

## Estimating input allocation for farm supply models

**A. Gocht**

**Institute for Food and Resource Economics ILR, Bonn University, Germany**

**Contact:**

**[alexander.gocht@ilr.uni-bonn.de](mailto:alexander.gocht@ilr.uni-bonn.de)**



**Paper prepared for presentation at the 107<sup>th</sup> EAAE Seminar "*Modelling of Agricultural and Rural Development Policies*". Sevilla, Spain, January 29<sup>th</sup> -February 1<sup>st</sup>, 2008**

*Copyright 2007 by A. Gocht. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.*

## Abstract

*When building an economic model for supply analysis the aim is to model a decision making process of one or more agents which fits the observed practice as good as possible. Hereby the modeller is often confronted with incomplete information about the production process; particular crop specific input data are rarely available. The problem of defining activity related technology inputs coefficients is not new. A good deal of literature comes from the mathematical programming perspective, where input coefficients were estimated using a standard linear regression function to fully represent the mathematical program. However this approach is a pure technical device and may result in an inconsistent model.*

*The author of the paper wants to investigate whether it is possible, employing proper estimation techniques, to simultaneously estimate all unknown coefficients of a mathematical farm supply model. This includes the estimation of parameters of the non linear cost function, used to calibrate and catch the simulation behaviour and the crop specific input coefficients. It is shown that a simultaneous estimation of all parameters improves the goodness of fit of the estimated parameters and that such an approach is technically feasible.*

**Key words:** farm supply model, input allocation, entropy, HDP

## 1. Introduction

The decision-making problem of an agricultural firm or sector in empirical analysis is often specified using multiple-output and constraint mathematical programming models. Here two trends can be discovered in the last decade. The first comprehends the issue of “How parameters of a programming supply model are derived”. In that area progress can be identified in a development away from pure normative programming models - where the model structure did not take historical data into account - towards supply models, where the advantages of econometric estimation and mathematical programming are combined. This enhancement was mainly possible due to ill-posed estimation methods proposed by Golan et al. (1997) or related techniques such as the highest posterior density (HPD) estimation Heckelee et al. (2005). The distinctive advantage of these estimation techniques is the possibility to overcome the ill-posedness by incorporating prior information and offered the practitioners the possibility to make full use of the available information. Recent “real-life” application from Buysse et al. (2007) and Jansson (2007) illustrate the use of the estimation methods and unambiguously depict the development away from calibration methods - such as positive mathematical programming (PMP, see e.g. Howitt (1995)) - towards econometric related type of estimated supply model using prior information.

The *second tendency* is the linkage of agricultural supply programmes with spatial related bio-physical models. These types of model interact with supply models over policy provoked land and input intensity changes, which requires that the supply model explicitly define technologies and it brings the coefficients of those technologies - fertilizer and pesticides per crop - into sharper focus.

These two tendencies gave impetus to raise the question; whether it is feasible to use the new estimation methods to further improve the explanatory power of the technology coefficients of the supply model. In this context particular the parameter of the input technology comes into focus, due to the linkage to environmental and bio-physical models. Unfortunately these input coefficients are often not directly observable and the information is often restricted to total farm or sector purchases of various input categories. Hence information about crop specific inputs are not collected during surveys but are often derived using engineering knowledge about the physiology of the crop and are estimated before the actual model is set up. Combining these two trends, the authors of the paper want to show that it is possible using proper estimation techniques to simultaneously estimate the coefficients of a farm supply model and the coefficient of the input allocation. We argue that this approach provides better estimates for input coefficients in an economically consistent theory formulation due to the first order condition applied in the input allocation model.

In order to develop our approach the paper is structured as following: In chapter one publication in the field of input-output and linear production coefficients estimation are reviewed. The second chapter will introduce the simultaneous model which is applied against a time series of German FADN data in North Rhine-Westphalia. The data aggregation and preparation steps are discussed in a subsequent chapter.

## 2. Production coefficients estimation

The term input allocation describes how aggregated input demand is “distributed” to production activities Britz (2005). The resulting activity input is called input coefficients. There is a long history of allocating inputs to production activities in agricultural analysis and the literature provides different approaches depending on the demand of the simulation model and the scope of the analysis. *Ad hoc* approach which infers the allocation based on information relying on sources such as published ‘industry standards’, agronomic field trials and expert opinions will not be further discussed.

### 1.1 Linear multiple regression function

The larger the regional coverage of a supply model becomes the lesser the manageable of ad-hoc determination of input coefficients. To overcome that and to keep the possibility to express regional or farm type variation of the production process several specifications of a linear multiple regression function were proposed, where total input use is treaded as dependent variable and the output of a firm as explanatory variable.

$$b = Ax$$

Herby,  $A$  is a  $n \times m$  matrix of technological coefficients with its elements  $a_{ij}$  representing the amount of input  $i$  required per unit of output  $j$ .  $x$  is a  $n \times 1$  vector of activity level.  $b$  is the available resource vector.

$$b_i^f = c_i + \sum_{j=1}^n a_{ij} x_j^f + u_i^f \quad \text{Equation 1}$$

Errington (1989) proposed a multiple regression or OLS technique applied to a system of derived demand as shown in Equation 1. He extended Equation 1 by adding a constant and an error term to account for deviation within the sample. He found very high  $R^2$  for most of the inputs whereas the constant for many of the inputs were not significant. However, through the OLS estimation he obtained negative estimates and in some cases the input coefficients sum up to greater than one, where the author concluded negative profits. Ray (1985) discussed several alternative estimation procedures based on mathematical programming. He emphasis that in view of the non-negativity demand for the input coefficients least square methods may lead to unacceptable estimates and outlined three alternative methods of inequality – constrained parameter estimation namely: inequality constraint least square, minimum mean absolute deviation and maximum absolute deviation. Midmore (1990) reviewed the findings by Errington and discussed the underlying technical assumptions and pointed out that if the commodity technology assumption<sup>1</sup> is implied and a correction for heteroscedasticity is applied, resulting from the size effect in production, the farm specific input coefficients can be estimated from regional farm business survey data. Further he noticed that using OLS techniques for (Equation 1) causes problems because the disturbance terms are not independent due to the accounting constraint involved, which dedicates that the columns of the coefficients matrix sum up to unity.

