share</i>	0.67	0.50	-0.58	0.56	-0.96	0.34
<i>mun_corporate_share</i>	6.03	0.00	2.88	0.00	-0.25	0.80
<i>saa_age</i>	-1.37	0.17	4.42	0.00	5.24	0.00
<i>saa_area</i>	0.52	0.60	-1.87	0.06	-2.19	0.03
<i>saa_pig/poultry</i>	-6.89	0.00	-3.49	0.00	0.03	0.98
<i>saa_corporate</i>	-1.98	0.05	-1.42	0.16	-0.43	0.67
<i>unempl_rate</i>	-2.45	0.01	4.25	0.00	5.68	0.00

Note: z-scores for testing the null hypothesis that coefficients of two different types are equal (Paternoster *et al.*, 1998).

Source: MSA COTNS database, Bretagne 2004-2014 - authors' calculations

The mixture probit model distinguishes three optimal types in the studied farm population, especially differing with respect to the effect of neighbouring farms' size. The first and the second types of farms in the mixture model are characterised by a significant positive and, respectively, negative impact of neighbouring farms' size on the probability to survive. In the third type, the effect is considerably smaller and not significant. The negative influence of

neighbouring farms' size on the probability to survive is found for the majority of farms (about 54%) while the positive impact is observed only for about 18% of farms. Computed z-scores show that these opposite effects are significantly different at a 1% level (see Table 4.3). For the pooled model, the effect of neighbouring farms' size is negative but insignificant. The different effects of neighbouring farms' size observed for the various groups may explain the insignificant impacts for the overall population, that is to say when such unobserved heterogeneity is not considered.

Referring to the discussion in Section 4.2, the two first types could mostly consist of business oriented farms where farm holders are mainly motivated by profit maximisation. The resulting negative impact of neighbouring farms' size on the probability to survive for type 2 may indicate that farms in this type are rather competitors for land, while the opposite effect for farms in type 1 may originate from positive spill overs of new technology adoption for these farms. For these two types of farms the impact of the average farm size at the SAR level is both positive and not significantly different from each other. This implies that average farm size at the SAR level might indeed capture unobserved spatial correlation while average farm size at the municipality level is more driven by actual spatial competition.

Contrary to the two first farm types, the impact of neighbouring farms' size is highly non-significant for the third type which accounts for about 28% of the farm population in Brittany. Referring again to the discussion in Section 4.2, this initially unexpected type could comprise farms characterised by prevailing non-pecuniary motives. It could be also the case of business oriented farms that have already reached their optimal economic size. The probability of survival for such farms may be therefore independent from the size of their neighbours. This result is in line with the impact of the average farm size at the small agricultural region, which has no significant effect on the probability to survive of this third farm type, contrary to the two first types. This suggests that the farming context has no specific influence on the persistence of such (third type) farms in the sector. This interpretation is confirmed by the positive impact of the time trend, meaning that the probability to survive increases for those (third type) farms over time, while the inverse trend is observed for farms that compete for land. This result is consistent with the evolution of farm size over the years: the larger the neighbours and the higher the competition for land, then the more difficult it becomes to innovate since new adoptions generally require more land.

The descriptive statistics for farm types are reported in Table 4.4. It should be noted that even if the differences between the means of some considered characteristics for two types are very small, the hypothesis of equality of means is rejected in all cases at the 0.1% level. This high level of significance could be due to the large size of the sample used for our estimation and

TABLE 4.4: Main characteristics of each type of farms identified with the mixture probit model for farm survival.

Variable code	Means			t-tests of equality of means		
	Type1	Type2	Type3	(1-2)	(1-3)	(2-3)
<i>age</i>	49.25	48.27	47.77	39.52	60.86	21.35
<i>area</i>	43.91	50.49	48.04	-63.66	-37.70	22.22
<i>agri_profit</i>	8.80	11.02	11.36	-77.33	-78.59	-9.54
<i>pig/poultry</i>	0.17	0.18	0.19	-17.96	-25.60	-9.20
<i>corporate</i>	0.43	0.48	0.45	-35.74	-11.45	21.80
<i>average_mun_age</i>	48.46	48.25	48.31	36.31	25.23	-9.63
<i>average_mun_area</i>	48.19	48.30	48.58	-3.16	-10.66	-7.85
<i>mun_pig/poultry_share</i>	0.18	0.18	0.19	-16.35	-21.09	-6.06
<i>mun_corporate_share</i>	0.46	0.46	0.47	8.35	-13.38	-21.35

Note: t-tests for the null hypothesis that the means of two different types are equal; the significance level are not reported since all t-statistics are significant at a level of 5%.

Source: MSA COTNS database, Bretagne 2004-2014 - authors' calculations

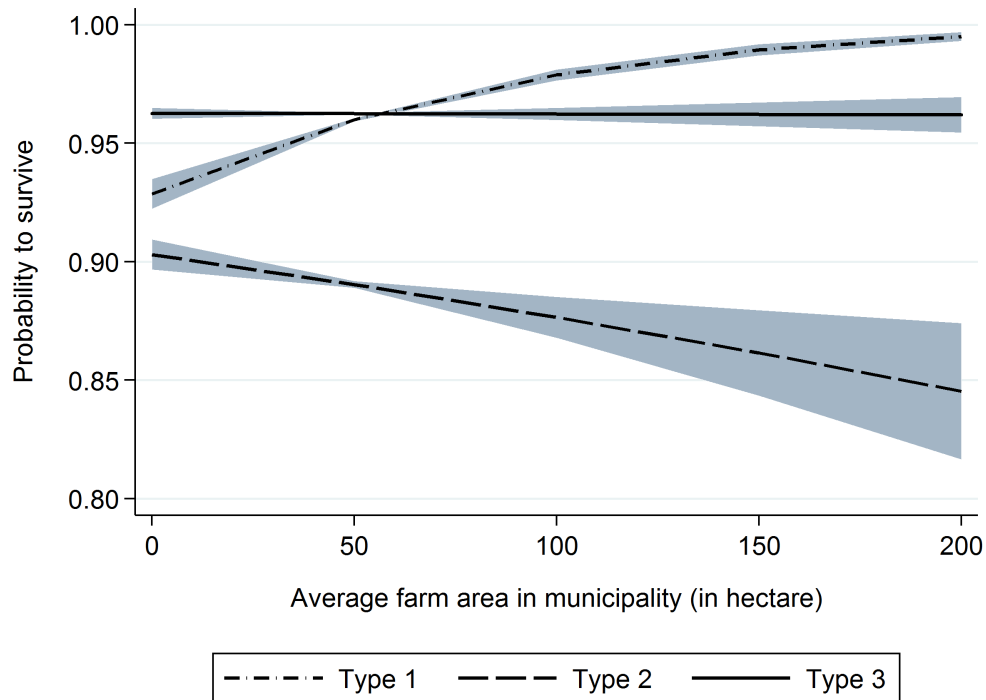
should therefore be interpreted with due care.⁴ Briefly, the results show that larger corporate farms are more likely to behave as competitors for land than farms in the other types. This result conforms with our discussion in Section 4.2 since such farms are more likely to be business oriented and thus mainly motivated by profit maximisation. Conversely, individual farmers with smaller operated farm sizes are more likely to benefit from positive spill overs of new technology adoption.

Figure 4.1 reports the probability that an average farm remains active from one year to the next with respect to the average farm size calculated at the municipality level. Three panels are provided, one for each type of farms. The figure shows that, overall, the probability to survive is lower for competitors for land and this probability decreases with neighbouring farms' size (farm type 2). The opposite effect is observed for farms that benefit from positive spill overs of new technology adoption (farm type 1). Figure 4.1 also shows that the probability to survive is higher and does not vary with the neighbouring farms' size for farms having mainly non-pecuniary motives or already that reached their optimal size (farm type 3).

The impacts of own farm's and farmer's characteristics on the probability to survive also vary according to the specific type a farm belongs to. For example, the first type of farms is characterised by a negative impact of the age of farm holder and of the total agricultural profit, in contrast to the other types for which the impacts are positive. While the result regarding total agricultural profit is more difficult to interpret, the negative impact means that the probability to survive decreases when farmers become old. This may be explained by the fact that young farmers may be more likely to adapt their production capacity using new technology. This result

⁴The t-test statistics are computed as: $t = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{(s_{\mu_1}^2/n_{\mu_1} + s_{\mu_2}^2/n_{\mu_2})}}$. As it can be seen from the formula, the larger the number of observations the larger the resulting t values.

