ERSA 39th European Congress 1999 Dublin

# A comparison of approaches to spatially explicit modelling of land use/cover change

R.A. GROENEVELD, E.C. Van Ierland, W.J.M. Heijman

Wageningen Agricultural University
Department of Economics and Management
Environmental Economics Group
P.O. Box 8130
6700 EW Wageningen
The Netherlands

**Abstract**

The process of land use and land cover change is closely related to a wide variety of socio-economic and environmental issues. Functions of land such as food production, housing, industry and habitat are in permanent competition, resulting in an optimal or suboptimal allocation of land use. At the same time, certain land use types are associated with the issue of climate change through carbon sequestration and the emission of trace gasses. Researchers in these issues have developed several approaches to modelling the land use allocation process. As each approach has its own benefits and its specific goal, these approaches need to be given a position in the wide field of land use modelling.

This paper describes and compares a number of approaches to the modelling of agricultural land use. A number of economic approaches are described, such as general equilibrium models, discrete choice models and linear programming models and models based on heuristic decision rules. The approaches are compared based on criteria of theoretical foundation and data requirements.

Many models tend to be of a highly detailed level, while appropriate data for parameterisation of the model is often not available. We will therefore demonstrate and compare two estimation/modelling approaches with an existing data set that is representative for many spatial data sets based on census data.

## 1. Introduction

Land use allocation is a complex process in which many biophysical and socio-economic factors play a role. On one hand there is the natural environment, including factors like soil fertility, precipitation and temperature, which determines local potential crop or cattle production or benefits from other land use types. On the other hand, socio-economic factors like crop prices, income, government policy, employment and technological development play a role, that influences the potential gains from alternative land uses. It is clear that modelling and analysing such a complicated system requires understanding of both natural and social sciences.

The issue of land use change has been studied with regard to topics such as deforestation (Angelsen and Kaimowitz, 1997;Kaimowitz and Angelsen, 1998;Chomitz and Gray, 1996), development policy (Schipper et al., 1998), natural resource management (Van De Putte, 1988), land use planning (Schotten et al., 1997) and environmental pollution (Jones and O'Neill, 1993). With regard to global warming, the models IMAGE (Alcamo et al., 1994) and CLUE (Veldkamp and Fresco, 1996) have been developed. IMAGE (*I*ntegrated *M*odel to *A*ssess the *G*reenhouse *E*ffect) is a multidisciplinary, integrated model to simulate the dynamics of the global society-biosphere-climate system. The CLUE model (*C*onversion of *L*and *U*se and its *E*ffects) has so far been used to analyse land use/cover change on a national or regional scale, for various countries and regions in the world.

Because of the importance of socio-economic factors, economic theory can offer a significant contribution to understanding and modelling the process of land use and cover change. Economic modelling uses causal relations between economic parameters (prices, technology, demand, etc.) to construct a consistent and reliable representation of the society under investigation. Socio-economic factors play an important role in the process of land use and land cover change. Spatially explicit economic modelling can take account of socio-economic factors explicitly and might be able to improve the understanding of land use/cover change.

*Aim of this study*

This paper explores and compares approaches to spatial explicit modelling of land use/cover change based on economic theory. The study is of limited extent and seeks to further integrate biophysical and economic aspects.

This research aim involves a number of research questions. First, how does economic theory deal with the subject of land use/cover change? How are spatial patterns of land use and their changes explained? Second, what approaches exist in economics for the modelling and explanation of land use/cover change? Third, how can economic theory on land use/cover change be incorporated in a spatially explicit modelling framework? And fourth, what are the advantages and disadvantages of the alternative approaches to land use modelling?

*Methodology*

Answering these questions requires a collection of previous studies and existing models of land use/cover change, but also the application of modelling approaches in an empirical context. The first step in the research was to examine the literature on land use modelling and spatial economics. The goal of this step was to explore and collect different approaches to land use modelling, and to compare these approaches on their characteristics.

In many study areas, the only data available are census data. Unfortunately, these data tend to be highly aggregated, so that detailed models need additional information in order to estimate the parameters. Therefore, model parameters have been estimated by means of an existing data set that is based on census data of Java. The empirical application followed two different approaches, namely aggregated production function estimation and parameter estimation by means of maximum entropy econometrics. Each approach has its own benefits and drawbacks: in both cases, models have to be adapted to the data structure and/or the estimation method. Once all parameters are estimated the question is whether the models produce realistic results. The model parameterisations have been used to simulate the process of land use allocation in Java. The results of the simulations and parameterisations offer appropriate material for comparison and evaluation of the two modelling/parameterisation approaches and to indicate possibilities for further development.

## 2. Results of the literature study

### 2.1 Introduction

This chapter discusses the modelling of land use/cover change based on economic theory and describes earlier studies in spatial economic modelling. It starts with a general discussion on the economic approach. From there, a number of earlier studies are described. In the last section of the chapter, the approaches to land use modelling are evaluated.

## 2.2 The economic approach to land use modelling

According to Hazell & Norton (1986), a sector model contains, implicitly or explicitly, five elements. The same five elements can be found in models of land use/cover change also: (1) a description of producer's economic behaviour; (2) a description of currently available and potential production functions, or technology sets, now and/or in the future; (3) a definition of the resource endowments held by each group of producers; (4) a specification of the factor and product markets and (5) a specification of the policy environment. One can add two elements to this list: (6) a specification of the time scale and (7) a specification of the spatial scale.

In most economic models, one assumes that economic subjects, i.e. producers and consumers, are optimisers, i.e. they maximise or minimise a given variable under given restrictions. This can imply profit maximisation or cost minimisation, but also the maximisation of utility, in which there can exist a trade-off between welfare and leisure.

Production functions can take the form of a continuous relation or of a set of production techniques. One important feature of the production function is whether the returns to scale are either diminishing, constant or increasing. When more than one production factor is used in a production function (e.g. when production is a function of labour and pesticides), the substitution elasticity between production factors becomes important.

The most important resource in land use modelling is land. Land can differ in quality, depending on a number of local circumstances, such as distance from markets (factor markets and product markets) and suitability for crop production. Land (or rather location) quality is also related to other resource endowments, such as labour force (population density) and fresh water.

Regarding to the specification of factor and product markets, the easiest thing to do in a model is to assume constant factor and product prices. This assumption allows a simple specification of the prices and avoids the existence of feedback mechanisms in the model. However, some models (like general equilibrium models) take several markets into account.

Government policy can influence an economy through instruments such as taxes, quotas and legal restrictions. In the case of land use modelling the designation of nature reserves is important.

A model can be either static or dynamic. A static model considers only one moment in time, while a dynamic model considers more moments at the same time. A comparative static model is run separately for a number of periods, where in each period one or more than one parameters are altered.

This specification refers to the number of regions or locations considered in the model and whether the topology is included. The distinction between regions and locations as spatial units is not strict. In this study a region is seen as an area, whose boundary is determined by the relative homogeneity of the area within. A location is a mere point on a map. As the number of points on a map is infinite, the points are aggregated to a finite number of grid cells.

In many land use models, several regions or locations have been taken into account and their topology is also included. The topology reveals itself in features such as distance from markets and natural resources, and local endowments of resources. External effects from other regions, e.g. air pollution, also have a spatial dimension and can be included in the topology.

## 2.3 Classification of studies

To be able to compare approaches to land use modelling, it is necessary to make some kind of classification. Four categories have been distinguished in this study: (1) Optimisation models, (2) General equilibrium models, (3) Discrete choice models and (4) Models based on heuristic decision rules.

Although many models are solved by means of optimisation techniques, optimisation models have been distinguished separately, as many optimisation models are more complex and developed primarily for scenario analysis. These models can be used to determine an optimal policy, but an optimisation model can also be developed for the analysis of ‘what-if’ scenarios. These optimisation models are based on the assumption that economic subjects (such as farmers and consumers) show optimising behaviour, and that the market comes close enough to the ideal of perfect competition to allow negligence of institutional issues. The approach requires that production costs and production relations be known. In the case of profit maximisation, prices of crops must also be known.

Unlike the aforementioned optimisation models, general equilibrium models take several markets into account instead of only the allocation of land use. Product markets, such as crop markets, and factor markets, such as the labour market, are included in the model.

Discrete choice models have originally been developed for the analysis and prediction of choices of individuals between mutually exclusive alternatives. The discrete choice method estimates per alternative the probability that a given individual will choose it. In land use modelling, a farmer on a small spot of land has also several alternatives (and let’s assume that because of the size of his spot of land, the alternatives are mutually exclusive). His choice can also be analysed by means of discrete choice modelling. If his set of probabilities is transposed to a higher scale (i.e., the model takes account of a large number of farmers), the distribution of crops is expected to be equivalent to the probabilities set.

Instead of mathematical calculations, heuristic decision rules can be used to make a model operative. A recent example is the cellular automata approach (Engelen et al., 1995), in which rules are applied that determine land use in a given location based on surrounding land use types.

## 2.4 Brief description of existing models

This chapter discusses earlier studies in land use modelling. In Appendix A the studies are compared with regard to features such as model category, time scale and spatial scale.

### 2.4.1 An optimisation model: The NERC-ESRC Land Use Programme (NELUP)

The Natural Environment Research Council-Economic and Social Research Council Land Use Programme (Moxey et al., 1995a;Moxey et al., 1995b;O'Callaghan, 1995) consists of three quantitative models, each describing a separate part of the process of land use: a group of hydrological models, an ecological model and a regional agricultural economic model.

In Moxey et al. (Moxey et al., 1995a), the economic model is applied to the catchment area of the River Tyne in Northern England. The model is written as a linear programming problem. Profit is maximised for a single-macro farm that represents all farms in the catchment area under restrictions of production possibilities, resource availability and a given rate of adjustment. The level of the adjustment coefficients represents the largest feasible change between two years for an enterprise level.

