seamless

System for Environmental and Agricultural Modelling; Linking European Science and Society

Working paper – Literature Review of Approaches to Estimate Structural Change

A. Zimmermann, T. Heckelei, I. Pérez

Partner involved: UBONN

Report no.: 16
October 2006
Ref: PD3.6.6.

universitätbonn

Logo's main partners involved in this publication

Sixth Framework Programme

SEAMLESS integrated project aims at developing an integrated framework that allows ex-ante assessment of agricultural and environmental policies and technological innovations. The framework will have multi-scale capabilities ranging from field and farm to the EU25 and globe; it will be generic, modular and open and using state-of-the art software. The project is carried out by a consortium of 30 partners, led by Wageningen University (NL).

Email: seamless.office@wur.nl
Internet: www.seamless-ip.org

Authors of this report and contact details

Name: Andrea Zimmermann — Partner acronym: UBONN
Address: Nussallee 21, D-53115 Bonn, Germany
E-mail: andrea.zimmermann@ilr.uni-bonn.de

Name: Thomas Heckelei — Partner acronym: UBONN
Address: Nussallee 21, D-53115 Bonn, Germany
E-mail: thomas.heckelei@ilr.uni-bonn.de

Name: Ignacio Pérez — Partner acronym: UBONN
Address: Nussallee 21, D-53115 Bonn, Germany
E-mail: ignacio.perez@ilr.uni-bonn.de

Disclaimer 1:

“This publication has been funded under the SEAMLESS integrated project, EU 6th Framework Programme for Research, Technological Development and Demonstration, Priority 1.1.6.3. Global Change and Ecosystems (European Commission, DG Research, contract no. 010036-2). Its content does not represent the official position of the European Commission and is entirely under the responsibility of the authors.”

Disclaimer 2:

Within the SEAMLESS project many reports are published. Some of these reports are intended for public use, others are confidential and intended for use within the SEAMLESS consortium only. As a consequence references in the public reports may refer to internal project deliverables that cannot be made public outside the consortium.

When citing this SEAMLESS report, please do so as:

Zimmermann. A, Heckelei, T., Pérez, I., 2006. Working paper – Literature Review of Approaches to Estimate Structural Change, SEAMLESS Report No.16, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.SEAMLESS-IP.org, 45 pp.

Table of contents

General information

Task(s) and Activity code(s):	Task 3.6, Activity 3.6.5
Input from (Task and Activity codes):	Task 3.6, Activity 3.6.5
Output to (Task and Activity codes):	Task 3.6, Activity 3.6.5
Related milestones:	M3.6.5

Executive summary

Changing economic and political conditions as well as dynamic processes within the agricultural sector lead to a continuous redistribution of resources between farms and to changes of the production systems on farms over time. This type of structural change is generally characterized by increasing average farm size and specialization. However, significant national and regional differences in development exist resulting from complex interdependencies between sectoral support policies, original farm structure, land market institutions, non-agricultural employment opportunities, technological developments etc. Given the objective of SEAMLESS to assess impacts and sustainability of agri-environmental policies, it is desirable to identify and be able to project the policy impact on structural change in agriculture. It is therefore envisaged to develop a module capable of forecasting regional shares of farm types depending on policies and a changed economic environment. This module will allow aggregation weights used in the up-scaling procedures from farm to market level to adjust endogenously (DOW).

This deliverable provides a review of the literature on modelling structural change in the agricultural sector. The objectives of the paper are: (1) to describe the methods that have been applied to explain and project structural change in agriculture, (2) to identify the major determinants of structural change and their relative contribution to explaining structural change and (3) to draw conclusions with respect to the methodology suitable for the purpose of SEAMLESS.

The major findings from the review are the following:

- The Markov chain analysis already identified in the DOW as the likely most suitable approach for SEAMLESS purposes is the most widely applied approach to model structural change.
- Statistically significant determinants differ from analysis to analysis, but do include economic variables depending on policy and technology scenarios.
- The Markov chain analysis is applicable to the EU-15 given the FADN data availability. Compared to rather focussed applications in the literature (regionally and with respect to specific farm types), the task is very ambitious, but a successful large scale cross-sectional analysis would also be of high scientific value.
- For the new member states, data availability does not seem to allow statistical analysis. An alternative approach (either transferring results from EU-15 or using simplified agent-based approaches) needs to be identified.
- Agent-based simulation models are increasingly used to model structural change, as they allow simulating a larger range of scenarios compared to statistical approaches. However, they still lack proof of tracking ability and valid responses to changed political and economic conditions. Their results, however, might lead to the inclusion of additional explanatory variables in the Markov chain analyses.

Specific part

1 Introduction

The focus of this paper lies in examining the empirical literature on structural change in the agricultural sector. With this purpose, different ways of predicting or analyzing structural change are summarized according to the methodology used. The objectives are: (1) to describe the methods that have been applied to explain and project structural change in agriculture, and (2) to identify the major determinants of structural change and their relative contribution to explaining structural change and (3) to draw conclusions with respect to the methodology suitable for the purpose of SEAMLESS.

Models dealing with structural change in agriculture are here classified according to their methodology (econometric and simulation models). This criterion is not excluding other ways of classification, e.g. according to the elements defining farm structure. Nevertheless, single modelling approaches and how structural change is defined in terms of its aspects of analysis are highly correlated. Thus the stratification along methodological aspects was deemed to be the most reasonable option.

The analysis is organised as follows. Firstly, earlier literature reviews and theoretical considerations on factors influencing structural change are examined in order to derive major causes determining structural change. Then, empirical studies on structural change are surveyed according to the methodology used. A distinction is made between *econometric models* and *simulation models*. Within the framework of econometric models, we further differentiate between regression analyses that mainly aim at quantifying the impact of specific aspects to different dimensions of structural change, Markov models, which have a more predictive nature, cohort analyses, and models of discrete choice. Among the simulation models multi-agent systems, as a newly established method, shall be looked at in more detail.

2 Structural change in Agriculture

2.1 Definition of structural change

One of the problems faced in the analysis of structural change in agriculture is the heterogeneity of the definition of *farm structure*. There is basically a general recognition of the complexity of this concept but no single widely accepted definition (Stanton 1993, Balmann 1997). Balmann, for example, defines it as: "who is producing what, in what amounts and by what means?" (Balmann 1997, p. 106). Nevertheless, from a wider perspective the concept of agricultural structure can be framed by looking at its main elements: farm size, resource ownership and control, managerial and technological requirements, tenure pattern, importance of part-time operations, degree of vertical integration in a given industry, organisation of production, ease of entry into farming as an occupation and manner of asset transfer to succeeding generations (Penn 1979, Tweeten 1984, Knutson et al. 1990). This list is not exhaustive but pretends to cover the main definitory elements of agricultural structure found in the literature.

The definition of *structural change* varies depending on the underlying definition of the *agricultural structure*. Basically there are two orientations: one relating to productivity changes (e.g. Oehmke and Schimmelpfennig 2004, Kim et al. 2005) and another relating to the structure of the industry. The first definition of structural change leads to the wide field of time series analyses (e.g. determination of structural breaks) which is extensively covered in the branch of general economics. In agricultural economics, however, the focus of the discussion often lies on changes in the structure of the industry. Nevertheless, in most studies both are evaluated together, since farm structure is usually not independent of production relationships. In this paper only the definition relating to the structure of the industry will be taken into account whereby the explicit focus of the studies varies with the methodology used to quantify structural change. The main aspects of structural change analysed in the literature can be summarized under the topics farm growth, entry and/or exit, and farm succession (the latter a special variant of farm exit). Thus, for our purposes structural change might simply be defined as the change in the number of farms in different farm types (as classified e.g. according to different size or activity measures, age cohorts, specialisation classes etc.).

When analysing structural change, an additional question relates to how different constraints affect these (or some of these) elements. With this purpose, researchers usually try to explain their changes motivated by the imposition of specific economic, social and environmental constraints within a certain time perspective (factors contributing to structural change). Moreover, important interactions between these elements might be observed (e.g. the effect of off-farm employment in family farm size), so that some of them might become explanatory variables for the rest. Here is where a second problem arises, namely the classification of farms into homogeneous groups, i.e. farm typologies. To date, no internationally accepted classification has emerged (Stanton 1993). The major measures refer to farm size (economic or physical size), labour requirement and share of household income provided by the farm. It is important to consider all these issues in a comprehensive analysis of structural change.

2.2 Factors contributing to structural change

Most studies about farm structures provide an enumeration of the factors assumed to determine structural change in agriculture. Here, only a short overview of these factors is given, leaving the in-depth discussion to others (e.g. Hallam 1991 and 1993, Goddard et al. 1993, Harrington et al. 1995).

In his analysis "Empirical Studies of Size, Structure, and Efficiency in Agriculture" Hallam (1993) distinguishes between normative and aggregate studies of size and scale, studies of firm efficiency, and models of firm growth and survival. Under normative studies he summarizes cost function, production function and profit function studies. To studies of firm efficiency belong relative efficiency models and frontier functions. Our particular interest lies in, to what Hallam refers to as "models of firm growth and survival".

By reviewing the literature on economies of size and scale in the agricultural sector, Hallam (1991) finds out that given the empirical evidence on these issues, firm size in production agriculture should be fairly constant with little entry and exit in the years ahead. As this is not the case in reality, he concludes that the differences between the actual changes in industry structure and those predicted by the theory of "economic size" should be explained by factors such as (1) external economies, e.g. pecuniary economies in purchase and sale of products, or economies realized as the industry expands and specializes, (2) technical change, (3) management and information, (4) differences in values and goals of the farmers, and (5) opportunity costs outside the agricultural sector.