FIGURE 4.1: Probability to survive for varying municipality-level average farm sizes by unobserved farm types (predicted margins with 95% confidence intervals).



Source: MSA COTNS database, Bretagne 2004-2014 - authors' calculations

is consistent with the positive impact of the average age of farm holders at the municipality level on the probability for a farm to survive. The younger the neighbours, the more competitive they are because of a possible relative higher motivation or capacity to innovate by adopting a new technology.

In addition to the fact that the mixture probit model enables identifying specific impacts of neighbouring farm size, it is also more accurate in predicting farm survival in Brittany. The superiority of the mixture model in particular comes from its better performance in the prediction of farm exit. The mixture model performs about 15% better in predicting farm exit in Brittany than the pooled estimation.

4.6.2 Farm growth model

Table 4.5 reports the estimated parameters for both the pooled estimation and the mixture of simple linear regressions for the farm growth process.

Here also, the results from the pooled estimation are conformed to our expectations. Turning to the mixture model, the farm population is divided into three types, characterised

TABLE 4.5: Estimated parameters for both the pooled and the mixture linear regression model for farm growth.

Variable code	Pooled	Mixture		
		type 1	type 2	type 3
<i>intercept</i>	4.0651 (5.6072)	-147.2279*** (1.0430)	-59.9425*** (14.6292)	-17.5979*** (1.0048)
<i>time_trend</i>	0.2524*** (0.0475)	0.0197* (0.0087)	0.4053*** (0.1218)	-0.0964*** (0.0087)
<i>age</i>	-0.711*** (0.0386)	-0.0976*** (0.0110)	0.0854 (0.1563)	0.0025 (0.0066)
<i>age_square</i>	0.0003 (0.0004)	0.0008*** (0.0001)	-0.0093*** (0.0017)	-0.0002*** (0.0001)
<i>ln_area</i>	5.6571*** (0.0372)	52.6593*** (0.0431)	4.8363*** (0.0325)	9.4797*** (0.0176)
<i>agri_profit</i>	0.3538*** (0.0100)	0.0040*** (0.0011)	0.3598*** (0.0151)	0.0100*** (0.0012)
<i>agri_profit_x_retirement</i>	0.1108*** (0.0143)	-0.0138*** (0.0017)	-0.1086*** (0.0300)	0.0123*** (0.0030)
<i>pig/poultry</i>	-9.7902*** (0.1809)	0.2651*** (0.0317)	-18.2485*** (0.4210)	0.1193*** (0.0355)
<i>corporate</i>	27.1756*** (0.1238)	0.4472*** (0.0234)	24.545*** (0.3401)	0.9315*** (0.0366)
<i>average_mun_age</i>	0.7470*** (0.0306)	0.0309*** (0.0058)	1.3864*** (0.0816)	0.0031 (0.0052)
<i>average_mun_area</i>	0.8599*** (0.0072)	0.0085*** (0.0011)	1.2661*** (0.0167)	0.0034** (0.0012)
<i>mun_pig/poultry_share</i>	12.267*** (0.6074)	-0.5257*** (0.1031)	21.3584*** (1.5000)	0.9872*** (0.1132)
<i>mun_corporate_share</i>	-31.8797*** (0.6083)	-0.3832*** (0.1049)	-50.9605*** (1.5183)	0.7515*** (0.1081)
<i>average_saa_age</i>	-0.4408*** (0.1172)	-0.1048*** (0.0216)	0.0917 (0.3054)	0.1777*** (0.0209)
<i>average_saa_area</i>	0.0585*** (0.0122)	-0.0113*** (0.0023)	0.0718* (0.0331)	0.0080*** (0.0021)
<i>saa_pig/poultry_share</i>	-4.3395*** (1.2088)	0.2592 (0.2054)	-25.2777*** (2.9371)	0.3612 (0.2179)
<i>saa_corporate_share</i>	-5.6855*** (1.1420)	-0.8635*** (0.2066)	8.6827** (2.9053)	0.2063 (0.2046)
<i>unempl_rate</i>	0.2425*** (0.0519)	0.1176*** (0.0095)	0.4492*** (0.1371)	-0.0651*** (0.0104)
ln(sigma)	3.4262*** (0.0032)	1.3233*** (0.0038)	3.7508*** (0.0046)	1.1468*** (0.0039)
Type shares		43.72%	27.68%	28.60%
Number of observations	294,288		294,288	
Log pseudo-likelihood	-1,425,877		-1,036,676	
AIC	2,851,794		2,073,434	
BIC	2,852,005		2,073,868	
AIC3	2,851,814		2,073,475	

Note: *, ** and *** indicate significance at 10%, 5% and 1% levels, respectively; standard errors in parentheses.

Source: MSA COTNS database, Bretagne 2004-2014 - authors' calculations

TABLE 4.6: Z-scores for testing the equality of the estimated farm types' coefficients from the mixture simple linear regression for farm growth.

Variable code	z-score (1-2)	p-value	z-score (1-3)	p-value	z-score (2-3)	p-value
<i>intercept</i>	5.95	0.00	89.51	0.00	2.89	0.00
<i>time_trend</i>	3.16	0.00	-9.39	0.00	-4.11	0.00
<i>age</i>	1.17	0.24	7.79	0.00	-0.53	0.60
<i>age_square</i>	-5.93	0.00	-8.20	0.00	5.31	0.00
<i>ln_area</i>	-886.48	0.00	-928.04	0.00	125.53	0.00
<i>agri_profit</i>	23.52	0.00	3.57	0.00	-23.11	0.00
<i>agri_profit</i> × <i>retirement</i>	-3.15	0.00	7.52	0.00	4.00	0.00
<i>pig/poultry</i>	-43.85	0.00	-3.06	0.00	43.47	0.00
<i>corporate</i>	70.69	0.00	11.14	0.00	-69.03	0.00
<i>average_mun_age</i>	16.57	0.00	-3.57	0.00	-16.92	0.00
<i>average_mun_area</i>	75.26	0.00	-3.13	0.00	-75.55	0.00
<i>mun_pig/poultry_share</i>	14.56	0.00	9.88	0.00	-13.54	0.00
<i>mun_corporate_share</i>	-33.23	0.00	7.54	0.00	33.97	0.00
<i>saa_age</i>	0.64	0.52	9.40	0.00	0.28	0.78
<i>saa_area</i>	2.51	0.01	6.32	0.00	-1.92	0.06
<i>saa_pig/poultry</i>	-8.67	0.00	0.34	0.73	8.71	0.00
<i>saa_corporate</i>	3.28	0.00	3.68	0.00	-2.91	0.00
<i>unempl_rate</i>	-2.41	0.02	-12.98	0.00	-3.74	0.00
ln(sigma)	408.72	0.00	-32.40	0.00	-433.94	0.00

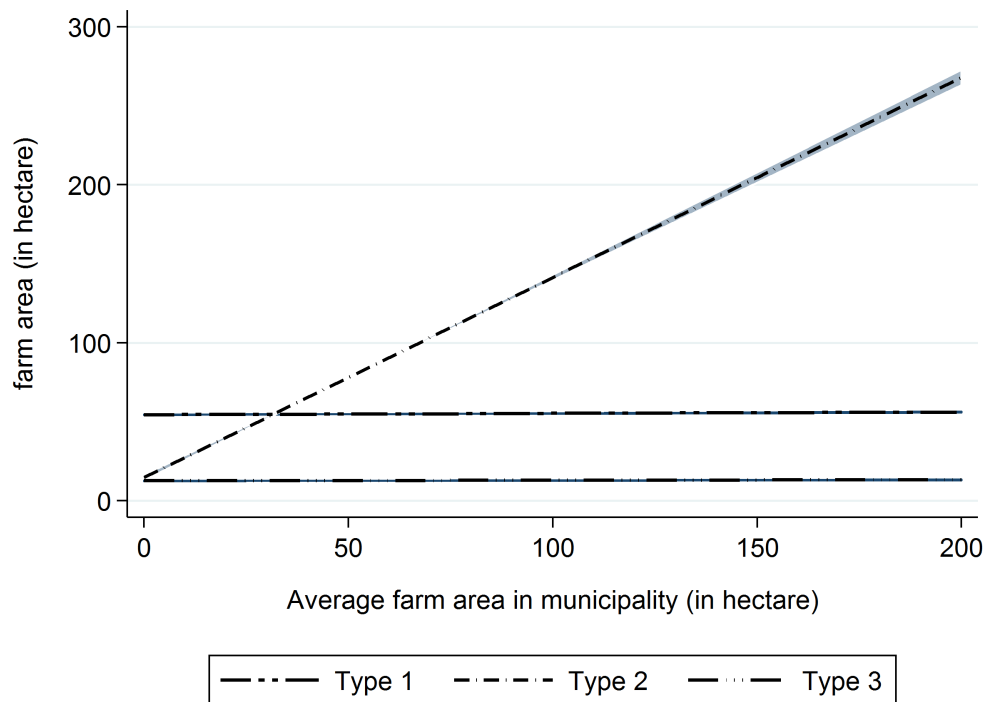
Note: z-scores for testing the null hypothesis that coefficients of two different types are equal (Paternoster *et al.*, 1998).