### 2.4.2 A general equilibrium model: The LUC model of Fischer et al.

Fischer et al. (1996) describes a comprehensive general equilibrium model of land-use and land-cover change dynamics, based on welfare analysis.

Supply is represented following a nested approach. The model distinguishes three aggregate sectors: agriculture, forestry and other. Aggregate sectors are divided in sectors, which are divided in sub-sectors. Sub-sectors consist of products. Some sub-sectors produce the same products, such as biofuel and other energy sources, producing power.

The representation of agricultural supply is based on a combination of non-linear optimisation and the revenue function approach. Supply is driven by profit maximisation on the level of representative firms. Demand is described by demand systems, which are mathematically derived from a micro-economic utility maximisation problem.

The study region is subdivided into compartments, reflecting structured entities, i.e., sub-systems, of the broader region under consideration and their economic and other interactions. As geographical data sets are mostly organised on rectangular grids, compartments are defined as collections of grid cells, and can possibly vary over time. Depending on scale, a compartment may correspond to a collection of farms, to a watershed, a zone in a country, or a group of provinces.

Compartments interact through commodity trade and financial markets, and flows of mobile resources and pollutants. They compete for allocation of limited public resources and foreign investment and are jointly affected by government policies. Human migration may generate demographic flows.

Land resources are described by site classes that are defined in terms of intrinsic land properties, such as temperature regime, moisture regime, land accessibility, etc. A location can be transferred between site classes by land improvement, land degradation or climate change.

Land-use is described in a nested way. At the highest level, major land uses are defined. Within each of the major land-use classes, several land uses are described by a list of land-use classes. Two processes of land-use/cover change are distinguished: *land conversion,* which is a transfer between major land uses, and *land modification,* which is a transfer between land-use types within a major land-use.

### 2.4.3 Discrete choice models

**A Spatial Model of Land Use in Belize**

Chomitz & Gray (1996;Chomitz and Gray, 1995) describes a multinomial logit model based on the classic Von Thünen model, that is applied to the issue of deforestation in Belize. The model is to a large extent similar to that of McMillen (McMillen, 1989), but its derivation from the Von Thünen model makes it interesting enough to be described here.

In the classical Von Thünen model, we have seen that the potential rent $R$ associated with devoting plot $i$ to use or commodity $k$ is[1]

$$R_{ik} = P_{ik}Q_{ik} - C_{ik}X_{ik} \tag{2.1}$$

where $P_{ik}$ denotes the output price, $C_{ik}$ denotes a vector of input prices and $X_{ik}$ denotes the optimal quantities of inputs and $Q_{ik}$ the potential output of $k$ at point $i$. Unfortunately, $P$, $C$, and $Q$ are unobserved. However, in some cases determining factors of price and productivity are observed and therefore a reduced-form model can be formulated. Therefore, following Von Thünen, prices and costs are determined by the distance from the market:

$$P_{ik} = e^{g_{0k} + g_{1k} D_i} \tag{2.2}$$

$$C_{ik} = e^{d_{0k} + d_{1k} D_i} \tag{2.3}$$

where output prices are assumed to decrease with distance ($g_{1k} < 0$) and input costs to increase ($d_{1k} > 0$). The production function for use $k$ is assumed to be a Cobb-Douglas function, which depends on $X_{ik}$ and a parameter $S_{ik}$ which is the product of agroclimatic and other variables. From these

[1] Assuming a static framework

equations, a loglinear equation is derived, which relates the potential rent to observed parameters such as distance and other biophysical parameters:

$$\ln R_{ik} = a_{0k} + a_{1k} D_i + a_{2k} \ln(s_{1i}) + \ldots + u_{ik} \equiv \mathbf{Z}_i \mathbf{A}_k + u_{ik} \tag{2.4}$$

where **Z** is the vector of independent variables and **A** is a vector of reduced form parameters. To estimate the model, we assume that land is devoted to the highest-rent use: point $i$ is devoted to use $k$ if

$$R_{ik} > R_{ij}, \; \forall\, j \neq k$$

If the disturbances are Weibull distributed and uncorrelated across uses $j$, then this equation is equivalent to a multinomial logit model in which the probabilities that plot $i$ is devoted to use $k$ is

$$\text{Prob}(i \text{ devoted to } k) = \frac{e^{\mathbf{Z}_i \mathbf{A}_k}}{\sum_j e^{\mathbf{Z}_i \mathbf{A}_j}} \tag{2.5}$$

The multinomial logit model allows us to estimate the coefficients in equation (2.4) provided that the coefficients of one use - for example, natural vegetation - are normalised to zero.

**The spatial hedonic model of Geoghegan et al.**

Although not strictly a logit model, the hedonic approach (Geoghegan et al., 1997;Bockstael, 1996), can be categorised as a discrete choice model. In this approach land use change is driven by changes in land prices for a given land use type. A Markov matrix is formulated of transition between land use types in which the probabilities of land use conversion are interpreted as discrete choice probabilities, as it is assumed that the choice of land use is mutually exclusive.

In the simplest characterisation of the problem, a parcel of land or cell in the landscape, denoted $j$, which is currently in state $a$, is converted to state $i$ at time $t$ if

$$W_{jit|a} - C_{jit|a} \geq W_{jmt|a} - C_{jmt|a}$$

for all land uses $m = 1, \ldots, M$ (including $a$). We define $W_{jit|a}$ as the present value of the future stream of returns to parcel $j$ in state $i$ at time $t$, given that the parcel was in state $a$ in time $t$-1, and $C_{jit|a}$ as the cost of converting the parcel from state $a$ to state $i$ (which will be 0 if $a = i$).

Not all factors affecting $W$ and $C$ are observable. Therefore, $W$ - $C$ is rewritten in a systematic portion $V$ and a random portion $h$. The model is estimated in two steps. In the first step, the value of land in alternative uses is estimated. The second step is the estimation of a probability of a given parcel being converted, conditional on its value in alternative uses and its costs of conversion.

### 2.4.4 Models based on heuristic decision rules

**Models based on cellular automata**

Cellular automata are examples of mathematical systems constructed from many identical components, each simple, but together capable of complex behaviour (Wolfram, 1984). A cellular automaton consists of an array of cells in which each cell can assume one of $k$ discrete states at any one time. Time progresses in discrete steps, and all cells change state simultaneously as a function of their own state, together with the state of the cells in their neighbourhood, in accordance with a specified set of transition rules (Engelen et al., 1995).

In Engelen et al. (1995) the concept of cellular automata is applied in exploring the impact of climate change on a small island. The philosophy behind the cellular automata approach is that the effects of land-use/cover change drivers, even macro-scale drivers such as climate change, are actually expressed at the micro-scale level.

*Application to a small island*
In exploring the impact of climate change on a small island, the modelling framework consists of two linked components: one for macro-level processes and one for those operating at micro-level.

At the macro-level, the modelling framework integrates several component sub-models, representing the natural, social and economic sub-systems. These are all linked to each other in a network of mutual, reciprocal influence.

The land demand module takes the growth coefficients calculated by the macro-level model and returns the amount of additional space required to carry out the corresponding activities. The total area of land required by each activity drives the micro-level part of the model.

At the micro-level, land-use change is calculated by means of cellular automata. In this case, the neighbourhood is a circular template of 113 cells. Each cell in the grid is in one of thirteen states, each representing a land-use. The suitability of a cell depends on aggregate, distance weighted push and pull effects of all the cells in the neighbourhood (locational suitability) and on its own physical, environmental and institutional characteristics.

**The Land Cover Model in IMAGE**
IMAGE (Integrated Model to Assess the Greenhouse Effect; see Alcamo et al. (1994)) is a modelling framework consisting of a number of models, that is developed to investigate linkages and feedbacks in the global society-biosphere-climate system, and to evaluate consequences of climate policies. The framework includes three systems, an Energy-Industry system, an Atmosphere-Ocean system and a Terrestrial Environment system. Thirteen world regions are considered, of which only three regions refer to countries: other regions refer to groups of countries.

The Land Cover Model in IMAGE (Zuidema et al., 1994) distinguishes only three land use types: agricultural land, where crops are grown; range land, where cattle is kept and exploited forest, where fuelwood is grown. The model is demand-driven: agricultural demand is the main driver of land use change. For crops and animal products, demand is calculated by the Agricultural Demand Model in the Terrestrial Environment system. For fuelwood, demand is calculated by the Energy-Economy Model in the Energy-Industry system. It is assumed that wood is only used as fuel.

A Terrestrial Vegetation Model in the Terrestrial Environment system translates local biophysical circumstances into potential crop productivity. By means of heuristic land use rules, potential productivity and land use demand are reconciled. There are eight land use rules:

1. Hierarchy of land use demands: (1) agricultural land (2) range land (3) exploited forest;
2. Agricultural land expands only when current land cannot satisfy demands;
3. New agricultural land is allocated adjacent to current agricultural land;
4. New agricultural land is allocated to land with highest pot. productivity;
5. Grassland expands only if it is replaced elsewhere by agricultural land, or if the number of animals in the region increases;
6. New grassland is allocated adjacent to current agricultural land, grassland or savanna;
7. Urban fuelwood demand in Africa, India, South Asia and East Asia is satisfied by clearing existing forests;
8. Agricultural land taken out of production will revert to its climate-potential land cover.

The allocation procedure starts with the allocation of agricultural land. Once agricultural land is allocated, range land is allocated according to the land use rules. Finally, fuelwood is allocated.

## 2.5 Evaluation of existing models

### 2.5.1 Introduction

This section compares the modelling approaches described in Section 2.4 by means of two criteria, namely the theoretical consistency and the data requirements of the approach.

As a model is a representation of reality, one can ask the question to what extent the mathematical functions in the model represent mechanisms in reality. Suppose a model uses a linear function to describe the influence of $x$ on $y$, can one expect $x$ to have a linear effect on $y$ in reality also? The extent to which this is true, is in this report referred to as the *theoretical consistency*. One could also describe this criterion as the level of causality in the model.

