European Commission

JRC SCIENTIFIC AND POLICY REPORTS

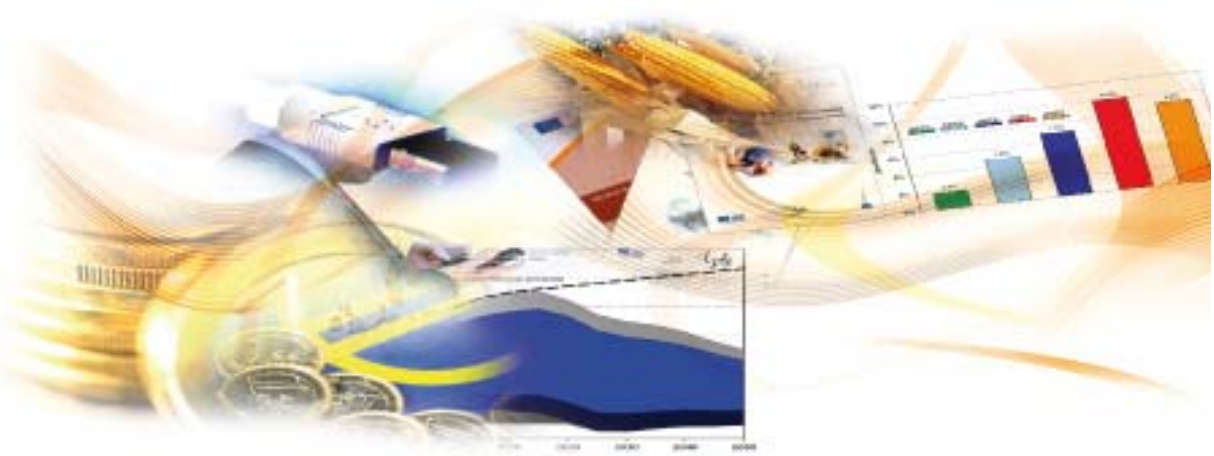
Modelling farm structural change

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

Authors:
Alexander Gocht, Norbert Röder, Sebastian Neuenfeldt, Hugo Storm, Thomas Heckelei

Editors:
María Espinosa, Sergio Gomez y Paloma

2012



Report EUR 25555 EN

European Commission
Joint Research Centre
Institute for Prospective Technological Studies

Contact information
Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: jrc.ipts.secretariat@ec.europa.eu
Tel.: +34 954 48 8458
Fax: +34 954 48 8434

http://ipts.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

This publication is a Scientific and Policy Report by the Joint Research Centre of the European Commission.

JRC 75524

EUR 25555 EN

ISBN 978-92-79-27047-5 (pdf)

ISSN 1831-9424 (online)

doi:10.2791/17545

Luxembourg: Publications Office of the European Union, 2012

Printed in Spain

Acknowledgements

This study has been developed under the project "Modelling the effects of the CAP on farm structural change" (Contract 151949-2010-A08-DE). The project team would like to acknowledge the advice and assistance they have received from Maria Espinosa and Sergio Gomez y Paloma (DG JRC/IPTS). The results of the project were presented at the expert workshop in the Institute for Prospective and Technological Studies (IPTS) on the 7th/8th February 2011. We are indebted to the participants, in particular to Thierry Vard and Hubertus Gay (DG-AGRI) for their constructive and helpful comments and recommendations.

Table of Contents

List of Tables

List of Figures

Executive Summary

Study objectives

This study had the following specific objectives. The first objective was to review the existing methodologies for ex-ante and ex-post modelling of structural change. The second objective was to develop prototypes of analytical tools for ex-post and ex-ante analysis of structural change in agriculture based on FADN (Farm Accountancy Data Network) and FSS (Farm Structure Survey) data. The ex-ante incorporation of structural change was conducted in the farm module of CAPRI (Common Agricultural Policy Regionalised Impact System).

These prototypes were developed and tested for selected German NUTS-2 regions.

The report structure proceeds along the above-defined objectives. The literature review is presented in part II of the report, while the prototype analytical tool is presented in part I, differentiating ex-post and ex-ante analysis.

Hence, the general objective was to develop tools to improve analysis, project farm structural change and enable CAPRI to consider the effects of both endogenously driven and policy-promoted structural change. In particular, structural change influences the efficiency of production in various dimensions and consequently affects the allocation of resources to different production activities.

Results

Eight main achievements of the current study are worthy of explicit mention. Whereas the first achievement is linked to a better understanding of the limits and opportunities of the underlying data sources, the remaining achievements concern methodological enhancements.

The **first** achievement is the identification of significant problems in linking structural change in agriculture to external drivers. These problems are related to the manner in which the FADN data are sampled and processed. An important issue is that the use of time-varying Standard Gross Margin (SGM) or Standard Output (SO) induces additional volatility in the observed changes in farm specialisation. Furthermore, the more or less automatic removal of farms that alter their 'type of farming' or Economic Size Unit (ESU) leads to a much lower mobility in the data when compared with reality. Finally, the most mobile farms (the relatively small ones) are not sampled. These peculiarities generate problems for empirical approaches when analysing structural change based on FADN micro data.

The **second** achievement is the development of a methodology to both analyse and quantify structural change in productive orientation on a continuous scale. Such methodology allows a finer separation of random and directed movements of direction and strength compared to analysis based at the farm typology level.

The **third** achievement is closely linked to the second achievement. To our knowledge, the mathematical framework of Multiplicative Competitive Interaction (MCI) models has not been applied in the context of modelling structural change in agriculture, particularly with respect to changes in productive orientation. MCI models use the individual farm as the level of observation, whereas Markov models use topologically and/or topographically delimited groups of farms. The higher number of observations in the MCI model may lead to more robust estimation results and permits the inclusion of additional explanatory variables without over-fitting the model.

The **fourth** achievement is linked to the intended use of the results of the ex-post analysis in ex-ante projections. We developed an approach to assess the quality of the ex-post analysis using an out-of-sample validation.

The **fifth** achievement is the estimation of non-stationary Markov transition probabilities using a Bayesian estimation framework that allows available data sources at the EU level (the Farm Structure Survey (FSS) and the FADN) to be combined in a consistent way. The combination of both data sets exploits their specific advantages while limiting the effect of their shortcomings. The empirical application was restricted to Germany; however as FSS macro data and FADN micro data are available for the entire EU, applications to other EU regions follow straightforwardly.

Achievement number **six** is an evaluation of the appropriateness of the Markov approach for predicting structural change in comparison to naïve prediction methods. The results show that the Markov prediction may outperform naïve prediction methods but that the quality of the prediction is critically dependent on the model specification. A higher in-sample fit does not necessarily lead to better out-of-sample prediction, which potentially indicates that the effects of specific explanatory variables may change over time.

The **seventh** achievement is a novel proposal to analyse the effects of explanatory variables in the Markov model. A common problem in the use of non-stationary Markov models is that there is no direct interpretation of the estimated coefficients; in addition, the calculated marginal effects cannot be intuitively interpreted. To alleviate this problem, we introduced a novel method of directly analyzing the effects of one variable on the movement of farm.

The **eighth** achievement is the development of a conceptual approach to implement farm type structural change in the baseline generation of CAPRI-farm and the implementation of the suggested approaches to one NUTS-2 region. The presented example demonstrates

that information from Markov projections can be used to derive a consistent farm grid for a simulation year that represents the on-going structural change.

Outlook

The following potential further research for the improvement of the ex-post analysis of structural change and its implementation in the ex-ante models should be considered.

First, the use of FSS micro-data will enable to isolate the effect of changing prices and policies from the effect of physical asset changes on the attribution of farms to 'types of farming' and 'Economic Size Unit'. The easiest way to isolate the effect of changes in physical assets is to base the valuation of the different activities on farm level on fixed coefficients (Standard Gross Margin / Standard Output) and to derive the sampling weights for the FADN farms based on these updated FSS data.

Second, if the analysis of structural change is to be emphasised more strongly in the analysis of FADN-data, the sampling procedure should be improved to prevent the automatic removal from the sample of farms that alter their type of farming.

Third, the use of MCI or multinomial logit models to directly estimate the shares of different farm types at the regional level should be investigated. In comparison to Markov models, a successful application of these types of models would greatly reduce the number of parameters to be estimated and therefore increase the robustness of the results.

Fourth, the following empirical questions remain open, independent of the chosen analytical tool. How relevant are different sub-processes (farm exit, farm entry, farm shrinkage and farm growth, both with and without change in specialisation) with respect to changes in overall farm specialisation on the regional level? Is the relative importance of these processes comparable across space and time?. To answer these questions, more studies with micro-level panel data are needed.

PART I: DESCRIPTION OF THE PROTOTYPE ANALYTICAL TOOLS TO ASSESS FARM STRUCTURAL CHANGE

Introduction

This part of the report ("Description of the prototype analytical tools to assess farm structural change") consists of three main chapters, in chapter 1 and chapter 2 are presented respectively the ex-post and ex-ante methodology, while chapter 3 concludes with the general remarks and potential further research. In Chapter 1, the methodology, data preparation steps, designs chosen for the estimation exercises and applied test statistics are discussed. Afterwards, the results are presented, and the chapter concludes with a summary and the conclusions of the ex-post analysis. In particular, two methodological approaches are considered. The first views structural change from the discrete perspective using a Markov approach, whereas the second uses the continuous perspective to evaluate the type of farming over time using MCI (Multiplicative Competitive Interaction) models.

The second chapter addresses the use of the analytical tool to assess structural change ex-ante. The baseline trend estimation for the farm types in CAPRI (Common Agricultural Policy Regionalised Impact Modelling System) is extended to include a consideration of the information regarding the evolution of the number of farms. The chapter starts with an explanation of the current baseline estimation and develops several approaches to include structural change in the baseline process. Subsequently, the empirical implementation of the suggested approaches is described. The farm types and the region used to demonstrate the approaches are presented, followed by the results section, in which the current implementation is compared to the suggested approach. This chapter ends with a conclusion.

In the third chapter the general conclusions and outlook of the ex-post and ex-ante analysis are presented.

1 Ex-post modelling of structural change

The research performed in this study ultimately aims to perform ex-ante policy impact analysis using methodologies generally applicable to the entire EU. The primary aim of this research is to develop a robust approach to predict structural change (defined as the change in production systems and farm entry/exit into one sector/farm typology). These robust estimates could be included in the baseline of policy assessment tools such as CAPRI. The second goal is to determine relevant factors for observed structural changes in the past. A general methodology was developed for including factors that affect structural change into an ex-ante policy evaluation model. The scientifically sound identification of relevant determinants for farm structural change can only be performed by statistical/econometric analysis of observed developments in the past. The chosen Markov analysis and the analysis of changes of the specialisation on the farm level derived from the MCI model are also appropriate for fulfilling two other criteria: (1) applicability to the entire farm population and (2) ability to project farm numbers between FSS survey years in the FADN grid (dependent on the farm type classes). In particular, an ex-ante policy assessment will profit from the inclusion of determinants of structural change, if these determinants have proven that they yield robust estimates and do not capture peculiarities observed in a particular period. For this reason, we do not focus the assessment of the quality and reliability of a given variable on the in-sample test as in ordinary regression analysis but rather focus on the ability to improve the modelling of structural change on the out-of-sample data. Therefore, we split the data into a training set, from which the parameters of the models were estimated, and a test set to calculate the fit statistics of the model.

1.1 Design of the out-of-sample prediction

In the following section, the design of the out-of-sample prediction is described. Here, the performance of different methods and data sources to project farm numbers in different farm types and size classes is compared. Two methods are of particular interest: one based on estimated stationary or non-stationary Markov probabilities and one based on continuous trajectories. The two approaches differ in principle because of the scale on which they operate. The Markov approach works primarily on a regional scale and uses farm level observations only to reduce the degrees of freedom involved in the estimation of the transition probability matrix. The MCI approach in the presented implementation works primarily on the farm level (estimating the shares of the different specialisations). These results are used in a second step to derive the new farm type for each farm and, with the help of weighting factors, to derive the share of a given type of farming in a region.

Stationary and non-stationary transition probabilities (TPs) are estimated from the FADN micro data and the corresponding macro data grid available from the EuroStat FSS (Farm

Structure Survey). A Bayesian estimation framework is employed to combine micro and macro data, thereby allowing two different data sources to be combined in an estimation. The continuous MCI approach is intended to explain the change of a farm's type of farming as the result of underlying changes in the relative importance of the different farm specialisations. The MCI approach was developed in marketing theory to explain the shares of ***discrete*** brands in a given market. Therefore, if the MCI approach is used on the farm level, farm size cannot be simultaneously incorporated as a dependent variable. If the MCI approach is used on a regional scale, the share of any group of farms can be used as a dependent variable. The performance of the MCI and the Markov chain approach is compared to naive extrapolations of farm numbers based on a linear trend prediction, a geometric trend projection and a constant prediction in which it is assumed that the farm structure is the same as it was in the last recorded year.

To evaluate the performance of the different Markov projections, out-of-sample predictions are performed for two different time periods. For the analysis, FADN data and the corresponding macro data grid from EuroStat are available from 1989 to 2008. In the first projection period, farm numbers are projected over a four-year interval beginning in 2003. An eight-year period beginning in 1999 was used in the second interval. In both cases, the predicted number of farms is compared to the observed number of farms in 2007, which is the last available FSS year. The years from 1989 to 2003 and from 1989 to 1999 are used to estimate the stationary and non-stationary TP FADN data for the shorter and longer projection periods, respectively. The linear and geometric predictions use the same macro data and time periods.

The regional coverage considered for the out-of-sample experiment differs between the projection methods (Table 1). The linear and geometric projections and the projection based on the stationary TP, which is estimated from FADN data, are performed for all FADN regions in Germany. Some FADN regions are excluded because of an insufficient number of observations (Hamburg (20), Bremen (40), Saarland (100) and Saxony-Anhalt (115)). The non-stationary projection and the stationary projection for the prediction period 1999-2007, however, are restricted to the West German FADN regions (excluding 20, 40, and 100), for which a time series from 1989 to 2008 is available.

Table 1: Characteristics of the different methodologies used in the out-of sample prediction

Projection Method		FADN Regions considered	Input Data (Dependent variable)	Data sources for explanatory variables	Quality measure used for the out-of sample prediction	Predicted variable
Naive Trend	**Linear** **Geometric** **Constant**	*Prediction period 2003-2007: FADN region: 10, 30, 50, 60, 70, 80, 90, 112, 113, 114, 116* *Prediction period 1999-2007: FADN region: 10, 30, 50, 60, 70, 80, 90*	FSS macro data	---	Mean Square Error (MSE); Mean percentage deviation (pDev)	Number of farms per region and size class
Markov	**Stationary**	ditto	FSS macro data + FADN micro data	---	ditto	ditto
	Non-Stationary	FADN region: 10, 30, 50, 60, 70, 80, 90	ditto	FADN, DeStatis		
MCI	**Stationary** **Non-Stationary**	All German FADN regions	FADN micro data	FADN, DeStatis	Root Mean Squared Error (RMSE)	Step 1) share of farm specialisation per farm Step 2) weighted to obtain shares of farm types per region

In the Markov approach, we evaluate the performance of the out-of-sample prediction based on the calculation of mean square error (MSE) and mean percentage deviation (pDev) for each prediction method in each region. The MSE is calculated for each prediction method and region as the mean over the square difference between the predicted farm numbers and the observed farm numbers in each class[1]. Similarly, the mean percentage deviation is calculated as the mean over all farm types and classes of the percentage deviation of the predicted farm number from the true value derived from FSS. The latter measure is not defined in cases in which the true number of farms in a class is equal to zero; in this situation, the observation is excluded from the calculation of the mean percentage deviation.

[1] The class definition (typology) is described in section 2.2.2.

The linear prediction is performed for each region, farm type and size class by estimating γ_1 and γ_2 of the linear function $n_t = \gamma_1 + \gamma_2 t$, 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. Macro data derived from FADN and the corresponding macro data grid available from EuroStat are used for the estimation. As in the Markov prediction, two time periods, one from 1989 to 1999 and one from 1989 to 2003, are considered in the estimation to predict farm numbers in 2007. The geometric growth rate is derived by estimating $\ln(n_t) = \lambda_1 + \lambda_2 t$. Farm numbers in $t+1$ are predicted using the estimated parameters $\hat{\lambda}_1$ and $\hat{\lambda}_2$ to calculate $\hat{n}_{t+1} = e^{\left(\hat{\lambda}_1 + \hat{\lambda}_2 (t+1)\right)}$. The data source and time periods are the same as those used for the linear prediction. An advantage of the geometric prediction is that in the linear case, the predicted farm number can become negative. However, problems arise 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.

We conduct out-of-sample predictions for the MCI approach analogously to the Markov approach for two different time periods. In the first case, we predict the change in the types of farming between 2003 and 2007 based on estimations derived from data recorded between 1989 and 2003. For the second out-of-sample prediction, only data from 1989 to 1999 are used to predict the farm structure in 2007. In contrast to the Markov approach, the MCI approach is purely based on the observed transitions at the farm level. Therefore, FADN regions with fewer records can remain in the sample. Only when the fit of the predictions is evaluated is it sensible to merge the data of smaller groups with similar larger groups to prevent deviations caused by small samples sizes that may affect the overall result.

As the MCI approach is based purely on the transitions observed in the FADN sample, it cannot depict changes in farm numbers, as exits from the sector are not recorded in FADN. Therefore, the evaluation of the out-of-sample prediction is based on the match of the predicted and observed farming structure, i.e., the share that the different types of farming have in the population. The root mean square error is chosen as an indicator and is calculated as follows:

(1): $$RMSE = \sqrt{\sum_r w_r \left(\sum_t \left(s^e_{t,r} - s^o_{t,r}\right)^2\right) / \sum_r w_r}$$

where t is the type of farming, w_r is the number of farms in region r, and s^e and s^o are the estimated and observed share, respectively. Table 1 provides a summary of the characteristics of all prediction methods with respect to the regional coverage that the data sources and the employed fit measure.

1.2 Data

1.2.1 Main data sources

Several aspects of the FSS and FADN databases differ (Table 2), which has implications for the analysis of structural change. First, FADN is not capable of depicting farm exits because it is a sample of the total population. Second, no inferences regarding the development of small farms can be made from FADN as small farms are not sampled. FADN is better suited to identify the influence of factors varying over time (e.g., market conditions) because of the high temporal resolution and the long time series. The high spatial resolution of FSS permits the formation of more reliable conclusions regarding factors varying over space (e.g., production conditions, availability of off farm income) and is less prone to problems regarding aggregation errors. An important practical difference regarding the development of methodologies to analyse structural change is that, within this study, the authors have access to FADN micro-data but not direct access to FSS micro-data. Therefore, to reduce the time demands associated with the implementation and debugging of the analytical programs, all analyses are first applied and tested using the FADN dataset. In particular, full access to FSS data would permit the development of incremental models capable of isolating the effect of changing SGM or SO.

Table 2: Differences between the FSS and FADN databases in Germany

	FSS	**FADN**
Type of data	Full population	Representative rotating sample of commercial farms; extrapolation to the population based on farm-specific weighting factors
Cut off limit	> 2 ha or > 8 LU	SGM: 1989-1998 (> 8 ESU) / 1999-2008 (> 16 ESU)
Sampling frequency	3-4 years interval	Annually
Time series	1999, 2003, 2007	1989-2008
Spatial resolution	LAU 2 (Local Administrative Unit)	Farms identified at NUTS 3 level but sampling scheme based on NUTS 1 level
Information	Structural	Structural and financial

The German FADN sample provides a unique opportunity to use this approach because the farms remain in the sample for a long duration (Table 3). More than 12,000 farms remained in the sample for at least 4 years.

Table 3: Length of time that farms remained in the German FADN sample (1989-2008)

Years in the sample	N° of farms
1	4334
2	2628
3	2642
4	1717
5	1391
6	1549
7	1133
8	1020
9	974
10	1366
> 10	3024

To retrieve the CAPRI classification of farm types, we use 8 data dimensions (Table 4). This classification fulfils the criteria of Equation (28) in Annex 3, in which the partial standard Gross Margin should equal unity. With the exemption of grazing livestock and permanent crops, the name of the dimension corresponds to the name of the specialised FADN types of farming on the 2-digit level. Extra dimensions are needed for grazing livestock. T4D corresponds to dairy cattle, whereas T4X sums the SGM of all male bovine, suckler cow, non-bovine and forage cropping activities. T4X is also needed to calculate the main specialisation (P1,…, P5). All decision variables for constructing the 2-digit FADN classification can be calculated with the proposed dimensions: e.g., for type of farming 71 (P1 = T13 + T14; P2 = T20; P3 = T30; P4 = T41 + T4D + T4X; P5 = T50).

Table 4: Classification of the FADN variables describing the farm structure to retrieve the 2-digit farm-classification and the CAPRI farm classification for Germany

Dimensions	FADN variables considered to calculate the share of the total SGM to define the step length[2]
T13	D/1-D/9; D/22; D/26-D/30
T14	D/10; D/11; D/14a; D/19; D/20; D/23-D/25; D/31- D/35
T20	D/14b; D/15-D/17; I/2
T30	G/1 - G/7
T41	J/2; J/4; J/6
T4D	J/7
T4X	J/1; J/9; J/10; J/3; J/5; J/8; (D/12; D/18, F/1; F/2)
T50	J/11-J18

The acronyms of the FADN variables are explained in Annex 1.

[2] In Annex 2, there is a description of the calculation of the step length.

1.2.2 Applied typology for the estimation

Data constraints and computational consideration of the Bayesian estimation approach need to be addressed to determine the degree of detail of specialisation and size classes for which farm numbers are projected, which limits the maximum number of classes. To maintain a sufficient degree of detail while keeping the number of classes manageable, farms belonging to classes characterised by very specialised production programs, such as *Specialist Horticulture*, *Specialist Vineyards, Specialist Fruit And Citrus Fruit, Specialist Olives* or *Various Permanent Crops Combined,* are excluded from the analysis. Additionally, farm types with only a small number of farms in the sample are combined with other farm types that have relatively similar production programs. Finally, five different farm types, as defined in Table 5, are considered in the analysis. As a second dimension, in addition to the five different farm types, three size classes, defined as (I) 16-<40 ESU, (II) 40-<100 ESU and (III) >100 ESU, are considered. The restriction to three size classes and the specific definition of size classes are given by the FADN clustering scheme employed for Germany over the period 1989 to 2007.

Table 5: Definition of farm types considered for prediction

ID	Name	TF14 Description	TF14 Code
1	COP crops	Specialist Cereals, Oilseed And Protein Crops; Specialist Granivores	TF13; TF5
2	Other crops	Specialist Other Field Crops; Mixed crops	TF14, TF6
(3	Horticulture	Horticulture, Permanent Crops	TF2; TF3)
4	Milk	Specialist Milk	TF41
5	Other livestock	Specialist Sheep and Goats; Specialist Cattle	TF42, TF43, TF44
6	Mix	Mixed Livestock; Mixed Crops and Livestock	TF7, TF8

Horticulture (ID 3) is only considered in the MCI approach.

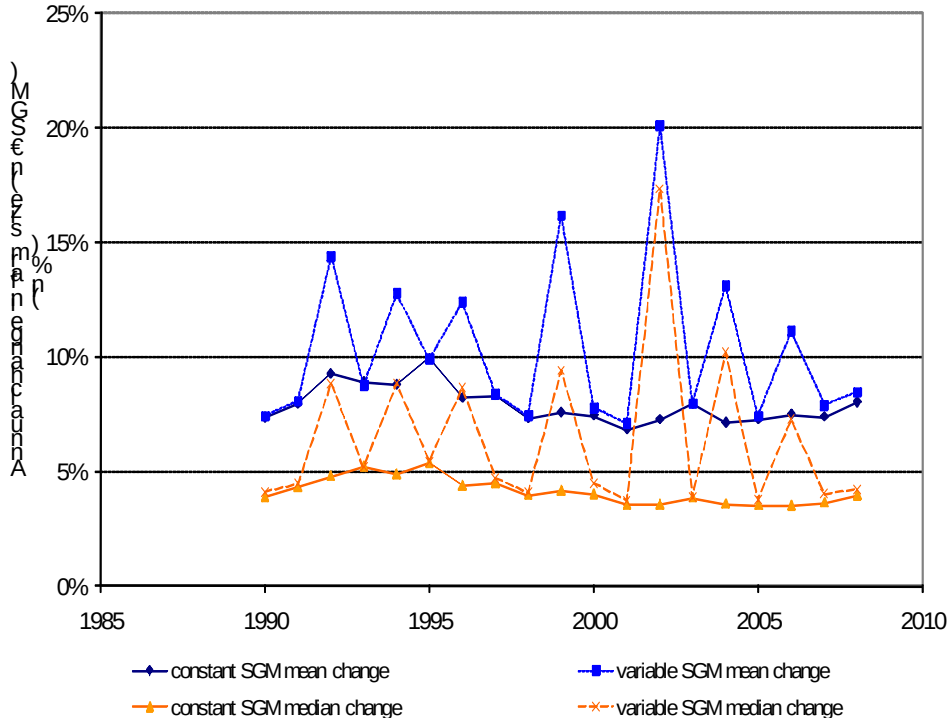
1.2.3 Treatment of the SGM

As solid information on structural change, i.e., the modification of the farm's physical layout (e.g., machines, building and labour), is lacking in FSS and FADN statistical data, we use the agricultural enterprises crop and livestock shares based on the Standard Gross Margin (SGM) as a proxy of structural change. The SGM of a crop or livestock item is defined as the value of output from one hectare or from one animal less the cost of variable inputs required to produce that output. The SGM is used to define the Community Typology for agricultural holdings according to the Commission Decision 85/377/ECC amended by the Commission Decisions 96/376/EC; 96/373/EC; 1999/725/EC and

2003/369/EC. In Germany, the SGM is calculated for each NUTS 2 region and is updated annually. In the FADN database, to determine a farm's farm type, the activities are weighted with a three-year average SGM. This procedure particularly dampens the impact of short-term price fluctuations. However, this average SGM is not a dynamic average but is kept constant for two or three years.

A pivotal issue in determining structural change over a longer time period is whether to use a time-variable or time-invariant (constant) SGM. A change to a time-invariant SGM could result in a different attribution (size and type of farming) for a given farm in both the FSS and FADN databases. Consequently, the FADN weighting factor must be adapted. Because time constraints, EuroStat was not capable of recalculating the grid based on constant SGMs within the time frame of this study.

Figure 1: Comparison of the average (mean and median) annual changes in farm size (measured in € SGM) of German farms over time if the farms' activities are weighted with a constant or variable SGM.

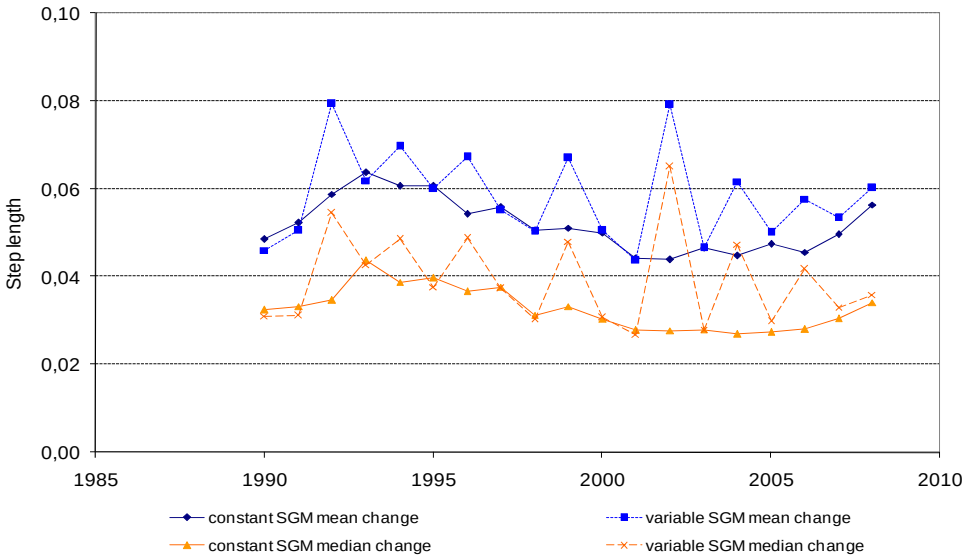


Source: Own calculation based on the German FADN-farms during the period 1995-2007. Only farms that remained in the sample for at least two consecutive years are included.

However, the use of a variable SGM leads to severe problems in the estimation process and to counterintuitive results. First, the use of variable SGMs introduces additional dynamics regarding structural change that are not mirrored by a change in physical assets. When the SGM was updated, the recorded changes were 30% to 300% larger than those in years without an update. The updating influences the observed dynamics with regard to the changes in farm size (Figure 1) and also the specialisation (Figure 2). In years without an update of the SGM, the dynamic is independent of the year selected to weight the

farm's activities. In the short run (interannual), most farms only marginally alter their structure in terms of both size and specialisation (see Annex 3, Figure 22-Figure 25). The distributions are right-skewed because the medians of the distributions are markedly lower than the means.

Figure 2: Comparison of the average (mean and median) annual change in the farm specialisation[a] of German farms over time if the farms' activities are weighted with a constant or variable SGM.



Source: Own calculation based on the German FADN-farms in the period 1995-2007. Only farms that remained in the sample for at least two consecutive years are included.

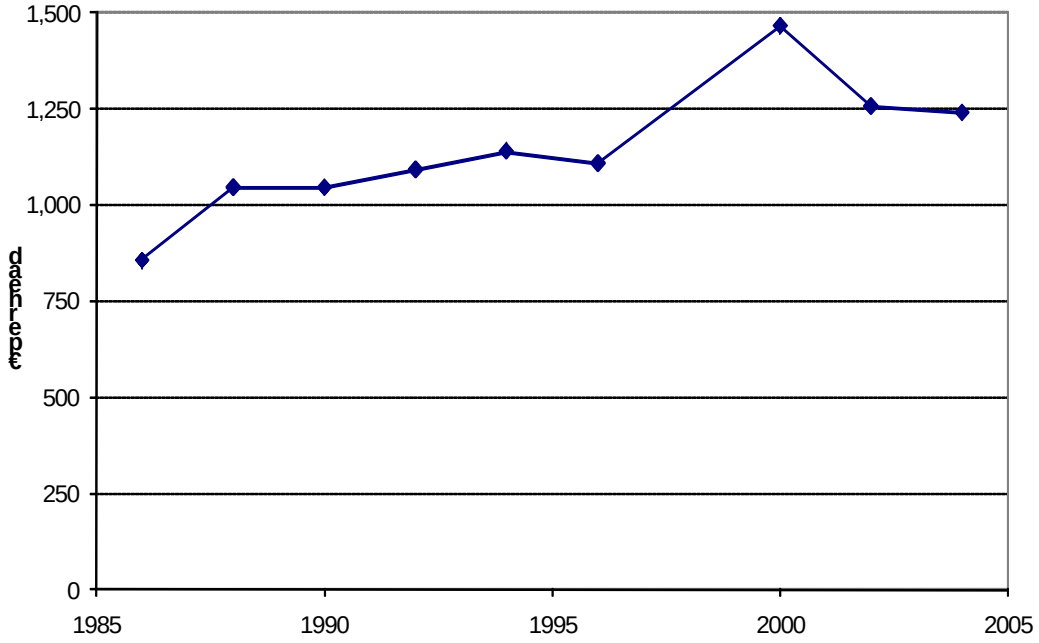
a) measured by the step length using Manhattan Block metrics (for the calculation see Annex 3).

Second, even if three-year averages of the SGM are used, the SGM varies quite significantly over time. Figure 3 depicts the development of the SGM per dairy cow over time for one German region. The price spike in 2000, which affected the SGM for 2002 and 2003, can clearly be observed. Only this price fluctuation would lead to an increase in the economic farm size of a pure dairy farm by more than 30% from 2001 to 2002 and a decrease by 15% from 2003 to 2004[3]. Consequently, an approach that determines structural change on the basis of a variable SGM would largely depict the changes in input and output prices but hardly any structural adjustments of the farms. Therefore, the calculations are based on the German NUTS 2 SGM in 2002. The weighting factors in the FADN sample were kept as stated for the MCI approach. In the Markov approach, in which valid weighting factors are required to derive the macro data, we use only fixed SGM for the time period of from 1999-2007, for which it is possible to recalculate farm weights based on the available FSS micro data. For years before 1999, it is not possible to

[3] For the FADN years 2001, 2002, 2003 and 2004, the following FSS SGM were used, respectively: 1996, 2000, 2000 and 2002.

recalculate weights based on a fixed SGM because of the lack of FSS micro data. Consequently, variable SGMs are considered in the period of 1989 to 1999, whereas in the period of 1999 to 2007, the SGMs in 2002 are considered to calculate the economic size, farm type and weighting factor for each observation.

Figure 3: Development of the FADN-SGM per dairy cow in Upper Bavaria (DE21) between 1986 and 2004.



Source: EuroStat.

The presented methodology is independent of the method used to weight the activities (variable or constant SGM or Standard Output (SO)). Nevertheless, a change from, e.g., variable to constant SGM will have implications for the obtained results. After the accounting year 2010, the typology for agricultural holdings will be based on the SO (Commission Regulation 1242/2008/EC) rather than the SGM. The classification in both databases (FSS and FADN) is based on the SGM. This classification cannot be easily transferred to a SO-based classification as SO values are currently only available for 2004 and a change of the economic weight of the activities (from SGM to SO) would require the recalculation of the weighting factors in the FADN database. Therefore, we use the SGM approach in this study.

The farms are attributed to a type of farming based on the principles laid down in Commission Decision (2003/369/EC). As the data set with the lowest resolution determines the overall achievable data resolution, we have to omit the information for size classes 5 and 6 for the years prior to 1999, and we will interpret the output for the merged classes (9 and 10) because these classes were merged prior to 2004 (FADN, 2010).

1.2.4 Explanatory variables

Table 6 lists the explanatory variables that were identified in previous studies as determinants of structural change. The explanatory variables used in the non-stationary estimation of TP and the continuous approach are derived from the FADN dataset and from the public German database DeStatis. The geographical and time period availability are specified. The identified variables are broadly distinguishable by time and regional varying variables. Not all of the variables described in Table 5 were considered in the empirical applications because of limited data availability or high multicollinearity between explanatory variables. The final set of variables considered in the estimations is described in Table 8 and Table 9.

Table 6: List of potential variables to be considered in an ex-post analysis and the likely data sources on the EU level

Group deter-minant	Indicator	Proxy	Data source	Variable definition (con-struction)	Unit	Geo-graphical resolution[e]	Years	Spatial coverage
Technology								
	Yields		FADN[a]	=Table_K_column_qq / Table_K_column_AA	tons / ha or tons per head	single farm, FADN-region	1989-2008	EU
			DeStatis		tons / ha or tons per head	NUTS 3	1974-2010	Germany
			EUROSTAT[b]	agr_r_crops	100 kg/ha	NUTS 2	1977-2010	EU
Farm structure								
	Initial farm size		FADN	=A27	Econ.size in EUR	single farm, FADN-region	1989-2008	EU
	Initial farm specialisation		FADN	=A30	4-digit Calculated by DG AGRI (c.f. Typology Regulation)	single farm, FADN-region	1989-2008	EU
	Farm size heterogeneity		FADN	Gini-Index; Shannon-Index		FADN-region	1989-2008	EU
	Stocking densities		FADN	=SE080/SE025	LU per ha	single farm, FADN-region	1989-2008	EU
	Share of mixed farms		FADN	=Sum(SYS02*(A30>=6000))/Sum(SYS02)	%	FADN-region	1989-2008	EU
Market conditions								
	Output prices		FADN	=(SA+CV+FU+FC-BV)/QQ	EUR per ton	single farm, FADN-region	1989-2008	EU
			DeStatis	Index	Index	NUTS 0 (MS)	1989-2008	Germany
	Input prices		DeStatis	Index	Index	NUTS 0 (MS)	1989-2008	Germany
	Prices		EUROSTAT	Agr-r-aacts_	EUR	NUTS 2	1974-2010	EU
	Land rent		FADN	=SE375/SE025	EUR / ha	single farm, FADN-region	1989-2008	EU
	Land rent		DeStatis		EUR / ha	NUTS 2	1989-2008	Germany

Group deter-minant	Indicator	Proxy	Data source	Variable definition (con-struction)	Unit	Geo-graphical resolution[e]	Years	Spatial coverage
CAP								
	1st pillar CAP payments		FADN	Table M	EUR	single farm, FADN-region	1989-2008	EU
			FADN	Table J	EUR	single farm, FADN-region	1989-2008	EU
	2nd pillar CAP payments		CATS (Clearance of Account database)		EUR	NUTS 3	2000-2005	EU
	2nd pillar CAP payments		Programmes Guidance, Guarantee, SAPARD, Objective 1, IFDR, LEADER, EU27 (Agrex, Agriview)		EUR	NUTS 1	2000-2007	EU
	Land rent		FADN	=SE375/SE025	EUR / ha	single farm, FADN-region	1989-2008	EU
			DeStatis		EUR / ha	NUTS 2	1989-2008	Germany
Natural resources								
	Share of grassland		FADN	=(K147AA+K150AA+K151AA)/SE025	%	single farm, FADN-region	1989-2008	EU
			CORINE		%	LAU 2	1990, 2000, 2010	EU
	Slope		USGSTOPO_30[d]		%	1 ha²	2000	EU
	Climate	LFA	FADN	=A39	1,2,3	single farm, FADN-region	1989-2008	EU
			Climate Data CRU-TS 1.2		Temperature	16 km²	1901-2000	EU
	Natura 2K %		EEA		%	scale 1:100 000	2010	EU
	Irrigated area		CORINE		%	LAU 2	1990, 2000, 2010	EU
Social and								

Group deter-minant	Indicator	Proxy	Data source	Variable definition (con-struction)	Unit	Geo-graphical resolution[e]	Years	Spatial coverage
demographical factors								
	Market distance	Population density	EUROSTAT	Table:demo_r_poar	Inhabitants per km²	NUTS 2	1990-2010	EU
			DeStatis			LAU2	1974-2010	Germany
		Population growth	EUROSTAT	Table:demo_r_poar	%	NUTS 2	1990-2010	EU
			DeStatis			LAU2	1974-2010	Germany
		Commuting time to larger cities	BBSR[c]		Min	LAU 2	2003	Germany
	Off-farm employment	Unemployment rate	EUROSTAT	Table: tgs00007	%	NUTS 2	1999-2007	EU
		Relation Off-farm to farm income	FSS		%	single farm, LAU 2	1999,2003, 2007	Germany
	Age		FADN	=C01YR	Year	single farm, FADN-region	1989-2008	EU

a) The variable definition/construction refers to the FADN data warehouse variables.
b) The variable definition/construction is derived from the EuroStat-Agriculture database http://epp.eurostat.ec.europa.eu/portal/page/portal/agriculture/data/database.
c) The Federal Institute for Research on Building, Urban Affairs and Spatial Development (Commuting times)
d) Earth Resources Observation and Science (Global 30-Arc Seconds elevation data)
e) The columns present the resolution at which representative results can be obtained.

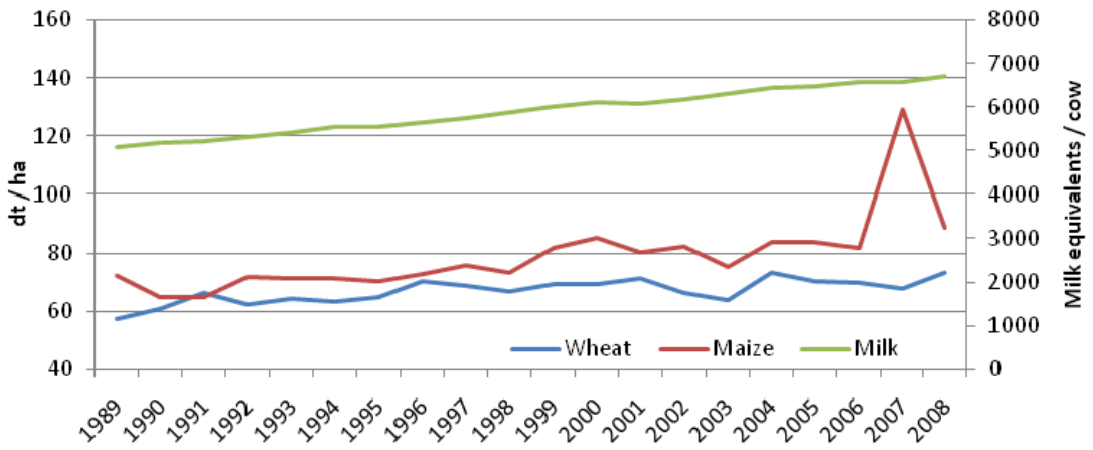
In the following section, we will investigate the characteristics of some explanatory variables that we consider crucial for a more detailed estimation of structural change.