Source: MSA COTNS database, Bretagne 2004-2014 - authors' calculations

by specific impacts of the explanatory variables. Even though the results show that the studied farm population may be divided into more than one homogeneous type (according to the chosen information criteria), farm size growth is found to be positively related to neighbouring farms' size for all the endogenously determined homogeneous types. The positive impact of neighbouring farms' size on own size growth whatever the farm type may indicate that farmers continually adjust their operated farm size during their lifespan in the sector. It could also suggest an indirect effect resulting from the negative impacts on farm survival, as shown in the previous section. The higher neighbouring farm size the higher exit rate around could be as shown by the farm survival model, leaving plots for the remaining farms in the sector. However, the magnitude of the impact of neighbouring farms' size differs from type to type, with the differences across types being significant at the level of 0.1% (see Table 4.6). This means that heterogeneity still exists in the impact of neighbouring farms' size on the evolution of own size.

The impact of the other explanatory variables included in the model are also different according to the farm type considered. This indicates the existence of different farm profiles not only in term of interactions with their neighbours but also in term of individual behaviours. For example, age of farm holders, share of farms specialised in pig and/or poultry in the municipal-

FIGURE 4.2: Total land used for varying average farm size in municipality by type of farms (predicted margins with 95% confidence intervals).



Source: MSA COTNS database, Bretagne 2004-2014 - authors' calculations

ity, average farm holder age, average farm size and the share of corporate farms in the small agricultural area have negative effects on farm size in type 1, while these variables have opposite impacts for farm types 2 and 3. Even though the square of farm holder age has the opposite sign suggesting non linear effect for farm type 1, the resulting coefficient is very small compared to the coefficient for farm holder age, which indicates a predominance of the (linear) negative impact on farm size. In farm type 2 which presents the highest impact of neighbouring farms' size, specialisations in pig and/or poultry as well as the share of these specialisations at the SAR level negatively influence farm growth. In contrast to farm types 1 and 2, farms in type 3 with the lowest neighbouring farm's size impact tend to decrease operated size over time as shown by the negative sign of the time trend. This result may suggest that these farms may face more constraints with regard to land acquisition. The negative impact of the unemployment rate in the employment area may confirm this assertion: as shown in the previous section, the higher the unemployment rate in the employment area the higher the probability to survive. Therefore, plots could be less available to increase operated size over time.

Figure 4.2 presents the evolution of farm size with respect to neighbouring farms' size for the three identified farm types. It shows that the impact of neighbouring farms' size is relatively low for farm types 1 and 3, while farm size varies almost proportionally to neighbouring farms'

TABLE 4.7: Main characteristics of types from the mixture of normal regression for farm growth.

Variable code	Means			t-tests of equality of means		
	Type1	Type2	Type3	(1-2)	(1-3)	(2-3)
<i>age</i>	47.30	45.55	51.23	52.21	-97.30	-132.96
<i>area</i>	54.52	79.30	12.85	-130.84	749.65	-107.48
<i>agri_profit</i>	12.29	12.59	6.44	-6.11	109.40	100.63
<i>pig/poultry</i>	0.15	0.24	0.18	-56.042	-19.48	33.51
<i>corporate</i>	0.47	0.77	0.21	-154.65	134.21	287.64
<i>average_mun_age</i>	48.18	48.21	48.56	-2.49	-37.26	-32.49
<i>average_mun_area</i>	49.36	51.93	43.81	-46.60	98.91	-185.24
<i>mun_pig/poultry_share</i>	0.19	0.19	0.17	-14.62	36.04	46.55
<i>mun_corporate_share</i>	0.47	0.48	0.44	-23.20	45.84	62.99

Note: t-tests for the null hypothesis that the means of two different types are equal; the significance level are not reported since all t-statistics are significant at a level of 5%.

Source: MSA COTNS database, Bretagne 2004-2014 - authors' calculations

size in type 2. In this latter case, the impact of neighbouring farms' size is almost 150 times higher and more than 350 times higher than it is for type 1 and type 3, respectively. This may reveal that farms belonging to types 1 and 3 have lower pecuniary motives than those in type 2, or that they may have already reached their optimal size implying that they are less sensitive to the evolution of neighbouring farms' size.

Descriptive statistics for each unobserved type are presented in Table 4.7. The figures show that farms in type 2 are on average much larger and more likely to be with a corporate legal form in comparison to types 1 and 3. Farms in type 2 are thus more likely to increase their operated size over time since they may have higher financial capacity. Conversely, farms belonging to type 3 are on average smaller and with older farm holders than those in the two other types, and are operated predominantly under individual legal status, with on average younger farmers as neighbours. This could explain why type 3 farms are more likely to decrease their operated size over time. As discussed in the previous section, older farmers may be less motivated to increase their farm size; in addition, they may face higher competition for land when surrounded by younger farmers. Furthermore, farms operated under individual legal status may face higher financial credit constraints.

4.7 Concluding remarks

The study conducted in this chapter underlines the importance of accounting for unobserved farm heterogeneity in spatial interdependence between farms when analysing farm structural change. This was made possible by a modelling approach that enables endogenously grouping farms within specific homogeneous types. This approach allows identifying specific relationships between farms via the impact of neighbouring farms' size, measured at the municipality level, on

farm survival and growth. The application to a panel of French farms located in Brittany shows that the relationship between farms in this region is rather in terms of competition for land than in terms of positive spill overs of new technology adoption. This results in a negative impact of neighbouring farms' size on the probability to survive for a majority of farms. However, for a about 18% of the farm population, the neighbouring farms' size has no significant impact on the probability to survive, suggesting the existence of potential non-pecuniary motives for these farms.

In contrast to the probability to survive, neighbouring farms' size is positively related to farm growth. Indeed, while three unobserved farm types are also endogenously identified in the case of farm growth, a positive effect of neighbouring farms' size is evidenced for all the three types of farms. This suggests that, even though neighbouring farms' size does not affect the probability of some farms to remain in business from year to year, farms in general tend to adjust their operated size over time. However, heterogeneity in the growth process is evidenced since the impact of neighbouring farms' size varies in magnitude according to farm types. Thus, this leads to identify three types of farms that are different from the types obtained from the survival model. Descriptive statistics show that the resulting farm types from the growth model are more related to some observed characteristics than the farms types from the survival model. These results confirm that neighbouring farms' size may differently influence farm survival and farm growth despite the fact that this two processes are related to each other. This also suggests that farms should not be considered as isolated entities and that agricultural policies should take into account potential relationships between farms.

While this study clearly adds to the existing literature, the analysis could be improved in three different ways. Firstly, the impact of neighbouring farms' size is investigated here by using the average farm size at the municipality as a proxy. As farms may compete for land in other municipalities in addition to their own municipality, investigating the impact of neighbouring farms' size using a spatial weighting matrix constructed at the farm level (using appropriate data sources that include the exact location of farms), could help estimate more efficiently the impact of neighbouring farms' characteristics. Secondly, the two models used for the present analysis (farm survival and farm growth) are estimated separately. However, it is clear that the survival and growth processes in farming are related to each other. Not accounting for this link may bias the estimation results. In this case, a sample selection model or a two-part model could be a way of improvement. Thirdly, some other factors, such as subsidies received by the farms and their neighbours, may have a significant impact on farm survival and growth as it has been shown by previous studies (see [Storm *et al.* \(2015\)](#) for a recent example). Including such variables in the analysis may thus improve the understanding of structural change in farming.