Unfortunately, theoretical consistency often implies a detailed model that needs a large amount of data before it can be calibrated. Collecting this data can be very difficult, if not impossible. Therefore it important to take data requirements into account as well as theoretical consistency.

### 2.5.2 Optimisation models

Like all models, optimisation models are based on a number of assumptions, that can be in accordance with reality to a small or large extent. Typical assumptions in optimisation models are optimising behaviour of producers and consumers, and that markets will translate all individual decisions to a global optimum. It will depend on the area under consideration whether these assumptions can be made.

Compared to general equilibrium models, many optimisation models assume that prices are constant. Strictly speaking, prices are hardly ever constant in reality, unless their level is enforced by law: there is always *some* fluctuation. However, in a small study area, whose influence on national prices can be neglected, the assumption of fixed prices may be justified.

Essential information in the optimisation approach to land use modelling is the relation between inputs and outputs. All models are based on production decisions: to find an optimal allocation of crop production under given restrictions. For this purpose, one should be able to calculate the crop production level under given supply of nutrients, moisture, temperature, etc. Unfortunately, this information is not always available, and in the countries examined by CLUE (Costa Rica, Honduras, China, Java) it is not.

### 2.5.3 General equilibrium models

General equilibrium models can be viewed as the ultimate application of neoclassical theory. In that sense, general equilibrium is a theoretically sound approach, better than optimisation modelling, as it is supported by the dominant school of economics. However, there can also be objections to it. As general equilibrium modelling is closely related to optimisation modelling, these objections are similar to those to optimisation modelling. Most general equilibrium models assume optimising behavour of individuals, full availability of information and perfectly competitive markets (i.e. no oligopolies, no government intervention). This might not be the case: individuals do not always optimise their profit or utility, information is seldomly fully available, markets can be dominated by one company, etcetera.

The theoretical consistency of general equilibrium models is in great contrast to the calibration possibilities: general equilibrium modelling requires a large amount of data, of which many data are not available in an empirical setting. This makes that general equilibrium models are difficult to apply for specific empirical studies on land use and land cover change.

### 2.5.4 Discrete choice models

It is possible to formulate discrete choice models in such a way that the assumptions behind them are reasonable. One can assume that the chance of conversion to a given land use type depends on the expected income from that land use type relative to the other land use types, as defined by the logit or the probit equation.

Formulating a discrete choice model in the theoretically soundest way (i.e. assuming that land use conversion depends on expected income) requires parameters that are not easily estimated. On the other hand, restricting the analysis to data that is better available will decrease the theoretical base of the approach. For Geoghegan's (Geoghegan et al., 1997) hedonic approach, parcel values are needed.

### 2.5.5 Heuristic decision rules

In one way, heuristic decision rules are propositions and assumptions, and can therefore be interpreted as a theory of itself. In another way, optimisation is also a decision rule, although it is not as heuristic as the *if...then* type of rules in the approach of Zuidema et al. (1994).

One can wonder to what extent this type of decision rules is a realistic representation of land use allocation, because empirical validation of these rules is not easy. In many cases the economic justification for the applied allocation rules is not very strong and rather 'mechanic' allocation rules are applied. The assumptions behind the IMAGE procedure, represented by the land use rules, are explicit, and can therefore easily be evaluated. However, it remains difficult to assess their reliability for long run scenario studies. The assumptions behind the cellular automata approach are less clear.

Both heuristic approaches in this study use demand projections on one hand, suitability estimations on the other hand and use their decision rules to abstract a land use pattern from these two parameters. The heuristic approach seems very flexible in terms of the type of data it requires, but still a sufficient amount of information (quantitative or qualitative) will be necessary to construct a reliable model. It seems that *a priori* information on the mechanisms in land use allocations is essential for this approach.

## 3. Parameterisation of land use models

### 3.1 Introduction

The models described in the previous section are able to simulate the process of land use and cover change at a highly detailed level. This level of detail has the disadvantage that data needed for parameterisation of the models are often not available. In many areas, the most detailed information comes from census data. These data tend to be highly aggregated, which implies that highly detailed models cannot be estimated. For example, crop production and total use of production factors are known, but not the production factor use per crop type.

This chapter estimates the parameters of two non-linear models by means of an existing data set based on census data from Java. The data set used is representative of most spatial data sets based on census data and maps of biophysical parameters. Inputs are not known per crop: only total inputs and crop output are known.

Before the estimation of parameters biophysical clusters of cells have been constructed, based on major biophysical parameters like temperature, precipitation and elevation, in order to distinguish between areas of different biophysical character. Within these clusters, two other estimation methods have been used. The first method estimates a production function where all crops are aggregated into one production parameter. In the second method, separate crop production functions have been estimated by means of an estimation method that is especially developed for parameter estimation from ill-posed or ill-defined data sets.

### 3.2 Construction of biophysical clusters

By statistical clustering techniques, n-dimensional observations (e.g., temperature, altitude and such) can be summarised in a one-dimensional classification, where the within-class variation is supposed to be small compared to the between-class variation. In this study four quantitative biophysical parameters have been used for the construction of biophysical clusters: (1) cloud cover in percentage of time; (2) total precipitation in mm; (3) average temperature in degrees Celsius; (4) mean altitude (as calculated by the Digital Elevation Model[2]) in m.

These parameters have been summarised by means of the FASTCLUS procedure in SAS (SAS Institute Inc., 1989). The FASTCLUS procedure constructs clusters by minimising the euclidian distance between observations and so-called cluster seeds. Cluster seeds can be considered as the centres of the clusters. As FASTCLUS uses the data without any preparation, it is highly recommended to standardise the data before starting the procedure as the parameters are not measured in the same units. Therefore, the parameters are first standardised so that all parameters have an average value of 0 and a standard deviation of 1.

*The clusters*

Table 3-1 shows the number of cells, the average value of the parameters and the standard deviation of the parameters per cluster.

| Cluster | # cells | Cloudiness | Precipitation | Temperature | Altitude |
|---|---|---|---|---|---|
| 1. | 79 | -0.83 | 0.05 | -0.98 | 0.96 |
| | | 0.51 | 0.98 | 0.50 | 0.66 |
| 2. | 17 | -2.24 | -0.07 | -2.62 | 2.70 |
| | | 0.66 | 0.92 | 0.50 | 0.59 |
| 3. | 149 | 0.81 | -0.64 | 0.64 | -0.59 |
| | | 0.56 | 0.51 | 0.39 | 0.30 |
| 4. | 84 | -0.20 | 1.10 | 0.32 | -0.41 |
| | | 0.40 | 0.69 | 0.48 | 0.42 |

**Table** Error! Style not defined.-Error! Bookmark not defined.**:** Cluster means of the four-cluster classification

One cluster (2) has an extremely high value for altitude and very low temperature and cloudiness. Cluster 1 can be interpreted as an area between the highest and the lowest parts of Java, with also low temperature and cloudiness, but not as low as in cluster 2. Clusters 3 and 4 have roughly the same altitude, but there are some differences in cloudiness and precipitation: cluster 3 is cloudier and has more precipitation than cluster 4.

Figure 3-1 gives a spatial presentation of the clusters on Java.

[2] The Digital Elevation Model (DEM) is a digital representation of a continuous variable over a two- dimensional surface by a regular array of z values referenced to a common datum. Its precise structure is beyond the scope of this study.

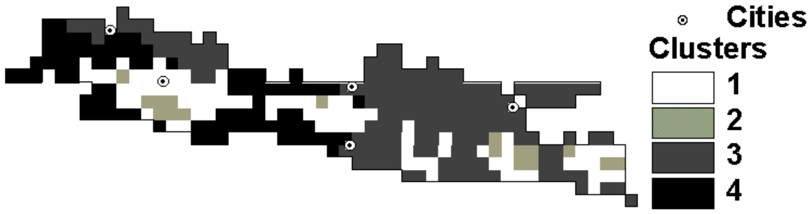


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Clusters in Java

## 3.3 Aggregated production functions

One way of estimating production function parameters from aggregate data is to aggregate crop production instead of estimating the distribution of inputs. In that case there is only one variable for output (which consists of the separate crops) and a number of variables for inputs. In this situation the function can be estimated by straightforward techniques such as OLS.

How should we aggregate separate crops? We could sum the physical weight of crop production, which would in some way imply that farmers strive for the production of biomass. But a more sensible assumption would be that farmers strive for the production of income. Therefore, the aggregate output is defined as the sum of crop production times the price of that specific crop.

