Institut für Lebensmittel- und Ressourcenökonomik der
Rheinischen Friedrich-Wilhelms-Universität Bonn

Methods of analysis and empirical evidence of farm structural change

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Grades

Doktor der Agrarwissenschaften

(Dr.agr.)

der

Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität

Bonn

vorgelegt am 08. April 2014

von

Hugo Storm

aus

Lahnstein

Referent:	Prof. Dr. Thomas Heckelei
Korreferent:	Prof. Dr. Robert Finger
Externer Gutachter:	Dr. Laurent Piet
Tag der mündlichen Prüfung:	15. August 2014
Erscheinungsjahr:	2014

Danksagung

Allen, die mich bei der Anfertigung dieser Arbeit auf unterschiedlichste Weise unterstützt haben, möchte ich an dieser Stelle Danke sagen! Meinem Doktorvater Prof. Dr. Thomas Heckelei für die Möglichkeit zur Promotion und der Freiheit, mich mit Dingen zu beschäftigen, die mich interessieren. Für die fachlich kompetente und lösungsorientierte Unterstützung und Förderung − gerade die Diskussionen über methodische Probleme waren immer eine besondere Bereicherung. Für die Flexibilität, die Promotion mit familiären Belangen und der Arbeit im elterlichen Betrieb kombinieren zu können. Vor allem aber für das immer faire und freundschaftliche Verhältnis, das besonders zur Motivation und Freude an der Arbeit beigetragen hat. All dieses ging weit über das Selbstverständliche hinaus.

Prof. Dr. Robert Finger danke ich für die Übernahme des Koreferats. Prof. Dr. Ron Mittelhammer für seine wertvollen Beiträge zu Kapitel zwei, für die Möglichkeit zu einem Aufenthalt an der WSU und die Erfahrung mit ihm ökonometrische Probleme zu diskutieren. Dr. Norbert Röder, Dr. Alexander Gocht, Sebastian Neuenfeldt, Dr. María Espinosa und Dr. Sergio Gomez y Paloma für die Hilfe bei der Datenbeschaffung, Diskussionen und Ideen für Kapitel drei. Dr. Klaus Mittenzwei für die Möglichkeit zu einem Aufenthalt am NILF, unzählige Ideen, Diskussionen und Hilfe für Kapitel vier und die produktive und freundschaftliche Zusammenarbeit.

Meinen ehemaligen und aktuellen Kollegen vom ILR für all die fachliche, technische und organisatorische Unterstützung, die freundschaftliche Atmosphäre und die schöne gemeinsame Zeit am Institut, in der Mensa, bei Betriebsausflügen, Institutsolympiaden, Feiern oder bei sonstigen Aktivitäten. Besonderer Dank gilt hier Dr. Andrea Zimmermann und das nicht nur als treuste Mensabegleiterin, sondern auch dafür, dass sie mir ihr Thema „vererbt“ hat, für die Einführung in die

Markow/Strukturwandel-Welt, unendlichen Diskussionen darüber und für unzählige beantwortete Fragen. Meinem ersten, zweiten und dritten Bürokollegen Dr. Christian Götz, Yinan Zhang und Sebastian Rasch für fachliche und weniger fachliche, aber immer nette und unterhaltsame Diskussionen.

Meinen Eltern möchte ich danken für den Spaß und das Interesse an der Landwirtschaft, das sie mir mitgegeben haben und die Möglichkeiten und Freiheiten meinen Interessen und Ideen nachzugehen. Ganz besonderes möchte ich Caroline für Ihre kontinuierliche Unterstützung, Motivation in schwierigen Phasen und das Mittragen all meiner Pläne, sowie meinem Sohn Tilbert für die nötige Ablenkung danken!

Kurzfassung

Ziel der Arbeit ist die Entwicklung und Anwendung von Methoden zur empirischen Analyse und Modellierung des Agrarstrukturwandels. Veränderungen der Agrarstruktur sind nicht allein für den Sektor bedeutend, sondern können weitreichende ökonomische, soziale und ökologische Konsequenzen für eine Region haben. Ein Verständnis des Strukturwandels ist somit wichtig für die Folgenabschätzung (agrar-)politischer Maßnahmen, sowie deren Gestaltung im Hinblick auf konkrete (agrar-)politische Ziele.

Ein häufig verwendeter methodischer Ansatz zur Untersuchung des Agrarstrukturwandels ist die Markowketten-Analyse. In dieser Arbeit wird ein Bayes'scher Schätzansatz entwickelt, der eine Kombination von einzelbetrieblichen und aggregierten Daten in der Schätzung von nicht-stationären Markowketten erlaubt. Die Datenkombination erfolgt auf eine, im Vergleich zu existierenden Ansätzen, konsistentere und transparentere Weise und es wird gezeigt, dass sie die Präzision sowie die numerische Stabilität des Schätzers erhöht. Darauf aufbauend wird ein Bayes'scher Ansatz zur Vorhersage des EU Strukturwandels entwickelt, der es erlaubt die verfügbaren Daten besser zu nutzen.

Darüber hinaus befasst sich die Arbeit mit Interdependenzen auf Betriebsebene und deren Bedeutung für den Strukturwandel. Es wird argumentiert, dass sich das Verhalten von Betrieben gegenseitig bedingt und die Annahme einer unabhängigen Entwicklung, wie sie der Markowketten-Analyse zugrundeliegt, zu Problemen führen kann. Es wird empirisch gezeigt, dass die Berücksichtigung von Interdependenzen zwischen Betrieben wichtig für eine konsistente Aggregation der Ergebnisse der Betriebsebene zur Politikfolgenabschätzung auf regionaler Ebene ist. Am Beispiel Norwegens wird gezeigt, dass zur Abschätzung der Effekte von Direktzahlungen die Charakteristika benachbarter Betriebe berücksichtigt werden müssen. Nach Wissen des Autors ist die Arbeit die erste, die empirisch die Bedeutung von Interdependenzen auf Betriebsebene für den Strukturwandel belegt. Mit Blick auf eine Politikfolgenabschätzung zeigen die Ergebnisse, dass Direktzahlungen, die ein Betrieb selbst erhält, einen positiven Einfluss auf das Überleben des Betriebs haben, während Direktzahlungen an benachbarte Betriebe einen negativen Einfluss haben. Zur Abschätzung des generellen Effekts von Direktzahlungen ist es somit notwendig, die Interdependenzen zwischen Betrieben zu berücksichtigen. Werden diese vernachlässigt, kann der Effekt von Direktzahlungen überschätzt werden.

***Schlüsselwörter**: Agrarstrukturwandel, Markowketten, Datenkombination, räumliche Abhängigkeit, Aggregation, Politikfolgenabschätzung*

Abstract

The dissertation aims to develop and apply new empirical methods to analyze and model farm structural change. Changes of the farm structure are not only important for the sector itself but may have broader economic, social and environmental consequences for a region. Understanding this process is important for assessing the impact of (agricultural-) policies.

A common approach to analyze farm structural change are Markov chains. The dissertation provides a Bayesian estimation framework that allows to more consistently and transparently combine individual and aggregated data in the estimation of non-stationary Markov models compared to existing methods. It is shown that the data combination improves precision and numerical stability of the estimation. Building on this, a Bayesian prediction framework for EU farm structural change is developed exploiting the available information more fully.

Secondly, farm interdependences and their importance for farm structural change are analyzed. It is argued that the assumption of independence between farm behavior as implied by the Markov approach may become problematic in specific applications. Empirical evidence is provided that these interactions are indeed important to consider for a consistent aggregation of farm level results when assessing policy effects at regional level. Specifically, it is shown for the case of Norway that it is important to consider neighboring farm characteristics when analyzing the influence of direct payments on farm survival. To the knowledge of the author, the study is the first to show empirically that spatial interdependence at farm level is important for farm structural change. With respect to policy assessment, the results indicate that direct payments a farm receives itself have a positive influence on farm survival while neighboring direct payments have a negative one. For an overall assessment of the policy effects it is thus necessary to consider the interdependencies between farms. Ignoring these interdependencies might lead to an overestimation of the effects of direct payments.

Keywords: *Farm structural change, Markov process, data combination, spatial dependence, aggregation, policy assessment*

Contents

List of tables

List of figures

Abbreviations

CAP	Common Agricultural Policy
ESU	European Size Units
EU	European Union
EU-15	The 15 Member States of the EU from 1995 to 2004
FADN	Farm Accountancy Data Network
FSS	Farm Structural Survey
GCE	Generalized Cross-Entropy
IIA	Independence of Irrelevant Alternatives
MASE	Mean Absolute Scaled Error
MH	Metropolis Hastings
PT	Parallel Tempering
SDEM	Spatial Durbin Error Model
SGM	Standard Gross Margin
SLX	Spatially Lagged Explanatory Variable Model
TP	Transition Probability

Chapter 1
Research Context[1]

1.1 Motivation and general research question

Individual farmers decide to change their specialization, the intensity level, the size of the farm or its organization in response to changes in their environment. The environment they respond to is complex and covers a wide range of issues such as personal, social, economic, natural or political factors (Boehlje 1992; Zimmermann et al. 2009). The decision of a single farmer is largely irrelevant for the overall sector. Collectively, however, the sum of the individual decisions transform the farm sector as a whole. This process is typically characterized by the term *farm structural change*.

The term is generally understood even by people with no or only a loose connection to agriculture. The drastic decline of farm numbers along with an

---

[1] The research presented in this dissertation is to the largest part supported by the project "*Development of a Bayesian estimator for non-stationary Markov transition probabilities and its application to EU farm structural change*" of the German Research Foundation (DFG), grant no. HE 2854/4-1. Chapter three has its origin in a joint research project "*Modelling the effects of the CAP on farm structural change*" (Contract 151949-2010-A08-DE) from the European Commission Joint Research Centre - Institute for Prospective Technological Studies (IPTS). Chapter four was primarily conducted during a research visit at the Norwegian Agricultural Economics Research Institute (NILF) in Oslo additionally funded by a travel grant from the German Academic Exchange Service (DAAD) financed by the Federal Ministry for Economic Cooperation and Development under project ID 56453885.

increase in average farm size that occurred in most developed countries in the last decades is the most common association that comes to mind. In the academic literature however, there is no universal definition of the term and plenty of authors provide alternative ones. By comparing different definitions Stanton (1993) identified (1) the farm businesses as a productive enterprise (2) the farm household and (3) the agriculture resources as three major elements of all definitions of structural change. Further he highlighted (1) “changing distributions within the sector”, (2) “production decisions and who makes them” and (3) “ownership of resources and control over their use (p.19)” as the three major issues farm structural change is concerned with. The term farm structural change thus covers a wide array of aspects. Nevertheless, in the most basic and widely used case, farm structural change analysis is indeed concerned with the size and number of farms in the population (Goddard et al. 1993:476). However, Stanton (1993) points out that “[n]o single frequency is adequate to describe farm structure. Hence, distributions of farms by sales class, land area, labor force, acres of key crops or numbers of livestock are all used in examining structure and change through time. What happens to these distributions remains a focus of public interest and debate (p. 19).”

But irrespective of the definition chosen and the aspect of farm structural change considered, its consequences are not only relevant for the sector itself but may have broader social, economic and environmental consequences for a region (Flaten 2002:436–438). Being able to understand and explain farm structural change at the aggregate level and the individual decisions that lead to it is thus crucial to assess how (agricultural-) policy affects this development.

A large body of literature is concerned with the analysis of farm structural change (see references in the following). The overall objective of this dissertation is to develop and apply methods to analyze and model farm structural change. The dissertation can be distinguished into two major parts. The first, consisting of chapter two and three, aims at improving the use of data information in the analysis of EU farm structural change by developing a Bayesian estimation approach for non-stationary Markov models combining farm level with aggregated data. Markov models are popular for the analysis of farm structural change and chapter two and three address a specific methodological gap indentified in the literature. In the second part, consisting of chapter 4, the assumption of independent farm behavior underlying the Markov approach is addressed. Specifically, the focus is on the importance of farm level spatial interaction between farms and their relevancy for policy assessment.

Even though the two parts stand on their own the reoccurring theme that is central throughout the dissertation is the distinction between strategic decisions at the individual farm level (e.g. size, specializations, survival decisions etc. by an individual farmer) and changes of the farm structure at the aggregated/regional level. What we are commonly interested in is how all individual decision together lead to transformations of the sector as a whole. For an understanding of farm structural change it is crucial, however, to clearly differentiate between the individual and aggregate level. In the first part the distinction plays a role at the data level. The unique feature of the Markov approach is that it allows deriving probabilities for individual farm level behavior from aggregated data. The estimation approach considered here provides the possibility to combine individual and aggregated data. In the second part the focus is on the interaction between farms on the individual level. These interactions are analyzed empirically and shown to be crucial for aggregating results from individual farm to regional level. Due to interaction between farms, the aggregated outcome is not simply the sum of individual decisions.

In the remainder of this introductory the contribution of the thesis to the literature is highlighted. Afterwards, a concluding section summarizes results and discusses limitations and further research potential.

1.2 **Contribution of the thesis**

In this section the three chapter of the dissertation are summarized. Additionally, the gaps in the literature addressed by each single chapter and the path of development from the first to the last chapter are highlighted.

1.2.1 *Bayesian Estimation of Non-Stationary Markov Models Combining Micro and Macro Data*

The analysis of farm structural change has a long tradition in agricultural economics with Cochrane (1958) as one of the earliest references. In the following years a large body of literature emerged considering multiple aspects of farm structural change using a large array of methodological approaches (see Zimmermann et al. 2009 for an extensive review). A popular approach is the Markov framework which allow analyzing the movement of individuals between predefined states over time (recent examples are: Huettelet et al. 2010, Huettel and Jongeneel 2011, Zimmermann and Heckelei 2012a, Zimmermann and Heckelei 2012b). For estimation, the latter three of the cited examples rely on the generalized cross-entropy (GCE) approach, proposed by Golan and Vogel (2000) and first applied in a Markov context by Karantininis (2002). The GCE approach allows including prior information in the estimation. In the Markov context, prior information is typically specified for the transition probabilities. Prior information can be based on previous studies and on external knowledge. The possibility to consider prior information is the strength as well as the major criticism of the GCE approach. The use of prior information allows estimating ill-defined systems but is often criticized to introduce subjective prior believes in estimation.

This criticism is addressed in a recent dissertation by Zimmermann (2012) who proposed to specify prior information in the GCE approach empirically based on additional data. For the EU, there are two types of data sources that provide information about farm structural change: the farm structural survey (FSS; Council Regulation (EC) No 1166/2008) and the Farm accountancy network (FADN; Council Regulation (EC) No 1217/2009). Each data source provides different information at different levels and temporal resolution. The FSS provides aggregated census data in which the total number of farms in each state is observed. Data obtained from FADN is an unbalanced panel in which the individual movement of farms between states can be identified for a sample of farms. In accordance to the literature we refer in the following to the aggregated and individual level data as “macro” and “micro” data, respectively. Zimmermann (2012) proposed combining the two data sources in a GCE estimation approach. The estimation is based on FSS macro data, while FADN micro data is used to specify the prior information on the transition probabilities avoiding an otherwise rather ad hoc specification.

Despite this contribution, however, several shortcomings of the GCE approach persist which are addressed in chapter two of this dissertation. One general shortcoming of the GCE approach is the rather in-transparent way prior information is specified and used in estimation. This makes it difficult for the researcher and the research community to assess the importance and influence of prior information on the final estimation results. Further, it is not possible to specify an ignorance (or non-informative) prior for cases where no prior information is available. For a detailed more technical discussion of the general limitations of the GCE approach we refer to Heckelei et al. (2008). An additional shortcoming of the approach proposed by Zimmermann (2012) is that it ignores the precision of prior information in the micro data. Thus micro and macro data is not weighted in estimation and the final results are independent of the size of the micro sample. Lastly, since FSS data is only available every two to three years, the approach requires interpolating FSS macro data to a yearly basis.

These limitations are addressed in chapter two by the developed Bayesian approach. The proposed framework is an alternative to the GCE approach and also allows combining micro and macro data in the estimation of transition probabilities. Similar to the entropy approach proposed by Zimmermann (2012), micro data is used to specify prior information on the transition probabilities. Specifically, a prior density is defined based on the micro data and combined with a macro data based likelihood function. In comparison to the GCE approach, this combination of prior and likelihood within the Bayesian framework is consistent and more transparent. Also, the approach implies a weighting of micro and macro data such that the precision of both is considered consistently. An additional feature of the proposed approach is that it can handle asynchronous micro and macro data, meaning that the time resolution of the combined micro and macro may differ. It is thus possible to combine, for example, yearly micro data with three yearly macro data. In the application based on FSS macro and FADN micro data it is thus not longer necessary to interpolate the FSS macro data to a yearly basis.

Apart from these improvements over the GCE approach, chapter two also contributes to the literature by proposing two different specifications for the transition probabilities. Specifically, it introduces an ordered logit specification as an alternative to the multinomial logit model used so far in the structural change Markov literature. It is argued that the ordered logit model is not only theoretically more appropriate for ordered choices but also empirically since it requires substantially fewer parameters to be estimated.

The proposed Bayesian estimation framework is evaluated with a Monte Carlo Simulation in order assess the influence of prior information on several performance indicators. Results show that prior information improves the numerical stability of the estimation approach, decreases the variance of the posterior and the mean square error of the posterior mean estimator. The effects become more pronounce the larger the micro sample size and the higher the number of Markov states considered.

Additionally, the proposed estimator is applied in a real world data setting using the same data combination of FSS macro and FADN micro data as proposed by Zimmermann (2012). In the application an example with unordered and ordered states using the multinomial and ordered logit model, respectively, is considered. The application illustrates how asynchronous data, here yearly FADN micro data and two to three yearly FSS macro data, can be combined. The results depict reasonable patterns for the estimated transition probabilities and indicate in what way the prior information (in the micro data) is updated using FSS macro data.

It should be pointed out that the approach proposed in this chapter is equally relevant for the analysis of issues in other disciplines in which micro and macro data is available for the estimation of Markov models (see section 2.1 for a more detailed discussion).

1.2.2 *Short term prediction of agricultural structural change using FSS and FADN data*

Chapter three applies the proposed Bayesian approach to address specific policy requirements and data insufficiencies. The work conducted in this chapter is in parts the result of the joint research project "*Modelling the effects of the CAP on farm structural change*" (Contract 151949-2010-A08-DE) from the European Commission Joint Research Centre - Institute for Prospective Technological Studies (IPTS). The broader aim of the project is to develop novel analytical tools for ex-post and ex-ante analysis of structural change using FADN data. The specific objective of chapter three is a short term prediction of farm structural change using FADN in combination with FSS data.

The Bayesian approach developed in chapter two helps to exploit the specific advantage of each data set while mitigating its disadvantages. FADN data is available on a yearly basis while FSS data is only available every two to three years. It is thus likely that FADN data is available for one to three years after the last available FSS year. FADN data provides micro level information, i.e. the movement of farms between states, but not information about farm entry or exit due to its sampling plan. FSS data, on the other hand, does not allow indentifying the movement of farms over time but provides aggregate information about the number of farms in the population in different states. Implicitly, this contains information about net farm exit/entry. Combining both data sources allows exploiting the micro level information in order to complete the missing macro data for the most recent years and to obtain information not available from any of the two single data sets alone.

The chapter extends the Bayesian approach developed in chapter two by improving the numerical Monte Carlo integration approach used in estimation. The computational implementation of the Bayesian framework is challenging. Integrating of the posterior density is intractable analytically. Instead, a Monte Carlo integration approach is employed for which a sample from the posterior density is obtained. In chapter two the sample is obtained using a Metropolis Hastings algorithm. In this chapter the algorithm is replaced by a Parallel Tempering algorithm (Liu 2008). The Metropolis-Hasting algorithm considers just one Markov chain to generate random outcome from the posterior. The Parallel Tempering approach runs several chains raised to different powers in parallel and allows swapping states between them. The algorithm is capable of escaping local minima more easily which increases the numerical stability of the sampling approach and the final estimation.

Additionally, the chapter contributes to the literature by developing a Bayesian prediction framework that provides an entire predictive distribution instead of only a point prediction. From this predictive distribution, point predictions as well as the variance of predictions can be derived.

The approach is evaluated in an out-of-sample prediction with respect to the completion of macro data information for the most recent years. For this, predictions of farm numbers in different states, specializations and time periods are considered. In each case the Markov states reflect three size classes defined in terms of the economic size of a farm and an artificial entry/exit class. The prediction is performed for these four states considering all farms irrespective of their farm size as well as for three different farm specializations, namely crop, livestock and mixed farms. All predictions are performed for seven (West) German regions for which a relatively long sample is available. Also, three different out-of-sample prediction periods are considered for each prediction. Each time the last FSS year is excluded from estimation and the prediction is performed for this year. Considering all these individual out-of-sample predictions different measures of the prediction quality are calculated. The results are compared to naive linear, geometric and constant (i.e. no change) predictions that are additionally performed for each considered situation.

The out-of-sample prediction results indicate that the proposed Markov prediction approach outperforms the geometric and linear prediction. It failed however, to clearly outperform the prediction of no change. These results indicate that structural change within two to three years is rather modest and the prediction of “no change at all” is difficult to compete with. Nevertheless, the proposed approach can be useful in order to predict farm numbers between FSS years or for longer prediction periods in which a prediction of no change becomes less plausible.

1.2.3 *Direct payments, spatial competition and farm survival in Norway*

Chapter four focuses on the importance of spatial interaction between farms for policy assessment. Particularly it looks at farm exit decision in Norway and the role of direct payments in this respect. The hypothesis explored in the paper states that farms interact with each other in multiple ways and that these interactions are important for farmers’ survival decision and need to be considered in policy assessment when aggregating results from the individual farm to the regional level.

The chapter thus addresses two limitations of the Markov approach. These two issues are first the interdependency between individual farm behaviors and secondly the aggregation of individual farm level results to the regional level. In the following, the importance and implications of both issues for the Markov approach are discussed. Afterwards the contribution of chapter four is highlighted in this respect.

Interdependence between farm behavior

The specification of the Markov model relies on the multinomial logit model or as proposed in chapter two on the ordered logit model. Even though the latter might be an improvement in situations with ordered states, both specifications still rely on the assumption of independent individual transitions between states. For the concrete example of farm structural change, this implies that farm transitions between different size or specialization classes are independent from each other. This assumption is problematic. Even though decisions to change farm size, specialization or to enter or exit are indeed taken at the individual level they are likely to be influence by decisions of other farms.

Interdependence between farm behavior can come in multiple forms. Flaten (2002) points out that farmers are part of a social rural network, which is important for social well being of the rural society. A strengthening or weakening of this network might thus affect individual farm decision, while decision of the individual farmer (i.e. to exit) might affect the social network[2]. Similarly, farmers are part of a corporate network of suppliers, wholesalers and processors on which they depend but which also depends on their individual decisions (Mosnier and Wieck 2010). Good access to up- and downstream industries is vital for farm productivity and survival. But also do these industries depend on the decision by individual farmers to change their specialization or to quit. Farmers are also part of a corporate network with other farmers important for technology adoption and knowledge transfer (Rogers 1995; Berger 2001). For example, Case (1992) and Holloway et al. (2002), found evidence that the probability of adopting a new technology increases with neighboring adoption. Consequently, an active corporate network may raise technology diffusion, which increases farm productivity and finally influences decisions of an individual farmer. Apart from these network effects farmers also compete on input and output markets. In most of the structural change literature, prices for inputs and outputs are takes as given, which often makes sense. For some goods, however, markets are local with only few farms participating such that their decisions matter directly for market

[2] This simulations interactions are also known as Manski's reflection problem (Manski 1993).

outcomes. On the output side this might hold for goods that are marketed locally. The most important case in this respect, though, is the land market (Leathers 1992; Margarian 2010). Due to the physical immobility of land and transportation costs which increases with the distance between farm and plot, potential buyers are typically limited to a small number of farms. Common to almost all these forms of market or network interactions is that they are spatial in nature, meaning that the interaction level decreases with distance. The importance of the location of farms in space, the immobility of land and the spatial interaction between farms is considered in the Agent based model literature on farm structural change (Balmann 1997; Happe 2004; Happe et al. 2006; Happe et al. 2008; Berger 2001; Freeman et al. 2009).

Ignoring the interactions in the Markov approach is problematic not only because they might be relevant explanatory variables itself but also because they are important for a consistent aggregation of the farm level results. A particular problem arises in the Markov approach with respect to the land market: In most regions, agricultural area is almost fixed in supply and fully employed such that the prerequisite for farms to grow is that other farms free resources by declining in size or exit the sector (Weiss 1999). In the classical application of the Markov approach where farms are grouped into different size classes, this interaction between farms and the resulting limitation to farm growth is not accounted for and difficult to do so. This implies that using estimated Markov transition probabilities farms may predict to transit to larger size classes without other farms giving up area in comparable quantities. We might thus predict a farm distribution in which more than the total available agricultural area is employed. In cases where size classes are defined in terms of economic size units (as in chapter 3), the interconnection is not as direct, since all farms may grow in terms of the economic size by intensifying production, for example by increasing livestock density per area. Nevertheless, most of livestock production remains, to some extent, coupled to agricultural area. At some point, growing in economic size units is likely, to go along with an increase in cultivated area.

Consistent aggregation of farm level results

Following from this discussion we conclude that imposing the independence assumptions in the Markov approach can lead to violations of land constraints that exist at the aggregated level when aggregating the individual level results. This particular problem is an example of a more general problem in the analysis of farm structural change - not limited to the Markov approach - which relates to the aggregation of individual farm level results to the aggregated regional level.

Empirical studies analyzing farm structural change may in general be classified in either farm level or regional level studies. Regional level studies analyze the change in total number of farms, whereas individual level studies consider changes at the farm level. Both types have advantages and disadvantages with respect to different purposes. Regional level studies (such as Goetz and Debertin 2001 or Breustedt and Glauben 2007) directly deliver the result at the aggregated regional level which is of interest in an assessment of policy effects on farm structural change. As pointed out by Gale (1994), however, the disadvantage of regional level studies is that aggregated statistics may mask quite complex behavior at the individual farm level (see also Ehrensaft et al. 1984:824 in this respect). Further, explanatory variables need to be defined at the regional level as well. Interpretation of the effects is thus only possible indirectly and statistical more complicated. Also identification might be more difficult since only the variation between regions can be used.