---

<sup>1</sup> Commodity technology assumption implies that the input structure of a commodity is the same, regardless of the industry (farm type in the context of agriculture) where it is produced Midmore (1990).

Dixon et al. (1992) viewed the estimation of linear programming model's technology coefficients as an application of random coefficients regression (RCR) and proposed an estimator, which restricts the predicted coefficient values to be non-negative.

All cited publications tried to overcome the problem of non-negativity of the estimated coefficients using standard estimation techniques, the first alternative formulation was offered by Moxey et al. (1994). He estimated Equation 1 by applying Bayesian econometrics. He argued that the use of Bayesian priors, applying a probability density function with zero probability to assign non-negativity, is a more natural mean of conducting constraint estimation and already argued to use additional a priori assumption based on information from other studies. In the same direction Léon et al. (1999) proposed the use of Maximum Entropy (ME) estimation techniques to directly introduce a priori information on the estimated coefficients to ensure nonnegativity. Léon et al. (1999) modified the system of linear equations (Equation 1) and defined a set of discrete "support values" within specified intervals. For each input  $i$  he set up the following reparameterisation of the unknown coefficient:

$$a_{ij} = z' p_{ij} \quad u_i^f = v' w_i^f$$

where the vector  $z$  and  $v$  are support values for defining the a priori information and  $p$  and  $w$  reflect the vectors of unknown probabilities. The model objective is to maximize the entropy measure done by maximizing equation

$$H(p, w) = -p \ln(p) - w \ln(w)$$

s.t.

$$b = xz' p + v' w$$

plus the normalisation constraints for the probabilities  $p$  and  $w$  to ensure that the probabilities sum to unity. Léon et al. (1999) presented different modifications of the basic model: First discussed the constrained estimation by adding a cross equation restriction on coefficient matrix  $a$  to ensure singularity. Further to account for zero observations he utilized a tobit variant of the linear model. He compared the Entropy results of the different model designs with those from classical estimation techniques, namely linear programming, OLS, and Bayesian estimation methods using accounting data from dairy/beef farms from France. Furthermore he tested the sensitivity of the GME outcomes compared to different designs of prior information. He concluded that the standard estimation techniques OLS, linear programming are no real alternative due to the problems aforementioned by the other authors and stated that it is difficult to discriminate between the Bayesian and the GME approach. This is not surprising due to the similarities of these estimation techniques as shown in Heckeley et al. (2005).

However the reviewed approaches to estimate input allocation are purely technical and lack any economical or behavioural concepts. The next subchapter sophisticates the estimation of input allocation by implying behavioural assumption together with economical considerations (functional form). Hereby the estimation of production function serves as the starting point.

## 1.2 Recovering allocatable input technology in production analysis

There exists a wide field of literature which investigated the compatibility of production theory of a firm with the possibility to recover the input allocation. Hereby we discuss the usefulness of the single equation multi product production function and briefly approaches based on duality.

### 1.2.1 Single equation multi-product production function

At the earlier time of production analysis an elegant way to avoid problems with missing input allocation was the use of *implicit representation of input per activity approach*. Hasenkamp (1976) used a single function such as  $f(y) = g(x)$  where outputs are denoted by  $y_i$ , and inputs by  $x_j$ . He represented output by a Constant-Elasticity-of-Transformation function<sup>2</sup> whereas the input function<sup>3</sup>, which measured aggregated input, was formulated as a Cobb-Douglas.

This multiple-output production function is a technical relationship that specifies possible output mixes that can be produced from each mix of input. The implicit representation of input allocation using a single function of this form was very popular, due to the assumed general representation of the multi-product multi-factor technologies Lence et al. (1998). However Mittelhammer et al. (1981) showed that the single joint production function is very restrictive. Let as shown this flaw by the following example:

$$f(y) = g(x)$$

where outputs are denoted by  $y_i$ , and inputs by  $x_j$ . Initialize the production function by a well known Cobb-Douglas formulation such as:

$$y_1 y_2^\delta = \alpha_0 x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3}$$

The production function is a technical relationship that specifies possible output mixes that can be produced from each mix of inputs. This production function does not explicitly explain how the allocation of inputs among alternative production activities affects output. Here a technology matrix A as in the linear regression context is not specified to explain the relation between inputs and outputs. Suppose a producer choose the amount of inputs to produce an amount of output. The inputs can be represented as:

$$x_j = x_{1j} + x_{2j}$$

where  $x_{kj}$  is the quantity of the jth input applied to produce the kth output. Thus, for the production function this will result in the following presentation, as

$$y_1 y_2^\delta = \alpha_0 (x_{11} + x_{21})^{\alpha_1} (x_{12} + x_{22})^{\alpha_2} (x_{13} + x_{23})^{\alpha_3}.$$

In the case that fertilizer (j=1) is applied to wheat, the production function (relationship) cannot determine whether wheat (k=1) or corn (k=2) production will be affected. Thus output determination by the inputs cannot be ensured. The single equation multi output production function approach can not be utilized to explain the relationship of inputs and output. Furthermore Just et al. (1983) p. 772 showed that the approach implies imposing arbitrary restrictions.