### 3.3.1 Functions to be estimated

Prices used in the aggregation of crop production are shown in Table 3-2.

| Crop | Price in Rp (1994) |
|---|---|
| Cassava | 180.33 |
| Maize | 344.42 |
| Soybean | 1159.15 |
| Rice | 390.90 |
| Groundnut | 2058.26 |

**Table** Error! Style not defined.-Error! Bookmark not defined.**:** Crop prices (Anonymous, 1996)

We relate total aggregate production to total inputs:

$$Q_b = \alpha_b \prod_f^4 I_{fb}^{\alpha_{fb}} \tag{3.1}$$

To estimate this function with OLS, we take the logarythm of both sides:

$$\ln Q_b = \ln \alpha_b + \sum_f^4 \alpha_{fb} \ln I_{fb} \tag{3.2}$$

### 3.3.2 Results

The results of the estimations are shown in Table 3-3. The coefficients have been calculated by means of normalised data: the values in this table are the rescaled values. The values between brackets are t-values of the null hypothesis $\alpha$=0. Note that the t-values of the intercept have the null hypothesis ln$\alpha$=0, which implies that they are negative if $\alpha < 1$ in the original calculations.

| | α | $R^2$ | Urea | Labour | Tractors | Land |
|---|---|---|---|---|---|---|
| 1 | 1589925 | | 0.083 | 0.006 | 0.045 | 0.878 |
| | 2.9 | 98% | 5.0 | 0.4 | 6.6 | 35.2 |
| 2 | 1246218 | | 0.083 | 0.013 | 0.062 | 0.895 |
| | 1.8 | 99% | 4.0 | 0.6 | 4.0 | 31.2 |
| 3 | 1836673 | | 0.082 | -0.036 | 0.093 | 0.892 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 5.2 | 94% | 5.4 | -2.2 | 18.1 | 42.1 |
| 4 | 1160586 | | 0.079 | -0.004 | 0.067 | 0.948 |
| | 11.5 | 99% | 6.1 | -0.4 | 6.9 | 28.3 |

**Table** Error! Style not defined.**-**Error! Bookmark not defined.**:** Results of OLS estimation of an aggregated production function

It is not surprising that land has by far the largest influence on total production and a coefficient close to unity. However, it is not equal to unity and an F-test indicates that this is not likely either (Table 3-4). This is consistent with economic theory, which assumes diminishing returns to scale for land input due to the fact that the best spots will be used first.

| Estimation | F | p(F) |
|---|---|---|
| All | 69.96 | 0.00 |
| Cluster 1 | 24.15 | 0.00 |
| Cluster 2 | 13.36 | 0.00 |
| Cluster 3 | 26.04 | 0.00 |
| Cluster 4 | 2.44 | 0.12 |

**Table** Error! Style not defined.**-**Error! Bookmark not defined.**:** F-values and P-values of H0: $\alpha_{land} = 1$

### 3.3.3 Discussion

The results are satisfying: the t-values and the $R^2$ indicate that the model fits well into the data and the results are consistent with theoretical expectations. One explanation for the low coefficient value of labour is that labour might be very abundant in Java, so that the input of labour is somewhere at the far end of the production curve. In that case the disturbance in the observations can cause estimations with negative coefficients where one doesn't expect them.

## 3.4 The maximum entropy approach

### 3.4.1 Introduction

Econometricians have often come upon the problem that the only data that is available is highly aggregated. For example, in the analysis of multiproduct-multifactor firms data on total inputs are available, but the specific inputs in a given product are often not known (Lence and Miller, 1998).

The maximum entropy approach (Golan et al., 1996) is especially developed for the estimation of parameters by means of limited data. The ME formalism is used when:

- the data are in the form of averages or aggregates where, as a result, probabilities must be used to represent partial information about individual outcomes
- we know something but we don't know everything
- we don't want to tell any more or any less than we know.

The basic idea of the ME approach can be made clear using Jaynes' dice problem (Golan et al., 1996). Suppose we have a six-sided die that can take on the values $k$ = 1, 2, ..., 6 and we want to estimate the probabilities $\mathbf{p} = (p_1, p_2, ..., p_6)'$ for each possible outcome of the die, but all we know is the average outcome $y$: we are not allowed to roll the dice, say, one hundred times in order to observe the frequency distribution of the sample.

This means that we want to estimate six unknown probabilities from two pieces of information: we know the average outcome $y$ and we know that all probabilities should sum to unity. Under these restrictions, still many combinations of **p** are possible. We might solve this problem by using prior or non-sample information to choose from the feasible set of solutions. In this case, we might expect the die to be roughly 'fair', i.e. all probabilities are the same, so the average outcome $y$ should be

somewhere near 3.5. However, if $y \neq 3.5$, the underlying distribution is not likely to be uniform. In this case, we follow the ME formalism and construct the following model.
We know that the expected outcome is equal to $y$:

$$\sum_{k=1}^{6} p_k x_k = y \tag{3.3}$$

We also know that the probabilities should sum to 1:

$$\sum_{k=1}^{6} p_k = 1 \tag{3.4}$$

And all probabilities are non-negative:

$$p_k \geq 0 \tag{3.5}$$

Under these restrictions, we want to select the probabilities that maximise

$$H(\mathbf{p}) = -\sum_{k=1}^{6} p_k \ln(p_k) \tag{3.6}$$

which is Shannon's entropy measure: under restrictions (3.4) and (3.5) $H$ is at its maximum value if all $p_k$ are equal. Under additional restriction (3.3) the maximum of $H$ will refer to the situation where all $p_k$ are as equal as possible, while still satisfying this restriction.

Estimation of parameters is done in a similar fashion. Like the die, a parameter can take on many values within a given range. We will call these values *support values* and assign a probability to each support value. Under maximum entropy, the expected value of the parameter will lie in the middle of the interval; additional information can distract the expected value to another place in the interval, i.e., additional information can put a restriction on the maximisation problem, thereby decreasing the uniformity of the probabilities.

In the example we used a uniform distribution as *a priori* probability distribution, but this is not always necessarily the case. A more general description of the ME formalism is the *cross-entropy* formalism, where we have an *a priori* distribution **q** and we want to find the distribution **p** that is closest to **q** while still satisfying the constraints. In the cross-entropy case, equation (3.6) is replaced by

$$I(\mathbf{p},\mathbf{q}) = \sum_{k=1}^{K} p_k \ln \frac{p_k}{q_k} = \sum_{k=1}^{K} p_k \ln p_k - \sum_{k=1}^{K} p_k \ln q_k \tag{3.7}$$

$I(\mathbf{p}, \mathbf{q})$ is the measure of cross entropy and can be interpreted as the difference between **q** and **p**. Instead of maximising general entropy, this variable is minimised.

### 3.4.2 Estimation of parameters by the maximum entropy

The previous section presented a brief decription of the ME approach. In this section the model is described that is constructed for the estimation of parameters.

In many economic studies of land use and agricultural production land is seen as a mere input like fertiliser and labour. We will formulate production as follows:

$Q = bL^l A^a$

where $Q$ represents total production, $L$ total labour input and $A$ the area of land where the crop is grown. In this case, scale effects of land are allowed, which is consistent with economic theory, that mostly assumes diminishing returns to scale. The rationale behind this assumption is that the best pieces of land will be chosen first for the production of crops, followed by slightly less suited pieces of land, etc.

Total production is related to four inputs, with land as a fourth production factor:

$$Q_{cn} = \boldsymbol{a}_c \left(A_{cn}\right)^{a_{c4}} \prod_{f}^{3} \left(\boldsymbol{f}_{cfn} I_{fn}\right)^{a_{cf}} \tag{3.8}$$

In this equation capital symbols denote known variables. Since we only know the total factor input $I_{fn}$, this parameter is multiplied by a distribution parameter $\boldsymbol{f}_{cfn}$. This parameter is non-negative and should add up to unity:

$$\boldsymbol{f}_{cfn} \geq 0 \qquad \forall\ c, f, n \tag{3.9}$$

$$\sum_c \boldsymbol{f}_{cfn} = 1 \quad \forall\ f, n \tag{3.10}$$

We should take the logarithms of both sides of equation (6.10) in order to get functions that are easier to estimate and add en error term to these functions:

$$\ln Q_{cn} = \boldsymbol{b}_c + \boldsymbol{a}_{c4} \ln A_{cn} + \sum_{f}^{3} \boldsymbol{a}_{cf} \ln\left(\boldsymbol{f}_{cfn} I_{fn}\right) + \boldsymbol{e}_{cn} \tag{3.11}$$

where $\boldsymbol{b}_c = \ln \boldsymbol{a}_c$. All parameters except $\boldsymbol{f}_{cfn}$ will be divided in support variables, that can be recognised by the additional index $v$. The corresponding probabilities are represented by the roman equivalents of the greek characters that represent the support values. An overview is given in Table 3-5.

| Parameter | Support | Probability |
|---|---|---|
| $\boldsymbol{a}_{cf}$ | $\boldsymbol{a}_{cfv}$ | $a_{cfv}$ |
| $\boldsymbol{b}_c$ | $\boldsymbol{b}_{cv}$ | $b_{cv}$ |
| $\boldsymbol{e}_{cn}$ | $\boldsymbol{e}_{cnv}$ | $e_{cnv}$ |

**Table** Error! Style not defined.-Error! Bookmark not defined.**:** Symbols of support parameters and probabilities

The relation between a parameter and its support parameters is as follows:

$$\boldsymbol{a}_{cf} = \sum_v \boldsymbol{a}_{cfv} a_{cfv} \tag{3.12}$$

$$\boldsymbol{b}_c = \sum_v \boldsymbol{b}_{cv} b_{cv} \tag{3.13}$$

$$\boldsymbol{e}_{cn} = \sum_v \boldsymbol{e}_{cnv} e_{cnv} \tag{3.14}$$

From equations (3.11) through (3.14) the following equation is derived:

$$\ln Q_{cn} = \sum_v \boldsymbol{b}_{cv} b_{cv} + \left(\sum_v \boldsymbol{a}_{c4v} a_{c4v}\right) \ln A_{cn} + \sum_{f}^{3} \left\{ \left(\sum_v \boldsymbol{a}_{cfv} a_{cfv}\right) \ln\left(\boldsymbol{f}_{cfn} I_{fn}\right) \right\} + \sum_v \boldsymbol{e}_{cnv} e_{cnv} \tag{3.15}$$

This is the consistency constraint of the ME model. Besides this constraint, additivity constraints are included in the model that require that all probabilities of a given parameter add up to unity:

$$\sum_{v} a_{cfv} = 1 \quad \forall\, c, f \tag{3.16}$$
$$\sum_{v} b_{cv} = 1 \quad \forall\, c \tag{3.17}$$
$$\sum_{v} e_{cnv} = 1 \quad \forall\, c, n \tag{3.18}$$

All probabilities must be non-negative:

$$a_{cfv} \geq 0 \quad \forall\, c, f, v \tag{3.19}$$
$$b_{cv} \geq 0 \quad \forall\, c, v \tag{3.20}$$
$$e_{cnv} \geq 0 \quad \forall\, c, n, v \tag{3.21}$$