Bibliography

- Andrews, R. L. and Currim, I. S. (2003). Retention of latent segments in regression-based marketing models. *International Journal of Research in Marketing* 20: 315–321.
- Aubert, M. and Perrier-Cornet, P. (2009). Is there a future for small farms in developed countries? evidence from the French case. *Agricultural Economics* 40: 797–806.
- Bakucs, Z., Bojnec, S., Fertő, I. and Latruffe, L. (2013). Farm size and growth in field crop and dairy farms in France, Hungary and Slovenia. *Spanish Journal of Agricultural Research* 11: 869–881.
- Bollman, R. D., Whitener, L. A. and Tung, F. L. (1995). Trends and patterns of agricultural structural change: A Canada–US comparison. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 43: 15–28.
- Bowman, M. S. and Zilberman, D. (2013). Economic factors affecting diversified farming systems. *Ecology and Society* 18: 33.
- Breustedt, G. and Glauben, T. (2007). Driving forces behind exiting from farming in Western Europe. *Journal of Agricultural Economics* 58: 115–127.
- Case, A. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics* 22: 491–508.
- Conradt, S., Bokusheva, R., Finger, R., Kussaiynov, T. *et al.* (2014). Yield trend estimation in the presence of farm heterogeneity and non-linear technological change. *Quarterly Journal of International Agriculture* 53: 121–140.
- Daniels, T. L. (1986). Hobby farming in America: Rural development or threat to commercial agriculture? *Journal of Rural Studies* 2: 31–40.
- Dong, F., Hennessy, D. A. and Jensen, H. H. (2010). Contract and exit decisions in finisher hog production. *American Journal of Agricultural Economics* 92: 667–684.
- Eastwood, R., Lipton, M. and Newell, A. (2010). Farm size. *Handbook of agricultural economics* 4: 3323–3397.
- Gaté, R. and Latruffe, L. (2016). Difficultés rencontrées lors de la transmission d'exploitations agricoles. le cas de la Bretagne. *Économie rurale* 1: 5–24.
- Goddard, E., Weersink, A., Chen, K. and Turvey, C. G. (1993). Economics of Structural Change in Agriculture. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 41: 475–489.

- Greene, W. (2004). Convenient estimators for the panel probit model: Further results. *Empirical Economics* 29: 21–47.
- Greene, W. H. (2012). *Econometric analysis (International edition)*. Pearson, 7th ed.
- Hallam, A. (1991). Economies of size and scale in agriculture: An interpretive review of empirical measurement. *Review of Agricultural Economics* 13: 155–172.
- Happe, K., Kellermann, K. and Balmann, A. (2006). Agent-based analysis of agricultural policies: an illustration of the agricultural policy simulator agripolis, its adaptation and behavior. *Ecology and Society* 11: 49.
- Harrington, D. H. and Reinsel, R. D. (1995). A synthesis of forces driving structural change. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 43: 3–14.
- Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, 475–492.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47: 153–161.
- Hess, S., Bierlaire, M. and Polak, J. (2006). A systematic comparison of continuous and discrete mixture models. Tech. rep.
- Holloway, G., Lacombe, D. and LeSage, J. P. (2007). Spatial econometric issues for bio-economic and land-use modelling. *Journal of Agricultural Economics* 58: 549–588.
- Holloway, G., Shankar, B. and Rahmanb, S. (2002). Bayesian spatial probit estimation: a primer and an application to hyv rice adoption. *Agricultural Economics* 27: 383–402.
- Holloway, L. (2002). Smallholding, hobby-farming, and commercial farming: Ethical identities and the production of farming spaces. *Environment and Planning A* 34: 2055–2070.
- Howley, P. (2015). The Happy Farmer: The Effect of Nonpecuniary Benefits on Behavior. *American Journal of Agricultural Economics* 97: 1072–1086.
- Howley, P., Dillon, E. and Hennessy, T. (2014). It's not all about the money: understanding farmers' labor allocation choices. *Agriculture and Human Values* 31: 261–271.
- Jackson-Smith, D. B. (1999). Understanding the Microdynamics of Farm Structural Change: Entry, Exit, and Restructuring among Wisconsin Family Farmers in the 1980s. *Rural Sociology* 64: 66–91.
- Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica* 65: 1335–1364.

- Latruffe, L. and Piet, L. (2014). Does land fragmentation affect farm performance? a case study from brittany, france. *Agricultural Systems* 129: 68–80.
- Lennox, C. S., Francis, J. R. and Wang, Z. (2012). Selection models in accounting research. *The Accounting Review* 87: 589–616.
- Leung, S. F. and Yu, S. (1996). On the choice between sample selection and two-part models. *Journal of Econometrics* 72: 197 – 229.
- Little, R. J. A. (1985). A note about models for selectivity bias. *Econometrica* 53: 1469–1474.
- Maybery, D., Crase, L. and Gullifer, C. (2005). Categorising farming values as economic, conservation and lifestyle. *Journal of Economic Psychology* 26: 59 – 72.
- McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*, 382. John Wiley & Sons.
- McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.
- Mzoughi, N. (2011). Farmers adoption of integrated crop protection and organic farming: Do moral and social concerns matter? *Ecological Economics* 70: 1536 – 1545.
- Paternoster, R., Brame, R., Mazerolle, P. and Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology* 36: 859–866.
- Pennings, J. M. and Garcia, P. (2004). Hedging behavior in small and medium-sized enterprises: The role of unobserved heterogeneity. *Journal of Banking & Finance* 28: 951–978.
- Piet, L. and Cariou, S. (2014). Le morcellement des exploitations agricoles françaises. *Économie rurale* : 107–120.
- Puhani, P. (2000). The heckman correction for sample selection and its critique. *Journal of Economic Surveys* 14: 53–68.
- Shapiro, D., Bollman, R. D. and Ehrensaft, P. (1987). Farm size and growth in canada. *American Journal of Agricultural Economics* 69: 477–483.
- Storm, H. and Heckelei, T. (2016). Direct and indirect spatial interaction of Norwegian farm growth. Selected paper presented at the AAEE annual meeting. July 31 - August 2, Boston (USA).
- Storm, H., Mittenzwei, K. and Heckelei, T. (2015). Direct Payments, Spatial Competition, and Farm Survival in Norway. *American Journal of Agricultural Economics* 97: 1192–1205.
- Sumner, D. A. and Leiby, J. D. (1987). An econometric analysis of the effects of human capital on size and growth among dairy farms. *American Journal of Agricultural Economics* 69: 465–470.

- Teillard, F., Allaire, G., Cahuzac, E., Léger, F., Maigné, E. and Tichit, M. (2012). A novel method for mapping agricultural intensity reveals its spatial aggregation: Implications for conservation policies. *Agriculture, Ecosystems & Environment* 149: 135 – 143.
- Train, K. E. (2008). Em algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling* 1: 40–69.
- Trujillo-Barrera, A., Pennings, J. M. E. and Hofenk, D. (2016). Understanding producers' motives for adopting sustainable practices: the role of expected rewards, risk perception and risk tolerance. *European Review of Agricultural Economics* 43: 359 –382.
- Weiss, C. R. (1999). Farm Growth and Survival: Econometric Evidence for Individual Farms in Upper Austria. *American Journal of Agricultural Economics* 81: 103–116.
- Willock, J., Deary, I. J., Edwards-Jones, G., Gibson, G. J., McGregor, M. J., Sutherland, A., Dent, J. B., Morgan, O. and Grieve, R. (1999). The Role of Attitudes and Objectives in Farmer Decision Making: Business and Environmentally-Oriented Behaviour in Scotland. *Journal of Agricultural Economics* 50: 286–303.

Chapter 5

General conclusion

5.1 Overview and main findings

Structural change in farming has long been the subject of considerable interest among agricultural economists, policy-makers and other stakeholders because the evolution of farm structures may have consequences for the viability of the sector and for other related sectors of the economy (Weiss, 1999; Breustedt and Glauben, 2007). A large body of the agricultural economics literature is thus concerned with the analysis of structural change in various farming contexts. Studies aim in particular at understanding the mechanisms underlying the changes in farm structures in order to identify key drivers of the observed trends and to generate prospective scenarios. Structural change is however a complex concept and remains difficult to analyse (Stanton, 1991). Finding a global accepted definition for farm structural change is difficult because of the multidimensional characteristics of the processes at hand. This has led to the development and application of a variety of modelling approaches in the literature. Among others, the Markov chain framework is particularly suitable to model farm size change over time, and thus has been widely used to study structural change in various farming context (Zimmermann *et al.*, 2009). Despite the efforts devoted to this type of analysis, several issues remain that need to be alleviated to better represent, and hence understand, this process.