Other authors determined similar factors as the major causes for structural change. All studies stress the importance of the following factors: technology, off-farm employment, public programs, market structure, human capital and economic forces (U.S. Congress 1985, Goddard et al. 1993, Harrington et al. 1995, Hallam 1991, Boehlje 1990). Furthermore, Goddard et al. (1993) add demographical aspects to the list of factors contributing to structural change. They also emphasized that these factors are not seen to be mutually exclusive but rather interrelated with each other (p. 477). The single aspects are briefly explained as following:

1) The *technology* model is based upon the concepts of economies of scale and size, and the adaptation and diffusion of technology. The literature on economies of size has focused fundamentally on the long run cost curve in agricultural production and the determinants that shape and shift that curve (Boehlje 1990). The adaptation and diffusion of technology refers to the concept of Cochrane's treadmill (Cochrane 1958). The concept focuses on the impact of technological innovation reducing real per unit cost of output at the farm level and with competition encouraging farmers to adopt new technologies. The first adopters of the new technology will gain from the first-mover advantage (Bremmer et al. 2004), but as adoption becomes widespread, prices of farm commodities will fall differently per farm size, triggering structural adjustments (Ahearn et al. 2002). According to Weiss (1999) it is frequently argued that this process of technologically induced farm growth is stronger for larger farms.
2) *Off-farm employment* is handled in two ways. On the one hand, it is seen either as a first step out of the sector or as a possibility to stay in the sector by co-financing the agricultural business. As opportunity costs increase due to better wage levels outside of agriculture, farmers tend to leave the sector until wages equalize (Hallam 1991). Another way to achieve comparable incomes with the non-farm sectors would be to enlarge the farm size (Harrington et al. 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) or farmers are even willing to subsidize their small farm at least in the short-run from other income sources (Harrington et al. 1995). Thus, structural change would tend to favour part-time farming.
3) *Public programs* are governmental policies that impact the agricultural sector in different ways according to their design. Examples often mentioned are tax policy, commodity programs, credit programs, general monetary and fiscal policies, and public research and extension efforts (Harrington et al. 1995, Goddard et al. 1993, U.S. Congress 1985).
4) *Human capital* refers to and is influenced by the managerial capability, level of schooling, public education programs. It is assumed that an increase in human capital allows the firm manager to more effectively process information used to allocate the firm's resources and to evaluate new technologies. Thus, an increase in human capital allows for effectively managing an increasing firm size (Goddard et al. 1993). On the other hand Hallam (1991) argues that the costs of managing large operations are often underestimated.
5) *Demographics* refer mainly to the age structure of farm operators and the shrinking number of entrants to the farming sector. One might argue that these aspects are a consequence rather

than a cause of structural change. However, the speed of change in a region might be heavily influenced by the age structure of the farmers. Goddard et al. (1993) also point to the changes in the demographical structure of the general population that might have some influence concerning the demand of agricultural products.

6) *Market structure* itself influences structural change. This point is derived from the industrial organization structure (Boehlje 1990). The way in which prices are set is determined by the nature of the market, so that the conduct of the industry is a function of its structure (polypoly, oligopoly, monopoly vs. polypsony, oligopsony, monopsony). The development of institutional arrangements, such as vertical integration and cooperatives has an (unclear) impact on structural change as well (Goddard et al. 1993).

7) *Economic forces. S*ector specific and macroeconomic factors such as input and output prices, changes in demand, and the interest rate are also supposed to have an impact on structural change (Hallam 1991, Goddard et al. 1993).

2.3 Selected modelling approaches

Following several modelling approaches applied to the analysis of structural change in agriculture are explained (see table 1). Two main categories can be differentiated: econometric and simulation (programming) models.

The first one covers Markov chains, regression models, cohort analyses and models of discrete choice. Among these models, three basic methodologies are identified. The Markov chain approach tries to retrieve specific patterns of structural change from historical development and exploits these results to make forecasts into the future. Regression models are mainly used to identify the impact of specific variables on farm growth and abandonment. Finally, cohort analyses may help to separate demographical factors that contribute to structural change from economic factors. Models of discrete choice have mainly been applied to analyse farm succession.

Table 1. Modelling approaches to structural change

Model	**Field of analysis**
Markov chains	Number of farms in farm types as defined by the researcher
Regression models	Farm growth
Cohort analyses	Total number of farms, labour development
Discrete choice	Total number of farms, farm succession
Recursive-programming approaches	Changes within pre-defined farm groups
Multi-agent systems	Cover almost everything, it depends on the model specification

Within simulation models, recursive programming and multi-agent models are selected as more common methodologies. Whereas the first type of models analyses economic behaviour based on the decomposition of complex decision problems into sequences of smaller and simpler problems stepwise, multi-agent models consist on simultaneous interactions between autonomous entities (agents) in a specific environment (rules).

3 Econometric models

3.1 Markov chains

The concept of the Markov chain was first established by the Russian mathematician A. A. Markov (1856-1922) in 1906. For illustrative purposes in a paper published in 1913 Markov applied his chains to the distribution of vowels and consonants in A. S. Pushkin's poem "Eugeny Onegin". At present, much more important applications of Markov chains have been discovered (Basharin et al. 2004). Despite the early introduction of the basic concept of the Markov chain process, only recently economists have recognized its importance as an economic analysis tool. Today it is generally accepted and often applied. In agricultural economics Markov chains are primarily used to predict changes in the structure of the farm or agro-industry. With respect to animal health economics Markov chains are also used in bio-economic models to predict the spread of a disease. The basic concept of the Markov chains shall be explained here briefly before a review of applications of this technique is given. Finally, it is outlined which explanatory variables have been proven to be significant for the process of structural change in these studies.

3.1.1 Concept of the Markov chains

According to Hallberg (1969) the development of many economic phenomena over time can be described as stochastic processes. In this sense, Markov chains can be used to predict future developments of certain variables. In particular, for the analysis of structural change in one sector the estimation of Markov chains is an often used approach (Disney et al. 1988).

In a Markov process the movement of firms from a specific firm category (e.g. a farm type) to another one is represented by transition probabilities.

In case of a first order Markov chain 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:

$$P\{s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots\} = P\{s_t = j | s_{t-1} = i\} = p_{ij}$$

where s_t $\{1,2,\ldots,N\}$ is a discrete, stochastic variable and i and j are the states (farm types) a specific farm can be in. The transition probability p_{ij} represents the probability of a movement from state i to state j.

Since the probabilities are not allowed to be less than zero and the process has to result in some state it follows that: $\sum_j p_{ij} = 1$ and $p_{ij} \geq 0$ for $i, j \geq 0$.

The single transition probabilities can be summarized in a transition matrix P $(N \times N)$:

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

The equation to be estimated then becomes:

(3.1) $$p_{ij} = \frac{m_{ij}}{\sum_{j=1}^{n} m_{ij}}$$

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.

From the transition probabilities predictions on future farm numbers in any state can be easily calculated:

(3.2) $$X_t = X_0 P^t$$

where 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 X_t is the t[th] configuration vector.

One of the strongest assumptions of the Markov model is that the transition probabilities do not change in the whole process, 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 achieved. This may represent a realistic assumption as long as all other factors remain the same, too. However, this assumption does not hold for most economic phenomena. Changes in exogenous variables, e.g. wages, prices, technology or policy, require the determination of non-stationary (varying) transition probabilities (Hallberg 1969), which require an econometric model "behind" the pure Markov chain. The non-stationary transition probabilities are, therefore, specified as functions of exogenous variables and parameters.

3.1.2 Markov studies in agriculture (methodological development)

The following discussion on Markov chain applications in agriculture follows a (loose) chronological order to reflect the methodological development of the approach. An overview of Markov studies is given in table 2.. However, only studies that contributed to the evolution of the approach are described in more detail in this section. References to even more Markov applications can be also found in the literature listed in table 2 (see especially Stavins and Stanton (1980)).

One of the first applications of a Markov model in the agricultural sector was accomplished by Judge and Swanson (1961). Their analysis provided a thorough explanation of the Markov chain theory and was applied to a sample of 83 hog-producing firms in central Illinois in the period 1946-1958. The yearly change of the size distribution (number of litters of hogs) was analysed.

A further early Markov study by Stanton and Kettunen (1967) already emphasized the effect of how new entrants into the market are modelled. Usually, a "no production" category is formed, from which entrants are assumed to enter the market and to which farms that exit the sector move. In this study it is shown that the number of assumed potential entrants to the industry has an important effect on both (short-run) projections and equilibrium solutions. However, it will not affect the estimated proportions of active firms falling in each size category (p. 634). Stanton and Kettunen concluded that if there are no barriers to entry to the industry under study, then the number of potential entrants should be estimated to be large relative to the number of given firms. Otherwise, the use of a large number of potential entrants will increase the rate of change, both in net exits and entries and in shifts within size classes, more rapidly than it should.

The afore mentioned stationary models are all solved via maximum likelihood estimation. This is generally possible, if information concerning the movement of individual firms among different states is available. This information is generally called "micro-data" (Stavins and Stanton 1980) or "survey data" (Zepeda 1995a). However, often only aggregate data without detailed information tracing the movement of individual firms among size categories is available.

Telser (1963) was the first to estimate transition probabilities based on aggregate (macro-) data. He developed a methodology for using least squares techniques to estimate stationary transition probabilities from aggregate data:

(3.3) $$m_{jt} = \sum_i m_{i(t-1)} p_{ij} + v_{jt}$$

where m_{jt} is the observed proportion of individuals in state j at time t and $m_{i(t-1)}$ is the observed share in state i at time $t-1$[1]. v_{jt} is a random variable (error term). Least squares estimates of p_{ij} are obtained by choosing a set of estimates $\hat{p}_{ij}$ so to minimize $\sum_t \hat{v}_{jt}^2$ subject to the r constraints $\sum_j \hat{p}_{ij} = 1 \ (j = 1,\ldots,r)$, where

(3.4) $$\hat{v}_{jt} = m_{jt} - \sum_i m_{i(t-1)} \hat{p}_{ij}$$

Nevertheless, the use of unrestricted least squares does not impose non-negativity to single transition probabilities or exclude values greater than 1. However, Telser could show that the least-squares estimates are consistent and can be weighted so that they are also asymptotically efficient. Also, a non-stationary model was calculated by expressing the transition probabilities as functions of exogenous variables.

Telser illustrated his estimation technique by estimating the transition probabilities of smokers changing from one cigarette brand to another over the time period 1925-1943. Relative advertising expenditures were used as explanatory variable for the non-stationary model. Although not referring to the agricultural sector his study was included here to reflect the methodological development of the Markov chain approach.

Krenz (1964) estimated a stationary Markov model from aggregate data via maximum likelihood estimators. For that purpose he imposed a kind of "rule-of-thumb" method (Stavins and Stanton 1980, p. 13). The imposed constraints were that farms in the largest category remain in this category, increases in the number of farms in any state come from the next smaller state, and decreases in size are not allowed, since those farms were expected to go out of business (Krenz 1964, p. 78). Stavins and Stanton (1980) pointed to the theoretical limitations of this approach since the behaviour pattern for the farms that should be investigated is, in fact, already postulated beforehand. However, a similar approach was used later on by Keane (1976 and 1991).