1.2.4.1 Technology

Technological development over time can be approximated by variables that measure yield or by a trend variable. One obstacle in the definition of a yield variable is that it must measure the average yield variation over time for the entire farm type. This is straightforward for farm types such as the *Milk* class but problematic for more aggregated classes such as the *Mix* class (ID 4 and 6, respectively, as reflected in Table 4). Two possibilities are identified for a potential definition of a yield variable. One possibility is to define yield in terms of monetary output rather than physical quantities, which makes aggregation over different production activities feasible. The required data are available from the FADN 'standard results' for each of the different farm types (such as SE135 *Total output crops & crop production*, SE216 *Cows' milk & milk products*, SE251 *Total*

livestock output or SE131 *Total output*). The drawback of this approach, however, is that that a substantial part of the fluctuation can be attributed to price fluctuation, which adds substantial noise to a variable intended to measure technological development over time. Another possibility would be to consider the yield for selected production activities as representative of the productivity of the entire farm type class. For *COP crops* and *Other crops,* the yield of wheat (SE110) or the yield of maize (SE115) could be assumed to be representative of the productivity in that year. For farm type *Milk* and *Other Livestock*, milk yield (SE125) could be used. For the *Mix* farm type, one of the measures or a combination of both could be selected. Figure 4 shows the development of the average wheat, maize and milk yields. With the exception of a spike in the maize yield in 2007, which was likely caused by an unrealistically high outlier in that particular year, the average yield of each product closely follows a linear trend with little variation between years. This rather low variation is particularly problematic because the error of assuming that one particular crop represents the technological development in an entire farm type is supposedly rather large. Stated differently, the noise in the approximation appears to be substantial relative to the actual variation. Based on this discussion, we decided to consider a simple trend variable (defined as *1,2,..,T*) that may be more appropriate to capture technical improvements. The interpretation of this variable is also more intuitive because a yield variable that follows a linear trend would capture all other linear effects that evolve over time.

Figure 4: Development of the average yield of wheat (dt/ha; SE110), maize (dt/ha; SE115) and milk equivalents (kg) per cow (SE115) (averaged over all Germany FADN regions)

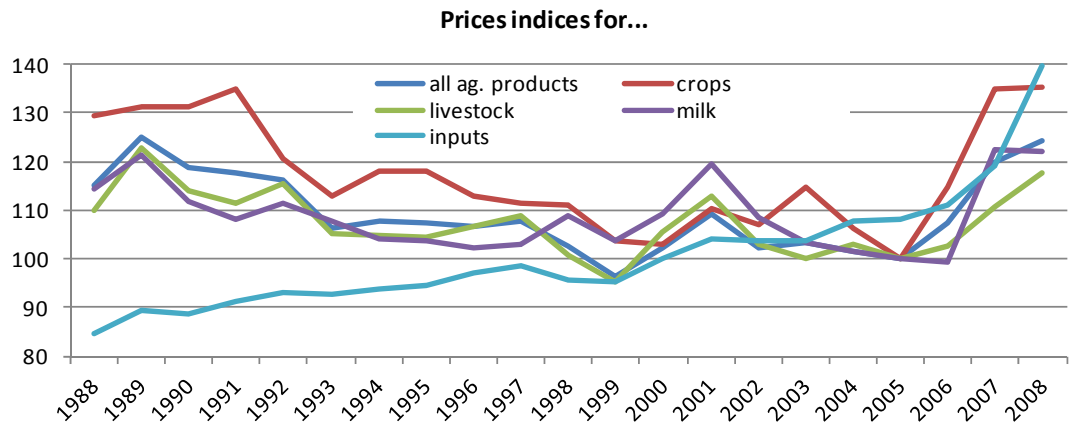


Source: Own calculation from the FADN database.

1.2.4.2 Prices, subsidies and land rent

Output and input prices are derived from DeStatis for the time period 1988 to 2008. The database provides different price indices for different categories, which can be used for the different farm types. An index for crops (*LANDWIRTPROD05*) is used for *COP crops* and *Other crops*, an index for milk (*MILCH01*) is used for *Milk*, and an index for livestock products (*TIERART001B*) is used for *Other livestock*. For the *Mix farm* type, an index for all agricultural products (*LANDWIRTPROD01*) is considered. All price indices are considered in the logarithm. A price index for agricultural input prices (*BETRIEBMILAND01*) is available for input prices. However, input prices are highly correlated with a trend variable (as defined above) (which can also be observed in Figure 5) such that input prices must be excluded from the analysis to avoid multicollinearity.

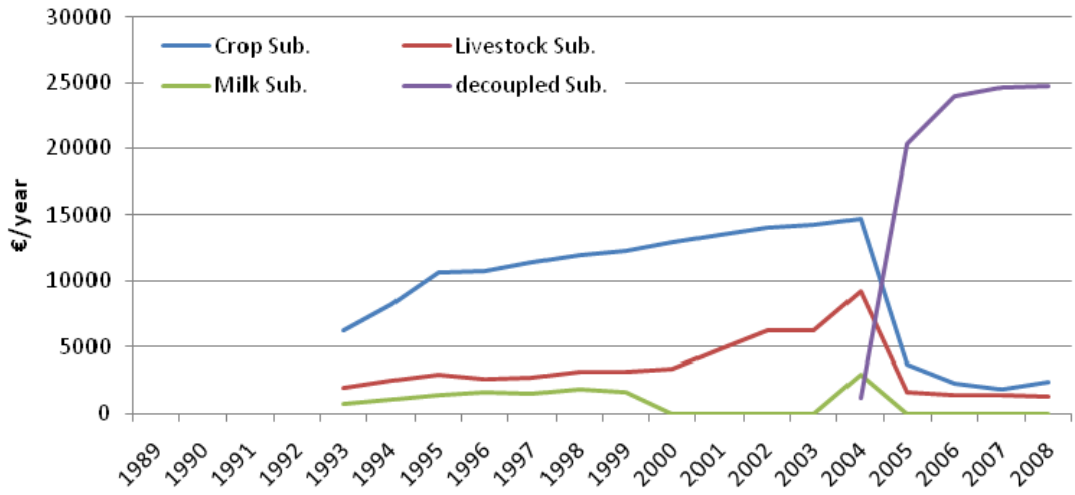
Figure 5: Output and input price indices for different farm products



Source: Destatis 2011.

First and second pillar payments can be derived from the standard results in FADN for the different farm types (see Figure 6). Three problems, however, limit the use of first and second pillar payments as explanatory variables. First, subsidies have only been reported since 1993, which would effectively reduce the time series by 5 years if a lag of one year is considered; this would result in a time series from 1995 to 1999, or 1995 to 2003, which is problematic from a computational point of view. A second problem is the break in the policy regime in 2003/2004, during which coupled payments were phased out and decoupled payments were introduced. Because the policy change occurred after the time period used for the estimation, it is not possible to consider this break in the estimation or prediction.

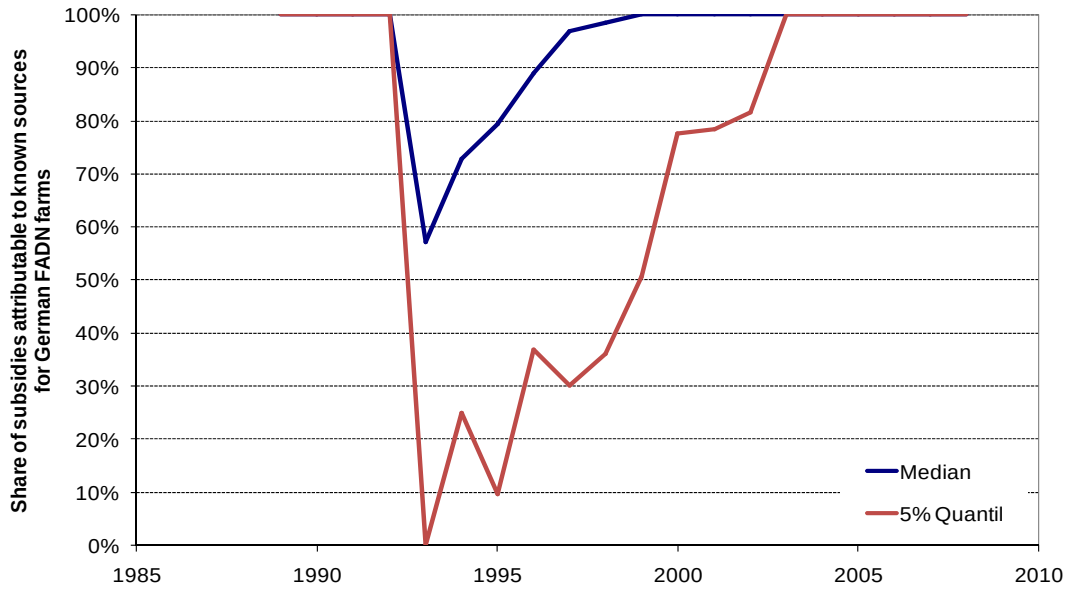
Figure 6: Average subsidies (€/year) received per farm (West German FADN regions)



Source: Own calculation from the FADN database.

Third, the quality of the recorded data changed substantially during the recorded period. Especially in the mid-1990s, a non-negligible share of subsidies can be attributed to neither the first nor the second pillar. Only from 2003 onward does the sum of subsidies under the different subheadings equal the stated total.

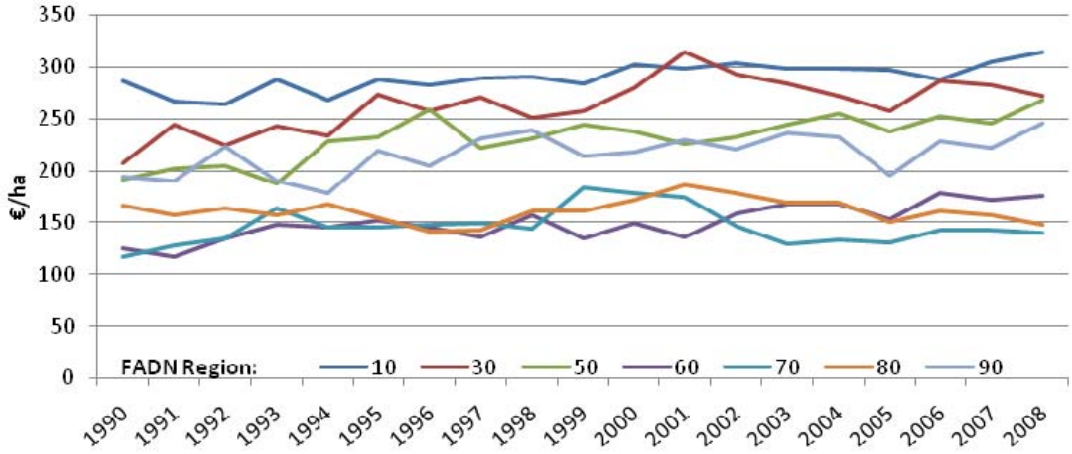
Figure 7: Share of subsidies attributable to known sources for German FADN farms



Source: Own calculation from the FADN database (Relation of the aggregate of SE610, SE615, SE621, SE622, SE623, SE625, SE626, SE630 to SE605).

The average land rent in a region is calculated from FADN as the rent paid (SE375) divided by the rented Utilised Agricultural Area (UAA) (SE030). Figure 8 depicts the development of the average land rent for the two selected regions, Lower Saxony and Bavaria, over time. In the estimation, the average land rent is considered on a log scale and lagged by one year.

Figure 8: Average land rent per ha (West German FADN regions)

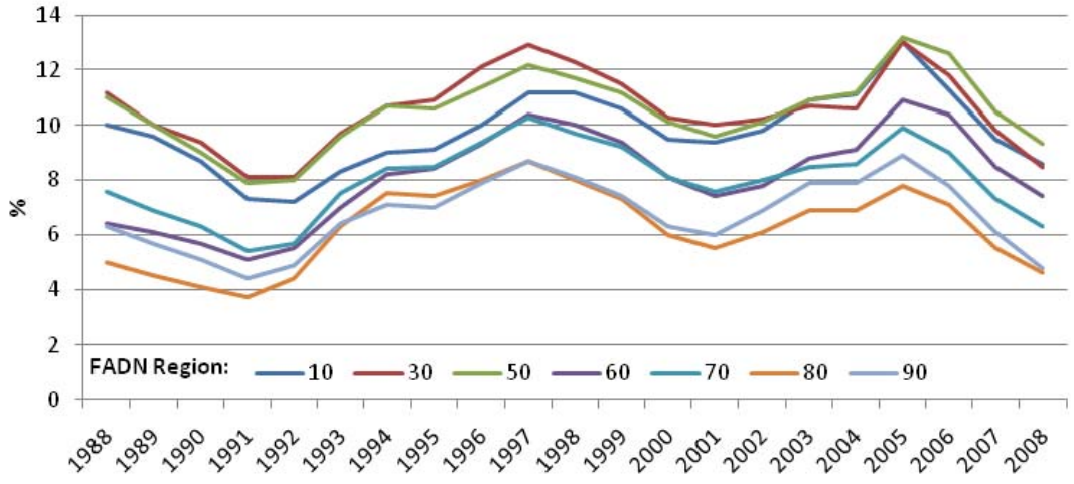


Source: Own calculation from the FADN database.

1.2.4.3 Socio-economic variables

The off-farm employment opportunities can be approximated by the unemployment rate (Figure 9). The unemployment rate (calculated for the civil dependent labour force) is available from Destatis for the German Laender for the time period 1991 to 2008 and from Statistisches Bundesamt (1993) for 1988 to 1991. Figure 9 clearly shows that the development of the unemployment rate is highly correlated among the FADN regions and the differences in the unemployment rate in 2008 can be largely attributed to a base effect (unemployment rate in 2008).

Figure 9: Unemployment rate of the civil dependent labour force



Source: Destatis 2011 and Statistisches Bundesamt (1993, p. 128).

Frequently, a change in the farm structure coincides with the handover of the farm business. Therefore, a change in the farmer's age could be a reasonable proxy for changes that are occurring in the farm structure or are going to occur in the near future, which makes a closer examination of the FADN-variable C01YR worthwhile (Table 7). For roughly 1,700 farms (7.7% of the total population), no birth year is stated. Most of these holdings are not family farms. Within the sample, more than 2,700 changes in the age of the farm manager are recorded. However, in nearly a quarter of these cases, the farm manager becomes older upon the transition. It seems unlikely (based on the theory and rationale) that the proportion of farms that is handed over from the younger to the older generation is this high, and therefore, we assume that this variable does not accurately reflect reality.

Table 7: Overview of some key figures regarding the age information in the German FADN sample (1989-2008)

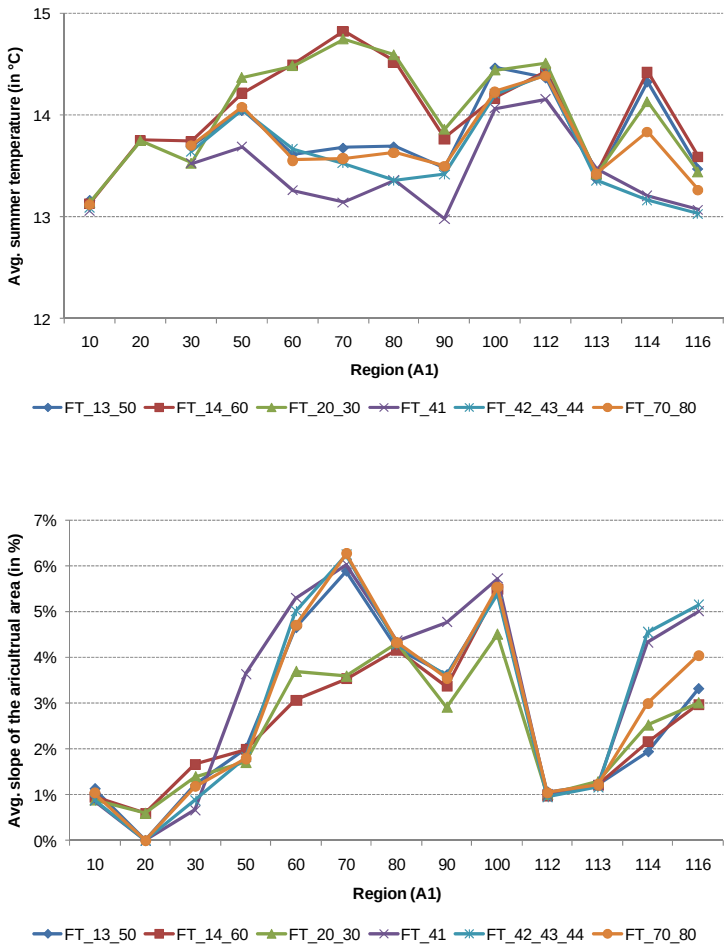
Farms	Observations
… in the sample	**21,778**
… for which no birth year of the farmer is stated	1,676
thereof family farms (A18 = 1)	523
… whose manager is older than 70 years or younger than 18 years	263
… observed at least twice	**17,444**
…for which the birth year of the farm manager changed	2,716
…for which the farmer becomes older when the birth year is changed	693

Source: Own calculation from the FADN database.

1.2.4.4 Natural conditions

The distribution and development of farm types is dependent on the natural conditions (e.g., temperature, precipitation, slope). For instance, in Bavaria (FADN region 90), dairy farming retreated in the last decade from areas suitable for arable farming (DICK and HETZ, 2011). Figure 10 shows the differences in the average conditions faced by the different farm types in the German FADN regions for two selected variables (average summer temperature and average slope). Marked differences exist among the farm types for these two factors, even within one region, and these factors determine plant growth and available field labour days. For instance, in the Southern and Western FADN regions (see Annex 2), dairy farms tend to be located in cooler and hillier regions when compared with cereal farms. It is reasonable to assume that the relative efficiency of the different farm types differs with regard to their sensitivity to these environmental factors. This may, *ceteris paribus*, imply that a dairy farm in an area with warmer and drier conditions is much more likely to expand cash cropping than a farm facing a cooler and more humid climate.

Figure 10: Average summer temperature and slope for the different farm types in the German FADN regions



Source: Own calculation based on the FADN database, BKG (2007) and DWD (2009).

1.2.5 Implementation of the data in the final estimation approaches

All time-varying explanatory variables are considered with a lag of one year because we expect that adjustments will occur with some delay. This implies that the first year needs to be omitted from the estimation if explanatory variables are not available prior to the first year considered.

The variable sets vary between the two estimation approaches (Markov and MCI), as the MCI (because of the higher number of observational units) is better suited to the depiction of the influence of time-invariant factors (e.g., climate[4], terrain, commuting time). As these factors can hardly be influenced by political decisions, their inclusion actually improves the trend prediction. The Markov approach focuses on the influence of time-variant variables and depicts time-invariant differences among regions using a regional dummy.

The selection (Markov and MCI approach) of the sets of explanatory variables is based on the Akaike Information Criterion (AIC). The AIC provides a trade-off between goodness of fit and the simplicity of the model. A general-to-specific selection process is adopted in which we start with the largest model, including all explanatory variables, and subsequently exclude variables as long as their exclusion results in a lower AIC. The specification with the lowest AIC is chosen as the final specification.

In all Markov models a constant and regional dummy variables are considered for all regions except one. The exclusion of the constant or a regional dummy variable is not considered in the model selection process, and thus the constant and all regional dummy variables are included in all models. In the MCI models, the constant and the lagged shares of the specialisation are excluded from the selection process.

Because of the described data problems regarding subsidies, the different public support regimes are depicted in the Markov approach by two different dummy variables. In particular, a dummy variable for the MacSharry reform (zero for $t \leq 1992$ and one afterwards) and a dummy variable for the Agenda 2000 (zero for $t \leq 1999$ and one afterwards) are considered. Despite these problems, the payments are implemented in the continuous approach as continuous variables.

[4] Climate is calculated as the average of the time series (see Table 9), and therefore, it is considered to be time-invariant.

Table 8 provides the final selection of all explanatory variables used in the Markov approach for the different farm types. The primary data source for the Markov out-of-sample prediction is the FADN database and the corresponding macro data grid available from EuroStat. Micro data can be derived directly from FADN because the farms, which are uniquely identified in the dataset, are observed over time. Macro data can be derived by considering the weights attached to each sample farm. These weighting factors are calculated based on the Farm Structure Survey (FSS), which is conducted every three or four years, and reflect the number of farms that a sample farm represents in the population. Using these weighting factors, it is therefore possible to recalculate the total number of farms in the population within classes of specialisation and size from the FADN sample. One obstacle is that, even though FADN data are available on a yearly basis, the weights are calculated based on the latest FSS information, which is only available every two to three years. This leads to breaks in the macro data such that farm numbers in the categories change only every two to three years. To address this factor, farm numbers are approximated between FSS years using a calculated geometric growth rate[5] (a linear approximation could have also been used, but a relative growth rate was assumed to be more appropriate).

Table 8: Time-varying explanatory variables for the different farm types selected in the Markov approach

Explanatory variables (availability)	Farm Types				
	COP crops	**Other crops**	**Milk**	**Other livestock**	**Mix**
Output prices (1988-2007)	Crop Price index	Crop Price index	Milk Price index	Livestock price index	Avg. Commodity price index
Yield	Trend				
Off-farm employment (1988-2007)	Unemployment rate				
Land rent (1991-2007)	Land rent				
Agenda 2000	Equal to zero before 2000, one otherwise				
MacSharry Reform	Equal to zero before 1993, one otherwise				

Table 9 lists the explanatory variables ultimately used in the MCI approach. In principle, the MCI approach uses the same data sources as the Markov approach. Some additional data sources are used to derive time-invariant information for the region. Whereas a price index is implemented in the Markov approach, relative price changes are used in the MCI models. As year-to-year changes in the structure of individual farms are fairly small and

[5] The geometric growth rate is calculated by $r = \sqrt[t_1 - t_0]{n_{t_1} / n_{t_0}}$, where n_t is the number of farms in t. Farm numbers are predicted between FSS years using $n_t = n_{t_0} \left(1 + r\right)^{t - t_0}$. For example, for the years between the FSS 1990 and 1993, the geometric growth rate is calculated by $r = \sqrt[1993-1990]{n_{1999} / n_{1990}}$, and farm numbers in 1991 are predicted by $n_{1991} = n_{1990} \left(1 + r\right)^{1991-1990}$.

largely random (Figure 23), we use a four-year period rather than a yearly interval in the MCI approach. An advantage of this interval is that the results can be directly compared with an analysis based mainly on FSS data.

Table 9: Variables used in the MCI approach

Type of variable	Variable	Comment	Estimation period (1989-1999)	Estimation period (1989-2003)
Price (1988-2008)				
		$ln(p_{t,t}) - ln(p_{t,t-1})$		
	Soft wheat	EuroStat: Agr-r-aacts	X	X
	Durum wheat	ditto	X	X
	Rape seed	ditto	X	X
	Flowers	ditto		X
	Sugar beet	ditto	X	X
	Grass	ditto		X
	Beef	ditto		X
Farm (1989-2008)				
	Farm size	log of the farms SGM		X
	Interest	Relationship of paid interest to SGM	X	X
	Age	Year		X
	Tenure	€ per ha		X
	Share of rented land	%		X
	Stocking density	LU per ha	X	
	Share of grassland on UAA	%		X
	Farm diversification	Shannon index based on 8 and 14 different specialisations	X	X
	Total subsidies	Relationship of total subsidies to SGM	X	X
	Subsidies 1st pillar	Relationship of subsidies to SGM	X	
	Subsidies 2nd pillar	Relationship of subsidies to SGM		X
	Subsidies agri-environment	Relationship of subsidies to SGM		X
Region				
	Summer temperature (Avg. of 1961-1990)	DWD (2009)		X
	Winter temperature (Avg. of 1961-1990)	ditto	X	X
	Annual precipitation (Avg. of 1961-1990)	ditto		X
	Winter precipitation (Avg. of 1961-1990)	ditto	X	X
	Water balance (Avg. of 1961-1990)	ditto		X
	Commuting time (2003)	BBSR	X	X
	Avg. Population density	EuroStat:	X	X

(1997-2008)	demo_r_d3dens		
Change in population density (1997-2008)	ditto	X	
Avg. Employment rate (1999-2008)	EuroStat: tgs00007	X	
Change in employment rate (1999-2008)	ditto		X
Share of grassland	BKG (2007)		X
Altitude	BKG (2007)	X	
Slope	BKG (2007)		X
Natura 2K %	EEA		

1.3 Models and Prediction

1.3.1 Markov approach

The estimation of stationary or non-stationary TP is based on a Bayesian framework developed by STORM et al. (2011). The framework allows micro and macro data to be combined for the estimation of stationary or non-stationary Markov TPs. Thus, the FADN micro data from the annually surveyed farms can be combined with the available macro data by considering the weighting factors of each FADN sample farm. Using these weighting factors, it is therefore possible to recalculate the total number of farms in the population and within classes of specialisation and size from the FADN sample. This combination of micro and macro data builds on ZIMMERMANN and HECKELEI (2012), who considered the same combination of data in a Generalised Cross Entropy approach to estimate and explain Markov TPs across European regions for the dairy sector and the entire agricultural sector, respectively.

1.3.1.1 Estimation of TP

The estimation framework is based on a Markov approach that models the movement of individuals among a finite number of predefined states, $i = 1,...,k$, as a stochastic process. The k states are mutually exclusive and exhaustive. The Markov process is characterised by a $(k \times k)$ 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

(2): $$\mathbf{n}_t = \mathbf{P}_t' \mathbf{n}_{t-1}.$$

In a non-stationary Markov process, the TPs change over time depending on exogenous variables. The specification of the TPs differs depending on the type of Markov states considered. If we assume that the Markov states do not have an order, the multinomial logit is an appropriate specification. In this case, the TPs are specified according to

(3): $$P_{ijt} = \frac{e^{\mathbf{z}_{t-1}\beta_{ij}}}{1 + \sum_{f=1}^{k-1} e^{\mathbf{z}_{t-1}\beta_{if}}} \quad \forall i, j,$$

where the vector $\boldsymbol{\beta}_{ij}$ denotes the unknown coefficients and $\boldsymbol{\beta}_{ik} = 0 \quad \forall i$ by definition. The vector $\mathbf{z}_{t-1}$ represents (lagged) exogenous explanatory variables including a constant. If,

however, the states have an order, the ordered logit model is a superior specification. Here, the TPs are specified according to

$$(4) \qquad P_{ijt} = \frac{e^{c_{ij} - \mathbf{z}'_{t-1}\boldsymbol{\beta}_i}}{1 + e^{c_{ij} - \mathbf{z}'_{t-1}\boldsymbol{\beta}_i}} - \frac{e^{c_{i,j-1} - \mathbf{z}'_{t-1}\boldsymbol{\beta}_i}}{1 + e^{c_{i,j-1} - \mathbf{z}'_{t-1}\boldsymbol{\beta}_i}} \qquad \forall\ j = 1, \ldots, k,$$

where the cut points c_{ij} for $j = 2, \ldots, k-1$ are estimated along with $\boldsymbol{\beta}_i$ and $c_{i,0} \equiv -\infty$, $c_{i,1} \equiv 0$, $c_{ik} \equiv \infty$. In cases where the states are ordered, the order logit model provides the advantage that the assumption of "Independence of Irrelevant Alternatives" (IIA) does not apply. Additionally, from a computational point of view, the ordered logit specification is preferable because only $k\,n_z + (k-2)\,n_z$ parameters need to be estimated, compared to $k(k-1)n_z$ parameters for the multinomial logit model, where n_z denotes the number of explanatory variables, including a constant. For the multinomial logit model, the stationary model arises as a special case if only a constant is considered as explanatory variable, that is, $z_{t-1} = 1\ \forall t$ such that $\mathbf{P}_t = \mathbf{P}\ \forall t$.

The Bayesian framework employs a data likelihood function that represents the available macro data. The available micro data are used to specify a prior density. The specification of the likelihood function is based on a likelihood concept derived by MACRAE (1977:187). The concept is appropriate for a first-order non-stationary Markov process for macro data observed for the entire group over time. MACRAE (1977) shows that, in this case, the group proportions are distributed as a weighted sum of independent multinomial random variables with a probability equal to the corresponding row in $\mathbf{P}_t$ and weights equal to the group proportions in $t-1$. The resulting likelihood function is given by

$$(5): \qquad L\left(\boldsymbol{\beta} \middle| \mathbf{n}_1, \ldots, \mathbf{n}_T\right) = \prod_{t=1}^{T} \sum_{\mathbf{H}_t \in \mathrm{H}_t} \prod_{i=1}^{k} \left(n_{i,t-1}!\right) \left(\prod_{j=1}^{k} P_{ijt}^{\ \eta_{ijt}} / \eta_{ijt}! \right),$$

where n_{it} are elements of the data vector $\mathbf{n}_t$, η_{ijt} denotes the (unobserved) number of individuals transiting from state i at time $t-1$ to state j at time t and $\mathbf{H}_t$ is a matrix with elements η_{ijt}. The summation in the likelihood expression (5) is over the set H_t of all combinations of matrices $\mathbf{H}_t$ for which the rows sum to $\mathbf{n}_{t-1}$ and columns sum to $\mathbf{n}_t$, that is

$$(6): \qquad \mathrm{H}_t = \left\{ \mathbf{H}_t \middle| \sum_h \eta_{iht} = n_{i,t-1},\ \ \sum_h \eta_{hjt} = n_{jt} \quad \forall i, j \right\}.$$

Because the set H_t rapidly increases in size with an increasing number of farms, the implementation of expression (5) for larger samples is challenging (or impossible) from a computational point of view. To overcome this problem, a large sample approximation proposed by HAWKES (1969) and BROWN and PAYNE (1986) that avoids the computation of the set H_t is employed. Denoting $\mathbf{n}_t^*$ and $\mathbf{P}_t^*$ as $\mathbf{n}_t$ and $\mathbf{P}_t$ without the last row and

column, respectively, these authors argue that, in large samples, it can be assumed that $\mathbf{n}_t^*$ is independent $(k-1)$-variate normal with mean vector $\mathbf{P}_t'^*\mathbf{n}_{t-1}$ and covariance matrix

(6): $$\operatorname{cov}\left(\mathbf{n}_t^*\right) = diag\left(\mathbf{P}_t^{*\prime}\mathbf{n}_{t-1}\right) - \mathbf{P}_t^{*\prime} diag\left(\mathbf{n}_{t-1}\right)\mathbf{P}_t^* = \mathbf{\Gamma}_t,$$

where $diag(\cdot)$ is a squared matrix with the argument vector as the main diagonal and zero off-diagonals. The large sample log-likelihood, L_{la}, can then be written as

(7): $$L_{la}\left(\boldsymbol{\beta}\middle|\mathbf{n}_1,\ldots,\mathbf{n}_T\right) = \sum_{t=1}^{T} -0.5\left(\log\left|\mathbf{\Gamma}_t\right| + \left(\mathbf{n}_t^* - \mathbf{P}_t^{*\prime}\mathbf{n}_{t-1}\right)'\left(\mathbf{\Gamma}_t\right)^{-1}\left(\mathbf{n}_t^* - \mathbf{P}_t^{*\prime}\mathbf{n}_{t-1}\right)\right).$$

To specify an appropriate prior density $p(\boldsymbol{\beta})$, the underlying sampling distribution of the micro observations needs to be considered. Let n_{it} be the number of individuals in state i at t, let $\mathbf{X}_t^i$ be the share of individuals in t that were in state i in $t-1$ and let $\mathbf{P}_{it}$ be the i-th row of $\mathbf{P}_t$. Each individual in the micro sample transits between states in accordance with the TP matrix $\mathbf{P}_t$. Therefore, as argued by MACRAE (1977:187) and equivalently to the case of aggregated data, each vector $\mathbf{X}_t^i$ has a multinomial distribution about mean $\mathbf{P}_t^i$ with size $\mathbf{n}_{it}$. The observed number of individuals $\mathbf{n}_t$ is then the weighted sum of vectors $\mathbf{X}_t^i$. Hence, the prior density can be represented as a likelihood similar to (5), except that now information regarding the individual transitions $\mathbf{N}_t$ is available, which makes the summation over the set N_t unnecessary. The likelihood used to specify the prior density thus simplifies to

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

where $\mathbf{P}_t$ is specified according to (3).

For Bayesian inference, it is necessary to sample from the posterior density $h\left(\beta\middle|\mathbf{d}\right)$, which is proportional to the product of likelihood $L\left(\boldsymbol{\beta}\middle|\mathbf{n}_1,\ldots,\mathbf{n}_T\right)$ and prior density $p(\boldsymbol{\beta})$. The sample is obtained via the Markov Chain Monte Carlo (MCMC) method, i.e., by using a random walk Metropolis Hasting (MH) algorithm. The MH sampler allows a (pseudo-) random sample to be produced from almost any target distribution known up to a normalising constant (cf. CHIB and GREENBERG, 1995). Using the obtained posterior sample, the posterior mean, which is the optimal Bayesian estimator under squared error loss, can then be derived directly by calculating the mean of the sample.

1.3.1.2 Prediction strategy

Prediction of farm numbers using the estimated TPs, $\mathbf{P}_t$ is straightforward and follows directly from the Markov process given in (2). Prediction of a four-year time period, for example, is achieved by

(9) $$\hat{\mathbf{n}}_{t+3} = \hat{\mathbf{P}}'_{t+3}\hat{\mathbf{P}}'_{t+2}\hat{\mathbf{P}}'_{t+1}\hat{\mathbf{P}}'_{t}\mathbf{n}_{t-1}.$$

As stated in the previous section, the aim is to predict the number of farms in 15 (3 sizes x 5 farm types) classes. When adding an entry/exit class, each farm is uniquely defined in each time period as one of the 16 (3 sizes x 5 farm types + Entry/Exit) classes. Defined in this way, the 16 classes are mutually exclusive and exhaustive and thus fulfil the assumption of the Markov process. To predict farm number in the 15 (3 sizes x 5 farm types) classes, it would, in principle, be possible to estimate a full (16x16) (3 sizes x 5 farm types + entry/exit) TP matrix that is used for prediction. However, from a computational point of view and in light of the sample size, it seems to be more appropriate to divide the estimation and prediction problem into several smaller problems. Therefore, a separate (4x4) TP matrix is estimated for each of the five farm types that represents the three size classes and an entry/exit class (referred to as "size TP matrices" in the following). In this case, the entry/exit class represents farms entering/exiting farming as well as farms entering/exiting from/to another farm type. These TP matrices for each farm type are then used to predict farm numbers in each of the size classes such that a prediction for all five farm types is finally obtained in three size classes.

Splitting the estimation problem into several small estimations provides important advantages for the prediction of farm numbers. Most important is that the number of parameters to be estimated is reduced substantially. This reduction is the result of two effects. First, an ordered logit specification can be used for the size TP matrices, if we are willing to assume that the classes are ordered from entry/exit, size class I, II and III. For the full (16x16) TP matrix, this assumption is not suitable, and the multinomial model, which requires more parameters to be estimated, seems to be the only legitimate choice. More importantly, in the case of the multinomial logit model, the number of parameters increases exponentially so that dividing the problem substantially reduces the number of parameters. Both effects together imply that, for the stationary model (i.e., only a constant as explanatory variable) for the full (16x16) TP matrix using the multinomial logit specification $k(k-1)n_z = 16 \times (16-1) \times 1 = 240$, parameters compared to $5 \times (k\, n_z + k\,(k-2)) = 5 \times (4 \times 1 + 4\,(4-2)) = 60$ need to be estimated for all five small TP matrices using the ordered logit model. When considering additional explanatory variables in the non-stationary approach, the gap widens further. In that case, different estimations for each farm type allow the different explanatory variables for each farm type to be considered (i.e., rather than a price index for all agricultural products, a price index for the corresponding farm type can be considered). To predict farm number in the non-stationary

approach, the actually observed explanatory variables are considered for the prediction period.

1.3.1.4 Handling of entry and exit

One problem commonly encountered in structural change analysis using the FADN database is the handling of farm entry/exit to/from the sector. The problem is that FADN provides no information on whether a farm that is no longer observed has been omitted from the sample or has ceased to farm. In addition, no information is provided as to whether a farm observed in FADN for the first time just started farming or if it already existed but is newly added in the sample. Therefore, neither macro nor micro data are directly available for the entry/exit classes. Two different strategies are applied for the macro and micro data to consider an entry/exit class. For macro data, the number of farms in the entry/exit class in each year and region is calculated as the difference between an assumed maximum number of farms and the observed number of farms in all classes considered in that year and farm type. The maximum number of farms is set to be equal to the highest number of farms observed over time in that region. No distinction is made between farm types to determine the maximum number of farms, i.e., the same maximum number of farms is considered for each of the five size TP matrices (because the absolute number of farms in the E/E class is irrelevant, a different maximum number of farms for each farm type could just as well be assumed). For example, the number of farms in the entry/exit class for the *COP crop* farm type in North Rhine-Westphalia in 2000 is equal to 29,232, calculated as the difference between the number of *COP crop* farms (in that year, region and any of the three size classes; 4,097) and the highest total number of farms observed for that region in any time period (33,320 in the year 1989).

For micro data, no information about entry/exit is available, and a non-informative prior is considered. The implementation of an ignorance prior for specific categories is implemented straightforwardly in the Bayesian framework derived above. If we define the entry/exit class to be the first Markov state $(k=1)$, ignorance of the entry/exit class results in a prior specification given by

(10) $$p(\boldsymbol{\beta}) = L(\boldsymbol{\beta}|\mathbf{N}_1,\ldots,\mathbf{N}_T) = \prod_{t=1}^{T}\prod_{i=2}^{k}\left(n_{i,t-1}!\right)\left(\prod_{j=2}^{k}\mathbf{P}_{ijt}^{\,n_{ijt}} / n_{ijt}!\right)$$

[note the differences in the indices i and j compared with (8)].

1.3.2 Continuous approaches for the description / explanation of structural change in agriculture

1.3.2.1 Introduction

The theoretical framework of the estimation of market shares in marketing research, comprehensively examined by, among others, COOPER and NAKANISHI (1988), is linked to the case of farm specialisation shares. This framework is used to estimate the influence of different variables (e.g., socio-economic, geographical and climate) as well as how and to what extent these variables affect farm specialisation. Therefore, a few models of different complexity are presented, and their utility for describing and explaining farm specialisation is shown.

1.3.2.2 Market share attraction and MCI models to analyse structural change

Market share models are widely used in marketing research to estimate market shares of brands or products and to investigate the effects of marketing instruments on own or competitor's market shares. For a comprehensive examination, see COOPER and NAKANISHI (1988) or FOK et al. (2002). The connection between market share and marketing activities can be described by two theorems. The fundamental theorem of KOTLER (1984) states that the market share of a brand or product is proportional to the marketing effort applied by the firm. BELL et al. (1975) postulate that attractiveness determines the market share. Consumers are attracted to different brands, and the most attractive brand has the greatest market share.

In the case of shares of different farm specialisations, the different production activities chosen by the decision maker (farmer or farm holder) determine the farm specialisation shares. Analogously to the market share case in which brands compete for shares of a limited market, the different activities compete for their share of the farmer's resources.