The overall aim of the thesis was to contribute to the improvement of the analysis and of the understanding of farm structural change, with an application to the French farming sector, given the importance of such changes over the last decades in this country. The specific objectives were threefold. Firstly, the thesis intended to empirically show that controlling for both observed and unobserved farm heterogeneity is important to represent more efficiently the ongoing process of structural change in farming. Secondly, the thesis sought to investigate the main drivers of the observed trends in the French farming sector over the last decades, accounting for both kinds of

heterogeneity. Lastly, the thesis aimed at analysing the interdependence between neighbouring farms, also under the assumption of a heterogeneous farm population.

As regards the first objective, while a lot of work in agricultural economics has been devoted to improve, on the one hand, the modelling approaches used for the analysis of structural change in farming and, on the other hand, the estimation techniques, less attention has been paid so far to take into account farm heterogeneity. However, heterogeneity is obviously present in farming because farmers' objectives as well as their social and economic environment are different and heterogeneous. As structural change in farming has its origins in individual farmers' decisions (Bollman *et al.*, 1995), studies which assume homogeneity may be misleading. Despite some attempts to take farm heterogeneity into account, the approaches used so far remain restrictive since they have only considered observed heterogeneity, particularly those applying Markov chain models (MCMs) to study farm structural change (*e.g.*, Zimmermann and Heckelei (2012)). The first research question addressed in this thesis therefore concerns the importance of controlling for possible unobserved heterogeneity in the transition process of farms. To answer this question, a mixture modelling approach was applied in order to endogenously divide the farm population into different homogeneous types, each type being characterised by a specific transition process.

To first test the usefulness of such a modelling approach, the so-called mover-stayer model (MSM) was applied in chapter 2 of the thesis. The MSM consists in a restricted version of the more general mixed-Markov chain model (M-MCM) where homogeneous types are first exogenously defined and set to two, namely the 'movers' and the 'stayers'. For the first time in the agricultural economics context, an empirical application of this model was performed using a sample of commercial farms from the French 'Réseau d'Information Comptable Agricole' (RICA) surveyed from 2000 to 2013.

To demonstrate the usefulness of taking unobserved heterogeneity into account, both the MSM and the MCM were applied to the same farm sample and both models' respective performance was compared. To simplify, the models' comparison was performed under stationary assumption. The results showed that, with respect to the MCM and even when using the simple MSM framework, accounting for unobserved heterogeneity in the transition process of farms enables to better estimate the observed transition probability matrix as well as to more accurately derive the distribution of farms across size categories in the long-run, without degrading any short-term analysis. In particular, assuming homogeneity leads to overestimate farm mobility since the stayer part of the farm population is not explicitly accounted for. This result is consistent with the findings in other strands of the economic literature, namely that, by relaxing the assumption of homogeneity, the MSM leads to a better representation of the underlying transition process. For instance, such results have been found by Blumen *et al.* (1955) when studying inter-industry mobility of workers; by Shorrocks (1976) and Major (2008)

when investigating income dynamics or; by [Cipollini *et al.* \(2012\)](#) when analysing firm dynamics. Robustness checks showed that these conclusions remain valid when using different limits for the definition of size intervals and/or using different numbers of size categories. Chapter 2 results thus showed that accounting for unobserved farm heterogeneity, even roughly, is crucial in the analysis of agricultural structural change.

With reference to the second objective of the thesis, chapter 3 of the thesis extended chapter 2 in two directions. Firstly, farm types were endogenously estimated, relaxing the predefined fixed number of farm types and ‘pure stayers’ assumptions. Secondly, transition probabilities were allowed to vary according to some time-varying variables, including agricultural policy. Many theoretical studies in the literature pointed out factors that may play an important role in driving farm structural change (*e.g.*, [Hallam \(1991\)](#); [Goddard *et al.* \(1993\)](#); [Harrington and Reinsel \(1995\)](#)). Since then, several empirical studies have been dedicated to identify the main drivers of structural change in different farming contexts or to investigate the impact of specific variables on this process (see [Zimmermann *et al.* \(2009\)](#) for a review). Focusing on agricultural policies, theories however provide ambiguous predictions as regards the impact of public support programmes ([Harrington and Reinsel, 1995](#); [Roberts and Key, 2008](#)). Furthermore, empirical investigations sometimes found opposite effects of agricultural policy measures on farm size change. For example, [Huettel and Jongeneel \(2011\)](#) found that EU milk quotas tended to increase farm growth while [Zimmermann and Heckelei \(2012\)](#); [Ben Arfa *et al.* \(2015\)](#) found an opposite effect. The second research question addressed in chapter 3 of the thesis was thus whether or not accounting for unobserved heterogeneity in the transition process of farms could help identifying specific impacts of the main drivers of farm structural change, especially focussing on agricultural policies. To tackle this question, a non-stationary version of the M-MCM was developed, where transition probabilities were specified as a multinomial logit function of a set of potential explanatory variables. Transition probability elasticities as well as structure elasticities were derived in order to analyse the impacts of the chosen exogenous variables on farm size change, focusing on public support provided to farmers in the frame of the European Union (EU) Common Agricultural Policy (CAP). The RICA data from 2000 to 2013 were again used for an empirical investigation in the French farming context.

The proposed non-stationary M-MCM allowed identifying two types of farms, each of which being characterised by a specific transition process. The transition probability matrix of the first recovered type shows that these farms may be considered as ‘almost stayer’ farms: their transition matrix is highly diagonal and thus their probability of changing from size category over time is very small. Farms belonging to the second type are more likely to change their size category over two consecutive years and can therefore be identified as ‘likely movers’. This suggests that the MSM applied in chapter 2 is an appropriate approach to efficiently represent transition process in farming and is thus a good choice especially when econometric applications

are limited, with regards to its simplicity.

The resulting estimated parameters of the transition probabilities showed that the impacts of the considered drivers depend significantly on the unobserved type of farms. Focusing on agricultural policy, the thesis also found a significant difference in the impact of farm subsidies from Pillar One of the CAP (production subsidies) compared to subsidies from Pillar Two (rural development subsidies), for both farm types. In particular, results indicated that farm growth is more likely to be influenced by subsidies from Pillar One. Moreover, aggregated at the farm population level, structure elasticities showed that the results from the mixture model are different from those obtained under the assumption of homogeneity. Overall, the homogeneous model tends to overestimate the impacts of subsidies, both from Pillar One and Pillar Two, on farm size change. Such results confirm that ignoring the potential unobserved heterogeneity among farmers as regards structural change may lead to incorrectly assessing the impact of driving factors of this process, which can be critical when drawing policy recommendations is at stake.

Chapter 4 extends chapters 2 and 3 in two by identifying specific relationships between neighbouring farms both in their probability to survive and to increase operated size. Concerning the third objective of the thesis, understanding interdependence between farms may give a better insight into the ongoing structural change in farming, and is thus crucial for an assessment of agricultural policy measures (Storm *et al.*, 2015). However, investigating interdependence between farms is challenging since different types of interactions between farms may co-exist. In particular, neighbouring farms can be viewed as competitors given land market regulations (Weiss, 1999) or as a source of motivation for new technology adoption (Case, 1992; Holloway *et al.*, 2002). However, farms may also be indifferent to their neighbours' behaviours as their activity may be governed by non-pecuniary motives (Daniels, 1986; Willock *et al.*, 1999; Holloway, 2002; Mzoughi, 2011). While Markovian approaches such as these used in chapters 2 and 3 enable mirroring farm interdependency by considering at the same time several dimensions of structural change, it is unable to identify specific types of interactions between farms. The third research question addressed in chapter 4 of the thesis was therefore to examine whether or not different types of interactions between farms exist in a specific farming context, and if accounting for unobserved heterogeneity could help identifying these specific types of farm relationships. The presence of unobserved heterogeneous behaviours was investigated both in the farm survival and farm growth processes. To do so, a probit model for farm survival, and a standard linear regression model for farm growth incorporating spatial effects, were developed within a mixture modelling framework in order to evidence potential heterogeneous relationships between farms. Furthermore, the models were designed in such a way that the panel dimension of the data could be taken into account in order to capture dynamics effects. A panel of French farms located in the Western region of Brittany and surveyed by the 'Mutualité Sociale Agricole' (MSA) from

2003 to 2013 was used for the empirical investigation.