Whereas, the early applications of Markov chains in agriculture dealt only with stationary transition probabilities, Hallberg (1969) was the first, who estimated non-stationary transition probabilities to model structural change. The non-stationary transition probabilities were estimated depending on some exogenous factors via least squares regression technique:

(3.5) $$\hat{p}_{ijt} = \hat{\alpha}_{ij} + \sum_{k=1}^{K} \hat{\beta}_{ijk} z_{kt}$$

with a set of exogenous variables z_{kt}, $k = 1,2,\ldots,K$. In order to meet the probability requirements (non-negativity and summing-up conditions), Hallberg estimates the n+nk parameters of a given row

[1] Note: whereas in the description of the Markov process, m_{jt} has been defined as 'number of movements between states', from now onwards m_{jt} refers to the 'share of farms in state j'.

in a single regression equation. Although Hallberg's restricted least squares approach ensures that all rows sum to unity, it does not deal directly with the constraint requiring that all single probabilities p_{ij} be greater than or equal to zero. If this was the case, Hallberg assumed negative probabilities to be zero, and probabilities greater than unity to be one. The approach was applied to plants manufacturing frozen milk products in Pennsylvania.

Stavins and Stanton (1980) estimated various models in order to find the best performing one. Among the estimated models there were a 'traditional' stationary micro-data model, a stationary macro-data model (of the Krenz type), and a micro-data non-stationary Markov model. In the last case, Stavins and Stanton refined Hallberg's approach and met the probability requirements without the use of *ad hoc* procedural assumptions. They specified the required equations such that each row in the transition probability matrix was handled as a separate multinomial logit model, using an exponential function to ensure that all predicted probability values would be positive and would sum to unity for each row. The multinomial logit model is described below and shall not be taken into account here.

Ethridge et al. (1985) estimated stationary as well as non-stationary transition probabilities for five size and four activity groups of cotton gin firms in West Texas and thus, were the first who incorporated another dependent variable than size class into their model. Industry structure projections were made for both the stationary and the non-stationary model. For the non-stationary model different scenarios with respect to the explanatory variables were imposed.

Disney et al. (1988) suggested quadratic programming methods or minimum absolute deviation (MAD) to avoid this problem. They used a minimum absolute deviation technique, where linear programming was used to minimize the sum of the absolute value of the deviations to calculate the transition probabilities. The hog-corn price ratio was included as explanatory variable:

$$\text{(3.6)} \qquad \min \sum_j \sum_t |f_{jt}| + \sum_j \sum_t |g_{jt}|$$

subject to the constraints:

$$m_{jt} = \sum_i m_{i(t-1)} p_{ij} + f_{jt} - g_{jt}$$

and

$$\sum_j p_{ij} = 1$$

where f_{jt} and g_{jt} are the positive and negative vertical deviations above and below the regression line for the set of observations. The hog-corn price ratio (HC) is included as explanatory variable by substituting p_{ij} with $a_{ij} + b_{ij}HC_t$. A similar approach was later used by Von Massow et al. (1992).

Chavas and Magand (1988) developed an approach to estimate the probability of entry/exit and the transition probabilities of the remaining firms (i.e. firm growth) separately. As this approach is also used by Zepeda (1995a), whose model is described below the model specification is left out here.

Zepeda (1995a) modelled the probability of net new entry separately from the transition probabilities of the existing firms. Thus, the explanatory variables affecting net new entries were allowed to be different than those affecting movements between existing firms and forcing net new entries to be proportional to the number of farms by size category is avoided. Both, net new entrants and the transition between states for existing firms were allowed to vary over time. To satisfy the non-negativity condition and force the transition probabilities to equal to unity, Zepeda used a multinomial logit model suggested by MacRae (1977). In her model of Wisconsin dairy farms, the following explanatory variables were included: the milk-feed price ratio for 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 only supposed to influence net new entries to the sector.

Zepeda (1995a) used a non-stationary Markov process to characterize size categories as a function of net new entries and movement between size categories suggested by Chavas and Magand (1988):

(3.7) $$n_{jt} = d_{jt} - h_{jt} + \sum_{i=1}^{s} p_{ijt} n_{i(t-1)}$$

Between $t-1$ and t the number of first-time entrants to each size category j, is defined by d_{jt} and the number of firms that leave farming is given by h_{jt}. The movement between size categories at time t equals the transition probability p_{ijt}, times the number of existing firms in category i in the last period, $n_{i(t-1)}$. Thus, the model structure is unlike many others since the transition probabilities pertain only to existing farms. Net new entries, a_{jt}, are given by subtracting exits h_{jt} from entrants d_{jt}: $a_{jt} = d_{jt} - h_{jt}$. Zepeda stated, that given that aggregate data on total farm numbers by size was used, the number of entrants and exits (i.e. net new entrants) were not observable. However, with her model, the corresponding estimates could be recovered.

Net new entrants and the transition probabilities for the movement between size classes can be generally specified as functions of explanatory factors $a_{jt} = f\lfloor Z_{\mathrm{j(t-1)}}, Z_{\mathrm{j(t-2)}}, \ldots \rfloor$ and $P_{ijt} = f\lfloor Z_{\mathrm{i(t-1)}}, Z_{\mathrm{i(t-2)}}, \ldots \rfloor$, where f denotes a general function of the independent variables, given by matrices Z and X. In the absence of a priori information on the functional form, in a first order Markov process net new entrants can be specified as

(3.8) $$a_{jt} = \gamma_{\mathrm{j}} Z_{\mathrm{j(t-1)}}$$

where the $\boldsymbol{\gamma}$'s are vectors of coefficients to be estimated.

To meet the probability requirements, the choice of functional form is limited (Maddala 1983). A multinomial logit model satisfies these conditions.

(3.9) $$P_{ijt} = \frac{\exp(X_{\mathrm{i(t-1)}}\beta_{\mathrm{ij}})}{1+\sum_{\mathrm{k=1}}^{\mathrm{s-1}} \exp(X_{\mathrm{i(t-1)}}\beta_{\mathrm{ik}})}, \; j = 1,\ldots,s-1,$$

(3.10) $$P_{ist} = \frac{1}{1+\sum_{k=1}^{s} \exp(X_{\mathrm{i(t-1)}}\beta_{\mathrm{ik}})}$$

where the β's are vectors of coefficients. Equation (3.9) measures the probability of movement between size category i and all the others apart from the last category. Equation (3.10) measures the probability of movement between each size category and the last category. The system to be estimated is given by combining the equations (3.7), (3.8), (3.9), (3.10) and adding error terms to the model. The model is estimated via a nonlinear seemingly unrelated regression estimator.

In a second application of the Markov chain approach Zepeda investigated the influence of technical change on the size distribution of dairy farms (Zepeda 1995b). Again a multinomial logit model was estimated. As explanatory variable only milk production per cow per year was used as proxy for technical change. Her results suggest that increases in the level of technology among continuing dairy farms enhance their ability to stay at the same size class versus growing in the short run. In the long run, however, the proportion of very large farms increases.

Karantininis (2002) used a maximum entropy formalism to estimate the non-stationary transition probabilities of the farm size distribution in Danish hog production which is shown below. As explanatory variables he used pork prices, pork feed prices, and input and output prices of other livestock. The output prices of other livestock were prices of milk, beef, eggs, and poultry meat, and the input prices were pig composite feeds, poultry and cattle feeds, fertilizer prices, and the interest rate. He first estimated a stationary Markov model that performed rather badly. In the first estimated non-stationary model it was found that the transition probabilities at the lower left off-diagonal were mostly non-zero (large farms are likely to decrease in size), which was introduced as prior information into a second non-stationary model. Furthermore, it was assumed that farms do not grow more than five size categories in each time period, and that it is most likely to remain in the same category as in the period before. This second non-stationary model revealed the best overall performance of the three Markov models. Karantininis concluded that the used generalised cross entropy (GCE) estimator was more efficient and could overcome many of the problems of traditional techniques, such as the ordinary least squares and multinomial logit (especially the problem of dimensionality). An additional advantage of the GCE approach is that it can deal with so-called ill-posed problems (large transition probability matrices, missing data points).

Jongeneel et al. (2005) were the first to conduct a cross-country analysis. In their analysis the development patterns of the dairy farm structure in the Netherlands, West and East Germany, Poland and Hungary were examined. The transition probabilities were calculated via maximum entropy estimation. The impact of each variable on the individual transition probabilities and size categories was like in Zepeda (1995a), Zepeda (1995b) and Karantininis (2002) evaluated in the form of impact elasticities. These “probability elasticities” measure the effect of a 1% change in the i^{th} explanatory variable on the probability that an existing farm will remain in the same size category or move to another size category in any period (Zepeda 1995b). Entry and exit to the industry were modelled by defining a “no production” category. However, no market entries were assumed for the dairy sector. Unlike Zepeda (1995b), who imposed as a restriction in her model that the transition probabilities might scale up or down in size only one size category, Jongeneel et al. (2005) included the size scaling as prior information that may be overruled by the data (p. 4). Country-specific policy (dummy) variables were used as explanatory variables for the different countries. Also, different size classes were imposed to reflect the differing structures in the countries analysed. For Hungary and Poland, subsistence farms (‘very low’ farm size classes) were taken into account, although they were likely to behave rather differently than conventional farms. Jongeneel et al. concluded that one option could be the exclusion of this special category from the analysis and to try to explain it in a different way, taking into account different explanatory variables, since several studies indicated that subsistence farming is rather isolated from and insensitive to market signals. The overall performance of the model seemed to be rather satisfactory, even though only few explanatory variables were taken into account (p. 14). However, it was also mentioned that the final estimated transition probabilities followed the prior information rather closely. Whereas, the use of prior information appeared to be crucial for obtaining plausible results, its impact in the cross entropy approach seemed to be so strong that the quality of the final estimates is to an important degree determined by the quality of the prior information. Jongeneel et al. suggested to look for other ways of including the prior information, in which it could be given less weight. They also mentioned the possibility of predicting future farm numbers in the different size categories by updating the Markov model with forecasts on the explanatory variables.