COOPER and NAKANISHI (1988) distinguish different specifications of market share attraction models in simple effects, differential effects and fully extended models. Models with simple effects estimate the impact of one or more explanatory variables on all the market shares of all brands. This means, for instance, that the price has the same effect on every brand's market share. Differential models allow the estimation of different effects of each explicative variable on every brand's market share, which implies that the price of one brand has a different effect on the brand's market share when compared with the effect of the price of another brand on its market share. In fully extended models, own and cross effects of each explanatory variable are estimated for each brand's market share,

which suggests that the own and the competitor's price have different effects on the brand's market share. Furthermore, the market share can be a linear, multiplicative or exponential function in the marketing mix variables on the right side of the equation (COOPER and NAKANISHI, 1988: 27).

The concept of market share models shall be applied for ex-post estimation of farm structural change. Therefore, shares of different farm specialisations (see 0) and socioeconomic variables are used rather than a brand's market share and its marketing instruments. The three model types introduced above would yield the following results in the case of farm specialisation. The simple effects model estimates one single effect for each explanatory variable for all different farm specialisations (e.g., age has the same effect on each type of specialisation). In the differential effects model, the influence of a given explanatory variable can differ among the farm specialisations (e.g., the price of cereals may have a different impact for cash cropping and granivore production). The fully extended model permits an analysis of the effects of the explanatory variables observed in other farms (e.g., age) on the farm specialisation shares of the investigated farm.

For the estimation of a brand's market share, the respective market share is recorded for each brand at every point in time and point of sale (respective submarket). For the estimation of farm specialisation, we record the respective shares of the different farm specialisations for each observed period and farm. Hence, a brand's market share and submarkets are replaced by farm specialisation and farms. In both cases, exogenous explanatory time-variant variables such as the unemployment rate are used. Furthermore, the different points of observation (submarkets, farms) can differ by static (e.g., slope) or time-variant explanatory variables (e.g., socio-demographic structure). In contrast to the market share case, we do not have a single variable that can be univocally attributed to a given farm specialisation. In the market share case, the marketing expenses of a given brand are such that a variable that can be referred univocally to a given firm. However, the attribution of a particular product price to a farm specialisation would result in pitfalls, as some prices could be attributed to several farm specialisations, e.g., the prices of cereals to the different cash crop specialisations (output) as well as granivore production (input). In addition, in the market share case, the prices can differ among different points of sale or submarkets, whereas we assume equal commodity prices for all farms. In contrast to the general market share case, the data in our application are characterised by many farms recorded for only a fairly limited time period, as opposed to a limited number of submarkets monitored over a long period. Because of these differences, some conclusions that can normally be drawn in the brand market share case cannot be made in the farm specialisation case. These differences will be highlighted in the next section.

1.3.2.3 Empirical implementation

We do not explore the simple effects model because it is very unlikely that the effect of a given explanatory variable is independent of the analysed farm specialisations. Therefore, we investigate the impact when a given explanatory variable has different (potential) effects on each farm specialisation (differential effect model). In the farm specialisation share case, "attraction" is proportional to the utility of a given farm specialisation. For example, a farmer may have the following production mix: maize, wheat and granivores. The regression estimation may imply a positive effect on cash cropping and a negative effect on pig fattening because of an increase in cereal price. This means that the "attractiveness" of cash cropping increases whereas the attractiveness of pig fattening decreases. Consequently, the farm specialisation share of cash cropping increases whereas that of pig fattening decreases. According to COOPER and NAKANISHI (1988: 128p), the differential effects model can be formulated generally as follows:

(11) $$s_i = \frac{A_i}{\sum_{j=1}^{m} A_j}$$

(12) $$A_i = e^{(\alpha_i + \varepsilon_i)} \prod_{k=1}^{K} f_k(X_{ki})^{\beta_{ki}},$$

where in our case:

A_i attraction of farm specialisation i

s_i farm specialisation share of specialisation i

m number of farm specialisations

X_{ki} value of the k-th explanatory variable explaining attraction of farm specialisation i

K number of explanatory variables

β_{ki} coefficient of the influence of the k-th explanatory variable on attraction of farm specialisation i

α_i intercept for specialisation i

f_{ki} positive, monotone transformation of X_{ki}

ε_i specification-error term.

The farm specialisation share can easily be derived from Equation (11) after the attraction of a given specialisation has been estimated. The share of a farm specialisation can decrease even when its attraction increases. This is the case when the attraction of a farm specialisation increases in absolute terms but decreases relative to other farm specialisations.

According to COOPER and NAKANISHI (1988: 28), market (specialisation) share models must comply with the two following basic conditions:

- Estimated market shares from the model are nonnegative.
- The sum of estimated market shares equals one (if the shares for all brands / specialisations are estimated).

Only two broad families of models fulfil these criteria: "Multiplicative Interaction" (MCI) and Multinomial Logit Models. For this study, we opted for a MCI model approach as it is more frequently adopted in the literature. This approach implies that a farm's utility is a multiplicative function of the explanatory variables. The utility and specialisation share are proportional, which means that if the utility of a specialisation increases, the specialisation share will also increase. The variables must be logarithmised to apply common linear estimation techniques.

The dummy regression formulation model of differential effects can be described as follows (COOPER and NAKANISHI, 1988: 129p):

(13) $$\ln s_{tt} = \alpha_1 + \sum_{j=2}^{m} \alpha_j' d_j + \sum_{u=2}^{T} \gamma_u D_u + \sum_{k=1}^{K} \sum_{j=1}^{m} \beta_{kt} d_j \ln X_{ktt} + \varepsilon_{tt}$$

In this formula, there are dummies for farm specialisation ($d_j = 1$, if $j = i$ and 0 otherwise) and farm ($D_u = 1$, if $u = t$ and 0 otherwise). Subscript t denotes the farm. All of the other variables are defined as in Equation (12). Recall that, in the farm specialisation case, the dimension submarket is replaced by farm. In contrast to the market share case, in which normally a limited number of points of sale are observed over a longer period, our dataset consists of many farms that are each recorded only for a limited time interval. This peculiarity of the dataset demands the estimation of a large number of farm dummies (γ_u) ($T > 7600$). COOPER and NAKANISHI (1988: 117p) emphasise keeping this dimension as a dummy variable to properly estimate the parameters. However, we refrained from the calculation of farm dummies, as the number of observations per farm is

very limited and the dummy would therefore, in the majority of cases, capture nearly all of the explanatory value of a given observation. Therefore, we drop the term $\sum_{u=2}^{T} \gamma_u D_u$ from Equation 13.

We do not investigate cross effects because we would have to extend the dataset by $(T-1) \times T \times k \times m$ variables, which would make the model impossible to estimate, and we do not believe that measurable and reliable cross effects of the explicative variables of other farms could be detected from the one in consideration in the given data set. Two potential sources of cross effects among farms are likely: the land market and the commodity market. The sample consists only of a limited number of farms (<2% of the population), for which we have only limited information regarding the distance between the farms in consideration. Therefore, any conclusion regarding cross effects arising from interactions in the land market within a given set of farms would be very ambiguous. Regarding market interactions, there is no obvious reason why Farm A should have a different influence on Farm C than Farm B, particularly if Farms A and B are nearly identical.

We assume that the farmer is going to decide on the magnitude of activities based on his experience in the previous years rather than on the current situation because the yields or prices are unknown when the farmer seeds or increases his stock. Therefore, we lagged all of the explanatory variables by one year.

To preserve the information of the full data set, zero and negative values (before taking the log) must be avoided. Therefore, we added a small constant before the transformation (e.g., 0.1) to all specialisation shares.[6] COOPER and NAKANISHI (1988: 153pp) state that deleting observations with zero values of the dependent variable (specialisation shares) or adding the constant only to zero-value observations would lead to biased estimations.

1.3.2.4 Model specification

The data set is extended by lagged farm specialisation shares as additional explanatory variables. In our case, we use four-year-lagged farm specialisation shares. In this first example, the model to be estimated consists of lagged farm specialisation shares as the only explanatory variable and can be formulated as follows:

[6] An extensive and sensitive analysis (constant between 0.001 and 3) concluded that the results of the estimation are fairly independent of the constant for the range 0.1 to 1.5.

(14) $\ln s_{tt} = \alpha_1 + \sum_{j=2}^{m} \alpha_j' d_j + \sum_{k=1}^{m} \sum_{j=1}^{m} \beta_{kt} d_j \ln s_{ktt} + \varepsilon_{tt}$

The definitions of the variables are the same as in (12) and (13). The four-year-lagged farm specialisation share variable ($\ln s_{ktt}$) occurs m-times in each equation. This model can be called the "autoregressive model". In addition to the average effect of each farm specialisation, the farm specialisation shares given four years ago determine the ex-ante simulation of the amounts of each farm specialisation share for every farm. Ultimately, $m \times m + m$ parameters must be estimated.

In the second model, the model of Equation (13) is expanded by additional explanatory variables:

(15) $$\ln s_{tt} = \alpha_1 + \sum_{j=2}^{m} \alpha_j' d_j + \sum_{k=m+1}^{K} \sum_{j=1}^{m} \beta_{kt} d_j \ln X_{ktt} + \varepsilon_{tt}$$

Each equation consists of lagged individual (e.g., age and stock density) and regional/aggregated (e.g., population density) variables. Compared with Equation (14), the farm specialisation shares for every farm are simulated by more variables than only the lagged farm specialisation shares. In conclusion, $(K - m) \times m + m$ parameters are going to be estimated.

In a last step, the two previous models are combined so that the farm specialisation share is the result of the four-year lagged specialisation share and the lagged explanatory variable:

(16) $$\ln s_{tt} = \alpha_1 + \sum_{j=2}^{m} \alpha_j' d_j + \sum_{k=1}^{K} \sum_{j=1}^{m} \beta_{kt} d_j (\ln X_{ktt} + \ln s_{ktt}) + \varepsilon_{tt}$$

In this model, $K \times m + m$ parameters must be estimated.

1.4 Results

The following section presents some key findings of the Markov and MCI approaches. As the goal of the study is to develop a robust approach to predict farm numbers rather than to explain changes observed in the past, we base both approaches for the evaluation of the different model specifications on the quality of the out-of-sample prediction. Therefore, we split the available sample into two parts: the first is the training set with which we estimated the model parameters. We applied these parameters to the second set, the test set, and compared the predicted values with the recorded results based on the 2007 FSS. We analysed structural change occurring within a four- and an eight-year interval, respectively. In the first case, all FADN data prior to 2003 were used in the test set. The obtained coefficients were applied to the data from 2003, and the results were compared with the reported structure for 2007. In the second case, we used only the data prior to 1999 in the test set and extrapolated the farm structure in 1999 to 2007. This estimation was only conducted for Western Germany, as FADN data for Eastern Germany are only available from 1995 onward.

In the non-stationary Markov model and the MCI models, the set of selected variables is determined within a stepwise regression framework. We applied a backward elimination procedure and removed variables from a model as long as a smaller variable set led to an increase in the Akaike Information Criterion (AIC) (see Section 1.2.5).

The ex-post analysis was conducted for nearly all German regions (see section 1.1) and the data were pooled over all regions rather than regionally differentiated specifications. In the case of the Markov approach, this means that we estimated a dummy variable fixed effects panel model. Pooling the data implies that a given explanatory variable has the same effect across all regions, which would lead to lower explanatory power in the test phase. However, the lower flexibility of the pooled model (i.e., more degrees of freedom) allows a more robust estimation. This is crucial, as the number of observations per combination of region, type of farming and farm size class is very limited.

This method also limits the possibility of a direct comparison of single region estimations and a panel approach. Estimations of single regions are performed for two regions and one farm type. The obtained results indicate a superior out-of-sample prediction quality, in terms of the MSE and pDev, of the panel approach (see Table 10). These results, however, need to be considered with care because it is difficult to obtain robust estimation results for non-stationary TPs for single regions. Pooling of the data and the use of the panel approach are therefore primarily considered because they provide more robust estimation results.

Table 10: Out-of-sample prediction fit measures for a fixed-effects panel estimation and separate estimation per regions

Region	Farm Type	Separate estimations per region MSE (in 1000)	Separate estimations per region pDev (in%)	Fixed-effects Panel model MSE (in 1000)	Fixed-effects Panel model pDev (in %)
Lower Saxony	COP crops	1,292	54	132	16
Bavaria	COP crops	472	40	239	34
	Mean	882	47	186	25

Source: Own calculation based on German FADN and FSS data
Note: Prediction period 2003-2007 (Estimation period 1989-2003)

Because of the limited amount of time, data and resources, the approaches differ in their focus in analysing structural change. Whereas the Markov approaches focus on farm growth, i.e., change in farm size, the MCI approach focuses on the analysis of the change in the productive orientation, i.e., the type of farming.

1.4.1 Markov

1.4.1.1 Stationary Markov

For each time period, region and farm type, the development of farm numbers is predicted in any of the three size classes using an estimated stationary Markov TP, a linear trend prediction, a geometric trend prediction and a constant prediction (which assumes that farm numbers remain constant). Therefore, for the constant estimation, farm numbers for 2007 are assumed to be the same as in 2003 and 1999 for the training sets 1989-2003 and 1989-1999, respectively. A discussion of the individual results of all of these cases is neither possible nor desirable, and it appears that any method can perform better or worse in a particular case. However, because the method that is most appropriate for a specific situation is not usually known in advance, we are interested in identifying the method that, 'on average', delivers the most 'accurate' prediction. As described in Section 0, we defined 'accuracy' in terms of the MSE and the mean percentage deviation (pDev), which are an absolute and a relative measure of the prediction quality, respectively. Table 11 shows each fit measure averaged over all size classes and considered regions to obtain the fit measures for the five farm types. To obtain one aggregate measure of the prediction quality, the average of the fit measures over all farm types is also reported.

Table 11: Mean square error (MSE) and mean percentage deviation (pDev) of the out-of-sample prediction for different prediction methods (averaged over all size classes)

Farm Type	Stat. Markov		Linear		Constant		Geometric	
	MSE (in 1,000)	**pDev** (in %)	**MSE (in 1,000)**	**pDev** (in %)	**MSE (in 1,000)**	**pDev** (in %)	**MSE (in 1,000)**	**pDev** (in %)
Prediction period (Test Set) 2003-2007 (Estimation period (Training Set) 1989-2003) Estimation performed for German FADN region: 10, 30, 50, 60, 70, 80, 90, 112, 113, 114, 116								
COP crop	34	17	31	20	26	17	393	48
Other Crop	46	39	223	49	16	23	1,045	68
Milk	167	24	2,291	86	494	21	7,447	65
Other Livestock[a]	28	23	88	57	29	31	959	615
Mix	134	31	893	76	106	31	21,609	84
Overall Mean	82	27	705	58	134	25	6,291	176
Prediction period (Test Set) 1999-2007 [Estimation period (Training Set) 1989-1999] Estimation performed for German FADN region: 10, 30, 50, 60, 70, 80, 90								
COP crop	333	36	404	42	455	47	2,332,534	3,191
Other Crop	1,903	124	1,974	107	1,189	107	123,649	514
Milk	4,137	84	5,437	141	9,944	97	60,761	444
Other Livestock	136	66	40	59	94	59	282	79
Mix	2,386	88	2,372	68	2,954	101	1,006,658	391
Overall Mean	1,779	80	2,046	84	2,927	82	704,777	924

Source: Own calculation based on the German FADN sample

a) Region 70 is excluded from the calculation of *pDev*. There is only one farm in the large size class, and the predicted value is relatively low, but compared with the small absolute number, the relative error is extremely large for all estimation methods.

For the shorter prediction period, the prediction based on the stationary Markov TP is better than the linear and geometric predictions in terms of the average MSE and the pDev. The stationary prediction is comparable to the constant prediction and is better in terms of the MSE (82 versus 134) and slightly worse in terms of the pDev (27% versus 25%). For the longer prediction period from 1999-2007, the relative performance of the stationary prediction increases slightly. For this time period, the stationary prediction is slightly better than the constant prediction in terms of both the MSE and the pDev. Furthermore, it is better than a linear prediction, although the difference in the performance of both methods decreases in comparison to the shorter prediction period. For both time periods, the geometric prediction has, on average, the worst performance of all methods considered. The geometric prediction method seems to easily exaggerate small trends observed in the data to result in very large errors. This is particularly problematic for the long prediction (1999 to 2007) period, for which small errors in the assumed exponential growth can result in rather extreme changes.

1.4.1.2 Non-stationary Markov

Similar to the stationary case, estimations for the non-stationary case are performed for the two different prediction periods from 2003 to 2007 and 1999 to 2007, respectively. In the final model specification (Table 12), the unemployment rate, the land rent and the trend variable, as well as the dummy variable for the MacSharry Reform, are selected for all farm types for the prediction period from 2003 to 2007. Output price variables are selected in the model for *Other crops* and *Other livestock*. The dummy variable for the Agenda 2000 is selected for the model for *COP crops*, *Other crops* and *Milk*. For the prediction period from 1999 to 2007, only the unemployment rate is selected for all farm types. The land rent is included for all but the *Milk* farm type, and the dummy variable for the MacSharry Reform is included for all but the *Other crop* farm type. The trend variable and the output price variables are selected for the model for *COP crop*, *Milk* and *Other livestock*.

Table 12: Final model specifications for the non-stationary Markov approach (in addition to the listed variables, regional dummy variables were considered for all except one region in all models)

Prediction period	Farm Type	Explanatory variables						
2003-2007	COP crops	Const	Unemp		Land rent	Trend	DAGENDA	DSHARRY
	Other crops	Const	Unemp	Price Crop	Land rent	Trend	DAGENDA	DSHARRY
	Milk	Const	Unemp		Land rent	Trend	DAGENDA	DSHARRY
	Other livest.	Const	Unemp	Price Live.	Land rent	Trend		DSHARRY
	Mix	Const	Unemp		Land rent	Trend		DSHARRY
1999-2007	COP crops	Const	Unemp	Price Crop	Land rent	Trend	---	DSHARRY
	Other crops	Const	Unemp		Land rent		---	
	Milk	Const	Unemp	Price Milk		Trend	---	DSHARRY
	Other livest.	Const	Unemp	Price Live.	Land rent	Trend	---	DSHARRY
	Mix	Const	Unemp		Land rent		---	DSHARRY

Source: Own calculation based on the German FADN sample

Table 13 (ignoring the last column for the moment) shows the resulting fit of the sample prediction measured in terms of the AIC for the final specification of the non-stationary model in comparison to the stationary model. Because the non-stationary model permits greater flexibility, as expected, a better sample prediction fit (i.e., a lower AIC) is obtained for the non-stationary approach in comparison to the stationary model in all farm types.

Table 13: Comparison of the Akaike Information Criterion (AIC) for different estimation methods

Prediction period	Farm Type	Stationary	Non-stationary	Non-stationary (without trend)
2003-2007	COP crops	10.50	9.86	9.89
	Other crops	11.70	10.09	10.33
	Milk	13.50	11.22	11.47
	Other livest.	10.40	8.53	9.15
	Mix	12.70	10.39	12.52
1999-2007	COP crops	11.10	10.19	10.59
	Other crops	10.80	10.12	10.12
	Milk	13.00	9.90	11.24
	Other livest.	8.80	7.21	8.20
	Mix	12.00	9.97	9.97

Source: Own calculation based on the German FADN sample

Table 14: Mean square error (MSE) and mean percentage deviation (pDev) of the out-of-sample prediction for differentt prediction methods (averaged over all size classes)

Estimation performed for German FADN region: 10, 30, 50, 60, 70, 80, 90												
Prediction period (Test Set) 2003-2007 (Estimation period (Training Set) 1989-2003)												
	Non-stationary Markov		Stationary Markov		Linear		Constant		Geometric.		Non-stationary (without trend)	
Farm Type	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)
COP crop	235	49	50	16	38	15	36	16	560	51	36	14
Other Crop	289	41	72	36	349	50	25	20	1,642	76	57	26
Milk	721	19	261	21	3,598	97	776	23	11,701	76	262	15
Other Livestock[a]	174	80	43	19	138	49	45	29	1,505	146	54	32
Mix	396	30	209	27	1,399	65	165	24	33,955	100	179	19
Overall Mean	363	44	127	24	1,104	55	209	22	9,873	90	117	21
Prediction period (Test Set) 1999-2007 (Estimation period (Training Set) 1989-1999)												
COP crop	9,304	332	333	36	404	42	455	47	2,332,534	3,191	731	54
Other Crop	1,799	121	1,903	124	1,974	107	1,189	107	123,649	514	1,799	121
Milk	15,979	134	4,137	84	5,437	141	9,944	97	60,761	444	7,196	70
Other Livestock	156	96	136	66	40	59	94	59	282	79	141	65
Mix	1,674	75	2,386	88	2,372	68	2,954	101	1,006,658	391	1,674	75
Overall Mean	5,782	151	1,779	80	2,046	84	2,927	82	704,777	924	2,308	77

Source: Own calculation based on the German FADN sample

a) Region 70 is excluded from the calculation of *pDev*. There is only one farm in the large size class, and the predicted value is relatively low, but compared with the small absolute number, the relative error is extremely large for all estimation methods.

Table 14 (ignoring the last two columns for the moment) shows a summary of the fit measures of the out-of-sample prediction for the non-stationary approach in comparison to the stationary approach, a linear prediction, a geometric prediction and a constant

prediction for the two time periods. In terms of percentage deviation averaged over all farm types, the non-stationary approach performs worse (44%) than the stationary approach (24%) and the constant prediction (22%) but better than the linear prediction (55%) and the geometric prediction (90%) for the prediction period from 2003 to 2007. Regarding the MSE averaged over all farm types, the order of the above mentioned specifications changes such that the non-stationary approach performs better than the linear and geometric prediction but worse than the stationary Markov prediction and the constant prediction. For the longer prediction period from 1999 to 2007, the linear prediction, the constant prediction and the stationary Markov prediction perform better than the non-stationary prediction with regard to the MSE and the pDev. Only the geometric prediction is worse than the non-stationary prediction.

For both time periods, the stationary prediction performs better than the non-stationary prediction. This is true despite the better model fit (i.e., lower AIC), discussed above (Table 13), for the non-stationary model compared with the stationary model. This finding indicates that the selected explanatory variables explain the patterns of structural change in the estimation period but the effects of some/all explanatory variables are different (in terms of sign and/or magnitude) in the prediction period. To further explore the differences in the effects of some of the explanatory variables, an alternative model specification can be used that is the same as that used above, except that the trend variable is excluded for all models. As expected, the exclusion of the trend variable increases the AIC for those models in which the trend was previously selected but still results in a lower AIC (i.e., better model fit) than the stationary model (Table 13, last column). Despite this worsening of the model fit, however, exclusion of the trend variable substantially increases the quality of the out-of-sample prediction (Table 14, last two columns) of the non-stationary approach in all cases. Without the trend variables, the non-stationary approach becomes the best approach for the prediction period 2003 to 2007 with regard to the overall MSE as well as the overall pDev. For the prediction period 1999 to 2007, the non-stationary approach without the trend variable is also the best approach with regard to the pDev and is only slightly worse than the linear prediction and the stationary Markov prediction in terms of the MSE.

In addition to the prediction of farm numbers in the different size classes, it is also relevant to consider the prediction quality of the different approaches for the total number of farms (Table 15). Overall, for the prediction period of 2003 to 2007, the stationary, non-stationary and constant prediction approaches perform similarly well. With regard to the pDev, the constant prediction is better than the stationary and non-stationary predictions. With regard to the MSE, the stationary prediction is best, followed by the non-stationary prediction without the trend variable and the constant prediction. The linear prediction and the non-stationary approach with the trend variable perform worse than the other three methods but are still substantially better than the geometric prediction with respect to the pDev and MSE. For the longer prediction period of 1999 to 2007, the linear prediction now has the best performance for both measures, followed by the

stationary prediction, the constant prediction and the non-stationary prediction without the trend variable. The non-stationary prediction is again only better than the geometric prediction.

Table 15: Mean square error (MSE) and mean percentage deviation (pDev) of the out-of-sample prediction for the total number of farms for different prediction methods

Estimation performed for German FADN region: 10, 30, 50, 60, 70, 80, 90												
Prediction period (Test Set) 2003-2007 (Estimation period (Training Set) 1989-2003)												
	Non-stationary Markov		Stationary Markov		Linear		Constant		Geometric.		Non-stationary (without trend)	
Farm Type	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)	MSE (in 1,000)	pDev (in %)
COP crop	467	22	251	8	162	8	71	6	2,377	31	158	10
Other Crop	2,100	36	273	23	262	19	82	14	3,960	66	366	21
Milk	2,511	11	1,041	16	3,266	26	2,914	14	46,878	69	1,709	8
Other Livestock	410	20	29	7	341	18	78	9	5,606	97	219	20
Mix	493	12	522	14	659	19	645	14	109,291	105	662	10
Overall Mean	1,196	20	423	14	938	18	758	11	33,622	73	623	14
Prediction period (Test Set) 1999-2007 (Estimation period (Training Set) 1989-1999)												
COP crop	18,744	124	1,475	27	1,060	27	2,226	40	6,952,589	1489	3,899	51
Other Crop	13,365	116	14,434	111	11,922	110	8,707	95	433,958	512	13,365	116
Milk	35,385	73	12,134	67	11,614	32	32,601	75	336,203	361	27,665	63
Other Livestock	1,012	40	49	23	87	38	54	29	906	64	172	46
Mix	4,162	39	6,417	44	3,214	29	8,780	54	3,213,045	415	4,162	39
Overall Mean	14,534	78	6,902	54	5,579	47	10,474	59	2,187,340	568	9,853	63

Source: Own calculation based on the German FADN sample

Analysis of the effects of explanatory variables

In addition to predicting farm numbers, which was the primary aim of this section, it is also interesting to analyse the effects of the explanatory variables on the development of the farm number.

Then, the value of the explanatory variable of interest is changed (e.g., the unemployment rate is increased by 10% or a dummy variable is switched from 0 to 1), and a new transition matrix is calculated while keeping the values of all other variables fixed. Finally, the relative difference from the reference situation between the two transition matrices is calculated, which allows us to analyse how a *ceteris paribus* increase of one variable changes the transition of farms between different classes.

Given the large number of combinations that arise for the six explanatory variables, five farm types, four states (three size classes+entry/exit) and two prediction periods, it is not possible to discuss every potential combination. Rather, we attempt to identify common effects across all farm types and prediction periods for each explanatory variable. For all cases, the results of the non-stationary approach excluding the trend variables, which delivered the best prediction results, are considered.

The unemployment rate was selected in all models, and with only a few exceptions, an increase in the unemployment rate increased the number of farms that enter from the entry/exit class while reducing the number of farms that exit. Furthermore, an increase in the unemployment rate leads to fewer farms moving to smaller size classes while increasing the number of farms that remain in the current size class or increase in size. With respect to the land rent, in almost all cases, an increase in the land rent reduces the number of farms that enter. The effects on the small and medium size classes are also consistent across almost all farm types and prediction periods. Here, an increase in the land rent increases the number of farms that move to larger size classes while reducing the number of farms that decrease in size or exit. The *COP crop* farm type is the only exception; for this farm type, the effects contradict the observations in the other cases such that an increase in the land rent increases farm entries as well as farm exits from the small class. Output prices were only selected in a few cases, and different effects are observed in each case, which makes it difficult to identify general effects across farm types. In addition, for each farm type, the corresponding price is considered (e.g., the milk price for the *Milk* farm type), which complicates the comparison of the effect across farm types. The effects of the dummy variable for the MacSharry reform are mixed, and no general pattern arises from the different farm types and prediction periods. The dummy variable for the Agenda 2000 was selected in the *COP crop*, the *Other crop* and the *Milk* farm types. Here, a shift in the dummy variable from zero to one leads to more small and medium farms increasing in size and fewer farms decreasing in size or exiting the farm type. Similarly, a shift in the dummy variable from zero to one reduces the number of large farms that decrease in size or exit and hence increases the number of large farms that remain in their current class. Further, it leads to reduced entry into the *Other Crop* and *Milk* farm types but increased entry into the *COP Crop* farm type.

1.4.2 MCI models

The goal of the MCI model application is to correctly predict the share of the different farm types for a given year. The development of farm numbers was not analysed because the MCI approach is based on observations on the farm level and exits from the sector are not recorded in FADN. In contrast to the Markov approach, we focus on the change in the type of farming.

First, we evaluate the quality of the in-sample prediction of an MCI approach including only the lagged farm specialisation share and the expanded model with explanatory variables. The set of explanatory variables was selected by a backward elimination algorithm based on the AIC. The selected indicators are listed in Table 9.

Table 16: Comparison of the in-sample prediction quality between a MCI approach and the benchmarka between 1989 and 2003 over a four-year period

	Constant	**MCI (only lagged)**	**MCI (explanatory variables)**
Observations		25,401	
Correct prediction of the farm type	84.7%	86.7%	86.7%
RMSE regarding the individual specialisation shares	0.0775	0.0331	0.0323
Total distance from the observed farm specialisation (halved Manhattan Block Distance) (quantile)			
10%	0.003	0.008	0.009
25%	0.026	0.026	0.026
50%	0.063	0.051	0.050
75%	0.126	0.083	0.081
90%	0.250	0.118	0.113

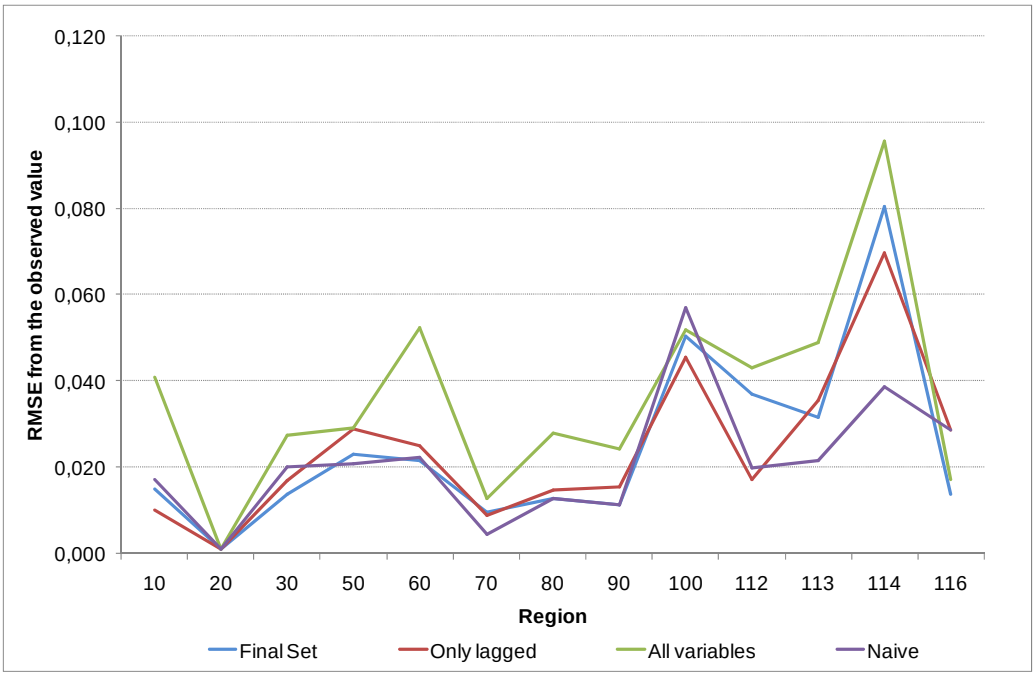
Source: Own calculation based on the German FADN sample
a) The benchmark assumption is that the shares of the different types of farming remain constant over the four-year period.

Table 16 shows some statistics for the period 1989-2003 when the farm type is projected over a period of four years. As a benchmark, we assume that the farms maintain their specialisation shares and consequently their farm type for four years. Examination of the specialisation shares reveals that the MCI without any explanatory variables approach performs better than the benchmark In particular, the likelihood of larger errors is reduced (e.g., smaller values for the 90% quantile). This reduction comes at the expense of the introduction of small errors for farms that do not alter their specialisation shares. Second, the potential of the MCI approach to correctly predict the type of farming is analysed. Therefore, we aggregate the different specialisation shares obtained from the MCI approach according to the FADN rules and Table 5. The MCI approach correctly predicts the farm type for more than 86.7% of the farms, which is a slightly higher rate than the benchmark. Although the MCI approach yields better results than a naive approach, the additional benefits of introducing explanatory variables are limited.

1.4.2.1 Performance of different specifications of MCI models

Figure 11 compares the root mean squared error (RMSE) with respect to the shares of the different types of farming for the German FADN regions in an out-of sample prediction. The coefficients are estimated for the period of 1989-2003 (training set) and applied to the farms present in the FADN sample in 2003 (test set) to predict the composition of the types of farming in 2007. The reduced set of variables (final set) clearly outperforms the full set of variables because of the overspecialisation of the full set. If the temporal dependency is only considered for the different specialisation shares and no further explanatory variables are included in the model (only lagged specialisation shares, Equation 14), the result is nearly as good as that for the final set. However, the best fit is obtained when it is assumed that the shares of the different types of farming remain constant over time (naive approach). The naive approach outperforms the other approaches, particularly in areas where the number of farms in the sample is small, e.g., Eastern Germany (regions 112-116). Because of the limited number of observations per region, the attribution of a "wrong" type of farming for even a single farm will have a large impact on the model goodness of fit.

Figure 11: Root mean squared error (RMSE) between the predicted and the observed shares of the different types of farming across the region for different specifications of the MCI model (training set: 1989-2003; test set 2003-2007)

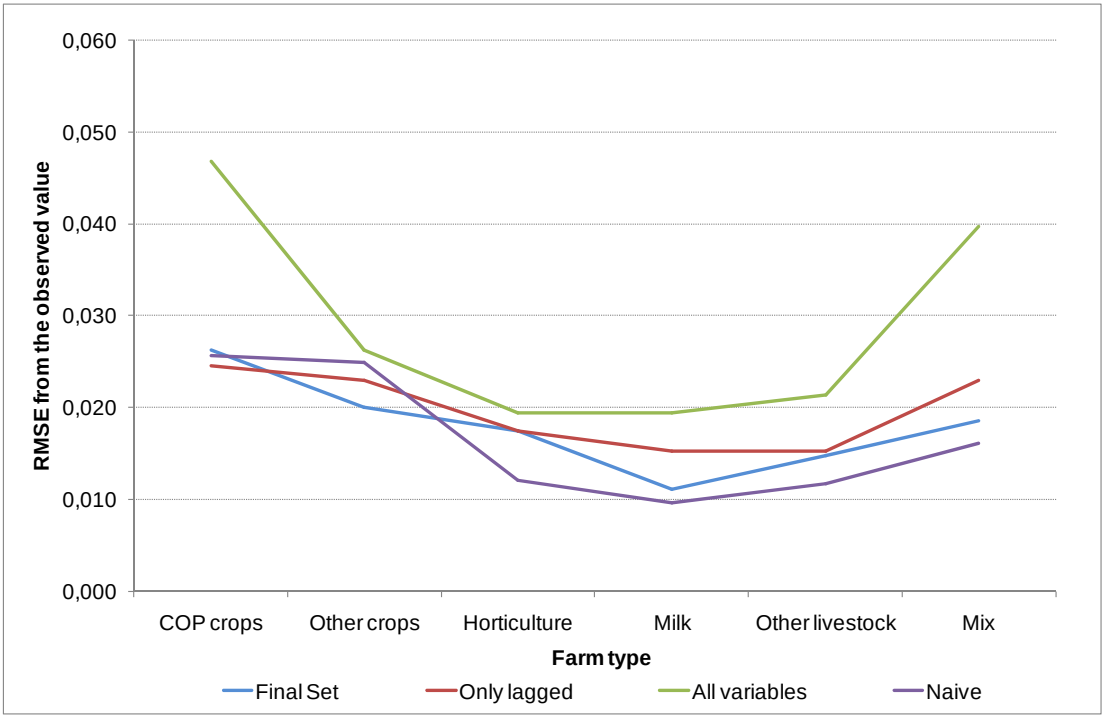


Source: Own calculation based on the German FADN sample

The *COP crops* farm type has the highest RMSE in the out-of-sample prediction (Figure 12). The naive approach outperforms the other model, which includes a reduced set of

explanatory variables (final set) in all cases except for the cash cropping farm types (*COP crops* and *Other crops*). In particular, if all variables are included for the out-of-sample prediction, the results for the *COP crops* and *Mixed* farm types are worse than those in the other models.

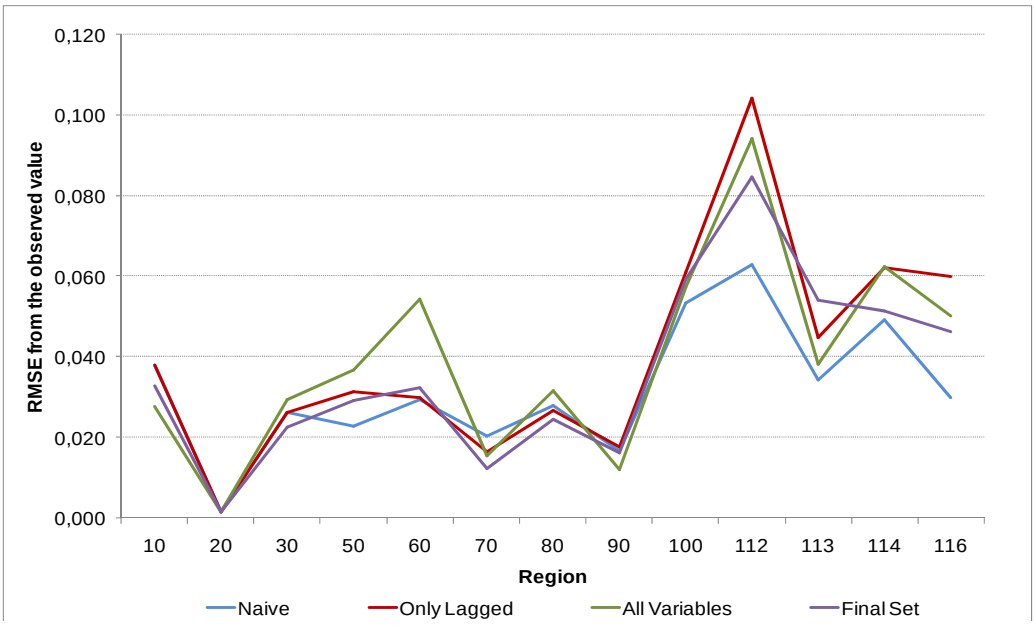
Figure 12: Root mean squared error (RMSE) between the predicted and observed shares of the different types of farming across the types of farming for different specifications of the MCI model (training set: 1989-2003; test set 2003-2007)



Source: Own calculation based on the German FADN sample

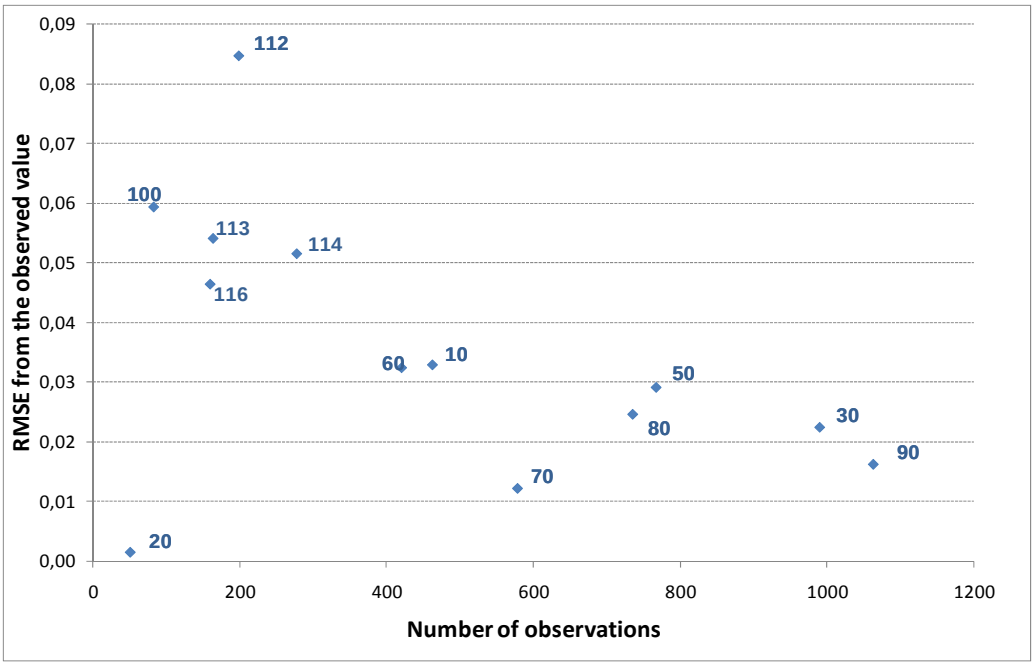
An examination of the 8-year prediction period (1999-2007) reveals that the naive approach slightly outperforms the final set of the MCI approach (Figure 13 & Figure 15), primarily because the shares of the different farm types are falsely predicted in FADN regions with few observations, e.g., the East German FADN regions (FADN code 112-116) (Figure 14). Improved performance is only achieved if the farming structure in a region is strongly dominated by a single type of farming (e.g., FADN Region 20). The specification including all variables does not perform consistently worse than the other specifications, which is in contrast to the short prediction period and to our prior expectation that the problem of overspecialisation should be positively correlated with the prediction period.