### 1.2.2 Approaches based on duality

Shumway et al. (1984) initiated a field of literature which investigated the compatibility of duality theory with the possibility to recover the input allocation. He discussed jointness and nonjointness in production. If corn supply responds to soybean and wheat prices then jointness in inputs for the pro-

<sup>2</sup> applied the following multi – output function  $f(x) = (\sum \delta_i y_i^c)^{1/c}$  whereas the parameter were chosen as  $0 < \delta_i; c > 1$  Hasenkamp (1976).

<sup>3</sup> The Cobb-Douglas input aggregated function was in the form as  $g(x) = A(\prod x_j^{\beta_j})$  whereas the CES function was chosen in the form  $g(x) = A(\sum \alpha_j x_j^\beta)^{1/\beta}$ .

duction is implied. Obviously inputs that are fixed but allocatable in a firm give rise to joint production since the quantity of a constraining input allocated to one specific output is not independent of a price of another output to which it could be also allocated. In a mathematical programming model jointness occurs due to the resource constraints on allocatable inputs which bind the activities or products into a single joint production system. When a farmer produces multiple products and allocatable fixed inputs are present, jointness is likely to occur. However jointness is less likely to be a problem for long run policy analysis than for short run and intermediate-run analysis. Regardless the source of jointness Shumway et al. (1984) discussed the problem of identifying the allocation of inputs among outputs using dual models. They concluded that even if allocation of inputs is observed the dual approach to production does not yield allocation equation for inputs.

Paris (1989) is responsible for correcting this misperception. Using a two stage shadow price approach, he demonstrated that a dual specification can be derived that contains all allocation equations. Although the envelope theorem does not directly recover input allocation equations from the multiple-output restricted profit function, it does recover the sum of the allocations and as long as the individual allocation equations sum to the total demand equation, they can be consistently estimated (Shumway (1995), page 185). Recent application of the duality based input allocation are Guyomard et al. (1996) and Lansink et al. (1997).

Lence et al. (1998) were obviously not aware of Paris's findings (1989) and took the critics of the dual approach of Shumway et al. (1984) paper to opt for another "less restrictive" approach to identify the allocation of inputs based on a production function in the absence of activity specific input allocation and under to use of non sample information.

### 1.3 Summary Literature review

The problem of defining activity related technology inputs coefficients is not new. A good deal of literature comes from the mathematical programming perspective, where input coefficients were estimated using a linear regression function (for allocatable variable and quasi fixed input) to completely represent the mathematical program (chapter 1.1). However this approach can be seen as pure technical device and it is not clear how the pre-determination of input output relationships will influence the final model. A second field comprehends recovering of input allocation based on production theory. Single equation multi-product production function was proposed but is very restrictive and output determination by inputs cannot be ensured. A further approach was to identify the allocation of inputs among outputs using dual models (chapter 1.2.2). This approach assumes optimality but is restrictive due to the functional form and does not take non sample or prior information into account.

If we assume that an allocation of inputs to outputs categories under a long run perspective is not only determine by plant-physiological reasoning, but as well by the aim of the manager to maximize profits and if we are confronted with the situation that only aggregated data on input use for a farm is available (which is often the case in policy analysis) and an explicit representation in the supply model is necessary then it seems to be straightforward to estimate the unknown coefficients of the farm supply model simultaneously, in order to account for the interrelationships of the economic model.

In the next chapter we investigate starting with a simple linear programming model under which assumptions the allocation of inputs can be derived. Shortcomings are discussed and the model extended to a mathematical farm supply programming model in the tradition of PMP (positive mathematical programming). Herby it is to emphasise that we clearly distinguish between allocatable inputs which are restricted by resource constraints and non restricted inputs such as fertilizer and seeding.

## 3. The theoretical model

In this chapter the theoretical model will be derived to simultaneously estimate the parameters of the supply model. The chapter will proceed along the following path: First we introduce a simple linear programming model and derive the optimality condition. Based on this we show in a subsequent chap-

ter under which conditions a linear model can be estimated using an information theoretic approach. Hereby it is always the aim to estimate the model parameters of a supply model in such a way that

- the economic assumption “profit maximization” holds
- that the model fits as best as possible to the observed data and
- available non-sample information are considered.

Afterwards, we discuss the shortcoming of the linear formulation and modify the model in a way that it fits to the requirements of a regional farm supply model.

#### 1.4 Linear model

We assume that a producer aims to maximise profits. This requires revenue exhaustion, i.e. marginal revenues must be equal to marginal costs simultaneously for all realised activities. This assumption can be represented by a normal linear programming model, where the producer tries to solve the following optimization problem

$$\max_x f(x) \quad s.t.$$

$$Rx \leq c, \quad [\lambda]$$

$$x \geq 0$$

where  $f(x)$  is the profit function represented by

$$f(x) = x[p'Y + s - w'A] \quad \text{Equation 2}$$

and  $x$  represents the vector of acreages for each of  $N$  land uses with  $j = 1, 2, \dots, N$ .