For illustrative purposes we will repeat here the maximum entropy approach chosen by Karantininis and Jongeneel et al. A stationary transition probability matrix (TPM) using generalised cross entropy (GCE) was developed by Lee and Judge (1996), and Golan et al. (1996). The transition between time t and t+1 in the stationary model can be formulated as follows:

$$(3.11) \qquad \mathbf{y}(\mathrm{t}+1)=\mathbf{x}'(\mathrm{t})\mathbf{P}+\mathbf{u}(\mathrm{t})$$

where $\mathbf{y}(t+1)$ is a KX1 vector of proportions falling in each of the K Markov states at time $t+1$, and $\mathbf{x}(t)$ are the sample proportions at time t. The TPM is $\mathbf{P} = (\mathbf{p}_1\, \mathbf{p}_2 \ldots \mathbf{p}_K)$ with each vector $\mathbf{p}'_k = (p_{1k}, p_{2k}, \ldots, p_{Kk})$. Finally, $\mathbf{u}(t)$ is a vector of disturbances with zero mean bounded within a specified support vector $\mathbf{v}$. For T transitions the model can be written more compactly:

$$\text{(3.12)} \qquad \mathbf{y}_T = (I_K \otimes X_T)\mathbf{p} + \mathbf{u}_T$$

$$(TK \times 1) = (TK \times K^2)(K^2 \times 1) + (TK \times 1)$$

where the TPM is now written as a vector: $\mathbf{P} = (\mathbf{p}'_1, \mathbf{p}'_2, \ldots, \mathbf{p}'_K)$, I_K is a $K \times K$ identity matrix and $\otimes$ denotes the Kronecker product. Each element of the $\mathbf{u}_T$ is parameterised as $u_{it} = \sum_m^M v_m w_{itm}$, where $\mathbf{w}$ is an M-dimensional vector of weights (in the form of probabilities) for each u_{it}, and $\mathbf{v}$ is an M-dimensional vector of supports. With $\mathbf{x}(t)$ being a vector of proportions, the support vector can be set to $(n_{it} \in [0,1])$ or to $\mathbf{v} = \left[-1/K\sqrt{T}, \ldots, 0, \ldots, 1/K\sqrt{T}\right]'$. By using GCE, any prior information about $\mathbf{P}$ can be incorporated in the form of a matrix of priors $\mathbf{Q}$. Prior information about the disturbance $\mathbf{u}_T$, call it w^o_{itm}, can be incorporated as well and are assumed to be uniformly symmetric about zero. The objective of the GCE estimator is to minimize the joint entropy distance between the data and the priors. Let H(•) be the measure of cross entropy, then the GCE is:

$$\text{(3.13)} \qquad \min_{\mathbf{p},\mathbf{w}} \left\{ H(\mathbf{P}, \mathbf{W}, \mathbf{Q}, \mathbf{W}^o) = \sum_i \sum_j p_{ij} \ln(p_{ij} / q_{ij}) + \sum_i \sum_t \sum_m w_{itm} \ln(w_{itm} / w^o_{itm}) \right\}$$

subject to the KXT data consistency constraints (equation (3.12)), the normalization constraints for both the transition probabilities (K constraints) and the error weights (KXT constraints): $\sum_j^K p_{ij} = 1, \sum_m^M w_{itm} = 1$; and the K^2 non-negativity constraints for $\mathbf{P}$ and the KXTXM constraints for $\mathbf{w}: \mathbf{P} \geq 0, \text{and } \mathbf{w} \geq 0$.

The non-stationary transition probabilities can be expressed as:

$$\text{(3.14)} \qquad p_{ij}(t) = f_{ij}(\mathbf{z}_{ij}(t), \beta_{ij}) + e_{ij}(t)$$

where $f_{ij}(\bullet)$ is a function relating each element $p_{ij}(t)$ of the non-stationary transition probability matrix (NSTPM) to a vector of explanatory variables $z_{ij}(t)$. The β_{ij} are parameters of the $f_{ij}(\bullet)$, and $e_{ij}(t)$ is the disturbance term. The Markov process can now be expressed as:

$$\text{(3.15)} \qquad \mathbf{y}(t+1) = \mathbf{x}'(t)\left[\boldsymbol{\beta}\, \mathbf{z}(t) + \mathbf{e}(t)\right] + \mathbf{u}(t),$$

where the disturbances $\mathbf{e}$ and $\mathbf{u}$ can be recovered separately.

Table 2. Markov chain analyses

Year	Analysis	Region	Focus	Data	Time period	Transition Probabilities	Methodology	Dependent Variables	Explanatory Variables
1961	Judge, G. G.; Swanson, E. R.	Illinois, USA	Hog production	Survey	1946-1958	Stationary	Maximum likelihood	Firm size (in number of litters of hogs)	
1962	Padberg, D. I.	California, USA	Wholesale fluid milk industry	Survey	1950, 1955, 1960	Stationary	Maximum likelihood	Firm size (in market shares)	
1964	Krenz, R. D.	North Dakota, USA	All farms	Aggregate	1935-1960	Stationary	Maximum likelihood	Firm size (in acres)	
1967	Stanton, B. F.; Kettunen, L.	New York, USA	Dairy farms	Survey	1960-1964	Stationary	Maximum likelihood	Firm size (in herd size)	
1969	Hallberg, M. C.	Pennsylvania, USA	Plants manufactoring frozen milk products	Survey	1944-1963	Non-stationary	Least squares	Firm size (in sales volume)	Wages, population, per capita income, farmgate price for milk, retail price
1976	Keane, M.	South of Ireland	Dairy farms	Aggregate	1968-1973	Stationary	Maximum likelihood	Firm size (in milk supply)	
1980	Stavins, R. N.; Stanton, B. F.	New York, USA	Dairy farms	Aggregate, Survey	1968-1977	Stationary, non-stationary	…, Multinomial logit	Firm size (in milk supply)	Milk-feed price ratio
1985	Ethridge, D. E. et al.	West Texas, USA	Cotton gin firms	Survey	1967-1979	Stationary, non-stationary	Maximum likelihood, least squares	Activity and size (in gin capacity)	Wages, energy costs, plant capacity, technical change
1985	Edwards, C. et al.	USA	All farms	Survey	1974-78	Stationary		Firm size (by acres, value of sales, tenure, standard industrial classification)	

Year	Analysis	Region	Focus	Data	Time period	Transition Probabilities	Methodology	Dependent Variables	Explanatory Variables
1987	Garcia, P. et al.	Illinois, USA	Cash grain farms	Survey	1976-1985	Stationary	Maximum likelihood	Firm size (gross value of farm product and tillable acres)	
1988	Disney, W. T. et al.	Southern states, USA	Hog production	Aggregate	1969-1982	Stationary, non-stationary	Minimum absolute deviation (MAD)	Firm size (in saled market hogs/year)	Hog-corn price ratio
1988	Chavas, J.-P.; Magand, G.	Different regions, USA	Dairy farms	Aggregate	1977-1984	Non-stationary	Multinomial logit	Entry/exit; firm size (in herd size)	Economies of size, sunk costs, market prices
1991	Keane, M.	Dairy co-operative society, Ireland	Dairy farms	Aggregate	1983-1989	Stationary	Maximum likelihood	Firm size (in milk supply)	
1992	Von Massow, M. et al.	Ontario, Canada	Hog production	Aggregate	1971-1989	Stationary, non-stationary	Minimization of median absolute deviation (MOMAD)	Firm size (in number of hogs marketed)	Hog-corn price ratio, interest rate, labor-capital price ratio
1995a	Zepeda, L.	Wisconsin, USA	Dairy farms	Aggregate	1972-1992	Non-stationary	Multinomial logit	Entry/exit; firm size (in herd size)	Milk-feed price ratio, interest rate, dairy termination program, debt, drought
1995b	Zepeda, L.	Wisconsin, USA	Dairy farms	Aggregate	1980-1992	Non-stationary	Multinomial logit	Firm size (in herd size)	Milk production per cow (proxy for technical change)
2002	Karantininis, K.	Denmark	Hog production	Aggregate	1984-1998	Non-stationary	Maximum entropy	Firm size	Input and output prices of pork and other livestock, fertilizer prices (as a proxy of energy costs), interest rate

Year	Analysis	Region	Focus	Data	Time period	Transition Probabilities	Methodology	Dependent Vari-ables	Explanatory Vari-ables
2005	Jongeneel, R. et al.	Netherlands, West Germany, East Germany, Poland, Hungary	Dairy farms	Aggregate	NL 1972-2003; WD 1971-2003; ED 1991-2003; PL 1996-2000; HU 2000, 2003	Non-stationary	Maximum entropy	Firm size (in herd size)	Technology shifter, level of aggregate milk production, dummy or dummy-trend variable indicating the switch in policy regime, farm gate price of milk

Source: Own table.

3.1.3 Explanatory power and significance of variables

Whereas in the section above the emphasis was put onto the development of the methodology, this section stresses the use of different exogenous variables and their significance for structural change as well as the overall explanatory power of the models.

Note that the studies differ with regard to several aspects (e.g. definition of structural change, modelling approach, time period under consideration, data quality and quantity, quantity of states assumed[2], etc.) and in fact cannot be compared directly with each other. Nevertheless, an overview of the exogenous variables assumed to affect structural change is given (table 3) and some conclusions are drawn with regard to the models' prediction ability.

Table 3. Significance of exogenous variables

Study	Year	Region	Focus	Time period	Dependent	Explanatory	Sign.
Chavas, J.-P.; Magand, G.	1988	USA	Dairy farms	1977-1984	Net entry; firm size	Economies of size Sunk costs Market prices	Yes Yes Yes
Disney, W. T. et al.	1988	Southern states, USA	Hog production	1969-1982	Farm size	Hog/corn price ratio	Yes
Ethridge, D. E. et al.	1985	West Texas, USA	Cotton gin firms	1967-1979	Activity and size	Wage rate Electricity rate Capacity Productivity Time Percentage of seedcotton ginned	Yes Yes No No Yes Yes
Hallberg, M. C.	1969	Pennsylvania, USA	Frozen milk product plants	1944-1963	Firm size	Wages Population Per capita income Milk price Retail price	Yes Yes Yes Yes Yes
Jongeneel, R. et al.	2005	Netherlands, West Germany, East Germany, Poland, Hungary	Dairy farms	NL 1972-2003; WD 1971-2003; ED 1991-2003; PL 1996-2000; HU 2000, 2003	Farm size	Technology shifter (trend) Milk production Policy dummy Milk price	Yes Yes Yes Yes
Karantininis, K.	2002	Denmark	Hog production	1984-1998	Firm size	Pork prices Milk price Egg price Cattle price Poultry price Pig feed Cattle feeds Poultry feeds Fertilizer prices Interest rate	Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes

[2] Between three (e.g. Zepeda 1995a) and nineteen (Karantininis 2001) states (i.e. farm types) were assumed in the mentioned studies.