Individual level studies (such as Kimhi and Bollman 1999; Weiss 1999; Gale 2003; Bragg and Dalton 2004; Hoppe and Korb 2006; Dong et al. 2010) in contrast consider explanatory variables at the individual farm level. This allows exploiting the variation between individual which eases identification and might make definition and interpretation easier and more direct. Understanding the driving forces of farm structural change at the individual level might be an important result of its own and the final purpose of an individual farm level study. For policy assessment, however, one is usually interested at the aggregated regional effect of a policy. For the required aggregation it is crucial to consider the interaction between farms discussed above, but accounting for this interaction is difficult and often impossible. This lead Roberts and Key (2008:628) to argue in favor of regional level studies over individual level studies for policy assessment: “[Farm level] studies [...] consider effects of payments on the growth or survival of individual farms, which cannot predict the effects of an increase in payments on aggregate farm structure. This is because studies of individual farms cannot account for how induced changes on one farm affect other, neighboring farms […].”

In social science these “aggregation problems” are well known and discussed under the term “emergence”, meaning that macro patterns arise from the interaction of individuals which could not be derived from the properties of the individuals (Emmeche et al. 1997; Schelling 2006; Epstein 2006).

The Markov approach cannot clearly be classified as either an individual or regional level study. Instead it is at the intersection between both, which has advantages as well as disadvantages. In most cases the input data either consists of individual level transitions (micro data) or the aggregated number of farms in different classes (macro data; see Zimmermann et al. 2009 for an overview of studies differentiated by data use). In both cases explanatory variables are usually defined at the aggregated level which has the advantage that data requirements are relatively low, but, as mentioned above, interpretation and identification is more problematic. The resulting transitions probabilities describe behavior at the individual farm level. The possibility to use macro (i.e. aggregated) data to derive information about the individual level behavior (the transition probabilities) is a unique an attractive feature, compared to other approaches applied in the structural change context. For some application this individual level behavior is of final interest. In other instances, such as policy assessment the individual level results need to be aggregated. The specific Markov application in Zimmermann and Heckelei (2012a), Zimmermann and Heckelei (2012b) and in chapter two and three add a new twist in the classification of the Markov approach as either a individual or regional level approach. Here, farm level micro data is introduced as additional information and combined with the aggregated macro data. In these specific applications individual farm behavior is thus derived from a combination of macro and micro data. This provides advantages as discussed above but explanatory variables remain to be defined at the regional level with the associated disadvantages. Also the problems arising from the independence assumption remain unchanged.

Contribution

Both issues, the interdependence between farms and the aggregation of farm level results, are addressed in chapter 4. With farm survival the chapter focuses on a narrow but important aspect for farm structural change. The analysis takes place at the individual farm level using a Norwegian data set.

The specific focus of the chapter is to identify the role of policy in farm survival and particularly the effect of direct payments. Within this context the importance of spatial interactions between farms and their role in policy assessment is considered specifically. In order to assess the influence of a change in the direct payments scheme on the aggregated/regional farm level the interdependence are considered when aggregating farm level results. The empirical analysis of farm level spatial interdependence is novel to the farm structural change literature, even though farm level spatial interdependence is highlighted in the theoretical structural change literature and considered in agent based models. The study thus contributes to the literature by showing that spatial interdependence between farms is indeed important empirically. Furthermore, results indicate that spatial interdependence is important for a policy assessment at the aggregated/regional level. Failure to account for it may lead to substantial overestimation of the policy effects.

The empirical analysis employs a Norwegian data set covering almost all Norwegian farms in 1999 and 2009, providing the production activities in the two years as well as some additional farm characteristics, including the location of each farm in space. Based on the production activities it is possible to derive the direct payments each farm receives. This spatially explicit data set at near census level covering more than 64.00 farms provides a unique opportunity to analyze spatial interdependence at farm level.

A spatial binary choice probit model is estimated. The binary dependent variable is defined as farm survival/exit between 1999 and 2009. As explanatory variables several own and neighboring characteristics of the farm and the holder are considered. With respect to the research question, the primary interest lies on own and neighboring direct payments. Two specifications of the spatial probit model are considered. First, a spatially lagged explanatory variable model (SLX) and second a spatial Durbin error model (SDEM). Both specifications consider spatially lagged neighboring characteristics, but the SDEM additionally allows for spatial correlation in the errors.

Results indicate that the most important variables to explain farm survival are variables related to the own size of a farm, such as the total labor input, the agricultural area or the total direct payments. All three variables are positively correlated with farm survival indicating that large farms are more likely to survive. All other variables add only little to the overall explanatory power of the model. Nevertheless, with respect to the research question of evaluating the effect of direct payments on the aggregated level the effects of the spatially lagged neighboring characteristics are crucial to consider. Results show that neighboring direct payments negatively influence own survival. The overall effect of a change in direct payments is thus a complex process and depends on the interaction between farms. This issue is explored in greater detail using policy scenario simulations. With respect to the overall importance of farm interdependence results further indicate that neighboring agricultural area and total labor input have, ceteris paribus, a positive influence on own survival. Findings of negative effects of neighboring payments but positive effects of neighboring farm area and labor input hint at the different channels through which farms interacts. The negative effects of direct payments can be seen as an indication of competition on the land market. The positive effects of neighboring area and labor input on the other hand show that farms gain from larger neighbors (as long as direct payment are kept constant). One explanation for this can be positive effects through an active corporate network, which is strengthened by large and potentially more active neighbors.

For policy assessment at the aggregated/regional level, scenario simulations are performed for the entire farm population based on the obtained regression results. Different policy scenarios such as an overall decrease of the payments rates by 10% or the abolishment of specific elements of the payment scheme that favors smaller farms are considered. In both cases, the change in the predicted survival probability before and after the policy change is derived for the entire population considering the neighboring relationship. The difference in the survival probability provides an assessment of the policy effects at the aggregated/regional level. Overall, the effects of direct payments on farm survival remain modest. The same simulations are repeated for a model that ignores the spatial interactions in the estimation and simulations. Results show that ignoring spatial interaction lead to an overestimation of the effects of direct payments.

1.3 **Conclusion**

Summary of results

The overall aim of the dissertation is to develop and apply methods to analyze farm structural change. The dissertation can broadly be distinguished into two major parts. In the first, a methodological gap in the literature with respect to the combination of micro and macro data in the estimation of Markov models is addressed. The second part looks at the importance of interactions between farms and thus addresses limitations of the Markov approach considered in the first part. The recurring theme in the dissertation is the relationship between the micro and macro level. Strategic decisions are taken by the individual farmer at the micro level. The aggregation of these individual decision lead to changes in the farm structure at the macro level. These different levels of analysis are considered in varying ways throughout the dissertation.

Specifically, in part one the relationship between the micro and macro level is reflected in the data considered for estimation. Here micro and macro data is combined in the estimation of Markov transition probabilities which describe behavior on the micro level. The dissertation contributes to the literature by providing a Bayesian estimation framework for non-stationary Markov models. The proposed Bayesian approach allows combining micro and macro data in the estimation more consistently and transparently than other methods previously applied in the literature. Based on Monte Carlo Simulation it is shown that adding micro data to a macro data based Markov estimation indeed improves the precisions of the estimates and the numerical stability of the approach. Additionally, a Bayesian prediction formwork is developed that enables a prediction of farm numbers in the EU in different categories, based on a combination of two data sources that allows deriving information not available from one data set alone.

In a second part, it is argued and shown empirically that the assumption of independence between farm behaviors on the micro level may become problematic for specific applications. Working on the individual farm level usually provides more information, but for policy assessment an aggregation to the regional level becomes necessary which needs to consider interactions between farms. Theoretically or in agent based models interactions between farms are considered in multiple ways in the literature. To the knowledge of the author, however, the study is the first to show empirically that spatial interactions at the farm level are indeed important for analysis of farm structural change. Specifically, it is shown for the case of Norway that neighboring characteristics are relevant for the influence of direct payments on farm survival. The empirical results indicate that direct payments a farm receives itself have a positive influence on farm survival while neighboring direct payments have a negative one. For an overall assessment of the policy effects it is thus necessary to consider the interaction between farms. Ignoring these interactions might lead to an overestimation of the effects of direct payments.

Limitations and outlook

Despite these contributions to the literature there are several remaining shortcomings of which some are more specific with respect to data or technical issues and some more general. Here the focus is on the more general shortcomings and an outlook for relevant future research is provided. The more detailed shortcomings are left to the specific sections in each individual chapter.

A major contribution of the work is the empirical analysis of the importance of farm interaction on the micro level. A particular challenge in this respect is to identify the concrete channels through which the interaction between farms occurred. One usually can infer from empirical approaches information about - in the best case - causal relationship between variables. The underlying mechanism that drives the relationship, however, remains often unobserved. Economic theory can provide explanations about the relationship, but in this context theory about the spatial interaction between farms is not very well developed, complex and sometimes conflicting. The explanations for the interactions thus remain partial.

Agent based models, on the other hand, approach the issue of aggregation and interaction from the opposite side. They naturally work at the micro level and agents are allowed to interact with each other. The aggregate regional results then emerge from the interaction between individuals. The aggregation problem is thus solved endogenously, which is one of the strength of the agent based model approach. The problem here is that the way farms interact in the first place, is based on assumption made in the model design. This model design can be based on economic theory or empirical evidence. Often, however, the specification is rather ad hoc with a weak empirical justification. This is problematic because even though the aggregated results emerge consistently they crucially depend on the correct specification of the micro level interactions.

The agent based model approach thus solves the aggregation problem naturally and allows defining the exact mechanisms of farm interaction. Empirical (econometric) approaches, on the other hand, provide empirical evidence of the importance of interdependence but are usually limited with respect to an explanation of the mechanisms that lead to the interdependence. The two approaches thus seem to complement each other and a combination of both could be fruitful in further research to exploit their individual advantages. One approach would be to implement alternative assumption about possible interaction mechanisms in the agent based model. The model outcomes can then be compared to the empirically observed patterns. This approach would help to shed light on the mechanisms that most likely lead to the relationships observed empirically. The obtained results might be useful for theory development and the approach provides an empirical justification of specific agent base model assumptions. The outlined approach can more clearly be illustrated using the empirical results presented in chapter 4. We found that neighboring direct payments have a negative influence on farm survival while neighboring cultivated area and labor input have a positive influence. From these empirical results, however, we can only conclude indirectly about the potential mechanisms through which farms interact. As argued in chapter 4, it is likely that the negative influence of direct payment hint at competition on the land market while the positive influence of cultivated area or labor input hint at positive influences due to corporate network effects. An agent based model can be used to explore the roll of alternative mechanisms of interactions. It may be compared which form of interactions leads to the observed pattern. Specifically, alternative versions of the agent based model can be considered which are based on different assumptions concerning the land market or corporate network effects. For each specification the regression applied to the empirical data can be repeated for the agent based model results and it can be explored which specification lead to similar patterns observed empirically. Even though, this strategy will not provide a direct proof of the underlying mechanisms of farm interaction it will nevertheless help to understand which specification is capable of reproducing the observed pattern and thus is most likely at work in reality. The combination of spatial econometric approaches, similar as the one proposed in this dissertation and agent based models would provide an empirical validation of the agent based model and could be helpful for theoretical development concerning farm level interactions.

1.4 **References**

Balmann A. 1997. Farm-based modelling of regional structural change: A cellular automata approach. *European Review of Agricultural Economics* **24**(1):85–108.

Benirschka M, Binkley JK. 1994. Land Price Volatility in a Geographically Dispersed Market. *American Journal of Agricultural Economics* **76**(2):185–195.

Berger T. 2001. Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics* **25**(2-3):245–260.

Boehlje M. 1992. Alternative models of structural change in agriculture and related industries. *Agribusiness* **8**(3):219–231.

Bragg LA, Dalton TJ. 2004. Factors Affecting the Decision to Exit Dairy Farming: A Two-Stage Regression Analysis. *Journal of Dairy Science* **87**(9):3092–3098.

Breustedt G, Glauben T. 2007. Driving Forces behind Exiting from Farming in Western Europe. *Journal of Agricultural Economics* **58**(1):115–127.

Case A. 1992. Neighborhood influence and technological change. *Regional Science and Urban Economics* **-** *Special Issue Space and Applied Econometrics* **22**(3):491–508.

Cochrane W. 1958. Farm Prices: Myth and Reality. Minneapolis: University of Minnesota Press.

Dong F, Hennessy DA, Jensen HH. 2010. Contract and Exit Decisions in Finisher Hog Production. *American Journal of Agricultural Economics* **92**(3):667–684.

Ehrensaft P, LaRamee P, Bollman RD, Buttel FH. 1984. The Microdynamics of Farm Structural Change in North America: The Canadian Experience and Canada-U.S.A. Comparisons. *American Journal of Agricultural Economics* **66**(5):823.

Emmeche C, Køppe S, Stjernfelt F. 1997. Explaining Emergence: Towards an Ontology of Levels. *Journal for General Philosophy of Science* **28**(1):83-117.

Epstein JM. 2006. Generative social science. Princeton: Princeton University Press.

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**(4):429–441.

Freeman T, Nolan J, Schoney R. 2009. An Agent-Based Simulation Model of Structural Change in Canadian Prairie Agriculture, 1960-2000. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* **57**(4):537–554.

Gale HF. 1994. Longitudinal Analysis of Farm Size over the Farmer's Life Cycle. *Review of Agricultural Economics* **16**(1):113.

Gale HF. 2003. Age-Specific Patterns of Exit and Entry in U.S. Farming, 1978-1997. *Review of Agricultural Economics* **25**(1):168–186.

Goddard E, Weersink A, Chen K, Turvey CG. 1993. Economics of Structural Change in Agriculture. *Canadian Journal of Agricultural Economics* **41**(4):475–486.

Goetz SJ, Debertin DL. 2001. Why Farmers Quit: A County-Level Analysis. *American Journal of Agricultural Economics* **83**(4):1010–1023.

Golan A, Vogel SJ. 2000. Estimation of Non-Stationary Social Accounting Matrix Coefficients with Supply-Side Information. *Economic Systems Research* **12**(4):447–471.

Happe K. 2004. Agricultural Policies and Farm Structures - Agent-Based Modelling and Application to EU-Policy Reform. Halle: Vauk.

Happe K, Balmann A, Kellermann K, Sahrbacher C. 2008. Does structure matter? The impact of switching the agricultural policy regime on farm structures. *Journal of Economic Behavior & Organization* **67**(2):431–444.

Happe K, Kellermann K, 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**(1): 49. [online].

Heckelei T, Mittelhammer RC, Jansson T. 2008. A Bayesian Alternative to Generalized Cross Entropy Solutions for Underdetermined Econometric Models. Discussion Paper 2008:2, Institute for Food and Resource Economics, University of Bonn.

Holloway G, Shankar B, Rahmanb S. 2002. Bayesian spatial probit estimation: a primer and an application to HYV rice adoption. *Agricultural Economics* **27**(3):383–402.

Hoppe RA, Korb P. 2006. Understanding U.S. Farm Exits. USDA Economic Research Report Number 21, June 2006.

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

Huettel S, Margarian A, von Schlippenbach V. 2010. Regional asymmetries in farm size. Paper presented at the 114th EAAE Seminar, April 15 - 16, Berlin, Germany.

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

Kimhi A, Bollman R. 1999. Family farm dynamics in Canada and Israel: the case of farm exits. *Agricultural Economics* **21**(1):69–79.

Leathers HD. 1992. The Market for Land and the Impact of Farm Programs on Farm Numbers. *American Journal of Agricultural Economics* **74**(2):291–298.

Liu JS. 2008. Monte Carlo strategies in scientific computing. New York: Springer.

Manski CF. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* **60**(3):531.

Margarian A. 2010. Coordination and Differentiation of Strategies: The Impact on Farm Growth of Strategic Interaction on the Rental Market for Land. *German Journal of Agricultural Economics* **59**(3 Special Issue):202–216.

Mosnier C, Wieck C. 2010. Determinants of spatial dynamics of dairy production: a review. Discussion Paper 2010:2, Institute for Food and Resource Economics, University of Bonn.

Roberts MJ, Key N. 2008. Agricultural Payments and Land Concentration: A Semiparametric Spatial Regression Analysis. *American Journal of Agricultural Economics* **90**(3):627–643.

Rogers EM. 1995. Diffusion of innovations. New York: Free Press.

Schelling TC. 2006. Micromotives and macrobehavior. New York: Norton.

Stanton BF. 1993. Farm Structure: Concept and Definition, in Hallam A. (Ed.), Size, Structure, and the Changing Face of American Agriculture, Boulder: Westview Press, 14–29.

Weiss CR. 1999. Farm Growth and Survival: Econometric Evidence for Individual Farms in Upper Austria. *American Journal of Agricultural Economics* **81**(1):103–116.

Zimmermann A. 2012. Empirical analysis of farm structural change at EU-level. Dissertation, Bonn: University of Bonn.

Zimmermann A, Heckelei T. 2012a. Differences of farm structural change across European regions. Discussion Paper 2012:4, Institute for Food and Resource Economics, University of Bonn.

Zimmermann A, Heckelei T. 2012b. Structural Change of European Dairy Farms - A Cross-Regional Analysis. *Journal of Agricultural Economics* **63**(3):576–603.

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

Chapter 2
Bayesian Estimation of Non-Stationary Markov Models Combining Micro and Macro Data[3]

Abstract. We develop a Bayesian framework for estimating non-stationary Markov models in situations where not only macro population data is available on the proportion of individuals residing in each state, but also micro-level sample data on observed transitions between states. Posterior distributions on non-stationary transition probabilities are derived from a micro-based prior and a macro-based likelihood using potentially asynchronous data observations, providing a new method for inferring transition probabilities that merges previously disparate approaches. Monte Carlo simulations demonstrate how observed micro transitions can improve the precision of posterior information. We provide an empirical application in the context of farm structural change.

Keywords: Markov process, transition probabilities, micro and macro data, data combination

JEL classification codes: C11, C81

[3] An earlier version of this part is published as Storm H, Heckelei T, Mittelhammer RC. 2011 Bayesian estimation of non-stationary Markov models combining micro and macro data. Discussion Paper 2011:2, Institute for Food and Resource Economics, University of Bonn.

2.1 Introduction

A new Bayesian framework for inferring the transition probabilities of non-stationary Markov models is developed in this paper. Non-stationary Markov models facilitate analysis of factors influencing the probability that an individual will transition between predefined states. Data used for estimating Markov models can either be panel data, where the specific movement of an individual between states is observed over time, or aggregated data, providing only the number of individuals residing in each state over time. Following Markov terminology, we refer to such panel data and aggregated data as *micro* and *macro* data, respectively. The overall objective of our approach is to combine micro and macro information into a unified and consistent methodology for estimating transition probabilities.

The idea of combining micro and macro data was considered previously in the context of a medical application by Hawkins and Han (2000). They analyzed macro data obtained in repeated independent cross sectional surveys within a city district together with limited micro data obtained from respondents who were 'coincidently' interviewed in two consecutive cross sectional surveys. The behavior under study was the benefits of an intervention program attempting to modify drug use-related behavior, and their Markov model was a two-state process relating to awareness, or not, of the health consequences of not bleaching shared drug needles. They defined a linear model, within the Classical statistical framework, that explained the binary marginal probabilities of being in one of the two awareness states in a certain time period (based on "standard observed proportion estimates" from aggregate data) as well as transition probabilities relating to transitions between the two states (from the micro data). Generalizations of Hawkins and Han's binary state model to multinomial transitions are conceptually possible, but the parameter dimensionality, as well as the complexity of the covariance structure and constraint set imposed by the sampling design, quickly renders their general linear model approach intractable as the number of states increase beyond two.

A recent alternative by Zimmermann and Heckelei (2012) utilizes a Generalized Cross Entropy (GCE) approach to combine micro and macro data, and has an advantage relative to Hawkins and Han of being dimensionally and computationally better suited for modelling multinomial Markov processes with a relatively large number of states. They utilize estimates of transition probabilities derived from observed micro transitions as reference probabilities in the GCE approach. However, treating the reference probabilities as priors on the transition probabilities, as they do, results in a relative weighting of micro and macro data information that is independent of the precision of the estimates underlying the prior information. In particular, the influence of the micro estimates on the final estimation result is the same no matter how large the micro sample, and thus no matter how precise the prior information is relative to population characteristics. In addition, the approach requires the specification of reference distributions for residuals, including the specification of support points, which determine the signal-to-noise ratios in the Markov transition equations *a priori*.

In contrast to the previous two Classical approaches, the Bayesian framework provides a flexible and tractable method of combining micro and macro data generating processes that is logically consistent and coherent within the tenets of the probability calculus while accommodating a relatively large number of Markov states. The rather complicated linkages between transition probabilities and observed Markov state outcomes, and the complex parametric constraints and covariance matrix structure of the combination of micro and macro data generating processes, are specified consistently as a matter of course in specifying the posterior probability distribution for the parameters of the transition equation. Moreover, the Bayesian framework allows prior information to be incorporated into the estimation of non-stationary Markov models within an established coherent probabilistic framework. In addition, the Bayesian methodology provides a natural and relatively straightforward way of combining data observations at either the macro or micro level that are asynchronous[4], which is in contrast to the methods offered heretofore. The approach is also applicable to both ordered and unordered Markov states, which is yet another flexible feature of the method. Overall, the Bayesian approach that we present offers a tractable full posterior information approach for combining micro and macro data-based

[4] By “synchronous” we mean both that observations over time occur in sequence without gaps (follow a tact) and that the micro and macro data are observed for the same time units.

information on non-stationary transition probabilities that allows the estimation of functional relationships linking transition probabilities with their determinants.[5]

Examples of empirical problems for which both macro and micro data are relevant and available exist in previous literature. One example is an analysis of European Union (EU) farm structural change, where structural change is defined as farm size or production specialization change over time (see Zimmermann et al. 2009 for a review of that strand of literature). In that application population data on the number of farms in specific size or specialization states is available from the Farm Structure Survey (FSS). Micro data, offering observed transitions of individual farms between the states, is available in the Farm Accountancy Data Network (FADN), albeit for a relatively small sample of farms. Another example is an analysis of voter transitions in political science. Here, macro data on the vote shares of candidates is available from official statistics, whereas micro data can be obtained from voter (transition) surveys (McCarthy and Tyan 1977; Upton 1978). Additional examples of similar data situations can be found in the context of Ecological inference problems, which are closely related to Markov processes (Wakefield 2004; Lancaster et al. 2006). In general the proposed approach is relevant for all situations in which the micro sample is relatively small compared to the macro data. If the micro sample is relatively large the macro sample does not contribute additional information such that an approach relaying exclusively on the micro data is sufficient.

The paper is organized as follows: First, the Bayesian framework for non-stationary Markov models is developed in section 2. Two different specifications of the transition probabilities, that of ordered and unordered Markov states, are discussed, appropriate likelihood functions and prior densities are defined, and issues relating to computational implementation are identified. Then the design and results of a Monte Carlo simulation experiment are presented in section three

---

[5] In their pedagogical contribution to the use of MCMC computational methodology Pelzer and Eisinga (2002) include an example of a Bayesian approach specifically designed for a two state Markov model which depends crucially on the characteristics of a Bernoulli process. The specification of prior information in their example is effectively ad-hoc, whereas our specification is fully consistent with the structure of the data generating process. Moreover, their example does not generalize to either stationary or non-stationary multinomial Markov processes.

and used to assess how the inclusion of prior information affects the posterior as well as the numerical stability of the sampling algorithm, and the degree to which estimator performance is improved under different micro sample sizes for both specifications. In section four the methodological framework is applied empirically in the context of an analysis of farm structural change in Germany. The application demonstrates how the framework can facilitate estimation in a situation where estimation with either micro or macro data alone would suffer from several limitations. Section 2.5 provides conclusions and a discussion of areas for further research.

2.2 Bayesian Approach for Non-Stationary Markov Models

Markov processes provide a conceptual model for the movement of individuals between a finite number of predefined states, $i=1,...,k$, within the context of a stochastic process. The k states are mutually exclusive and exhaustive. A Markov process is characterized by a $(k \times k)$ transition probability (TP) matrix[6] $\mathbf{P}_t$. The elements P_{ijt} of $\mathbf{P}_t$ represent the probability that an individual moves from state i in time $t-1$ to j in time t. The $(k \times 1)$-vector $\mathbf{n}_t$ denotes the number of individuals in each state i at time t and evolves over time according to a (first order) Markov process

$$\mathbf{n}_t = \mathbf{P}_t' \mathbf{n}_{t-1}. \tag{1}$$

In a non-stationary Markov process, the TPs change over time periods[7] $t=0,1,...,T$. Data used for estimating a non-stationary Markov process can either be macro or micro level. In the case of macro data, only the aggregate numbers of individuals in the states, $\mathbf{n}_t$, is observed at each time period. For micro data, the movement of each individual between states is also observed over time. Thus, the $(k \times k)$-matrix $\mathbf{N}_t$ with elements n_{ijt} representing the number of individuals that transition from state i at $t-1$ to j in t, is directly observed.

---

[6]Bold letters are used for vectors or matrices.

[7]Depending on the problem context, one could also consider only two time periods observed over various regions, or a combination of multiple time and regional observations.

In this section we assume data observations are synchronous, as defined in footnote 1, both for ease of exposition and to be consistent with precedence in the literature. However, the proposed approach is considerably more flexible in that asynchronous data can be analyzed in a straightforward way, and in the empirical application in section 4, macro data available only every two to three years will be combined with yearly micro data. Similarly, the reverse case, where macro data has a higher temporal resolution than the micro data, can be considered as well.

The structural specification of the TP matrix $\mathbf{P}_t$ depends on the underlying behavioral model. In the following subsection we review TP matrix specifications corresponding to ordered as well as unordered Markov states to define notation and establish the foundation for the definition of the posterior. Then the data likelihood function $L(\mathbf{n}_1,...,\mathbf{n}_T|\boldsymbol{\beta})$, representing the macro data, and a prior density $p(\boldsymbol{\beta})$, representing the micro data are defined and combined into the posterior distribution for the TPs.[8] The last subsection presents computational methodology relating to the use of the posterior distribution for inference purposes.

Specification of the Transition Probability Matrix

For appropriate specification of the TPs, the nature of the relationship between Markov states need to be considered, and we discuss two different behavioral models that differentiate between ordered and unordered Markov states. We argue that for ordered Markov states the ordered logit model is superior to the more common multinomial logit model with respect to both model assumptions and from a computational point of view.

In cases where the states of the Markov process are unordered, the multinomial logit model is a suitable specification for the TPs[9]. The specification based on the

[8] In his dissertation, Rosenqvist (1986) introduces the conceptual rudiments of combining micro and macro data in a prior-likelihood framework. However, the analysis was restricted to stationary processes with synchronous observations and the micro and macro data observations were assumed to be disjoint. Our Bayesian framework is not constraint by any of these assumptions and moreover, we provide a tractable empirical method of implementation.