Chapter 4 results showed that, as mentioned above, the three types of relationships between farms can coexist when the probability of surviving two consecutive years is considered. However, the relationship between farms rather pertains to a competition for land than to positive spill overs of new technology adoption. This leads to a negative impact of neighbouring farms' size on the probability to survive for a majority of farms. For another group of farms, the relationship between neighbours is rather positive suggesting positive spill overs of new technology adoption. Finally, a last group of farms gathers those that are not affected by neighbouring farms' size. As previously argued, farms belonging to this group may be characterised by weaker financial motives than those in the two other groups. In contrast to the probability to survive, the size of neighbouring farms is positively related to the growth of the farm considered. Indeed, while three farm groups were also identified when the growth process was considered, a positive effect of neighbouring farms' size was evidenced for all these groups. This means that even though neighbouring farms' size does not affect the probability of certain farms to remain active two consecutive years, farms in general tend to adjust their operated size over time. However, the impact of neighbouring farms' size varies in magnitude across the three groups, which also gives evidence of heterogeneity in the growth process of farms. These results confirmed that neighbouring farms' size differently influences farm survival and growth, and that different types of interactions between farms coexist in farming. This suggests that farms should not be considered as isolated entities and that policy measures should take into account such relationships between farms.

5.2 Recommendations

Some methodological and policy recommendations may be put forward from the results obtained in this thesis. From a methodological point of view, the following main recommendations could be drawn:

- Future analyses of structural change in farming should pay more attention to potential unobserved heterogeneity in the transition process of farms. Ignoring such heterogeneity in farmers' behaviours may limit the validity of the analysis. In particular, ignoring heterogeneity in the transition process of farms would lead to overestimate farms' mobility as it has been shown in other strands of economics (*e.g.*, [Blumen *et al.* \(1955\)](#); [McFarland \(1970\)](#); [Major \(2008\)](#)). From a statistical point of view, this would lead to inaccurate predictions and projections of future farm distribution across size categories.
- The impact of causal factors of structural change should be investigated while taking into account the heterogeneity in farmers' behaviours. While former studies generally

estimated a mean effect, this thesis showed that some factors driving structural change may have opposite impacts depending on unobserved farm types, which may result in an insignificant impact when aggregated at the population level. Accounting for heterogeneity in farmers' behaviours, especially by using a mixture modelling approach, thus enables assessing more efficiently the impact of factors affecting structural change in farming than it would be done under the homogeneous assumption; indeed, accounting for heterogeneity allows identifying specific impacts on specific homogeneous farm types that coexist in the population.

- Analysis of structural change in farming should investigate interdependence between farms and take into account possible heterogeneity in the relationships between farms. While it is clear that the work done on this matter by Storm et al. (2015) is very important in the analysis of structural change in farming, the mean effect approach adopted by the authors could be viewed as restrictive since farms may behave differently given the characteristics of their neighbours. As underlined in this thesis, identifying different types of relationships between farms is important to understand the trend of farm structural change in some regions and can help assessing more efficiently the impact of causal factors of the process, among which agricultural policies.
- As a consequence of the previous recommendations, more efforts should be devoted to the availability of individual level data for the analysis of structural change in farming. Regarding Markovian-like analysis, as long as the transition-process of farms may be considered as unique (i.e., farms are homogeneous in the way they change their category of size), estimating the resulting transitions using aggregate data is consistent as shown by Lee *et al.* (1965), Lee *et al.* (1970) and MacRae (1977). But, if more than one transition processes coexist, only individual level data can enable capturing such kind of heterogeneity efficiently, to the best of our knowledge. Individual level data are thus more convenient for the analysis of structural change in farming context characterised by potential heterogeneity in farmers' behaviours. Such type of data should therefore be used more extensively, and some effort should be provided by researchers and other stakeholders in agriculture to generate and give access to such data.

Recommendations to policy-makers and other stakeholders in the farming sector may be threefold:

- Firstly, support programmes to farmers should take into account the fact that some farms are more likely to move across categories of sizes (the so-called 'likely movers') than others (the so-called 'likely stayers'). If policy or support programmes aim at slowing down structural change in farming either to promote family farming or to keep the multifunctional

nature of agriculture by restricting concentration and farm specialisation for example, they should be drawn in such a way to encourage ‘likely stayers’ behaviours because mobility in farming is much more related to ‘likely movers’ behaviours. In the case where such different farms’ behaviours coexist, the efficiency of policy or support programmes would depend on which type dominates in a specific farming context.

- Secondly, priority should be given to support from Pillar Two of the CAP to the expense of Pillar One if agricultural policy measures aim at slowing down the ongoing process of structural change in farming in the EU. Indeed, while the budget for the direct payments of Pillar One is the most important element of the CAP expenditures in the EU ([Matthews, 2011](#)), this thesis showed that Pillar One coupled and decoupled subsidies to farms are more likely to encourage farm growth than supports awarded through Pillar Two, which is a common feature in agricultural economics. Likewise, [Piet et al. \(2012\)](#) evidenced that subsidies from Pillar Two contribute to reduce size inequality across farms more than Pillar One. Some modifications operated through the CAP reform in 2013 are in line with this recommendation. Among others, some new regulations introduced to develop a more sustainable EU agriculture, such as green direct payments to improve farms’ environmental performance and redistributive payments to increase the support to small and medium-sized farms ([European Union, 2013](#)), could serve as barriers to the ongoing farm growth process in the EU.
- Lastly, policy or support programmes to farmers should take into account interdependence between farms and different types of interactions that could coexist in specific farming contexts. Depending on the type of relationships that dominates in a context (namely competition for land vs. positive spill overs of new technology adoption), the expected impact of support programmes could be either under- or over-estimated. For example, a new technology could spread more rapidly in a farming sector where relationships between farms are less affected by competition for fixed inputs such as land. Even if new technology adoption is supported by public programmes, farmers who face high level of competition for land would not be able to benefit from the new technology. In the case where farmers have some non-pecuniary motives for farming, the technology would be rejected if it does not meet their expectations or is contrary to their interests as argued [Mzoughi \(2011\)](#).

5.3 Discussion and directions for further research

This thesis contributes in several ways to research in agricultural economics regarding the analysis of structural change in farming. However, additional work would be required in order to strengthen and extend the contributions. Some shortcomings and suggestions have been

mentioned in due places in the specific chapters of the thesis. In this section, the focus is more general, and possible avenues for further research are proposed.

Firstly, the main contribution of the thesis is the empirical evidence that accounting for unobserved heterogeneity in modelling structural change in the farming context is crucial. However, there is, as yet, no formal framework in agricultural economics that could help identifying the potential sources of this heterogeneity and guide the empirical measurement of how they may affect farms' transition process. The findings of this thesis stress the need for developing such formal studies. Theoretical studies could in particular give ground to why different transition processes may exist in farming and provide guidance on how they could be identified. Within the mixture modelling approach, such theoretical background could then be very helpful in specifying farm-type membership by setting intuitions on the main determinants of farm types to consider and their a priori potential effects.

Secondly, in this thesis, a mixture modelling approach was applied to capture unobserved heterogeneity both in the transition process of farms and in the relationships between farms. Nonetheless, there exist several other modelling frameworks which could be applied to capture or control for unobserved heterogeneity. Especially in the case of panel data, modelling strategies such as fixed and random effect approaches, or individual and random parameter approaches, are alternatives to the mixture modelling approach (Greene, 2012). Despite the usefulness of the mixture approach, its stability and robustness should be compared to such modelling alternatives. Another possible way of improvement is the combination of a fixed or random effect model with the mixture modelling approach, following the line developed by, for example, Bago d'Uva (2005) or Deb and Trivedi (2013). This could enable verifying whether or not resulting farm types are totally homogeneous, since the choice of the optimal number of farm types is mainly based on a trade-off between the number of parameters to be estimated and the total number of farms.