Figure 13: Root mean squared error (RMSE) between the predicted and the observed shares of the different types of farming across the region for different specifications of the MCI model (training set: 1989-1999; test set 1999-2007)



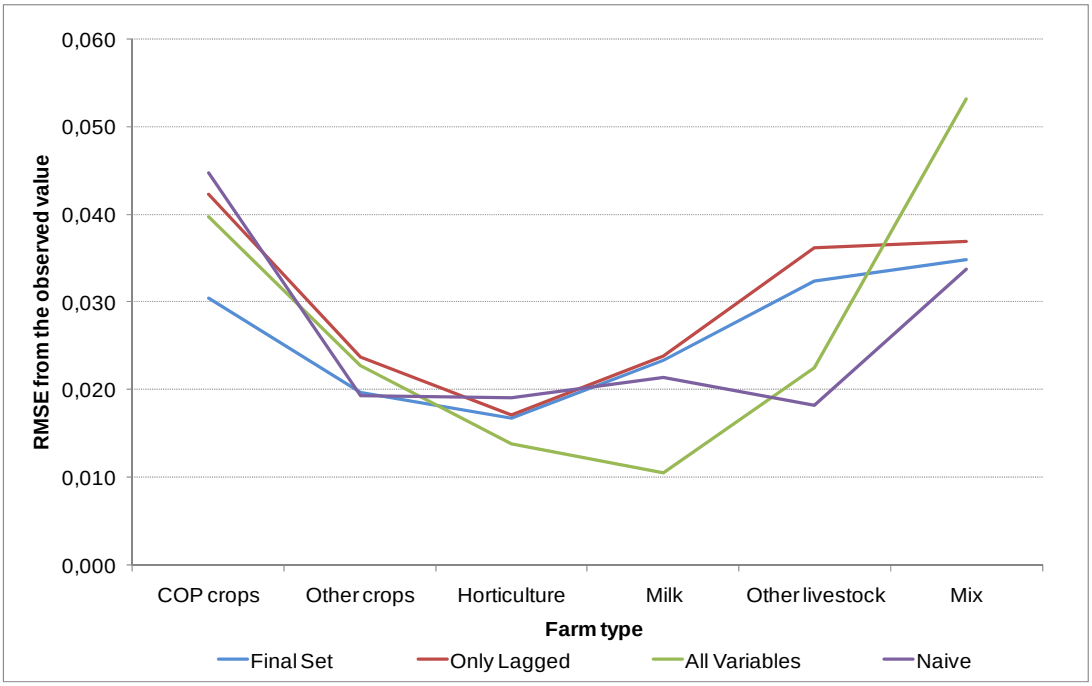
Source: Own calculation based on the German FADN sample

Figure 14: Dependence of the root mean squared error (RMSE) on the number of FADN farms per region for the final set of the MCI model (training set: 1989-1999; test set 1999-2007)



Source: Own calculation based on the German FADN sample

Figure 15: Root mean squared error (RMSE) between the predicted and the observed shares of the different types of farming across the types of farming for the "final set" specification of the MCI model and a "naive" estimation (training set: 1989-1999; test set 1999-2007)



Source: Own calculation based on the German FADN sample

Analysis of the effects of explanatory variables

Table 17 and Table 18 present the results of a sensitivity analysis performed for the market and policy variables selected with the training set 1989-2003 and applied to the test set 2003-2007. These tables are based on the final set of variables, i.e., the removal of any of the stated variables resulted in an increase in the AIC. These tables depict the impact of a 10% increase or decrease of the respective variable on the share of a given farm type. Two main conclusions can be drawn based on these tables. First, for most farm types, the response to an increase or decrease in the respective variable is rather symmetrical, i.e., if an increase in the variable leads to more farms belonging to that particular farm type, a decrease in the value of the variable leads to a smaller number of farms. Second, the impact of a change in the commodity prices seems to be more severe than the change in the share of subsidies distributed by the various instruments of the CAP. Some results can be explained quite intuitively, e.g., a higher beef price leads to more *Other livestock* farms (without dairy cattle) and less dairy farms. However, other results are difficult to explain, e.g., the result observed for durum wheat. On the one hand, the marked increase in the AIC upon the removal of the variable from the MCI models estimating the specialisation shares at the farm level indicates a high relevance; on the other hand, not a single farm changes its farm type in response to a change in the price. We suspect that the effect of prices such as that for durum wheat is not causal but rather

that their respective development are correlated with unknown factors influencing farm structural change.

Table 17: Impact of a 10% increase in the respective variable on the number of farms per type compared with the reference scenario for the period of 2003-2007 (test set) based on the period 1989-2003 (training set)

Variable	COP crops	Other crops	Horti-culture	Milk	Other livestock	Mix
Policy (share of)						
Agri-environmental subsidies[a]	-2.5%	1.9%	0.1%	-	0.6%	0.1%
1st pillar subsidies[a]	-	-	-	-	-	-
Total subsidies[a]	2.7%	-1.1%	-	-	-0.2%	-1.0%
Prices						
Beef	12.8%	0.2%	0.3%	-2.2%	13.4%	-9.0%
Durum wheat	-	-	-	-	-	-
Soft wheat	-	-	-	-	-	-
Rape seed	-1.7%	1.5%	-	-	-0.3%	0.2%
Sugar beet	12.9%	0.7%	0.3%	-1.8%	11.1%	-9.6%
Grass	-3.2%	3.2%	-0.5%	0.8%	-4.3%	0.2%
Flowers	-5.1%	2.4%	-2.2%	1.1%	-4.3%	2.5%
Farm economics						
Tenure	-	-	-	-	-	-
Share of interest payments[a]	-0.3%	-	-	-	-	0.2%

Source: Own calculation based on the German FADN sample
a) share of the respective variable in the farm's total SGM
- No influence: The net effect of changing this variable on the number of farms in this farm type is zero (number of farms leaving and entering the farm type is identical) although within farms, the share of specialisation might change.

Table 18: Impact of a 10% decrease in the respective variable on the number of farms per type compared with the reference scenario for the period of 2003-2007 (test set) based on the period of 1989-2003 (training set)

	COP crops	Other crops	Horti-culture	Milk	Other livestock	Mix
Policy (share of)						
Agri-environmental subsidies[a]	2.3%	-2.0%	-	-	-	-0.2%
1st pillar subsidies[a]	-	-	-	-	-	-
Total subsidies[a]	-3.0%	2.3%	-	0.1%	0.1%	0.3%
Prices						
Beef	-8.5%	2.9%	-3.7%	1.4%	-7.6%	5.7%
Durum wheat	-	-	-	-	-	-
Soft wheat	1.5%	-1.9%	0.4%	2.2%	-10.1%	-0.6%
Rape seed	1.6%	-0.5%	-	-	0.7%	-1.0%
Sugar beet	-8.8%	2.5%	-3.5%	1.3%	-6.2%	5.9%
Grass	2.5%	-1.7%	0.1%	-0.9%	6.0%	-0.9%
Flowers	3.6%	1.9%	0.1%	-1.4%	7.1%	-3.6%
Farm economics						
Tenure	-	-	-	-	-	-
Share of interest payments[a]	-	0.3%	-	-	-	-0.2%

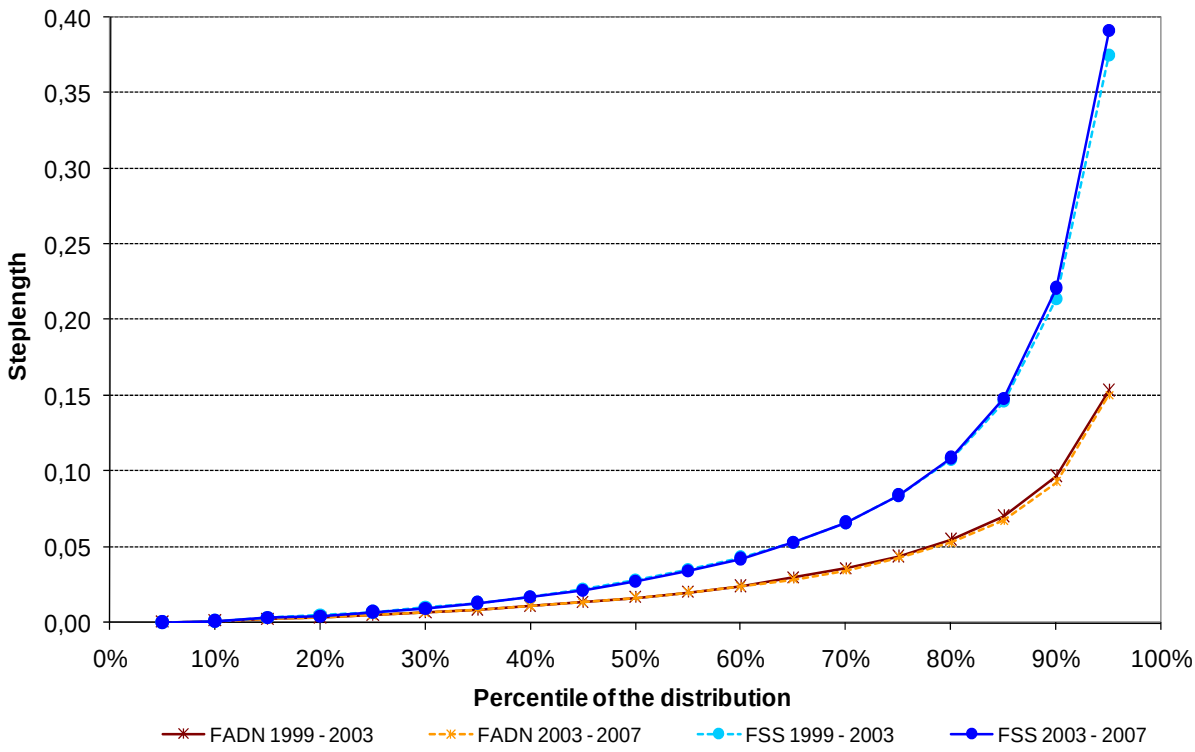
Source: Own calculation based on the German FADN sample
a) share of the respective variable in the farm's total SGM

In some cases, the number of farms increases irrespective of whether a lower or a higher value is chosen for the variable, e.g., the number of farms belonging to *other crops* increases independently of the direction in which the beef price is altered. If the transitions are analysed in greater detail, it is apparent that a higher beef price leads to more *mixed* farms converting to *other crops*, whereas in the case of a lower beef price, excess *COP crops* farms become *other crop* farms.

1.4.3 Comparison of the FADN sample with the FSS population

Although the MCI approach could reasonably predict the development of a given farm, the overall quality of the MCI and the Markov approaches for predicting structural change is, in general, mediocre. Even on a medium time horizon of four to eight years, naïve approaches that simply project the current composition regarding farm size classes or farm types perform nearly quite as good or better. There are several reasons for the mediocre predictive quality.

Figure 16: Distribution of the Step length aggregated over four years for the periods of 1999-2003 and 2003-2007 for FSS and FADN

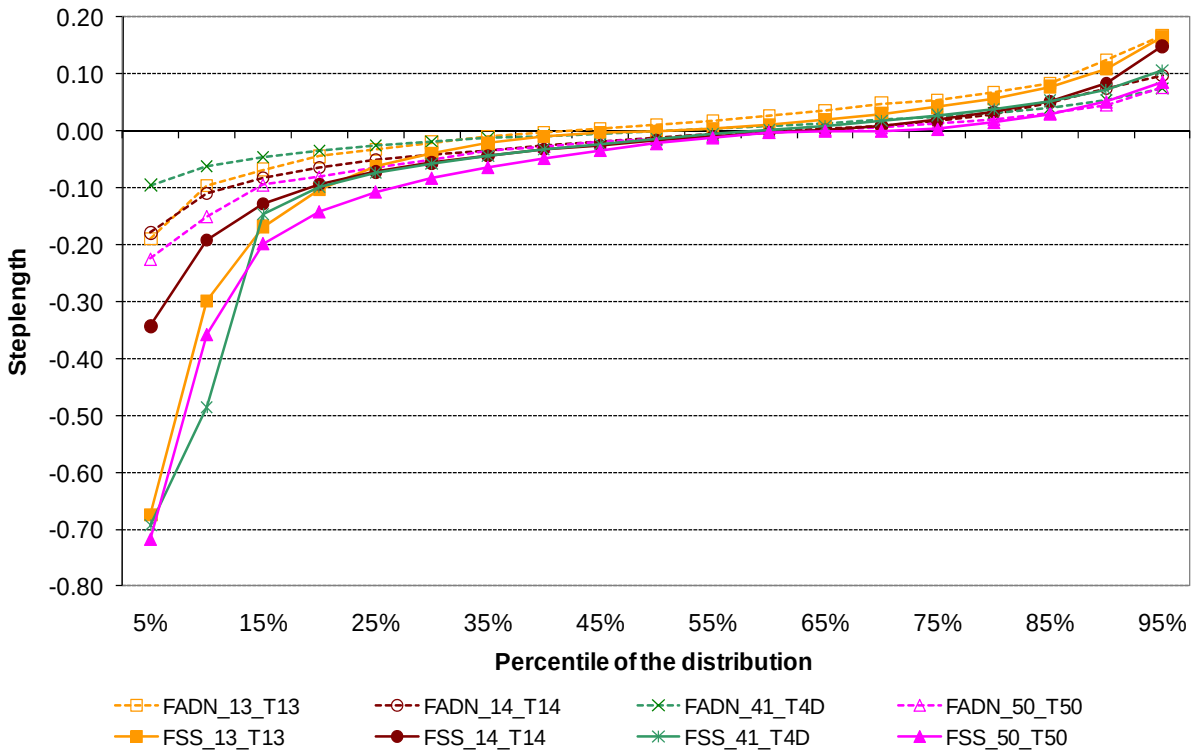


Source: Own calculation based on the sample of German FADN and FSS farms

First, the sample is still too small for out-of-sample predictions for a medium time horizon. The extension of the training set from 1989-1999 to 1989-2003 reduced the MAD

(Median Absolute Deviation) of the test set by nearly a third (not shown). Second, up to the late 1990s, the data for many variables recorded in the FADN contain a fairly large amount of implausible data (cf. Figure 7). The third reason for the mediocre quality is more severe. The recorded behaviour regarding the structural adjustment of the FADN farms differs substantially from that observed in the FSS population. In the FADN sample, there are relatively fewer farms that significantly change their productive orientation when compared with farms that do not (or only moderately) change their productive orientation (Figure 16). Up to a Steplength of 0.05, the distribution is comparable for the farms recorded in FSS and FADN. The picture changes for larger Steplengths. Step-lengths beyond 0.15 are observed for less than 5% of the FADN farms, whereas changes of this magnitude are recorded for more than 15% of the FSS farms. The reason for this bias is the sampling protocol, as farms that change their productive orientation or farm size so significantly that they would have to be attributed to a different type of farming are removed from the sample. This sampling bias affects the MCI and the Markov estimation, as FADN micro data are used for both estimations.

Figure 17: Comparison of the change in the dominant specialisation of selected specialised farm types within 4-year intervals (average of the two periods 1999-2003 and 2003-2007) for FSS and FADN

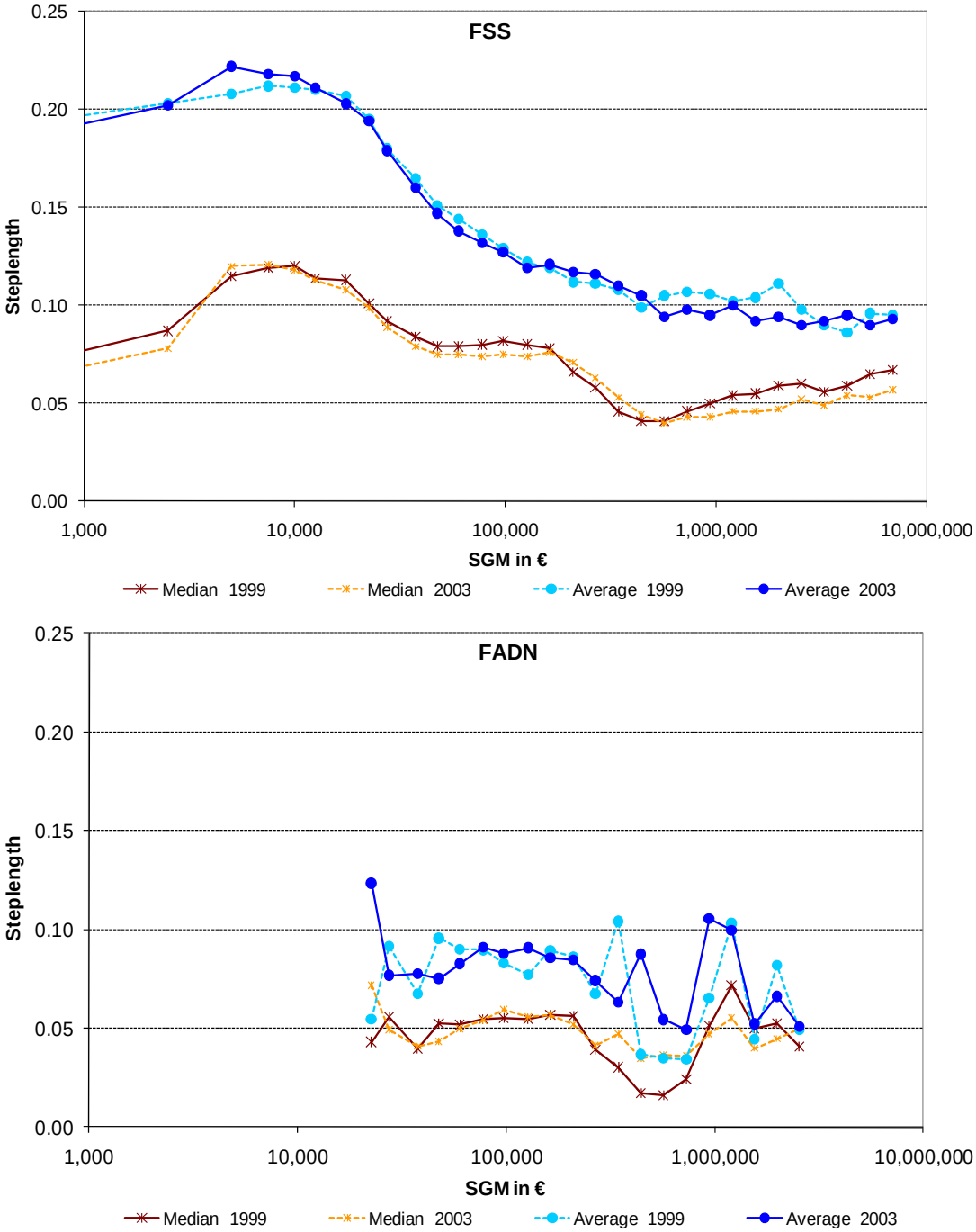


Source: Own calculation based on the sample of German FADN and FSS farms

Figure 17 illustrates that the bias introduced by the removal of farms that substantially alter their farming structure is not symmetrically distributed. This bias is exemplified for

the development of the main specialisation of different specialised types of farming (based on the 2-digit FADN typology). Although the share of farms that maintain or even increase the share of their main specialisation is comparable for both data sets, the share of farms quitting or nearly abandoning their main specialisation is substantially smaller in the FADN sample.

Figure 18: Median and average 4-year Steplength for the periods of 1999-2003 and 2003-2007 as a function of farm size for FSS and FADN farms



Source: Own calculation based on the sample of German FADN and FSS farms

A fourth reason for the reduced dynamics in the FADN sample is that the most dynamic segment, the small farms, is missing from the sample (Figure 18). In the observed periods, farms with an SGM below 7 ESU are responsible for roughly 80% of the farm exits, but roughly a quarter of all farms growing into a larger size class also belong to this group. However, even for farms of comparable size, the FADN sample has only roughly two-thirds of the dynamics observable in the FSS population.

1.5 Summary

The previous chapters have shown that the Markov and MCI approaches are both quite useful for the in-sample prediction and explanation of structural change in agriculture. However, the tools only barely or do not outperform naive approaches for out-of-sample prediction. This poor performance can be partly attributed to the peculiarities of the data set used. First, agriculture is not a very dynamic segment of the German economy. Therefore, the assumption of no change or business as usual is relatively accurate. Second, at least for predictions for a medium time horizon, the number of observations is still fairly limited. For the Markov approach and an eight-year prediction period, the prediction is based on only 70 observations (7 West German FADN regions and 10 periods). The situation is even worse for explanatory macro variables such as prices or general economic factors, e.g., the unemployment rate. For prices, there is only one price for each commodity per year. In addition, the prices for the different commodities are correlated with each other (e.g., rape seed and soft wheat $r^2 = 0.52$). Correlations are also present for factors such as the unemployment rate, for which the temporal development across the different FADN regions is synchronised by general economic cycles. Third, the MCI and the Markov approaches assume that the micro transitions, (i.e., changes in farm size and/or type of farming) observed in the FADN sample are somehow representative of the whole population. However, this is clearly not the case, as the dynamic subset of farms is automatically removed from the FADN sample because of the sampling protocol. This is a particular problem for the MCI approach, as the MCI, in contrast to the Markov approach, cannot compensate for the effect of biased micro transitions. Fourth, for the estimation period (training set), the data for some decisive variables are unreliable (e.g., subsidies or age). Fifth, the current estimations assume that a change in the explanatory variables will have an effect on structural change in the next year. This is rather unlikely; it is much more reasonable to assume that the structural change is more the result of the medium-term development of the explanatory variables. However, the time periods over which the development of the explanatory and response variables must be aggregated to obtain reasonable results remain an open question. Sixth, the influence of some variables on structural decision changes over time, e.g., prior to 2003, the first pillar clearly favoured some activities via coupled payments (e.g. fattening bulls), which are no longer supported after the implementation of the SFP (Single Farm Payment).

These problems not only limit the usefulness of the prediction but also demand caution when interpreting the coefficients derived in the regressions for the explanatory variables. A striking example is the durum wheat price in the MCI models. Although the cultivation of durum wheat is negligible in Germany, the removal of this indicator from the list of explanatory variables would result in a strong increase in the AIC (and therefore a worse model fit).

2 Ex-ante simulation of structural change

2.1 Introduction

The aim of this section is to discuss ex-ante approaches for implementing structural change in the CAPRI model. We make a distinction between the inclusion of structural change in the baseline and during simulation. The current report focuses on the inclusion of structural change in the baseline. With regard to the inclusion of structural change during simulation in the current version of CAPRI, the farm-type supply models interact with the market model to balance demand and supply using endogenous prices under the assumption of fixed factors of production such as land[7], capital and labour. When the ex-post approach proves that variables exist to explain structural change, the relationship should be established in the simulation. Then, factor endowment is no longer fixed. When the variable is also endogenous in the supply model, such as prices in CAPRI, we have a simultaneous dependence during the simulation that can be solved by iterative processing. The inclusion of structural change during simulation should be assessed in future projects to fully integrate structural change in the CAPRI model.

CAPRI's baseline is not the outcome of a simulation; rather, it is a consistent forecast combining trends and other information. A farm type in the baseline is characterised by a land endowment and input and output coefficients. The forecast depends on the projection of land use and production at the regional NUTS-2 level. Thus, the farm type baseline does not rely on expert information or other trend values derived at the farm type level because time series are generally not available. To improve this situation, we use information from the Markov and the continuous ex-post estimations. We can use both approaches to project the change in the future, using assumptions regarding the development paths of all relevant explanatory variables. For the baseline, the Markov approach gives forecasts for the potential number of farms in a farm type and hence for the factor endowment change, and the continuous approach gives forecasts for the evaluation of production branch shares[8] in a farm type.

This chapter proceeds along the following path. Section 2.2 introduces the baseline concept for the farm types, and Section 2.3 presents the methodology how the baseline estimation is extended to account for structural change (estimated in the ex-post analysis). Subsequently, the empirical implementation and the results of the suggested approaches

---

[7] We should note that, with the introduction of the land supply function in the CAPRI model, land is not a fixed factor. Therefore, there are behavioural functions for agricultural land supply and the transformation between arable land and grassland. However, the land cannot move between farm types.

[8] A production branch is defined as share of a group of activities, for example, all cereal and protein crop activities, valued by the standard gross margin relative to the total standard gross margin of a farm type.

are described in Section 2.4. The presentation of the farm types and the region used to demonstrate the approaches follows the results section, in which the current implementation is compared to the suggested approach. The chapter ends with the conclusions in Section 2.5.

2.2 The baseline concept of farm types in CAPRI

The baseline relies on the combination of two sources: the projection values of historical trends and expert information. A three-step "top down" approach to trends at the Member State (MS) level and NUTS-2 level as well as expert information on the EU and MS level are combined to ensure consistency of results across regional aggregations. The outcome of that process is a set of variables (e.g., hectares for the crop, herd sizes, input-output coefficients) for each NUTS-2 region in the EU at a certain point in the future that are used to calibrate the baseline.

Because time series are not recorded at the farm-type level, the most likely approximation of a farm type in the baseline is obtained by considering the NUTS-2 values as fixed and multiplying the base period value of a farm type by the ratio between the projected value and the base year value at the NUTS-2 level.

To illustrate this, we organise all values of the mathematical programming supply model for a specific farm type (*f*) with *f*=1,2, .. F in a four-dimensional matrix *d*, with the dimensions time (*t*), columns (*c*), and rows (*r*) and a regional dimension (*nts*). As a reminder, d is always defined over the four dimensions, even if indices are not presented in the formula. A detailed description of the column and rows (representing the variables in the programming models) is given in the CAPRI manual (BRITZ and WITZKE, 2008). The approximation[9] of the farm type baseline values, assuming 2004 as base year, in a certain NUTS-2 region and for the specific baseline in 2020 is calculated as:

(17): $$d^{a}_{2020,c,r} = \frac{d_{2020,c,r,nts}}{d_{2004,c,r,nts}} \times d_{2004,c,r} \forall f \in F$$

Note that $d^{a}_{2020,c,r}$ contains all necessary values such as yields, cropping area and nutrient intake for the mathematical programming model (*f*) in the baseline. However, it is likely that inconsistencies, such as mismatching production values or unsolvable technical relationships, occur if Equation 17 (linear scaling) is used without any further modification. Thus, we try to find a new set of $d^{v}_{c,r}$ for each farm type *f*, subscripted with *v*, which satisfies our consistency constraints and is similar to the approximated values in Equation 17. To demonstrate this approach, we consider the consistency constraint that production must equal the cropping area multiplied by the yield[10]. The following

[9] The approximation index (*a*) in a Bayesian estimation framework is sometimes called prior value.

[10] The baseline estimation normally consists of more constraints; here, we have chosen only two constraints as a didactic example.

constraint defines gross production (*g*) for all products (*o*) for all production activities (*j*) in d. Because a production activity, such as wheat, produces more than one output (such as corn and straw), yields are indicated by the combination of *j* and *o*. Note that, in d, the land use and herd sizes are indicated by the column *j* and the row item *l* (level). For mathematical correctness, the transposition (indicated by subscript T) of matrix d must be considered.

(18): $$d^{v}_{"g",o} = d^{v}_{j,"l"} d^{v}_{l,o}{}^{T} \qquad \forall f \in F, t = 2020$$

Another constraint ensures that the production at the farm type level sums to the NUTS-2 fixed production *(fix)* defined in the CAPRI baseline.

(19): $$d_{"g",o,nts}{}^{fix} = \sum_{f=1}^{N} d^{v}_{j,"l",f} d^{v}_{j,o,f}{}^{T} \qquad \forall t = 2020$$

If we assume that $d^{v}_{c,r}$ are realisations of a random variable, with N ($d^{a}_{c,r}, \sigma^2$), the most likely values for each farm type can be obtained by solving the following minimisation problem, subject to the constraints in Equation (18) and (19) (HECKELEI et al. 2005),

(20): $$\min \mathrm{vec}\left(d^{v}_{c,r} - d^{a}_{c,r}\right)' \times \aleph^{-1} \mathrm{vec}\left(d^{v}_{c,r} - d^{a}_{c,r}\right) \qquad \forall f \in F, t = 2020$$

where $\aleph$ is the covariance matrix and the prior vector $d^{a}_{c,r}$ for all elements of d, such as the approximation for land use $d^{a}_{j,"l"}$ with *j* cropping activities and the approximation for yields ($d^{a}_{j,o}$), with *o* products. Deviations from the prior values are allowed but are penalised using the quadratic loss function. All variables $d^{v}_{c,r}$ for all *f* are solved simultaneously using the given and fixed NUTS-2 region constraints indicated as *fix*. The estimated matrix $d^{v}_{c,r}$ for each *f* is then a consistent set of values for the baseline.

2.3 Considering the structural change in the baseline estimation (link between ex-post and ex-ante methodology)

In this section, the standard estimation approach is first explained, in which we consider the information regarding the number of represented farms derived from the Markov approach in the baseline as certain and given. Subsequently, the standard approach is extended by relaxing this assumption. The number of represented farms in the baseline enters the estimation in the form of prior expectations and is therefore estimated endogenously. Furthermore, the approach is improved by also considering the information about the "type of farming" and "economic size" of a farm type as a consistency constraint, which introduces the possibility of including the projection results derived from the continuous ex-post estimation approach in the baseline estimation in CAPRI.

2.3.1 Standard approach

The inclusion of structural change in the baseline can be achieved by calculating the prior values in Equation (17) using the projected number of represented farms in each farm type N, obtained from the Markov approach. To achieve the inclusion of structural change, the values for each activity (cropping area and herd sizes) are calculated for each farm type

for the year 2020. In Equation (18), total farm type activity levels ($d_{2004,f,"t"}$) are converted into average values per represented farm type in the base year by dividing by the total number of farms in the base year (N_{2004}). This average is multiplied by the projected number of farms in the year 2020 (N_{2020}) to obtain the most probable farm areas

and herd sizes $d^{a}_{2020,f,"t"}$.

(21): $$d^{a}_{2020,f,"t"} = \frac{d_{2004,f,"t"}}{N_{2004}} * N_{2020} \forall f \in F$$

The resulting approximation matrix $d^{a}_{2020,f,"t"}$, together with the other elements in matrix d calculated in Equation (17), enters the estimation similarly to Equations (19) and (20).

2.3.2 Extension with endogenous number of represented farms

We can also consider the projected number of represented farms in the baseline N_{2020} for each farm type as an approximation with an error and a standard deviation. This change is

indicated by denoting N_{2020} as N^{a}_{2020}. Our estimation is extended by the new variable for the evolution on the number of farms (N^{v}_{2020}) and the average farm values for farm area and herd sizes indicated by $d^{a}_{2020,f,"l"} = \frac{d_{2004,f,"l"}}{N_{2004}}$ $\forall f \in F$ to:

(22): $$d^{v}_{"g",o} = d^{v}_{f,"l"} N^{v} d^{v}_{f,o}{}^{T} \qquad \forall f \in F, t = 2020$$

where production sums to the NUTS2 production by:

(23): $$d_{"g",o,nts} = \sum_{f=1}^{N} d^{v}_{f,"l",f} N^{v} d^{v}_{f,o,f}{}^{T} \qquad \forall t = 2020$$

and the object function, using the number of represented farms N^{a}_{2020}, extends towards:

(24): $$\min \quad \mathrm{vec}\left(d^{v}_{c,r} - d^{a}_{c,r}, N^{v}_{2020} - N^{a}_{2020}\right)' \times \aleph^{-1} \mathrm{vec}\left(d^{v}_{c,r} - d^{a}_{c,r}, N^{v}_{2020} - N^{a}_{2020}\right)$$

$$\forall f \in F, \forall t = 2020.$$

Note that the estimation results now refer to the average farm type area and herd sizes $d^{v}_{f,"l"}$. Values can be calculated by multiplying the results by N^{v}, total.

2.3.3 Model with consistency constraints for type of farming and economic size

The model can further be extended by adding constraints to keep each farm type in its "type of farming" and "economic size" definition. This can be interpreted as a safety net during estimation in case large changes from the NUTS-2 level lead to drastic deviations in Equation (24). To obtain this safety net, it is necessary to calculate endogenously the

production branch share ($p^{b,v}$) as the ratio of the partial standard gross margin and the total SGM as given in Equation 25. The ratios are then not allowed to violate a set of defined ranges in manner similar to crop rotation restrictions. The ranges must be implemented in the estimation in the form of additional constraints for all types of farming according to the rules outlined in EU Commission Decision (85/377/EEC) as well as for the economic size classes. An example is provided in GOCHT and BRITZ (2011).

(25): $$p^{b,v} = \frac{\sum_{j \in B_b} (s_j d^{v}_{j,"l"})}{\sum_{j} (s_j d^{v}_{j,"l"})} \qquad \forall f \in F; b \in P1..P5 \quad .$$

Equation (25) calculates the five production branch shares, where *s* is the specific standard gross margin of production activity (*j*) and *b* refers to the five production branches (P1-P5), to classify the farm type in a certain type of farming.

To ensure that a farm type is consistent with the economic size of the farm type, the total standard gross margin must be converted to the economic size measure given in Equation (26) during the estimation, where 1200 € equals an Economic Size Unit (ESU):

(26): $$\mathrm{ESU}^{v} = \frac{\sum_{j}(s_j d^{v}_{j,r,t})}{1.200} \qquad \forall f \in F$$

As an example, if a farm type belongs to the ESU group greater than 100 ESU, the constraint:

$$\mathrm{ESU}^{v} > 100 \qquad \forall f \in \text{with ESU} > 100 \text{ in the base year}$$

must be satisfied during the baseline estimation, in addition to the constraints for partial standard gross margin shares, to ensure that the type of farming is maintained. The objective function remains as given in Equation (24).

2.3.4 Model to include the continuous ex-post results

The outcome of the continuous ex-post estimation can be included by also adding the partial shares to the objective function. Similar to the model with an endogenous number of represented farms, $p^{b,v}$ is included in the objective function, and the deviation is penalised from the projected shares ($p^{b,a}$) from the ex-post continuous approach, which results in an objective function of the form:

(27): $$\min \quad \mathrm{vec}\left(d^{v}_{c,r} - d^{a}_{c,r}, N^{v}_{2020} - N^{a}_{2020}, p^{b,v} - p^{b,a}\right)' \times$$

$$\aleph^{-1}\,\mathrm{vec}\left(d^{v}_{c,r} - d^{a}_{c,r}, N^{v}_{2020} - N^{a}_{2020}, p^{b,v} - p^{b,a}\right)$$

$$\forall f \in F, t = 2020$$

subject to the technical constraints and the constraints to conserve the type of farming definition and the economic size of a farm type.

2.4 Results including structural change during baseline estimation

This section presents the baseline estimation for the farm types when taking the projected number of farms for 2020 into account. We use a NUTS-2 region in Lower Saxony for the numerical presentation. The section begins with the description of the empirical implementation of the baseline estimation problem outlined in the previous section and three structural change implementations. The first implementation refers to the standard approach for structural change as outlined in Section (2.3.1). The second implementation refers to the model with consistency constraints for type of farming and economic size as outlined in Section (2.3.3). A third model, as an extension of the second model, takes the SGM shares in the baseline estimation as given in Section (2.3.4). The model implementation with an endogenous number of represented farms as outlined in Section (2.3.2) was not considered in the empirical implementation. The ex-post applications do not provide information regarding the moments of the error distribution for the number of farms ex-ante, which would be needed to parameterise the penalty function. How to derive such information from the applied Markov estimation has not yet been solved, as the estimation approach is not a normal OLS (ordinary least square) estimation.

Each model implementation produces several thousand values because all farm types in a NUTS-2 region are solved simultaneously, and for each farm type many values (e.g., yields, activity levels, fodder input and animal requirements) must be estimated. For the presentation of results, we use the utilised agricultural area (UAA), the economic size class ($\mathbf{ESU}$) and the standard gross margins ($\mathbf{p}^{b,a}$) as well as the livestock density. The implementations for structural change are compared with the current naïve farm-type baseline.

2.4.1 Empirical implementations

The estimation for the naïve baseline consists of the following equations given in Table 19. The first equation is the objective function [see also Equation (20)]. As described above, all values of the mathematical programming supply model for a specific farm type (f) are organised in a four-dimensional matrix d, with the dimensions time (t), columns (c) and rows (r) and a regional dimension (nts). A variable entering the objective function is described by a row and a column combination. The variables used are given in Annex 4 (Ex-ante projection) . For all variables, the approximation values ($d^{a}_{2020,c,r}$) are calculated based on Equation (17). The equations in Rows 2-5 in Table 19 endogenously calculate the gross production, UAA, livestock units per UAA and the total area of certain crop groups. These variables enter the objective function. Deviation from the approximation is

penalised during the estimation. Equation REQS_ ensures that the animal requirement for energy, crude protein and fibre is consistent with the estimated yield development. Equation STRA_ ensures that yield changes have the same relationship with the straw yield changes for cereals. Equation EFED_ ensures that non-tradable fodder such as silage maize and grass is consumed by the animals in a farm type. Equation NT2GROF_ adds the restriction that all farm types together produce and do not require more than was given at the NUTS-2 region. Comparable restrictions are enforced for the area and herd sizes in Equation NT2LEVL_. Equation NT2FEEDI_ ensures that cereals and protein-rich tradable fodder consumed in the farm types sum to the NUTS-2 level values. The endogenous calculation of the partial SGM (p) and the ESU is performed via the last four equations and is performed in the naive baseline estimation only to obtain results that can be later compared to the structural change implementations.

The *standard approach* for structural change [see Section (0)] uses the same equations as in Table 19 and also the approximations for yields and inputs. An exception is the calculation of the approximation of the animal herd and cropping area, which is performed using Equation (21); here, the number of farms in the base year and the predicted farm number in 2020 are used. The resulting herd sizes and cropping areas are up- or down-scaled to NUTS-2 values. Then, the approximation for UAA (Utilised Agricultural Area), LU (Livestock Unit) per UAA and the area of each crop group as well as gross production are calculated for each farm type. The second approach is an extension of the standard approach in which equations are added that keep a farm type in its specialisation and economic size class (see Section 2.2.1.3). This is necessary because the new farm structure based on the projected farm numbers can result in a violation of the structure and size of the farm type. To implement this approach, the shares of the partial SGM and the ESU are calculated using the first four equations in Table 20. In addition, five equations enter the model as constraints for the three ESU classes and 36 equations for the type of farming. During the estimation for each farm type, the related type of farming and ESU constraints are activated. The third approach extends the second approach by adding the partial SGM shares into the objective function as a variable. This is technically achieved by adding, for all partial standard gross margins p, a column ($d^{F}_{2020,c,r}$); as a corresponding row, we defined "*l*" as LEVL. We need to define our approximation for the new variables. The partial standard gross margins of the base year are considered to provide the best information regarding the future shares in the baseline, as the ex-post estimation revealed that the past shares provide the best information for future development.