Before formulating the entropy function we have to ask ourselves what the distributions would look like under complete uncertainty. The probabilities $a_{cfv}$, $b_{cv}$ and $e_{cnv}$ can be assumed to distributed equally. We are not sure whether this is true for the distribution of inputs. If five crops are grown in the area, should we assume *a priori* that each crop uses one fifth of the inputs? One objection to that assumption is that some crops, for example sweet potato, have been left out of the estimation because the production of these crops is very low compared to other crops. The selection of crops is somewhat arbitrary, but it would affect the *a priori* distribution if we assume it to be equal to 1/n, where n is the number of crops. Using production levels to indicate the relevance of a crop can be problematic, as a choice should be made what measure should be taken: the weight of production of the income it generates? In this study the *a priori* distribution is based on the sown area of the crops, so that the *a priori* fraction of total inputs is equal to the fraction of the sown area. When the *a priori* distribution of probabilities is not uniform, the problem should be formulated as a *cross-entropy* model, where the difference between the estimated probability distribution and the *a priori* probability distribution is minimised. For $a_{cfv}$, $b_{cv}$ and $e_{cnv}$ we can assume a uniform *a priori* probability distribution, i.e. the *a priori* probabilities are equal over the support values:

$w_v = w_{v'}\ \forall\ v \neq v'$ and

$$\sum_{v} w_v = 1$$

where $w_v$ denotes the *a priori* probability of support parameter $v$. For the distribution of inputs over crops, we assume that the fraction of inputs applied to crop $c$ is proportional to the fraction of harvested area that is used for crop $c$:

$$q_{cn} = \frac{A_{cn}}{\sum_{c} A_{cn}}$$

Now that we have determined the *a priori* probability distribution, we can formulate the objective function:

$$E = \sum_{c}\sum_{v} b_{cv} \ln\frac{b_{cv}}{w_v} + \sum_{c}\sum_{v}\sum_{f}^{4} a_{cfv} \ln\frac{a_{cfv}}{w_v} + \sum_{c}\sum_{n}\sum_{f}^{3} f_{cfn} \frac{f_{cfn}}{q_{cn}} + \sum_{c}\sum_{n}\sum_{v} e_{cnv} \ln\frac{e_{cnv}}{w_v} \tag{3.22}$$

The variable $E$ is a measure of cross-entropy and is to be minimised. In the literature cross-entropy is normally denoted by the symbol $I$, but as this would cause confusion with the symbol for inputs in this study, the symbol $E$ is chosen.

**Adding restrictions of optimising behaviour**

We can add further restrictions expressing the assumption that farmers will maximise their profit. If this is a reasonable assumption, we can abstract more information from the data.

In the optimum the value of marginal production should equal the price of inputs:

$$P_c \frac{\partial Q_c}{\partial i_{cf}} = P_f \tag{3.23}$$

In case we don't know the value of $P_f$, we can derive from this condition that the value of marginal production of input *f* must be equal for all crops. After all, if we have two crops 1 and 2 then

$$\left.\begin{array}{l} P_1 \dfrac{\partial Q_1}{\partial I_{1f}} = P_f \\ P_2 \dfrac{\partial Q_2}{\partial I_{2f}} = P_f \end{array}\right\} P_1 \frac{\partial Q_1}{\partial I_{1f}} = P_2 \frac{\partial Q_2}{\partial I_{2f}} \tag{3.24}$$

Let's take a look at the production function that is being estimated in the ME model and write this for simplicity as

$$Q_c = \boldsymbol{b} L_c^{l} U_c^{u} T_c^{t} A_c^{a}$$

where $L_c$ represents labour input, $U_c$ represents urea input, $T_c$ represents tractor input and $A_c$ represents land input in crop *c*. Unfortunately, we only know the price of urea, so equation (6.24) can only be applied to this factor:

$$P_c \frac{\partial Q_c}{\partial U_c} = P_c \boldsymbol{u} \boldsymbol{b} L_c^{l} U_c^{u-1} T_c^{t} A_c^{a} = P_c \boldsymbol{u} U_c^{-1} \boldsymbol{b} L_c^{l} U_c^{u} T_c^{t} A_c^{a} = P_c \boldsymbol{u} \frac{Q_c}{U_c} = P_U$$

In this restriction there are two variables: the coefficient of urea $\boldsymbol{u}$, and urea input $U_c$, which is subject to the restriction

$$\sum_c U_c = U$$

For the other inputs we will have to apply equation (6.25):

$$P_c \frac{\partial Q_c}{\partial L_c} = P_{c'} \frac{\partial Q_{c'}}{\partial L_{c'}} \Rightarrow P_c \boldsymbol{l}_c \frac{Q_c}{L_c} = P_{c'} \boldsymbol{l}_{c'} \frac{Q_{c'}}{L_{c'}}$$

$$P_c \boldsymbol{t}_c \frac{Q_c}{T_c} = P_{c'} \boldsymbol{t}_{c'} \frac{Q_c}{T_c}$$

$$P_c \boldsymbol{a}_c \frac{Q_c}{A_c} = P_{c'} \boldsymbol{a}_{c'} \frac{Q_{c'}}{A_{c'}}$$

The first order conditions are included in the model with an error term to allow for (small) deviations from the optimum. In terms of the model the equations are as follows:

$$P_c \left( \sum_v \boldsymbol{a}_{cfv} a_{cfv} \right) \frac{Q_{cn}}{\boldsymbol{f}_{cfn} I_{fn}} + \sum_v \boldsymbol{e}_{cfnv} e_{cfnv} = P_{c'} \left( \sum_v \boldsymbol{a}_{c'fv} a_{c'fv} \right) \frac{Q_{c'n}}{\boldsymbol{f}_{c'fn} I_{fn}} \tag{3.25}$$

$$P_c \left( \sum_v \boldsymbol{a}_{c1v} a_{c1v} \right) \frac{Q_{cn}}{\boldsymbol{f}_{c1n} I_{1n}} + \sum_v \boldsymbol{e}_{c1nv} e_{c1nv} = P_1 \tag{3.26}$$

where urea is input 1.

**The distribution of land**

Land is a special case. We don't know the price of land, but we do know its input distribution over the crops. Secondly, land has a particular relationship with production that allows further examination of this factor. The ratio of production to land ($Q_c/A_c$) is a known parameter: we call it yield. Therefore, the first order condition for land can be rewritten as

$$P_c \boldsymbol{a}_c q_c = P_{c'} \boldsymbol{a}_{c'} q_{c'}$$

Within one cluster the parameter $\boldsymbol{a}_c$ is assumed to be constant. If we assume that prices do not vary within one cluster either, we can rewrite the abovementioned equation so that all constants are on one side of the equation:

$$q_c = \frac{P_{c'} \boldsymbol{a}_{c'}}{P_c \boldsymbol{a}_c} q_{c'}$$

If we assume that both $P_c$ and $\boldsymbol{a}_c$ are constant: therefore, the ratio between yields should be constant also. With the correlation between crop yields, we can test these assumptions. Appendix B shows the correlation coefficients between crop yields. These tables show that the correlations between the yields are very poor. This indicates that the assumptions made cannot hold for the situation in the clusters. This implies that adding the first order condition for land input as a restriction to the model will cause major problems in solving the model. Therefore, the first order condition for land input has not been included.

**Support values**

Now the model equations have been specified, appropriate support values must be chosen. Two features are relevant in this case: the choice of the appropriate interval and the choice of the appropriate number of discrete values in the interval. Mostly five is the most convenient number of values in the range that yields an acceptable accuracy. Therefore, five discrete values will be used, which leaves us the question of the appropriate interval.

According to Golan et al. (1996) the range of the interval influences the estimations, though this influence is not as big for the parameters as it is for the disturbance term. Golan et al. (1996) gives the advise to set the interval of the disturbance term at $\{-3\boldsymbol{s}, 3\boldsymbol{s}\}$ where $\boldsymbol{s}$ represents the standard deviation of the explained variable, in this case $q_{cn}$. We have already specified the interval of the distribution parameter in equations (3.9) and (3.10): its interval is by definition $\{0, 1\}$.

The choice of parameter intervals is difficult if inputs and outputs are measured in different ranges. At the same time, the model will be solved more easily if all variables have roughly the same scale. Therefore, parameters are rescaled so that they all have a mean value of 1. In this case the choice of the parameter interval is more clear, though still somewhat arbitrary. Two things should be taken into consideration:

1. Should the parameters be allowed to be negative?
2. What should be the expected value of the interval?

The intercept $\boldsymbol{b}_c$ can be negative, as $\boldsymbol{a}_c$ will be derived from it by $\boldsymbol{a} = e^{b}$. But we don't want our coefficients to be negative, as this would imply that production goes down as inputs go up. In fact this consideration is another piece of information we use in the estimation of parameters. If we want our coefficients to be positive, the interval might be something like $\{0, x\}$ where $x$ is any positive value. However, this type of interval might be contradictory to the second consideration, i.e. the expected value of the coefficient. If the *a priori* probability distribution is uniform, the expected value of a parameter will be equal to the middle of the interval, or (lower bound + upper bound)/2. The expected value of the interval {0, $x$} would therefore be $x/2$, but maybe we don't want to assume that. If we know absolutely nothing about the value of a coefficient, we would rather assume it to be zero. A solution to this problem is to give all parameters the interval {-$x$, $x$}, and to place a lower bound at the coefficients:

$$\sum_v \boldsymbol{a}_{cfv} a_{cfv} \geq 0 \qquad (3.27)$$

In the estimations all coefficients have an interval of {-10, 10}, but it is recommended to check the model on the influence of the size of the intervals.

**Results**

Appendix C shows the estimation results. It can be seen that the price of the crops is to some extent compensated by a lower or higher intercept. Per production factor the coefficients are roughly the same for all crops. The coefficient of land indicates that the relation between land and production is close to linear.