Study	Year	Region	Focus	Time period	Dependent	Explanatory	Sign.
Stavins, R. N.; Stanton, B. F.	1980	New York, USA	Dairy farms	1968-1977	Firm size	Milk-feed price ratio	Yes
Von Massow, M. et al.	1992	Ontario, Canada	Hog production	1971-1989	Farm size	Hog/corn price ratio Interest rate Labour/capital price ratio	Yes Yes Yes
Zepeda, L.	1995a	Wisconsin, USA	Dairy farms	1972-1992	Net entry; firm size	Relative prices	Yes
					Net entry	Dairy termination program Drought Debt Interest rate	No Yes Yes Yes
Zepeda, L.	1995b	Wisconsin, USA	Dairy farms	1980-1992	Firm size	Milk production per cow	Yes

Source: Own table.

All explanatory variables used in the Markov studies relate more or less to the factors contributing to structural change outlined in section 2.2. Most often variables concerning technological change, economic factors like prices and interest rates, and policy variables have been taken into account. In no study appeared human capital or demographical aspects as explanatory variables. Only Zepeda (1995a) introduces a "new" variable, namely drought into her analysis.

It can be seen from table 3 that nearly all factors have proven to be significant for at least one of the size categories of each study. In the cases where no significant impact could be estimated reasons for this phenomenon were identified. Zepeda (1995a) argued that the Dairy Termination Program had no influence on the net entry of her model since Wisconsin, the state under study, had a relatively low participation rate in this program. Also, Ethridge et al. (1985) concluded that the capacity and productivity had no significant impact due to the fact that 3-year moving averages were taken for these variables and, in addition, that large changes occurred in neither of these variables during the period taken into account (p. 15).

Nonetheless, nearly all authors emphasised that their study most likely represents only a part of the problem and suggested to include more or other variables that might have some relevance.

This leads to the question of the overall performance (or 'explanatory power') of the examined models. The explanatory power of a model generally decreases with the prediction error, i.e. the difference between actual versus predicted outcomes of the dependent variable. The more can be explained of the variance of the outcomes, the better the fit of the model. The most common measure of the goodness of fit is the coefficient of determination, denoted by R^2 (ratio explained variance/total variance; $0 \leq R^2 \leq 1$). A zero value of R^2 indicates the poorest fit and a unit value the best fit that can be attained.

In table 3 the R^2's of the different Markov models are displayed. In some cases (depending on the estimation method) no 'traditional' R^2's, but so-called pseudo- R^2's (McFadden) were calculated. In general, the R^2's of the non-stationary models were quite good, whereas the stationary models (where calculated) performed rather badly (Karantininis (2002), Von Massow et al. (1992) and Hallberg (1969)). Only the stationary models estimated by Disney et al. (1988) reveal surprisingly good values for the R^2's (0.96 and 0.97). Similar values for the models estimated with the exogenous variables suggest that their influence is rather limited.

For both maximum entropy approaches (Karantininis (2002) and Jongeneel et al. (2005)) the incorporated prior information seems to considerably affect the overall quality of the model as indicated by the fact that the final estimates closely follow the prior information matrix.

Another measure for drawing conclusions regarding the explanatory power of a model is the calculation of mean squared prediction errors based on within sample forecasts where the transition probabilities estimated from the entire sample are used to predict the number (or proportion) of farms in the sample years. The root of the mean square prediction errors are then compared to the actual outcomes of the sample. These calculations were conducted by Hallberg (1969) who found a good prediction accuracy in accordance to the high values of the R^2's, Von Massow et al. (1992), very high prediction errors in the no production class, and Zepeda (1995a, 1995b), good performance in accordance with R^2.

An even better measure for the prediction capabilities of a model is given by out-of-sample predictions. Out-of-sample predictions can be used to show how well a model predicts into the future or outcomes in another system rather than the outcomes of the original data used to estimate the model.

Hallberg (1969) conducted out-of-sample predictions only for the stationary model. Thereby, estimates were calculated from different (early) time periods of the sample and then used to predict the later farm structure of the same sample. A comparison of the predicted values with the actual ones showed that none of the cases yielded reasonably good estimates. Stavins and Stanton (1980) found that the "simple regression specification" via least squares did not account for a very large proportion of the variation in the transition probabilities over time as indicated by low R^2's. Nonetheless, the estimated parameters of early sample years and the actual values of the exogenous variable were used for out-of-sample predictions. The predicted distribution showed approximately the correct shape.

Table 4. Explanatory power of the models

Study	Measure of fit	Performance	Comments
Chavas, J.-P.; Magand, G.	R^2	0.67-0.99	
Disney, W. T. et al.	R^2	0.94-0.97	4 models (with and without exit category, with and without explanatory variable), whereby the model without exit category but with explanatory variables performed best
Ethridge, D. E. et al.	R^2	0.32-0.72	Model performed well when sufficient observations were available, otherwise the results suffered from data limitations
Hallberg, M. C.	R^2 (non-stationary)	0.89-0.99	
	Within sample prediction (non-stationary)		Good performance
	Out-of-sample prediction (stationary)	No measure, graphical analysis in the original study	None of the cases yields reasonably good estimates for the 1963 structure; estimates better the shorter the forecasting time horizon
Jongeneel, R. et al.	Pseudo- R^2	0.82-0.93	The estimated matrix in all cases rather closely followed the prior matrix indicating that the quality of the final estimates is mainly determined by the quality of the prior information matrix
Karantininis, K.	Pseudo- R^2	0.07, 0.26, 0.49	Stationary, non-stationary I, non-stationary II (with incorporation of prior information from non-stationary I); non-stationary II closely followed the prior matrix
Stavins, R. N.; Stanton, B. F.	R^2	0.00-0.70	Values not comparable to other studies (variation refers to annual transition probabilities which could be calculated from the micro-data)
	Out-of-sample prediction	No measure, graphical evaluation in original study	Good performance
Von Massow, M.	Within sample	Stationary: 11-	Non-stationary models perform better than the stationary

Study	Measure of fit	Performance	Comments
et al.	prediction (root mean square percentage error)	33%, 63% (in-active class); Non-stationary (3 models): 9-20, 46-62% (inactive class)	one; each non-stationary model is estimated depending on one explanatory variable (hog-corn price ratio, interest rate, labour-capital ratio) of which the labour-capital ratio model performed best
Zepeda, L. (1995a)	R^2	0.9905-0.9986	
	Within sample prediction (pre-diction error in any year)	2.2-7.2% of farms	
Zepeda, L. (1995b)	R^2	0.88-0.99	
	Within sample prediction (pre-diction error in any year)	2-11% of farms	

Source: Own table.

Having defined a set of significant variables and established a good fit of the model, forecasts on the future farm structure can be made. Therefore, the transition probabilities need to be updated by fore-casts (or assumed values) of the explanatory variables. This is usually possible as long as e.g. policy changes and their influence on those variables are foreseeable. Thus, it prevails that the shorter the time horizon under consideration, the better the forecasting ability of the estimated model.

3.2 Regression models

These models are characterised through regressions on a number of explanatory variables. Most of the models try to explain firm growth/size or focus especially on entry and exit of firms from the sector. Since there is a vast amount of studies like this, only a few ones shall be taken into account here. 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's law states that the growth rate of firms is determined by random factors independent of size.

The basic equation to test Gibrat's law is:

$$\ln S_{it} - \ln S_{it-1} = \alpha + \beta \ln S_{it-1} + u_{it} .$$

Where S_{it} is the size of firm i at time t, and u_{it} 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 subsumed within the random process. Not all of the models mentioned below explicitly refer to Gibrat's law, in some only the effects of a number of explanatory variables other than firm size on structural change are tested.

Usually, not only farm growth, but also farm entry and exit (farm survival), are taken into account in the analyses based on Gibrat's law as those contribute substantially to structural change in the farm sector (Weiss 1999). Moreover, ignoring farm exit in the analysis would result in a problem of sample selection bias. In order to measure farm growth, farm size must be compared between two specific points in time. However, measures of farm growth are meaningful only for surviving farms. Farms exiting between the points in time over which growth is measured are normally excluded from the sample (as non-surviving farms) (Kostov et al. 2005). Thus, declining small farms are less likely to remain as survivors than declining large farms since smaller farms will hit a critical minimum farm size much sooner than larger farms with the same (negative) growth rate. Hence, growth rates estimated only on survivors will be biased towards finding relatively lower growth rates for the larger holdings (Weiss 1999). One possible solution to circumvent the problem of sample selection bias is a two-step estimation procedure, as devised by Heckman (1979). In step 1, the probabilities of farm survival are estimated from the complete sample (including surviving and non-surviving farms). These probabilities are subsequently used to obtain an additional variable which is introduced as a correction factor for the estimation of the growth model in step 2, where the estimation of the growth model is based upon a sample including only the surviving farms. In both steps other explanatory variables can be included. A discussion of this and other methods to solve the problem of sample selection bias can be found in Kostov et al. (2005).

Shapiro et al. (1987) tested the relationship between farm size and growth in Canada from 1966 until 1981. They found out that small farms seem to grow faster than large farms, what implies the rejection of Gibrat's law. Larger farms also experience more stable growth rates in comparison to small farms. Shapiro et al. also found 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) also 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 in the model 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 founds out 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. Furthermore, he determines two turning points which suggest a polarization of growth rates: small and very large farms grow faster than farms in the middle-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 was predicted correctly, whereas the occurrence of growth and renewal was overall 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.

Two similar studies on the U.S. farm sector focus on the explanation of productivity, farm size, and part-time farming (Evenson and Huffman 1997 and Ahearn et al. 2002). Among the explanatory variables used we can find prices, governmental policies, technology, human capital, infrastructure and climatic and geographical variables (Evenson and Huffman 1997, Ahearn et al. 2002). Variables on research and extension are used as proxies for technological change. Infrastructure is represented by access to highways and represents a new aspect in comparison to the variables treated in chapter 2 introduced by Ahearn et al. (2002). Evenson and Huffman found that from 1950 until 1982 changes in farm size in the US are dominated by input price changes rather than by technology or government programs. The Evenson and Huffman model fits well in the sense of having a system R-square of 0.70 and a large share of the estimated structural coefficients significantly different from zero. According to Ahearn et al., an increase in highway infrastructure has a significantly positive effect on agricultural productivity and off-farm work. Furthermore, Ahearn et al. could show that governmental policies have important effects on agricultural productivity and farm structure. The direction of these effects depends on the program involved, e.g. an increase in government commodity program payments increases agricultural productivity and farm size but has no significant effect on off-farm employment. Also, the set-aside acres of land that were diverted from production as a requirement of commodity program participation have a significant and negative impact on productivity. Ahearn et al. suggest that farm size and off-farm work are substitutes. An increase in off-farm work by the farmer would reduce productivity and farm size. A large share of the estimated coefficients in their model was significantly different from zero and the share of variation explained is quite good, 62% for the productivity, 71% for the size, and 63% for the off-farm participation equation. In table 5 only the significant variables explaining firm size are listed.