$Y$	$N \times N$ Matrix of yields
$p$	$N$ Dimensional vector of prices
$A$	$N \times M$ Matrix of input coefficients
$w$	Vector of input prices with $i = 1, 2, \dots, M$
$R$	$1 \times N$ Matrix of coefficient constraints, where as in $r_{1j} = 1$ for $j = 1, 2, \dots, J$
$c$	Vector of $c_1$ total land available and $c_2 = 0$
$\lambda$	Shadow value row vector with $1 \times M$

The model implies that the producer maximizing the total gross margins subject to resource constraints. The linear programming model could be solved if all coefficients (yield, input use, prices etc.) are available. The model can be equivalently described using the Lagrange method, where we assume that the optimal land allocations satisfy the land and political constraints in equality. This result in the first order conditions as

$$L(\lambda, x) = x[p'Y + s - w'A] + \lambda(Rx - c)$$

$$\frac{dL}{dx} = p'Y + s - w'A - \lambda R = 0$$

Model 1

$$\frac{dL}{d\lambda} = Rx - c = 0$$

where gamma ( $\lambda$ ) represents a row vector called the Lagrange multiplier. Denoting all observed variables with the subscript “o” we can write the first order condition in terms of resource constraints as

$$p^o ' Y^o + s^o - w^o ' A - \lambda R^o = 0 \quad \text{Equation 3}$$

$$R^o x^o - c^o = 0 \quad \text{Equation 4}$$

Further we add the data information on the aggregated input accounts as  $b^o$ , whereas the technology matrix A is not observable. Only total farm or sector purchases of various input categories are obtainable ( $b^o$ ).

$$Ax^o - b^o = 0 \quad \text{Equation 5}$$

The one dimensional matrix b comprehends inputs position such as total fertilizer use per farm, seeding, machineries, plant protection and others. For these positions no opportunity costs are attached because we assume that the farmer has no constraints for these inputs. In comparison, the allocatable resource constraints generating duals (shadow values) and hence are included in the first order condition of Equation 3. To estimate the model the technology matrix A, the duals of the resource constraints of the fixed but allocatable inputs have to be determined.

### 1.5 Estimation of a simple linear model under information theoretic approach

If we have more resource constraints than variables – since the number of nonzero activities in a linear programming is upper-bounded by the number of constraints – we could estimate a coefficients that results in a supply model which reproduces the observed practice. Because the model is underdetermined and ill-posed an appropriate information theoretic approach such as entropy or HDP has to be employed.

To illustrate this we are using Howitt’s illustrative programming model<sup>4</sup> and deriving the first order condition and setting up the following estimation model, whereby the assumption is that for the underlying data a set of true parameters  $\Psi$  of output prices, yields, subsidies, input coefficients and input prices of the model, exists that satisfy the Equation 3 and reproduce the observed practice and the dual values of the model as good as possible. The observed values, depicted by the subscript “o” in

Equation 3 are considered as the realisation of a random variable such as  $z$  that is conditional on  $\Psi \rightarrow L(z, \Psi)$ . It is assumed that there exists information or prior beliefs on the possible realisation of  $\Psi$ . The prior can be expressed as a prior density function  $p(\Psi)$ . This formulation is a Bayesian approach, where the model parameters are treated as stochastic outcomes. Hence it can be distinguished between the prior density  $p(\Psi)$ , which summarizes information collected as non sample information and the Likelihood function  $L(z, \Psi)$ . The combination of the prior density and the Likelihood function based on Zellner (1971), (p. 14) results in  $h(\Psi | z) \propto p(\Psi)L(z | \Psi)$  where  $\propto$  denotes proportionality. The posterior density ( $h$ ) allows drawing statistical inference about the parameter  $\Psi$  using probability statements. To obtain point estimates loss criteria has to be defined. The following model relates observed values to the true value.

$$\max_{\Psi} \{h(\Psi | z) \propto p(\Psi)L(z | \Psi)\} \quad \text{Equation 6}$$

This object function ( $h$ ) can be interpreted as the joint posterior density of the model parameters and is defined via the prior density  $p(\Psi)$ . This is multiplied by the likelihood function that assigns zero weights to values of  $\Psi$  that do not satisfy the model constraints and positive constant weights to values of  $\Psi$  that are compatible with the data and the model relationships Heckelee et al. (2005). Hence the values for  $\Psi$  that maximizes  $h$  is the Highest Posterior Density (HPD)-estimate of  $\Psi$ . For

---

<sup>4</sup> Howitt’s illustrative programming model (January 1994) can be downloaded from the Dept. of agricultural Economics website in U. C. DAVIS, CA, (e-mail rehowitt.ucdavis.edu)



Howitt's illustrative example the point estimates for an calibrated model can be obtained by taking the given technology matrix of the linear formulation ( $R^{prior}$ ), the resource input costs ( $w^{prior}$ ) of the linear formulation and the dual values of the resources ( $\lambda^{prior}$ ) as prior mean. The assumed prior density function has the form like  $vec(\Psi) \sim N(vec(z), \Sigma)$ . The covariance matrix  $\Sigma$  is set equal to a diagonal matrix with elements  $(vec(z)0.05)^2$ , the square taken element wise. Taking natural logs and restricting the objective function to terms that are relevant for the optimization we derive the following estimation problem (see also Equation 9).

$$\begin{aligned} & \text{Minimize} \\ & vec((\lambda - \lambda^{prior}), (R - R^{prior}), (w - w^{prior}))' \\ & \times \Sigma^{-1} vec((\lambda - \lambda^{prior}), (R - R^{prior}), (w - w^{prior})) \end{aligned} \quad \text{Equation 7}$$

subject to Equation 3 and Equation 4. The results of this estimation approach is illustrated in the next table using the observed land allocations and observed gross margins from Howitt's data set. The didactic example comprehends three crops and the observed allocation is depicted in Table 1.