Table 3-6 compares the estimated factor input per crop to the estimated average factor input per crop in the period 1984 - 1993 (source: (Anonymous, 1996)). Labour input per crop is only available in Rupees per hectare, so both estimated and observed labour ditribution have been normalised to an average of 1, in order to make the figures comparable.

| Urea input (kg/ha) | Cassava | Groundnut | Soybean | Maize | Rice |
|---|---|---|---|---|---|
| Estimated | 229.8 | 221.8 | 257.7 | 255.5 | 266.3 |
| Average 1984 - 1993 | 70.8 | 51.6 | 61.9 | 139.1 | 243.8 |
| Labour (normalised) | Cassava | Groundnut | Soybean | Maize | Rice |
| Estimated (number of workers) | 1.00 | 0.99 | 1.01 | 1.00 | 1.00 |
| Average 1984 - 1993 (Labour costs) | 0.79 | 0.99 | 0.86 | 0.56 | 1.81 |

**Table** Error! Style not defined.-Error! Bookmark not defined.**:** Estimated and observed distribution of urea and labour

We see that the distribution of urea and labour are very close to its *a priori* value compared to the average of observed input distributions. An observed distribution of tractor inputs is not available, so Figure 3-2 gives an indication of the variation in the ratio between *a priori* input and estimated tractor input. The figure shows the maximum and minimum values of the ratio and the {-σ, σ}-interval of the mean.

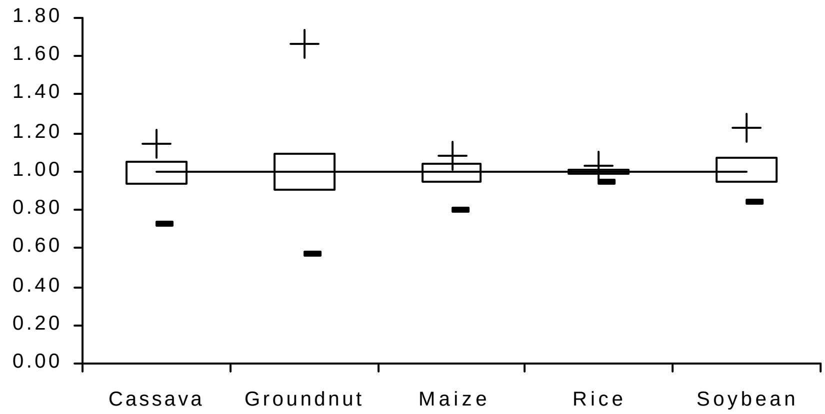


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Ratio between *a priori* and estimated tractor input

The figure clearly shows that the estimated distribution of input factors has remained close to the *a priori* distribution, especially for the more important crops like rice and maize. Less significant crops like groundnut and soybean appear to allow for wider variations, though even for these crops a large majority of the estimations remains close to their *a priori* value.

### 3.4.3 Discussion

Although some parameter values have resulted from the estimation procedure, it is questionable whether these estimations are reliable enough for application in a model. The small difference between estimated input distribution and the *a priori* distribution and the large difference between (national) average inputs per crop and the estimated inputs indicate that the estimated production functions might not be too different from the actual relationships.

# 4. Running the parameterised models

## 4.1 Introduction

In the previous chapter production functions have been estimated based on aggregate data, using two methods: that of estimating an aggregate production function and that of estimating the distribution of production factors along with the production function coefficients. In this chapter the estimated production functions are applied in relatively straightforward land use models of Java.

This chapter discusses the models and the results of both approaches. Finally, the two methodologies are compared in the discussion.

## 4.2 A land use model of aggregated production

### 4.2.1 Structure of the model

*Estimated production*

The calculation of the estimated production is very straightforward and only involves a recalculation of the production as predicted by the model:

$$T_{n0} = \alpha_n \prod_f I_{fn0}^{\alpha_{fn}} \tag{4.2.1}$$

*Optimised production*

As tractors, labour and land are exogenous in the model, urea input is the only free variable left. As mentioned before, urea use is optimal if the marginal production of an input equals the price of that input. By means of this rule, we can calculate the optimal urea input. The first order condition for urea use looks as follows:

$$\frac{\partial T}{\partial I_{4n0}} = P_4 \Leftrightarrow$$

$$\alpha_{4n} I_{4n0}^{\alpha_{4n}-1} \alpha_n \prod_f^3 I_{fn0}^{\alpha_{fn}} = P_4$$

We can rewrite this equation in the following form:

$$I_{4n0} = \left[ \frac{P_4}{a_{4n} a_n \prod_f^3 I_{fn0}^{a_{fn}}} \right]^{\frac{1}{a_{4n}-1}} \tag{4.2.2}$$

By equation (4.2.2) the optimal urea use is calculated, which is used in equation (4.2.1) to calculate the optimal turnover.

### 4.2.2 Results

Figure 4-1, Figure 4-2 and Figure 4-3 show the observed production levels and the estimated and optimal production levels as calculated by the model.

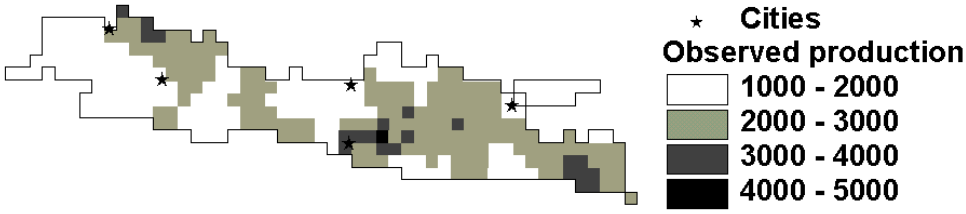


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Observed production in thousands Rp per km$^2$

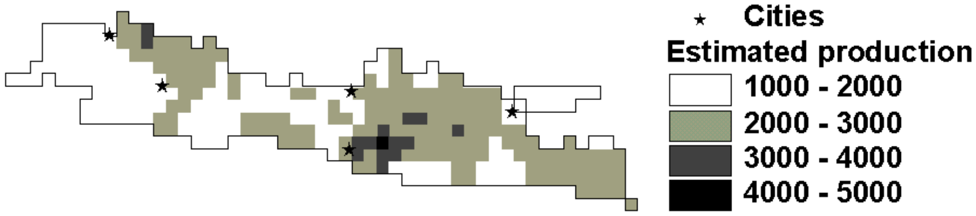


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Production estimated by the model in thousands Rp per km$^2$

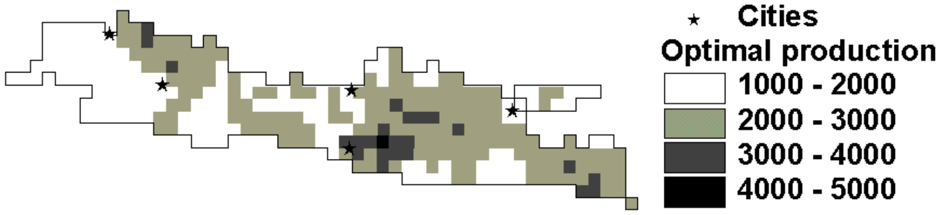


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Production under optimal urea input in thousands Rp per km$^2$

Figure 4-4, Figure 4-5 and Figure 4-6 show the differences between observed, estimated and optimal production.

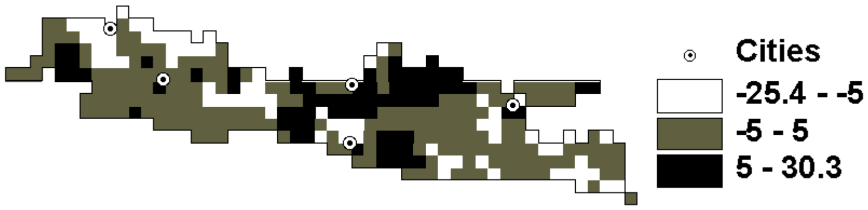


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Difference between observed and estimated production

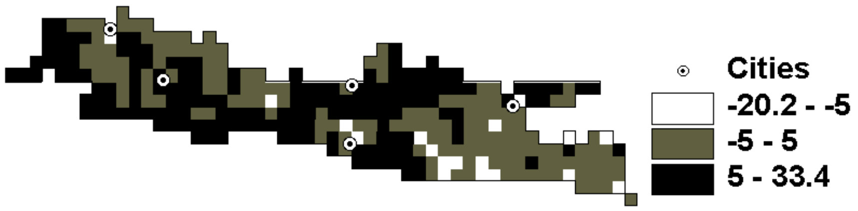


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Difference between observed production and production under optimal urea input

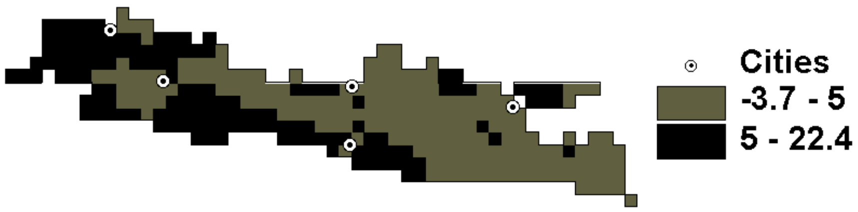


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Difference between estimated production and production under optimal urea input

The figures show that the differences are not shockingly large. In about 57% of the grid cells, the predicted production remains within a distance of 5% from the observed value. About 90% of the grid cells lies within a 10% distance from the observed value.

The production levels under optimised urea input differ somewhat more from the observed values. Still, 45% lies within a distance of 5% difference from the observed value. Under optimised urea input, production is more likely to be overestimated than it is to be underestimated: 50% of the gridcells have a production that is more than 5% higher than observed, while the percentage of grid cells with production more than 5% lower than observed is only 5%. There does not seem to be a clear spatial pattern in these differences.