Sumner et al. (1987) analysed effects of human capital on size and growth. Their study relates to a sample of southern dairy farms in the US. 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 (see section 3.3) 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.

Table 5 summarises the mentioned analyses including a column defining whether the explanatory variables were significant to the study.

Table 5. Significance of Variables in the Regression Analyses

Analysis	Year	Region	Focus	Time period	Dependent	Explanatory	Sign.
Ahearn et al.	2002	USA	All farms	1960-1996	Firm size, productivity, part-time farming	Productivity	Yes
						Off-farm employment	Yes
						Research	Yes
						Extension	Yes
						Specialization	Yes
						Commodity payments	Yes
						Price ratio machinery / hired farm labour	Yes
						Geoclimatic region	Yes
Bremmer et al.	2004	Netherlands	Arable farming and horticulture	1990-2000	Farm renewal, firm growth	Farmer's age	No
						Succession	No
						Off-farm employment	No
						Firm size	Yes
						Family labour input	Yes
						Solvency	Yes
						Mechanization	Yes
						Profitability	No
Evenson and Huffman	1997	USA	All farms	1950-1982	Firm size, productivity, part-time farming	Specialization	Yes
						Off-farm employment	Yes
						Public research	Yes
						Private research	Yes
						Public extension	Yes
						Schooling	No
						Wage (in manufacturing)	Yes
						Wage ratio (hired farm labour/manufacturing)	Yes
						Price ratio machinery / hired farm labour	Yes
						Price ratio fertilizer / hired farm labour	No
						Government price support (for crop and milk)	Yes
						Government crop diversion payments	Yes
						Trend	Yes
						Geoclimatic region	Yes
Shapiro et al.	1987	Canada	All farms	1966-1981	Firm size	Firm size	Yes
Sumner and Leiby	1987	Southern USA	Dairy farms	1982, 1977, 1987	Firm size and growth	Farmer's age	Yes
						Experience	Yes
						Schooling	Yes
						Management	Yes
Weiss	1999	Upper Austria	All farms	1980-1990	Entry, exit, firm growth	Farm size	Yes
						Human capital	Yes
						Off-farm employment	Yes
						Farmer's age	Yes
						Farmer's family status	Yes

Source: Own table.

In summary, Shapiro et al. (1987) and Weiss (1999) rejected Gibrat's law, since small farms grow faster than large farms.

3.3 Cohort analyses

Farmers of a certain gender and occupational category (full-time, part-time, hired, family) belonging to a cohort are defined by specifying the period during which they were born. Their number can be followed and simulated through time by cohort analysis (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).

Age cohort analyses in agriculture are often used to predict labour developments (De Haen and Von Braun 1977, Pavel 1997, Bauer 2000). Few are explicitly used to predict future farm numbers (Steele and Gaffney 1998). 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 (De Haen and Von Braun 1977). 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 (Steele and Gaffney 1998). 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) and Pavel (1997) found out that for the work force decrease in West Germany a considerable part (about 60 %) are due to age, death, and disability.

The basic equation for an age cohort analysis is:

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

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 net-autonomous change of the cohort size.

Generally, 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.

3.4 Models of discrete choice

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 for discrete choice analyses 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 formulae (Pietola et al, 2006):

$$\begin{gathered} y_i{}^* = \alpha + \beta z_i + u_i \\ where\ \ y_i = 1\ if\ \ y_i{}^* > 0,\ else\ y_i = 0 \end{gathered}.$$

$y_i{}^*$ 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) method. As explanatory variables served 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. Four different R^2-based measures revealed values between 50 and 80 per cent.

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 (1974) 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 (Eckstein and Wolpin, 1989; McFadden, 1989; Pakes and Pollard, 1989; Keane, 1993) 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 expiry of an early retirement programme, and a dummy for the existence of a spouse. However, some parameters associated with prices and subsidies were not significant at the five percent level.

4 Simulation models

4.1 Recursive programming approach

Day et al. (1978, p.2) define recursive programming as "a general approach to modelling economic behaviour based on the decomposition of large, complex decision problems into sequences of smaller, simpler decision problems conditioned by past decisions and observed changes in the decision-maker's environment". This implies that in a recursive programming model small (decomposed) problems are stepwise solved until the whole system is optimised.

According to Happe (2004, p. 25) these models can be looked at as ancestors of the multi-agent approach in the sense, that farm agents are assumed to represent whole farm types or farming regions. The farm agents are heterogeneous with respect to factor capacities, technical coefficients and the definition of the objective function in the set of linear equations underlying each farm agent. A thorough explanation of the approach as well as several applications and further references can be found in Day et al. (1978).

These models will be, however, not described here in further detail since usually only changes within certain farm types (e.g. acreage, labour use, prices) are examined, but the number of farms in a farm type remains the same.

4.2 Multi-agent systems

When speaking about multi-agent systems one first may define what exactly an agent is. Although there is much debate on the correct definition of agenthood, we will follow the definition of Jennings (1999) that states: "an agent is an encapsulated computer system that is situated in some environment, and that is capable of flexible, autonomous action in that environment in order to meet its design objectives".

An agent-based model, therefore, consists of autonomous decision-making entities (agents), an environment through which agents interact, rules that define the relationship between agents and their environment, and rules that determine sequencing of actions in the model.

Autonomous agents are composed of rules that translate both internal and external information into internal states, decisions or actions. According to Happe (2004), agent-based models can include both economic agents (e.g. farms, markets) and agents that can represent other social and environmental institutions (e.g. politicians, non-farm agents, land). After having defined the initial attributes of agents and objects (characteristics, behavioural rules, internally stored information about the agent itself and other agents), the structure evolves over time without further intervention by the modeller (Happe 2004). Agent-based models are usually implemented as multi-agent systems, a concept originated in the computer sciences that allows for an efficient design of large and interconnected computer programs. A set of global equilibrium conditions is not employed in these models, in contrast to modelling techniques such as conventional mathematical programming or econometrics (Parker et al. 2001a, p.1).

The implemented decision rules result in a particular property of agent-based systems that is called self-organisation. Self-organisation means the ability of multi-agent systems to generate complex structures that change endogenously, or 'from within'. In the same manner the speed of change is determined from within and not set externally (Happe 2004, p. 21).

For agricultural sector modelling especially agent-based models in combination with land-use/land-cover change models (LUCC) are of interest. According to Parker et al. (2001a, p.1), an agent-based model of LUCC consists of two key components. The first is a cellular model that represents the landscape under study. This cellular model may draw on a number of specific spatial modelling techniques, such as cellular automata or spatial diffusion models. The second component is an agent-based model that represents human decision making and interactions.

A thorough literature review of multi-agent systems in agriculture is provided by Happe (2004). For an overview of agent-based models see also Berger and Parker (2001, pp. 27) as well as Parker et al. (2001b, pp.79).

In the field of structural change mainly Balmann and Happe pioneered the work with agent-based approaches, whereby e.g. Balmann (1994) could demonstrate the existence of path dependencies in agricultural structures with the use of a cellular automaton. Happe (2004) and Freeman (2005) provide a detailed insight into the methodology. The model used by Happe, AgriPolis, is calibrated to the small German region Hohenlohe. In her study, two types of agents, farm agents and market agents are modelled. In her model one farm agent corresponds to one farm. The objective function of every farm agent is assumed to be a farm household income maximisation. The farm agents, thus, maximise total household income (farm and off-farm income) under specific constraints: farm factor endowments (land, labour, fixed assets, liquidity), the situation on input and output markets, overall framework conditions (opportunities for off-farm employment, interest rate levels, access to credit) and the political framework conditions. The farm agent is able to react to changes in its environment and its own state by adjusting its organisation in response to available factors endowments and observable actions of other farm agents (Happe et al. 2004). Production planning, investment, and the decision to continue or quit farming are based on expectations about future developments of prices, costs, technologies, investment possibilities and policies. Farm agents can undertake various actions and show a specific behaviour. They differ mainly with respect to specialisation, farm size, factor endowment, production technology, personal characteristics of the farmer and managerial ability (no learning imposed in the current version). Agent interactions occur only indirectly by competing on factor and product markets. The most important actions undertaken by a farm agent are renting land (renting additional land and disposing of unprofitable land), investment, production, farm accounting, and the decision whether to quit farming or to stay in the sector. The decisions to be made by a farm agent are also summarized in figure 1.