Table 19: Equations in the baseline estimation

Level of application	Equation name in CAPRI	Function	Type of equation
Farm types	SSQ_	The objective function, which minimises the deviation between the approximation and the estimated variable	Objective function
Farm types	GROF_	The gross production (yield x activity level)	Calculation of variables entering SSQ_
Farm types	LU_	Calculates the Livestock Units per UAA	
Farm types	GRPLVL	Calculates the area of a group of crops	
Farm types	AREA_	Calculates the UAA for each farm type	
Farm types	REQS_	Requirements balance on linear regression results	Constraint
Farm types	STRA_	Keep estimated straw yield to estimated main yield relation	
Farm types	EFED_	Ensures that produced fodder is used in the farm type	
All Farm Type in a NUTS-2	NT2GROF_	Consistency with upper regional level for gross production on farm	
All Farm Type in a NUTS-2	NT2LEVL_	Area activity levels and herd sizes must sum up to NUTS-2	
All Farm Type in a NUTS-2	NT2FEEDI_	Feed must be consistent with NUTS-2	
Farm types	ETSGM_	Calculates the total standard gross margin	For reporting
Farm types	ETSGM_ST_	Calculates the partial standard gross margin	
Farm types	EESU_	Calculates the ESU	
Farm types	ESHARE_	Calculates the partial standard gross margin shares	

Table 20: Additional equations for maintaining the ESU and type of farming during estimation

Category of equations with the corresponding equation names in CAPRI	Function	Type of equation
ETSGM_	Calculates the total standard gross margin	Calculation of variables entering the constraints below
ETSGM_ST_	Calculates the partial standard gross margin	
EESU_	Calculates the ESU	
ESHARE_	Calculates the partial standard gross margin shares	
eruleESU_2LL_	If farm type is in Size class 1, ESU must be greater than 0 ESU	Constraints
eruleESU_2LL_1	If farm type is in Size class 1, ESU must be less than 16 ESU	
eruleESU_3_	If farm type is in Size class 2, ESU must be greater than 16 ESU	
eruleESU_3_1	If farm type is in Size class 2, ESU must be less than 100 ESU	
eruleESU_4_1	If arm type is in Size class 4, ESU must be greater than 100 ESU	
erule1_1_, erule13_1_, rule14_1_	Ensures type of farming for "Specialist cereals, oilseed and protein crops" (FT 13) and "General field cropping" (FT 14)	Constraints
erule2_1_	Ensures type of farming for "Specialist horticulture" (FT 20)	
erule3_1_, erule31_1_, erule32_1_, erule33_1_, erule34_1_, erule34_2_,erule34_3_	Ensures type of farming for "Specialist vineyards" (FT 31), "Specialist fruit and citrus fruit"(FT 32), "Specialist olives" (FT 33)" and "Various permanent crops combined" (FT 34)	
erule4_1_, erule41_1_, rule41_2_,erule42_43_1_, erule42_43_2_,erule42_43_3_, erule44_1_, erule44_2_	Ensures type of farming for "Specialist dairying" (FT 41), "Specialist cattle-rearing and fattening" (FT 42) + "Cattle-dairying, rearing and fattening combined" (FT 43), and "Sheep, goats and other grazing livestock" (FT 44)	
erule5_1_	Ensures type of farming for "Specialist granivores" (FT 50)	
erule6_1_, erule6_2_, erule61_1_, erule61_2_, erule62_1_, erule62_2_, erule63_1_, erule63_2_	Ensures type of farming for "Mixed farming systems" (FT61 FT62 FT63)	
erule7_1_, erule7_2_, erule7_3_, erule71_1_, erule71_2_, erule72_2_	Ensures type of farming for "Mixed livestock holdings" (FT 7)	
erule81_1_, erule81_2_	Ensures type of farming for "Mixed holdings" (FT 8)	

The naïve baseline approach, as applied in the current CAPRI version; the standard approach that derives the structure of a future farm from the Markov projection; the

second approach, which is extended by constraints for type of farming and ESU class; and the third approach, in which the deviation of the partial SGM is penalised are implemented in the current CAPRI trend projection model (CAPTRD). All three versions can be optionally applied to the NUTS-2 regions. The code was implemented in the current trunk version of CAPRI. Because of refactoring work in CAPTRD, it is possible to run the trend projection for farm types in a separate task.

2.4.2 Data

As the NUTS-2 region we selected *Lueneburg (Lower Saxony)*, with the EUROSTAT number DE93, which includes, for several specialisations, the largest size classes with greater than 100 ESU. Such size classes particularly increase the number of farms ex-ante. In addition, this region also includes a residual farm type that represents a large share of the holdings that are mostly affected by structural change. In the Farm Structure Survey (FSS), 17,900 holdings were recorded in DE93 in the year 2003. FADN recorded 11,000 holdings, which is 38% less than the population in FSS because of the applied threshold of 16 ESU in Germany.

Table 21: Number of holdings for DE93

			Baseline 2020	
Statistic	**Number of Holdings in 1,000 in 2003**	**Farm Types in DE93 in the base year 2004**	**Number of Holdings in 1.000**	**Number of Holdings in 1.000**
FSS	17.9	General field cropping + Mixed cropping 16-100 ESU	1.2	0.48
FSS without <16 ESU	11.0	General field cropping + Mixed cropping >100 ESU	1.2	1.09
FADN	11.8	Specialist dairying 16-100 ESU	3.1	0.68
		Specialist dairying >100 ESU	0.7	0.92
		Mixed livestock 16-100 ESU	0.4	0.06
		Mixed crops-livestock 16-100 ESU	1.3	0.42
		Mixed crops-livestock >100 ESU	0.7	0.57
		Residual Farm Type	8.6	5.10
		CAPRI Farm Type total	17.2	9.41

The number of holdings can be confirmed by considering in the FSS statistics only farm groups above 16 ESU, which yields approximately 11,000 holdings. The CAPRI farm type layer includes eight farm types in DE93, which represents 17,200 holdings. The difference from FSS results from farm groups in FSS without UAA. These farm types are not considered during the selection routine for CAPRI. Table 21 also presents the number of farms predicted for the year 2020 derived from the stationary Markov approach and shows that, for all farm groups, the number of holdings is projected to decline, with the

exception of the farm group 'specialist dairying' greater than 100 ESU. The number of holdings declines by 45% until 2020; however, the UAA derived from the NUTS-2 baseline in CAPRI is almost identical (see Table 21). This trend can be confirmed using the latest available data until 2007. Compared with 2000, we had already observed a 20% reduction in 2007.

Table 22: Partial SGM shares, UAA and livestock density for all farm types in the DE93 base year

Farm Types in DE93	P1	P2	P3	P4	P5	ESU	UAA in 1.000 ha	Livestock Unit per UAA
General field cropping + Mixed cropping 16-100 ESU	0.73	0.03	0.02	0.11	0.12	43	60	0.44
General field cropping + Mixed cropping >100 ESU	0.71	0.12	0.01	0.06	0.09	143	161	0.36
Specialist dairying 16-100 ESU	0.13			0.86	0.01	69	173	1.40
Specialist dairying >100 ESU	0.17			0.81	0.01	133	74	1.53
Mixed livestock 16-100 ESU	0.21	0.02		0.50	0.29	64	19	1.71
Mixed crops-livestock 16-100 ESU	0.40	0.01	0.01	0.39	0.21	56	69	1.05
Mixed crops-livestock >100 ESU	0.43			0.33	0.23	145	80	1.15

Table 22 summarises the key characteristics for the farm types in the base year. Columns P1 to P5 present the partial SGM. Note that Table 30 provides the link between the production activity (j) and the partial SGM. The ESU of the farm types ranges from 43 to 145 ESU. Specialist dairying farms with ESU between 16 and 100 are using the largest share of UAA in that region. The last column describes the Livestock Unit per UAA.

2.4.3 Results

Table 23 presents the development of UAA. In the base year, the UAA in DE93 is 812,000 ha. The trend projection at NUTS-2 predicted a decrease in the UAA by 2.3% to 793,700 ha. The naive approach takes the percentage change between the baseline and base year per activity at the NUTS-2 level and applies it to the base year activity levels of the farm types to define the approximation (prior information) for the estimation. The estimated crop activity levels and, hence, the UAA deviate between -12.8% for general field cropping and mixed cropping farm types (16-100 ESU) and 8.5% for the specialist dairying farm types (>100 ESU). The application of the structural change standard approach results in a new area distribution over the farm types. All farm types with an ESU class greater than 100 ESU show increased UAA; the remaining farm types show decreased UAA. This is in line with the finding that the number of holdings declined, whereas the UAA at the NUTS-2 level remains practically unaltered (2.3% reduction).

Table 23: Evaluation of UAA for the different implementations

	Base Year	Naive Baseline		Structural Change					
				Standard approach		Type of farming & ESU constraints		Endogenous partial SGM	
	UAA in 1,000 ha	UAA in 1,000 ha	% to BASM	UAA in 1,000 ha	% to BASM	UAA in 1,000 ha	% to STD	UAA in 1,000 ha	% to STD
General field cropping + Mixed cropping 16-100 ESU	59.7	52.0	-12.8	31.3	-47.6	31.3	-0.1	31.3	0.1
General field cropping + Mixed cropping >100 ESU	161.4	142.1	-12.0	182.6	13.1	186.8	2.3	182.4	-0.1
Specialist dairying 16-100 ESU	173.3	182.2	5.1	68.2	-60.6	67.5	-1.1	67.1	-1.7
Specialist dairying >100 ESU	74.5	80.8	8.5	195.3	162.1	194.4	-0.5	191.9	-1.7
Mixed livestock 16-100 ESU	18.8	18.5	-1.5	4.7	-75.2	4.7	0.4	4.7	-0.2
Mixed crops-livestock 16-100 ESU	69.4	65.3	-6.0	33.1	-52.3	33.2	0.2	32.7	-1.3
Mixed crops-livestock >100 ESU	80.4	77.1	-4.1	104.8	30.4	102.9	-1.8	115.9	10.7
Residual Farm Type	183.2	175.3	-4.3	173.4	-5.3	172.7	-0.4	167.4	-3.5
NUTS-2 DE093	812.0	793.3	-2.3	793.3		793.3		793.3	

Although a higher number of holdings was only predicted for the farm type specialist dairying (>100 ESU), scaling to the predicted NUTS-2 activity levels results in increases in UAA for the general field cropping and mixed cropping (>100 ESU) and mixed crops-livestock (>100 ESU) farm types as well. The introduction of constraints to ensure compliance with type of farming and ESU class restrictions leads to smaller changes in UAA when compared with the standard approach. We can conclude that at least one constraint is binding and forces a further adjustment during the estimation of the estimated production activity levels. The structural change estimation with endogenous partial SGM leads to a further adjustment of the UAA between 10.7% and -1.7% compared with the standard approach. The combination of declining numbers of farms and only a moderate decline in agricultural activity results in an increase in the average farm size represented by the ESU per farm given in Table 24.

Table 24: Evaluation of ESU for the different implementations

	Base year	Naive Baseline *		Structural Change					
				Standard approach		Type of farming & ESU constraints		Endogenous partial SGM	
	ESU	ESU	% to Base Year	ESU	% to Naive	ESU	% to STD	ESU	% to Naive
General field cropping + Mixed cropping 16-100 ESU	43.5	39.9	-8.2	59.8	37.6	57.6	-3.7	55.4	-7.3
General field cropping + Mixed cropping >100 ESU	143.2	136.6	-4.6	199.8	39.5	188.5	-5.6	183.6	-8.1
Specialist dairying 16-100 ESU	69.0	63.5	-8.0	107.2	55.3	100.0	-6.7	100.0	-6.7
Specialist dairying >100 ESU	133.1	125.1	-6.0	201.9	51.7	201.1	-0.4	193.8	-4.0
Mixed livestock holdings 16-100 ESU	63.7	58.8	-7.7	96.8	52.1	99.3	2.5	96.7	-0.2
Mixed crops-livestock 16-100 ESU	56.0	50.9	-9.1	83.5	48.9	88.4	5.9	80.8	-3.2
Mixed crops-livestock >100 ESU	145.2	143.1	-1.5	223.8	54.1	234.2	4.6	267.8	19.7
Residual Farm Type	45.0	42.1	-6.4	66.3	47.3	68.4	3.2	68.0	2.6

* Holdings in the base year

The naive baseline is calculated by taking the number of farms in the base year as a reference. As the production declines at the NUTS-2 level, the ESU declines in all corresponding farm types in the naive approach. If the structural change standard approach is applied, the ESU increases, as expected, for all farm types between 37.7% and 55.3%. In the standard approach, this leads to a violation of the economic size class with 107.2 ESU for specialist dairying farms (16-100 ESU). The application of the constraints and the consideration of the partial SGM during the estimation provide a production activity distribution without violating the economic size of the farm types. However, it is obvious that the farm type specialist dairying (16-100 ESU) is at the boundary with 100 ESU. Many environmental indicators, such as nitrogen surplus or greenhouse gas emissions, are closely related to the stocking density of a farm. Therefore, the baseline estimation approach should also consider this aspect. As shown in

Table 29, the LU/UAA is part of the endogenously calculated variable in the model and enters the objective function, which minimises the deviation from the approximation of LU/UAA. In Table 25, the estimation of the stocking density is presented. As expected, the farm types specialist dairying (cattle) and mixed livestock (cattle and pig fattening) have the highest stocking density. General field cropping + mixed cropping have a stocking density of approximately 0.4 LU/UAA. Decoupling of the premiums and the dairy reform decreases the cattle activity and increases pig and poultry activities at the NUTS-2 level, which can also be observed in the naive approach. These stocking densities are taken during the structural change implementations as an approximation (prior information).

Table 25: Evaluation of LU/UAA for the different implementations

	Base year	**Naive Baseline**		**Structural Change**					
				Standard approach		**Type of farming & ESU constraints**		**Endogenous partial SGM**	
	LU/UAA	**LU/UAA**	**% to Base Year**	**LU/UAA**	**% to Naive**	**LU/UAA**	**% to Naive**	**LU/UAA**	**% to Naive**
General field cropping + Mixed cropping 16-100 ESU	0.4	0.5	3.6	0.5	8.7	0.5	2.2	0.5	0.0
General field cropping + Mixed cropping >100 ESU	0.4	0.4	16.9	0.5	9.5	0.4	-7.1	0.4	-4.8
Specialist dairying 16-100 ESU	1.4	1.1	-23.6	1.1	0.0	1.0	-3.7	1.0	-3.7
Specialist dairying >100 ESU	1.5	1.1	-25.4	1.1	-2.6	1.1	-1.8	1.1	-4.4
Mixed livestock holdings 16-100 ESU	1.7	1.6	-7.7	1.6	-1.3	1.6	0.6	1.5	-2.5
Mixed crops-livestock holdings 16-100 ESU	1.1	1.0	-4.1	1.0	3.0	1.1	9.9	1.0	1.0
Mixed crops-livestock >100 ESU	1.2	1.2	2.3	1.2	1.7	1.3	11.9	1.3	12.7
Residual Farm Type	0.7	0.6	-16.8	0.6	1.8	0.6	5.4	0.6	3.6
NUTS-2 DE093	0.9	0.8	-10.7	0.8		0.8		0.8	

Compared with the naive approach, the standard approach leads to a deviation of between -2.6% and 9.5%. Application of the type of farming and ESU constraints results in small increments for some farm types (e.g., for the farm type mixed crops-livestock (>100 ESU) and in a correction of the stocking density from 1.2 to 1.3 LU/UAA). A similar response can be observed for the structural change implementation with endogenous partial SGM.

Although some adjustments are required during estimation for the different structural change implementations, we can conclude that the stocking density for all farm types is well recovered with the estimation framework compared with the naive baseline. In Table 26, the partial SGM P1, P4 and P5 are represented. We did not consider P2 and P3, which are less relevant in that region, because they relate to permanent crops, fruits and vegetables. For both ESU classes of the general field cropping + mixed cropping farm type, the partial SGM share P1 is violated for the standard approach. Even at this early stage, the naive approaches for general field cropping and mixed cropping (>100 ESU) lead to a violation of a minimum two-thirds P1 share of the total standard gross margin. This violation is removed with the inclusion of the constraints for type of farming and ESU (0.67). The approach with endogenous partial SGM further closes the gap between the estimated variable and the approximations from the base year. For the mixed crops-livestock farm type, the partial SGM share P4 must be greater than one-third of the total SGM. This is not recovered during estimation in the naive baseline estimation or with the standard approach. With the explicit consideration of these constraints, the partial SGM share P4 increases to one-third.

Table 26: Evaluation of the partial SGM shares for the different implementations

			Standard approach	Type of farming & ESU constraints	Endogenous partial SGM
			P1		
	Base year	**Naive Baseline**	**Structural Change**		
General field cropping + Mixed cropping 16-100 ESU	0.73	0.66	0.64	0.67	0.71
General field cropping + Mixed cropping >100 ESU	0.71	0.63	0.61	0.67	0.67
Specialist dairying 16-100 ESU	0.13	0.19	0.17	0.18	0.18
Specialist dairying >100 ESU	0.17	0.24	0.23	0.23	0.23
Mixed livestock holdings 16-100 ESU	0.21	0.22	0.21	0.20	0.21
Mixed crops-livestock 16-100 ESU	0.40	0.40	0.37	0.34	0.37
Mixed crops-livestock >100 ESU	0.43	0.40	0.38	0.35	0.36
Sum ABS deviation to Base Year		0.31	0.38	0.36	0.28
Sum % deviation to Base Year		110	111	121	103
			P4		
General field cropping + Mixed cropping 16-100 ESU	0.11	0.09	0.11	0.09	0.11
General field cropping + Mixed cropping >100 ESU	0.06	0.05	0.05	0.04	0.05
Specialist dairying 16-100 ESU	0.86	0.80	0.81	0.80	0.81
Specialist dairying >100 ESU	0.81	0.75	0.76	0.76	0.76
Mixed livestock holdings 16-100 ESU	0.50	0.45	0.47	0.47	0.48
Mixed crops-livestock 16-100 ESU	0.39	0.35	0.39	0.41	0.39
Mixed crops-livestock >100 ESU	0.33	0.28	0.30	0.33	0.33
Sum ABS deviation to Base Year		0.29	0.18	0.20	0.13
Sum % deviation to Base Year		84	39	69	26
			P5		
General field cropping + Mixed cropping 16-100 ESU	0.12	0.14	0.15	0.15	0.14
General field cropping + Mixed cropping >100 ESU	0.09	0.11	0.12	0.11	0.11
Specialist dairying 16-100 ESU	0.01	0.01	0.01	0.01	0.01
Specialist dairying >100 ESU	0.01	0.02	0.02	0.02	0.02
Mixed livestock holdings 16-100 ESU	0.29	0.33	0.32	0.33	0.30
Mixed crops-livestock 16-100 ESU	0.21	0.25	0.24	0.24	0.23
Mixed crops-livestock >100 ESU	0.23	0.25	0.26	0.26	0.26
Sum ABS deviation to Base Year		0.15	0.17	0.16	0.11
Sum % deviation to Base Year		126	133	140	75

The sum of the absolute deviation and percentage deviation is also presented in Table 26. The implementation with endogenous partial SGM outperforms the other estimation approaches.

2.5 Conclusion

The chapter describes and explains the introduction of structural change in the baseline process using the farm type layer in CAPRI. Three different implementations are compared with the current (naive) approach. The standard approach for structural change calculates the approximation of the animal herd and cropping area using the projected number of farms in the baseline. The second approach is an extension of the standard approach and adds equations that retain a farm type in its specialisation and economic size class. The third approach, which is also an extension of the standard approach, endogenously estimates the partial SGM shares and minimises their deviation from the given information in the base year. We apply a Bayesian motivated estimation framework, which treats the available information for each farm type as a random variable in the mathematical programming model. The base year observations, the relative changes from the trend forecast at NUTS-2 and the projected number of holdings provide the approximation (prior) information. The consistency and definition-based conditions provide the data information. Their combination gives posterior estimates that fulfil the top-level NUTS-2 disaggregation requirement while exhausting the information content of the raw data. The structural change approaches were successfully implemented into CAPRI and tested for one German NUTS-2 region. The comparison of the three approaches with the naive baseline demonstrates that the decrease in the number of farm holdings in the baseline leads to an overall increase in the farm size, represented by the ESU, for all farm types in the region. It is necessary to use additional constraints for the economic size during the estimation to ensure that the ESU class limits are not violated. The results from a growth model, using the development of the economic size of a farm group ex-ante, could be utilised in the form of percentage change for the farm groups to further refine the assumptions regarding the distribution of the overall farm size increase. This would allow capturing the effect that farms change their size while maintaining their initial size class while ensuring that the change in average size depends on initial size class and farm type.

Currently, a linear scaling is used to make the approximation for the farm size consistent with the NUTS-2 production activity levels. Furthermore, the comparison demonstrates that the type of farming can be violated during the estimation, particularly if a farm type is already close to one of the bounds that define its type of farming in the base year. The partial SGM shares of a farm type are a good indicator of the farm type production structure. Because constraints for the type of farming allow a relatively wide adjustment of the production structure of a farm type, the endogenous included partial SGM shares can better ensure that the production program is as similar as possible to that in the base year. The estimation also demonstrates that the structural change implementation recovers the livestock density of a farm type. The main aim of introducing structural change into the CAPRI farm type baseline is to improve policy impact assessments and hence build a

more reliable farm type baseline. The presented example demonstrates that information from Markov projections can be used to derive a consistent farm grid that represents on-going structural change and farm growth. In order to extend this approach to the EU-27, a complete database regarding the future evolution of farm types in the EU-27 is needed, however currently is not available. One empirical drawback is the increased computing time. The estimation with endogenous partial SGM requires up to 12 minutes for a single NUTS-2 region. One potential solution is the used of parallel processing which could diminish the computing time when extending the approach to more NUTS-2 regions.

As mentioned in the introduction (section 2.1) in order to fully integrate structural change in the CAPRI model, the inclusion of structural change during simulation should be assessed in future developments of the model.

3 General conclusions and outlook

The report presents some new insights with respect to the analysis of structural change in European agriculture using the two farm level databases currently available at EU level (FADN and FSS). The two methodologies applied in the ex-post analysis are the Markov chain for the discrete approach and MCI models in the case of the continuous approach.

The methodologies have been applied to German FADN farms. Furthermore, the ex-post results were incorporated in the baseline of the farm module of CAPRI, and therefore, the naive approach currently implemented in CAPRI has been improved.

Within the study, several aspects emerged that require further in-depth analysis in upcoming research activities. These aspects could be broadly attributed to the following topics:

- data characteristics
- methodological issues in the ex-post analysis
- methodological issues in the ex-ante analysis
- empirical questions in the ex-post analysis

Regarding the data characteristics, the following issues should be mentioned in the context of structural change analysis.

- The development of models explaining structural change is complicated by two peculiarities of the data set. First, the development of many variables, viewed as important for explaining structural change in agriculture (e.g., type and level subsidies) does not gradually change over time but is characterised by marked shifts generate by discrete events. In ex-post analyses based on an in-sample validation, this behaviour can be easily accounted for by the introduction of structural breaks. However, this behaviour makes the ex-post (using out-of-sample) validation and the forecast very difficult as the timing, size and direction are unknown in the model. Furthermore, many variables that are assumed to have an influence on structural change are co-integrated, e.g., the development of unemployment rates in different regions and the development of prices for different agricultural commodities. The only way to at least partly address the co-integration problem is to increase the number of observations (e.g., higher spatial resolution or broader spatial scope) to increase the variability observed in the data.
- The data quality is not constant over time and across items. For example, in the German FADN sample, the degree of detail to which subsidies are recorded changes over time, and the age of the farmer often contain implausible data. It is well known that the quality of the physical data is lower than that of the accountancy data, as fewer generally valid cross checks on the plausibility of the data are possible.

- FSS micro data could be used rather than FADN micro data to analyse structural change. In contrast to FADN, FSS contains less information per observation and is recorded at greater temporal intervals. However, the higher number of observations in FSS permits a higher level of detail regarding the analysis of explanatory variables with pronounced spatial patterns (e.g., regional economic or demographic development and abiotic factors such as climate, inclination and soils). A major advantage of the FSS dataset is that it contains information on farms exiting the sector (as the entire population is sampled), whereas FADN does not. Therefore, the use of FSS micro data would be useful for Markov approaches, as it would allow detailed transition matrices to be derived directly for each region. These transition matrices could be used to analyse the influence of the explanatory variables on structural change. However, because FSS micro data are usually not accessible, a practical way to exploit the advantage of both data sources (FSS and FADN) is to combine them in the analysis as it has been demonstrated in the Bayesian Markov stimation used in this study. Further research in this respect could improve the combination of the two data sources, for example, by specifically considering the different temporal resolutions of the two datasets (i.e., yearly FADN versus FSS data available every two to three years).
- A crucial aspect in the construction of econometric models analysing structural change over time is the selection of an appropriate base unit to determine the farm typology and the economic size. The Standard Output (SO)/ Standard Gross Margin (SGM) are used in order to define the farm size ("Economic Size Unit") and the typology of the farm ("type of farming"). Currently, the activity data are weighted within FSS and FADN using a new set of SGM/SO every three to four years. As a result, the development of the 'type of farming' and 'Economic Size Unit' (ESU) reflects not only the development of physical assets but also market prices and support policies (e.g., coupled payments included in the SGM). To analyse the physical development of farms, the SGM should be constant over time. Currently, the type and size of the farms in the population (FSS) are determined using different SGM/SO for every point in time. Consequently, the number of FSS farms represented by a single FADN farm (weighting factor) is determined by the variable SGM/SO. To obtain consistent information, a recalculation of the 'types of farming' and ESU using fixed SGM/SO in both FADN and FSS is necessary.
- The FADN German micro data are characterised by a marked bias regarding the volatility of the farms' productive orientation over time because farms are removed from the FADN sample if they change their 'type of farming' or 'ESU'. The only way to overcome this problem is to adjust the sampling protocol.

Regarding methodological aspects, the following issues of the ex-post analysis require further research and analysis:

- What is the appropriate temporal lag and averaging structure for fairly variable information such as prices and subsidies to detect their impact on structural change?
- Is the MCI (Multiplicative Competitive Interaction), the MNL (Multinomial Logit) or a mixed specification more appropriate for isolating the influences of the different variables?
- Could MCI or MNL models be used to directly analyse and project the shares of "*k*" different farm types on the regional level? In comparison to Markov models, a successful application would reduce the required number of observations as only the k-shares and not the k^2 transitions would be calculated in the MCI models. In comparison to an MCI approach at the farm level, the results of an MCI application at the regional level may be more stable as the changes observed at the regional level are much smaller and less discrete than those on the farm level.
- Will the use of FSS micro data or a combined data set (FSS / FADN) lead to similar and even more stable results? This question is particularly important in the context of an out-of-sample validation.
- A methodology must be developed to allow the grouping of farms throughout the EU-27 to a limited number of farm typologies to allow the estimation of structural change models across countries (or regions). Two different approaches can be envisaged. In the first, farms would be grouped primarily based on the observable variables at different levels at the starting point (the method currently chosen in the FADN or CAPRI methodology). In the second approach, the classification would be based primarily on the observed development (behaviour) within a certain time. Thus, for the classification, the level of a given set of specialisations is not as important as the magnitude and direction of the changes.

Regarding the methodological aspects of the ex-ante analysis, the current implementation of structural change in CAPRI could be extended in the following ways to improve its depiction of structural change.

- The implementation of an intrinsic growth model and a module depicting the shift of partial SGM (e.g., derived from an MCI model) would add a reasonable degree of flexibility to the model. The potential adaptations of a given farm type would no longer be limited by rigid bounds predetermined by the initial farm type. Although the previous point would already affect the baseline, the implementation of a module that endogenously simulates structural change in response to a policy shift affects only the simulation behaviour of CAPRI. The inclusion of this module is rather straightforward. The main constraint is derived from the lack of robust coefficients for the relationship between prices or certain policy-related drivers and structural change throughout the EU 27.

- The inclusion of structural change during simulation should be assessed in future projects to fully integrate structural change in the CAPRI model.

Regarding the empirical aspects in the ex-post analysis the following open questions should be further analysed:

- How relevant are different sub-processes for explaining structural change with respect to farm specialisation on the regional level? These sub-processes are farm-exit, farm-shrinkage with and without change in specialisation, and farm growth with and without change in specialisation.
- How is the change in farm assets (e.g. number of ha) related to the initial farm type?
- What are realistic bounds regarding the changes in average farm size and specialisation for different types of farming?

Futher insights in the last two questions mentioned are particularly important for the incorporation of structural change in ex-ante simulation models because this information would be incorporated in the models as constraints regarding the feasible adjustments of the farm types.

The following roadmap could be envisaged to analyse (ex-post) and implement "structural change" in the models (ex-ante) for the EU-27.

- Address the problems regarding data quality mentioned above;
- Solve the data issues mentioned above, particularly with respect to the accessibility to FSS to allow the recalculation of the farm types based on constant SGM (or SO);
- Modify the sampling protocol such that farms that modify their 'type of farming' or 'ESU' are not automatically removed from the sample;
- Develop algorithms to check the FADN data routinely for implausible developments over time, particularly for the development of physical assets;
- Run more applications (case studies) to determine and assess the best econometric estimations and derive robust estimates (Markov, MCI and growth models) covering the entire EU-27 for variables framing possible developments in the ex-ante models;
- Determine (based on a protocol) the farm typologies and farm sizes in each EU-27 region in which the farm numbers have to be estimated.

References

Bell, D. E., Keeney, R. L., Little, J. D. C., 1975. A Market Share Theorem: Journal of Marketing Research 12, 136–141.

Bundesamt für Kartographie und Geodäsie BKG 2007. DGM25 (Digitale Terrain Model) 1:25 000, Frankfurt/Main, Germany.

Bremmer, J., Lansink, A.O., Olson, K.D., Baltussen, W.H., Huirne, R.B., 2004. Analysis of Farm Development in Dutch Agriculture and Horticulture, Strategic Decision Making and Firm Development in Dutch Horticulture, 23-39.

Britz, W., Witzk., P., 2008. CAPRI model documentation 2008: version 2. http://www.capri-model.org/docs/capri_documentation.pdf

Brown, P.J., Payne., C. D., 1986. Aggregate data, ecological regression, and voting transitions. Journal of the American Statistical Association. 81 (394), 452–460.

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

Cooper, L. G., Nakanishi, M., 1988. Evaluating Competitive Marketing Effectiveness: Market-Share Analysis: Kluwer Academic Publishers, Boston. pdf-edition 2010.

Dick, J., Hetz, W., 2011. Milchquotenzu- und -abgänge (Saldo) seit Einführung der Quotenbörse, München, Germany. http://www.lfl.bayern.de/iem/milchboerse/27526/ linkurl_0_1_0_1.pdf.

Deutscher Wetterdienst (DWD), 2009. Monthly average temperatures and rainfall for the period 1961-1990 (1x1 km grid), Germany.

Farm Accountancy Data Network (FADN), 2010. Clustering rules per member state and accounting year. http://ec.europa.eu/agriculture/rica/annex004_en.cfm. Last accessed: 27.06.2011.

Faith, D. P., Minchin, P. R., Belbin, L., 1987. Compositional dissimilarity as a robust measure of ecological distance: Vegetatio 69, 57-68.

Fok, D., Franses, P. H., Paap, R., 2002. Econometric analysis of the market share attraction model, Advances in Econometrics, Volume 16, Emerald Group Publishing Limited, 223-256.

Gocht, A., Britz. W., 2011. EU-wide farm type supply models in CAPRI - How to consistently disaggregate sector models into farm type models: Journal of Policy Modeling 33(1), 146-167.

Greene, W.H., Hensher, D. A., 2010. Modeling ordered choices, Cambridge University Press, Cambridge.

Hawkes, A.G., 1969. An Approach to the Analysis of Electoral Swing: Journal of the Royal Statistical Society Series A General, 132(1), 68–79.

Heckelei, T., Mittelhammer, R. C., Britz, W., 2005. A Bayesian Alternative to Generalized Cross Entropy Solutions to Underdetermined Models, Contributed paper presented at the 89th EAAE Symposium "Modelling Agricultural Policies: State of the Art and New Challenges“, February 3-5, Parma, Italy.

Huettel, S., Margarian, A., 2009. Agricultural Economics: Structural change in the West German agricultural sector. 40, 759-772.

Kareiva P.M., Shigesada, N., 1983. Analyzing Insect Movement as a correlated random walk: Oecologia 56, 234-238.

Kotler, P., 1984. Marketing Management: Analysis, Planning, and Control, Fifth Edition, Englewood Cliffs, NJ: Prentice-Hall, Inc.

Laube, P., Dennis, T., Forer, P., Walker, M., 2007. Movement beyond the snapshot – Dynamic analysis of geospatial lifelines: Computers, Environment and Urban Systems, 31, 481-501.

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

Papastamatiou, P., Cartamil, Y.P., Lowe, D.P., Meyer, C.G., Wetherbee, B. M., Holland, K. N., 2011. Scales of orientation, directed walks and movement path structure in sharks, Journal of Animal Ecology, 80, 864-874.

Pietola, K., Vare, M., Lansink, A.O., 2003. Timing and Type of Exit from Farming: Farmers' Early Retirement Programmes in Finland: European Review of Agricultural Economics 30, 99-116.

Statistisches Bundesamt, 1993. Statistisches Jahrbuch für die Bundesrepublik Deutschland 1993, Metzler-Poeschel, Stuttgart.

Storm, H., Heckelei, T., Mittelhammer, R.C., 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.

Weiss, C.R., 1999. Farm Growth and Survival: Econometric Evidence for Individual Farms in Upper Austria, American Journal of Agricultural Economics, 81, 103-116.

Zimmermann, A., Heckelei T., 2012. Structural change of European dairy farms – a cross regional analysis: Journal of Agricultural Economics, 63:576-603.

Annexes

Annex 1

Description of the FADN variables

Annex 1: Description of the FADN variables

FADN-Variable	Description
D/1	Common wheat and spelt
D/2	Durum wheat
D/3	Rye
D/4	Barley
D/5	Oats
D/5	Oats
D/6	Grain maize
D/8	Other cereals
D/9	Protein crops
D/10	Potatoes
D/11	Sugar beet
D/12	Fodder roots and brassicas
D/14a	Fresh vegetables, melons, strawberries - outdoor - open field
D/14b	Fresh vegetables, melons, strawberries - outdoor - market garden
D/15	Fresh vegetables, melons, strawberries - under glass
D/16	Flowers - outdoor
D/17	Flowers - under glass
D/18	Forage plants
D/19	Seeds and seedlings
D/21	Fallow land without subsidies
D/22	Set-aside areas under incentive schemes - fallow land with no economic use
D/23	Tobacco
D/24	Hops
D/26	Soja, rape and turnip
D/27	Sunflower
D/29	Linseed (oil flax)
D/30	Other oil seed crops
D/34	Aromatic, medicinal and culinary plants
D/35	Industrial plants not mentioned elsewhere
F/1	Permanent grassland and meadow - pasture and meadow
F/2	Permanent grassland and meadow - rough grazings
G/1	Fruit and berry plantations - temperate climate
G/3	Olive plantations
G/4	Vineyards - quality wine
G/5	Nurseries
G/6	Other permanent crops
G/7	Permanent crops under glass
I/2	Mushrooms
J/1	Equidae
J/2	Bovine under one year old - total
J/3	Bovine under 2 years - males
J/4	Bovine under 2 years - females
J/5	Bovine 2 years and older - males
J/6	Heifers, 2 years and older
J/7	Dairy cows
J/8	Bovine 2 years old and over - other cows
J/9	Sheep
J/11	Pigs - piglets under 20 kg
J/12	Pigs - breeding sows over 50 kg
J/13	Pigs - others
J/13	Pigs - others
J/14	Poultry - broilers
J/15	Laying hens
J/16	Poultry – others

FADN variables according to Commission Decision 2003/369/EC

Annex 2

Description of the German FADN regions

Annex 2: Description of the German FADN regions

FADN-Code	Name	Region
10	Schleswig-Holstein	North
20	Hamburg	North
30	Niedersachsen	North
40	Bremen	North
50	Nordrhein-Westfalen	Centre
60	Hessen	Centre
70	Rheinland-Pfalz	Centre
80	Baden-Württemberg	South
90	Bayern	South
100	Saarland	Centre
110	Berlin	East
120	Brandenburg	East
130	Mecklenburg-Vorpommern	East
140	Sachsen	East
150	Sachsen-Anhalt	East
160	Thüringen	East

Annex 3

Trajectory analysis of the FADN data

Annex 3: Trajectory analysis of the FADN data

In this Annex we present an approach on how structural change, i.e., change in farm specialisation, could be measured and analysed on farm level. Furthermore, we present some first results based on the German FADN sample. In order to model structural change from the continuous perception two aspects and related methods should be considered:

- Change in farm size. Can be observed by a change in the total production or total standard gross margin (SGM) of a farm or farm group. A potential econometric methodology for the data analysis is the use of a dynamic panel data (DPD) estimator.
- Change in farm specialisation. Farm specialisation is expressed by the share of a certain production branch on the overall production. The development over time and given that a specialisation is multi-dimensional, multiplicative competitive models (MCI) can be considered for the analysis of potential drivers explaining the change in specialisation.

In contrast to the previous two aspects, farm exit must be perceived as a discrete event and can be analyzed with a Logit-model. As the calculation of farm exit (e.g., PIETOLA et al., 2003; HUETTEL and MARGARIAN, 2009) and change in farm size (farm growth rates) (e.g., BREMMEr et al. (2004); WEISS (1999)) based on micro level data is well established in agricultural economics, the analysis undertaken in this study is focused on the change in farm specialisation. In the next section, the methodology to calculate the change in farm specialisation later used as explanatory variable in the analytical estimation framework is explained. The explanatory model and its mathematical formulation in the form of a MCI-model are described in the literature review (section 3.4 in part II of the report).

The basic concept of the continuous approach is to analyse the agents' movement in time on the level of the single farm and to draw inferences for groups of farms in a second step.

The difference in measurement intervals between FADN and FSS (annual vs. 3-4 year interval) is important as it cannot be assumed that the movement of individual farms are purely directional and not random. In particular, we are interested in whether the magnitude and direction of the movement is conditional on, e.g., initial farm structure or region as this might crucially influence the model specification like the need to introduce regional or structural dummy variables or to model structural change by either a fixed or variable slope regression. Questions regarding the analysis of movements in space are widely addressed in geography or biology (e.g., LAUBE et al, 2007; PAPASTAMATIOU et al., 2011)). A movement in the data space is in the mathematical sense equivalent to a movement in a geographical space.