[9] A multinomial probit model could be an appropriate alternative for the error structure specification, but is left for future work.

multinomial logit model assumes that the transition of individuals between different states can be represented by a random utility model. The utility that would accrue to individual l upon moving from state i in $t-1$ to j in t is denoted as $U_{ijtl} = V_{ijt} + \varepsilon_{ijtl}$, where the deterministic component of utility is specified as $V_{ijt} = \mathbf{z}'_{t-1}\mathbf{b}_{ij}$, with $\mathbf{z}_{t-1}$ being a vector of lagged exogenous variables. The deterministic part varies only over time and not over individuals because aggregated (macro) data is considered. Consequently, the deterministic component of utility reflects exogenous variables that affect the utility of all individuals alike. The random error ε_{ijtl} varies over time and individuals. It is assumed that an individual chooses a transition that maximizes her utility U_{ijtl}. The assumption that ε_{ijtl} are *iid* random draws from a Gumbel distribution result in a multinomial logit specification for each row of $\mathbf{P}_t$.

If the Markov states are ordered, an ordered choice model is an appropriate specification for the underlying behavioral model. In this case it is assumed that there exists an unobserved continuous latent variable Y^*_{itl} for each individual l that determines the outcome of the observed variable Y_{itl} according to

$$Y_{itl} = j \quad \text{if} \quad c_{j-1} < Y^*_{itl} \leq c_j \quad \forall\, i, j = 1, \ldots, k \tag{2}$$

where the c_j's are the thresholds for each Markov state, with $c_o \equiv -\infty$ *and* $c_k \equiv \infty$. The index i indicates that an individual was in state i at $t-1$. The unobserved latent variable Y^*_{itl} consists of a deterministic part $\mathbf{z}'_{t-1}\boldsymbol{\beta}_i$ plus a random part ε^*_{itl}. The vector of unknown parameters $\boldsymbol{\beta}_i$ are allowed to differ between the k different states in $t-1$. As in the preceding multinomial logit model, the deterministic part varies over time but not over individuals. Assuming that ε^*_{it} are *iid* random draws from a logistic distribution[10] results in an ordered logit model for each row of $\mathbf{P}_t$.

One important difference between the ordered logit and the multinomial logit model is that only one error term, instead of one error term for each alternative, is considered for each individual. This implies that the assumption of “Independence

---

[10] Assuming that the ε^*_{it} are random draws from a normal distribution would result in a probit (see footnote 6).

of Irrelevant Alternatives" (IIA) does not apply to the ordered logit model. This is more appropriate whenever the alternatives are ordered since in this case it can be expected that the error associated with one state is more similar to the error of an alternative close to it than to an alternative further away (Train 2009). Also from a computational point of view, the ordered logit specification is often preferable since only $k\,n_z + (k-2)k$ parameters[11] need to be estimated, as compared to $k(k-1)n_z$ parameters for the multinomial logit model.

A further advantage of the ordered choice model is that the interpretation of the latent variable is often straightforward. For example, in the case of farm structural change noted in the introduction, where Markov states refer to firm size classes, the latent variable can be interpreted directly as farm size (see section 4). In the medical context where classes refer to different stages of illness, the latent variable can be interpreted as the degree of illness. However, the decision between an ordered and unordered choice model is not always straightforward and can depend on the problem context as well as decision makers' behavioral characteristics. In the voter transition example, one could regard the candidates as unordered choices, but alternatively one could also argue that they are ordered according to a one-dimensional political spectrum ("right" to "left"), in which case both models have justification and the choice between the two must be guided by additional theoretical and/or substantive behavioral arguments.

Posterior

The posterior is defined as the joint density of a micro data prior and macro data likelihood. Since micro and macro data are interdependent, the likelihood is the conditional density of the macro data given the micro data. The prior density represents information derived from a sample of micro observations on state transitions. It should be pointed out that the distinction between prior and likelihood is somehow artificial. Both are likelihood specification representing two different data sets. Also they are sampled at the same time which usually

[11] If a constant is included and c_1 is normalized to zero $(k-2)k$ cut points need to be estimated in addition to one parameter for each explanatory variable and state ($k\,n_z$) .

distinguishes prior and likelihood information. The labeling is thus more a convention and is motivated from the works in the context of the entropy estimation by Zimmermann and Heckelei (2012), mentioned above, using micro data to specify the support and prior densities in an entropy estimation based on macro data.

The foundation for the likelihood function is provided by the first-order non-stationary Markov process proposed by MacRae (1977). For the specification of a macro data based likelihood function MacRae (1977) points out that the nature of the likelihood specification depends critically on whether the state proportions, $\mathbf{x}_t$, are observed over time for the entire population of size N, which she refers to as *perfect observations*, or whether the state proportions, $\mathbf{y}_t$, are only a random sample of size $M_t < N$ drawn and observed at each time period, referred to as *imperfect observations*. In the case of perfect observations the distribution of $\mathbf{x}_t$ is fully characterized by $\mathbf{x}_{t-1}$. However, for imperfect observations the distribution of $\mathbf{y}_t$ also depends on earlier observations, $y_{t-2},...,y_0$, which provide additional information on y_t. For the latter case MacRae (1977) proposed a limited information likelihood approach which is appropriate whenever macro data is available for only a sample of the population. In the following, we focus on the case of perfect observations, i.e., a census type of macro data set, which characterizes the type of data available in our empirical application provided in section 4.

MacRae (1977) shows that in the case of perfect observations, the state proportions are distributed as a weighted sum of independent multinomial random variables with probabilities equal to the corresponding rows in $\mathbf{P}_t$ and weights equal to the state proportions in $t-1$. The resulting likelihood function is given by

$$L\left(\boldsymbol{\beta}\middle|\mathbf{n}_0,\mathbf{n}_1,...,\mathbf{n}_T\right)=\prod_{t=1}^{T}\sum_{\mathbf{H}_t\in\mathbb{H}_t}\prod_{i=1}^{k}\left(n_{i,t-1}!\right)\left(\prod_{j=1}^{k}P_{ijt}^{\eta_{ijt}}/\eta_{ijt}!\right). \tag{3}$$

The n_{it}'s are the elements of the data vector $\mathbf{n}_t$. The matrix $\mathbf{H}_t$ is of dimension $(k\times k)$ and has entries η_{ijt} denoting the (unobserved) number of individuals transitioning from state i at time $t-1$ to state j at time t. The summation involving $\mathbf{H}_t$ in likelihood expression (3) is over the set $\mathbb{H}_t$ of all matrices $\mathbf{H}_t$

having rows that sum to corresponding elements in $\mathbf{n}_{t-1}$ and columns that sum to the corresponding entries in $\mathbf{n}_t$, so that

$$\mathbb{H}_t = \left\{ \mathbf{H}_t \left| \mathbf{1}'_k \mathbf{H}_t = \mathbf{n}'_t,\ \mathbf{H}_t \mathbf{1}_k = \mathbf{n}_{t-1} \right. \right\}, \tag{4}$$

with $\mathbf{1}_k$ being a $(k \times 1)$ vector of ones. The set of matrices represented by $\mathbb{H}_t$ is the collection of all conceptually possible outcomes of between-states transition numbers when moving from observed state distribution $\mathbf{n}_{t-1}$ in time $t-1$ to the observed state distribution $\mathbf{n}_t$ in time t. With micro data available we observe that some transitions have occurred at the micro level. Let $\mathbf{N}_t^*$ denote the micro data i.e. a matrix of observed transitions with n_{ijt}^* being the number of state *i*-type units in time $t-1$ that we observed to be state *j*-type unites in time t. The likelihood of the event of moving from $\mathbf{n}_{t-1}$ to $\mathbf{n}_t$ changes given that certain ways of transitioning to achieve $\mathbf{n}_t$ are ruled out by the $\mathbf{N}_t^*$ observations. Particularly, the set of all possible combination is now defined as

$$\mathbb{H}_t^* = \left\{ \mathbf{H}_t : \mathbf{1}'_s \mathbf{H}_t = \mathbf{n}'_t, \mathbf{H}_t \mathbf{1}_s = \mathbf{n}_{t-1} \text{ and } \mathbf{H}_t \geq \mathbf{N}_t^* \right\} \tag{5}$$

such that the likelihood becomes

$$L\left(\boldsymbol{\beta} \middle| \mathbf{n}_0, \mathbf{n}_1, ..., \mathbf{n}_T; \mathbf{N}_1^*, ..., \mathbf{N}_T^* \right) = \prod_{t=1}^{T} \sum_{\mathbf{H}_t \in \mathbb{H}_t^*} \prod_{i=1}^{k} \left(n_{i,t-1}! \right) \left(\prod_{j=1}^{k} P_{ijt}^{\eta_{ijt}} / \eta_{ijt}! \right) \tag{6}$$

The number of elements in set $\mathbb{H}_t$ or $\mathbb{H}_t^*$ increases exponentially with the number of states, making the implementation of expression (3) or (6) for larger samples challenging (or intractable) from a computational point of view. For example, in the case of only three states and 200 observations, there are over 2.5 million combinations of (3×3)-matrices possible if approximately the same number of individuals reside in each of the three states. For the unconditional likelihood (3) this dimensionality problem can be approached using a large sample approximation that avoids the computation of the set $\mathbb{H}_t$ (see Hawkes 1969 and Brown and Payne 1986). The large sample approximation used the property that the multinomial distribution can be approximated with a multivariate normal distribution in large samples. In our case each i-th row $\mathbf{H}_{it}$ of $\mathbf{H}_t$ is multinomial with size $n_{i,t-1}$ over $1,...,k$ categories. If $n_{i,t-1}$ is large $\mathbf{H}'_{it}$

is approximately multivariate normal with mean $\boldsymbol{\mu}_i = \mathbf{P}_{it}^{*} n_{i,t-1}$, where $\mathbf{P}_{it}^{*}$ denotes the i row of $\mathbf{P}_t$ without the element of the last column, and covariance matrix $\mathbf{V}_i = n_{i,t-1}\left[diag\left(\mathbf{P}_{it}^{\prime *}\right) - \mathbf{P}_{it}^{\prime *}\mathbf{P}_{it}^{*}\right]$, where $diag\left(\cdot\right)$ denotes a square matrix with the argument vector as the main diagonal and zero off-diagonal elements. Since transitions between observations are independent, each row of $\mathbf{H}_t$ is independent and the probability of $\mathbf{H}_t$ is approximately equal to a multivariate normal random $k(k-1)\times 1$ vector $M_t = \left[\mathbf{H}_{1t}^{*} .... \mathbf{H}_{kt}^{*}\right]^{\prime}$ with mean $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^{\prime} ... \boldsymbol{\mu}_k^{\prime}]^{\prime}$ and variance

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & 0 & \dots & 0 \\ 0 & \mathbf{V}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{V}_k \end{bmatrix}. \tag{7}$$

Defining $B = \left[I_1^{*} ... I_k^{*}\right]$, with $\mathbf{I}_i^{*}$ being an identity matrix of size $k-1$, we have $\mathbf{BM}_t = \mathbf{n}_t$. Using the each linear transformation of a multivariate normal random variable is also multivariate normal it follows that $\mathbf{n}_t$ is multivariate normal with mean $\mathbf{B}\boldsymbol{\mu} = \mathbf{P}_t^{*}\mathbf{n}_{t-1}^{*}$ and variance

$$\mathbf{BVB}^{\prime} = diag\left(\mathbf{P}_t^{\prime *}\mathbf{n}_{t-1}^{*}\right) - \mathbf{P}_t^{\prime *}\mathbf{n}_{t-1}^{*}\mathbf{P}_t^{*} = \boldsymbol{\Gamma}, \tag{8}$$

where $\mathbf{P}_t^{*}$ and $\mathbf{n}_t^{*}$ is equal to $\mathbf{P}_t$ and $\mathbf{n}_t$ without the last column and row, respectively. Therefore, the probability of $\mathbf{n}_t$ given $\mathbf{n}_{t-1}$ can be approximated by a normal density such that $P\left(\mathbf{n}_t \middle| \mathbf{n}_{t-1}\right) \approx \phi\left(\mathbf{n}_t ; \mathbf{P}_t^{*}\mathbf{n}_{t-1}, \boldsymbol{\Gamma}\right)$. From this it follows that (3) can be approximated by a large sample log-likelihood, L_{la}, given by

$$\begin{aligned} &L_a\left(\boldsymbol{\beta} \middle| \mathbf{n}_0, \mathbf{n}_1, ..., \mathbf{n}_T\right) = \\ &\qquad \sum_{t=1}^{T} -0.5\left(\log\left|\boldsymbol{\Gamma}_t\right| + \left(\mathbf{n}_t^{*} - \mathbf{P}_t^{*\prime}\mathbf{n}_{t-1}\right)^{\prime}\left(\boldsymbol{\Gamma}_t\right)^{-1}\left(\mathbf{n}_t^{*} - \mathbf{P}_t^{*\prime}\mathbf{n}_{t-1}\right)\right). \end{aligned} \tag{9}$$

When considering the micro observations, $\mathbf{H}_{it}$ is still multinomial with size $n_{i,t-1}$ over $1,...,k$ categories except that the constraint $\mathbf{H}_{it} \geq \mathbf{N}_{it}^{*}$ need to be considered. As argued above the approach is intended for situation in which the micro data is only available for a fraction of the observation in the macro data. In these situations the limits imposed by $\mathbf{H}_{it} \geq \mathbf{N}_{it}^{*}$ are hardly binding such that $\mathbf{H}_t$ can still be approximated by a multivariate normal. From this it follows that the large sample log-likelihood approximation in (9) remains valid for (6). The validity of this large sample approximation is assessed in the Monte Carlo simulations considering different sizes of the micro sample.

The specification of the prior density $p(\boldsymbol{\beta})$, considers the underlying sampling distribution of the micro observations. Recall that n_{it} is the number of individuals that were in state i at time t, let $\mathbf{X}_t^i$ be the vector of shares across states in t for individuals who were in state i in $t-1$, and let $\mathbf{P}_{it}$ be the i-th row of $\mathbf{P}_t$. The propensity of each individual in the micro sample to transition between states is in accordance with the appropriate elements of $\mathbf{P}_t$. Analogous to the case of macro data, the distribution across states in t of individuals who were in state i in $t-1$ is multinomial around mean $\mathbf{P}_{it}$ with size n_{it}. The observed number of individuals in each of the k states in t, n_{it}, $i=1,...,k$, is then the corresponding weighted sum of vectors $\mathbf{X}_t^i$, $i=1,...,k$. Therefore, the prior density can be represented as a likelihood similar to (3), except that now information about the individual transitions n_{ijt} is available, making the summation over the set $\mathbb{H}_t$ unnecessary because the actual transitions are observed. Hence the likelihood simplifies to

$$p(\boldsymbol{\beta}) = L(\boldsymbol{\beta} | \mathbf{N}_1,...,\mathbf{N}_T) = \prod_{t=1}^{T} \prod_{i=1}^{k} (n_{i,t-1}!) \left(\prod_{j=1}^{k} \mathbf{P}_{ijt}^{\,n_{ijt}} / n_{ijt}! \right), \tag{10}$$

where the $(k \times k)$-matrix $\mathbf{N}_t$ has elements n_{ijt} representing the number of individuals that transition from state i at $t-1$ to j in t. We emphasize that for the case of aggregated data discussed above, the distribution of $\mathbf{n}_t$ differs between imperfect and perfect observations, while for micro observations, this distinction does not apply. In the latter case, the distribution of $\mathbf{x}_t$ is fully characterized by $\mathbf{x}_{t-1}$ regardless of whether a sample or the entire population is observed. The fundamental difference is that in the case of micro observations, individuals in the sample in time period t are all the same as in $t-1$ which is usually not the case for imperfect macro data. Consequently, information earlier than $\mathbf{x}_{t-1}$ contains no additional information about $\mathbf{x}_t$.

Computational Implementation

In order to conduct inference in the model depicted above, integrating and/or taking expectations based on the posterior density

$h(\boldsymbol{\beta}|\mathbf{d}) \propto L(\boldsymbol{\beta}|\mathbf{n}_0, \mathbf{n}_1, ..., \mathbf{n}_T)\, p(\boldsymbol{\beta})$ or on its approximation $h_a(\boldsymbol{\beta}|\mathbf{d}) \propto L_a(\boldsymbol{\beta}|\mathbf{n}_0, \mathbf{n}_1, ..., \mathbf{n}_T)\, p(\boldsymbol{\beta})$ is required. An analytical approach to such computations is generally intractable. Instead a Monte Carlo integration approach is implemented based on sampling from the posterior density via a Metropolis Hastings (MH) algorithm.[12] For our purposes, we evaluate the optimal Bayesian estimator under quadratic loss, the posterior mean, by calculating the mean of an *iid* sample from $h(\boldsymbol{\beta}|\mathbf{d})$ for sufficiently large sample sizes.

Specifically, a random walk MH algorithm with a multivariate normal generating density is employed.[13] The variance of the proposal density is adjusted such that an acceptance rate in the interval $[.2, .3]$ is obtained. In cases where the number of parameters to be estimated is large, a "Block-at-a-Time" algorithm proposed by Chib and Greenberg (1995) is employed in which the parameters to be estimated are divided into blocks.

2.3 Monte Carlo Simulation of Prior Information Effects

In this section we analyze the influence of prior information, in the form of a sample of micro observations, on the posterior distribution and associated estimators' performance as well as on the behavior of the sampling algorithm. Based on an underlying population of $n_{ind} = 10{,}000$ individuals, four different scenarios are considered regarding the availability of prior information, including a case of no micro observations, and micro samples of *n* = 100, 500, and 1000. The scenarios are further distinguished by the number of Markov states ($k = 3, 4, 5$). Data is generated for $T = 100$ time periods and $n_z = 6$ explanatory variables including a constant. All simulations are undertaken for a Markov

[12] An interesting alterative to the simple random walk MH sample would be the development of a data augmentation sample algorithm, in the spirit of Albert and Chib (1993), for a non-stationary Markov model using aggregated data. Our first implementation of such an algorithm, building on Musalem et al. (2009) who proposed a concept to consider aggregated data in an simple ordered logit model, suffered, however, form slow convergence problems. Convergence problems are known for the Albert and Chib (1993) algorithm and could be overcome using alternatives such as those proposed by Frühwirth-Schnatter and Frühwirth (2007) or Scott (2011). These algorithms, however, focus on simple multinomial logit models and are not directly transferable to the Markov case using aggregated data.

[13] To mitigate computer overflow problems the Metropolis acceptance ration is calculated as $\alpha\left(\boldsymbol{\beta}^{(r)}, \boldsymbol{\beta}^{can}\right) = \min\left[\exp\left(\ln h\left(\boldsymbol{\beta}^{can}|\mathbf{d}\right) - \ln h\left(\boldsymbol{\beta}^{(r)}|\mathbf{d}\right)\right), 1\right]$.

model based on either the multinomial logit specification or the ordered logit specification discussed above, and are performed using Aptech's GAUSS™ 11.

Data Generating Process

The data generating process distinguishes between the two different behavioral models, based on the multinomial logit and ordered logit specification discussed above. In both cases the parameterization is chosen so that the deterministic part constitutes roughly one third of the model's total variation. Furthermore, in both cases n_{ind} individuals are considered that transition over time between the k states in accordance with the underlying behavioral model. The initial state of each individual in $t=1$ is randomly chosen with probability equal to $u_i \; \forall\, i=1,...,k$, where the probability is the same for all individuals and given by $u_i = \tilde{u}_i \Big/ \sum_{h=1}^{k} \tilde{u}_h$ with $\tilde{u}_i \sim iid\ \mathcal{U}(0,1)$, where $\mathcal{U}(a,b)$ denotes the continuous uniform distribution on the interval *a* to *b*.

In the multinomial logit model each individual l chooses the state of the next period based on the utility, U_{ijtl}, associated with a specific transition from state i in $t-1$ to j in t. The utility $U_{ijtl}=V_{ijt}+\varepsilon_{ijtl}$ consists of a deterministic part $V_{ijt}=\mathbf{z}'_{t-1}\mathbf{b}_{ij}$ and an individual random part ε_{ijtl} and is generated by drawing the elements of the (lagged) exogenous variables $\mathbf{z}_{t-1}$ from $\mathcal{N}(1,4)$ and the elements of the $(n_z \times 1)$ "true" parameter vectors $\mathbf{b}_{ij}$ from $\mathcal{U}(-1,1)$. Since only differences in utilities are relevant, the parameters of the last alternative are set to zero, $\mathbf{b}_{ik}=\mathbf{0}\; \forall\, i=1,...,k$, in order to identify the model. To obtain a logit model, the ε_{ijtl} are drawn from a Gumbel (type I extreme value) distribution, specified by $F_g(\varepsilon_{ijtl};0,3)=\exp\left(-e^{-\varepsilon_{ijtl}/3}\right)$. In each time period an individual chooses the transition that maximizes utility, moving from state i in $t-1$ to state j in t if $U_{ijtl}=Max\left(U_{i1tl},U_{i2tl},...,U_{iktl}\right)$.

For the ordered logit model, the transition between states is based on a latent index value $Y^*_{itl}=\mathbf{z}'_{t-1}\boldsymbol{\beta}_i+\varepsilon^*_{itl}$ consisting of a deterministic part $\mathbf{z}'_{t-1}\boldsymbol{\beta}_i$ and a random part ε^*_{itl}. The index value is generated by drawing the elements of the (lagged) exogenous variables $\mathbf{z}_{t-1}$ from $\mathcal{N}(1,4)$ and the elements of the $(n_z \times 1)$ *true* parameter vectors $\boldsymbol{\beta}_i$ from $\mathcal{U}(-1,1)$. The random errors ε^*_{itl} are *iid* random draws from a logistic distribution, specified by

$F_l\left(\varepsilon_{itl}^*;0,2.3\right)=\left(1+\exp\left(-\varepsilon_{itl}^*/2.3\right)\right)^{-1}$. The latent index value determines the outcome of Y_{itl} for each individual in each time period according to (2).

Using the above sampling design a micro dataset for n_{ind} individuals and T time periods is obtained for both the multinomial logit and the ordered logit specification, and represents the full population of individuals under study. For the specification of the prior density, random samples of size 100, 500, and 1000 are drawn without replacement from these micro datasets. The population is transformed into macro datasets by summing up the number of individuals in each state in each time period.

In order to avoid dependency of the results on a specific set of parameters, $n_{true}=10$ *true* models are generated using the data generating process. For each of the n_{true} true models the process is repeated $n_{rep}=20$ times with the same parameters, but with new draws of the random errors ε_{ijtl} or ε_{itl}^* in each repetition.

Performance Measures

The influence of prior information is assessed by a comparison of measures characterizing features of the posterior density, including performance of the posterior mean of the density, representing the minimum quadratic risk estimate of β. The effect of prior information on the numerical stability of the sampling algorithm is also analyzed. For the Monte Carlo simulation a fixed burn-in period and a fixed sample size is employed for the MH sampler. Even though appropriate burn-in periods and sample sizes are found using graphical measures in trial runs for each scenario and resulted in substantially large burn-in periods, it still cannot be guaranteed that the MH sample will converge correctly for every simulation run. Therefore, Box-Whisker-Plots are employed to detect outliers among the sum of squared errors of the $n_{true}n_{rep}$ simulations as an indication that the MH sample had not converged appropriately. Measures characterizing the posterior density and performance measures relating to the estimator are then calculated based on only those runs that were not designated as outliers.

The effect of prior information on the spread of the posterior is assessed based on posterior variances, and is calculated on the basis of the posterior sample outcomes. The total variance of the posterior density is calculated by summing over the posterior variances of all n_z parameters for each run, and then the mean over all $n_{true}n_{rep}$ simulation runs (outliers excluded) is calculated to obtain one scalar value measure of the total variance.

The analysis of the influence of prior information on the Bayes estimator (posterior mean) is based on the mean square error (MSE) criterion, calculated as the mean sum of squared errors between estimated and true parameter values, where the mean is calculated over all of the $n_{true}n_{rep}$ simulation runs not detected as outliers. The MSE is further decomposed into variance and bias components, where the squared bias is again summed over all parameters. The distribution of the sum of squared errors together with the number of outliers detected for each scenario provides an assessment of the numerical stability of the MH sampler, and the effects of prior information on that numerical stability.

Results of the Monte Carlo Simulation

Results for the multinomial logit model of a Monte Carlo simulation to analyze the influence of prior information, in the form of a micro sample, on the posterior and the posterior mean estimator. Results indicate that prior information reduces the variance of the posterior and improves the performance of the mean posterior estimate in terms of the MSE.

The results of the Monte Carlo Simulations for the multinomial logit model are presented in figure 2.1. Results show that incorporating prior information in the form of a micro sample decreases the total variance of the posterior density, and more so the larger the micro sample. The variance reduction effect of prior information becomes even more pronounced the greater the number of Markov states being considered. Similarly, prior information decreases the MSE of the estimator, and a greater number of Markov states accentuate this effect. Decomposing the MSE into bias and variance suggests that the MSE is primarily determined by the variance of the estimator. In all scenarios the share of the squared bias is only 4 to 9 % of total MSE.

The distribution of the sum of squared errors, as depicted in the Box-Whisker-Plots in figure 2.1, provides information about the numerical performance of the MH sampling algorithm. Results show that more simulation runs are detected as outliers in the no prior information scenario (i.e. micro sample with 0 obs.), especially when considering $k=4$ or $k=5$ Markov states. This observation indicates problems relating to the numerical stability of the MH sampler, in the sense that the algorithm does not converge correctly for some simulation runs. When considering a micro sample as prior information, substantially fewer simulation runs are detected as outliers, indicating that the use of prior information improves the numerical stability of MH sampler.

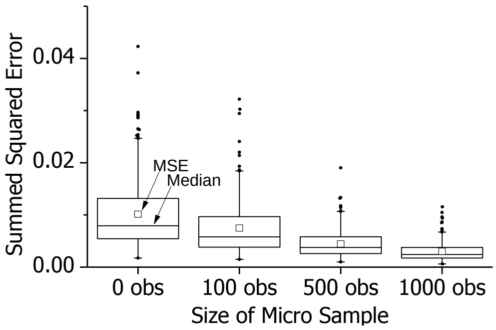
Comparable results are obtained for the ordered logit model as depicted in figure 2.2. Similar to the multinomial logit simulation, results indicate that prior information reduces the variance of the posterior density, and more so the larger the micro sample considered. The same can be observed for the MSE, which decreases with increasing micro sample size. If prior information is considered the MSE is mainly determined by the variance of the estimator such that the share of the squared bias is only 4 to 6 % of total MSE in all scenarios. For the no prior information scenarios, however, the bias share is substantially larger, being between 23 and 28 %.