Table 1 Allocation of the simple didactic LP Program

	LP	C_LP	A_LP
COT	2,65	1,49	1,49
WHT	0,00	0,62	0,62
RI	0,00	0,54	0,54

Source: Own calculations based on Howitt's PMP example

The observed land allocation for the three crops is the same as in column C\_LP. Column "LP" represents the optimal allocation using the original LP program from Howitt (1994). The C\_LP represents the optimal solution using the calibration constraints and the last column depict the solution where the technology matrix, dual values and the input costs are estimated based on prior information taken from Howitt's example. Table 2 depicts the differences between the original and the adjusted technology matrix of the allocatable inputs. An overview on the dual values and the cost adjustments is given in Appendix 1. The adjustment was done under the assumption that the land technology coefficient kept unchanged by 1.

Table 2 Comparison of the technology matrix amongst LP/C\_LP (Original) and A\_LP (Adjusted)

		Original	Adjusted
COT	LAND	1	1
	WATER	3,00	3,37
	CAPITAL	409,40	465,55
WHT	LAND	1	1
	WATER	1,84	1,67
	CAPITAL	564,52	495,98
RI	LAND	1	1
	WATER	5,70	4,89
	CAPITAL	962,96	886,36

Source: Own calculations based on Howitt's PMP example

The adjusted linear programming (A\_LP) exactly reproduce the observed data due to the first order condition in Equation 3. Comparing the original and the adjusted values it can be seen that the estima-

tion framework tries to minimize the quadratic deviation of the posterior and the prior settings. Even although the model could be estimated in such a way that it exactly reproduces the observed practice such a simple LP calibration approach has some unattractive properties that it rarely meets the requirements for applied policy analysis.

- i. Often a suitable number of empirical justified constraints are rarely available. Particular in farm group models which represents a greater population of farms.
- ii. Further the adjustment of the crop allocation to price and policy changes will be discontinuous and consequently does not realistically represent the behaviour of an aggregated production model. A continuously adjustment would be more suitable.
- iii. Economical concepts such as return to scale or diminishing costs are not considered<sup>5</sup>.
- iv. The LP formulation does not allow a concept for catching the supply response estimated from time series or multiple observations. No behaviour parameters are included. For instance priority believes on elasticities are not possible to include.

The small linear sample showed that allocatable input use could be adjusted in order to reproduce the observed practice. It should be emphasised here, that the differences of the priors versus the posterior point estimates of the technology matrix will increase when the input cost and dual are taken as certain outcomes. Nevertheless, the exercise demonstrated that as long as data can be seen as outcome of a random variable the model parameters can be estimated when taking relevant non-sample information into account. In the next chapter this small example will be extended that it meets the requirements of a farm supply model appropriate for policy analysis.

## 1.6 Farm supply model

To heal the mentioned problems of the linear example a quadratic term may be added to the objective function in Equation 2. Such a non-linear term can be interpreted analogue to the Positive Mathematical Programming (PMP) method introduced by Fajardo et al. (1981) and Howitt (1995). The term will be used to ensure that the response behaviour will be continuously and the observed practice can be reproduced. In order to capture *unknown* cost positions, aggregation errors and more the following quadratic cost function is added to Equation 2.

$$C^v = d'x + \frac{1}{2}x'Qx$$

And the first derivative of the compound objective function<sup>6</sup> results in

$$p^o'Y^o + s^o - w^o'A - \lambda R^o - d - x^oQ = 0 \quad \text{Equation 8}$$

---

<sup>5</sup> Could be only added using linear approximation which includes the need for additional constraints or/and variables in the linear programming.

<sup>6</sup> Note that for positive semi-definiteness of the Hessian matrix it is sufficient that Q is positive semi definite, which can be satisfied by the Cholesky factorisation with the upper triangular matrix U, and that all elements of Q are non-negative.  $Q = LL'$

The aim of the paper can now be precisely expressed as estimating the parameters of the first derivative of the non-linear cost function  $(\lambda, d, Q)$  simultaneously with the input coefficient matrix  $(A)$ . Further we assume that the producer is a price taker in outputs and variable input prices. For the sake of convenience we summarize the full model and adding the time index  $t$  and subscript all observed data with “o”.

$$p_t^o Y_t^o + s_t^o - w_t^o A_t - \lambda_t R_t^o - d - x_t^o Q = 0$$

$$R_t^o x_t^o - c_t^o = 0 \quad (\lambda)$$

$$A_t x_t^o - b_t^o = 0$$

$$Q = LL'$$

Model 2

#### 4. The empirical model

Considering Model 2 it is obvious that we are confronted with an underdetermined and ill-posed problem. The following parameters have to be estimated: the  $M \times T$  unknown vectored of dual values  $(\lambda)$ , an  $N \times 1$  vector of linear parameters of the cost function  $(d)$ , and the  $M \times N \times T$  matrix of technology  $A$  for variable inputs as well as the  $N \times N$  quadratic Matrix  $Q$ . In farm group models resource constraints equation (second equation in Model 2) are often less than the number of parameters to estimate  $(K)$ . In our farm supply model only land rent and set-aside is considered. If  $M < K$  the system of equations underdetermines the unknowns and thus there exists, assuming the equation is consistent and a solution exists, an infinite number of solution that solves the system<sup>7</sup>. However to be able to identify a unique solution for the proposed Model 2 one may build (based on Equation 6) the following maximization problem, similar to the formulation in the linear example. As in Equation 6 the highest posterior density values can be found when the arguments are found that maximize the prior beliefs  $p(\Psi)$  subject to the Likelihood function. This can be formulated as a maximization problem.