Thirdly, in the thesis it was assumed that the probability of farms to belong to a specific unobserved type is fixed over time. In other words, it was supposed that farms do not change behaviour during their entire lifespan. However, farms' behaviour may vary over time because of changes in the farm's situation or farmers' motivations. The mixture modelling approach as applied in the thesis could thus be extended to enable specific segments in the population to grow or decline over time. As regard the mixed-Markov chain model (M-MCM), one could use the latent Markov chain model first proposed by Poulsen (1990) and extended, on the one hand, by Van de Pol and Langeheine (1990) as the mixed Markov latent class model and, on the other hand, by Vermunt (2010) as the mixture of latent Markov chain models. Another approach could consist in specifying time-varying probabilities for a farm to belong to a specific type. In this case, a parametric form for the mixing distribution (*i.e.*, the posterior probability to belong to a specific farm type) could be employed as in Provencher *et al.* (2002), but using time-varying

explanatory variables in the type membership specification as in [Morduch and Stern \(1997\)](#). In the latter case, special attention has to be paid to the specification method as well as to the choice of the explanatory variables since both may influence the results ([Heckman and Singer, 1984](#); [Lindsay and Lesperance, 1995](#); [Wedel, 2002](#)). In this respect, the fractional multinomial logit developed in [Mullahy \(2015\)](#) could be a more appropriate specification method than the standard logit generally used since it allows accounting for outcomes being probabilities and can sometimes actually take extreme values 0 and 1 ([Papke and Wooldridge, 1993, 2008](#)).

Fourthly, to estimate the M-MCM, it is assumed that farms are independent conditional on their specific unobserved type, since the parameters of the model were estimated using the maximum likelihood estimation techniques. We may infer that this assumption does not affect the results in chapter 2 and 3 of the thesis because of the sampling method used in the ‘Réseau d’Information Comptable Agricole’ (RICA). Indeed, the RICA database is a stratified sample where sampled farms may be located far from each other. Because of that, it is unlikely that farms considered in the data used for the empirical applications actually compete for plots as neighbours. The independence assumption thus holds in this case. However, for validation of the modelling approach irrespective from the sample used, one potential improvement would therefore consist in specifying the M-MCM in such a way that one could account for interactions between neighbouring farms. A possible improvement could be, building on the experience of spatial econometrics, to directly incorporate the characteristics of neighbours in the specification of farm transition probabilities.

Fifthly, as is common in agricultural economics, this thesis dealt with structural change in farming by focusing only on the evolution of the total farm number and the average farm size. However, as argued in the general introduction of the thesis, structural change is a multidimensional concept and the indicators that depict this process are numerous. Structural change in farming can be therefore analysed through various other aspects such as the evolution of production systems, legal status, nature of labour force or capital invested. The proposed modelling approach in this thesis can be also applied to test the difference in farms’ behaviours through these characteristics. This could provide additional and insightful understanding of structural change in farming. Furthermore, in chapters 2 and 3, a continuous variable (standard output) is discretised to define Markov states. Robustness checks showed that the results slightly depend on the bound used for the category of sizes. To avoid such a limit, the modelling approach could be developed within the continuous parameterisation technique as proposed by [Piet \(2011\)](#).

Sixthly, as mentioned in the introductory chapter of the thesis, farm structural change may have important consequences for farming activities and for other economics sectors as well, given their interconnection ([Jones *et al.*, 1995](#)). The thesis dealt with spatial interdependence between farms, but only partially, as it left unexplored the interactions with the upstream and downstream sectors. The latter analysis could be complicated within a Markov chain or other

usual econometric models in agricultural economics since it would imply to consider multiple interactions between farms and other types of agents. In this case, agent based models (ABM) could be a good alternative. ABM could do a better job since such a modelling approach enables incorporating multiple interactions between different types of agents evolving in the same economic environment (Kremmydas, 2012; Bargigli and Tedeschi, 2013). ABM could provide thus more insightful analysis of farm structural change by integrating different types of agents who directly or indirectly participate to this process. However, to do this, more efforts should be devoted to make available the resources needed for the application of such a modelling framework. As mentioned in the introductory chapter 1, one of the main disadvantages of ABM is that this modelling approach requires much more data than usual econometric models such as the MCM (Zimmermann *et al.*, 2009).

Lastly, this study faced important data limitations. Data used from both RICA and ‘Mutualité Sociale Agricole’ (MSA) for the empirical investigations are limited in the information that they provide about farms. While RICA provides a large set of variables on the physical and structural characteristics of farms and their economic and financial features, individual-data MCM applications using this database remain limited. Indeed, RICA being a rotating panel, farms that enter, respectively leave, the sample in a given year cannot be considered with certainty as actual entries into, respectively exits from, the agricultural sector. This forces to work in chapter 2 and 3 only on movements of farms across size categories considering no entry in nor exit from farming, which is a very restrictive assumption. Furthermore, RICA being a stratified sample, the weighting factors within the database can be used to infer results at the population level when the database is used in its cross-sectional dimension. However, when this database is used as a panel, as it was the case in this thesis, these weighting factors cannot be used because they are not relevant in the inter-temporal dimension. Conversely, the MSA can be actually considered as an almost exhaustive database since it collects information about all individuals who declare a non-salaried farming activity in France and on their farms. However, contrary to the RICA, this database contains only few variables related to the physical and structural characteristics of farms as well as to their economic and financial features. Finally, both databases are restrictive in terms of information required to evidence interdependency between farms since the precise geographical location of the farmstead and plots are not recorded. Both sources should therefore be improved in some specific ways to enable more efficient analyses of farm structural change in France. Some efforts could also be devoted to generate datasets that combine the particular strengths of these two databases while improving their specific limits.

Bibliography

- Bago d’Uva, T. (2005). Latent class models for use of primary care: evidence from a British panel. *Health Economics* 14: 873–892.
- Bargigli, L. and Tedeschi, G. (2013). Major trends in agent-based economics. *Journal of Economic Interaction and Coordination* 8: 211–217.
- Ben Arfa, N., Daniel, K., Jacquet, F. and Karantininis, K. (2015). Agricultural Policies and Structural Change in French Dairy Farms: A Nonstationary Markov Model. *Canadian Journal of Agricultural Economics/Revue canadienne d’agroéconomie* 63: 19–42.
- Blumen, I., Kogan, M. and McCarthy, P. J. (1955). *The industrial mobility of labor as a probability process*. No. VI in Cornell Studies in Industrial and Labor Relations. Ithaca (New York): Cornell University.
- Bollman, R. D., Whitener, L. A. and Tung, F. L. (1995). Trends and patterns of agricultural structural change: A Canada–US comparison. *Canadian Journal of Agricultural Economics/Revue canadienne d’agroéconomie* 43: 15–28.
- Breustedt, G. and Glauben, T. (2007). Driving forces behind exiting from farming in Western Europe. *Journal of Agricultural Economics* 58: 115–127.
- Case, A. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics* 22: 491–508.
- Cipollini, F., Ferretti, C. and Ganugi, P. (2012). Firm size dynamics in an industrial district: The mover-stayer model in action. In Di Ciaccio, A., Coli, M. and Angulo Ibanez, J. M. (eds), *Advanced Statistical Methods for the Analysis of Large Data-Sets*, Studies in Theoretical and Applied Statistics. Springer Berlin Heidelberg, 443–452.
- Daniels, T. L. (1986). Hobby farming in America: Rural development or threat to commercial agriculture? *Journal of Rural Studies* 2: 31 – 40.
- Deb, P. and Trivedi, P. K. (2013). Finite mixture for panels with fixed effects. *Journal of Econometric Methods* 2: 35–51.
- European Union (2013). Overview of CAP reform 2014-2020. Agricultural Policy Perspectives Brief, N°5.
- Goddard, E., Weersink, A., Chen, K. and Turvey, C. G. (1993). Economics of structural change in agriculture. *Canadian Journal of Agricultural Economics/Revue canadienne d’agroéconomie* 41: 475–489.
- Greene, W. H. (2012). *Econometric analysis*. (International edition). Pearson, 6th ed.