The number of outliers detected by the Box-Whisker-Plots is used again to assess the numerical stability of the MH sampler. The results are consistent with the findings in the multinomial logit case, where performance of the MH sampler improves the larger the micro sample size considered as prior information. It is worth noting that the numerical problems in cases without prior information persist in the ordered logit model compared to the multinomial logit model even though substantially fewer coefficients need to be estimated (e.g. 25 compared to 120 for $k=5$).

Overall the results suggest that without prior information, alternative individualized sampling strategies or extensions of the simple MH sampler (e.g. Parallel Tempering (Liu 2008) or Multiple Try Method (Liu et al. 2000)) should be considered for successful sampling from the posterior, which could not be automated for the Monte Carlo simulations. This suggests that through prior information, the computational demands with respect to the sampling algorithm are reduced and that more precise estimation can be achieved with the simple MH sampler in both the multinomial and the ordered logit model with a moderately sized micro sample. The Monte Carlo results also show that despite the fact that the large sample approximation does not explicitly consider the conditioning of the macro data based likelihood on the micro data still leads to an improvement of the performance of the estimator.

Figure 2.1 Results for the multinomial logit model of a Monte Carlo simulation to analyze the influence of prior information, in the form of a micro sample, on the posterior and the posterior mean estimator.

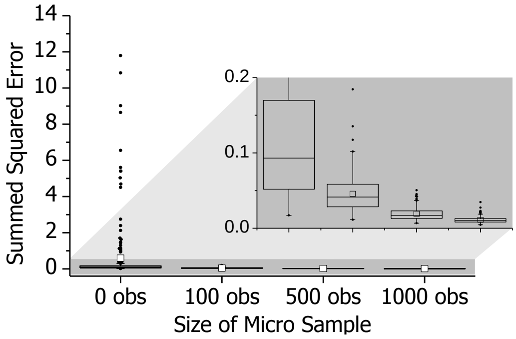
Number of Markov states: k=3



Measures	Size of micro sample 0	100	500	1000
Variance[a], Posterior	0.00592	0.00496	0.00312	0.00219
MSE[a], Estimator	0.00892	0.00672	0.00414	0.00275
Sq. Bias[a], Estimator	0.00042	0.00041	0.00024	0.00014
Outlier[b]	12	9	7	9

Sample: 50,000; Burn-In: 100,000; Blocks: 1;
σ: 1/800; num. o coef.: 36

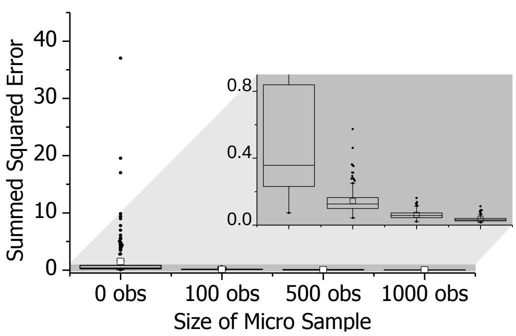
k=4



Measures	Sizeof micro sample 0	100	500	1000
Variance[a], Posterior	0.03585	0.02433	0.01340	0.00891
MSE[a], Estimator	0.09394	0.04425	0.01802	0.0106
Sq. Bias[a], Estimator	0.0080	0.021	0.00108	0.0045
Outlier[b]	34	0	9	11

Sample: 100,000; Burn-In: 200,000; Blocks: 1;
σ: 1/870; num. of oef.: 72

k=5



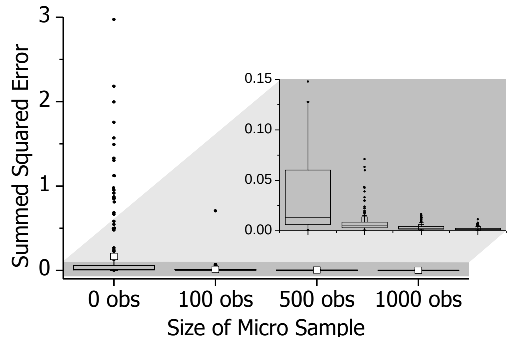
Measures	Size of micro sample 0	100	500	1000
Variance[a], Posterior	0.18702	0.10570	0.04839	0.02992
MSE[a], Estimator	.392	0.13130	0.0586	0.03296
Sq. Bias[a], Estimator	0.03678	0.00758	0.00336	0.00179
Outlier[b]	3	1	7	9

Sample: 250,000; Burn-In: 500,000; Blocks: 2;
σ: 1/580; num. of coef.: 120

[a] Calculated without simulation runs detected as outliers. [b] Note that due to the illustration the number of outliers cannot be derived from the figures directly. Source: Own estimations.

Figure 2.2 Results for the ordered logit model of a Monte Carlo simulation to analyze the influence of prior information, in the form of a micro sample, on the posterior and the posterior mean estimator

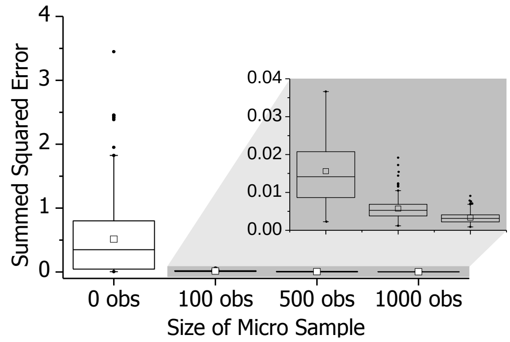
Number of Markov states: k=3



	Size of micro sample			
Measures	0	100	500	1000
Variance[a], Posterior	0.00557	0.00352	0.00197	0.00139
MSE[a], Estimator	0.02149	0.00530	0.00290	0.00168
Sq. Bias[a], Estimator	0.00532	0.00022	0.00018	0.00008
Outlier	33	18	14	15

Sample: 20,000; Burn-In: 50,000; Blocks: 1; σ: 1/730; num. of coef.: 21

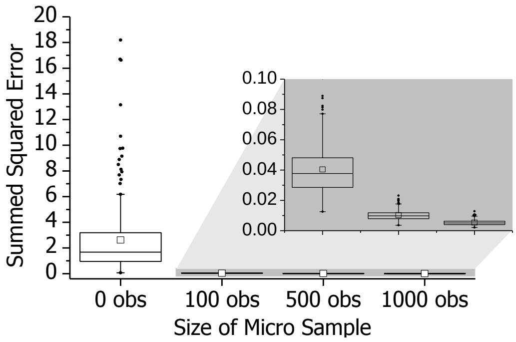
k=4



	Size of micro sample			
Meures	0	10	500	1000
Variance[a] Posterior	0.02031	0.00906	0.00435	0.00274
MSE[a], Estimator	0.43143	0.01457	0.00526	0.00320
Sq. Bias[a], Estimator	0.11923	0.00069	0.00023	0.00020
Outlier	8	5	11	5

Sample: 30,000; Burn-In: 70,000; Blocks: 1; σ: 1/700; num. of coef.: 32

k=5



	Size of micro sample			
Measures	0	10	500	1000
Variance[a], Posterior	0.05809	0.02361	0.00851	0.00483
MSE[a], Estimator	1.97938	0.03864	0.00979	0.00516
Sq. Bias[a], Estimator	0.45399	0.00196	0.00047	0.00029
Outlier	15	7	9	7

Sample: 50,000; Burn-In: 100,000; Blocks: 1; σ: 1/750; num. of coef.: 45

[a] *Calculated without simulation runs detected as outliers.* [b] *Note that due to the illustration the number of outliers cannot be derived from the figures directly. Source: Own estimations.*

2.4 Empirical Application: Structural Change in German Farming

The Bayesian estimation framework developed in section 2.2 is used to combine micro and macro data from two different data sources in an empirical analysis of structural change in German farming. The application demonstrates how the approach facilitates estimation of non-stationary TPs in a situation in which estimation with either macro or micro data alone would be substantially debilitated. Further, it illustrates how asynchronous data, in this case consisting of yearly micro data and macro data available only every two to three years, can be consistently combined in estimation. The application provides an alternative inferential approach to Zimmermann and Heckelei (2012) mentioned in section 1, who were the first to consider using the same data sources to analyze farm structural change, using a generalized cross entropy approach to estimation.

Both the multinomial logit and the ordered logit model of the TPs are applied to provide two different perspectives on the evolution of structural change. The multinomial logit model is applied in an analysis of changes in farm specialization (for example, the transition from a crop producing to a milk producing farm). In this case the states constitute five different farm types as well as an entry/exit class (see table 2.1). The entry/exit class is used to represent farms that enter or quit farming. The six states are mutually exclusive, and with the entry/exit class included, are also exhaustive. Since no clear order can be assumed for the farm types, the multinomial logit model is the appropriate

Table 2.1 Definition of farm types and size classes

	State	**Description**
Farm types considered in the multinomial logit model	E/E	Entry/Exit class
	COP crops	Specialist Cereals, Oilseed And Protein Crops; Specialist Granivores
	Other crops	Specialist other field crops; Mixed crops
	Milk	Specialist milk
	Other livestock	Specialist sheep and goats; Specialist cattle
	Mix	Mixed livestock; Mixed crops and livestock
Size classes considered in the ordered logit model	E/E	Entry/Exit class
	Small	16 -< 40 Economic Size Units (ESU)
	Medium	40 -< 100 Economic Size Units (ESU)
	Large	>100 Economic Size Units (ESU)

Note: In the FSS and the FADN farm are classified by type of farming and size classes based on the concept of Standard Gross Margin and Economic Size Units (ESU) (Commission Decision 85/377/ECC and following amendments); Source: Own table specification. The second analysis perspective concerns the transition of farms between an entry/exit class and three classes representing different sizes of operation. Here an ordering (entry/exit, small, medium, large) of the states can be assumed such that the ordered logit model can be applied. The four states are again mutually exclusive and exhaustive.

Sources for Micro and Macro Data

Two different data sources, namely the Farm Structural Survey (FSS) and the Farm Accountancy Data Network (FADN), provide the macro and micro data, respectively. The FSS is a census of all agricultural holdings (above a specific size limit) conducted every two to three years. The available FSS data do not allow tracking an individual farm over time so that only macro data can be derived from the survey. The FADN provides detailed farm level information from a sample of farm holdings on a yearly basis. Using information associated with farms that remained in the sample over several years, micro data on transitions between predefined states can be derived. The advantage of FADN is that it provides more detailed information with a higher temporal resolution compared to the FSS

The stratified sampling plan applied in FADN aims to obtain a sample of farms that encompass different farm types and size classes. However, the sample is not necessarily fully representative of the transitions between these farm types and classes. While the macro data derived from the FSS is less detailed and available only every two to three years, the information that it contains is representative of the entire population. An additional limitation of the micro data derived from the FADN is that no information about entry or exit of farms to or from the sector can be derived. The reason is that no distinction is made between farms that quit farming and farms that are simply not selected by the sampling scheme (the same applies for entry). In contrast, in the FSS data, because the total number of farms in the population is assessed, information about entry and exit can be derived. This is commonly accounted for in Markov-type models by adding a catch-all entry/exit category. The number of farms in this entry/exit class[14], which is

[14] One might also categorize this class as the number of farms that are inactive or that are idle.

unobservable, is defined as a residual between an assumed maximum number of farms (e.g. 20% more than the maximum number of farms observed in any year during the estimation period[15]) and the observed number of farms in the particular year.

Table 2.2 Available FADN and FSS years

Year	**FADN years** (t)	**FSS years** (λ)
1989	0	
1990	1	0
1991	2	
1992	3	
1993	4	1
1994	5	
1995	6	2
1996	7	
1997	8	3
1998	9	
1999	10	
2000	11	4
2001	12	
2002	13	
2003	14	5
2004	15	
2005	16	6
2006	17	
2007	18	7
2008	19	

Source: FADN data base.

[15] The assumed maximum number of farms was chosen ad hoc. Note that this value can be chosen arbitrarily without its value impacting the main results of principal interest. It only influences the absolute size of the TPs in the row of the entry/exit state that are defined in combination with the number of farms in the entry/exit state. The choice of the "20% more than the maximum observed number of farms" could be motivated from a Bayesian perspective by viewing the choice of the maximum number of farms in a hierarchical Bayesian formulation. A uniform prior density between 0 and 40% could be defined to represent prior beliefs about the number of individuals thought to be idle or potential farming entrants. In this instance, since no information about the true maximum number of farms is available in the data the optimal Bayesian estimation under squared error loss would be 20%, equivalent to the mean of the posterior density.

Both datasets are available at a regional level for the entire EU 27. However, the specific example is restricted to seven West German *Laender*[16] for which a relatively long time period is available. Here, FADN data is available from 1989 to 2008 on a yearly base while the FSS data is available from 1990 to 2007 for every two or three years (table 2.2).

Implementation

Estimation of TPs would in principle be possible with either micro or macro data alone. However, each approach would have substantial limitations. If only macro data were used one would need to address the problem that FSS data is only available every two or three years. If only FADN micro data were used no information about entry and exit of farms can be obtained. Only information about transitions between states, conditional on the farm being active and remaining active, can be derived. This is particularly problematic given that the rapid decline of farm numbers is the most obvious pattern of structural change observed in the last decades and hence of central interest. The combination of micro and macro data allows exploiting the advantages of each data source while mitigating their disadvantages. Using the framework delineated in section 2, it is straightforward to analyze both macro data available only every two or three years and yearly micro data in a consistent way. Moreover, it is possible to exploit the information in the macro data concerning entry and exit while using a non-informative prior for the entry/exit transitions.

In consideration of macro data being available only every two to three years, the large sample likelihood function (9) can be adjusted to apply to the available data as

$$L_a\left(\boldsymbol{\beta}\middle|\mathbf{n}_\lambda, \forall\lambda\in\Lambda\right)= \sum_{\lambda\in\{\Lambda-0\}} -0.5\left(\log\left|\boldsymbol{\Gamma}_\lambda\right| + \left(\mathbf{n}_\lambda^* - \boldsymbol{\Pi}_\lambda^{*\prime}\mathbf{n}_{\lambda-1}\right)'\left(\boldsymbol{\Gamma}_\lambda\right)^{-1}\left(\mathbf{n}_\lambda^* - \boldsymbol{\Pi}_\lambda^{*\prime}\mathbf{n}_{\lambda-1}\right)\right), \tag{11}$$

---

[16] Baden-Württemberg, Bavaria, Hesse, Lower Saxony, North Rhine-Westphalia, Rhineland-Palatinate, Schleswig-Holstein

where $\mathbf{n}_\lambda$ denotes the observed macro data in the FSS years $\lambda \in \Lambda$ with Λ being a set of all FSS years for which a pair of sequential observations are available such that $\mathbf{n}_\lambda$ and $\mathbf{n}_{\lambda-1}$ are both observed, $\lambda = 0$ begins the first of the FSS years, and $\lambda - 1$ refers to the FSS year previous to λ (see table 2.2). Further, $\mathbf{\Pi}_\lambda$ represents the TPs between FSS years which are calculated by multiplying the yearly TPs, represented by $\mathbf{P}_t$, accordingly. For example the first TP matrix between FSS years (1990 to 1993) is calculated as $\mathbf{\Pi}_1 = \mathbf{P}_2\mathbf{P}_3\mathbf{P}_4$ and the second (1993 to 1995) is defined by $\mathbf{\Pi}_2 = \mathbf{P}_5\mathbf{P}_6$. The remaining years follow accordingly based on the mapping of FSS and FADN years given in table 2. As we had done previously, $\mathbf{n}^*_\lambda$ represent $\mathbf{n}_\lambda$ without the last row and $\mathbf{\Pi}^*_\lambda$ represent $\mathbf{\Pi}_\lambda$ without the last column. The definition of $\mathbf{\Gamma}_\lambda$ follows from (8) where FADN years (t) are replaced by FSS years (λ) and $\mathbf{P}^*_t$ by $\mathbf{\Pi}^*_\lambda$. A non-informative prior distribution with respect to the entry/exit class, defined as the first state ($k = 1$), is obtained by adjusting (10) to (note the difference for the index i, j)

$$p(\boldsymbol{\beta}) = \prod_{t=1}^{T} \prod_{i=2}^{k} \left(n_{i,t-1}!\right) \left(\prod_{j=2}^{k} \mathbf{P}_{ijt}^{\;n_{ijt}} \,/\, n_{ijt}! \right). \tag{12}$$

For the multinomial and the ordered logit model two different model specifications are chosen. For the multinomial logit model the observations are pooled across different regions. For the ordered logit model, which requires fewer parameters, a fixed effects panel model is estimated by including regional indicator variables for all (except one) regions. Policy indicator variables are used as explanatory variables in both cases to model the effects of major shifts in EU agricultural policy on structural change. Specifically, these variables include an indicator for the *Mac Sherry Reform* in 1993 (zero before 1993, one otherwise), an indicator for the *Agenda 2000* in 2000 and an indicator for the *Mid Term Review* in 2003 in addition to a constant and, in the ordered logit model, the regional indicator variables.[17]

[17] The mean posterior estimator is calculated based on a sample of 100,000 draws from the posterior, after a burn-in-period of 200,000 iterations. The variance of the multivariate normal proposal density is $\left(1/350 \times \mathbf{I}\right)$ and $\left(1/400 \times \mathbf{I}\right)$ which resulted in an acceptance rate of 0.26 and 0.24 for the multinomial logit model and the ordered logit model, respectively.

Results

Table 2.3 provides the estimated TP matrix (averaged over all regions and time periods) between the five farm types and the E/E class obtained from the multinomial logit model. The TP matrix displays a reasonable pattern of magnitudes. As expected we obtain relatively high diagonal elements for the TP matrix, indicating that most farms remain in their current farm type. TPs between the substantially different farm types of crop (*COP crop* and *Other crop*) and livestock (*Milk* and *Other livestock*) enterprises are near zero while higher TPs are obtained for transitions between the two relatively similar crop farm types and the two livestock farm types. Further we observer relatively high TPs between all farm types and the *Mix* farm type which represents farms without one major specialization such that movement to or from any other class is likely if one branch of a farm gains importance.

Table 2.3 Comparison of transition probabilities (TPs) between farm types and between size classes calculated from FADN micro data and estimated TPs using FADN micro and FSS macro data (averaged over all regions and time periods).

Transition probabilities for transition between farm types

	Calculated TP from the FADN micro data						**Estimated TP using FADN micro and FSS macro data**					
	E/E	COP Crop	Other Crop	Milk	Other Livest.	Mix	E/E	COP Crop	Other Crop	Milk	Other Livest.	Mix
E/E	---	---	---	---	---	---	91	2	2	2	2	1
COP Crop	---	84	5	0	0	11	13	74	4	0	0	9
Other Crop	---	6	87	0	0	7	5	3	85	0	0	6
Milk	---	0	0	96	2	2	4	1	0	92	2	2
Other Livest.	---	0	0	14	72	14	9	0	0	14	60	17
Mix	---	6	4	3	2	85	4	4	4	3	3	83

Transition probabilities for transition between size classes

	Calculated TP from the FADN micro data				**Estimated TP using FADN micro and FSS macro data**			
	E/E	Small	Medium	Large	E/E	Small	Medium	Large
E/E	---	---	---	---	90	4	6	0
Small	---	90	10	0	11	85	5	0
Medium	---	5	91	4	0	7	86	7
Large	---	0	9	91	15	0	5	79

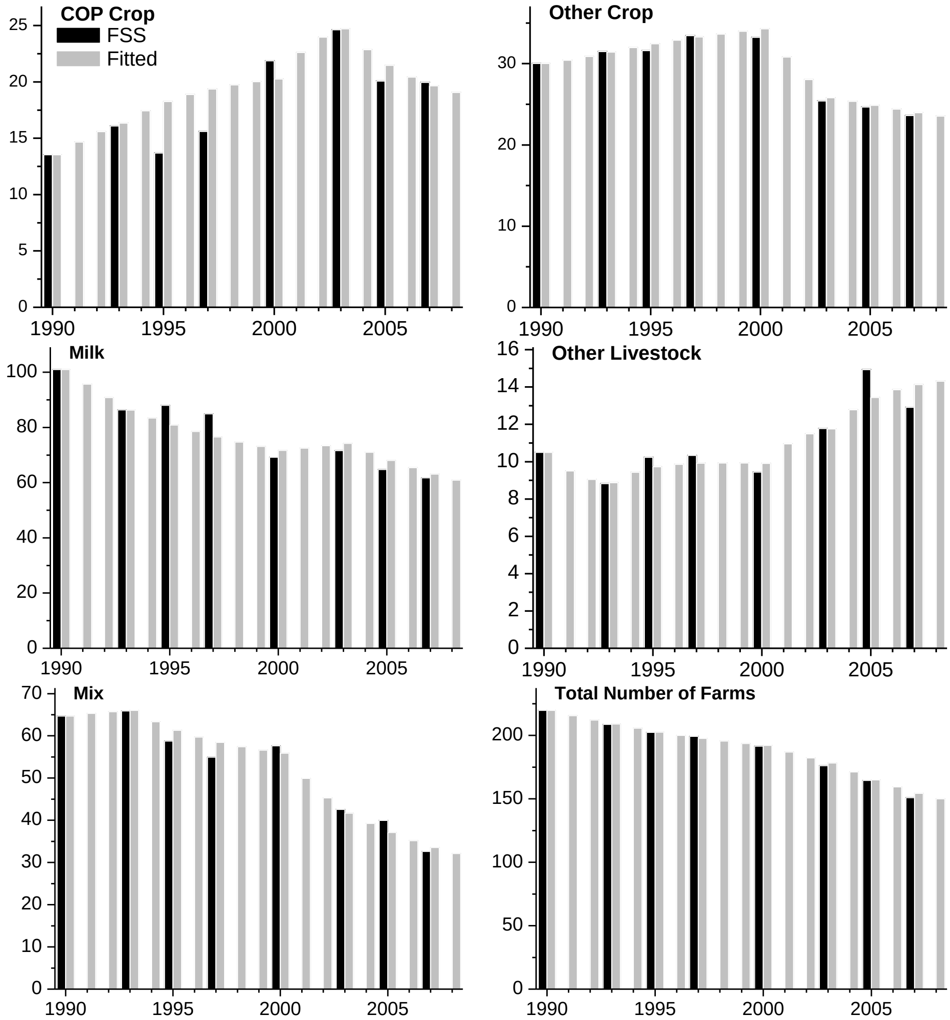
Source: Own estimations.

Comparisons with TP matrices calculated from the FADN micro data illustrates how prior information is updated using the macro data information (upper part of table 2.3). Although the two TP matrices are not directly comparable[18], the general pattern described above is already contained in the calculated TP matrix, which is then updated by the information in the FSS macro data. In addition to the results on the TPs, figure 2.3 provides a comparison between the observed numbers of farms in the FSS years with the yearly fitted values. It suggests that the combination of FSS data with yearly FADN data is well suited to recover the observed farm numbers and to provide yearly estimates for the number of farms between FSS years.

Table 2.3 (lower part) provides a TP matrix for the three size classes and the entry/exit class estimated using the ordered logit Markov approach in comparison to a TP matrix for the three size classes calculated from the FADN micro data (both averaged over all regions and time periods). Again the estimated TPs depict reasonable patterns and indicate how prior information is updated using FSS macro data. As expected, farms are most likely to remain in their current size class or transit to the immediate neighboring one. Farm entry is most likely to happen in the small or medium class and only very rarely in the large size class. Only with respect to farm exit results do not match the intuitive expectation. Naturally one would expect that farm exit rates are highest for small farms and decline for the medium and large class. Estimated exit TP, however, are largest for the large size class followed by the small and the medium size class. This might indicate that results overestimated the true exit rate from the large class while the exit rate from the medium class is underestimated. Nevertheless, the comparison between observed number of farms in the FSS years and the fitted values based on the estimated TPs shows that total exits rates are well matched (figure 2.4).

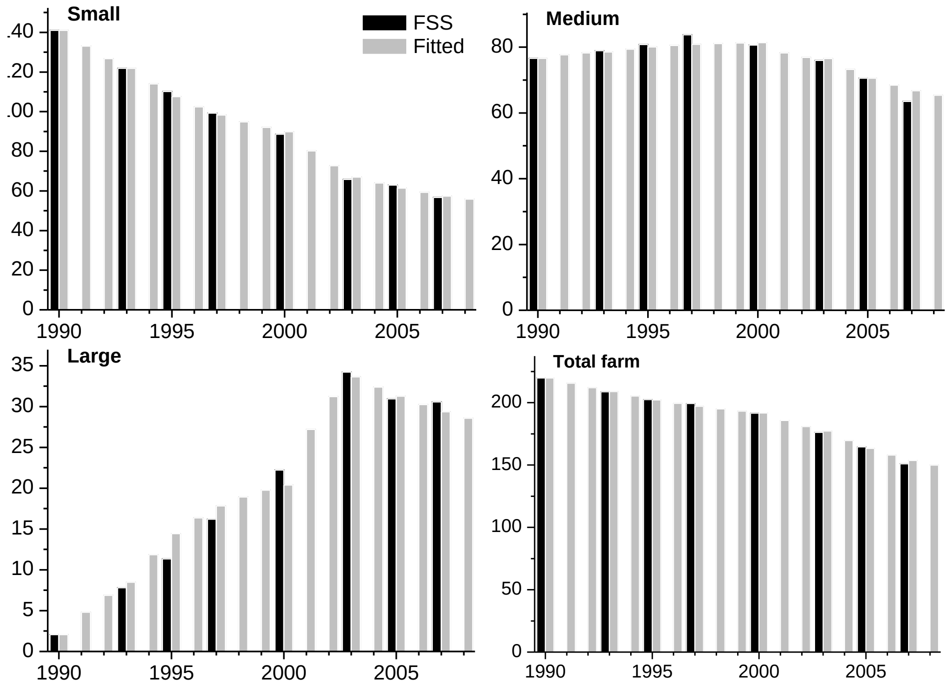
[18] As noted above no information about entry and exit is provided in FADN such that the calculated TP matrix gives the probability that a farm moves to another state conditional on the farm being active before and remaining active.

Figure 2.3 Number of farms (in 1000) observed in the FSS dataset and fitted values of the Markov multinomial logit model. Results aggregated over all considered regions and differentiated between the five different farm types and the total number of farms.



Source: Own calculations.

Figure 2.4 Number of farms (in 1000) observed in the FSS dataset and fitted values of the Markov ordered logit model. Results aggregated over all considered regions and differentiated between the three different size classes and the total number of farms.



Source: Own calculations.