$$\max p(\Psi)$$

Equation 9

$$s.t. \quad g(\Psi, x^o, \lambda^o)$$

To obtain an applicable model the Bayesian framework of Equation 9 will be added to Model 2. Assuming a multivariate normal density function for  $p(\Psi)$  in the form of

$$p(\Psi) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\Psi - \Psi^{prior})' \Sigma^{-1} (\Psi - \Psi^{prior}) \right]$$

and since the value that maximizes some function also maximizes the natural logarithm, we can take the logarithm of the objective function and by reducing the remaining to the elements which are important for the maximization of the objective function it becomes

min

$$vec(\Psi - \Psi^{prior})' \times \Sigma^{-1} vec(\Psi - \Psi^{prior})$$

---

<sup>7</sup> Assuming a linear system, as we investigated in chapter 1.5, the number of non-zero observation cannot exceed the number of resource constraints. To overcome this a quadratic term was introduced.

where  $\Sigma^{-1}$  is the covariance metrics and  $\Psi$  are the true values of the data generating process. The observed values  $z = (x^o, p^o, Y^o, w^o)$  (crop allocation, output prices, yield, input prices are outcomes of a random variable and hence an error term is added to this observation. Assuming there exist a set of true parameters  $\Psi(p, Y, A, w, d, Q)$  that satisfy the first order condition. It is further assumed that subsidies ( $s$ ), price index, set-aside rate, and the total land constrain are known with certainty and that the producer has naïve price expectations for output and input prices ( $p, w$ ) and that yield is modelled over time via a linear trend. This prepositions leads to the following model<sup>8</sup>.

min

$$\text{vec}(\varepsilon_t^x - 0, \varepsilon_t^Y - 0, \varepsilon_t^w - 0, (A_t - A_t^{\text{prior}}), (\lambda_t - \lambda_t^{\text{prior}}))'$$

$$\times \Sigma^{-1} \text{vec}(\varepsilon_t^x - 0, \varepsilon_t^Y - 0, \varepsilon_t^w - 0, (A_t - A_t^{\text{prior}}), (\lambda_t - \lambda_t^{\text{prior}}))$$

s.t.

all equations from Model 2

$$x_t^o = x_t + \varepsilon_t^x$$

$$Y_t^o = \beta_0 + \beta_1 T_t + \varepsilon_t^Y \quad \text{Model 3}$$

$$p_t = p_{t-1} + \varepsilon_t^p$$

Model 3 is estimated using Conopt4 which is part of the mathematical software package GAMS. The prior information for land duals was taken, where available, from the farm accounting network, the prior for the coefficient matrix A were taken from management handbooks.

## 5. Data Preparation

The model presented in chapter 4 is applied to a set of accounting data from North Rhine Westphalia, taken from the Farm Accounting Data Network (FADN)<sup>9</sup>, which belongs to the data network of the EU designed for the purpose of estimating costs and income level for the European agriculture. The farm groups are built up - by reason of confidentiality limits - on at least 15 farm accounts from 1997 to 2005. The farms are selected identically over time, which ensures a consistent data set.

Table 3 Number of accounts per farm group aggregated by the principle type of farming

	Specialist permanent Crops	Specialist grazing livestock	Mixed crops- livestock	Mixed cropping		SUM
NutsII	Specialist horticulture		Mixed livestock	Specialist field crops	Specialist granivore	
DEA1		25	6	16	13	71
DEA2	1	14	5	16	2	40
DEA3		73	2	26	65	166
DEA4	1	23	2	47	35	112
DEA5		51	2	6	13	72
	2	186	17	111	128	461

<sup>8</sup> Prior information about elasticities was also included in Model 2. For the sake of clarity the author did not present the equation but directly refers to Jansson (2007), page 29, equation (16).

<sup>9</sup> The aggregated FADN were provided by vTI – Farm Economics; Johann Heinrich von Thünen-Institute - Federal Research Institute for Rural Areas, Forestry and Fisheries

The specialized field crop farms in region DEA4 are selected to simulate and present the proposed input allocation for chapter 6. This particular farm group was chosen, because it is well represented during the simulation years from 1997 to 2005. Based on the information of each single farm account the production structure, prices, costs were mapped into the positions presented in the Appendix 3. Built up on the farms an average group was aggregated. Some production data of the selected farm group DEA4 is given in Appendix 2.

## 6. Results

In order to compare the effect of a simultaneously estimation of the input allocation two models were estimated and compared in this chapter. The first model is Model , for the remainder of the text indicated by “fixed model”, where the input allocation was pre-estimated and enters the model with the condition  $A_t = A_t^{prior}$ . In the second model (“sim model”)  $A_t$  is determined by the objective and the Likelihood function (constraints of Model ). The hypothesis for the simulation is that due to the optimality condition the simultaneously estimated input allocation (“sim model”) will be less restrictive than its counterpart. Furthermore we assume that the “sim model” produces better estimates regarding the observed values (crop allocation, prices and yields).

Table 4 shows the share of explained variation for crop allocation, prices and yields. In the most cases the simultaneous model specification considerably improved the goodness of fit for land allocation and prices. The fit for yields is in general lower due to the linear trend in both model specifications.

Table4: R<sup>2</sup> measure for the “fixed” and “sim” model - prior versus posterior fit

R <sup>2</sup>	fixed model			sim model		
	Crop allocation	Output prices	Yields	Crop allocation	Output prices	Yields
SWHE	0,58	0,30	0,76	0,83	0,44	0,77
RYEM	1,00	0,69	0,45	1,00	0,69	0,46
BARL	0,46	0,25	0,71	0,63	1,00	0,71
OATS	0,97	0,13	0,80	1,00	0,31	0,80
MAIZ	0,75	0,77	0,53	1,00	0,81	0,52
OCER	0,87	0,51	0,79	1,00	0,63	0,79
RAPE	0,70	0,94	0,56	0,99	0,97	0,56
PULS	1,00	0,96	0,34	1,00	0,97	0,33
POTA	0,30	0,96	0,48	1,00	0,96	0,45
SUGB	0,54	0,12	0,63	0,75	0,16	0,62
MAIF	1,00	0,19	0,76	1,00	0,40	0,80
OFAR	1,00	0,52	0,85	1,00	0,80	0,85
NONF	0,99	0,63	0,51	1,00	0,74	0,51
FALL	0,99	0,00	0,00	1,00	0,00	0,00
OSET	1,00	0,00	0,00	1,00	0,00	0,00

Source: Own calculations

Figure 1 depicts the prior versus posterior values for soft wheat allocation. It can be seen that the “sim” model outrange the “fixed” model estimates due to the possibility to adjust the input allocation between the crops within the total input constraints.