A3.1 Methodology

The EU-farm typology regarding farm specialisation is based on the relative shares of activity groups on the farm's total SGM. It is constructed on the base of a multi-dimensional data space derived from the SGM of the different agricultural activities[11]. If one introduces residual activity groups on all levels, the typology is constructed in a way that at a given level of topical resolution each activity is attributed to one and only one dimension, e.g., on the first level winter wheat is attributed to P1, on the second to P11, …., which means that the different partial SGM (P1-P5) must add to unity. Hence the following formula is generally valid:

(28): $$\sum_{g} s_{g,l} = 1; \forall\, l\,,$$

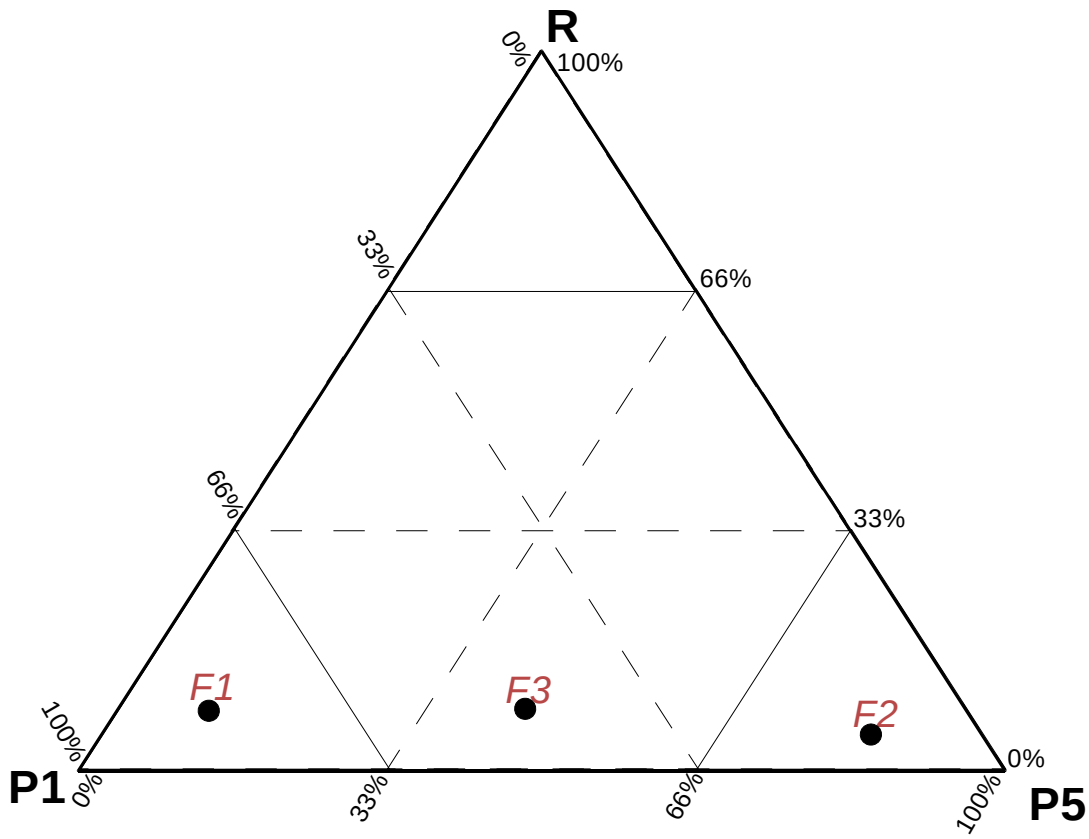
where s is the share of the g's group of activities on l level of topical resolution.

Figure 19 illustrates the location of the three farms (F1, F2 and F3) in a three dimensional data space[12], defined by the share derived from arable cropping (P1) and granivore production (P5) and a residual type (R). In each corner the share of one specific specialisation on the farm's SGM reaches 100%. For example, F1 derives nearly 80% of its SGM from arable cropping. F3 is mixed farm where both P1 and P5 contribute more than 33% to the farm's SGM. The larger the distance between two farms in Figure 19, the smaller is their similarity in farming structure.

[11] P1 = General cropping; P2 = Horticulture; P3 = Permanent crops; P4 = grazing livestock; P5 = granivores

[12] The data dimensions indicate the relative contribution of certain group of activities to a farm's total SGM. The data dimensions are not equivalent to types of farming but are the basis for their calculation.

Figure 19: Example of the location of three farms (F1, F2 and F3) in a three dimensional data space



Similarly one can depict the development of a given farm in time (Figure 20). Regarding this development of the farm's structure, a crucial question is whether this development is directional or if it is just the effect of chance events (random walk). In order to discriminate a random walk from a directional movement, we compute for farm *i* the relation between the distance from start to end point and the sum of the individual distances moved at time *t*:

(29): $$S_i = \frac{\left|\overrightarrow{F_{i,1}F_{i,T}}\right|}{\sum_{t=1}^{T}\left|\overrightarrow{F_{i,t}F_{i,t+1}}\right|}; \forall i.$$

If the movement is clearly directional S_i has a value of 1 and in case of pure random walk it is 0 (KAREIVA and SHIGESADA, 1983).

Figure 20: Development of farm F1 from t = 1 to t = 6

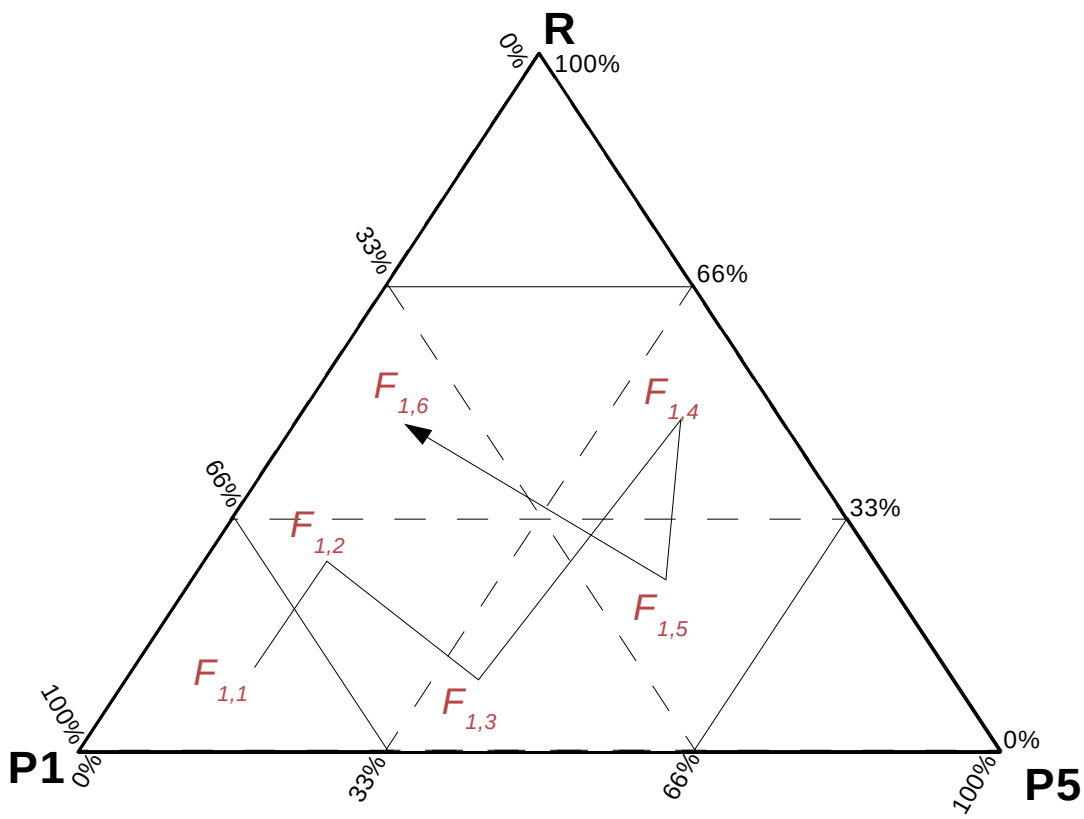
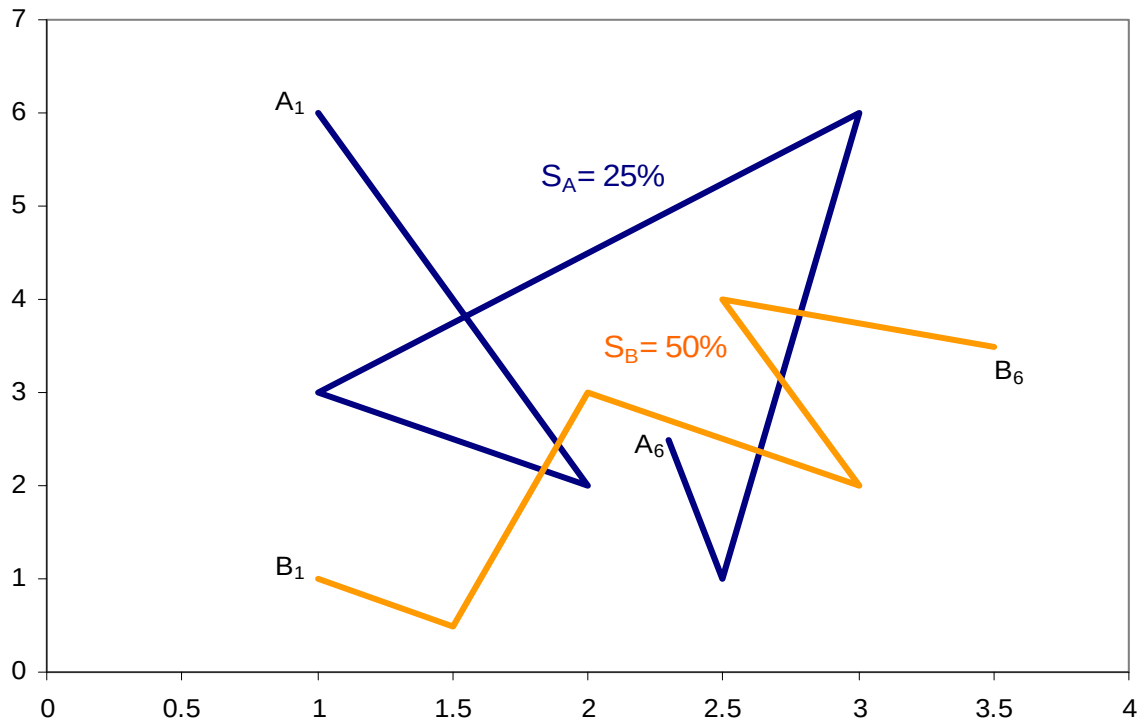


Figure 21 illustrates the S_i for two trajectories consisting of 6 points of observation.

Figure 21: Illustrative example of two trajectories over 6 periods



A variety of metrics are available in order to measure the distance between two observations. Table 27 shows that the choice of the metrics influences the obtained result. In particular, the classical Euclidean distance has the severe disadvantage, that despite the fact that in all cases farm "A" has nothing in common with its counterparts, the distance is larger when the counterpart is more specialised. The Manhattan Block distance and the spherical distance behave better in this respect. The use of the Manhattan Block distance is also widely advocated for in ecology for reason of numerical stability, where a comparable problem is frequently analyzed, measuring the similarity between two or more samples of vegetation communities (relevés) (e.g. FAITH et al., 1987).

Table 27: Illustrative example calculating the distance between farm A and three other farms

	Dimensions	Partial SGM of Farm A	Farm B	Farm C	Farm D
Activity groups	P1	1	0	0	0
	P2	0	1	0.5	0.25
	P3	0	0	0.5	0.25
	P4	0	0	0	0.25
	P5	0	0	0	0.25
Distance to Farm A					
	Euclidean		1.41	1.22	1.12
	Manhattan Block		2	2	2
	Spherical		1.57	1.57	1.57

The *Manhattan Block* distance between observations *a* and *b* is defined as:

(30): $$D_{MB}(a,b) = \sum_g \left| s_{a,g} - s_{b,g} \right| .$$

Therefore the distances are always measured parallel to the axis. According to our definition of the data space (Equation (28)), the shares (s_g) of a given observation add up to unity. Therefore, we could divide the D_{MB} by 2 as an increase in one direction is inevitable correlated to the same reduction in the other directions.

The use of the spherical distance is based on the following thoughts. Equation (31) is obviously similar to the definition of the surface of a *g* dimensional sphere, with unit radius:

(31): $$\sum_g \left(s_g\right)^2 = 1 .$$

If (s_g) shares are substituted by $\left(\tilde{s}_g\right)^2$ $\tilde{s}_g$ could therefore be treated as Cartesian coordinates describing the surface of a multidimensional sphere:

(32): $\sum_g (\tilde{s}_g)^2 = 1$; where $s = (\tilde{s}_g)^2$.

As s must be positive only one orthant of the sphere is defined. The distance between the two points a, b on a sphere is defined as:

(33): $D_{Sp}(a,b) = r * \cos^{-1}(\vec{a} \circ \vec{b})$.

Due to the fact that r is equal to 1, this reduces to the arccosine of the scalar product[13] of the two position vectors. As the sphere is defined only in one orthant, a division by $\pi/2$ standardizes D_{Sp} to the range 0 to 1. While D_{MB} treats changes in s (shares of each activity) comparable independently where they occur in the data space, D_{Sp} reacts very sensitive to the loss / additions of dimensions (Table 28).

Table 28: Illustrative example calculating the distances between farms using standardised Manhattan Block and Spherical Distance

	Dimensions	**Partial SGM of**			
		Farm A	**Farm B**	**Farm C**	**Farm D**
Activity groups	**P1**	0.50	0	0.55	0.40
	P2	0.40	0.25	0.45	0.05
	P3	0.10	0.25	0	0.10
	P4	0	0.25	0	0
	P5	0	0.25	0	0
Standardized distance			B	C	D
	Manhattan Block				
	A		0.65	0.10	0.10
	B			0.75	0.65
	C				0.33
	Spherical				
	A		0.69	0.20	0.07
	B			0.78	0.66
	C				0.36

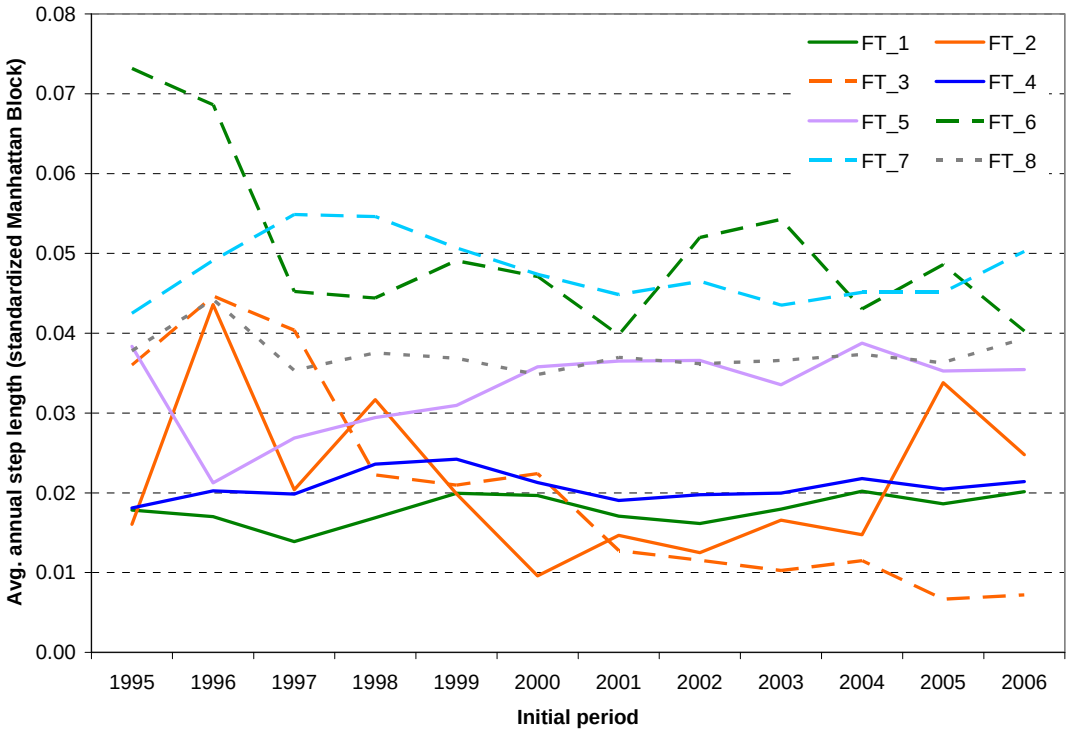
In the next section, it is presented the results regarding changes in farm specialisation based on FADN data for the period 1995-2007.

[13] The scalar product is the multiplication of two vectors defined as: $\vec{a} \circ \vec{b} = \sum_i^n a_i b_i$

A3.2 Results

In the following section we present some results to highlight the potential of the analysis of the trajectories. We focus the analysis on three aspects, the direction, the strength, and the randomness of the movement. As we are interested in structural change, i.e., the modification of the farm's physical layout, we use a constant SGM (2002) for all the years in order to level the effect of changing prices and costs between the years. The results are based on the complete German FADN farms in the period between 1995 and 2007. This analysis differs in two aspects from the results presented in the main text. First, the farms are aggregated according to the 1-digit FADN typology. Second, the changes in farm specialisation are analysed based on five data dimensions reflecting the 1-digit differentiation of the FADN-typology (General Type of farming)[14], i.e., the shares of P1 to P5.

Figure 22: Development of the average step length per farm type (1-digit FADN typology) measured by the Manhattan-Block-Metrics between 1995 and 2007



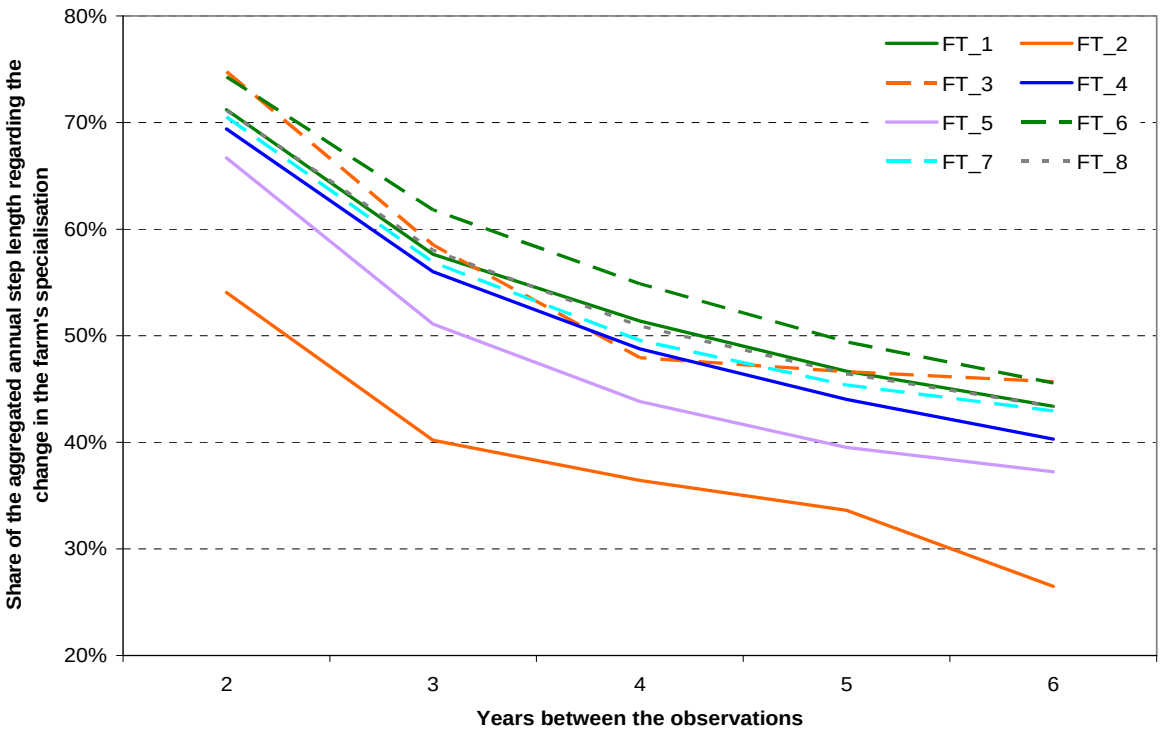
Source: Own calculation based on the German FADN-farms in the period 1995-2007. Only farms that remained at least 6 years in the sample were considered.

[14] FT1: Specialist field crops; FT2: Specialist horticulture; FT3: Specialist permanent crops; FT4: Specialist grazing livestock; FT5: Specialist granivore; FT6:Mixed cropping; FT7:Mixed livestock; FT8: Mixed crops livestock.

Figure 22 shows the development of the average step length of one year interval between 1995 and 2007. The step length is measured by the Manhattan Block metrics, therefore the figures correspond to a reallocation of activity shares between the specialisations in the same magnitude. For most farm types the average step length is fairly constant over time. The step length varies between over 0.07 for mixed plant production farms (FT 6) in 1995 and slightly below 0.01 in farms with permanent crops (FT 3) in 2006. Generally the dynamics is higher for mixed farm types (FT 6, FT 7 and FT 8) and for the farms specialised in granivore production (FT 5).

After we look at the step length, we will briefly analyze is the consistency of the development of the farms over time. We therefore calculate *S* according to equation (29). Figure 23 depicts the development of *S* as one increases the investigated interval from two to six years. Independent of the chosen interval, FT 2 has the lowest values for *S*. This means the farms belonging to this type show comparatively erratic movement through the data space. For a farm belonging to FT 2 the cumulated distance between the individual observations is nearly four times as long as the direct distance from the start to the end point if a six year time horizon is analysed. FT 6 is not only the type with the longest inter-annual step length but the farms belonging to this type show a comparatively directional movement in the data space.

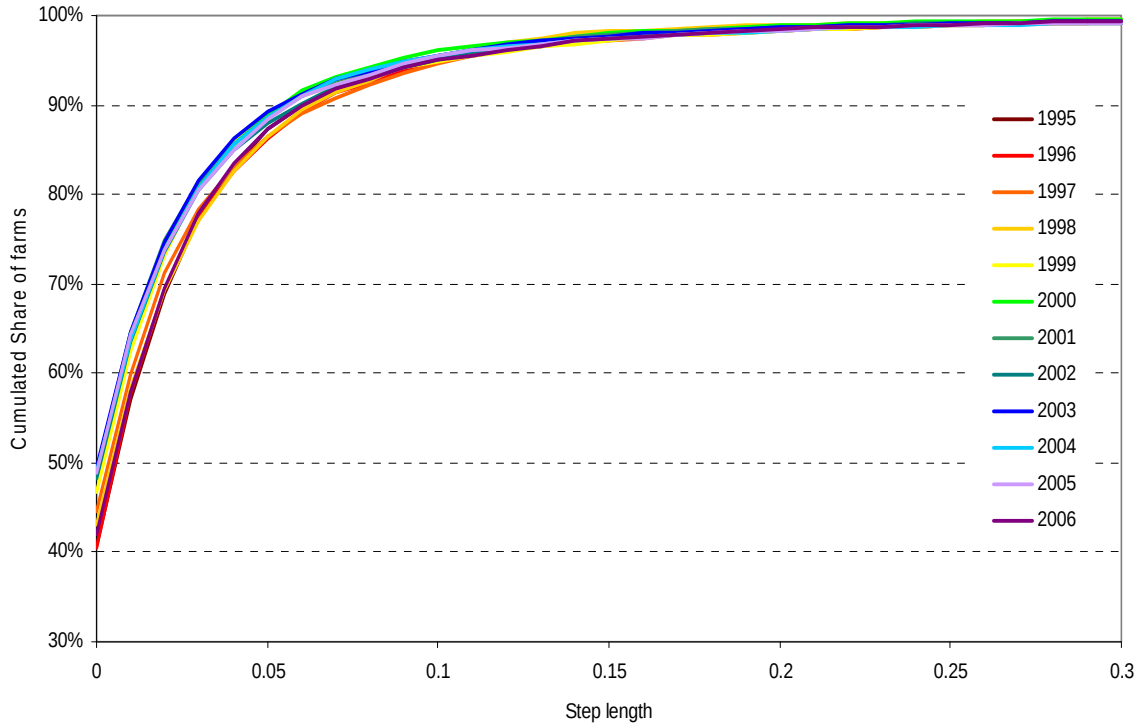
Figure 23: Relation of the one year to multi-year step length per farm type (1-digit FADN typology) measured by the Manhattan-Block-Metrics



Source: Own calculation based on the German FADN-farms in the period 1995-2007, Only farms that remained at least 6 years in the sample

The distribution of the annual step length is relatively constant over time (Figure 24). Depending on the year between 40% and 50% of the farms do not change their productive orientation at all (step length = 0). On an inter-annual basis shifts in the production orientation exceeding 10% (= step length > 0.1) are registered only for roughly 5% of the FADN farms. The overall distribution of the step length follows an inverse exponential function.

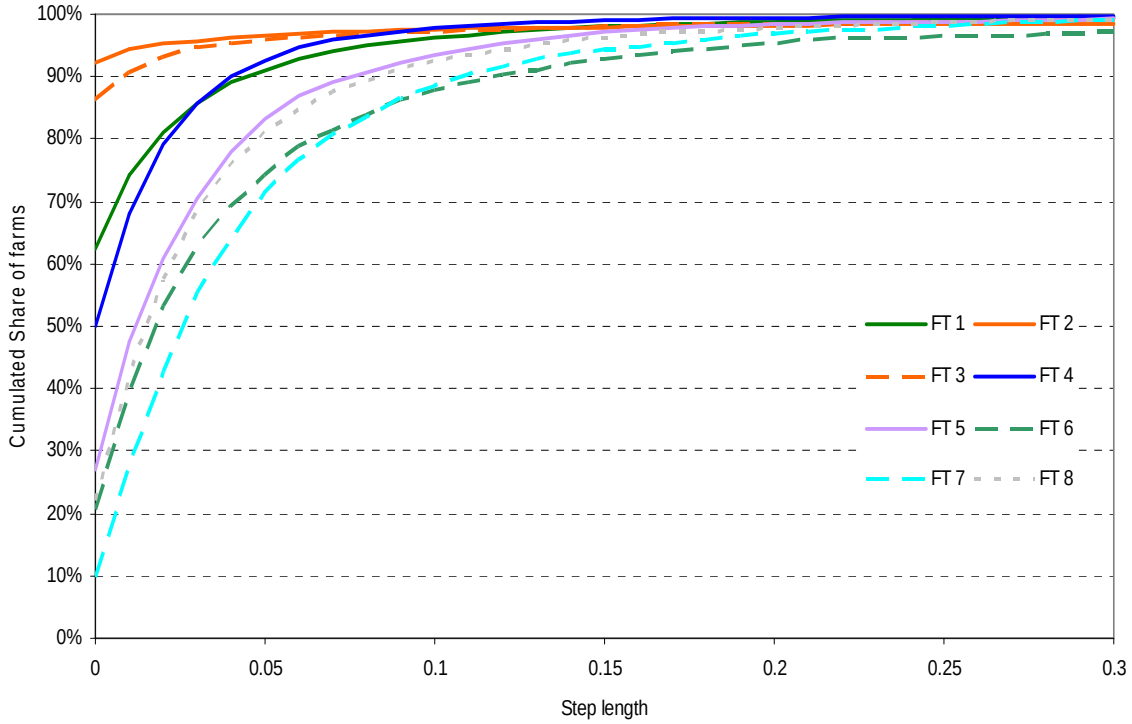
Figure 24: Distribution of cumulated share of farms as a function of the annual step length for the different years (measured by Manhattan Block distance)



Source: Own calculation based on the German FADN-farms in the period 1995-2007

Figure 25 indicates clear differences in the distribution of the inter-annual step length between the farm types. Mixed farm types and the specialized granivore farms are quite dynamic with respect to their specialization, however farms belonging to FT 2 or FT 3 hardly alter the shares of the different activities. These differences in the distribution of step length between the farm types are mirrored by the development of the step length over time (Figure 22). Comparing Figure 22 and Figure 25, it can be concluded that in particular the oscillations in Figure 22 regarding FT 2 are caused by some “outliers”.

Figure 25: Distribution of cumulated share of farms as a function of the annual step length for the different farm types (measured by Manhattan Block distance)

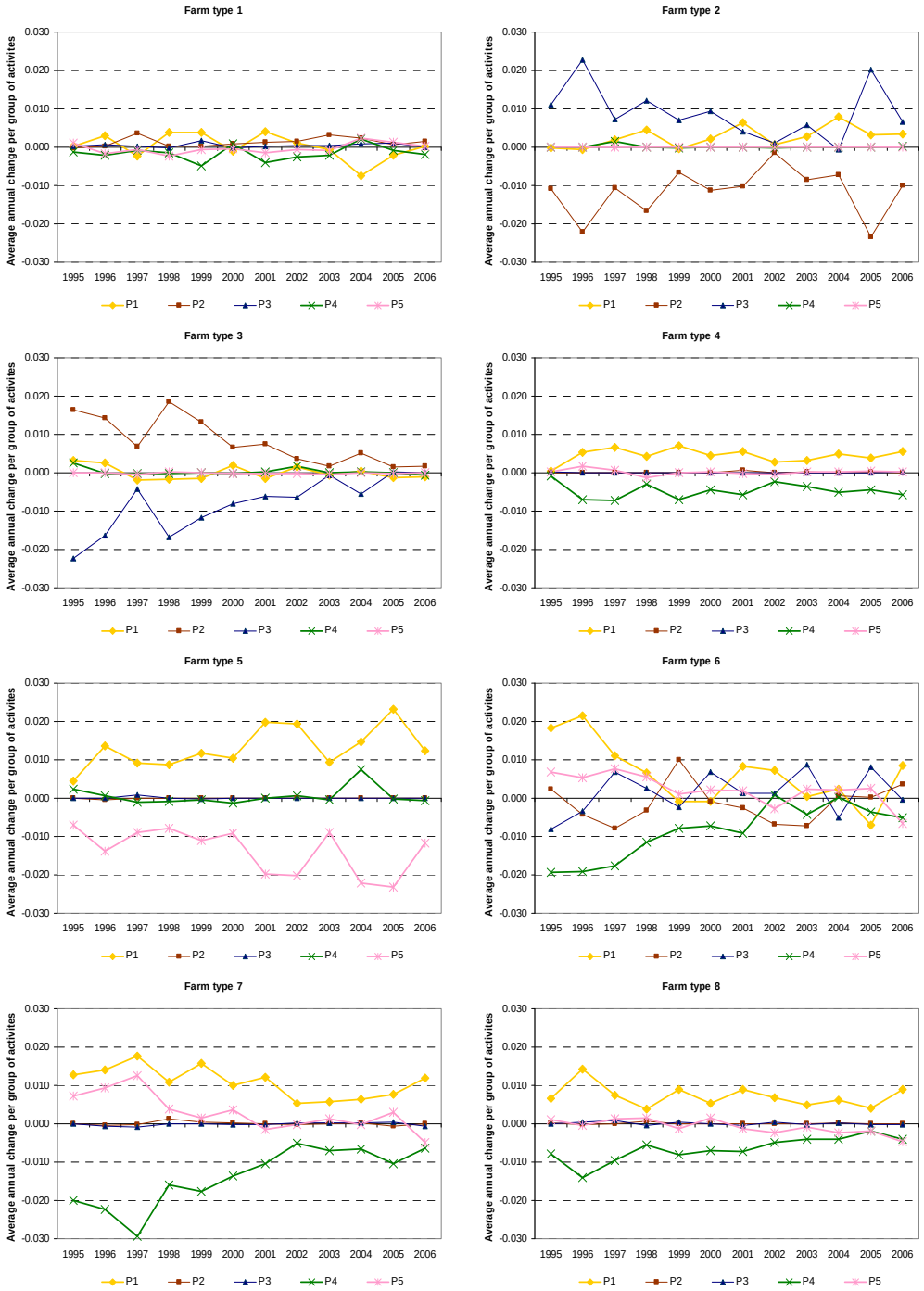


Source: Own calculation based on the German FADN-farms in the period 1995-2007

After briefly addressing the topic regarding the magnitude of change, we will now turn towards the direction of change.

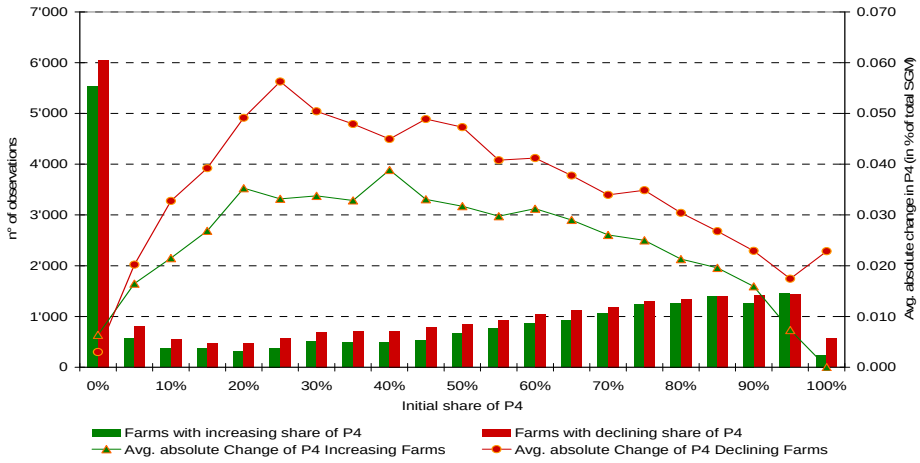
Figure 26 show the average change for the different specialisations over time. In all the years, there is an increase in arable cropping (P1) at the expense of the grazing livestock activities (P4) in FT 4, FT 7 and FT 8. For FT 4, there are no significant changes over time, while it is observed a shift in the dynamic for FT 7. In FT 1 the composition of the specialisation is fairly constant over time. For both FT 2 and FT 3 we can see an increasing diversification. While FT 2 substitutes P2 (horticulture) by P3 (permanent crops), the substitution process is the reverse for FT 3. In contrast to the other farm types, the development for FT 6 is characterized by large inter-annual fluctuations and the interpretation should be further analysed.

Figure 26: Development of the average annual change in the share of the P1-P5 activity groups for the eight general farm types.



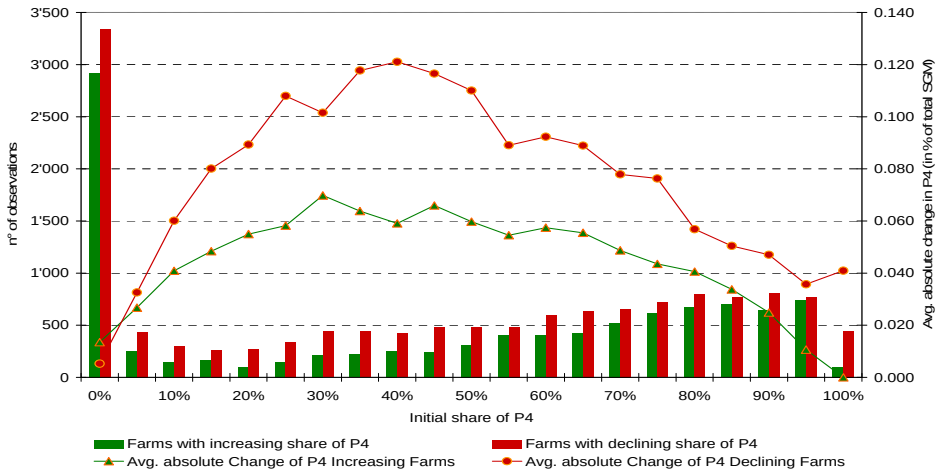
Source: Own calculation based on the German FADN-farms in the period 1995-2007

Figure 27: Number of farms and average development of farms regarding the share of P4 over 1 year in dependence of the initial share of P4 on the total SGM



Source: Own calculation based on the German FADN-farms in the period 1995-2007. Only farms that remained at least 6 years in the sample

Figure 28: Number of farms and average development of farms regarding the share of P4 over 4 year in dependence of the initial share of P4 on the total SGM



Source: Own calculation based on the German FADN-farms in the period 1995-2007. Only farms that remained at least 6 years in the sample

Figure 27 and Figure 28 depict the development of the share of P4 depending on the initial share of P4 on the total SGM. While Figure 27 depicts the development based on a one year interval, Figure 28 is based on rotating four year intervals (1995-1999, 1996-2000, etc.). Irrespective of the initial share of P4, the number of farms which reduce the share of P4 exceeds the number where P4 increases. In addition, the average change is larger in the shrinking farms. The share of P4 is especially reduced in the farms where the initial share of P4 lies between 15 and 70%.

Comparing Figure 27 and Figure 28, it can be noted that for the farms where P4 declines, the peak of the change is shifted to higher initial shares of P4 from roughly 25% in Figure 27 to 40% in Figure 28. This implies that the changes for the lower shares are of a more erratic movement as both graphs are based on the same subsample.

Annex 4

Ex-ante projection

Annex 4: Ex-ante projection

Table 29: Matrix d for dimension Columns and Rows in CAPRI for farm type during baseline estimation

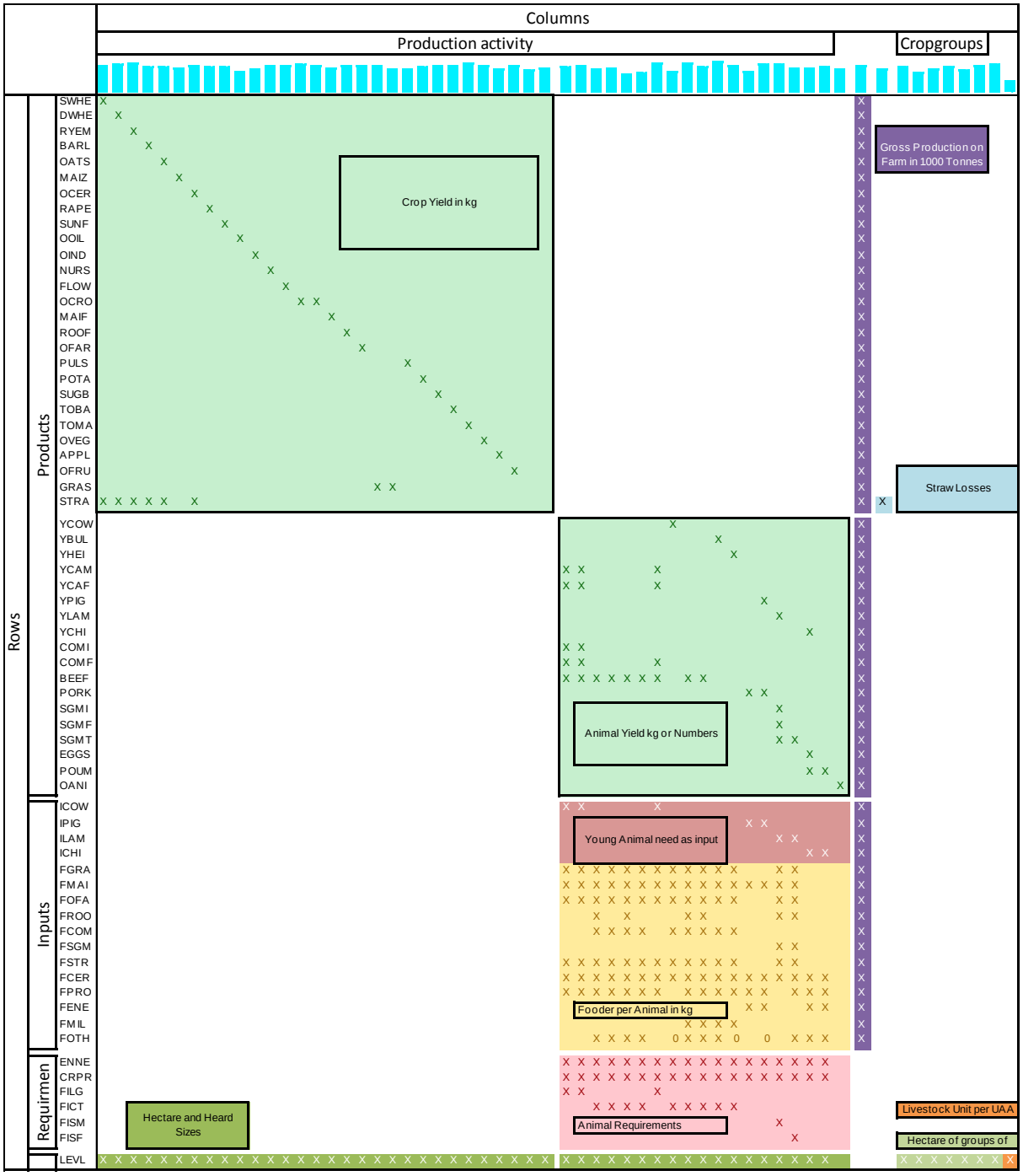


Table 30: Calculation of the partial SGM in CAPRI Farm Types

abbreviation	CAPRI activity long text	P1	P13_14	P2	P3	P4	P5
SWHE	Soft wheat production activity	■	■				
DWHE	Durum wheat production activity	■	■				
RYEM	Rye and meslin production activity	■	■				
BARL	Barley production activity	■	■				
OATS	Oats and summer cereal mixes without triticale	■	■				
MAIZ	Grain maize production activity	■	■				
OCER	Other cereals production activity including triticale	■	■				
RAPE	Rape production activity	■	■				
SUNF	Sunflower production activity	■	■				
SOYA	Soya production activity	■	■				
OOIL	Other seed production activities for oil industry	■	■				
OIND	Other industrial crops production activity	■					
NURS	Nurseries production activity				■		
FLOW	Flowers production activity			■			
OCRO	Other crops production activity	■					
MAIF	Fodder maize production activity	■					
ROOF	Fodder root crops production activity	■					
OFAR	Fodder other on arable land production activity	■					
GRAE	Gras and grazings production activity extensive					■	
GRAI	Gras and grazings production activity intensive					■	
PARI	Paddy rice production activity	■	■				
PULS	Pulses production activity	■	■				
POTA	Potatoes production activity	■					
SUGB	Sugar beet production activity	■					

abbreviation	CAPRI activity long text	P1	P13_14	P2	P3	P4	P5
TEXT	Flax and hemp production activity	■					
TOBA	Tobacco production activity	■					
TOMA	Tomatoes production activity			■			
OVEG	Other vegetables production activity			■			
APPL	Apples pears and peaches production activity				■		
OFRU	Other fruits production activity				■		
CITR	Citrus fruits production activity				■		
NONF	Non food production activities on set aside	■	■				
FALL	Fallow land						
OSET	Set aside obligatory	■	■				
VSET	Set asice voluntary	■	■				
BULL	Male adult fattening activity low final weight					■	
BULH	Male adult fattening activity high final weight					■	
SCOW	Suckler cows production activity					■	
HEIR	Heifers raising activity					■	
CAMF	Calves male fattening activity					■	
CAFF	Calves female fattening activity					■	
CAMR	Calves male raising activity					■	
CAFR	Calves female raising activity					■	
PIGF	Pig fattening activity						■
SOWS	Sows for piglet production						■
SHGM	Sheep and goats activity for milk production					■	
SHGF	Sheep and goats activity for fattening					■	
HENS	Laying hens production activity						■
POUF	Poultry fattening activity						■

Table 31: List of CAPRI files modified

In the course of this study, the following CAPRI files were created or modified:

gams/captrd.gms

gams/sets.gms

gams/captrd/data_structural_change.gms

gams/captrd/equations.gms

gams/captrd/equations_farm.gms

gams/captrd/est_nuts2_grid.gms

gams/captrd/estimate_farm.gms

gams/captrd/estimate_farm_start.gms

gams/captrd/farms_loop.gms

gams/captrd/save_results_captrd.gms

gams/captrd/sets_captrd.gms

gams/captrd/structural_change_contraints.gms

gams/captrd/structural_change_def_support_bound.gms

gams/captrd/structural_change_def_support1.gms

gams/farmtype/frmb_consis.gms
gams/farmtype/frmb_consis_new_sets.gms
gams/farmtype/frms_gm_model.gms

PART II: LITERATURE REVIEW ON THE STATE OF THE ART IN EX-POST AND EX-ANTE ANALYSIS OF STRUCTURAL CHANGE

Introduction

Agricultural sector models are typically restricted to endogenously model medium term decisions on production, variable input use and land allocation. More strategic, medium to long term investment decisions to enter or leave the business or to fundamentally change farm size, specialisation or production intensity of the farming system are typically not considered. However, these decisions occur regularly and are highly relevant for the overall impact of policies on the agricultural system and resulting farm structures have policy relevance by themselves. Therefore, modelling of farm structural adjustments in an ex-ante policy modelling exercises is considered highly desirable. As a first step towards establishing a module on farm structural change within an agricultural sector model this second part of the report reviews the literature with respect to ex-post estimation approaches able to measure and explain transitions between different farm typology classes but also looks into the few studies relevant for ex-ante analysis of structural change at the sectoral level. Furthermore, relevant determinants for the model specification shall be identified based on the results of existing empirical studies. The review is a modified and extended version of an earlier review by Zimmermann et al. (2009). It is modified to fit the objectives of the study and to take into account some methodological and empirical advances in the last two years.