2.5 **Conclusion**

We propose a Bayesian framework for analyzing non-stationary Markov models that allows micro and macro data to be combined in estimation. In contrast to earlier approaches for combining micro and macro data offered in the literature, the Bayesian framework offers a general full posterior information approach for combining micro and macro data-based information on TPs and allows the estimation of functional relationships that link TPs with their determinants. Our Monte Carlo simulations show how prior information, in the form of a micro sample of data, can improve the accuracy of posterior information on the parameters of interest as well as the numerical stability of the estimation approach.

An application of the approach in the context of farm structural change underscored the advantages of the approach in an empirical setting. The combination of micro and macro data based on the proposed framework allows one to take advantage of information in each data set while mitigating the respective disadvantages of using either data set in isolation. Moreover, it was shown that the approach allows combining two dataset with different temporal resolution (yearly micro data in combination with macro data available only every two or three years). In this respect the proposed framework could also be useful for deriving TPs for shorter time intervals (e.g., months) from TPs for longer intervals (e.g., years). Such problems arise in several areas of inquiry such as network theory (Estrada 2009), land use change (Takada et al. 2010), chronic disease analysis (Charitos et al. 2008) or the analysis of credit risk (Jarrow 1997) (see Higham and Lin 2011 for a general discussion of the problem).

The general findings and the proposed approach are subject to some limitations. First, the likelihood specification presented here is applicable for aggregated data observed for the entire population. For other situations alternative likelihood specifications, such as MacRae's (1977) limited information likelihood specification, need to be considered for use in the proposed Bayesian framework. Secondly, the number of model parameters increases with the number of Markov states, often limiting the number of states that can be feasibly considered in empirical applications. The proposed ordered logit approach moderated this problem significantly, but other model specifications based on continuous Markov chains, such Piet (2010), could provide further improvement in this respect.

Overall, this paper contributes to the existing literature by providing an analysis framework that allows for combining micro and macro data information relating to non-stationary Markov models in a way that is consistent with the established tenets of the probability calculus and leads to a minimum loss estimator that is based on full posterior information. The approach is relevant for a broad range of empirical applications in which macro data is available at the population level while micro data is only available for a subsample and one is interested in quantifying the effect of factors that cause individuals to switch between predefined states.

2.6 **References**

Albert JH, Chib S. 1993. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* **88**(422):669–679.

Brown PJ, Payne CD. 1986. Aggregate data, ecological regression, and voting transitions. *Journal of the American Statistical Association* **81**(394):452–460.

Charitos T, de Waal PR., van der Gaag LC. 2008. Computing short-interval transition matrices of a discrete-time Markov chain from partially observed data. *Statistics in Medicine,* **27**(6): 905–921.

Chib S, Greenberg E. 1995. Understanding the Metropolis-Hastings Algorithm. *The American Statistician,* **49**(4): 327–335.

Estrada E. 2009. Information mobility in complex networks. *Physical Review E,* **80(**026104):1-12.

Frühwirth-Schnatter S, Frühwirth R. 2007. Auxiliary mixture sampling with applications to logistic models. *Computational Statistics & Data Analysis* **51**(7):3509–3528.

Hawkes AG. 1969. An Approach to the Analysis of Electoral Swing. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **132**(1):68–79.

Hawkins DL, Han C. 2000. Estimating Transition Probabilities from Aggregate Samples Plus Partial Transition Data. *Biometrics* **56**(3):848–854.

Higham NJ, Lin L. 2011. On pth roots of stochastic matrices. *Linear Algebra and its Applications* **435**(3):448–463.

Jarrow RA. 1997. A Markov model for the term structure of credit risk spreads. *Review of Financial Studies* **10**(2):481–523.

Lancaster GA, Green M, Lane S. 2006. Reducing bias in ecological studies: an evaluation of different methodologies. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169**(4):681–700.

Liu JS, Liang F, Wong WH. 2000. The Multiple-Try Method and Local Optimization in Metropolis Sampling. *Journal of the American Statistical Association* **95**(449):121–134.

Liu JS. 2008. Monte Carlo strategies in scientific computing. New York: Springer.

MacRae EC. 1977. Estimation of Time-Varying Markov Processes with Aggregate Data. *Econometrica* **45**(1):183–198.

McCarthy C, Ryan TM. 1977. Estimates of Voter Transition Probabilities from the British General Elections of 1974. *Journal of the Royal Statistical Society. Series A (General)* **140**(1):78–85.

Musalem A, Bradlow ET, Raju JS. 2009. Bayesian estimation of random-coefficients choice models using aggregate data. *Journal of Applied Econometrics* **24**(3):490–516.

Pelzer B, Eisinga R. 2002. Bayesian estimation of transition probabilities from repeated cross sections. *Statistica Neerlandica* **56**(1):23–33.

Piet L. 2010. A structural approach to the Markov chain model with an application to the commercial French farms. Paper presented: 4èmes journées de recherches en sciences sociales, 9-10.12.2010, Rennes, France.

Rosenqvist G. 1986. Micro and macro data in statistical inference on Markov chains. Dissertation Swedish School of Economics and Business Administration: Helsingfors.

Scott SL. 2011. Data augmentation, frequentist estimation, and the Bayesian analysis of multinomial logit models. *Statistical Papers* **52**(1): 87–109.

Takada T, Miyamoto A, Hasegawa SF. 2010. Derivation of a yearly transition probability matrix for land-use dynamics and its applications. *Landscape Ecology* **25**(4):561–572.

Train KE. 2009. Discrete choice methods with simulation. New York: Cambridge University Press.

Upton GJG. 1978. A Note on the Estimation of Voter Transition Probabilities. *Journal of the Royal Statistical Society, Series A (General)* **141**(4):507–512.

Wakefield J. 2004. Ecological inference for 2×2 tables. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **167**(3):385–445.

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

Zimmermann A, Heckelei T. 2012. Structural Change of European Dairy Farms - A Cross-Regional Analysis. *Journal of Agricultural Economics* **63**(3):576–603.

Chapter 3
Short term prediction of agricultural structural change using FSS and FADN data[19]

Abstract: A Bayesian framework for short term prediction of farm numbers is developed that allows combining two asynchronous data sources in a single estimation. Specifically, the approach allows combining aggregated FSS macro data, available every two to three years, with individual farm level FADN micro data, available at a yearly base. A Bayesian predictive distribution is derived from which point predictions such as mean and other moments can be obtained. The proposed approach is evaluated in an out-of-sample prediction exercise of farm numbers in German regions and compared to linear, geometric and constant predictions. Results show that the proposed approach outperforms the linear and the geometric prediction and performs similar to the prediction of no change. The approach may be used for short term prediction as well as to complete the information within the sampling period.

Keywords: Bayesian prediction, Markov transitions, Asynchronous data, Structural Change

JEL classification: Q19, C11, C53

---

[19] An earlier version of this chapter is part of a project report Gocht A, Röder N, Neuenfeldt S, Storm H, Heckelei T. 2012. Modelling farm structural change: A feasibility study for ex-post modelling utilizing FADN and FSS data in Germany and developing an ex-ante forecast module for the CAPRI farm type layer baseline. JRC Scientific and Policy Reports, 25555 EN.

3.1 Introduction

Detailed up-to-date information about farm structural change, defined as the change in farm size or farm specialization, and the farm structure is of great interest for policy makers and stakeholders and provides the basis for policy analysis.

In the EU two major data sources, namely the Farm Structural Survey (FSS) and the Farm Accountancy Data Network (FADN) provide information at a regional level for all EU member states that can be used for the analysis of farm structural change. In this paper we aim to combine both data sources for a more precise prediction of farm structural change. The developed approach allows completing information on farm numbers in size and specialization classes not available, for the most recent years or for years between FSS years. In this paper the focus is on completing the information for most recent years, however, it should be stressed that the approach may as well be used to complete information between FSS years.

The FSS is a census of all agricultural holdings conducted every ten years with three intermediate sample surveys conducted in-between (Council Regulation (EC) No 1166/2008). FSS data is thus available every two to three years offering aggregated information about the total number of farm holdings in different size or specialization classes[20]. In the following we refer to this aggregated data as *macro data*. On the other hand, FADN data is available on a yearly basis and provides information about individual farms for a sample of farms. Different from FSS, individual farms can be indentified such that it is possible to track the development of one farm in the sample over several years. This type of data allows observing the movement of farm between classes for the analysis of farm structural change and we will refer to it in the following as *micro data*. The sample of FADN farms shall represent all relevant farm types and farm sizes in each region. The corresponding stratified sampling plan usually implies that farms

[20] The individual level (micro) FSS data is processed by the individual member states and typically not accessible for confidentiality reasons, whereas FSS macro data is publically available.

in less common farm types or size classes are overrepresented while farms in common farm types or size classes are underrepresented[21].

Given the shorter intervals with which FADN data is collected and the shorter release time, FADN data is generally the more recent information on farm numbers in classes compared to FSS. Therefore, we might have FADN data for up to three more years after the last available FSS year. The aim of this chapter is to exploit this information together with all other available FADN micro and FSS macro data from previous years to predict farm numbers in size and specialization classes for years after the last FSS year.

This objective addressed in this chapter is motivated by the particular need of the European Commission to predict farm numbers in classes between FSS years. This need resulted in the joint research project "*Modelling the effects of the CAP on farm structural change*" (Contract 151949-2010-A08-DE) from the European Commission Joint Research Centre - Institute for Prospective Technological Studies (IPTS). The work and results presented in this chapter reflect in parts the outcome of this project.

Particularly, in this chapter a prediction framework is developed in which farm structural change, defined as the transition between size classes, is modeled as a non-stationary Markov process. The non-stationary transition probabilities (TP) are estimated using the Bayesian estimation framework developed in chapter two that allows combining micro and macro data in a single estimation.

Methodologically, this paper contributes to the literature in three ways: 1) it allows to consistently combine the bi- or triennial FSS with the yearly FADN data in the estimation of yearly TPs , thereby improving upon previous approaches with data interpolation as in Zimmermann and Heckelei (2012b). The approach developed in chapter two allows merging such asynchronous data sources in a single estimation explicitly reflecting their connection in the data generating

[21] For each sample farm, however, a weight is calculated using the information in FSS about the total number of farms in each farm type, size class and region. With these weights the FADN sample can be aggregated to match FSS results on the population level and information about the total number of farms in each farm type or size class (macro data) can be derived. Even though these macro data can be derived from FADN each year, the weights still reflect only the last available FSS year.

process; 2) A Parallel Tempering (PT) approach (Liu 2008) for sampling from the posterior is implemented replacing the simple Metropolis-Hasting sampler used in the chapter two. The PT sampler converges more reliably than the simple Metropolis-Hasting sampler when faced with a multimodal posterior distribution; 3) The employed Bayesian approach offers a predictive distribution for the number of farms from which point predictions and predictive uncertainty can be derived.

The approach is illustrated and evaluated in an out-of-sample prediction for seven (West) German Regions for which a relatively long sample is available. Farm numbers are predicted for different size classes, with and without differentiation of specialization classes. Specifically, three (economic) size classes and an entry/exit class are considered in four different situations. First, we perform a prediction at an aggregated level where farm numbers in different size classes and the entry/exit class are predicted without any distinction by farm specialization. Then the prediction for the three size classes and entry/exit is repeated at a more disaggregated level for three different farm specializations, namely crop, livestock and mixed farms. In each case, three different time periods are considered in the out-of-sample prediction. The predictions based on the Markov approach are compared to simple constant, linear and geometric predictions of farm numbers.

Even though we choose seven West Germen regions for illustrative purposes, it should be pointed out that the approach can be directly transferred to other EU member states for sufficiently long series of FSS and FADN, currently available in at least the EU-15 member states.

The remaining structure of the paper is as follows: the next section 3.2 develops the estimation and prediction framework and derives an appropriate measure to assess the performance of the Bayesian Markov approach compared to its simple alternatives. Section 3.3 discusses the specific implementation, including the setup of the out-of-sample prediction, the selection of explanatory variables and the implementation of the PT sampling algorithm. Results are presented afterwards followed by conclusions.

3.2 **Method**

Bayesian estimation framework

The number of farms in different classes is modeled as a Markov process. In a Markov process, the movement of individuals between a finite number of predefined, mutually exclusive, and exhaustive states, $i = 1,...,k$, is a stochastic process. In the following we consider a situation in which the states represent an entry/exit and three different farm size classes ($k = 4$). The Markov process is characterized by a $(k \times k)$ transition probability (TP) matrix $\mathbf{P}_t$. The elements P_{ijt} of that matrix give the probability that an individual moves from state i in $t-1$ to j in t. The $(k \times 1)$ vector $\mathbf{n}_t$ denotes the number of individuals in each state i and develops over time according to a first order Markov process

$$\mathbf{n}_t = \mathbf{P}_t' \mathbf{n}_{t-1}. \tag{13}$$

In a non-stationary Markov process the TPs change over time depending on exogenous variables. The specification of the TPs, $\mathbf{P}(\boldsymbol{\beta})$ differs depending on the type of Markov states considered. If we assume that the Markov states do not have an order, the specification is based on the multinomial logit model, whereas an ordered logit model is suitable for our case where transitions between size classes are considered (see chapter two).

For the estimation of the non-stationary TPs a Bayesian estimation framework is employed that allows combining macro and micro data in the estimation of non-stationary Markov TPs. For a detailed description we refer to chapter two. The general idea of the framework is that a macro data based likelihood function is combined with a micro data based prior density. Both likelihood and prior are therefore data based and represent the two different available data sources we aim to combine in a consistent manner. Similarly as in chapter two we will combine FSS macro data, available every two to three years, with the FADN micro data, available at a yearly base (see table 3.1).

The prior density is combined with the likelihood function to a posterior density which is used for deriving the marginal density of individual parameters. Since the required integration is not traceable analytically, Monte Carlo Integration is employed. For this, the simple Metropolis-Hastings (MH) algorithm used in chapter two (section 2.2) to draw a sample from the posterior is replaced by a Parallel Tempering (PT) sampling algorithm (Liu 2008). The general idea of the PT approach is to run multiple copies of the original chain raised to different powers (i.e. temperatures) in parallel and allow exchanges between them. The advantage of the PT approach is that the 'heated' chains (raised to powers smaller one) are able to escape local modes more easily such that it becomes easier to sample from multimodal posterior distributions like those found in the specific application.

In our particular case we adopt the following setup of the PT sampler. We consider I parallel chains with temperatures $1 = T_1 < T_2 < ... < T_I$. The PT sampler consists of parallel and swapping steps. In each parallel step r the current states, $x_1^{(r)}, x_2^{(r)}, ..., x_I^{(r)}$, of all I chains are updated in simple MH steps using a random walk MH sample with a multivariate normal proposal density. After every five

Table 3.1 Available FADN and FSS years

Year	FADN years (t)	FSS years (λ)
1989	0	
1990	1	0
1991	2	
1992	3	
1993	4	1
1994	5	
1995	6	2
1996	7	
1997	8	3
1998	9	
1999	10	
2000	11	4
2001	12	
2002	13	
2003	14	5
2004	15	
2005	16	6
2006	17	
2007	18	7
2008	19	

Source: FADN data base.

parallel steps a swapping step is conducted, in which a swap between all neighboring chains is proposed. Denoting neighboring chains as i and $i+1$, a swap of states $x_i^{(r)}$ and $x_{i+1}^{(r)}$ is accepted with probability

$$\min\left\{1, \exp\left(\pi\left(x_i^{(r)}\right) - \pi\left(x_{i+1}^{(r)}\right)\right)^{T_i - T_{i+1}}\right\}, \tag{14}$$

where $\pi\left(x_i^{(r)}\right)$ denotes the log posterior density evaluated at state $x_i^{(r)}$ of chain i. Swaps are first considered for the last two chains and then going back in steps to the first two neighboring chains. With such a setup it is generally possible that the state of the last chain, $x_I^{(r)}$, is swapped to the first chains within one pass through all neighboring chains. This setup was found to be more efficient in our specific application compared to the approach proposed by Liu (2008) in which only one pair of neighbors are selected at random to swap states in each swap step.

The performance of the PT tempering crucially depends on the chosen number of parallel chains, I, as well as on the chosen temperatures $1 = T_1 < T_2 < ... < T_I$ and the covariance matrices of the multivariate normal proposal densities to be selected for each specific sampling. The temperatures require to be chosen such that a sufficiently large temperature range is covered and the hottest chain can easily escape local modes. On the other hand, the differences between neighboring chains' temperatures need to be small enough such that a sufficient amount of swaps are accepted. The specific implementation of the PT approach is described in section 3.3.

Prediction methods

The Markov process specified in (13) may be directly used for prediction of farm number in different states. The number of farms in k states in the last observed year t is denoted by a $(k \times 1)$ vector $\mathbf{n}_t$. Our aim is to predict farm numbers $\hat{\mathbf{N}} = \hat{\mathbf{n}}_{t+1}, ..., \hat{\mathbf{n}}_{t+\tau}$ in k states for τ years starting from the last observed year t. Taken the TPs $\mathbf{P} = \left(\mathbf{P}_{t+1}, ..., \mathbf{P}_{t+\tau}\right)$ as given, prediction to $t+\tau$ follows directly from (13) by

$$\hat{\mathbf{n}}_{t+\tau} = \left(\prod_{j=t+1}^{t+\tau} \mathbf{P}_j\right)' \mathbf{n}_t \,. \tag{15}$$

With (15) the predicted farm number $\hat{\mathbf{N}}$ are thus a function of the TPs, $\mathbf{P}$. The TPs $\mathbf{P}$ are itself a function of the unknown parameter $\boldsymbol{\beta}$, thus we can write $\hat{\mathbf{N}} = \hat{\mathbf{N}}(\mathbf{P}(\boldsymbol{\beta})) = \hat{\mathbf{N}}(\boldsymbol{\beta})$. The specification of the functional relationship $\mathbf{P}(\boldsymbol{\beta})$ is based the ordered logit specification (see chapter two).

The Bayesian estimation framework provides several ways of how the prediction may be implemented. One possibility is to derive point estimates of $\boldsymbol{\beta}$ such as the posterior mean, which is the optimal Bayesian estimator under squared error loss. Here we employ an alternative prediction strategy directly using the sample outcomes of the joint posterior of $\boldsymbol{\beta}$. This provides the advantage that a complete Bayesian predictive distribution is derived for each state and year in an intuitive and straightforward way. Technically, each sample outcome $\boldsymbol{\beta}_{(l)}$, $l = 1,...,L$ from the posterior is used to predict farm numbers based on (15) obtaining a sample of predictions $\hat{\mathbf{N}}_{(l)} = \hat{\mathbf{N}}(\boldsymbol{\beta}_{(l)})$. This sample can be regarded as a sample from the predictive distribution $f\left(\hat{\mathbf{N}} \middle| \mathbf{d}\right) = \hat{\mathbf{N}}(\boldsymbol{\beta}_{(l)})h(\boldsymbol{\beta}|\mathbf{d})$. The predictive distribution may itself be the final result or alternatively the mean, variance and the quintiles of the predictive distribution may be calculated from the sample.

Prediction measures

The prediction quality of the described approach is compared to the simple linear, constant and geometric prediction based on the Mean Absolute Scaled Error (MASE). The MASE is proposed by Hyndman and Koehler (2006) who argue that the MASE is superior to other commonly used forecast measures such as the (Root) Mean Square Error (which is not scale free) measures based on relative errors, such as the Mean Relative Absolute Error, or relatives measures, such as the relative Mean Absolute Error. The MASE has a clear interpretation, is scale free and defined in all relevant situations (not defined only in the irrelevant case where historical data shows no variation). It is calculated by dividing the absolute prediction error $e_t = \left|\hat{Y}_t - Y_t\right|$, where $\hat{Y}_t$ is a prediction of Y_t, by the average one-step naive forecast in the sample period,

$$MASE = mean \frac{e_i}{\frac{1}{n-1}\sum_{i=2}^{n}\left|Y_t - Y_{t-1}\right|}. \tag{16}$$

Therefore, a MASE less than one indicates a better prediction than the average one-step naive forecast within-sample. In our specific case the MASE is calculated for the predictions of farm numbers in $t+\tau$ over all regions and size classes, without considering the artificial E/E. The average one-step naive forecast is calculated over all observed FSS years. This needs to be considered in the interpretation of the absolute size of the MASE since the step-length of the out-of-sample prediction might differ from the step length of the naive one-step forecast (two or three years). It is, however, irrelevant for a relative comparison of the MASE between different prediction methods being the primary purpose of the out-of-sample prediction.

3.3 **Implementation**

Setup of out-of-sample prediction

In the out-of-sample prediction, farm numbers are predicted for different size and specialization classes. The classification of farms is based on the one in FSS and FADN with size and specialization classes based on their economic size and the relative importance of different production activities (Commission Decision 85/377/EEC). The physical units of production (hectare or livestock units) are valued by the corresponding Standard Gross Margins (SGM) calculated for each region on a regular basis by the member states. The sum of all production activities valued by the SGM determines the economic size of a farm, expressed in Economic Size Units (ESU), while the share of each production activity on total ESU determines the farm specialization.[22]

In the out-of-sample prediction, four different situations are distinguished. On the one hand the prediction is performed for all farms (excluding horticulture and permanent crops *TF14*: 20, 31, 32, 33, 34) irrespectively of their farm type. Additionally, the prediction is repeated for three different farm specializations,

[22] From the accounting year 2010, the typology for agricultural holdings is based on Standard Output (Commission Regulation (EC) No. 1242/2008) instead of SGM. The main differences among the SGM and SO is that the SO excludes direct payments and the cost of variable inputs. Moreover, the unit used to measure SO is the Euro and not the Economic Size Unit (1.200 Euro). The change will have no effect on the general applicability of the proposed prediction approach.

namely crop farms (*TF14*: 12, 14, 60), livestock farms (*TF14*: 41, 44, 45, 50, 70) and mixed farms (*TF14*: 80). In each of the four cases, three different size classes (small < 40ESU, medium <100ESU and large >100ESU) and an entry/exit class are considered. The entry/exit class is an artificial class required by the Markov approach and representing farms that enter or quit farming (Stokes 2006).

For each of the four cases, three different out-of-sample prediction periods are considered. In each prediction period the last FSS year is excluded from estimation and macro data instead predicted for this year. The prediction can then be compared to the observed macro data. By considering three different time periods, each time excluding an additional FSS year in estimation, it can be evaluated how the approach would have performed in previous periods. Table 3.2 presents the different prediction periods and the corresponding FADN and FSS data used.

For each individual prediction, a panel of seven West-German regions is considered in estimation (FADN regional codes: 10 (Schleswig-Holstein), 30 (Lower Saxony, 50 (North Rhine-Westphalia), 60 (Hesse), 70 (Rhineland-Palatinate), 80 (Baden-Württemberg) and 90 (Bavaria)).

These 12 different Bayesian Markov predictions (three time periods for each of the four cases) are compared to a constant, linear and geometric prediction. The linear prediction employs a least squares estimation of γ_1 and γ_2 of the linear function $n_t = \gamma_1 + \gamma_2 t + \varepsilon_t$, where n_t is the number of farms in time t. Using the estimates $\hat{\gamma}_1$ and $\hat{\gamma}_2$, farm numbers for $t+1$ are then predicted by $\hat{n}_{t+1} = \hat{\gamma}_1 + \hat{\gamma}_2 (t+1)$ and for the following years accordingly. For the estimation

Table 3.2: Out-of-sample prediction periods and corresponding data considered for estimation

Prediction period	FADN data considered in estimation	FSS data considered in estimation
2000-2003	1989-2003	1989-2000
2003-2005	1989-2005	1989-2003
2005-2007	1989-2007	1989-2005

Source: Own table.

only FSS macro data is employed. The geometric growth rate is derived by a least squares estimation of $\ln(n_t) = \lambda_1 + \lambda_2 t + \varepsilon_t$. Farm numbers in $t+1$ are predicted using the estimated parameters $\hat{\lambda}_1$ and $\hat{\lambda}_2$ to calculate $\hat{n}_{t+1} = e^{(\hat{\lambda}_1 + \hat{\lambda}_2 (t+1))}$. Data source and time periods are the same as those used for the linear prediction. An advantage of the geometric over the linear prediction is that predicted farm numbers cannot become negative. Problems arise, however, in the geometric prediction in cases in which no farms are observed in a particular time period. In these cases, the dependent variable is not defined, and we omit the observation from the estimation. The constant prediction assumes that farm numbers do not change during the prediction period, such that the predicted value is equal to the last observed value for each farm type and region.

Identification of potential explanatory variables

To select a set of explanatory variables for the estimation of the non-stationary TPs, first a set of factors that potentially drive farm structural change are identified based on theoretical considerations and the literature analyzing factors influencing farm structural change (Breustedt and Glauben 2007; Zimmermann et al. 2009; Piet et al. 2012; Zimmermann and Heckelei 2012a; Zimmermann and Heckelei 2012b). The identified factors may broadly be categorized in six general categories, include technology, the initial farm structure, market conditions, natural resource factors, social and demographical factors and agriculture policy (see table 3.3). For each potential factor, specific explanatory variables are identified that allow approximating that factor.

The model is specified as a dummy variable fixed effects model with a regional dummy variables included for each region except one. These dummy variables capture all time invariant factors such as the initial farm structure (farm size/capacity, size heterogeneity), natural conditions (share of absolute grassland, slope, temperature, population density etc.) that remain rather stable over the time period considered. For off-farm employment opportunities the unemployment rate and for the age structure of the farm population the percentage of farmers above 60 years old are considered as explanatory variables. Agricultural policy is considered by three dummy variables indicating major shifts in EU Agricultural Policy in 1993 (MacSharry reform), 2000 (Agenda 2000) and 2003 (Midterm reform).