Figure 1: Plot prior versus posterior for soft wheat

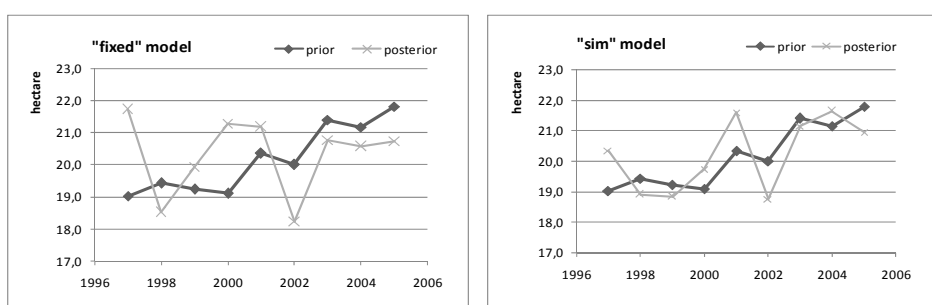
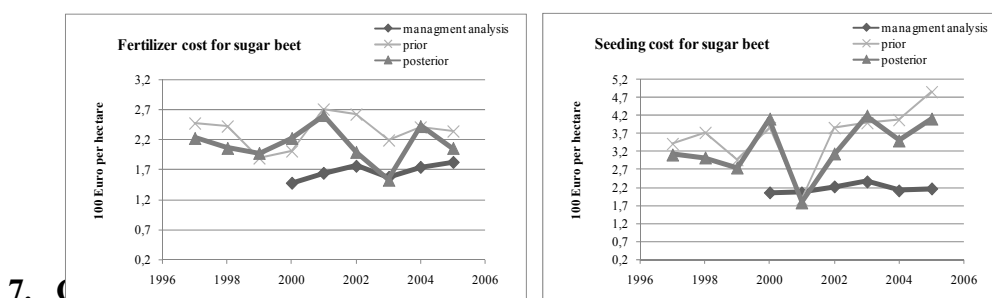


Figure 2 presents the deviation of the input allocation coefficients between the prior values, the posterior and data derived from a survey of several farms in the investigated region. It can be seen that the input allocation for sugar beet does not deviate too much from the prior values. However, the posterior fits well, at least for the fertilizer coefficients, to the survey data of the region. It should be emphasised again that the changes are due to the simultaneously estimated input coefficients under the optimality condition.

Figure 2 Comparison of estimated input costs for sugar beet (prior, posterior and survey data)



The papers point at how input allocation per crop under optimality condition can be estimated utilizing a farm sample from FADN. By employing proper estimation techniques all unknown coefficients of the mathematical farm supply model could be estimated. Furthermore it could be revealed that a simultaneous estimation of all parameters considerably improves the goodness of fit of the estimates. However, it is necessary to further develop this approach. The sensitivity of the posterior input allocation estimates has to be further investigated. Furthermore, additional expert data for other crops have to be collected to compare the posterior estimates and to be able to draw conclusion on whether such an approach will lead to a better representation of input coefficients in farm supply models.

## 8. Acknowledgments

The author wishes to thank T. Jansson, who kindly provided his routines for estimation supply models. Due to his work it was possible to quickly setup a estimation framework to investigate the input allocation under optimality condition. Further the author appreciates the financial support of the Schaumann foundation for the 107<sup>th</sup> EAAE conference.

## 9. References

Britz, W.2005. CAPRI Modelling System Documentation; Bonn University, Agricultural and Resource Economics.