The next section focuses on the determinants of structural change based on earlier literature reviews and draws some attention to theoretical considerations. In section 2, the main developments of Markov chain modelling are presented. Section 3 addresses the econometric farm growth, cohort and entry/exit models and introduces the multiplicative competitive interaction models not previously considered for structural change analysis but showing some potential. Section 4 goes into more recent advances in modelling transition probabilities before section 5 looks at ex-ante simulation studies. Section 6 summarizes key results and concludes.

1 Factors contributing to structural change in agriculture

Most studies on farm structure provide an enumeration of the factors assumed to determine structural change in agriculture. Here, a brief overview of these factors is given, leaving the in-depth discussion to others (see Reimund et al., 1977; Hallam, 1991; Hallam, 1993; Boehlje, 1992; Goddard et al., 1993; Harrington and Reinsel, 1995). Factors should not be seen as mutually exclusive but are rather interrelated, as several authors point out (see U.S. Congress, 1985; Van Dijk et al., 1986; Goddard et al., 1993; Harrington and Reinsel, 1995; Hallam, 1991; Boehlje, 1992). Here we present a non-exhaustive list of the main determinants of structural change derived from theory. Their empirical relevance will be the focus of the subsequent sections.

Technology. The technology model is based upon the concepts of economies of scale and the adoption and diffusion of technology. It refers to the concept of Cochrane's treadmill (Cochrane, 1958) and focuses on the impact of technological innovation reducing per unit costs of output at the farm level. The first adopters of the new technology will gain from the first-mover advantage as long as output prices remain largely unchanged. But as adoption spreads, prices of farm commodities will fall and competition increases by forcing others to adopt the new technology or to exit the industry, triggering structural adjustments (Harrington and Reinsel, 1995).

Off-farm employment is handled in two ways. On the one hand, it could be seen as a first step out of the sector. As opportunity costs increase due to better wage levels outside of agriculture, farmers tend to leave the sector until wages equalize (Hallam, 1991) or try to achieve comparable incomes by enlarging the farm business (Harrington and Reinsel, 1995). On the other hand, off-farm employment provides a method to keep on farming at small scales if the off-farm income complements the household income (Goddard et al., 1993; Gebremedhin and Christy, 1996) or farmers are even willing to subsidize their small farm at least in the short-run from other income sources (Harrington and Reinsel, 1995).

Policy. Structural changes in agriculture are also driven by the general institutional and legal environment as well as by specific public programs which impact the agricultural sector in different ways according to their design. Examples often mentioned apart from agricultural sector policies are tax policies, commodity programs, credit programs, general monetary and fiscal policies, and public research and extension efforts (Harrington and Reinsel, 1995; Goddard et al., 1993; U.S. Congress, 1985).

Human capital refers to and is influenced by the managerial capability, the level of schooling, and public education programs. It is assumed that an increase in human capital would allow the firm manager to more effectively process information used to allocate the firm's resources and to evaluate new technologies (Boehlje, 1992; Goddard et al., 1993).

Demographics refer mainly to the age structure of farm operators and the shrinking number of entrants to the farming sector. Although being a consequence rather than a cause of structural change, the age structure is believed to determine the speed of change in a region (Harrington and Reinsel, 1995). Reimund et al. (1977) and Goddard et al. (1993) also point to general changes in the demographical structure which impacts the agricultural sector through changes in the demand of agricultural products.

Market structure itself influences structural change. This point refers to the Structure-Conduct-Performance approach and is derived from the industrial organization literature (Van Dijk et al., 1986; Boehlje, 1992). The way in which prices are set is determined by the nature of the market, i.e. the degree of market power exercised on the supply or demand side, so that the conduct of the industry is a function of its structure. The development of institutional arrangements, such as vertical integration and cooperatives, has an (so far unclear) impact on structural change as well (Goddard et al., 1993).

Social setting. Sociological aspects and discussions of structural change in agriculture usually refer to the concept of the family farm (Peterson, 1986; Boehlje, 1992). The sociological model as described by Boehlje (1992) refers to the motivations to maintain a family farm-based agriculture. Boehlje distinguishes between aspects coming from society and the farmer's household. He argues that from the societal perspective the maintenance of a family farm-based agricultural structure is important to efficient production, community viability, and food supply. From the individual perspective the motivations are primarily related to the independent lifestyle, family bonding and relationships. In multigenerational family farm operations the objective is frequently identified as providing an opportunity for a future generation to farm.

Economic environment. Several sector specific and macroeconomic factors such as input and output prices, demand changes, and the interest rate are supposed to have an impact on structural change (Hallam, 1991; Goddard et al., 1993). However, most of the aforementioned points could also be expressed in economic terms, so that in fact the economic environment could be regarded as the heading subsuming the other factors.

2 Markov models

The estimation of Markov chains has a long tradition in the analysis of structural change in agriculture and is a widely accepted approach to predict the number of farms in certain farm types. The chapter is divided into four parts. Firstly, the general concept of the Markov chains is introduced, then stationary and non-stationary Markov chain studies in the farm structural change literature are discussed and finally the findings are summarized.

2.1 Concept

In a Markov chain the movement of firms from a specific firm category (e.g. a farm type) to another one is seen as a stochastic process which can be represented by transition probabilities. Usually, the movement of farms between several farm types is supposed to follow a first order Markov chain, i.e. it is assumed that the probability of the movement of a farm at time t to another farm type in the period $t+1$ is independent of earlier periods.

(34): $$n_{j(t)} = \sum_{i=1}^{N} n_{i(t-1)} p_{ij},$$

where the number of farms n in farm type j at time t depends on the number of farms in all farm types i in the period before $(t-1)$ multiplied by their respective transition probabilities p_{ij} to move from farm type i to farm type j in one time period. The probability constraints, non-negativity ($p_{ij} \geq 0$) and summing-up to unity ($\sum_{j=1}^{N} p_{ij} = 1$) must hold. The single transition probabilities can be collected in a transition probability matrix P $(N \times N)$:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}.$$

If micro-data is available, i.e. data from which the exact number of movements from one farm type to another can be derived, the elements in P can be estimated as

(35) $$\hat{p}_{ij} = m_{ij} / \sum_{j=1}^{N} m_{ij},$$

Where m_{ij} denotes the number of movements of firms from state i to state j during the time period under discussion and N is the total number of states. Anderson and Goodman (1957) have shown that the above given approximation of the true p_{ij} is, in fact, the maximum

likelihood estimate. If only macro-data, i.e. the number of farms per farm type and year is given, the Markov chain is usually estimated according to equation (34) by replacing the number of farms n by farm type shares y and adding an error term.

The estimated transition probabilities can be used to predict future farm numbers in any state:

(36) $$X_t = X_0 P^t,$$

where the row vector X_0 is the initial starting state vector or the initial configuration of individuals in the N states, where x_{0i} represents the number of individuals in state i during time period $t = 0$, and the row vector X_t is the tth configuration vector.

One of the strongest assumptions in this form of the Markov model is that the transition probabilities do not change over time, i.e. they are said to be stationary. This implies that the process of structural change follows the same path until an equilibrium solution is reached. Stationarity may represent a realistic assumption as long as all other factors remain constant, but it generally does not hold for economic phenomena. Changes in exogenous variables require the determination of non-stationary (time-varying) transition probabilities. In the case of micro-data availability non-stationary transition probabilities can be obtained by applying equation (35) on an annual base:

(37) $$\hat{p}_{ij(t)} = m_{ij(t)} / \sum_{j=1}^{N} m_{ij(t)} .$$

However, equation (37) cannot be used to detect which factors and to what extent these factors have actually influenced the structural process in question. Thus, an econometric model 'behind' the pure Markov chain is required. The non-stationary transition probabilities are, hence, specified as functions of (potentially lagged) exogenous variables and parameters and regressed against these in a second estimation step:

(38) $$p_{ij(t)} = f_{ij}(Z_{(t)}, \beta_{ij}),$$

where f_{ij} is the function of the vector of explanatory variables $Z_{(t)}$ and the matrix of parameters β_{ij} which relates the exogenous variables to the transition probabilities. In the case of macro-data, equation (38) can directly be substituted into equation (34) by changing the stationary p_{ij} to non-stationary $p_{ij(t)}$.

2.2 Stationary Markov chain models

Generally, the first Markov chain studies of the agricultural sector deal with micro-data used to estimate stationary transition probabilities via the maximum likelihood method following Anderson and Goodman (1957). Publications which refer to this type of Markov models are Judge and Swanson (1961), Padberg (1962), Stanton and Kettunen (1967), Edwards et al. (1985), and Garcia et al. (1987). Krenz (1964) is the first who estimated a stationary Markov model from macro-data. However, in order to do so he simply applied the micro-data maximum likelihood estimator (35) and replaced the single farm movements by farm type shares calculated from the aggregated data. Additionally, a number of constraints had to be imposed to ensure meaningful results. Stavins and Stanton (1980) point to the theoretical limitations of this approach since the behavioural pattern for the farms that should be investigated is, in fact, already postulated beforehand. Also, Lee et al. (1977) and MacRae (1977) have shown that the maximum likelihood function in case of macro-data is in fact rather complex, such that the approach chosen by Krenz suffers from a weak econometric foundation as well. Nonetheless, a similar approach was used later on by Keane (1976), Keane (1991) and Tonini and Jongeneel (2002) and the imposition of constraints on the transition probabilities became rather popular among applied Markov studies. Recent applications of stationary Markov chain models are Jongeneel and Tonini (2008) and Piet (2008). Jongeneel and Tonini introduce mobility indices based on Shorrocks (1978) to the Markov chain literature. The mobility indices give information on mobility level and direction of farms between the different farm types. Piet presents a continuous version of the Markov chain model which gives more insight in the full size distribution of farms and allows the reconstruction of transition probabilities for any desirable size class. The stationary Markov chain models are summarized in the A.

2.3 Non-stationary Markov chain models

The non-stationary Markov chain applications in the agricultural economics literature are split into two-step approaches and approaches estimating the Markov chain and the influence of exogenous variables simultaneously. An overview of the non-stationary Markov chain applications is provided in Table 32.

2.3.1 Two-step approaches

Hallberg (1969) was the first, who calculated non-stationary transition probabilities in order to predict structural change depending on exogenous variables. The first estimation step follows equation (37) and for the second estimation step a restricted least squares procedure is applied. However, the least-squares approach suffers from the fact that it is not possible to ensure the probability constraints when making predictions with the

estimated coefficients. Other micro-data models applying two-step procedures are Salkin et al. (1976), Stavins and Stanton (1980), Ethridge et al. (1985), and Rahelizatovo and Gillespie (1999).

Among other modelling exercises, Stavins and Stanton (1980) represent the transition probabilities as multinomial logit functions of explanatory variables and coefficients in the second estimation step. The multinomial logit formulation has the advantage that the probabilities automatically sum to unity and are positive. An ordinary least squares estimator is used to estimate a linearised version of the model.

Rahelizatovo and Gillespie (1999) are the first to conduct a cross-regional analysis where the regional dummy variable reveals a significant influence on most transition probabilities. Other factors significantly affecting structural change among dairy farms in Louisiana are found to be input and output prices, technology expressed as productivity, financial conditions, and agricultural policies that have provided incentives for early retirement and reduction in milk production. Decreasing milk prices are predicted to increase the number of farms quitting the sector. With regard to policy plans to decrease dairy waste disposal into the Tangipahoa River, Rahelizatovo and Gillespie also discuss environmental concerns of the predicted structural change towards larger farm entities. In fact, they predict (without having implemented the relevant policy change in their model) that some producers might discontinue production facing increased investments into waste disposal facilities.

Stokes (2006) employs a generalised cross-entropy estimator (GCE) based on Lee and Judge (1996) and Golan et al. (1996) to estimate time-varying transition probabilities from macro-data. Afterwards the influence of other explanatory variables is analysed by regressing the most interesting transition probabilities against these variables linearly. The prior transition probability matrix for the GCE approach is obtained by firstly estimating a stationary transition probability matrix with a uniform prior. The model is applied to Pennsylvanian dairy farms. Stokes finds that milk prices, price volatility, land values, and the dairy termination program strongly impact the probability for exit from dairying in Pennsylvania. Dairy farm size growth is found to be inhibited by milk price volatility and land values, but responds positively to higher milk prices. Growth and contraction are also positively related to productivity. Concerning the transition probabilities Stokes follows that if the status quo is maintained, there will be fewer, larger dairy farms, with the rate of decline estimated to be about 2.0 percent to 2.5 percent annually over the next two decades. Another contribution of Stokes to the Markov chain literature is the presentation of the linkage between an analytical model of the firm and the Markov chain model saying that as long as the farmers' decisions are consistent with a dynamic planning horizon and the uncertainty faced is Markovian, the size of the firm will also be Markovian as the properties from the underlying sources of uncertainty are inherited through the optimization process.

2.3.2 Simultaneous estimation of Markov chain and exogenous influence

Based on Telser (1963), Disney et al. (1988) are the first in the field of agricultural economics who estimate non-stationary transition probabilities from macro-data. In their study on the hog production industry in southern states of the USA they find that both total farm numbers and the size distribution of pork farms are highly sensitive to different hog-corn price ratio scenarios. A methodologically similar approach was later used by Von Massow et al. (1992).

Chavas and Magand (1988) develop an approach to estimate the probability of net entry and the transition probabilities of the remaining firms separately. Equation (34) is therefore redefined as:

(39) $$n_{j(t)} = a_{j(t)} + \sum_{i=1}^{N} n_{i(t-1)} p_{ij(t)},$$

with $a_{j(t)}$ representing net new entries. As explanatory variables for the transition probabilities pertaining to continuing farms economies of size, sunk costs and market prices are chosen. The vector $a_{j(t)}$ is specified as a function of the same variables with slight adaptations in the variable definition ($a_{j(t)} = f(Z_{j(t-1)}, \alpha_j)$). The transition probabilities are estimated within a multinomial logit framework considering four size classes.

Zepeda (1995a) takes up the approach of Chavas and Magand (1988) and models the probability of net new entry separately from the transition probabilities of the existing firms. In her model of Wisconsin dairy farms the milk-feed price ratio is assumed to affect both net new entries and state transitions. The interest rate (to reflect the cost of capital), a dummy policy variable (farmers are paid to exit the sector), the amount of debt and a dummy variable for drought are supposed to influence only net new entries. Zepeda concludes from her analysis that farmers respond symmetrically to price changes when entering or quitting dairy farming, but they are more responsive to price decreases than price increases when changing the herd size. It is also found that under none of the calculated price scenarios any small- or medium-sized farms would exist in the long run. A similar approach has also been applied to hog production firms in the United States in Gillespie and Fulton (2001).

In a second application, Zepeda investigates the influence of technical change on the size distribution of dairy farms (Zepeda, 1995b). The model is applied to four size classes only, without considering entries or exits. As proxy for technical change the milk production per cow and per year is used. Steady state probabilities and elasticities measuring the effect of the explanatory variables on the transition probabilities referred to as ‘probability elasticities’ are calculated. Zepeda finds that increases in the level of

technology among continuing dairy farms enhance their ability to stay the same size versus growing in the short run, but in the long run increase the proportion of very large farms.

Karantininis (2002) is the first who applies a generalised cross-entropy (GCE) formalism for a Markov chain estimation of the agricultural sector (Lee and Judge (1996); Golan et al. (1996)). His study focuses on the farm size distribution of Danish hog producers. Although a non-stationary cross-entropy formulation according to equation (34) with the probabilities being substituted by equation (38) is shown in his article, Karantininis applied an instrumental variables techniques (IV-GCE) developed by Golan and Vogel (2000) in order to determine the impact of exogenous factors on structural change. The IV-GCE procedure is much simpler to apply, but does not allow the estimation of different transition probability matrices for each point in time as possible in traditional non-stationary Markov studies. Nonetheless, the IV-GCE approach is mostly referred to as 'non-stationary', which is thought to reflect the fact that explanatory variables are considered and their impact on the transition probabilities can be measured by elasticities. Using the uniform distribution as prior for the transition probabilities, Karantininis firstly estimates a stationary Markov model which is found to perform rather badly. Information gained from a pre-estimated non-stationary model with a rather simple matrix of prior transition probabilities is introduced as prior information in the main estimation. This second non-stationary model reveals the best overall performance as measured by the pseudo-R^2 of the three Markov models. Karantininis uses pork prices, pork feed prices and input and output prices of other livestock as explanatory variables. Most of the elasticities for pig prices are found to be positive in most of the upper off-diagonals and negative in most of the lower off-diagonal elements meaning that increases in pig prices reduce the probability of firms downsizing, and increase the probability of them increasing in size. Non-stationary Markov chain studies applying IV-GCE estimators according to Karantininis (2002) can also be found in Jongeneel et al. (2005), Tonini (2007), Tonini and Jongeneel (2008), and Huettel and Jongeneel (2011).

Jongeneel (2002) analyses farm structure changes of Dutch dairy farms with a GCE estimator. Unlike Karantininis (2002), Jongeneel estimates time-varying transition probabilities which are simultaneously represented as linear functions of exogenous variables and coefficients.

2.4 Summary

Markov chain applications to the agricultural sector advanced from stationary micro-data approaches in the early studies to non-stationary macro-data models related to exogenous factors via two-step or simultaneous estimation procedures. Accordingly, the estimation techniques applied changed from maximum likelihood over linear model specifications to the representation of the transition probabilities as multinomial logit functions. Recently,

cross-entropy techniques making use of a priori information given by the researcher and tackling the problem of ill-posedness became popular in Markov chain estimations.

Concerning regional and farm type coverage, with few exemptions only a single region and production orientation (mostly dairy or pig farms) are considered in the Markov chain studies analysed. The maximum number of farm types for which transition probabilities have been estimated is 19 (18 size classes and the artificial entry/exit class; Karantininis, 2002).

As far as estimation results are concerned, most of the more recent studies predict further farm number decreases of small to medium sized farms, whereas the number of large farming entities is mainly predicted to increase. The explanatory variables used in the non-stationary Markov studies relate to the factors contributing to structural change outlined in section 0.2.4. Most often variables concerning technological change, economic factors like prices and interest rates, and policy variables have been taken into account, whereas human capital or demographical aspects did not appear in any study as explanatory variables. Only Zepeda (1995a) introduces a 'new' variable, namely drought, in her analysis.

With regard to the explanatory power of the Markov chains, most authors who conducted stationary as well as non-stationary analyses found that the non-stationary models performed much better in predicting the farm type distribution than the stationary ones (e.g. Hallberg, 1969, Stavins and Stanton, 1980, Von Massow et al., 1992, Karantininis, 2002). The R^2 values tend to attest the models a rather high explanatory power. Where low R^2 values are reported, these mainly refer to single transition probabilities or are attributed to the estimation technique applied (Salkin et al., 1976; Stavins and Stanton, 1980). A number of studies conducting within sample predictions found a good prediction accuracy of the models applied (Hallberg, 1969; Garcia et al. (1987); Zepeda, 1995a and 1995b; Tonini and Jongeneel, 2002). An exemption is Von Massow et al. (1992) who found very high prediction errors in the 'no production' class. Out-of-sample predictions conducted by Hallberg (1969) for the stationary model revealed a rather poor fit when compared to the actual values. Stavins and Stanton (1980) found that the out-of-sample predicted distribution showed approximately the correct shape if the multinomial logit model specification was applied to estimate the transition probabilities. In many cross-entropy approaches the incorporated prior information is found to considerably affect the overall quality of the model as indicated by the fact that the final estimates closely follow the prior information matrix.

Table 32: Non-stationary Markov studies in the agricultural economics literature

Year	Author	Region	Specialisation	Data type	Time Series	Transition Probabilities	Methodology	Number of States	Dependent Variable	Explanatory Variables	Performance
1969	Hallberg	Pennsylvania, USA	Frozen milk products plants	Micro	1944-1963	Stationary, non-stationary	Maximum likelihood + least squares (2-step)	4 +entry/exit	Firm size (in sales volume)	Wages, population, per capita income, farm-gate price for milk, retail price	R^2: 0.89-0.99
1976	Salkin et al.	Oklahoma, USA	Cotton warehouses	Micro	1964-1973	Stationary, non-stationary	Least squares + geometric model (2-step)	5 + entry/exit	Firm size (in warehouse capacity)	Time	R^2: 0.002-1.0 (linear model), 0.47-1.0 (geometric model)
1980	Stavins and Stanton	New York, USA	Dairy farms	Micro (stationary, 2-step non-stationary), macro (stationary)	1968-1977	Stationary, non-stationary	Maximum likelihood + multinomial logit (2-step)	9 + entry/exit	Firm size (in milk supply)	Milk-feed price ratio	R^2: 0.00-0.70
1985	Ethridge et al.	West Texas, USA	Cotton gin firms	Micro	1967-1979	Stationary, non-stationary	Maximum likelihood + least squares (2-step)	12 (including new entrants, dead gin firms and 5 size classes of inactive and active farms, respectively)	Activity and size (in gin capacity)	Wages, energy costs, plant capacity, technical change	R^2: 0.32-0.72
1988	Disney et al.	Southern states, USA	Pig farms	Macro	1969-1982	Stationary, non-stationary	Minimum absolute deviation	4 (+ entry/exit)	Firm size (in saled market hogs/year)	Hog-corn price ratio	R^2: 0.94-0.97
1988	Chavas and Magand	Different regions, USA	Dairy farms	Macro	1977-1984	Non-stationary	Multinomial logit	4	Net entry; firm size (in herd size)	Economies of size, sunk costs, market prices	R^2: 0.67-0.99

Year	Author	Region	Specialisation	Data type	Time Series	Transition Probabilities	Methodology	Number of States	Dependent Variable	Explanatory Variables	Performance
1992	Von Massow et al.	Ontario, Canada	Pig farms	Macro	1971-1989	Stationary, non-stationary	Minimization of median absolute deviation	5 + entry/exit	Firm size (in number of hogs marketed)	Hog-corn price ratio, interest rate, labour-capital price ratio	Within sample prediction (root mean square error): Stationary 11-33%, 63% (entry/exit); non-stationary 9-20%, 46-62% (entry/exit)
1995a	Zepeda	Wisconsin, USA	Dairy farms	Macro	1972-1992	Non-stationary	Multinomial logit	3	Entry/exit; firm size (in herd size)	Milk-feed price ratio, interest rate, dairy ter-mination program, debt, drought	R^2: 0.9905-0.9986, within sample prediction (error in any year): 2.2-7.2%
1995b	Zepeda	Wisconsin, USA	Dairy farms	Macro	1980-1992	Non-stationary	Multinomial logit	4	Firm size (in herd size)	Milk pro-duction per cow (proxy for technical change)	R^2: 0.88-0.99, within sample prediction (error in any year): 2-11% of farms
1999	Rahelizatovo and Gillespie	Louisiana, USA	Dairy farms	Micro	1981-1995	Non-stationary	Maximum likelihood + SUR (2-step)	4 + entry/exit	Firm size (in productivity)	Milk and feed prices, milk pro-duction per cow, interest rate, debt/-equity ratio, policy dummies, regional dummies	R^2: 0.63-0.80
2001	Gillespie and Fulton	17 US states	Pig farms	Macro	1988-1997	Non-stationary	Multinomial logit	3	Firm size (in number of hogs)	Regional dummies, hog-corn price ratio, interest rate, corporate farm laws, meat pro-cessing ca-pacity, per-centage of land in farms	R^2: 0.75-0.97

Year	Author	Region	Specialisation	Data type	Time Series	Transition Probabilities	Methodology	Number of States	Dependent Variable	Explanatory Variables	Performance
2002	Karantininis	Denmark	Pig farms	Macro	1984-1998	Stationary, non-stationary	GCE, IV-GCE	18 + entry/exit	Firm size (in number of hogs)	Input and output prices of pork and other livestock, fertilizer prices, interest rate	Pseudo-R^2: 0.07, 0.26, 0.49
2002	Jongeneel	Netherlands	Dairy farms	Macro	1972-1999	Non-stationary	GCE + linear explanation function	3 + entry/exit	Firm size (in herd size)	Milk output, milk price, policy dummy, trend	-
2005	Jongeneel et al.	Netherlands, Germany, Poland, Hungary	Dairy farms	Macro	NL: 1972-2003, W-DE: 1971-2003, E-DE: 1991-2003, PL: 1996-2000, H: 2000/2003	Non-stationary	IV-GCE	NL, E-DE, H: 7 + entry/exit, W-DE: 6 + entry/exit, PL: 4 + entry/exit	Firm size (in herd size)	Trend, milk output, milk price, quota dummy, auction dummy	Pseudo-R^2: NL: 0.84, W-DE: 0.92, E-DE: 0.89, PL: 0.93, H: 0.82
2006	Stokes	Pennsylvania, USA	Dairy farms	Macro	1980-2003	Non-stationary	GCE + SUR (2-step)	6 + entry/exit	Firm size (in herd size)	Milk price, milk price volatility, productivity, interest rates, land values, policy dummy for exit probabilities	-
2007	Tonini	Poland, Hungary	Dairy farms	Macro	PL: 1995-2005, H: 2000-2003	Stationary, non-stationary (PL), stationary (H)	IV-GCE	PL: 8 + entry/exit, H: 7 + entry/exit	Firm size (in herd size)	Trend, milk producer price, price for concentrates for cattle	Pseudo-R^2: PL: 0.048/0.051, H: 0.000
2008	Tonini and Jongeneel	Poland	Dairy farms	Macro	1995-2006	Non-stationary	IV-GCE	8 + entry/exit	Firm size (in herd size)	Trend	Pseudo-R^2: 0.34
2008	Huettel and Jongeneel	Germany, Netherlands	Dairy farms	Macro	W-DE: 1971-2005, E-DE: 1991-2005, NL: 1972-2006	Non-stationary	IV-GCE	W-DE: 6 + entry/exit, E-DE and NL: 7 + entry/exit	Firm size (in herd size)	Milk price, milk yield, policy dummy	Pseudo-R^2: W-DE: 0.82/0.80, E-DE: 0.58, NL: 0.82/0.90

3 Other econometric models

There exists a vast amount of econometric models apart from Markov chains that deal with structural change in agriculture. These models are characterised by regressions on a number of explanatory variables. The regression analyses can thematically be divided into three model variants. Most of the regression models are related to analysing farm growth, specifically testing Gibrat's law, others are cohort analyses which concern the number of farm holders and the reasons for entering or leaving the sector and the last selected variant of models considers farm succession explicitly. Applications of the model types are summarized in Table 33.

3.1 Farm growth

Most of the models reviewed in this section try to explain farm growth or size or focus especially on entry and exit of farms to or from the sector. Many of the studies on growth and size distribution of farms rely on a simple stochastic model which is usually a variant of Gibrat's law (Gibrat, 1931). Gibrat's law states that the growth rate of firms is determined by random factors and independent of firm size. The basic equation to test Gibrat's law is:

(40): $$\ln S_{i(t)} - \ln S_{i(t-1)} = \alpha + \beta \ln S_{i(t-1)} + u_{i(t)},$$

where $S_{i(t)}$ is the size of firm i at time t, and $u_{i(t)}$ is the random effect. Gibrat's law is true if $\beta = 0$ (Weiss, 1999). The main weakness of the law is that systematic factors that are of primary interest from a social science perspective are comprised under the random process. Therefore, the equation given above is often extended to take into account other factors than size as well and on the left hand side of equation (40) it is common to include also farm entry and exit (farm survival).

Shapiro et al. (1987) test the relationship between farm size and growth in Canada from 1966 until 1981. They find out that small farms grow faster than large farms implying the rejection of Gibrat's law. Larger farms also experience more stable growth rates in comparison to small farms. Shapiro et al. also find that the probability of exit is greater than the probability of entry at any size, and that the probability of either of them is highest for small farms.

Weiss (1999) takes into account the two interrelated determinants 'entry/exit' and 'firm growth' of continuing farms. He adds a number of other socioeconomic factors to the elementary stochastic model of Gibrat's law in his analysis on Upper Austrian farm households from 1980 to 1990. Factors assumed to have an impact on farm growth and

survival are human capital, off-farm employment and other individual and farm-specific characteristics. Weiss splits up his estimation into the branches full-time and part-time farming, but analyses also all farms together. He finds that a large proportion of the variance in the data cannot be explained with the specified econometric model and suggests other important determinants which may have an influence on the unexplained variation (e.g. farm income, farmer's attitude towards risk, etc.). The estimated negative relationship between part-time farming and farm expansion/survival supports the assumption that part-time farming promotes the restructuring of the farm sector. He further finds that the effect of farmer's age on the probability of survival is positive for young farmers and becomes negative for farmers over 51. Moreover, the existence of a farm successor has a positive impact on farm survival. With regard to human capital, agricultural specific schooling and general schooling are examined. An increase in agricultural specific schooling increases the probability of farm survival and farm growth. General schooling has a positive impact on farm survival, but the effect on farm growth is seen to be insignificant.

Weiss furthermore includes aspects concerning the family status of the farmer and derives interesting insights. If the farm operator is married, this has a positive impact on survival and growth of the firm. Also, an increase in the number of family members increases farm survival and growth. If the operator is female, this has a negative impact on farm survival and farm growth. Generally, the effect of all these factors seems to be higher for full-time farms. Gibrat's law is rejected since farm growth is less than proportionate to farm size. As Shapiro et al., Weiss estimates that smaller farms grow faster than larger farms. He determines two "centres of attraction" which suggest a polarisation of growth rates: small and very large farms grow faster than farms in the medium size class.

Bremmer et al. (2004) analyse the structural change in arable farming and horticulture in the Netherlands with regard to farm renewal and farm growth. Renewal covers all changes at the firm requiring the application of new knowledge and includes diversification and innovation. Explanatory variables have been selected in order to reflect personal characteristics of the farm operator, firm structure, and firm performance. The farm operator is characterised by age, time horizon (long if successor exists or age below 50, short otherwise), labour input of family members, off-farm income and education. Firm structure is reflected by the variables soil type, location, farm size, solvency and mechanisation. Profitability is the only variable in the category performance. Personal characteristics are shown to have a weak impact on farm growth. Thus, age, succession, and off-farm income have no influence, and family labour input is negatively correlated with farm growth. Firm development (profitability) is correlated with neither firm growth nor renewal. The results show that firm structure has a larger impact on firm development than personal characteristics and performance. The degree of mechanization has the largest marginal impact on both farm growth and renewal, since a high degree of mechanization implies high investments in the past, encouraging firm renewal and firm growth. Firm growth is found to be independent of firm size. However, the authors

conclude that the present models do not provide a satisfactory explanation for firm growth and renewal. In general, a large proportion of no-changes is predicted correctly, whereas the occurrence of growth and renewal is predicted incorrectly. According to the authors this might be due to data limitations as most firms provided only five or six observations and firm growth and renewal took place in a limited number of years. For further research they suggest to include the decision making process in the model. Separate estimation of the model for arable farming and protected horticulture shows that firm size has a positive impact on firm growth in arable and a negative impact in horticultural farming.

Sumner and Leiby (1987) analyse effects of human capital on size and growth. Their study employs a sample of southern dairy farms in the United States. Variables included are age (supposed to reflect general experience, life-cycle, and cohort effects), experience (measures the tenure of the farm operator, where, for a given age, more dairy experience means less general experience), schooling (representative for general human capital), and management (as an indicator of dairy-specific information or techniques). Cohort analyses are conducted for age, experience, and schooling cohorts. From the econometric analysis the authors conclude that the considered variables indeed may affect farm size and growth. However, the effects remain unclear and further work in this field is suggested.

3.2 Number of farm holders

Farmers of a certain gender and occupational category (full-time, part-time, hired, family) belonging to a cohort, i.e. group, are defined by specifying the period during which they were born. Their number can be followed and simulated through time by cohort analyses (De Haen and Von Braun, 1977). This method depends on population dynamics and the life cycle of farmers. Projections are made by assuming that historical patterns of changes in the number of farmers by age cohort will continue into the future (Olson and Stanton, 1993). The basic equation for an age cohort analysis is:

(41): $$H_{a+1}(t+1) = H_a(t)\, ps_{a,a+1}\, pe_{a,a+1} - NA_{a,a+1}(t, t+n)\,,$$

where $H_a(t)$ is the number of holders in the cohort of age a at time t, $ps_{a,a+1}$ is the probability to survive during age interval a to $a+1$, $pe_{a,a+1}$ is the probability to maintain the earning capacity during age interval a to $a+1$, and NA is the non-autonomous change of the cohort size. Age cohort analyses in agriculture are usually used to predict labour developments (De Haen and Von Braun, 1977). With a cohort analysis the autonomous changes in the farm structure can be separated from non-autonomous changes. Autonomous events are demographic factors such as ageing, death, disability, and retirement through ageing. Non-autonomous changes are those changes that can be attributed to all other factors (e.g. new entrants, change of occupation, early retirement).

They are usually interpreted as arising from changes in social and economic circumstances. The autonomous component of the decrease in the number of farmers in a specific age cohort can be inferred from general population statistics. The residuals (the non-autonomous change) that follow from the cohort analysis are then explained using econometric methods which may include several explanatory variables that were already outlined in the previous sections. De Haen and Von Braun (1977) predicted that for the work force decrease in West Germany a considerable part (about 60 %) are due to age, death, and disability.

EU-wide age cohort analyses have been carried out within the SEAMLESS project by Garvey (2006). Garvey (2006) finds that for the explanation of the non-demographic part, i.e. the non-autonomous change of the number of farm holders, only the regional unemployment rate appears to be useful. His analysis shows that a percentage point increase in regional unemployment generally leads to a 1.5 percentage point increase in net-entry to the farming sector among young farmers. For farmers between 35 and 55 years a 0.8 percent increase of net-entry is found in case of a one percentage point increase of the unemployment rate. In general, net-entry among young farmers appears to be more sensitive to regional unemployment changes than entry or exit for more middle-aged farmers.

The age cohort approach could theoretically be used in analyses of structural change to approximate the number of farms in a region. However, this approach makes sense for regions in which one farm corresponds to one farm holder (family farm structure). For regions where this is not the case, e.g. in Eastern Europe, the age cohort approach is not suitable. Furthermore, the methodology is not suitable for modelling aggregate change of farm numbers in specialisation and size classes as the underlying decisions are mainly determined by other factors than age structure.

3.3 Farm succession

In the context of farming systems, models of discrete choice have mainly been used to explain switches from conventional to organic farming (a literature review is provided by Acs et al., 2005). However, there exist a number of studies that concentrate on the estimation of farm survival by analysing the probability of farm succession. These studies are normally formulated as problems of discrete choice where the model generally includes characteristics of the individual (e.g. age, number and age of children) and relative attributes of competing choices (e.g. expected utility). Examples are the studies by Kimhi and Nachlieli (2001) and Pietola et al. (2003). Generically, we can represent a discrete choice model according to the following formula (Pietola and Heikkilä, 2006):

(42):
$$y_i^* = \alpha + \beta z_i + u_i$$
$$where\ \ y_i = 1\ if\ \ y_i^* > 0,\ else\ y_i = 0\ .$$

y_i* is a latent response variable defined in practice and unobservable. What we observe is the dummy variable y_i representing a certain choice. From the previous relations the choice probability relation and the likelihood function can be derived.

Kimhi and Nachlieli (2001) estimated a binary choice model for Israeli farms in which a variable w_t is defined as the tendency to declare a successor in period t. The model was estimated via probit and SNP (semi-nonparametric) methods. The age of the farm owner, an education dummy, off-farm employment, the age difference between farm owner and eldest child, the number of daughters and sons, a regional dummy, farm size, a production dummy, and a dummy for an already existing (declared) successor served as explanatory variables. Four different R^2-based measures revealed values between 50 and 80 per cent. Kimhi and Nachlieli (2001) found that the probability of having a successor rises with the age of the operator (up to age 68), his/her level of schooling, and age of the oldest child. The number of children and the parents' off-farm employment did not have a significant influence on the probability of succession. Also, succession probabilities were found to be much higher in farms located in Northern regions of the country and fruit or vegetable farms have higher probabilities for succession than farms with more land and/or poultry enterprises.