The differences between estimated production and production under optimised urea input are much smaller. In West Java, Yogyakarta and the southern part of Central Java the urea input under optimised urea input is significantly higher than observed.

## 4.3 A land use model based on ME estimations

In the ME estimations production functions are estimated of each crop. Therefore, the model based on these estimations optimises the distribution of inputs whose total input is given and calculates the production of each crop.

### 4.3.1 Structure of the model

As in the maximum entropy estimations land is included as an input like urea or machinery, the model itself is relatively straightforward. The model maximises profit defined as turnover minus costs:

$$p = \sum_y \left[ \sum_c \sum_n \left( P_{cy} a_{cn} \prod_f I_{cfny}^{a_{cnf}} - \sum_f I_{cfny} P_{fy} \right) \right] \quad (4.3.1)$$

such that

$$\sum_c I_{cfny} \leq S_{fny} \quad (4.3.2)$$

$$I_{cfny} \geq 0 \quad (4.3.3)$$

where the symbols denote the following:

$a_{cn}$ Intercept of crop $c$ in cell $n$
$a_{cnf}$ Coefficient of factor $f$ for crop $c$ in cell $n$
$I_{cfny}$ Input of factor $f$ in crop $c$ in cell $n$ in year $y$
$P_{cy}$ Price of crop $c$ in year $y$
$P_{fy}$ Price of input factor $f$ in year $y$
$S_{fny}$ Stock of input factor $f$ in cell $n$ in year $y$

The value of $a_{cn}$ and $a_{cnf}$ in a given cell $c$ depends on the cluster the cell is assigned to.

### 4.3.2 Results

Figure 4-7 shows the absolute area of rice according to the model.

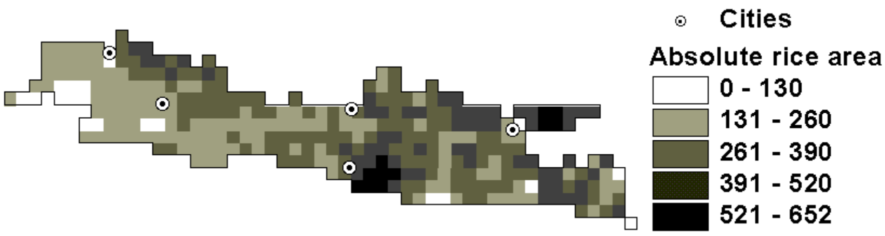


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Rice area in square kilometers according to the model

On first sight this figure looks reasonable, however, if we look at the relative rice area (Figure 4-8) we see that the solution found is probably a corner solution: all land is devoted to rice production.

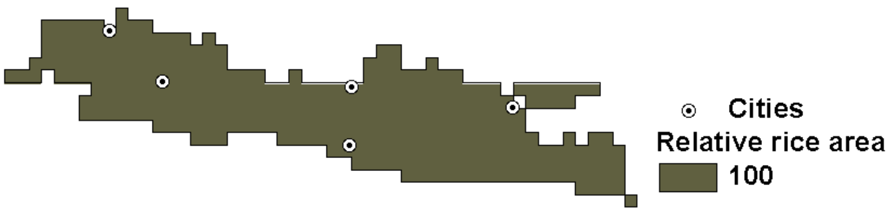


**Figure** Error! Style not defined.-Error! Bookmark not defined.**:** Rice area in % of agricultural area according to the model

Apparently, in its present form, the model has no interior solution. In Section Discussion and conclusions a number of explanations for these results will be discussed, as well as possibilities for improvement of the results.

## 5. Discussion and conclusions

In this paper we have seen a wide range in economic models of land use and cover change. Although many models are based on standard assumptions of economic modelling, recent developments, like cellular automata models, have added different approaches to economic land use models. The models discussed in this paper have a high level of detail, which is convenient in a theoretical sense, but has its drawbacks in empirical applications, as in many study areas census data are the most detailed source of information. These data are in most cases highly aggregated and only available at regional scales, which is quite high for detailed models of land use and cover change.

The estimation and running of the aggregated model yields satisfying results: the results of the model run indicate that it can reproduce existing patterns of agricultural activity quite well. Unfortunately, it does not model the production of separate crops, but for the analysis of phenomena that depend on land use intensity without making distinction between crops the methodology should be useful. The approach can be improved if spatial characteristics like distances, transport costs and local price differences are known.

The detailed model shows promising opportunities for the parameter estimation of detailed land use models, although it can be improved on many aspects, as it tends to yield conrner solutions in its present form. Practical aspects of land use and cover change not included in the model as yet are conversion costs, conversion time, the valuation of risk and the level of subsistence farming. Instead of modelling *land use*, the detailed model can be designed to analyse *land use change* and subsistence farming can be taken into account by estimating a minimum level of production of each crop, or a minimum level of food ingredients.

**Acknowledgements**

The research described in this paper has been carried out under the Dutch National Research Programme on Global Air Pollution and Climate Change in co-operation with the CLUE research group at Wageningen Agricultural University. The authors wish to thank the CLUE group for the kind permission to use their database of Java for the empirical applications in this paper.

## References

Anonymous (1996). *The BPS website.* Badan Pusat Statistik, Republik Indonesia , Jakarta. URL: www.bps.go.id

Anonymous (1996). *Vademekum Sumber Daya*. Direktorat Jenderal Tanaman Pangan Dan Hortikultura/Direktorat Bina Usaha Tani Dan Pengolahan Hasil, Jakarta.

Alcamo, J., G. J. J. Kreileman, M. S. Krol, G. Zuidema, G. J. Van Den Born, A. F. Bouwman, B. J. De Haan, K. Klein Goldewijk, O. Klepper, J. Krabec, R. Leemans, J. G. J. Olivier, A. M. C. Toet, H. J. M. De Vries, H. J. Van Der Woerd, R. A. Van Den Wijngaart, J. G. Van Minnen, M. Vloedbeld, M. Jonas, and K. Olendrzynski (1994). *IMAGE 2.0: Integrated modeling of global climate change*. Kluwer Academic Publishers, Dordrecht.

Angelsen, A. and D. Kaimowitz (1997). *What Can We Learn from Economic Models of Tropical Deforestation?* Jakarta.

Bockstael, N. (1996). Modeling Economics and Ecology: The Importance of a Spatial Perspective. *American Journal of Agricultural Economics*, 78:1168-1180.

Chomitz, K. M. and D. A. Gray (1995). *Roads, Land, Markets and Deforestation: A Spatial Model of Land Use in Belize*. Policy Research Working Paper, No. 1444, World Bank, Policy Research Department, Washington, D.C.

Chomitz, K. M. and D. A. Gray (1996). Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize. *World Bank Economic Review*, 10:3, 487-512.

Engelen, G., R. White, I. Uljee, and P. Drazan (1995). Using Cellular Automata for Integrated Modelling of Socio- Environmental Systems. *Environmental Monitoring and Assessment*, 34:203-214.

Fischer, G., Y. Ermoliev, M. A. Keyzer, and C. Rosenzweig (1996). *Simulating the Socio-Economic and Biogeophysical Driving Forces of Land-Use and Land-Cover Change: The IIASA Land-Use Change Model*. IIASA Working Paper, No. WP-96-010, International Institute for Applied Systems Analysis, Laxenburg.

Geoghegan, J., L. A. Wainger, and N. Bockstael (1997). Spatial Landscape Indices in a Hedonic Framework: an Ecological Economics Analysis Using GIS. *Ecological Economics*, 23:3, 251-264.

Golan, A., G. Judge, and D. Miller (1996). *Maximum entropy econometrics: Robust estimation with limited data.* 1stst edition, John Wiley & Sons, Chichester.

Hazell, P. B. R. and R. D. Norton (1986). *Mathematical programming for economic analysis in agriculture*. 1st edition, MacMillan Publishing Company, New York.

Jones, D. W. and R. V. O'Neill (1993). Land Use in the Presence of an Atmosphere Externality, with and without Corrective Taxes. *Journal of Regional Science*, 33:4, 457-480.

Kaimowitz, D. and A. Angelsen (1998). *Economic Models of Tropical Deforestation: A Review*. Center for International Forestry Research, Bogor.

Lence, S. H. and D. J. Miller (1998). Estimation of multi-output production functions with incomplete data: A generalised maximum entropy approach. *European Review of Agricultural Economics*, 25:2, 188-209.

McMillen, D. P. (1989). An Empirical Model of Urban Fringe Land Use. *Land Economics*, 65:2, 138-145.

Moxey, A. P., B. White, and J. R. O'Callaghan (1995). The Economic Component of NELUP. *Journal of Environmental Planning and Management*, 38:1, 21-33.

Moxey, A. P., B. White, R. A. Sanderson, and S. P. Rushton (1995). An Approach to Linking an Ecological Vegetation Model to an Agricultural Economic Model. *Journal of Agricultural Economics*, 46:3, 381-397.

O'Callaghan, J. R. (1995). NELUP: an Introduction. *Journal of Environmental Planning and Management*, 38:1, 5-20.

SAS Institute Inc. (1989). *SAS/STAT® User's guide, Version 6*. Fourthst edition, SAS Institute Inc., Cary, NC.

Schipper, R. A., Jansen, H. G. P., Bouman, B. A. M., Hengsdijk, H., Nieuwenhuyse, A., and Sáenz, F. (1998). *Evaluation of development policies using integrated bio-economic land use models: applications to Costa Rica*. Paper presented at the AAEA congress, July 31 98. Guálipes.

Schotten, C. G. J., R. J. Van De Velde, H. J. Scholten, W. T. Boersma, M. Hilferink, M. Ransijn, P. Rietveld, and R. Zut (1997). *De Ruimtescanner, geïntegreerd ruimtelijk informatiesysteem voor de simulatie van toekomstig ruimtegebruik*. 711901 002, National Institute of Public Health and the Environment, Bilthoven.