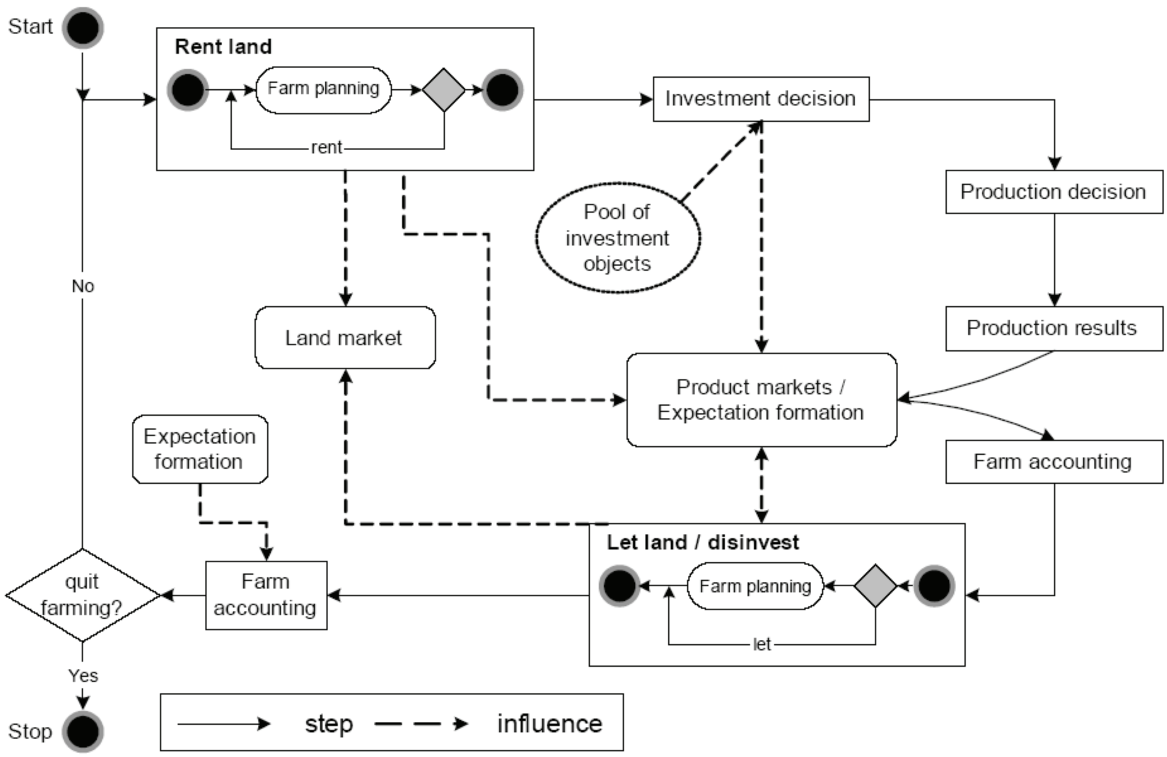
The market agent coordinates the behavioural response of markets. It is his responsibility to bring together supply and demand (products or production factors) and to determine a price for the netput. More specifically, in AgriPoliS, there is a land market agent, the auctioneer, and a product market agent. Unlike the farm agent that meets all the criteria mentioned in the agent definition above, in this case the market agents can only be considered as very basic agents, whose sole objective is to coordinate the actions of farm agents on the markets for products, land, capital, and labour (Happe et al. 2004). The spatial dimension of AgriPolis is represented by usage of a cellular automaton. Typical farms are modelled and up-scaled to represent the regional characteristics of the test region. Farm typologies depend on the farm type (professional vs. non-professional), specialization and size.

Happe uses sensitivity analysis as a form of validation where varying input factors are used to prove whether the effects agree with prior knowledge about the system. However, Happe argues that sensitivity analysis remains a difficult task when dealing with complex simulation models like AgriPoliS. She uses the statistical techniques of Design of Experiments (DOE) and meta-modelling which provide a way of carrying out simulation experiments systematically and take account for parameter interactions (p. 92). Experimental design provides a way to investigate some aspects of a simulation model systematically and to bring statistical aspects into the analysis of results. It represents a way to understand relationships between some parameters in the model.
After having simulated a number of possible parameter constellations (for discussion see Happe) the simulation results are analysed by applying a so-called meta-model which establishes a functional relationship between sensitivity and various factors. The meta-model is defined as a regression model where the independent variables are factor levels and the dependent variable is the simulation response. Eventually, the meta-model is validated by determining the degree to which the meta-model represents the underlying simulation model correctly. This can be done either by running additional simulation scenarios and comparing results with meta-model predictions, or by analysing residuals.

Technological change, interest rate, managerial ability, shadow prices on land, and the size of the region are selected by Happe for the DOE analysis in the reference scenario as key factors for structural change (policy settings of the Agenda 2000). The results are presented in a graphical analysis and by the outcome of the meta-model. The meta-model achieves an adjusted R^2 of 0.995. The results show that as expected technological change, interest rate levels and the managerial ability have an impact on structural change as shown by their influence on the average economic land rent. Particular emphasis is given to the managerial ability of the farmer by exploring the effect of heterogeneous managerial skills. In terms of modelling, this factor is introduced by altering the agent's costs, such that a farm agent with high (low) managerial ability is assumed to produce at lower (higher) costs. According to their managerial ability, farm agents respond quite differently to different interest rate levels which were used as proxy for overall economic framework conditions.

Two sets of policies are analysed with regard to structural change. The first group comprises three fundamentally different policies, a retirement payment scheme, fully decoupled payments and a stepwise phasing out of direct payments. Each of these policies aims at facilitating structural adjustment. The second set of policies is inspired by the policy debate on decoupling direct payments in the Common Agricultural Policy. Three ways of decoupling direct payments are analysed with respect to their impact on structural adjustment. These policies are compared with respect to their impact on structural change, the pace of structural change and their impact on factor use, farm size, incomes, efficiency and governmental expenses.

Figure 1. Events for a farm agent in one planning period



Source: Happe 2004, p. 54.

Freeman (2005) analyses the impact of agricultural stabilization and support programmes of the past few decades in Canada. The basic characteristics of his modelling approach are quite similar to the ones of Happe (2004) (e.g. farm agents interact through the land market). Unlike Happe, who considers all types of specialisations in her model, Freeman represents only crop producers. Farms differ with respect to operator's age, farm size, tenure, finance and management. The focus of the study lies

on farm expansion and contraction, whereas in AgriPoliS also information on specialisation classes is available.
Freeman assumes, like Happe, a significant impact of managerial attributes on the evolution of regional agricultural structures which is established by implementing four distinct managerial classes with correspondingly different risk preferences. Furthermore, a strong demographical component was established by assuming three life phases which are characterised by different levels of risk aversion: an entry/establishment phase, a development/growth phase, and an exit/retirement phase. Whereas AgriPoliS is calibrated to the base year 2000/2001 with the political framework conditions of the Agenda 2000, Freeman's model is adapted to the year 1960 such that only past development is replicated. The spatial and producer profile is initialized to represent an agricultural region typical of one that would be found in the dark brown soil zone of Saskatchewan in the year 1960.
Major policy programs for the simulation period 1960-2000 were incorporated such that a validation through direct comparison of the simulation results to available census and survey data was possible. In addition, a number of drivers of structural change were discussed in the context of the simulation results.
Two scenarios are analysed, a base scenario with the policy settings that actually have taken place in the simulation period and a zero transfer scenario that simulates the structural evolution of the idealized study region in the absence of any government interventions through stabilization programs and ad-hoc stabilization payments. Results for both scenarios are reported and compared to each other for the total number of farm units, the mean farm size, the distribution of the farm size, land values, the proportion of farm land leased from non farming owners, and farm debt. Results show that in the base scenario the model was able to correctly replicate the observed structural developments of the study region. By comparing the results for both the base and zero transfer scenarios, the net results of the package of stabilization programs and ad-hoc payments could be estimated. However, while the transfers certainly have had an impact the particular consequences remain somehow unclear in comparison to their original intention.
The drivers of structural change which impact on the model are analysed in more detail: the entrepreneurial behaviour and farm household expectations, the cost of the production and productive efficiency, path dependency and the farm life-cycle, and government transfers. Freeman found out that economies of scale are not necessary to replicate the structural shifts that occurred over the period studied. Instead, managerial ability proves to be an important factor in the sustainability and growth of an individual farm. Freeman concludes that "finally, it appears that opportunity and luck often play a more important role than individual ability. In fact, it is possible that a less productive farm agent may succeed while a more productive farm agent may fail, and this is due to the immobility of farmland. (p. 104)"

In the following section advantages and disadvantages of the agent-based approach will be briefly discussed.

An advantage of the agent-based modelling approach is the greater flexibility in comparison to other approaches. This flexibility mainly arises from the bottom-up way of modelling, which is a typical characteristic of the approach. The greater flexibility is on the one hand due to flexible behavioural foundation on the individual level (e.g. bounded rationality, heterogeneous objectives and abilities) and on the other hand to the modelling of flexible general frameworks and conditions (e.g. imperfect markets, bilateral or multilateral exchange relationships) (Happe 2004, Balmann et al. 2001). According to Happe (2004), it is possible to implement axiomatic assumptions given by theory, but other assumptions can also be made specific to the problem to be studied.
A second advantage that arises from the bottom-up approach refers to the execution of computational experiments. This is what Berger et al. (2001) refer to as 'computational laboratory'. Agent-based models are able to serve as computational laboratories that allow for thought experiments and may structure the exploration of dynamic interactions (Berger et al. 2001, p. 9). The computational laboratory property arises from the property of self-organization that is inherent to the agent-based system and is defined by spontaneous order and emergence, and the endogenous change of structures (Balmann et al. 2001).

A third advantage lies in the possibility of interdisciplinary modelling by linking human decision-making processes with models of biophysical processes (Berger et al. 2001, p. 10). At least theoretically, this opens the path to simulate the impact of technological innovations on structural change. It also includes the opportunity of integrating spatial aspects. Particularly in agriculture, land use takes a central position since spatial aspects have a direct effect on farm decision-making and on the economics of the farm. The link between economic and spatial models may thus support a better understanding of interdependencies between agent behaviour and space in land use systems (Happe 2004).

Disadvantages are closely interrelated with the advantages described before. The first refers to the point flexibility and may be called complexity. The greater flexibility with respect to assumptions (e.g. different objective functions for the agents, market imperfections) requires the modeller to choose assumptions quite carefully. According to Happe (2004), there is a risk of choosing the "wrong" assumptions. Another critical point made is the number of assumptions that are modelled (Happe 2004). Couclelis 2001 expresses the danger of over-specification by stating: "Agent-based modelling meets an intuitive desire to explicitly represent human decision making when modelling systems where we know for a fact that human decision making plays a major role. However, by doing so, the well-known problems of modelling a highly complex, dynamic spatial environment are compounded by the problems of modelling highly complex, dynamic decision-making units interacting with that environment and among themselves in highly complex, dynamic ways (Couclelis 2001, p. 4)".

Another disadvantage is given by the data requirements if the system shall be calibrated to a specific region with real-world data. According to Happe (2004) this concerns the accessibility of individual farm accountancy data which often is available only for aggregates of selected farm variables.

In general, the ability of agent-based models to track observed structural change has not been shown in any meaningful way up to date. Another shortcoming is the current lack of methodologies to statistically validate the responses of the system to changes in institutional or economic conditions. These shortcomings could be partially addressed by reflecting behavioural and parameter uncertainty through simulations under different specifications allowing to represent the plausible range of model outcomes. Another opportunity might be the use of meta-models (Happe 2004) to describe the relationship between parameter settings and model results using regression techniques. This approach might lead to some sort of statistical validation of agent-based models in the future by comparing outcomes of these meta-models with similar reduced form models estimated using observed structural change.