Technological developments as well as market conditions are represented by standard gross margins (SGM) for different production activities as explanatory variables. SGMs are provided by EuroStat (Commission Decision 85/377/EEC) at regional level for all relevant production activities and member states. SGMs are calculated by member states based on a period of several years to reduce the effects of short term price or yield fluctuations. Therefore, SGMs should reflect longer-term changes in productivity as well as in input or output prices that affect the attractiveness of different production activities. For our purpose we aggregated the different individual SGMs into five SGM indices to reflect major production activities in different farm specializations. Specifically, SGMs indices

Table 3.3: Factors identified to potentially influence farm structural change and corresponding explanatory variables

General Category	Factors	Approximated by
Technology	Yields	Index of Standard gross margins (SGM) for different farm specializations. Specialist COP (*SGM13*), Specialist other filed crops (SGM14), Specialist Milk (SGM41), Specialist sheep/goats/cattle (SGMLive) Specialist Grainivores (SGM50) *Source: FADN*
Initial Farm structures	Farm size/capacity	Fix effects
	Size heterogeneity	Fix effects
Market conditions	Input/output prices (price ratios)	SGMs (see Technology)
Natural resource factors	Share of grassland	Fixed effects
	Slope	Fixed effects
	Temperature	Fixed effects
Social and demographical factors	Population density/growth	Fixed effects
	Off-farm income opportunities	Unemployment rate (*Unemp*) *Source: DeStatis*
	Age structure	Percentage of farmers aged above 60 (*Above60*) *Source: FADN*
Agricultural Policy	Agricultural Policy	Dummy variables for mayor policy reforms (MacSharry reform, Agenda 2000 and Midterm reform)

Source: Breustedt and Glauben 2007*;* Zimmermann, Heckelei and Domínguez 2009*;* Piet et al. 2012*;* Zimmermann and Heckelei 2012a*;* Zimmermann and Heckelei 2012b.

are calculated for *Specialist COP (SGM13)*, *Specialist other filed crops (SGM14)*, *Specialist Milk (SGM41)*, *Specialist sheep/goats/cattle (SGM40)* and *Specialist Grainivores (SGM50)*. It is assumed that the SGMs affect transitions of farms between size classes in two different ways. On the one hand SGMs reflect the productivity of production factors in different activities. Hence an increase of the SGM of one specialization should increase the attractiveness of the corresponding farm types. This in turn draws production factors and finally farms into those farm specializations. Also, an increase in the SGM should lead to an increase of the ratio of on-farm to off-farm income possibilities such that farm entries/exits should be increased/decreased. On the other hand, changes in SGMs directly affect the transitions between states because the classification of farms in size classes is based on the SGMs. Therefore, changes in SGMs have a direct effect on the change between classes. An increase in SGM, for example, increases the economic size of a farm even though the physical layout stays the same; hence the farm should move to a higher size class. These two effects, movements in the physical unites as well as in the valuation of each unit, render an interpretation of the causal relationship between SGM and farm structural change problematic but this is irrelevant for the prediction of farm numbers.

The set of explanatory variables is further restricted using the high correlation between individual explanatory variable. Particularly, three SGM indices (*Specialist other filed crops (SGM14)*, *Specialist sheep/goats/cattle (SGM40)* and *Specialist Granivores (SGM50))* are excluded which are highly correlation to the other two SGMs (table 3.4). Even though high correlations among explanatory variables are irrelevant for prediction they add little to the overall explanatory power of the model and are therefore excluded in order to limit the numerical complexity which increase with each additional explanatory variable.

Table 3.4: Correlation matrix of explanatory variables

	SGM13	SGM14	SGM41	SGM40	SGM50	Unemp	Above60
SGM13	1	**0.84**	-0.13	0.25	0.10	0.41	0.36
SGM14		1	0.11	0.49	0.34	0.40	0.40
SGM41			1	**0.85**	**0.82**	0.06	0.10
SGM40				1	**0.92**	0.19	0.35
SGM50					1	0.19	0.23
Unemp						1	-0.16
Above60							1

Source: Own calculation.

Implementation of the Parallel Tempering sampler

For sampling from the posterior we found an implementation of the PT approach using $I = 30$ parallel chains to be suitable for delivering robust sample results. The selection of temperatures and covariance matrixes of the proposal densities requires a substantial amount of manual fine tuning for each individual estimation. Temperatures are chosen such that the swap acceptance rate is above 20% for most of the pairs and at least 2-3% such that swaps between all chains are possible. The covariance matrices of the multivariate normal proposal densities are specified as diagonal matrices with equal variance for all parameters within one chain but different across chains such that an acceptance rate between 20-30% is obtained for most chains. In order to ease convergence of the sample we set a supports for each parameter usually ranging from $[-8,8]$ to avoid that the sampler drifts away and gets stuck in areas of very low density. In individual cases the support is increases when trail runs indicate that a substantial marginal probability is place near the edge of the chosen support of a parameter, such that the final result are not affected by the chosen support. Starting values for all parameters in all chains are drawn randomly from a uniform distribution with the specific support chosen for the parameter. For the final estimation a burn-in period of two million draws and a sample of one million draws are used. Computations are performed using Aptech's GAUSS™ 12 on an Intel® Xeon® E5-2690, where computation time for one estimation is around 1.6 hours using around half of the available CPU.

Table 3.5: Mean, 5% and 95% Quintiles of the marginal posterior density for the first 10 of 64 coefficients estimated in five identical runs using different random starting values. Estimation is for the prediction of crop, livestock and mixed farms combined for the prediction period from 2005 to 2007.

Coef.	1. Run	2. Run	3. Run	4. Run	5. Run
Mean of the marginal posterior density					
1	-3.08	-3.01	-3.08	-3.14	-3.01
2	0.86	0.86	0.87	0.86	0.85
3	3.81	3.75	3.77	3.78	3.79
4	-2.93	-2.89	-2.88	-2.95	-2.89
5	1.02	1.02	1.02	1.03	1.01
6	-0.58	-0.59	-0.59	-0.58	-0.60
7	-0.64	-0.65	-0.64	-0.64	-0.64
8	0.88	0.90	0.89	0.87	0.89
9	1.81	1.75	1.80	1.84	1.76
10	1.58	1.57	1.57	1.58	1.58
5% Quintiles of the marginal posterior density					
1	-3.43	-3.52	-3.49	-3.52	-3.46
2	0.69	0.62	0.71	0.71	0.08
3	3.13	3.37	3.18	3.00	3.28
4	-3.17	-3.14	-3.16	-3.16	-3.16
5	0.73	-0.15	0.54	0.64	-0.06
6	-0.81	-0.74	-0.74	-0.72	-0.85
7	-0.78	-0.81	-0.78	-0.81	-0.77
8	0.60	0.69	0.69	0.67	0.53
9	1.25	0.54	1.43	1.44	0.83
10	1.37	1.42	1.41	1.25	1.45
95% Quintiles of the marginal posterior density					
1	-2.18	-1.13	-2.43	-2.41	-0.13
2	1.19	1.08	1.15	1.33	1.01
3	4.00	3.95	3.92	3.93	4.01
4	-1.11	-1.32	-1.30	-1.00	-0.08
5	1.34	1.31	1.25	1.22	1.38
6	-0.45	-0.15	-0.47	-0.43	-0.14
7	-0.45	-0.47	-0.49	-0.52	-0.42
8	1.17	1.19	1.17	1.20	1.16
9	2.01	2.03	2.03	2.06	2.01
10	1.71	1.82	1.71	1.71	2.09

Source: Own estimation.

In order to assess the convergence of the PT sampler, each estimation is repeated several times using a different set of random starting values. The results of the different runs are compared and it is checked if the marginal posterior densities are sufficiently similar between the runs. Specifically the mean, as well as the 5% and 95% quintile of the marginal posterior densities are compared. To illustrate the results of the convergence check, table 3.5 shows exemplary estimation results for 10 coefficients obtained in five estimations using different starting values. They illustrate that despite small variation induced by sampling noise, all runs converged to very similar values indicating that the sampler has indeed converged and does not get caught up in local modes. Similar results are obtained for all of the 64 coefficients and for all other estimations but are not shown here due to space limitations.

3.4 **Results**

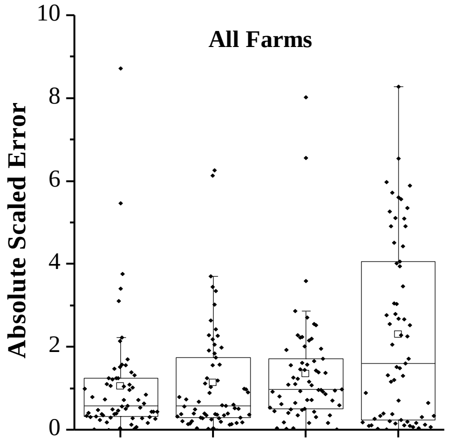
To assess the quality of the different prediction approaches different measures based on the Absolute Scaled Error are considered. In the out-of-sample prediction we obtain prediction results in each of the four cases for seven regions, three time periods and three size classes[23]. For each single prediction, the Absolute Scaled Error is calculated and then summarized across predictions by the mean and median Absolute Scaled Error as a measure of central tendency as well as the standard deviation and the 3rd quartile as measures of spread. The 3rd quartile is used as we are only interested in how far the Absolute Scaled Error deviates from zero.

Figure 3.1 depicts the performance measures of the different prediction method for the four different cases considered. The Markov approach clearly outperforms the geometric prediction in all four cases with respect to all measures. Compared to the linear prediction and the prediction of no change the picture is less clear. With respect to the mean Absolute Scaled Error, the Markov prediction outperforms the constant and the linear prediction in case of ‘all’ farms and livestock farms while it is outperformed by the constant and linear prediction in case of crop farms and the constant prediction in case of mixed farms. With respect to the median Absolute Scaled Error the Markov prediction is slightly inferior to the prediction of no change which has either a very similar or slightly lower median Absolute Scaled Error. Compared to the linear prediction, the

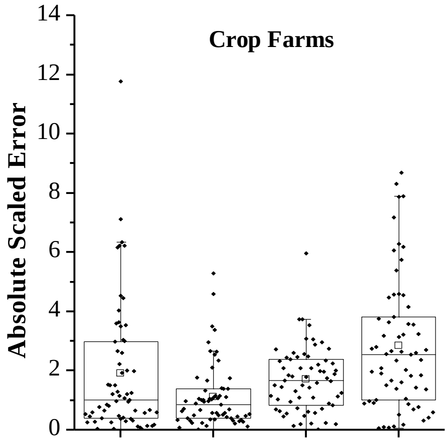
[23] The prediction for the entry/exit class is not considered since it is a no observable artificial class (see section 3.1).

Markov prediction is superior except for the case of Livestock farms where the linear prediction is slightly better.

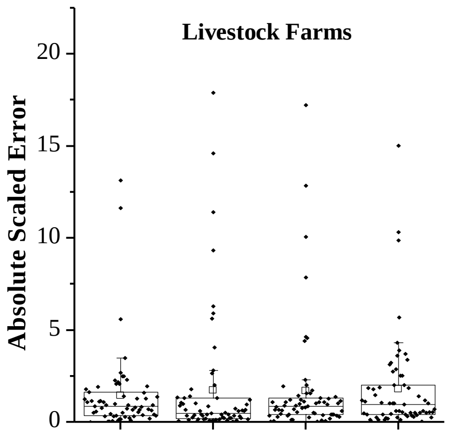
Figure 3.1 Box-Whisker-Plot of Absolute Scaled Errors for different prediction methods in four cases. Absolute Scaled Errors are displayed for each prediction in three size classes, three prediction periods and seven regions considered. In each case the Markov prediction is calculated as the mean of the posterior predictive distribution



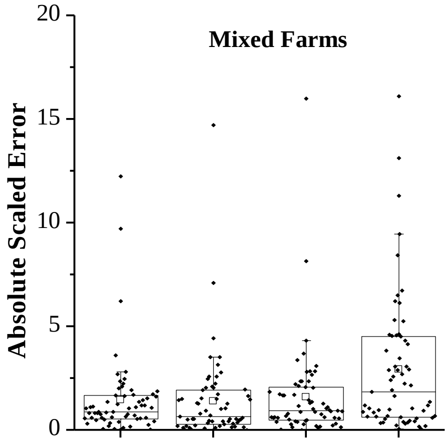
	Markov	Const	Linear	Geom.
Mean	1.05	1.14	1.36	2.30
Median	0.59	0.58	0.97	1.61
std	1.38	1.32	1.35	2.18
Q3	1.26	1.75	1.72	4.06



	Markov	Const	Linear	Geom.
Mean	1.90	1.12	1.71	2.85
Median	1.01	0.86	1.68	2.54
std	2.23	1.07	1.09	2.22
Q3	2.99	1.38	2.39	3.82



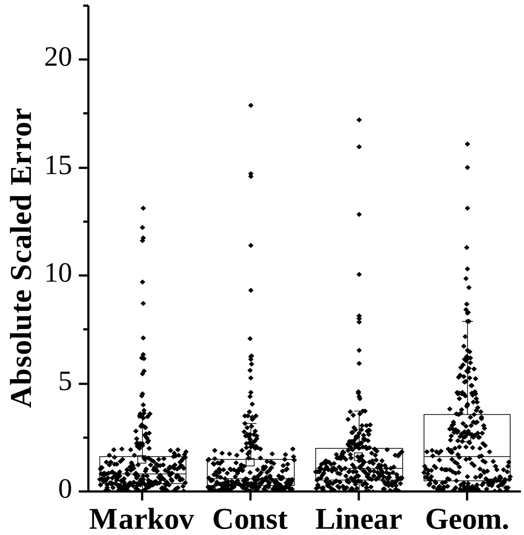
	Markov	Const	Linear	Geom.
Mean	1.44	1.73	1.68	1.78
Median	0.87	0.48	0.86	0.94
std	2.23	3.42	3.00	2.62
Q3	1.62	1.31	1.33	2.03



	Markov	Const	Linear	Geom.
Mean	1.46	1.40	1.59	2.93
Median	0.87	0.65	0.94	1.84
std	2.02	2.13	2.27	3.28
Q3	1.68	1.92	2.08	4.53

Source: Own calculations.

Figure 3.2 Box-Whisker-Plot of Absolute Scaled Errors for different prediction methods in the out-of-sample prediction. Absolute Scaled Errors are displayed for a prediction of crop, livestock and mix farms as well as for a prediction of all farms combined. In each case farm number are predicted in three size classes, three prediction periods and seven regions. The Markov prediction is calculated as the mean of the posterior predictive distribution.



	Markov	Const	Linear	Geom.
Mean	1.46	1.35	1.59	2.46
Median	0.85	0.62	1.08	1.64
std	2.00	2.19	2.07	2.64
Q3	1.61	1.52	2.00	3.59

Source: Own calculations.

For an overall assessment the individual results of the four cases are combined to obtain an overall measure of the prediction quality. The results are given in figure 3.2. The geometric prediction performs worst on all measures, followed by the linear prediction. The Markov prediction and the prediction of no change perform very similar with the prediction of no change being slightly better. The prediction of no change has a slightly lower mean and median Absolute Scaled Error and a slightly lower 3rd quartile while the Markov prediction has a slightly lower standard deviation. The results indicate that overall the Markov prediction is not able to clearly outperform the prediction of no change.

3.5 **Conclusion**

Overall, the paper contributes to the literature by extending the Bayesian estimation approach for non-stationary Markov model developed in chapter two by implementing a Parallel Tempering sampler that allows obtaining more robust sampling results. Additionally, a Bayesian prediction framework is derived that allows obtaining a full predictive distribution from which point predictions as well as all other moments of the prediction can be derived. Further, by relying on the Bayesian approach developed in chapter two, asynchronous data can be considered directly without the need of interpolating macro data as in previous studies.

In the paper a prediction framework was developed that enables a short term prediction of farm number combining all the available FSS macro and FADN micro data in one prediction. The results of out-of-sample prediction show that the developed approach outperforms naive linear and geometric predictions but is not able to outperform a prediction of no change. One needs to keep in mind, however, that the farm structure was rather stable within the short prediction period of two to three years. Moreover, further relevant drivers of farm structural change might have been missed due to limited data availability on potential explanatory variables.

Even though the focus of the paper and the out-of-sample prediction was a short term prediction of farm numbers, the proposed approach is useful for other purposes as well. As mentioned in the introduction, one application is to improve the projection of farm numbers in each farm typology (characterized by a type of farming and Economic class) in between the FSS years. This exercise is important in order to update the FADN weights which are a relevant input for agricultural policy analysis. Furthermore, the approach is useful to study the drivers of farm structural change by analyzing the influence of explanatory variables on the non-stationary transition probabilities using all available FSS and FADN data.

3.6 **References**

Breustedt G, Glauben T. 2007. Driving Forces behind Exiting from Farming in Western Europe. *Journal of Agricultural Economics* **58**(1):115–127.

Hyndman RJ, Koehler AB. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* **22**(4):679–688.

Liu JS. 2008. Monte Carlo strategies in scientific computing. New York: Springer.

Piet L, Latruffe L, Le Mouel C, Desjeux Y. 2011. How do agricultural policies influence farm size inequality? The example of France. *European Review of Agricultural Economics* **38**(4):

Stokes JR. 2006. Entry, Exit, and Structural Change in Pennsylvania's Dairy Sector. *Agricultural and Resource Economics Review* **35**(2):357–373.

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

Zimmermann A, Heckelei T. 2012a. Differences of farm structural change across European regions. Discussion Paper 2012:4, Institute for Food and Resource Economics, University of Bonn.

Zimmermann A, Heckelei T. 2012b. Structural Change of European Dairy Farms - A Cross-Regional Analysis. *Journal of Agricultural Economics* **63**(3):576–603.

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

Chapter 4
Direct payments, spatial competition and farm survival in Norway[24]

Abstract: We argue that farm survival is influenced by neighboring farmer's characteristics, and in particular by direct payments neighboring farmers receive. The paper shows empirically that these interdependencies are crucial for an assessment of the effects of direct payments on farm survival. Using spatially explicit farm level data for nearly all Norwegian farms, a spatial probit model is estimated in order to explain farm survival from 1999 to 2009. We show that ignoring spatial interdependencies between farms leads to a substantial overestimation of the effects of direct payments on farm survival. To our knowledge, this article is the first attempt to empirically analyze the importance of neighboring interdependencies for the effects of direct payments on farm survival.

Keywords: direct payments, farm structural change, land market, policy assessment, spatial competition

JEL classification: C21, C25, Q12, Q13

---

[24] This chapter is the basis for a paper accepted for publication in the American Journal of Agricultural Economics entitled: Storm H, Mittenzwei K, Heckelei T. (accepted). Direct payments, spatial competition and farm survival in Norway. DOI: 10.1093/ajae/aau085.

4.1 Introduction

In Norway, as in many other industrial countries, direct payments are often legitimized as a way to maintain a vital agricultural sector and, in particular, to prevent the abandonment of farms. It is often argued (e.g. Breustedt and Glauben 2007 for the EU or Goetz and Debertin 2001 for the US) that agricultural support increases farm profitability and with it reduces farm exits. Both studies analyze the effects of income support on net regional farm exit. These aggregate regional effects, however, might mask potential different reaction at the individual level (Gale 1994, Ehrensaft et al. 1984). Additionally, regional level studies rely on explanatory variables defined at the regional level making definition, interpretation and identification more complicated. With respect to direct payments, for example, one can only identify the aggregated effect of the average payment level in a region on net exit which likely differs from the individual level effects.

Individual farm level studies in contrast allow a direct analysis of the effects of farm characteristics and payments on farm survival, for example Key and Roberts (2006) who employ different survival modelling approaches to farm exits. For an overall assessment of the effects of payments, however, individual farm level effects need to be aggregated. We argue that this requires considering the interdependence between farms. As this link is missing in empirical farm level studies to date, Roberts and Key (2008:628) suggest regional level studies for policy assessment.

In this article we aim to explicitly consider these interdependencies in estimation and aggregation of the farm survival effects induced by a policy change. The objective is to empirically analyze the effect of direct payments on farm exit rates controlling for spatial farm interdependence using individual farm level data of nearly all Norwegian farms for 1999 and 2009. It is shown that ignoring the spatial interdependencies between farms in aggregation leads to an overestimation of the effects of direct payments on farm survival. To our knowledge this article is the first attempt to empirically analyze the role of neighboring characteristics for an assessment of the effects of direct payments on farm survival.

The importance of neighboring characteristics for an empirical evaluation of policies has been pointed out by Holloway, Lacombe and LeSage (2007:39–40) albeit in a different context.

In the farm structural change context, which subsumes the analysis of farm survival, the importance of neighboring interaction has long been acknowledged. Specifically, agent-based models of regional farm populations recognize the importance of land immobility, the location of farms in space and the interdependence of farms via competition on the spatial land market (Balmann 1997; Balmann et al. 2006; Happe et al. 2006; Happe et al. 2008; Freeman et al. 2009). However, econometric studies concerned with spatial interaction in farm structural change are rare. Huettel und Margarian (2009) consider different theoretical frameworks of strategic competition on the land market but do not empirically model interaction between farms when analyzing the impact of current and past regional farm structure on farm structural change. Weiss (1999) is aware of the importance of farm interdependence and the competition for land and labor, but does not consider them in his empirical analysis of farm survival and farm growth in Upper Austria.

In general the importance of spatial interdependencies in agricultural markets is long recognized. A classical paper by Sexton (1990), for example, devises a theoretical spatial competition model of the pricing behavior of processors under various conditions. The topic is picked up in two more recent articles by Graubner et al. (2011a) and Graubner et al. (2011b). Benirschka and Binkley (1994) consider spatial correlation in explaining land prices, however, on the regional level only. In other areas such as land use/cover change models, spatial dependencies and interactions on the land market are widely recognized (see Irwin und Geoghegan 2001 and Verburg et al. 2004 for a review). Gellrich und Zimmermann (2007), for example, focus on drivers of land abandonment in the Swiss mountains. In some respect land abandonment is similar to farm survival since the reasons for both likely overlap. Their approach, however, considers spatial correlation between regions leading to different interpretations of the spatial correlation compared to our approach at farm level.

One reason for the scarcity of empirical models analyzing spatial farm level interdependencies is the very limited availability of spatially explicit farm level data for representative samples at country scale. The data source for Norway thus provides a unique opportunity to empirically analyze their importance for farm survival. We estimate a spatial binary choice model to explain farm survival using own as well as neighboring farm characteristics. The regression results are then used for policy scenario simulations that explore the effects of a change in the direct payment system and the influence of spatial dependence for policy evaluation.

4.2 Theoretical Background and Hypothesis

One important source of interdependence between farms is the interaction of farms on the land market. Farms compete for a (in most cases) fix supply of agricultural land. This implies that farms may only grow if other farms decline in size or exit (Weiss 1999). Our arguments in the following rest on the assumption that transactions on the land market (either via rental agreements or land sale) are driven by the relative difference between farms' willingness to pay (WTP) for land. Farm WTP for one unit land is equal to the marginal value product of land, i.e. the residual return to land after cost for all other production factors are accounted for. Each unit of cultivated land ties labor and capital, therefore, WTP for land can also be interpreted as the difference between the on-farm income per area unit and the forgone off-farm income induced by cultivating that area unit. If farmers derive non-pecuniary utility from being self employed or see farming as a "way of life" (Key und Roberts 2009), WTP may also be larger than that difference. Given that each farm is located in a specific point in space and land is immobile, WTP for a specific plot is reduced by transportation costs rising proportional to the distance between plot and farm. Transactions on the land market occur if the relative differences in WTP for a specific plot exceed transaction costs. Focusing on farm survival, the article studies the special case in which WTP of one farm is lower than WTP of a competitor for every available plot. In this case the farm quits by renting out or selling all its land.

WTP for land differs between farms due to different characteristics. Of particular interest is the effect of direct payments on WTP and finally on farm exit. Key und Roberts (2006:391) found empirical evidence that payments have a significant positive effect on farm survival. They argue that relieve of liquidity constraints increases the possibility to bid up prices on the land market and helps farms to achieve a more efficient scale of operation. Payments also improve the relative profitability of farming compared to alternative occupations. Additionally, they reduce income uncertainty and the risk of bankruptcy. They therefore induce farmers to invest more aggressively (Vercammen 2007). Consequently, we expect a positive influence of direct payments on WTP.

It remains unclear, however, whether the absolute amount of payments or measured in relative terms (e.g., on a per labor hour basis) is more relevant. This issue is similar to the question whether total farm income or the on-farm wage rate is more important. We expect that this question depends on labor market conditions. Under perfect labor market conditions total farm income is less important, because small farms can complement their total income by off-farm employment at the on-farm wage rate. With imperfect labor markets, however, it might not be possible to complement income by off-farm employment and farmers may need to quit farming in order to take on a full off-farm employment. In this case total farm income matters more than the on-farm wage rate. Accordingly, we expect direct payments per labor input/total direct payments to be more important for WTP under fully functioning/imperfect labor markets.

On the other hand total farm income or total payments are a measure for the absolute size of a farm which is important in multiple aspects. Larger farms are likely to use labor more efficiently due to scale effects (Flaten 2002), adopt new technologies earlier (Weiss 1999) and, giving their larger collateral, face lower borrowing costs (Roberts and Key 2008). A crucial aspect in this respect is that farm size can be measured in several ways. Total income or total payments for example reflect the economic size of farm while total cultivated area or total labor input reflect the input side of production. In general, all measures are expected to be highly correlated and each capture size effects. Total area and total direct payments, however, might be a more direct measure of farm collateral while total labor use might be more relevant to asses scale effects. Following from this, we expect that all three measures are important in determining farmers WTP with each representing somewhat different effects.

Beside these factors of primary interest here, are many others that might be important for determining farms WTP for land. To limit further discussion, we restrict attention to those variables available in the empirical application. The productivity of a farm for example should have a positive influence on on-farm income and hence on WTP. The share of lease to total land should, ceteris paribus, have a negative effect on farm net worth and hence increase capital cost and decrease WTP. Further difference in WTP may arise due to different legal requirements for specific production types or specific policies for single specializations[25]. Equally important as farm characteristics are attributes of the farm holder (see among other Weiss 1999, Key and Roberts 2006). However, the age of the farm holder is the only variable available in the empirical application. Key und Roberts (2006) argued that older farms posses more information about the strength of the firm, more financial liquidity and over time are able to obtain a certain scale of production. On the other hand farm development beyond a certain age of the holder is strongly dependent on the availability of successor. Before retirement farm size might increase if a successor is available or if not might decrease in order to prepare for an exit (Mann et al. 2013). The theoretical effect of age on farm growth and survival is thus unclear.

Due to competition on the land market we expect that farm size is positively related to own WTP but negatively related to neighboring WTP. Given such interaction on the land market we expect the effects of neighboring characteristic to be the opposite as the effects of own characteristics. For total direct payments, for example, this means we expect a positive influence on WTP and on own growth/survival but a negative influence of neighboring payments. Farm growth and survival therefore depends on the relative difference between WTP between farms, i.e. farms occupy that area for which their difference between on and off-farm income exceeds that of their competitors. With respect to payments, this implies that changes in payments only have an effect if they change the relative difference in WTP between farms. When changes in payments are the same for all farms, as in the case of decoupled payments or coupled payments when farms production program are exactly the same, the relative difference remains constant and will leave farmers growth or survival decision unaffected. One can also think of the effect of a full capitalization of payments in land rents (Latruffe and Le Mouël 2009). If changes in payments differ (e.g., due to different participating

[25] Within the study period, for example, the government had a large milk quota-buy-out program in place, which might have had an effect on milk farms but no direct effect on other farms.

rates (Roberts and Key 2008:630), different farm specializations or because per unit subsidy rates discriminate between land and herd sizes as is the case in Norway) the relative difference changes, leading to growth/decline and an increase/decrease of the likelihood of survival for favored/non-favored farms.