- Hallam, A. (1991). Economies of Size and Scale in Agriculture: An Interpretive Review of Empirical Measurement. *Review of Agricultural Economics* 13: 155–172.
- Harrington, D. H. and Reinsel, R. D. (1995). A synthesis of forces driving structural change. *Canadian Journal of Agricultural Economics* 43: 3–14.
- Heckman, J. and Singer, B. (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* 52: 271–320.
- Holloway, G., Shankar, B. and Rahmanb, S. (2002). Bayesian spatial probit estimation: a primer and an application to HYV rice adoption. *Agricultural Economics* 27: 383–402.
- Holloway, L. (2002). Smallholding, Hobby-Farming, and Commercial Farming: Ethical Identities and the Production of Farming Spaces. *Environment and Planning A* 34: 2055–2070.
- Huettel, S. and Jongeneel, R. (2011). How has the EU milk quota affected patterns of herd-size change? *European Review of Agricultural Economics* 38: 497–527.
- Jones, W., Freshwater, D. and Fiarchuk, R. (1995). Why the structure of agriculture still matters to national governments. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 43: 223–240.
- Kremmydas, D. (2012). Agent based modeling for agricultural policy evaluation: A review. AUA Working Paper Series No.2012-3 .
- Lee, T., Judge, G. and Zellner, A. (1970). *Estimating the parameters of the Markov probability model from aggregate time series data*. Amsterdam: North Holland.
- Lee, T. C., Judge, G. G. and Takayama, T. (1965). On Estimating the Transition Probabilities of a Markov Process. *Journal of Farm Economics* 47: 742–762.
- Lindsay, B. G. and Lesperance, M. L. (1995). A review of semiparametric mixture models. *Journal of statistical planning and inference* 47: 29–39.
- MacRae, E. C. (1977). Estimation of time-varying Markov processes with aggregate data. *Econometrica* 45: 183–198.
- Major, K. (2008). Income disparities among Hungarian micro-regions: The mover-stayer model. *Acta Oeconomica* 58: 127–156.
- Matthews, A. (2011). Post-2013 eu common agricultural policy, trade and development. a review of legislative proposals. international centre for programme on agricultural trade and sustainable development (ictsd), issue paper no 39 .
- 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. *American Sociological Review* 35: 463–476.

- Morduch, J. J. and Stern, H. S. (1997). Using mixture models to detect sex bias in health outcomes in Bangladesh. *Journal of Econometrics* 77: 259–276.
- Mullahy, J. (2015). Multivariate fractional regression estimation of econometric share models. *Journal of Econometric Methods* 4: 71–100.
- Mzoughi, N. (2011). Farmers adoption of integrated crop protection and organic farming: Do moral and social concerns matter? *Ecological Economics* 70: 1536 – 1545.
- Papke, L. E. and Wooldridge, J. (1993). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates.
- Papke, L. E. and Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* 145: 121–133.
- 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).
- Piet, L., Latruffe, L., Le Mouel, C. and Desjeux, Y. (2012). How do agricultural policies influence farm size inequality? the example of France. *European Review of Agricultural Economics* 39: 5–28.
- Pol, F. Van de and Langeheine, R. (1990). Mixed Markov latent class models. *Sociological methodology* 20: 213–247.
- Poulsen, C. S. (1990). Mixed Markov and latent Markov modelling applied to brand choice behaviour. *International Journal of Research in Marketing* 7: 5–19.
- Provencher, B., Baerenklau, K. A. and Bishop, R. C. (2002). A finite mixture logit model of recreational angling with serially correlated random utility. *American Journal of Agricultural Economics* 84: 1066–1075.
- Roberts, M. J. and Key, N. (2008). Agricultural Payments and Land concentration: A Semiparametric Spatial Regression Analysis. *American Journal of Agricultural Economics* 90: 627–643.
- Shorrocks, A. F. (1976). Income Mobility and the Markov Assumption. *The Economic Journal* 86: 566–578.
- Stanton, B. F. (1991). *Farm Structure: Concept and Definition*. Department of Agricultural Economics, Cornell University Agricultural Experiment Station, New York State College of Agriculture and Life Sciences, Cornell University.
- Storm, H., Mittenzwei, K. and Heckelei, T. (2015). Direct Payments, Spatial Competition, and Farm Survival in Norway. *American Journal of Agricultural Economics* 97: 1192–1205.

- Vermunt, J. (2010). Longitudinal research using mixture models. In Montfort, K. van, Oud, J. H. and Satorra, A. (eds), *Longitudinal Research with Latent Variables*. Springer Berlin Heidelberg, 119–152.
- Wedel, M. (2002). Concomitant variables in finite mixture models. *Statistica Neerlandica* 56: 362–375.
- Weiss, C. R. (1999). Farm growth and survival: Econometric evidence for individual farms in Upper Austria. *American Journal of Agricultural Economics* 81: 103–116.
- Willock, J., Deary, I. J., Edwards-Jones, G., Gibson, G. J., McGregor, M. J., Sutherland, A., Dent, J. B., Morgan, O. and Grieve, R. (1999). The Role of Attitudes and Objectives in Farmer Decision Making: Business and Environmentally-Oriented Behaviour in Scotland. *Journal of Agricultural Economics* 50: 286–303.
- Zimmermann, A. and Heckelei, T. (2012). Structural change of European dairy farms. A cross-regional analysis. *Journal of Agricultural Economics* 63: 576–603.
- Zimmermann, A., Heckelei, T. and Domínguez, I. P. (2009). Modelling farm structural change for integrated ex-ante assessment: Review of methods and determinants. *Environmental Science and Policy* 12: 601–618.

RÉSUMÉ

Prise en compte de l'hétérogénéité inobservée des exploitations agricoles dans la modélisation du changement structurel : illustration dans le cas de la France.

Le changement structurel en agriculture suscite beaucoup d'intérêt de la part des économistes agricoles ainsi que des décideurs politiques. Pour prendre en compte l'hétérogénéité du comportement des agriculteurs, une approche par les modèles de mélange de chaînes de Markov est appliquée pour la première fois en économie agricole pour analyser ce processus. La performance de cette approche est d'abord testée en utilisant une forme simplifiée du modèle, puis sa forme générale est appliquée pour étudier l'impact de certaines mesures de politique agricole. Pour identifier les principaux canaux d'interdépendance entre exploitations voisines dans les processus du changement structurel, une approche de mélange non-Markovienne a été appliquée pour modéliser la survie et l'agrandissement des exploitations agricoles. Trois principales conclusions découlent de cette thèse. Tout d'abord, la prise en compte de l'hétérogénéité dans les processus de transition des exploitations agricoles permet de mieux représenter le changement structurel et conduit à des prédictions plus précises de la distribution des exploitations, comparé aux modèles généralement utilisés jusqu'ici. Deuxièmement, l'impact des principaux facteurs du changement structurel dépend lui aussi des types non-observables d'exploitations mis en évidence. Enfin, le cadre du modèle de mélange permet également de révéler différents types de relations inobservées entre exploitations voisines qui contribuent au changement structurel observé à un niveau global ou régional.

Mots clés : Politiques agricoles, algorithme Espérance-Maximisation (EM), exploitations agricoles, processus Markovien, modèles de mélange, interdépendance spatiale, changement structurel, hétérogénéité inobservée.

ABSTRACT

Accounting for unobserved farm heterogeneity in modelling structural change: evidence from France.

Structural change in farming has long been the subject of considerable interest among agricultural economists and policy makers. To account for heterogeneity in farmers' behaviours, a mixture Markov modelling framework is applied to analyse this process for the first time in agricultural economics. The performance of this approach is first investigated using a restrictive form of the model, and its general form is then applied to study the impact of some drivers of structural change, including agricultural policy measures. To identify channels through which interdependency between neighbouring farms arises in this process, the mixture modelling approach is applied to analyse both farm survival and farm growth. The main conclusions of this thesis are threefold. Firstly, accounting for the generally unobserved heterogeneity in the transition process of farms allows better representing structural change in farming and leads to more accurate predictions of farm-size distributions than the models usually used so far. Secondly, the impacts of the main drivers of structural change themselves depend on the specific unobservable farm types which are revealed by the model. Lastly, the mixture modelling approach enables identifying different unobserved relationships between neighbouring farms that contributes to the structural change observed at an aggregate or regional level.

Keywords: Agricultural policy; Expectation-Maximisation (EM) algorithm; farms; Markovian process; mixture models; spatial interdependence; structural change; unobserved heterogeneity



AGROCAMPUS OUEST • Institut supérieur des sciences agronomiques, agroalimentaires, horticoles et du paysage
65 rue de Saint-Brieuc – CS84215 – F-35042 Rennes Cedex
Tél. : 02 23 48 50 00
www.agrocampus-ouest.fr