5 Conclusions

With Markov chains and multi-agent models, two methods could be identified as generally suitable for the problem to be solved here, i.e. the prediction of future numbers of farms in certain farm type classes. While regression and cohort analyses provide insights to potentially relevant variables explaining structural change, they mainly focus on differences in farm-household decisions related to socio-demographic variables. They do not take into account the interaction between farms leading to the aggregate outcome, i.e. the structural change. Markov models in combination with a simulation model like CAPRI delivering forecasts on explanatory variables as well as agent-based systems are suited to simulate changes in the agricultural structure for certain policy scenarios. Due to the explicit modelling of decision making, spatial interaction, and technology, multi-agent systems represent a theoretically interesting approach in the SEAMLESS context. Thereby, a significant advantage arises from the opportunity to simulate future development paths under conditions not observed before ("computational laboratory" allowing, for example, to directly simulating impacts of technological innovations). However, the complexity to be modelled, the data requirements that arise from this complexity, and the current absence of any validation procedures prevent its usage in a pan-European context. The Markov chain approach by contrast appears to be more suitable in this case, due to the fact that it is less demanding in terms of data and offers a straightforward approach of handling the problem. However, the envisaged dimension of such an approach with respect to regions and farm types goes considerably beyond what has been done in the literature to this day.

From the literature a number of variables could be identified that theoretically impact on farm structure. These are the seven aspects discussed in chapter 2.2 (technology, off-farm employment, government programs, human capital, demographics, market structure, and other economic forces). Sometimes climatic variables or geographical information is included.

In the Markov chains only part of these variables are used to date. Among these are mainly variables concerning technological change, government programs, and other economic forces. In the regression analyses a large range of variables is employed with emphasis put on socio-demographic variables, whereas the cohort analyses focus mainly on demographical factors. The agent-based models described implicitly cover the whole range of variables with the emphasis put on heterogeneity of human capital and technology.

However, the studies all identified statistically significant impacts of exogenous variables, which include a set of economic indicators such as prices and profitability of enterprises. The set of included variables differs substantially between the different models depending on regional and farm type focus.

From the discussion above implications for the Markov chain estimation within SEAMLESS-IF can be derived. These are mainly given through the limited variables used in earlier Markov models, their limited regional coverage, and the predominately focussed definition of the farm typology. Thus, besides finding an adequate specific methodology, major challenges and possibilities emerge from the dimension to be modelled and can be summarised as follows:

1) To include a broader number of explanatory variables, particularly with respect to demographical developments, but also heterogeneity in technology and management which proved to be important in agent-based models.
2) To extend the approach to a meaningful, cross-sectional analysis across NUTS 2 regions in EU-15.
3) To deal with a broad typology according to the SEAMLESS definitions, comprising not only size categories but also different specialisations. The additional inclusion of spatial differentiation and intensity classes within regions is however deemed impossible. Time series information on spatially referenced farms is not available and the complexity is already considerable and unprecedented without this addition.

The data requirements for the Markov analysis seem to preclude the application of the approach to the new member states. An alternative methodology need to be explored, either by extrapolating the results of the statistical analysis for the EU-15 to EU-10, or by looking into opportunities to define less complex agent-based simulation models potentially allowing to at least identify tendencies in policy or technology impacts on structural change.

Further deliverables in task 3.6, activity 3.6.5 foresee a closer look at the data by defining the number of farms in the typology and detecting past development paths (PD3.6.7) and first estimation results from the Markov chain analysis (PD3.6.10).

References

Ahearn, M., Yee, J., Huffman, W., 2002. R&D, Productivity, and Structural Change in U.S. Agriculture, 1960-1996. National Bureau of Economic Research Summer Institute, Conference on Research in Income and Wealth, Cambridge, Massachusetts.

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

Balmann, A., 1994. Pfadabhängigkeiten in Agrarstrukturentwicklungen. Duncker and Humbolt, Berlin.

Balmann, A., 1997. Farm Based Modelling of Regional Structural Change. European Review of Agricultural Economics 25 (1), 85-108.

Balmann, A., Happe, K., Kellermann, K., Kleingarn, A., 2001. Adjustment Costs of Agri-Environmental Policy Switchings. 7th International Conference of the Society for Computational Economics, June 28-29.

Basharin, G.P., Langville, A.N., Naumov, V.A., 2004. The Life and Work of A. A. Markov. Linear Algebra and its Applications 386, 3-26.

Bauer, K., 2000. Modelling Agricultural Labour Response to Market Developments during Transition. IDARA - Working Paper.

Berger, T., Parker, D.C., 2001. Examples of Specific Research. In: Parker, D.C., Berger, T., Manson, S.M. (Eds.), Agent-Based Models of Land-Use and Land-Cover Change. LUCC Report Series No. 6, Irvine, California, USA, pp. 27-35.

Boehlje, M., 1990. Alternative Models of Structure Change in Agriculture and Related Industries. University of Minnesota, Department of Agricultural and Applied Economics, Staff Paper Series P40-41.

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. Strategic Decision Making and Firm Development in Dutch Horticulture, 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.

Couclelis, H., 2001. Why I No Longer Work with Agents: A Challenge for ABMs of Human-Environment Interactions. In: Parker, D.C., Berger, T., Manson, S.M. (Eds.), Agent-Based Models of Land-Use and Land-Cover Change. LUCC Report Series No. 6, Irvine, California, USA, pp. 3-5.

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.

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.

Eckstein, Z., Wolpin, K.I., 1989. The Specification and Estimation of Dynamic Stochastic Discrete Choice Models. Journal of Human Resources 24, 562-598.

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.

Evenson, R.E., Huffman, W.B., 1997. Long-Run Structural and Productivity Change in U.S. Agriculture: Effects of Prices and Policies. Yale University, Economic Growth Center, Center Discussion Paper 773.

Freeman, T.R., 2005. From the Ground Up: An Agent-Based Model of Regional Structural Change, University of Saskatchewan, Saskatchewan, Canada.

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.

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.

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. (Eds.), 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.

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

Happe, K., Balmann, A., Kellermann, K., 2004. The Agricultural Policy Simulator (AgriPoliS) - An Agent-Based Model to study Structural Change in Agriculture. IAMO - Institut für Agrarentwicklung in Mittel- und Osteuropa, IAMO, Discussion Paper 71.

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

Heckman, J.J., 1979. Sample Selection Bias as a Specification Error. Econometrica 47 (1), 153-162.

Jennings, N.R., 1999. Agent-Based Computing: Promise and Perils. University of London, Department of Electronic Engineering, Queen Mary & Westfield College.

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. XIth International Congress of the EAAE, Copenhagen, Denmark.

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

Karantininis, K., 2002. Information Based Estimators for the Non-stationary Transition Probability Matrix: An Application to the Danish Pork Industry. Journal of Econometrics 107, pp. 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.

Keane, M.P., 1993. Simulation Estimation for Panel Data Models with Limited Dependent Variables. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics, Vol. 11. Elsevier, Amsterdam, pp. 545-571.

Kim, C., Schluter, G., Schaible, G.M.A., Hallahan, C., 2005. A Decomposed Negative Binomial Model of Structural Change: A Theoretical and Empirical Application to U.S. Agriculture. Canadian Journal of Agricultural Economics 53, 161-176.

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

Knutson, Penn and Boehm, 1990. Agricultural and Food Policy, 2nd Ed.Prentice Hall.

Kostov, P., Patton, M., Moss, J., McErlean, S., 2005. Does Gibrat's Law Hold amongst Dairy Farmers in Northern Ireland? XIth International Congress of the EAAE, Copenhagen, Denmark.

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 & Sons, Inc., New York, pp. 153-164.

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

Maddala, G.S., 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge.

McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Analysis. In: Zarembka, P. (ed.), Frontiers in Econometrics. Academic Press, New York.

McFadden, D., 1989. Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration. Econometrica 57, 995-1026.

Oehmke, J., Schimmelpfennig, D.E., 2004. Quantifiying Structural Change in U.S. Agriculture: The Case of Research and Productivity. Journal of Productivity Analysis 21, 297-315.

Olson, K.D., Stanton, B.F., 1993. Projections of Structural Change and the Future of American Agriculture. In: Hallam, A. (Eds.), 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.

Pakes, A., Pollard, J., 1989. Simulation and the Asymptotics of Optimization Estimators. Econometrica 57, 1027-1057.

Parker, D.C., Berger, T., 2001b. Synthesis and Discussion. Agent-Based Models of Land-Use and Land-Cover Change. Irvine, California, USA, pp. 79-88.

Parker, D.C., Berger, T., Manson, S.M., 2001a. Agent-Based Models of Land-Use and Land-Cover Change. Report and Review of an International Workshop, LUCC Report Series No. 6.

Pavel, F., 1997. Analyse der Bestandsentwicklung landwirtschaftlicher Familienarbeitskräfte in Westdeutschland. University of Bonn, Agricultural and Resource Economics.

Penn, J.B., 1979. The Structure of Agriculture: An Overview of the Issue, Structural Issues of American Agriculture, ESCS, USDA, Report 438.

Pietola, K., Väre, M., Oude Lansink, A., 2003. Timing and Type of Exit from Farming: Farmer's Early Retirement Programmes in Finland. European Review of Agricultural Economics 30 (1), 99-116.

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

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

Stanton, B.F., 1993. Farm Structure: Concept and Definition. In: Hallam, A. (Eds.), Size, Structure, and the Changing Face of American Agriculture. Westview Press, Boulder, Colorado, pp. 14-29.

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.

Steele, S.R., Gaffney, P., 1998. A Regional Analysis of the Changing Structure of Agricultural Land Holdings. University College, CAPRI - Working Paper.

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. (Eds.), Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld. Stanford University Press, Stanford, pp. 270-292.

Tweeten, L., 1984 Causes and Consequences of Structural Change in the Farming Industry Washington D.C.:National Planning Association, Food and Agriculture Committee, Report 207.

U.S. Congress, O.o.T.A., 1985. Factors Contributing to Structural Change in Agriculture. In: U.S. Congress, O.o.T.A. (Eds.), 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.

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.

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.

Glossary

Structural change	There are a lot of different ways to define structural change. For our purpose structural change is best defined as the movement of farms between certain farm types including exit and entry (e.g. size classes, specialisation categories).
Markov chain	Methodology describing movement of elements across classes over time with transition probabilities. Often used to model structural change in an industry.
Cohort analysis	Concept of modelling labour use in agriculture over time depending on age structure and related exit probabilities.
Multi-agent system	System explicitly representing interaction of multiple agents.