- Buyse, J., Fernagut Bruno Harmignie, O., Frahan, d., Henry, B., Lauwers, L., Polome Philippe Van Huylenbroeck, G., Meensel, V., and Jef., 2007. Farm-based modelling of the EU sugar reform: impact on Belgian sugar beet suppliers. *Eur Rev Agric Econ* 34: 21-52.
- Chambers, R.G., and Just, R.E. 1989. Estimating Multioutput Technologies. *American Journal of Agricultural Economics* 71: 980-995.
- Dixon, B.L., and Hornbaker, R.H. 1992. Estimating the technology coefficient in linear programming models. *American journal of agricultural economics* 74: 1029-1039.
- Errington, A. 1989. Estimating Enterprise Input-Output Coefficients from Regional Farm Data. *Journal of Agricultural Economics* 40: 52-56.
- Fajardo, D., B. A. McCarl; R. L. Thompson, and R. L. Thompson,. 1981. A Multicommodity Analysis of Trade Policy Effects: The Case of Nicaraguan Agriculture. *American Journal of Agricultural Economics* 23-31.
- Golan, A., Judge, G., and Miller, D. 1997. *Maximum entropy econometrics* Wiley,Chichester [u.a.].
- Guyomard, H., Baudry, M., and Carpentier, A. 1996. Estimating crop supply response in the presents of farm programmes: application to the CAP. *European Review of Agricultural Economics* 23: 401-420.
- Hasenkamp, G. 1976. A Study of Multiple-Output Production Functions. *Journal of Econometrics* 4: 253-262.
- Heckelei, T., Mittelhammer, R., and Britz, W. 2005. *A Bayesian Alternative to Generalized Cross Entropy* Parma : Monte Universita Parma Editore,Italy, Parma.
- Howitt, R.E. 1995. Positive Mathematical Programming. *American Journal of Agricultural Economics* 77: 329-342.
- Jansson, T. 2007. Estimation of supply response in CAPRI; Bonn University, Agricultural and Resource Economics, Discussion Paper 2007:2
- Just, R.E., Zilbermann, D., and Hochman, E. 1983. Estimation od Multicrop Production Function. *American Journal of Agricultural Economics* 65: 770-780.
- Lansink, A.O., and Peerlings, J. 1997. Effects of N-Surplus Taxes: Combining Technical and Historical Information. *European Review of Agricultural Economics* 24: 231-47.
- Lence, S.H., and Miller, D.J. 1998. Estimation of multi-output production functions with incomplete data: A generalised maximum entropy approach. *European Review of Agricultural Economics* 25: 188-209.
- Léon, Y., Peeters, L., Quinqu, M., and Surry, Y. 1999. The Use of Maximum Entropy to Estimate Input-Output Coefficients From Regional Farm Accounting Data. *Journal of agricultural Economics* 50: 425-439.
- Midmore, P. 1990. Estimating Input-Output Coefficients from Regional Farm Data. *Journal of Agricultural Economics* 41: 108-111.

Mittelhammer, R.C., Scott C. Matulich; Bushaw, and Bushaw, D. 1981. On implicit Forms of Multiproduct-Multifactor Production Functions. *American Journal of Agricultural Economics* 63: 164-168.

Moxey, A., and Tiffin, R. 1994. Estimating linear production coefficients from farm business survey data: A Note. *Journal of Agricultural Economics* 45: 381-385.

Paris, Q. 1989. A Sure Bet on Symmetry. *American Journal of Agricultural Economics* 71: 344-351.

Ray, S.C. 1985. Methods of estimating the input coefficient for linear programming models. *American journal of agricultural economics* 67: 660-665.

Shumway, C.R. 1995. Recent Duality Contributions in Production Economics. *American Journal of Agricultural Economics* 20: 178-194.

Shumway, C.R., Pope, R.D., and Nash, E.K. 1984. Allocatable Fixed Inputs and Jointness in Agricultural Production: Implications for Economic Modeling. *American Journal of Agricultural Economics* 66: 72-78.

Zellner, A. 1971. *An introduction to Bayesian inference in econometrics* Wiley, New York



## Appendix 1

Table 4 Comparison of duals values between the original and the adjusted version

	Original	Adjusted
LAND	28,3	28,6
WATER	5,2	5,3
CAPITAL	0,0	0,1

Source: Own calculations based on Howitt's PMP example

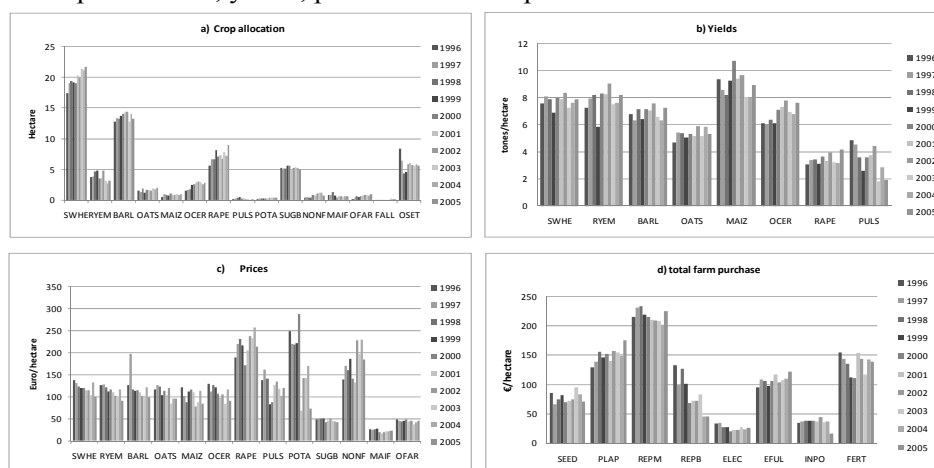
Table 5 Comparison of the resource input costs

		Original	Adjusted
COT	LAND	264,3	413,1
	WATER	25,6	30,3
	CAPITAL	0,1	0,1
WHT	LAND	112,0	96,4
	WATER	25,6	24,2
	CAPITAL	0,1	0,1
RI	LAND	196,9	180,5
	WATER	25,6	24,2
	CAPITAL	0,1	0,1

Source: Own calculations based on Howitt's PMP example

## Appendix 2

Figure 1 Observed crop allocation, yields, prices and total input use



## Appendix 3

Table 6 Crops modelled

Crops	Description
SWHE	Soft Wheat
RYEM	Rye
BARL	Barley
OATS	Oats
MAIZ	Maize
OCER	Other Cereales
RAPE	Rapeseed
PULS	Pulses
POTA	Potatoes
SUGB	Suger beet
MAIF	Fodder Maize
OFAR	Silage grass
NONF	Ind. Rapseed
FALL	Obligatory set-aside
OSET	fallowed land

Table 7 Inputs

Input	Description
SEED	Seeding
PLAP	Plant Protection
REPM	Repairments machinery
REPB	Repairments buildings
ELEC	Electricity
EFUL	Fuel
INPO	Other inputs
FERT	Fertilizer