Van De Putte, R. (1988). Farming systems and Land Use Modelling for Watershed Management. *ITC Journal*, 1, 83-86.

Veldkamp, A. and L. Fresco (1996). CLUE: a conceptual model to study the conversion of land use and its effects. *Ecological Modelling*, 85:253-270.

Wolfram, S. (1984). Cellular Automata as Models of Complexity. *Nature*, 311:419-424.

Zuidema, G., G. J. Van Den Born, J. Alcamo, and G. J. J. Kreileman (1994). Simulating changes in global land cover as affected by economic and climatic factors. *Water, Air and Soil Pollution*, 76:1-2, 163-198.

## Appendix AComparison of previous work

| Source | Name | Category | Decision level | Number of land use types | Smallest aggregation unit |
|---|---|---|---|---|---|
| Von Thünen (1826) | The Isolated State | Optimisation | -- | -- | -- |
| McMillen (1989) | | Discrete Choice | Bottom | >1 | Site |
| Martínez (1992) | Bid-Choice Land-Use Model | Discrete Choice | Bottom | >1 | -- |
| Alcamo et al. (1994) | IMAGE | Heuristic | Bottom | >1 | Cell |
| Crihfield (1994) | | Optimisation | Top | 1 | -- |
| Engelen et al. (1995) | | Heuristic | Bottom | 1 | Cell |
| Folmer et al. (1995) | ECAM | General Equilibrium | Top | > 1 | Country |
| Moxey et al. (1995) | NELUP | Optimisation | Top | 1 | Region |
| Chomitz & Gray (1996) | | Discrete Choice | Bottom | >1 | |
| Fischer et al. (1996) | LUC | General Equilibrium | Top | 1 | Compartments |
| Schipper et al. (1996) | REALM | Optimisation | Top | 1 | Farm types |
| Geoghegan et al. (1997) | | Discrete Choice | Bottom | 1 | Cell |
| Schotten et al. (1997) | Land Use Planner | Discrete Choice | Bottom | >1 | Cell |

| Source | Spatial scale levels | Time scale | Focus | Production functions |
|---|---|---|---|---|
| Von Thünen (1826) | 1 | Static | Optimal land allocation | Land use type |
| McMillen (1989) | 1 | Static | Urban fringe land use | Land use type |
| Martínez (1992) | 1 | Static | Land use change | Land use type |
| Alcamo et al. (1994) | 2 | Dynamic | Climate change | Land use type |
| Crihfield (1994) | 1 | Dynamic | Strip mining | Land use type |
| Engelen et al. (1995) | 3 | Static | Small island state | Land use type |
| Folmer et al. (1995) | 1 | Dynamic | CAP Reform | Mathematic |
| Moxey et al. (1995) | 1 | Static | Land use change | Land use type |
| Chomitz & Gray (1996) | 1 | Static | Deforestation | Land use type |
| Fischer et al. (1996) | 3 | Dynamic | Land use change | Continuous function |
| Schipper et al. (1996) | 1 | Static | Land use analysis | Land use type |
| Geoghegan et al. (1997) | 1 | Static | Residential value | Land use type |
| Schotten et al. (1997) | 1 | Static | Policy analysis | Land use type |

| Source | Scenario environment | Data[3] | Application area |
|---|---|---|---|
| Von Thünen (1826) | -- | Prices, costs | Germany |
| McMillen (1989) | -- | Land quality, adjustment costs | Chicago, USA |
| Martínez (1992) | -- | Land quality, prices | -- |
| Alcamo et al. (1994) | climate change | Demand, land quality | Earth |
| Crihfield (1994) | 9 policies | Prices, costs, interest | Illinois, USA |
| Engelen et al. (1995) | climate change | Land quality, population, demand | St. Lucia, Caribbean |
| Folmer et al. (1995) | CAP reform | SAM, historical parameters | European Union |
| Moxey et al. (1995) | CAP Reform | I-O rel., prices, costs, resources | Tyne catchment, GB |
| Chomitz & Gray (1996) | -- | Land quality, LUTs, distances | Belize |
| Fischer et al. (1996) | | Prices, quantities, etc. | China |
| Schipper et al. (1996) | 8 policy and econ. scenarios | Prices, land quality, environment | Neguev settlement, Costa Rica |
| Geoghegan et al. (1997) | | Housing prices, land qual. | Patuxent Watershed, USA |
| Schotten et al. (1997) | CPB[4]-scenarios | Demand, land quality | The Netherlands |

~~**Appendix A**~~

[3] Not all data requirements have been mentioned

[4] Centraal Planbureau (Central Planning Office)

## Appendix ACorrelation between yields per cluster

| | Cassava | Groundnut | Maize | Rice | Soy |
|---|---|---|---|---|---|
| Cassava | 1.00 | 0.13 | 0.52 | 0.47 | -0.19 |
| Groundnut | 0.13 | 1.00 | -0.02 | -0.01 | 0.14 |
| Maize | 0.52 | -0.02 | 1.00 | 0.64 | -0.16 |
| Rice | 0.47 | -0.01 | 0.64 | 1.00 | -0.16 |
| Soy | -0.19 | 0.14 | -0.16 | -0.16 | 1.00 |

**Table B-1:** Correlation between yields in cluster 1

| | Cassava | Groundnut | Maize | Rice | Soybean |
|---|---|---|---|---|---|
| Cassava | 1.00 | -0.61 | 0.68 | 0.69 | -0.16 |
| Groundnut | -0.61 | 1.00 | -0.52 | -0.66 | -0.04 |
| Maize | 0.68 | -0.52 | 1.00 | 0.79 | -0.10 |
| Rice | 0.69 | -0.66 | 0.79 | 1.00 | -0.12 |
| Soybean | -0.16 | -0.04 | -0.10 | -0.12 | 1.00 |

**Table B-2:** Correlation between yields in cluster 2

| | Cassava | Groundnut | Maize | Rice | Soybean |
|---|---|---|---|---|---|
| Cassava | 1.00 | 0.16 | 0.37 | 0.34 | 0.15 |
| Groundnut | 0.16 | 1.00 | 0.12 | 0.39 | 0.08 |
| Maize | 0.37 | 0.12 | 1.00 | 0.50 | 0.13 |
| Rice | 0.34 | 0.39 | 0.50 | 1.00 | 0.18 |
| Soybean | 0.15 | 0.08 | 0.13 | 0.18 | 1.00 |

**Table B-3:** Correlation between yields in cluster 3

| | Cassava | Groundnut | Maize | Rice | Soybean |
|---|---|---|---|---|---|
| Cassava | 1.00 | 0.06 | 0.50 | 0.38 | 0.22 |
| Groundnut | 0.06 | 1.00 | -0.22 | -0.04 | 0.31 |
| Maize | 0.50 | -0.22 | 1.00 | 0.41 | -0.14 |
| Rice | 0.38 | -0.04 | 0.41 | 1.00 | -0.09 |
| Soybean | 0.22 | 0.31 | -0.14 | -0.09 | 1.00 |

**Table B-4:** Correlation between yields in cluster 4

## Appendix ACoefficient estimates by the ME approach

| | Cassava | Groundnut | Maize | Rice | Soybean |
|---|---|---|---|---|---|
| Intercept | 13509 | 1097 | 2343 | 4591 | 1317 |
| Int*price | 1216 | 1618 | 977 | 1579 | 1518 |
| Urea | 0.067 | 0.046 | 0.071 | 0.044 | 0.055 |
| Labour | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 |
| Tractors | 0.020 | 0.014 | 0.021 | 0.013 | 0.017 |
| Land | 0.919 | 0.958 | 0.966 | 0.976 | 0.905 |
| $R^2$ | 98% | 99% | 99% | 97% | 98% |

**Table C-1:** Production coefficients in cluster 1

| | Cassava | Groundnut | Maize | Rice | Soybean |
|---|---|---|---|---|---|
| Intercept | 11090 | 830 | 1715 | 6403 | 967 |
| Int*price | 998 | 1224 | 715 | 2203 | 1115 |
| Urea | 0.066 | 0.041 | 0.068 | 0.042 | 0.052 |
| Labour | 0.068 | 0.045 | 0.067 | 0.042 | 0.055 |
| Tractors | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 |
| Land | 0.803 | 0.969 | 0.909 | 0.828 | 0.880 |
| $R^2$ | 96% | 98% | 100% | 94% | 95% |

**Table C-2:** Production coefficients in cluster 2

| | Cassava | Groundnut | Maize | Rice | Soybean |
|---|---|---|---|---|---|
| Intercept | 11022 | 1018 | 2004 | 3261 | 995 |
| Int*price | 992 | 1501 | 836 | 1122 | 1148 |
| Urea | 0.061 | 0.046 | 0.069 | 0.041 | 0.052 |
| Labour | 0.059 | 0.046 | 0.065 | 0.038 | 0.050 |
| Tractors | 0.030 | 0.024 | 0.034 | 0.020 | 0.026 |
| Land | 0.836 | 0.824 | 0.835 | 0.965 | 0.875 |
| $R^2$ | 99% | 97% | 96% | 98% | 97% |

**Table C-3:** Production coefficients in cluster 3

| | Cassava | Groundnut | Maize | Rice | Soybean |
|---|---|---|---|---|---|
| Intercept | 9590 | 981 | 1720 | 2311 | 1158 |
| Int*price | 863 | 1447 | 717 | 795 | 1335 |
| Urea | 0.051 | 0.037 | 0.059 | 0.034 | 0.043 |
| Labour | 0.043 | 0.031 | 0.047 | 0.028 | 0.036 |
| Tractors | 0.069 | 0.049 | 0.077 | 0.044 | 0.058 |
| Land | 0.891 | 0.872 | 0.907 | 1.031 | 0.812 |
| $R^2$ | 99% | 97% | 98% | 98% | 98% |

**Table C-4:** Production coefficients in cluster 4