Finally, it is important to point out that spatial interactions on the land market are not the only way farms interact which each other. One other important type of interaction is technology adoption and knowledge transfer (Rogers 1995; Berger 2001). Case (1992) and Holloway et al. (2002), for example, found evidence that the probability of adopting a new technology increases with neighboring adoption. Consequently, an active corporate network raises technology diffusion and with it farm productivity. Neighboring farms are also important to maintain an active network of suppliers and processors. Overall, an active corporate network should thus increase farm profitability and hence WTP for land. Similarly as with payments, the effects of an active corporate network on farm size could cancel out if all farms benefit similarly (WTP would increase for all farms alike). However, larger farms are more likely to adopt a new technology (Feder and Slade 1984) and might also be more capable in maintaining an active corporate network of suppliers and processors. Therefore, WTP of small farms that benefit from larger neighbors might increase more as WTP of large farm with small neighbors. Based on this reasoning neighboring size can also have a positive influence on own WTP and hence farm size and survival. Which effects dominate in the end, the negative due to competition on the land market or the positive due to an active corporate network, remains an empirical question. In general, all cases where we do not find the opposite sign of neighboring characteristics compared to own characteristics hints at interaction between farms other than the competition on the land market.

4.3 Empirical Model and Estimation

The empirical investigation explores the effects of own and neighboring direct payments on farm survival using a spatial probit model considering the exit decision of almost all Norwegian farms between 1999 and 2009 (a description of the data is available in appendix A). The latent variable $\mathbf{y}^*$ underlying the probit model, determines the outcome of the observed survival ($y_i = 1$ if $y_i^* > 0$) or exit decision ($y_i = 0$ if $y_i^* \le 0$). The model can be interpreted as a latent utility model reflecting the difference between own and neighboring WTP for land discussed in the theoretical part. The latent variable $\mathbf{y}^*$ is specified to be a linear function of own characteristics $\mathbf{X}$ and neighboring characteristics $\mathbf{WX}$, with $\mathbf{W}$ being a spatial weighting matrix defined below. For estimation purposes, two different model specifications are considered. The first specification is a spatially lagged explanatory variable model (SLX) of the form $\mathbf{y}^* = \mathbf{X\beta} + \mathbf{WX\theta} + \boldsymbol{\varepsilon}$ with $\varepsilon \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ which assumes iid normal errors. The second specification is a spatial Durbin error model (SDEM) of the form $\mathbf{y}^* = \mathbf{X\beta} + \mathbf{WX\theta} + \mathbf{u}$ with $\mathbf{u} = \rho\mathbf{Wu} + \boldsymbol{\varepsilon}$ which relaxes the assumptions of the SLX model by allowing for spatially autocorrelated errors (LeSage and Pace 2011:22). Reasons for choosing the SLX and SDEM specification over the more common spatial autoregressive model (SAR) are laid out in appendix B.

The SLX model is estimated using standard probit maximum likelihood estimation techniques. We then test for spatial error dependence using three different test principles compared in Amaral et al. (2013). All three tests clearly reject the H_0 of no spatial autocorrelation. Since autocorrelation may lead to estimation bias the test results render the SDEM model more appropriate. Estimation of a SDEM probit model for over 60,000 observations, however, is challenging from a computational perspective. Most existing estimation techniques such as McMillen (1992), Beron and Vijverberg (2004) or LeSage and Pace (2009) are only applicable for relatively small samples of a couple of thousand observations (see Pace and LeSage 2011 for a detailed discussion of the limitations with respect to large samples). Therefore, Pace und LeSage (2011) proposed a simulated maximum likelihood framework capable of handling large samples (a detailed description of the implementation of the estimator is provided in appendix C).

4.3.1 *Dependent and Explanatory Variables*

The dependent variable in the analysis represents farm survival in 2009 of all farm active in 1999 and is equal to one if a farm is still active in 2009, zero otherwise. We consider a farm as active if at least one production activity is observed in the payment data base. Explanatory variables are derived from the payment data base as well as from the 1999 farm census. As discussed above, the most important variables relate to different types of farm income. Farm income is divided into market returns and direct payments which are of particular interest. Since actual market returns for each farm are unobserved, we consider an average market return to labor for each production activity derived from the reference farms data collection (NILF 2000 and NILF 2009)[26]. These average returns are used to reflect the difference in market returns arising from different production programs. The direct payments per farm are calculated rather accurately using actual payment rates, observed production activities and eligibility rules. Total income is then equal to the sum of the two and farmers' on-farm wage rate is approximated as the ratio of direct payment and market returns over an estimated total labor use. Additionally, we obtain a measure for the potential change in the on-farm wage rate under the condition that a farm maintains its 1999 size and production program. The reasoning is that changes in either size or production program might already be the result of changes in income opportunities that we aim to measure. A detailed description of all income variables is provided in Storm und Mittenzwei (2013).

As discussed in the theoretical section it remains undecided whether the total income or the on-farm wage rate is more important for farmers WTP for land and hence farm survival. In the empirical application we thus include total direct payments (*dpay99)* and total market return in 1999 (*mReturn99*) as a measure of total farm income as well as direct payment and market returns per labor use in 1999 (*dpay99/reqLabo* and *mReturn99/reqLabo*). Furthermore, we add the change in the latter two (*C.DPayLabo* and *C.mRetLabo*) as measures of the on-farm wage rate.

Additionally total agricultural area (*area*), total observed labor input in 1999 (*obsLabo99*) and estimated labor use for 1999 (*reqLabo99*)[27] are included. All three, together with total income, are measures for the absolute size of a farm. In line with the discussion above, we finally include the age of the farm holder[28]

[26] It is important to recognize that 'market returns' also substantially depend on policy decisions since market prices are strongly affected by administrative prices.

[27] See Storm and Mittenzwei (2013) for detailed information about the estimation of the labor requirements.

[28] For observations where age is missing in the data base we imputed the mean age. The age is missing for example for all farms where the owner is not a natural person. In total we have 495 or 0.77% missing observations for age.

(*age*), the ratio of leased to total agricultural area (*landLease/Tot*), the ratio between observed labor input, estimated labor requirements (*laboObs/Req*) as a measure of farm productivity, and dummy variables indicating if a farm has milk cows (*hasMilk*), sheep (*hasSheep*), sows (*hasSows*), poultry (*hasPoultry)* as a rough measure to reflect specialization specific policy environments. Descriptive statistics along with the variable code of all explanatory variables are provided in appendix D. For the analysis all variables are z-standardized.

4.3.2 *Spatial Weighting Matrix*

Estimation of SLX and SDEM models, requires the specification of a spatial weighting matrix $\mathbf{W}$, which approximates neighboring relations between farms. This task is challenging in general and in particular given Norway's heterogeneous farming regions, varying from dense small scale berry production to wide extensive sheep grazing areas. We expect that neighboring relations and the size of the local land market, i.e. the distance between farms and fields farmers compete for, to differ between these regions. From the 1999 farm census, data about the driving distance from the farmstead to the furthest field is available and we expect that this data carries information about the local structure of the farm sector and the distance over which farms compete for land. Using this data the median driving distance to the furthest field in each municipality is derived and neighbors of a farm are defined as all farms that are within a radius of this distance. The median is used to eliminate the influence of potential outlier and zero observations that cannot be distinguished from missing values. Further, the maximum number of neighbors is set to 20 (nearest neighbors) in order to prevent farms from having a very large number of neighbors and neighboring farms are weighted by their inverse distance.

One common criticism of spatial regression models, particularly in a micro-data environment (Bell and Dalton 2007), is that $\mathbf{W}$ is defined rather arbitrary and does not necessarily represent the true neighboring relation. Even though we base our definition on empirical data this criticism remains valid. The importance of the definition of $\mathbf{W}$ for the final results is controversial. LeSage and Pace (2011) argued that in most cases the results are less sensitive to the definition of $\mathbf{W}$ as commonly believed. Others, such as Holloway et al. (2007), found that the spatial correlation in an SAR model depends heavily on the definition of $\mathbf{W}$. They used a Bayesian model comparison approach in order to determine the most likely neighboring relationships. Such a rigorous treatment of model uncertainty goes beyond the scope of our article. We explore the sensitivity of our results to the definition of $\mathbf{W}$ by repeating the estimation of the SLX model using two very different definitions. Results (reported in appendix E) show that the final conclusions are largely unaffected by the definition of $\mathbf{W}$.

4.4 Regression Results

In the following the results with and without spatial interactions are presented. Distinguishing between the two models, we can explore the effects of including and ignoring spatial interactions on the influence of changes in direct payments. The regression results for the non-spatial model as well as the results for the spatial model using the SLX and SDEM model specification are reported in table 4.1. It can be seen that the coefficients with respect to the non-spatially lagged variables differ only slightly between the three specifications. Therefore we discuss the non-spatial results first and highlight the differences with respect to the spatial model in the following.

The non-spatial regression results are presented in the left panel of table 4.1. Except for the market return and its square, the “has poultry” dummy (*hasPoultry*) and the squared estimated labor requirement (reqLabo99^2), all explanatory variables are highly significant. Insignificant squared terms are dropped from the model. Statistical significance is comparatively easy achieved with more than 60,000 observations, but says little regarding relevance. A measure of the explanatory power of the overall model is the percentage of

Table 4.1: Regression Results for the Non-Spatial Probit SLX and SDEM Model to Explain Farm Survival

	Non-spatial probit		SLX		SDEM	
Variable	Coef	p-value	Coef	p-value	Coef	p-value
const	0.3931	0.0000	0.3948	0.0000	0.3974	0.0000
age	0.5656	0.0000	0.5581	0.0000	0.5631	0.0000
age^2	-0.6596	0.0000	-0.6499	0.0000	-0.6548	0.0000
area	0.2533	0.0000	0.1920	0.0000	0.1886	0.0000
area^2	-0.1331	0.0000	-0.1190	0.0000	-0.1176	0.0000
obsLabo	0.2784	0.0000	0.2622	0.0000	0.2626	0.0000
obsLabo^2	-0.1174	0.0000	-0.1100	0.0000	-0.1103	0.0000
reqLabo	0.1411	0.0000	0.1291	0.0000	0.1334	0.0000
mRet	0.0043	0.7039	0.0090	0.4286	0.0107	0.3250
dpay	0.6197	0.0000	0.7421	0.0000	0.7507	0.0000
dpay^2	-0.3382	0.0000	-0.3477	0.0000	-0.3518	0.0000
laboObs/Req	-0.0425	0.0000	-0.0396	0.0000	-0.0394	0.0320
landLease/Tot	-0.0455	0.0000	-0.0441	0.0000	-0.0442	0.0287
mretrun/reqLabo	0.1141	0.0000	0.1006	0.0000	0.0972	0.0000
dpay/reqLabo	0.0629	0.0000	0.0723	0.0000	0.0738	0.0000
C.DPayLabo	0.1311	0.0000	0.0967	0.0000	0.0951	0.0000
C.mRetLabo	0.0780	0.0000	0.0638	0.0000	0.0586	0.0000
hasMilk	-0.1885	0.0000	-0.2254	0.0000	-0.2270	0.0000
hasPoultry	0.0071	0.2799	0.0061	0.3554	0.0067	0.5520
hasSheep	0.0220	0.0010	0.0209	0.0031	0.0205	0.0229
hasSows	0.0455	0.0000	0.0421	0.0000	0.0433	0.0084
W_mRet	---	---	-0.0179	0.0539	-0.0185	0.0665
W_dpay	---	---	-0.2708	0.0000	-0.2718	0.0000
W_area	---	---	0.0721	0.0000	0.0742	0.0003
W_reqLabo	---	---	0.0617	0.0000	0.0624	0.0188
W_landLease/Tot	---	---	-0.0371	0.0000	-0.0373	0.0520
W_FarmWage	---	---	0.0345	0.0015	0.0341	0.0186
W_C.inco	---	---	0.0498	0.0000	0.0509	0.0000
W_hasMilk	---	---	0.0774	0.0000	0.0761	0.0015
W_hasPoultry	---	---	0.0094	0.1084	0.0102	0.5515
W_hasSheep	---	---	0.0186	0.0090	0.0177	0.4407
W_hasSows	---	---	0.0144	0.0163	0.0130	0.5541
rho	---	---	---	---	0.1199	0.0000
n		*64488*		*64488*		*64488*
% Correct predictions Model		*72.59*		*72.63*		*72.64*
% Correct predictions Naive		*62.72*		*62.72*		*62.72*
Total Gain[a]		*9.88*		*9.91*		*9.92*

Note: The dependent variable is equal to one if the farm stays active between 1999 and 2009 and zero otherwise. Spatially lagged variables are denoted with a leading "W_";

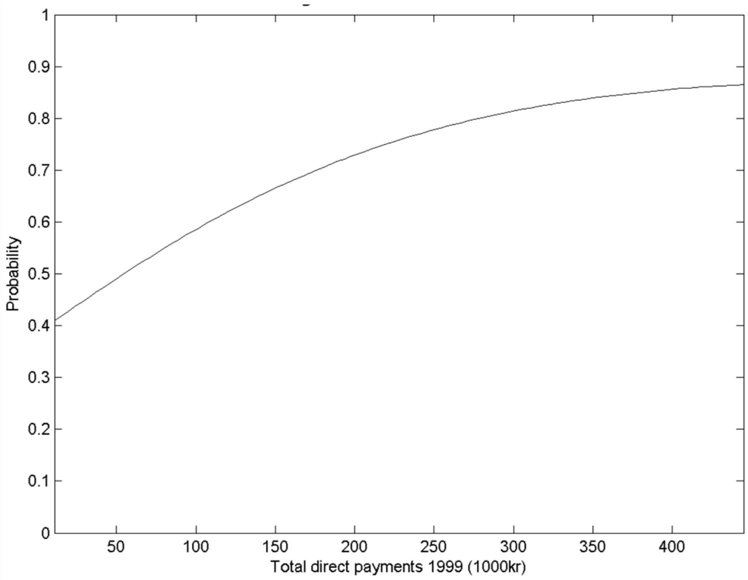
[a]*Change in "% Correct" compared to naive specification.*

correctly predicting the binary choice. With the non-spatial model we are able to correctly predict the exit/survival decision in 72.64% compared to 62.72% using a naïve model. To assess the explanatory power of individual variables, we can explore how the percentage of correct prediction changes with or without the variable under consideration. Overall we found that the variables related to farm size (*area*, *obsLabo99*, *reqLabo99*, and *dpay99*) are most important explaining farmers' exit/survival decision (see appendix F). A model with all variables except these variables would predict 67.58% of the choices correctly, which is around half the gain of the full model over the naïve model. Due to their high correlation (see appendix G) each size variable can explain more or less the same, with the percentage of correct predictions of a model with only one of the size variables being only slightly lower than of one with all included. The importance of the remaining variables, including the on-farm wage rate, is relatively evenly distributed with each variable adding only little to the overall explanatory power of the model (see appendix H). As discussed above, the greater importance of the absolute size of a farm compared to the on-farm wage rate or changes in the later, might indicate potential imperfections on the labor market. These might render the potential on-farm income per person or family, proxied by the absolute size, as more important than the on-farm wage rate per hour.

Overall, all variables related to the absolute size of a farm show a positive influence (some with decreasing rate) between farm size and survival. Figure 4.1, which shows the survival probability for an 'average' farm for varying direct payments, illustrates that the effect of payments levels out with increasing size. This indicates, as mentioned above, that beyond the size sufficient to provide for the family, additional payments have a minor effect on the probability of survival.

With respect to policy design, we could draw the conclusion from the non-spatial findings that increasing direct payments would be a good approach to increase the survival probability (of at least relatively small) farms. In the following we explore how this conclusion is affected when considering spatial interaction between farms.

Figure 4.1 Probability for an 'average' farm to stay active between 1999 and 2009 for varying total direct payments



Note: The x-axis represents the 2.5% to 97.5% quintile of the observed total direct payments. All other variables are held constant at their means.

The spatial regression results for the SLX and SDEM model are reported in the right panel of table 4.1. For model specification we included all variables, except the squared terms, as spatially lagged variables[29]. The regression results for the SLX and SDEM model are almost identical despite the significant spatially autocorrelated errors with $\rho = 0.12$ in the SDEM model. This implies that ignoring the spatially autocorrelation in the errors does not result in a substantial bias of the SLX model.

[29] Since the spatially lagged variables show less variation we summarize variables that are highly correlated and measure related aspects. Specifically, the two variables for the on-farm wage rate *mReturn99/reqLabo* and *dpay99/reqLabo* are summarized to one variable *W_FarmWage99*. Similarly, the two variables for the change in on-farm wage rate *C.DPayLabo* and *C.mRetLabo* are summarized to one variable *W_C.inco99*. The spatially lagged observed labor input is excluded from the model specification because of a high correlation to the estimated labor requirement that does not allow identifying both variables. The general model results and conclusions, however, are unaffected by the choice of which to exclude from the model.

The results of the non-spatial variables discussed before seem to be robust with respect to the inclusion of spatial lagged explanatory variables as they stay almost unchanged. The percentage of correct predictions improves only slightly indicating that the spatial lagged variables have only little explanatory power. Further, the relative importance of different variables stays unaffected such that the findings discussed above similarly hold for the spatial model. With respect to the research question, however, considering the spatial effects is crucial. Above we concluded that, irrespective of the measurement, the absolute size is the most important factor in explaining farm survival and, for the relevant range, the larger the size the higher the survival probability. These findings remain valid for the spatial model, but the effect of the absolute size of neighboring farms is somewhat more complicated. When considering only one spatially lagged variable for neighboring farm size, we found a negative influence between neighboring farm size and own survival irrespective of which variable (*w_dpay99*, *w_uaar* or *w_laboreq99*) is used. As discussed above, all three measures of the absolute size of the farm are highly correlated and the same holds for the spatially lagged absolute size measures. Nevertheless, the large sample size is sufficient to identify different coefficients for the three variables (table 4.1). Farms with larger neighbors in terms of area and labor use have a higher survival probability, while farms with larger neighbors in terms of total direct payments have a lower survival probability. As discussed in the theoretical section, a reason for this finding could be the multiple ways farms interact which each other. On the one hand farms gain from an active cooperative network due to technology diffusion or easier access to suppliers or processors. The larger the neighboring farms in terms of the cultivated area and/or the total labor use, the more likely it is that farms are situated in an active corporate network with the positive effects that follow from this. On the other hand farms compete with their neighbors for the limited resource land on local land markets. Hence, farms having neighbors with higher direct payments (everything else equal) have increased the attractiveness to rent out/sell the farm, limit growth prospects and therefore decrease survival probability.

With respect to policy design, these findings have important implications. The non-spatial results imply that increasing direct payments increases survival probability and that an increase for all farms may reduce farm exits. However, the spatial results imply that increasing own direct payments increase survival probability, but negatively affect the survival of neighboring farms. The overall effects of a change in direct payments are ambiguous and need to consider the actual neighboring relations between farms in the population. This issue is explored in the next section.

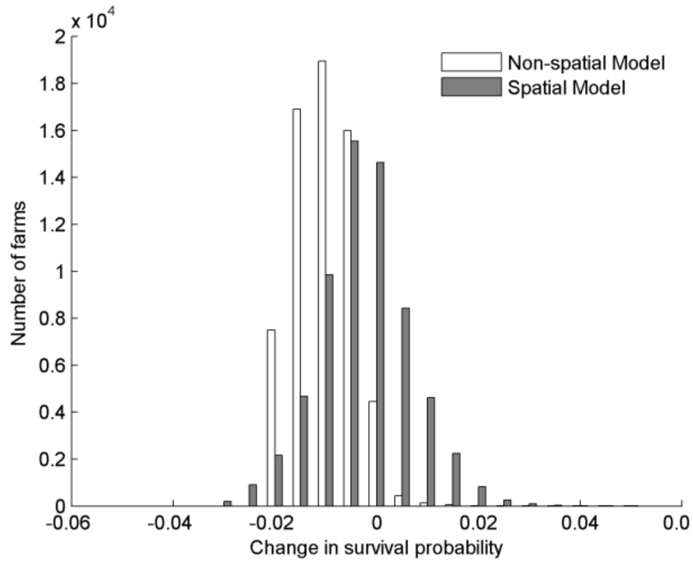
4.5 Policy Scenario Simulation

In order to illustrate the importance of spatial dependencies for policy design and evaluation, and to provide empirical evidence for specific changes in the Norwegian payment scheme, simulation experiments based on the entire population are performed. In particular, we calculate the change in farm survival probability for two different direct payment policy scenarios. Therefore, predicted survival probabilities are derived based on observed explanatory variables and for the same variables except that total direct payments are now calculated for each farm under the new policy regime. The difference between these survival probabilities is plotted as a histogram in figure 4.2 and 4.3. The differences between the spatial and non-spatial model is highlighted in each histogram by showing separate results for each model. The first scenario (figure 4.2) considers a general reduction of all payment rates by 10%, irrespective of farm types, sizes or location. The second scenario (figure 4.3) assumes the elimination of the structural dimension of the payments by providing support through flat per animal and per ha payment rates. In the current policy regime, several payment schemes differentiate payments rates according to farm size, in the sense that rates for the first unit (animal head or area) are higher than for the last. Assuming flat rates equal to the lowest rates currently paid, this scenario implies a 30% reduction of total direct payments with small farms being more affected in relative terms.

With an average decrease in survival probability equal to 1.04 and 0.26 percentage points for the non-spatial and spatial model, respectively, the effects of the first scenario (figure 4.2) seem modest in general but particularly when considering the spatial dependence. The 95% prediction interval for both values, equal to [0.89, 1.18] and [0.2, 0.31] is relatively small due to the large sample. The prediction intervals are calculated using a bootstrap procedure described in appendix I. Instead of the average decrease in survival probability the result can also be expressed in terms of number of predicted farm exits. Here, predicted exits increase by 964 and 171 for the non-spatial and spatial model, respectively (farm population size 64.488 in 1999 and 40.445 in 2009). Figure 4.3 shows results of the second scenario in which the structural dimension of direct payments is abolished. The effects are more substantial with an average decrease in predicted survival probability equal to 4.00 (95% prediction interval [4.52 ,3.47]) and 1.60 (95% prediction interval [1.77,1.42]) and an increase in predicted farm exits equal to 4,046 and 1,474 for the non-spatial and spatial model, respectively.

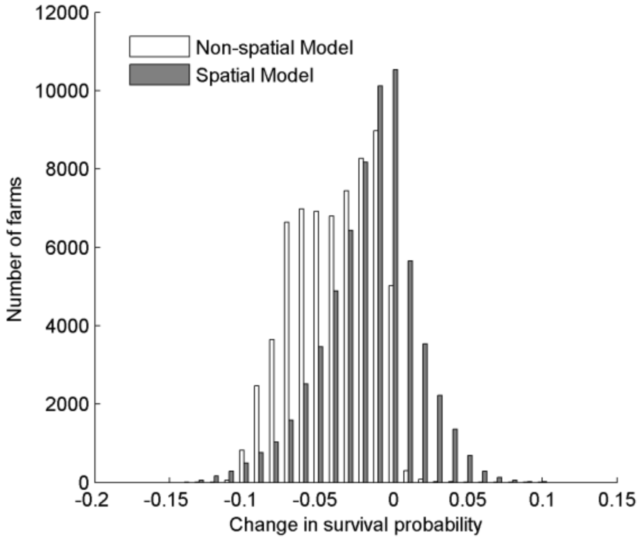
Several conclusions follow from this. The first scenario shows generally moderate effects on farm survival. Interestingly, the spatial model indicates (figure 4.2) that a substantial share of farms is unaffected or even gains from an overall decrease in payments. The neighboring effects of payments seem to outweigh the negative effect of a decrease in own payments. Further, with respect to the importance of spatial dependence for policy design, both scenarios indicate that the non-spatial model leads to a substantial overestimation of the effects of payments. This is especially evident in the second scenario, where despite a large overall reduction

Figure 4.2 Change in individual farm survival probability for a 10% reduction of all direct payment rates



Source: Own illustrations.

Figure 4.3 Change in individual farm survival probability for a abolishment of the structural dimension of direct payments in which rates are set equal to the lowest rates currently paid



Source: Own illustrations.

in payments of about 30%, survival probabilities decrease only by 1.6 percentage points on average. These results are important for Norwegian policy makers and the society as a whole as a basis for decision on policy design. A spatial representation of the results at regional level is available in the appendix J and K.

4.6 **Conclusion**

This article for the first time considers spatial interdependence between farms in an empirical assessment of the effects of direct payments on farm survival. We found that higher direct payments for neighbors decrease own survival probability. Ignoring this spatial interdependence led to a substantial overestimation of the overall effects of direct payments on farm exits. An assessment of the effect of a change in the support regime should therefore not be based on the assumption of independent farm behavior. Instead, policy changes should be assessed for the entire farm population considering the spatial interdependence. Policy simulation results show rather modest effects of reductions in direct payments and survival probabilities for a substantial share of farmers which remain unaffected or even increase.

In addition, we found that the total economic size of a farm is more important for farm survival than on-farm wage rates. Imperfect labor markets and family farm structures in Norway are likely rendering a farm's total income potential as highly relevant for farm survival. Results indicate that direct payments may help smaller farm across thresholds for survival to some extent, but the probability of survival of larger farms is basically unaffected.

4.7 References

Amaral PV, Anselin L, Arribas-Bel D. 2013. Testing for spatial error dependence in probit models. *Letters in Spatial and Resource Sciences* **6**(2):91–101.

Balmann A, Dautzenberg K, Happe K, Kellermann K. 2006. On the dynamics of structural change in agriculture: Internal frictions, policy threats and vertical integration. *Outlook on Agriculture* **35**(2):115–121.

Balmann A. 1997. Farm-based modelling of regional structural change: A cellular automata approach. *European Review of Agricultural Economics* **24**(1):85–108.

Bell KP, Dalton TJ. 2007. Spatial Economic Analysis in Data-Rich Environments. *Journal of Agricultural Economics* **58**(3):487–501.

Benirschka M, Binkley JK. 1994. Land Price Volatility in a Geographically Dispersed Market. *American Journal of Agricultural Economics* **76**(2):185–195.

Berger T. 2001. Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics* **25**(2-3):245–260.

Beron K, Vijverberg W. 2004. Probit in a Spatial Context: A Monte Carlo Analysis, in: Anselin L. (Ed.), Advances in spatial econometrics. Berlin, Heidelberg, New York: Springer, 169–195.