Pietola et al. (2003) analysed the timing and type of exit from farming in relation to early retirement programmes in Finland. Three choice alternatives were assumed: exit and close down of the farm operation, exit and transfer of the farm to a new entrant, or the continuation of farming. These three alternatives are mutually exclusive such that two binary indicators (exit and transfer) were used to identify them, whereas the third choice of continuation was observed if neither exit nor transfer occurred. McFadden's R^2 was 0.68 and 0.65 for two estimated models (a model which controls for serial correlation by simulating the sequence of interrelated choice probabilities using the Geweke-Hajivassiliou-Keane (GHK) simulation technique and multinomial probit, respectively). Explanatory variables were the farmer's age, a regional dummy, land and forest area, output prices, subsidy rates, the level of saved pension, a dummy which indicates the expiration of an early retirement programme, and a dummy for the presence of a spouse. However, some parameters associated with prices and subsidies were not significant at the five percent level. The results of the study suggest that the timing and type of exit decision respond elastically to farmer and farm characteristics and the political and economic environment. More specifically it is predicted that an increase of the minimum age of eligibility for early retirement will first slow down structural development, since farmers cannot exit as early as before. However, as the exit decision is delayed and the farmers' age increases, the probability of transferring the farm to a new entrant will decrease. This result is in line with Kimhi and Nachlieli (2001), who predict a decreasing probability for farm succession for farms with farm holders being older than 68 years as well.

Table 33: Overview of other (than Markov) econometric models

Analysis	Year	Region	Focus	Time period	Dependent	Explanatory	Performance
Farm growth							
Bremmer et al.	2004	Netherlands	Arable farming and horticulture	1990-2000	Farm renewal/ farm growth	Farmer's age, succession, off-farm employment, firm size, family labour input, solvency, mechanisation, profitability	R^2: 0.36/0.30 (both); 0.28/0.33 (arable); 0.78/0.32 (horticulture)
Shapiro et al.	1987	Canada	All specialisations	1966-1981	Firm size; entry/exit	Firm size	R^2: 0.10-0.80
Sumner and Leiby	1987	Southern USA	Dairy farms	1982, 1977, 1987	Firm size and growth	Farmer's age, experience, schooling, management	
Weiss	1999	Upper Austria	All specialisations	1980-1990	Entry, exit, firm growth	Farm size, human capital, off-farm employment, farmer's age, farmer's family status	
Number of holders (age cohort analyses)							
De Haen and Von Braun	1977	West German regions	All specialisations	1965-1975	Number of holders	Autonomous events, non-autonomous events	
Garvey	2006	EU15	All specialisations	1995-2000	Number of holders	Autonomous events, non-autonomous events	R^2: 0.47-0.63
Farm succession (discrete choice analyses)							
Kimhi and Nachlieli	2001	Israel	All specialisations	1994-1995	Tendency to declare a successor in period t	Farmer's age, education, off-farm employment, age difference between holder and eldest child, number of children, regional dummy, farm size, production dummy, dummy for declared successor	R^2: 0.49-0.83
Pietola et al.	2003	Finland	All specialisations	1993-1998	Exit and close down; exit and transfer to new entrant; continuation	Farmer's age, regional dummy, land and forest area, output prices, subsidy rates, saved pension, expiry of early retirement programme, marital status	R^2 (McFadden): 0.65-0.68

3.4 Multiplicative competitive interaction models

Multiplicative Competitive Interaction (MCI) Models are not yet applied for analyzing farm structural change, but constitute a tool which is potentially fruitful in this context. MCI models were developed by Nakanishi and Cooper (1974). These models are basically multinomial logit models with the purpose to predict the distribution of market shares among brands and electoral votes in ballots (cf. review by Cooper 1993). DeSarbo et al. (2002) present an extended application where they model the prescription share of different physicians for various drugs in response to different marketing strategies of the producers. This problem is comparable to the estimation of the share of different specialisations in farms and the linkage of these shares to potential drivers. Each farmer like each physician can be regarded as a market. However, not the different producers compete for a market share on the physician's prescription but different specialisation like dairy, cash cropping, … compete for the farmer's resources e.g. labour, capital and land. Analogous to these classical agricultural resources the physician's prescriptions constitute a limited resource as both the number of patients and the physician's prescription budget are limited. In case of the different specialisations in agriculture the drivers are not influenced by a solitary agent, who intends to maximize the share of a specialisation but are a result of the overall agricultural business environment and the local conditions, regarding both natural factors and agricultural structure. This feature clearly deviates from the cases analyzed in the MCI literature. In the MCI literature the drivers can at least partially be influenced by a single agent with specific intentions. However, we do not regard this as a general problem regarding the overall applicability of the model in the context of analyzing structural change.

The presentation of the MCI models follows DeSarbo et al. (2002). In the simple MCI model the share s of a single brand (specialisation) i on the total market (farm portfolio) is:

(43): $$s_i = \frac{\exp(\alpha_i + \varepsilon_i)\prod_{k=1}^{K} X_{ik}^{\beta_k}}{\sum_{i=1}^{M}\exp(\alpha_i + \varepsilon_i)\prod_{k=1}^{K} X_{ik}^{\beta_k}}$$), with

K = 1,…, K the different explanatory variables

M = 1,…, M the different specialisations

ε = the error term

α = a general regression parameter varying only by brand (specialisation) *i (effectiveness coefficient)*

β = a specific regression parameter varying by explanatory variable X_k

This model can be extended to a *differential effects* MCI model, where

(44): $$s_i = \frac{\exp(\alpha_i + \varepsilon_i)\prod_{k=1}^{K} X_{ik}^{\beta_{ik}}}{\sum_{i=1}^{M}\exp(\alpha_i + \varepsilon_i)\prod_{k=1}^{K} X_{ik}^{\beta_{ik}}}).$$

Here β depends not only on the explanatory *k* but also on the brand (specialisation) *i*. However, the *differential effects* MCI assumes constant cross-elasticity for all brands (i≠j). This assumption is relaxed in the *fully extended* MCI*:*

(45): $$s_i = \frac{\exp(\alpha_i + \varepsilon_i)\prod_{k=1}^{K}\prod_{j=1}^{M} X_{ik}^{\beta_{ijk}}}{\sum_{i=1}^{M}\exp(\alpha_i + \varepsilon_i)\prod_{k=1}^{K}\prod_{j=1}^{M} X_{ik}^{\beta_{ijk}}}).$$

Unfortunately, the number of estimated parameters increases from K+M in case of the *simple* MCI model to 2MK+M in the *fully extended* version.

4 Recent advances in estimating transition probabilities

4.1 Cross sectional scope

Traditional Markov analysis focused on one or a very small number of regions. Consequently only time variant influences of structural change could be identified. Recent attempts extent the cross regional coverage of Markov studies in order to indentify regional (time-invariant) influence of structural change. First attempts in this respect are undertaken by Zimmermann and Heckelei (2008). Building on their approach, Huettel and Margarian (2009) analyze the influence of the initial regional structure on the development of structural change. They argue that farm specific factors are not sufficient to fully explain farm development and different regional patterns of farm structural change. Mainly through the interactions on the land market, the behaviour of one farmer affects the development of others. They calculate stationary TPs for two time periods for 327 NUTS III regions for West Germany using micro data from the Research Data Centre of the Federal Statistical Office and the statistical offices of the German Laender. In a second step, the calculated TPs are used to calculate mobility indices that summarize farm mobility in a region. With a descriptive comparison of the mobility indices between regions it is analysed how the mobility indices differ between regions with different characteristics. In addition the regional TPs of the two time periods are regressed on explanatory variables using a log odds ratio model. To account for interdependences of the equations a generalized least squares estimation technique is employed to jointly estimate the set of seemingly unrelated regressions.

Zimmermann and Heckelei (2012a) analyse regional differences in farm structural change but with a larger cross-sectional scope. Micro and macro data from the FADN dataset are used to investigate regional differences in structural change across the EU15 from 1990 to 2005. They also consider a broader definition of structural change which not only includes changes in farm size but also changes in the production specialization of a farm. Therefore farms are classified into 31 Markov states that consist of 10 different production specialization category differentiated into 3 size category and one artificial entry/exit category. In the study stationary TPs are estimated using the Generalized Cross Entropy approach. For the estimated TP matrix mobility indices are calculated which are then regressed on explanatory (structural) variables. Through differentiating two dimension of the farm typology it was possible to calculate an overall mobility index as well as specific mobility indices measuring size mobility and production specialization mobility. Results showed that regional characteristics explain a substantial share of the regional differences of farm structural change.

A second paper by Zimmermann and Heckelei (2012 b) maintains the cross-sectional focus on regions in the EU15 but looks specifically at the size related structural development in the dairy sector. As in the first paper, FADN data is used but for a somewhat smaller time period reaching from 1995 to 2005. Six Markov categories are

considered which consist of five size categories and an artificial entry/exit category. The non-stationary TPs are estimated using the Generalized Cross Entropy estimator. The estimated TPs that vary over time and region are than regressed on explanatory variables in a pooled regression without calculating mobility indices as an intermediate step. Apart from other region specific characteristics representing initial farm structure, market conditions, and resource endowments as well as social and demographical factors are used as explanatory variables. The results confirm the empirical findings of the more recent studies that regional-specific variables including initial farm structure significantly affect the TPs and hence structural change.

4.2 Combination of micro and macro data

Another advancement of the studies by Zimmermann and Heckelei (2012a) and Zimmermann and Heckelei (2012b) is the combination of macro and micro data in the estimation of TPs. Both studies use FADN micro data from the annually surveyed farms in combination with macro data which is available by considering the aggregation weight attached to each sample farm derived from the Farm Structure Survey (FSS). The micro data offer the advantage of observed transitions between classes and detailed farm level data whereas the macro dataset provides aggregate information at population level. As mentioned above, if micro data is available, the TPs need not to be estimated but can be calculated directly. The obtained information on the TPs is then used as prior information in a macro data Generalized Cross Entropy estimation approach similar to the approach proposed by Karantininis (2002) and Stokes (2006). The advantage of combining micro and macro data in estimation is that all the available information is considered in one estimation approach. This idea is further pursued by Storm et al. 2011). They develop a Bayesian framework for non-stationary Markov models as an alternative to the Generalized Cross Entropy estimator. The aim is to incorporate the prior information in a more transparent manner and consistent with probability theory. It is specifically analysed how a sample of micro observations can be used to specify a prior density and later be combined in a Bayesian framework with the likelihood function representing the macro data approach. Preliminary Monte Carlo simulation results demonstrate the feasibility of the approach and the increased precision of estimation results for increasing micro samples in a mean square error sense.

4.3 Advancement towards a continuous Markov approach

Another strand of literature concerns advancement of the classical Markov application towards a continuous Markov approach. The above mentioned paper by Piet (2008) was the first attempt in this direction and is advanced by a recent paper of the same author (Piet 2010). The basic idea is to assume a (parametric) distribution for the individual farms with respect to the underlying continuous variable (such as farm size) together with

a continuous transition probability density. This approach tackles two important problems of the classical Markov approach. Firstly, the number of parameters to be estimated is reduced. Instead of estimating a set of independent (except for adding up conditions) transition probabilities, a continuous transition probability density is estimated which can, for simple distribution functions, be characterised by only two parameters. Secondly, the definition of (size) classes can be done flexibly and the number of parameters is independent of the number of classes considered. The main drawback of the approach is the need to assume an appropriate parametric distribution for the individual farms. This is especially problematic if macro data where farms are already classified into (size) classes is the only available source of information. So a useful application requires census type data or representative samples of individual farms. The advantages of simple parametric distributions come at the cost of less flexibility in matching the observed data. Further advancements in this respect are highly desirable.

4.4 Relevance of determinants of structural change from recent papers

The above mentioned recent advancements in the cross-sectional scope of Markov applications allow identifying the influence of regional characteristics on structural change not possible in previous applications. Huettel and Margarian (2009) showed that the initial average farm size and the distribution of land affect structural change in West Germany. Zimmermann and Heckelei (2012a) further extended the cross-sectional coverage to the EU15 (but also derived results for Germany, France, Spain and Italy separately). By considering farm size change and specialization change as two dimensions on structural change a broader definition of structural change is employed. Beside farm size and size heterogeneity they also indentified the unemployment rate, as a measure of off-farm employment alternatives, as a relevant factor for regional differences of structural change. With respect to specialization changes, the share of farms with a mixed production program was found to have a significant positive influence on the mobility to change the specialization. Results by Zimmermann and Heckelei (2012b) confirmed the general findings of other recent studies with respect to the influence of the average farm size, the farm size heterogeneity and the influence of the unemployment rate. Additionally the initial stocking density, the average land rent as well as natural resource factors such as the share of grassland, the slope and the temperature in a region were identified as relevant factors. Further, the population density and the population growth are found to be additionally relevant social and demographical factors explaining regional differences in structural change.

One common observation of the Markov application is that the effects of regional characteristics can differ in different time periods (observed in Huettel and Margarian (2009) by comparing results between two different time periods) or between broader regions (observed in Zimmermann and Heckelei (2012a) by a comparison of results for the EU15, Germany, France, Spain and Italy). These observations correspond to findings

outside the Markov literature by Röder and Kilian (2011). They only considered farm exit as a measure for farm structural change. Using data for German municipals for a time period from 1999 to 2007 the results indicate that the strength of influence, and for some variables even the direction of influence, depends on the regional context. For example, they found that the characteristics of farm land distributions can have ambiguous effects on farm exit depending on the regional context. They argue that this observation might explain the ambiguous findings for the influence of some factors on structural change in the literature.

5 Ex-ante simulation of structural change

For ex-ante policy impact assessment obviously more is needed than to just identify relevant determinants of structural change and to quantify their impact in econometric exercises. In principle, all of the approaches described above qualify for using the estimated models as simulation models to project the future development of farm structure. For this it is required that development paths of all relevant explanatory variables until the desired simulation year are made available by other statistical forecasting exercises or at least are defined by assumption in a scenario analysis.

Here we restrict our attention, however, to those approaches that potentially allow to project a farm structure typology at full sectoral and regional coverage. More precisely, we will explore the literature regarding research relevant for adjusting the farm type structure in a sectoral model like the farm type version of CAPRI (Gocht and Britz 2010) to future simulation years. This excludes all econometric models considered above which only provide information on partial aspects of farm structural change, such as those focusing on entry/exit decisions, or which do not (at least implicitly) take into account the interaction of farms at the regional level relevant for the adjustment of the whole sector, such as growth models at individual farm level.

The consideration of ex-ante analysis of structural change within or in connection to a partial equilibrium model does introduce another complication beyond the need to project exogenous explanatory variables mentioned above: there exist simultaneous dependence between the outputs of the market model and structural variables. For example, decisions on investments in production capacities or primary factors in the form of land, labour and physical capital employed in different farm types at least partially depend on relative price developments for the main outputs of these farm types. These product prices in turn depend on market outcomes influenced by capacity decisions of a sufficiently large supply region.

The simultaneous dependence of structural developments and output prices in the context of agricultural sector models have been addressed in the past by (1) recursive dynamic model formulation, (2) by a fully dynamic solution and (3) by ignoring the simultaneity and focusing on exogenous determinants identified by estimating Markov transition probabilities.

Recursive dynamic programming models have already been introduced in the 1960s to represent dynamic adjustments of production capacities at farm level and later also picked up regional interdependence and structural elements (Day and Cigno, 1978). Typically, yearly iterative steps allow adjusting production capacities at farm, farm type or regional level to supply model or market equilibrium model outcomes of the previous year. The Dynamic Analysis and Prognosis System (DAPS, translated from German) is a rather sophisticated example of such an approach developed by Bauer (1979) and later extended

and applied by Loritz-Hoffmann (1988). For a good overview on structure and assumptions see Bauer (1989). Land, labour and equipment constraints in a profit maximisation context generate shadow prices depending on the economic profitability of production activities. The larger these shadow prices, the higher is the economic incentive to relax these constraints, i.e. to invest in the expansion of these production factors. In DAPS, the extent of the delayed adjustment was determined by empirically estimated behavioural functions.

Even though it was not implemented in DAPS, such recursive dynamic formulations may be incorporated in market equilibrium models with endogenous output prices. When farm types are distinguished, then differential developments of prices for certain agricultural product groups will cause differentiated expansion or contraction of production capacities in different specialisations over the time horizon of simulation. However, this type of model does not naturally lend itself to incorporating the development of farm size distributions. One can certainly distinguish farm types also by size, but typical model specifications will not allow identifying what part of primary factor adjustments can be attributed to farm size changes and what part to changes in the number of farms within a type.

An example of a recursive dynamic programming that explicitly considers endogeneous prices and different farm size groups is the Finnish agricultural sector model DREMFIA (e.g. Lehtonen 2001 und 2004; Lehtonen et al. 2007). This spatial (dis-) equilibrium model of the agricultural sector generally follows the same modelling philosophy as DAPS, but maximises the sum of producer and consumer surplus to obtain annual market equilibriums instead of a pure supply specification. Its extended version furthermore includes an endogenous investment and technology diffusion model (Lehtonen, 2001) which is connected to farm size by letting production cost decrease with increasing farm size. The model seems to stop short, however, of representing farm size development as an endogenous variable, but instead allows statements on how farm sizes – and with it the number of farms – has to develop to achieve a certain farm income level under different policy scenarios. The previous sentence deliberately uses the words “seems to” as the available literature cited above does not allow to fully clarify the status of the farm size distribution as an endogenous or exogenous variable.

Weiss (2007) included developments in entry/exit, farm size and specialisation endogenously in the agricultural model FAMOS by adjusting the weights of the 8000 represented farms over time based on estimated logistic functions. The original problem identified by the author in the introduction is similar to the current status of the CAPRI model with farm type layer: lack of adjustment of farm type representativeness in simulations across several years which ignores the quite relevant changes in farm size and specialisations observed in reality. Due to the strong differentiation of farm groups desired in ex-ante projection activities, a fully simultaneous Markov estimation was not possible. However, the logistic estimates yielded probabilities of changing, for example, the farm

size group analogous to a non-stationary Markov transition probability matrix. A simple simulation exercise four years into the future (out-of-sample projection) showed that the use of the estimated logistic models and corresponding adjustments in production capacities of certain farm groups performed considerably better in matching observed structural variables compared to keeping the aggregation weights fixed. A limitation of this approach is that product prices could not be included as explanatory variables as there was no variation in the data set. This limits the usefulness for longer term projections in connection with a market equilibrium model.

In general, we see that ex-ante projections of structural change at the sectoral level with sufficient, i.e. policy relevant level of product and farm type differentiation, are very rare. All existing approaches have considerable limitations, the most important of which is the lack of simultaneous feedback between different output prices and structural variables.

6 Conclusions

This second part of the report presents a literature review with relevance to ex-post and ex-ante analysis of structural change at the agricultural sector level while maintaining a policy relevant detail of analysis. With a resulting focus on Markov chain models able to estimate and explain transition probabilities for farm movement between farm typology classes, different estimation techniques depending on the available data were distinguished. In addition, the review aimed at a general overview on the empirical relevance of potential explanatory variables to be included in the project's intended estimation exercises, thereby considering other econometric exercises aimed at partial aspects of structural change. Finally, model specifications mainly targeted at simulating ex-ante structural change under economic and policy scenarios are reviewed.

The review of the Markov chain literature shows considerable variety with respect to specific approaches applied. Apart from data availability, it seems that computational requirements exponentially increasing with farm types and observations have restricted most studies in the past with respect to the regional scope and class resolution. However, developments in estimation based on cross-entropy formulations allowed for some extensions in the more recent literature. An interesting possibility is the combination of micro and macro data in estimation using cross-entropy formulation or Bayesian methods currently under development.

Turning to the identification of relevant determinants we can conclude that up to about three years ago Markov chain models mainly identify statistical relevance of technological change, government programs, and prices of outputs and inputs related to the specific farm specialisation class considered. Price variables typically capture incentives within the class (key output prices) and those of alternative uses of the resources (land values and interest rates). The limitation to single or very few regions of analysis explains that theoretically relevant variables with little variation in time were not considered. But most recent papers concentrate on regional and farm structural differences and are able to show their relevance in explaining variation of structural change across regions.

In the regression analyses of farm growth and exit/entry decisions, a different and large set of variables is employed with a strong emphasis on socio-demographic determinants. Variables selected and found statistically significant in both types of analyses have a considerable overlap because continuation and growth is the alternative decision to exiting the sector. Farmer's age and education as well as farm size appear relevant in most of the studies presented. The very few cohort analyses focus on autonomous demographical drivers of labour use in agriculture. The key lesson to be learned here is the importance of the age-structure of farm holders for the aggregate exit pattern of farms confirming the significance of farmer's age in strategic decisions on farm continuation. This variable as well as measures on education are so far ignored in Markov chain analyses and could be relevant for cross regional variation in aggregate exit decisions. Although important

aspects of farm structural change were identified by the considered farm growth, exit/entry and cohort analyses, the conceptual lack of interaction between farms or farm types makes it difficult to infer deeper general insight or aggregate impacts. Despite their theoretical importance for the strategic farm decisions modelled, the conditions outside of agriculture (job opportunities, interest rates, alternative uses of land, etc.) rarely proved to have statistical influence or were not even considered. Again, most likely a limited variation of these variables due to the small regional coverage might explain this observation and has also recently been successfully addressed.

An interesting approach with some similarity to the estimation of transition probabilities are the multiplicative competitive interaction models, potentially allowing for more flexible and general model specifications that allow taking into account better the non-linear nature of relationships between explanatory variables and the structural change phenomenon emphasized in recent publications.

Regarding the consideration of ex-ante analysis of structural change, the literature review has proofed that ex-ante projections of structural change at the sectoral level with sufficient, i.e. policy relevant level of product and farm type differentiation, are very rare. All existing approaches have considerable limitations, the most important of which is the lack of simultaneous feedback between different output prices and structural variables.

References

Acs, S., Berentsen, P.B.M., Huirne, R.B.M., 2005. Modelling Conventional and Organic Farming: A Literature Review. NJAS Wageningen Journal of Life Sciences 53 (1), 1-18.

Anderson, T.W., Goodman, L.A., 1957. Statistical Inference about Markov Chains. Annals of Mathematical Statistics 28, 89-110.

Bauer, S., 1989. Some lessons from the Dynamic Analysis and Prognosis System (DAPS). In: Bauer S. and Henrichsmeyer W. (edts.): Agricultural Sector Modelling. Vauk Verlag, Kiel.

Bauer, S.,1979. Quantitative Sektoranalyse als Entscheidungshilfe für die Agrarpolitik: Ein quantitatives Analyse- und Prognosesystem für die Landwirtschaft in der Bundesrepublik Deutschland (DAPS). Duncker & Humblot, Berlin.

Boehlje, M., 1992. Alternative Models of Structure Change in Agriculture and Related Industries. Agribusiness 8 (3), 219-231.

Bremmer, J., Oude Lansink, A.G., Olson, K.D., Baltussen, W.H., Huirne, R.B., 2004. Analysis of Farm Development in Dutch Agriculture and Horticulture. In: Bremmer, J. (Ed.), Strategic Decision Making and Firm Development in Dutch Horticulture. PhD-thesis, Wageningen University and Agricultural Economics Research Institute, Wageningen, Netherlands, pp. 23-39.

Chavas, J.P., Magand, G., 1988. A Dynamic Analysis of the Size Distribution of Firms: The Case of the US Dairy Industry. Agribusiness 4 (4), 315-329.

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

Cooper, L.G., 1993. Market share models. In: Eliashberg, J. and Lilien, G.L. (eds.): Handbooks in operations research and management science (5): 259–311, North Holland, New York.

Day, R.H., Cigno, A., 1978. Modelling Economic Change: The Recursive Programming Approach, North-Holland, Amsterdam.

De Haen, H., Von Braun, J., 1977. Mobility of Agricultural Labour and Fluctuating Regional Labour Markets: A Demographic and Economic Analysis with Application to West-Germany. European Review of Agricultural Economics 4 (3), 215-243.

DeSarbo, W. S., Degeratu, A. M., Ahearne, M. J., Saxton, K. M., 2002. Disaggregate market share response models. International Journal of Research in Marketing 19, 253–266.

Disney, W.T., Duffy, P.A., Hardy, W.E.J., 1988. A Markov Chain Analysis of Pork Farm Size Distributions in the South. Southern Journal of Agricultural Economics 20, 57-64.

Edwards, C., Smith, M.G., Peterson, R.N., 1985. The Changing Distribution of Farms by Size: A Markov Analysis. Agricultural Economics Research 37 (4), 1-16.

Ethridge, D.E., Roy, S.K., Myers, D.W., 1985. A Markov Chain Analysis of Structural Changes in the Texas High Plains Cotton Ginning Industry. Southern Journal of Agricultural Economics 7 (2), 11-20.

Garcia, P., Offutt, S.E., Sonka, S.T., 1987. Size Distribution and Growth in a Sample of Illinois Cash Grain Farms. American Journal of Agricultural Economics 69 (2), 471-476.

Garvey, E., 2006. Estimations of Labour Inputs using FADN Data, SEAMLESS Report No. 15, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.SEAMLESS-IP.org, 35 pp, ISBN no. 90-8585-043-6.

Gebremedhin, T.G., Christy, R.D., 1996. Structural Changes in U.S. Agriculture: Implications for Small Farms. Journal of Agricultural and Applied Economics 28 (1), 57-66.

Gibrat, R., 1931. Les Inégalités Economique, Sirey, Paris.

Gillespie, J.M., Fulton, J.R., 2001. A Markov Chain Analysis of the Size of Hog Production Firms in the United States. Agribusiness 17 (4), 557-570.

Gocht, A., Britz, W., 2010. EU-wide farm type supply models in CAPRI - How to consistently disaggregate sector models into farm type models. Journal of Policy Modeling. 33(1): 146-167

Goddard, E., Weersink, A., Chen, K., Turvey, C.G., 1993. Economics of Structural Change in Agriculture. Canadian Journal of Agricultural Economics 41, 475-86.

Golan, A., Judge, G., Miller, D., 1996. Maximum Entropy Econometrics, John Wiley and Sons, Chichester.

Golan, A., Vogel, S., 2000. Estimation of non-stationary social accounting matrix coefficients with supply side information. Economic Systems Research 12, 447-471.

Hallam, A., 1991. Economies of Size and Scale in Agriculture: An Interpretative Review of Empirical Measurement. Review of Agricultural Economics 13 (1), 155-172.

Hallam, A., 1993. Empirical Studies of Size, Structure, and Efficiency in Agriculture. In: Hallam, A. (Ed.), Size, Structure, and the Changing Face of American Agriculture. Westview Press, Boulder, Colorado, pp. 204-231.

Hallberg, M.C., 1969. Projecting the Size Distribution of Agricultural Firms - An Application of a Markov Process with Non-Stationary Transition Probabilities. American Journal of Agricultural Economics 51, 289-302.

Harrington, D.H., Reinsel, R.D., 1995. A Synthesis of Forces Driving Structural Change. Canadian Journal of Agricultural Economics Special Issue 43, 3-14.

Huettel, S., Jongeneel, R., 2008. Structural Change in the Dairy Sectors of Germany and the Netherlands – A Markov Chain Analysis. Paper presented at the 12th EAAE Congress, August 26-29, Ghent, Belgium.

Huettel, S., Jongeneel, R., 2011. How has the EU milk quota affected patterns of herd-size change? European Review of Agricultural Economics, (Advance Access).

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

Jongeneel, R., 2002. An Analysis of the Impact of Alternative EU Dairy Policies on the Size Distribution of Dutch Dairy Farms: An Information Based Approach to the Non-Stationary Markov Chain Model. Paper presented at the 10th EAAE Congress, August 28-31, Zaragoza, Spain.

Jongeneel, R., Longworth, N., Huettel, S., 2005. Dairy Farm Size Distribution in East and West: Evolution and Sensitivity to Structural and Policy Variables: Case-Studies of the Netherlands, Germany, Poland and Hungary. Paper presented at the 11th EAAE Congress, August 24-27, Copenhagen, Denmark.

Jongeneel, R., Tonini, A., 2008. Dairy Quota and Farm Structural Change: A Case Study on the Netherlands. Paper presented at the 107th EAAE Seminar, January 29-February 1, Seville, Spain.

Judge, G.G., Swanson, E.R., 1961. Markov Chains: Basic Concepts and Suggested Uses in Agricultural Economics. Agricultural Experiment Station Research Report AERR-49. University of Illinois. Department of Agricultural Economics, Urbana, Illinois.

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

Keane, M., 1976. Projections of Change in Creamery Supplier Structure. Irish Journal of Agricultural Economics and Rural Sociology 6 (2), 177-184.

Keane, M., 1991. Changes in the Size Structure of Irish Dairy Farms. Irish Journal of Agricultural Economics and Rural Sociology 14, 67-74.

Kimhi, A., Nachlieli, N., 2001. Intergenerational Succession in Israeli Family Farms. Journal of Agricultural Economics 52 (2), 42-58.

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

Lee, T.C., Judge, G.G., 1996. Entropy and Cross-Entropy Procedures for Estimating Transition Probabilities from Aggregate Data. In: Berry, D.A., Chaloner, K.M., Geweke, J.K. (Eds.), Bayesian Analysis in Statistics and Econometrics - Essays in Honor of Arnold Zellner. John Wiley and Sons, Inc., New York, pp. 153-164.

Lee, T.C., Judge, G.G., Zellner, A., 1977. Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, North-Holland Publishing Company, Amsterdam.

Lehtonen, H., 2001. Principles, structure and application of dynamic regional sector model of Finnish agriculture. PhD thesis, Systems Analysis Laboratory, Helsinki University of Technology, Agrifood Research Finland, Economic Research (MTTL). Publ. 98.

Lehtonen, H., 2004. Impacts of de-coupling agricultural support on dairy investments and milk production volume in Finland. Acta Agric. Scand. C: Food Econ. 1, 46-62.

Lehtonen, H., Bärlund, I., Tattari, S., Hilden, M.. 2007. Combining dynamic economic analysis and environmental impact modelling: Addressing uncertainty and complexity of agricultural development. Environmental Modelling & Software 22, 710-718

Loritz-Hoffmann, J., 1988. Die Auswirkungen agrarpolitischen Instrumenteneinsatzes auf die agrarsektoralen Anpassungsprozesse und ihre Abhängigkeit von den gesamtwirtschaftlichen Rahmenbedingungen. Dissertation: University of Bonn

MacRae, E.C., 1977. Estimation of Time-Varying Markov Processes with Aggregate Data. Econometrica 45 (1), 183-198.

Nakanishi, M., Cooper, L.G., 1974. Parameter Estimation for a Multiplicative Competitive Interaction Model — Least Square Approach. Journal of Marketing Research 11: 303–311.

Olson, K.D., Stanton, B.F., 1993. Projections of Structural Change and the Future of American Agriculture. In: Hallam, A. (Ed.), Size, Structure, and the Changing Face of American Agriculture. Westview Press, Boulder, Colorado, pp. 587-627.

Padberg, D.I., 1962. The Use of Markov Processes in Measuring Changes in Market Structure. Journal of Farm Economics 44 (1), 189-199.

Peterson, E.W.F., 1986. Agricultural Structure and Economic Adjustment. Agriculture and Human Values 3 (4), 6-15.

Piet, L., 2008. The Evolution of farm size distribution: revisiting the Markov chain model. Paper presented at the 12th EAAE Congress, August 26-29, Ghent, Belgium.

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.

Pietola, K., Heikkilä, A., 2006. Possibilities to Switch into Capital-Intensive Technologies in Small Scale Dairy Production. Contributed paper, 96th EAAE Seminar, January 10-11, Tänikon, Switzerland.

Pietola, K., Väre, M. Oude Lansink, A., 2003. Timing and type of exit from farming: farmers' early retirement programmes in Finland. European Review of Agricultural Economics 30(1): 99-116.

Rahelizatovo, N.C., Gillespie, J.M., 1999. Dairy Farm Size, Entry, and Exit in a Declining Production Region. Journal of Agriculture and Applied Economics 31 (2), 333-347.

Reimund, D.A.; Moore, C.V.; Martin, J.R., 1977. Factors Affecting Structural Change in Agricultural Subsectors: Implications for Research. Southern Journal of Agricultural Economics 9 (1), 11-19.

Röder, N., Kilian, S., 2011. Which parameters determine the development of farm numbers in Germany? Jahrbücher f. Nationalökonomie u. Statistik 231 (1):1–21.

Salkin, M.S., Just, R.E., Cleveland, J.O.A., 1976. Estimation of Nonstationary Transition Probabilities for Agricultural Firm Size Projection. Annals of Regional Science 10, 71-82.

Shapiro, D., Bollman, R.D., Ehrensaft, P., 1987. Farm Size and Growth in Canada. American Journal of Agricultural Economics 69 (2), 477-483.

Shorrocks, A.F., 1978. The Measurement of Mobility. Econometrica 46 (5), 1013-1024.

Stanton, B.F., Kettunen, L., 1967. Potential Entrants and Projections in Markov Process Analysis. Journal of Farm Economics 49, 633-642.

Stavins, R.N., Stanton; B. F., 1980. Using Markov Models to Predict the Size Distribution of Dairy Farms, New York State, 1968-1985. Cornell University. Department of Agricultural Economics, Ithaca, New York.

Stokes, J.R., 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.

Sumner, D.A., Leiby, J.D., 1987. An Econometric Analysis of the Effects of Human Capital on Size and Growth among Dairy Farms. American Journal of Agricultural Economics 69 (2), 465-470.

Telser, L.G., 1963. Least-Squares Estimates of Transition Probabilities. In: Christ, D.F. (Ed.), Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld. Stanford University Press, Stanford, pp. 270-292.

Tonini, A., 2007. Agriculture and Dairy in Eastern Europe after Transition focused on Poland and Hungary, Wageningen, Netherlands.

Tonini, A., Jongeneel, R., 2002. Dairy farm size restructuring in Poland and Hungary. In: Hinners-Tobraegel, L, Heinrich, J. (Eds.), Agricultural Enterprises in Transition: Parallels and Divergences in Eastern Germany, Poland and Hungary. Studies on the Agricultural and Food Sector in Central and Eastern Europe Vol. 15, Wissenschaftsverlag, Kiel, pp. 317-339.

Tonini, A., Jongeneel, R., 2008. The Distribution of Dairy Farm Size in Poland: A Markov Approach based on Information Theory. Applied Economics 40, 1-15.

U.S. Congress, O.o.T.A., 1985. Factors Contributing to Structural Change in Agriculture. In: U.S. Congress, O.o.T.A. (Ed.), Technology, Public Policy, and the Changing Structure of American Agriculture: A Special Report for the 1985 Farm Bill. Technology, Public Policy, and the Changing Structure of American Agriculture: A Special Report for the 1985 Farm Bill. Washington, DC, pp. 107-118.

Van Dijk, G., Hoogervorst, N., Bauwens, N., Tijdink, T., 1986. A Framework for Agricultural Structural Policy Analysis. Sociologia Ruralis 16 (2), 109-127.

Von Massow, M., Weersink, A., Turvey, C.G., 1992. Dynamics of Structural Change in the Ontario Hog Industry. Canadian Journal of Agricultural Economics 40, 93-107.

Weiss, C.R., 1999. Farm Growth and Survival: Econometric Evidence for Individual Farms in Upper Austria. American Journal of Agricultural Economics 81, 103-116.

Weiss, F., 2007. Modellierung landwirtschaftlichen Strukturwandels in Österreich: Vergleich einer Modellprognose mit den Ergebnissen der Strukturerhebungen (1999-2005). Diskussionspapier DP-33-2007, Institut für nachhaltige Wirtschaftsentwicklung, Universität für Bodenkultur Wien.

Zepeda, L., 1995a. Asymmetry and Nonstationarity in the Farm Size Distribution of Wisconsin Milk Producers: An Aggregate Analysis. American Journal of Agricultural Economics 77 (2), 837-852.

Zepeda, L., 1995b. Technical Change and the Structure of Production: A non-stationary Markov Analysis. European Review of Agricultural Economics 22, 41-60.

Zimmermann, A., Heckelei, T., 2008. Farm Structural Change in German Regions - An Empirical Analysis Using Micro and Macro Data. Poster presented at the XIIth EAAE Congress, August 26-29, Ghent, Belgium.

Zimmermann, A., Heckelei, T., Pèrez Dominguez, I. (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. 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: 576–603.

Annexes

Annex 1

Stationary Markov studies in the agricultural economics literature

Annex 1: Stationary Markov studies in the agricultural economics literature

Year	Author	Region	Specialisation	Type of Data	Time Series	Methodology	Number of States	Dependent Variables	Performance
1961	Judge and Swanson	Illinois, USA	Pig farms	Micro	1946-1958	Maximum likelihood	6 + entry/exit	Firm size (in number of litters of hogs)	-
1962	Padberg	California, USA	Wholesale fluid milk industry	Micro	1950-1955, 1955-1960	Maximum likelihood	3 + entry/exit	Firm size (in market shares)	-
1964	Krenz	North Dakota, USA	All farms	Macro	1935-1960	Maximum likelihood	6 + entry/exit	Firm size (in acres)	-
1967	Stanton and Kettunen	New York, USA	Dairy farms	Micro	1960-1964	Maximum likelihood	3 + entry/exit	Firm size (in herd size)	-
1976	Keane	South of Ireland	Dairy farms	Macro	1968-1973	Maximum likelihood	6 + entry/exit	Firm size (in milk supply)	-
1985	Edwards et al.	USA	All farms	Micro	1974-78	Maximum likelihood	8; 8 + entry/exit	Firm size (by acres, value of sales, tenure, standard industrial classification)	-
1987	Garcia et al.	Illinois, USA	Cash grain farms	Micro	1976-1985	Maximum likelihood	11 for each size measure (entry/exit not considered)	Firm size (gross value of farm product/tillable acres)	Within sample prediction (average root mean square error): 3.1/11.7%
1991	Keane	Dairy co-operative society, Ireland	Dairy farms	Macro	1983-1989	Maximum likelihood	7 + entry/exit	Firm size (in milk supply)	-
2002	Tonini and Jongeneel	Poland	Dairy farms	Macro	1981/1987, 1998-2001	Maximum likelihood	4 + entry/exit, 6 + entry/exit	Firm size (in herd size)	Within sample prediction (average prediction error): 0.25%
2008	Jongeneel and Tonini	Netherlands	Dairy farms	Macro	1972-2006	GCE	7 + entry/exit	Firm size (in herd size)	Pseudo-R^2: 0.33-0.38
2008	Piet	France	All farms	Macro	1980-2005	Nonlinear least-squares	Continuous	Firm size (in utilised agricultural area)	R^2: 0.99

European Commission
EUR 25555 – Joint Research Centre – Institute for Prospective Technological Studies

Title: 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

Authors: Alexander Gocht, Norbert Röder, Sebastian Neuenfeldt, Hugo Storm, Thomas Heckelei
Editors: María Espinosa and Sergio Gomez y Paloma

Luxembourg: Publications Office of the European Union

2012– 166 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online)

ISBN 978-92-79-27047-5 (pdf)

doi:10.2791/17545

Abstract

The present study aims to develop a prototype analytical tool to assess structural changes at the farm level in EU-27 using the Farm Accountancy Data Network (FADN) combined with the Farm Structure Survey (FSS). For the purpose of this study, farm structural change is related to the change in production systems, therefore a change in farm size and farm entry/exit into one sector/farm typology.

In the ex-post analysis of structural change two methodologies are presented, one in which structural change is analysed from a discrete perspective using a Markov approach, whereas the second uses the continuous perspective to evaluate the type of farming over time using MCI (Multiplicative Competitive Interaction) models. The methodolgies are applied in selected German regions and the goodness of fit in the out of sample prediction is compared.

In the ex-ante methodology, the existing farm module of CAPRI (Common Agricultural Policy Regionalised Impact System) is expanded by considering the findings of the statistical ex-post-analysis when projecting farm-type structural change in the baseline trends.

Results show that the Markov prediction may outperform naïve prediction methods but that the quality of the prediction is critically dependent on the model specification. A higher in-sample fit does not necessarily lead to better out-of-sample prediction, which potentially indicates that the effects of specific explanatory variables may change over time.
In addition, introducing structural change into the CAPRI farm type baseline improve policy impact assessment and hence a more reliable and consistent farm grid for simulations is constructed.

LF-NA-25555-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-27047-5

9 789279 270475