Breustedt G, Glauben T. 2007. Driving Forces behind Exiting from Farming in Western Europe. *Journal of Agricultural Economics* **58**(1):115–127.

Case A. 1992. Neighborhood influence and technological change. *Regional Science and Urban Economics - Special Issue Space and Applied Econometrics* **22**(3):491–508.

Ehrensaft P, LaRamee P, Bollman RD, Buttel FH. 1984. The Microdynamics of Farm Structural Change in North America: The Canadian Experience and Canada-U.S.A. Comparisons. *American Journal of Agricultural Economics* **66**(5):823.

Feder G, Slade R. 1984. The Acquisition of Information and the Adoption of New Technology. *American Journal of Agricultural Economics* **66**(3):312.

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**(4):429–441.

Freeman T, Nolan J, Schoney R. 2009. An Agent-Based Simulation Model of Structural Change in Canadian Prairie Agriculture, 1960-2000. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* **57**(4):537–554.

Gale HF. 1994. Longitudinal Analysis of Farm Size over the Farmer's Life Cycle. *Review of Agricultural Economics* **16**(1):113.

Gellrich M, Zimmermann NE. 2007. Investigating the regional-scale pattern of agricultural land abandonment in the Swiss mountains: A spatial statistical modelling approach. *Landscape and Urban Planning* **79**(1):65–76.

Gibbons S, Overman HG. 2012. Mostly Pointless Spatial Econometrics? *Journal of Regional Science* **52**(2):172–191.

Goetz SJ, Debertin DL. 2001. Why Farmers Quit: A County-Level Analysis. *American Journal of Agricultural Economics* **83**(4):1010–1023.

Graubner M, Balmann A, Sexton RJ. 2011a. Spatial Price Discrimination in Agricultural Product Procurement Markets: A Computational Economics Approach. *American Journal of Agricultural Economics* **93**(4):949–967.

Graubner M, Koller I, Salhofer K, Balmann A. 2011b. Cooperative versus non-cooperative spatial competition for milk. *European Review of Agricultural Economics* **38**(1):99–118.

Happe K, Balmann A, Kellermann K, Sahrbacher C. 2008. Does structure matter? The impact of switching the agricultural policy regime on farm structures. *Journal of Economic Behavior & Organization* **67**(2):431–444.

Happe K, Kellermann K, 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**(1):49.

Holloway G, Lacombe D, LeSage JP. 2007. Spatial Econometric Issues for Bio-Economic and Land-Use Modelling. *Journal of Agricultural Economics* **58**(3):549–588.

Holloway G, Lapar, Ma. Lucila A. 2007. How Big is Your Neighbourhood? Spatial Implications of Market Participation Among Filipino Smallholders. *Journal of Agricultural Economics* **58**(1):37–60.

Holloway G, Shankar B, Rahmanb S. 2002. Bayesian spatial probit estimation: a primer and an application to HYV rice adoption. *Agricultural Economics* **27**(3):383–402.

Huettel S, Margarian A. 2009. Structural change in the West German agricultural sector. *Agricultural Economics* **40**(6, Suppl. s1):759–772.

Irwin EG, Geoghegan J. 2001. Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment* **85**(1-3):7–24.

Kelejian HH, Prucha IR. 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *The Journal of Real Estate Finance and Economics* **17**(1):99–121.

Key N, Roberts MJ. 2006. Government Payments and Farm Business Survival. *American Journal of Agricultural Economics* **88**(2):382–392.

Key N, Roberts MJ. 2009. Nonpecuniary Benefits to Farming: Implications for Supply Response to Decoupled Payments. *American Journal of Agricultural Economics* **91**(1):1–18.

Latruffe L, Le Mouël C. 2009. Capitalization of Government Support in Agricultural Land Prices: What do we Know? *Journal of Economic Surveys* **23**(4):659–691.

Lee L. 2004. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica* **72**(6):1899–1925.

LeSage JP, Pace RK. 2009. Introduction to spatial econometrics. Boca Raton, London, New York: CRC Press.

LeSage JP, Pace RK. 2011. Pitfalls in Higher Order Model Extensions of Basic Spatial Regression Methodology. *The Review of Regional Studies* **41**(1):13-26.

Mann S, Mittenzwei K, 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:**109–138.

Manski CF. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* **60**(3):531.

McMillen DP. 1992. Probit with Spatial Autocorrelation. *Journal of Regional Science* **32**(3):335–348.

NILF. 2000. Referansebrulsberegninger 2000. Norsk institutt for landbruksøkonomisk forskning (NILF), Oslo.

NILF. 2009. Referansebrulsberegninger 2009. Norsk institutt for landbruksøkonomisk forskning (NILF), Oslo.

Norwegian Agricultural Authority (NAA). 2011. Direct payments database, Oslo.

Pace RK, LeSage JP. 2011. Fast Simulated Maximum Likelihood Estimation of the Spatial Probit Model Capable of Handling Large Samples. Available at SSRN: http://ssrn.com/abstract=1966039.

Pace RK, Zhu S. 2012. Separable spatial modeling of spillovers and disturbances. *Journal of Geographical Systems* **14**(1):75–90.

Roberts MJ, Key N. 2008. Agricultural Payments and Land Concentration: A Semiparametric Spatial Regression Analysis. *American Journal of Agricultural Economics* **90**(3):627–643.

Rogers EM. 1995. Diffusion of innovations. New York: Free Press.

Sexton RJ. 1990. Imperfect Competition in Agricultural Markets and the Role of Cooperatives: A Spatial Analysis. *American Journal of Agricultural Economics* **72**(3):709.

Statistics Norway. 2011. Structure of agriculture. Holdings, by size of agricultural area in use and county. Available at: http://www.ssb.no/a/english/kortnavn/stjord_en/tab-2012-11-28-01-en.html, downloaded 15.03.2011.

Storm H, Mittenzwei K. 2013. Farm survival and direct payments in the Norwegian farm sector. Discussion paper 2013-5, Norwegian Agricultural Economics Research Institute, Oslo.

Verburg PH, Nijs TC de, van Ritsema Eck J, Visser H, Jong K de. 2004. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems* **28**(6):667–690.

Vercammen J. 2007. Farm bankruptcy risk as a link between direct payments and agricultural investment. *European Review of Agricultural Economics* **34**(4):479–500.

Weiss CR. 1999. Farm Growth and Survival: Econometric Evidence for Individual Farms in Upper Austria. *American Journal of Agricultural Economics* **81:**103–116.

4.8 **Appendix**

Appendix A: Description of the Data Source

The analysis is based on data from the Norwegian Direct Payment Register for the years 1999 and 2009. The register contains information about agricultural area by crop and number of animals by type of animal (126 different crop and animal activities are distinguished) for every farm that applies for direct payments[30]. Eligibility for direct payments is subject to certain conditions, one of which is a minimum economic size of the farm (measured by turn-over) in order to prevent “hobby-farms” from receiving subsidies. As a consequence, the total numbers of acreage and/or animals may be somewhat underestimated when compared with other official sources such as slaughter statistics or the decennial total farm census.

Individuals and legal entities managing agricultural area or keeping animals eligible for direct payments may apply for subsidies by filling in data in the register. The register links the amount of acreage and animals with business identification and property numbers. Additionally, farmers’ social security numbers are available containing the birth date.

As the unit of analysis we rely on the property number. Property units present in 1999, but not in 2009 are assumed to have left the sector. Some potential measurement errors arise from this choice: We disregard if farms split their activities in different business units. Small farms may incidentally have left the sector in 2009, but applied for subsidies in 2008 and 2010.

Table A-1 shows the number of farms covered in the database for the two measures mentioned above and compared to the number of farms recorded in other statistics.

[30] Because of missing observations due to mergers of municipalities it was necessary to exclude 11 municipalities from the analysis. Municipality codes 529, 716, 718, 1154, 1214, 1418, 1514, 1569, 1572, 1576, and 1842.

Table A-1. Number of Farms for Various Accounting Measures

	1999	2009
Property number (NAA)	66,892	45,460
Business number (NAA)	66,832	45,420
Number of farms (Statistics Norway)	70,740	47,688

Source: Norwegian Agricultural Authority 2011 and Statistics Norway 2011

Table A-1 reveals that there are small differences between the measures to identify farms. For all practical purposes regarding the analysis, the number of properties and the number of businesses appears to be the same. Further, the numbers are somewhat lower than the number of farms provided by the Statistical Office (Statistics Norway) due to certain size limits regarding the eligibility of direct payments.

Appendix B: Background information on model choice

In many spatial econometric studies the spatial autoregressive (SAR) model of the form

(17) $$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ with } \boldsymbol{\varepsilon} \sim N\left(0, \sigma^2 I\right),$$

is employed. The SLX and SDEM model is chosen over the more common SAR model since it allows greater flexibility with respect to the direct and indirect effects of explanatory variables. As shown by Pace and Zhu (2012), the indirect effects in the SAR model have the same signs as direct effects and the ratio between indirect and direct effects is constant across variables. This is an undesirable property for our purposes because we generally expect the direct effects of own payments to differ from the indirect effects of neighboring payments as discussed in the theoretical part of the article.

Additional reasons for our choice come from a rather forceful argumentation by Gibbons and Overman (2012) to consider the SLX model as a more credible alternative to the SAR model in many situations. They argue that the SAR model may suffer from an identification problem which is insufficiently addressed in the applied literature. This identification problem is similar to Manski's reflection problem (Manski 1993). Without additional information, restrictions or theoretical arguments it is hardly possible to determine if exogenous changes are caused by neighboring outcomes, $\mathbf{w}_i'\mathbf{y}$, and not by changes in neighboring characteristics $\mathbf{w}_i'\mathbf{X}$ (Gibbons and Overman 2012, p. 181). In other words, for identification we require variation in $\mathbf{w}_i'\mathbf{y}$ that is not caused by variation in $\mathbf{w}_i'\mathbf{X}$ or $\mathbf{w}_i'\boldsymbol{\varepsilon}$. In most cases estimation of the SAR model is based on (quasi) ML advocated by LeSage and Pace (2009) for which Lee (2004) showed that it provides consistent estimates under the condition that the spatial model is the true data generating process. Gibbons and Overman (2012) however argued that this assumption is difficult to defend since the spatial weighting matrix **W** is usually not known with certainty. Alternative to the ML approach, the identification problem of the SAR model can be addressed by a (2SLS) IV estimation approach proposed by Kelejian and Prucha (1998). Considering the reduced from of the SAR model given by

(18) $$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \rho\mathbf{w}_i'\mathbf{X}\boldsymbol{\beta} + \rho^2\mathbf{w}_i'\mathbf{W}\mathbf{X}\boldsymbol{\beta} + \rho^3\mathbf{w}_i'\mathbf{W}^2\mathbf{X}\boldsymbol{\beta} + [...] + v_i,$$

with $v_i = \rho\mathbf{w}_i'\mathbf{v} + u_i$ illustrates that the spatial lags $\mathbf{w}_i'\mathbf{X}, \mathbf{w}_i'\mathbf{W}\mathbf{X}, \mathbf{w}_i'\mathbf{W}^2\mathbf{X}$ etc. could be used as instruments. Gibbons and Overman (2012), however, argued that this approach also requires **W** to be known such that the exclusion restriction on the spatial lags $\mathbf{w}_i'\mathbf{X}, \mathbf{w}_i'\mathbf{W}\mathbf{X}, \mathbf{w}_i'\mathbf{W}^2\mathbf{X}$ etc. are justified. Concluding from these combined obstacles Gibbons and Overman (2012, p. 183) "ague for estimation of reduced form SLX models in $\mathbf{x}_i$ and spatial lags of $\mathbf{x}_i$, rather than direct estimation of the SAR or SD *[Spatial Durbin Model]* model". Further they state that "we believe that in many situations this 'reduced from' approach is simply more credible. The composite reduced form parameter describing the influence of neighbors' characteristics *or* outcomes is itself a useful and policy-relevant parameter. With this in hand judgment can be made based on theory and institutional context about the likely channels through which the effects operate (p. 183-184)." Following these lines of argumentation and considering that the main interest of the study is anyway on the influence of neighboring characteristics, the SLX model is chosen. As suggested by Gibbons and Overman the channel through which the influence occurred cannot be measured empirically but theoretical arguments can be provided which channels are most likely (here interaction on the land market, social or corporate networks etc).

Appendix C: Technical implementation of the SDEM model estimation using the GHK algorithm

The major difference between a standard and spatial probit model is that the likelihood function is based on multivariate instead of a univariate truncated normal distribution due to the dependence between observations. This increases computational needs particularly for large samples. For these cases Pace und LeSage (2011) proposed a simulated maximum likelihood framework capable of handling large sample sizes. Their approach is based on the GHK (Geweke-Hajivassiliou-Keen) algorithm to approximate the intractable multivariate integral of the multivariate truncated normal distribution. The general idea of the GHK algorithm in this context is to replace the joint multivariate truncated normal density by a product of conditional densities. This product of conditional densities has a sequential order in the sense that each conditional density only depends on prior variables in the sequence. Using specific realizations of the random variables allows calculating the sequence of conditional densities. By repeating the calculation R times, each time with different realizations of the random variables, a numeric approximation of the multivariate truncated normal distribution can be obtained. One obstacle of the approach with respect to large samples is that the number of operations required for the GHK algorithm depends on the number of non-zeros in the Cholesky lower triangular matrix of the covariance matrix[31]. Pace und LeSage (2011) argued, however, that in most spatial application each observation might only depend on a limited number of neighbors such that the sparsity of the variance-covariance matrix can be exploited in order to reduce the computation burden of the GHK sampler. They further propose to adopt the GHK algorithm to rely on a Cholesky decomposition of the precision matrix (i.e. the inverse variance-covariance matrix) instead of the variance-covariance matrix since in many situations it has greater sparsity. It

[31] For a dense variance-covariance matrix there are $n(n+1)/n$ non-zeros elements.

should be pointed out that the approach proposed by Pace and LeSage (2011) cannot only be used to estimate the SDEM model but also to other spatial model specification such as the SAR model.

In our specific implementation for the SDEM model we also rely on the precision matrix being equal to[32] $\Psi = (\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})$. As recommended by Pace und LeSage (2011) the sparsity of the precision matrix or variance-covariance matrix can be increased by an appropriate ordering of the observations. In our implementation we use the Matlab (Version R2013a) build in function *symamd*() for a symmetric approximate minimum degree permutation applied to the precision matrix to reorder the observations. For the GHK algorithm, we follow Pace und LeSage (2011) and employed scrambled Halton sequences where we skipped the first 1,000 values and used only every 101st value (Matlab default). For each likelihood evaluation we used $R = 15$. Optimization is performed with the Matlab Optimization Toolbox using a constrained maximization solver with an interior-point algorithm. Derivates are approximated numerically using forward differences. With our implementation it is possible to estimate the SDEM model with 64,488 observations in around 5.2h hours using Matlab Parallel Computing Toolbox with 12 workers on a Intel® Xeon® E5-2690 (2 processors) where we parallelize the R repetitions of the GHK sampler. This is lower as the speed reported in Pace und LeSage (2011), who claim to estimate a sample of size 100.000 in around four minutes on a standard laptop computer without parallelization, but since our focus is on a single estimation no further improvements of the implementation is pursued.

---

[32] The variance-covariance matrix is given by $Var(u) = \Gamma_\rho \Gamma'_\rho$ with $\Gamma = (\mathbf{I} - \rho\mathbf{W})^{-1}$, see for example Beron und Vijverberg 2004, S. 170–173.

Appendix D. Descriptive Statistics and Definition of Variable Codes (n=64488)

	Codes	**Units**	**Mean**	**Median**	**Max.**	**Min.**	**Std. Dev.**
Age of farm holder	*age*	year	48.83	49.00	97.00	7.00	11.58
Farm area	*area*	daa[a]	153.50	121.00	3411.00	0.00	132.45
Observed labor input	*obsLabo*	hour	2215.46	1900.00	52330.00	0.00	1827.00
Estimated labor requirement	*reqLabo*	hour	1950.39	1454.92	44452.84	9.79	1719.36
Total direct payments	*Dpay*	1000 Nkr	167.02	128.47	1252.47	0.01	132.06
Total market return	*mRet*	1000 Nkr	-33.87	-24.20	1403.76	-2606.99	66.27
Ratio observed over estimated labor requirement	*laboObs/Req*	ratio	1.37	1.13	83.32	0.00	1.33
Ratio leased area over total area	*landLease/Tot*	ratio	0.27	0.13	1.50	0.00	0.33
Dummy if farm has milk cows	*hasMilk*	binary	0.33	0.00	1.00	0.00	0.47
Dummy if farm has sheep	*hasSheep*	binary	0.33	0.00	1.00	0.00	0.47
Dummy if farm has poultry	*hasPoultry*	binary	0.01	0.00	1.00	0.00	0.08
Dummy if farm has sows	*hasSows*	binary	0.05	0.00	1.00	0.00	0.22
Tot. market ret. per labor req. in 1999	*mretrun/ reqLabo*	1000 Nkr/hour	-0.01	-0.02	0.24	-0.58	0.03
Tot. direct pay. per labor req. in 1999	*dpay/ reqLabo*	1000 Nkr/hour	0.09	0.09	0.42	0.00	0.03
Change in market returns per labor 99-09 structure equal to 1999	*C.mRetLabo*	1000 Nkr/hour	-0.05	-0.04	0.29	-0.15	0.03
Change in direct pay. per labor 99-09 structure equal to 1999	*C.DPayLabo*	1000 Nkr/hour	0.10	0.09	0.28	-0.17	0.04

[a] *daa = 1/10 ha.*

Appendix E. Sensitivity of SLX Regression Results with Respect to three Different Definitions of the Neighboring Relationships

	All within 2km radius		5 Nearest Neigh		Median dist. to furthest fields	
Variable	Coef	p-value	Coef	p-value	Coef	p-value
const	0.3954	0.0000	0.3953	0.0000	0.3948	0.0000
age	0.5615	0.0000	0.5567	0.0000	0.5581	0.0000
age^2	-0.6516	0.0000	-0.6475	0.0000	-0.6499	0.0000
area	0.1408	0.0000	0.1504	0.0000	0.1920	0.0000
area^2	-0.1063	0.0000	-0.1105	0.0000	-0.1190	0.0000
obsLabo	0.2500	0.0000	0.2538	0.0000	0.2622	0.0000
obsLabo^2	-0.1055	0.0000	-0.1081	0.0000	-0.1100	0.0000
reqLabo	0.1244	0.0000	0.1291	0.0000	0.1291	0.0000
mRet	0.0107	0.3453	0.0090	0.4300	0.0090	0.4286
dpay	0.8222	0.0000	0.8086	0.0000	0.7421	0.0000
dpay^2	-0.3591	0.0000	-0.3588	0.0000	-0.3477	0.0000
laboObs/Req	-0.0382	0.0000	-0.0386	0.0000	-0.0396	0.0000
landLease/Tot	-0.0405	0.0000	-0.0420	0.0000	-0.0441	0.0000
mretrun/reqLabo	0.0894	0.0000	0.0926	0.0000	0.1006	0.0000
dpay/reqLabo	0.0829	0.0000	0.0813	0.0000	0.0723	0.0000
C.DPayLabo	0.0661	0.0000	0.0752	0.0000	0.0967	0.0000
C.mRetLabo	0.0525	0.0001	0.0574	0.0000	0.0638	0.0000
hasMilk	-0.2467	0.0000	-0.2449	0.0000	-0.2254	0.0000
hasPoultry	0.0058	0.3791	0.0055	0.4026	0.0061	0.3554
hasSheep	0.0307	0.0000	0.0290	0.0001	0.0209	0.0031
hasSows	0.0375	0.0000	0.0389	0.0000	0.0421	0.0000
W_mRet	-0.0223	0.0480	-0.0053	0.5805	-0.0179	0.0539
W_dpay	-0.3040	0.0000	-0.2653	0.0000	-0.2708	0.0000
W_area	0.0633	0.0000	0.0774	0.0000	0.0721	0.0000
W_reqLabo	0.0886	0.0000	0.0517	0.0004	0.0617	0.0000
W_landLease/Tot	-0.0441	0.0000	-0.0482	0.0000	-0.0371	0.0000
W_FarmWage	0.0490	0.0000	0.0272	0.0027	0.0345	0.0015
W_C.inco	0.0639	0.0000	0.0394	0.0000	0.0498	0.0000
W_hasMilk	0.0765	0.0000	0.0892	0.0000	0.0774	0.0000
W_hasPoultry	0.0059	0.3131	0.0034	0.5631	0.0094	0.1084
W_hasSheep	0.0018	0.8075	0.0043	0.5643	0.0186	0.0090
W_hasSows	0.0304	0.0000	0.0213	0.0004	0.0144	0.0163
n		64488		64488		64488
% Correct predictions Model		72.80		72.78		72.63
% Correct predictions Naive		62.72		62.72		62.72

Note: Neighbors defined as 1) all farms with a radius of 2km, 2) five nearest farms and 3) all farms within a radius of the regional median furthest driving distance to field.

Appendix F. Correct Predictions of Farm Survival between 1999 and 2009 with Different Model Specification with Respect to the Absolute Size of a Farm

	Naive	All other non-spatial explanatory variables					Full Model
			and Area	and obs. Labor	and est.req Labor	and direct payments	
% Correct	62.72	67.58	71.82	71.48	71.85	72.49	72.59
Diff. to full M.	-9.88	-5.01	-0.78	-1.11	-0.75	-0.11	0.00

Note: Presented results refer to the non-spatial binary choice probit model.

Appendix G. Correlation Coefficients between Different Measures of the Absolute Farm Size

	Area	Obs. labor input	Est. labor requirement	Total direct payments
Area	1	0.44	0.65	0.62
Obs. labor input		1	0.78	0.70
Est. labor requirement			1	0.85
Total direct payments				1

Source: Own calculation.

Appendix H. Correct Predictions of Farm Survival between 1999 and 2009 with Different Model Specification with Respect to the On-Farm Wage

	Naive	All non-spatial explanatory variables except			Full Model
		on-farm wage	changes in on-farm wage	on-farm wage and changes in on-farm wage	
% Correct	62.72	72.53	72.34	72.17	72.59
Diff. to full M.	-9.88	-0.07	-0.26	-0.43	0.00

Note: Presented results refer to the non-spatial binary choice probit model.

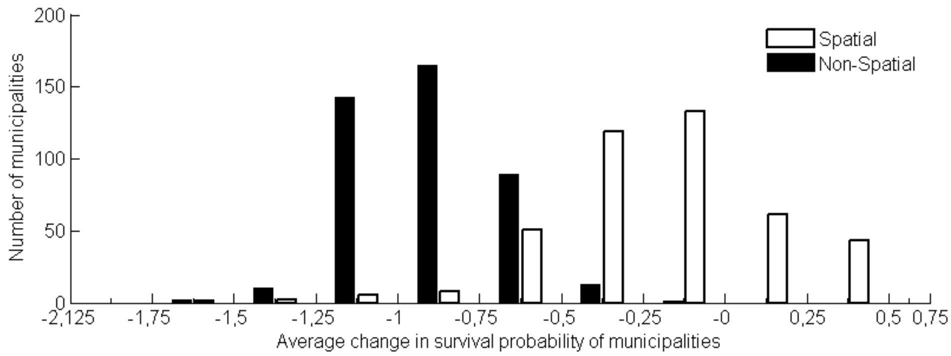
Appendix I: Description of the Bootstrap to derive the interval of the difference in predicted probabilities

In order to derive the interval of the average change in predicted survival probabilities a bootstrap procedure is applied to the scenario calculations. The following steps are performed specifically:

1. Relying on the asymptotic normality of the ML estimator a random draw from $\tilde{\boldsymbol{\beta}}_r \sim \mathcal{N}\left(\hat{\boldsymbol{\beta}}, \boldsymbol{\Gamma}\right)$, with $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\Gamma}$ being the ML estimate and covariance, respectively, is obtained.
2. For $\tilde{\boldsymbol{\beta}}_r$ the scenario simulation are performed and the mean change in survival probability between the baseline and the scenario is calculated.
3. Step 1 and 2 are repeated for $r = 1, \ldots, R$.
4. Based on the bootstrap sample of size $R = 2000$ the 95% interval of predicted probabilities is calculated as the 2.5% and 97.5% quantile of the sample.

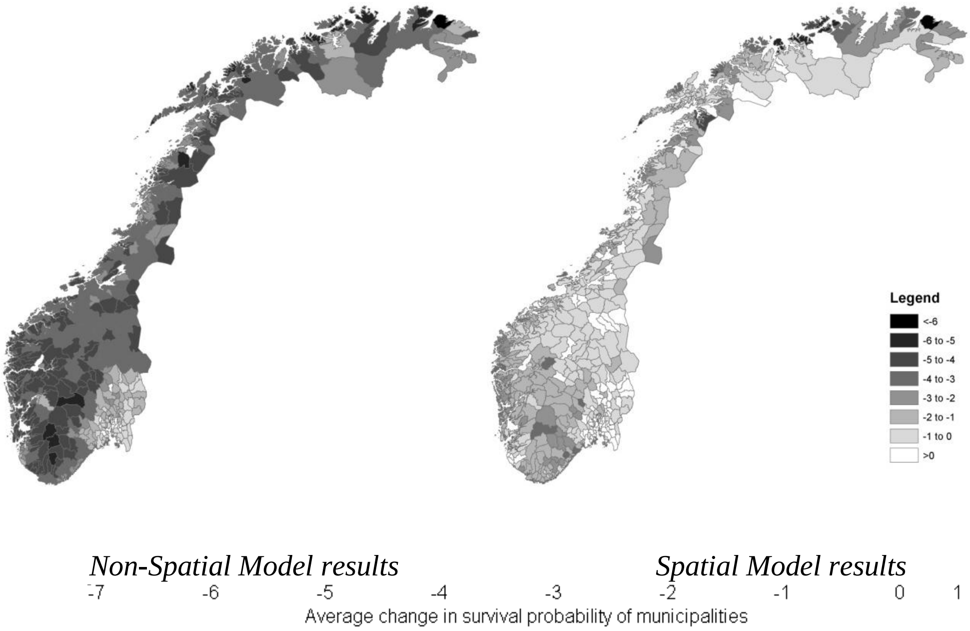
Appendix J. Average change in survival probability of municipalities for a 10% reduction of all direct payment rates

Non-Spatial model results *Spatial model results*



Source: Own illustrations. Shape files were derived from the GADM database (www.gadm.org), version 2.0.

Appendix K. Average change in survival probability of municipalities for an abolishment of the structural dimension of direct payments in which rates are set equal to the lowest rates currently paid



Source: Own illustrations. Shape files were derived from the GADM database (www.gadm.org), version 2.0.