Estimating parameters of a constrained non-linear optimisation model using multiple observations*

*Torbjörn Jansson***
*Marcel Adenäuer***

Abstract
This article implements a method for estimating the behavioural parameters of a non-linear agricultural supply model using multiple observations of prices and model outcomes. The focus is on developing a workable method for calibrating the crop sector of the modelling system CAPRI to the outcomes of simulation experiments with farm models. Lacking real data, the work is carried out for a didactic model using a synthetic data set.

Keywords: non-linear programming, PMP
JEL-classification: C13, F11

1 Introduction

Calibration methods for constrained mathematical programming models have been a vividly discussed topic since the publication of "Positive Mathematical Programming" (PMP, Howitt 1995). Less attention has been paid to the estimation of the parameters influencing the simulation behaviour of such models. Heckelei and Wolff (2003) estimate first order conditions of constrained optimisation models using synthetic data sets. The contribution of this paper is twofold: Firstly, we propose a new formulation for the CAPRI (Britz 2005) crop supply model, incorporating endogenous yields and input use, and secondly, we adapt the method suggested by Heckelei and Wolff for the estimation of key behavioural parameters of the proposed model.

The work is part of a project that aims at developing a technique for calibrating the supply response of the CAPRI supply models to that of farm level models, by generating a set of simulation results with the farm models which serves as a sample of observations in an estimation of some of the parameters in CAPRI. In the extension, the developed technique should for the basis for estimation of the parameters of CAPRI using a time series of observations of prices, production and input use.

---

* Paper submitted to the Ecomod conference on Regional and Urban Modeling, Brussels 2006

** Institute for food and resource economics, Bonn University, Germany. Corresponding author T. Jansson, jansson@agp.uni-bonn.de

In this article it is assumed that sufficient information for identifying the technical coefficients of the yield and input use functions is supplied from an exogenous source. Thus only a subset of all parameters of the supply model need to be determined using the sample of simulation results from the farm models. Furthermore, the application has the character of explorative research, because neither the new CAPRI supply model nor the external sources of technical coefficients or the farm level models that will eventually be used for estimating it yet exist. However, this article develops assumptions and an algorithm that are sufficient for estimating the model, and the technique is demonstrated using a didactic model. The method is based on explicit estimation of the Kuhn-Tucker conditions as well as sufficient second order conditions.

The work is outlined as follows: First we describe the microeconomic assumptions underlying the new supply model, and the optimisation problem of the producers is formulated. Then, we describe what kind of data and *a priori* information is available and which parameters are to be determined in the estimation. Having done that, we propose a method for computing and estimating the parameters. Finally, the method is evaluated using synthetic data. Some weak points are pointed out, and upcoming research needs are indicated.

2 A microeconomic model of crop supply

The producer of crops in this didactic-size suggestion for new CAPRI regional supply models is assumed to allocate inputs to production activities in order to maximise the sum of profits, restricted by technological and political constraints.

There are three types of inputs: "land" (l), "plant protection" (x) and "other variable inputs" (a), which are to be explicitly allocated to the set of production activities containing cereals, oilseeds, potatoes, fodder, set-aside (voluntary and obligatory) and fallow land. In what follows, land and plant protection are treated separately from other inputs, and the term "inputs" will refer only to the latter. There is only one such other input, called "REST", in this application.

The profit function (1) is the sum of profits of the individual production activities (the first term), and a term that is a quadratic function of activity levels. It is the parameter vector c and the square matrix B of that quadratic term that are to be estimated using the multiple observations. Because of the relationship to PMP, that term is henceforth referred to the PMP-term.

The technological constraints, except for the land constraint (5) and the non-negativity constraints (7), come in the form of a functional relation f of yield to use of land, plant protection and land, and a land constraint (2). Other inputs enter the yield function in a Leontieff manner, so that only the limiting factor of inputs and plant protection determines yield. This been formulated using the auxiliary variable g which enters the yield function and Leontieff technology constraints (eq. 3 and 4). The lower case Greek letters α, β, χ, δ, γ and η denote technical coefficients that need to be determined using other information than simulation experiments with the farm level models.

The only political constraint is the set-aside requirement (6) with the activity SETA = “set-aside”, ARB the set of activites subject to the set-aside constraint (cereals, oilseeds and set-aside itself) and the (exogenous to the estimation) parameter r the regional set-aside rate—the general set-aside rate corrected by the fraction of small producers in the region.

$$\max_{l,x,a} \sum_j l_j \left(y_j p_j + s_j - x_j v_j - \sum_i a_{ij} w_i \right) - \left(\sum_j c_j l_j + \tfrac{1}{2} \sum_{jk} l_j l_k b_{jk} \right) \tag{1}$$

$$\text{subject to} \quad y_j = f(l_j, x_j, g_j) = \alpha_j + \beta_j \frac{l_j}{L} + g_j \tag{2}$$

$$\chi_j (x_j)^{\delta_j} \geq g_j \tag{3}$$

$$\gamma_{ij} + \eta_{ij} a_{ij} \geq g_j \quad \forall\, i \tag{4}$$

$$\sum_j l_j \leq L \tag{5}$$

$$r \sum_{j \in ARB} l_j - l_{'SETA'} \leq 0 \tag{6}$$

$$l_j,\, x_j,\, a_{ij} \geq 0 \tag{7}$$

The formulation can be simplified. Because the prices of all inputs and of plant protection are strictly positive, profit maximisation requires the inequalities 3 and 4 to be satisfied with equality. Then choosing the level of one input determines the use of all other inputs, so that one input can be chosen as *numeraire*. For this purpose we chose plant protection and substituted the left hand side of (3) for g_j everywhere.

Furthermore, the land constraint (5) can be replaced by an equality by introducing the activity “fallow land”, and the set-aside constraint (6) by redefining the variable $l_{'SETA'}$ = “set-aside” as $l_{'OSET'} + l_{'VSET'}$ with $OSET$ = “obligatory set-aside” and $VSET$ = “voluntary set aside” and the corresponding change in the set ARB. Finally, the non-negativity constraint for activity levels can be omitted if we restrict our interest to production activities that are actually observed in the region, refraining from estimating the dual values of the same constraint for activities that are not currently observed, and similarly for inputs and plant protection. Making those modification yields the equality constrained regional supply model given by equations (8-12), the objective function repeated for reference.

$$\max_{l,x} \sum_j l_j \left(y_j p_j + s_j - x_j v - \sum_i a_{ij} w_i \right) - \sum_j c_j l_j - \tfrac{1}{2} \sum_{jk} l_j l_k b_{jk} \tag{8}$$

subject to $$y_j = \alpha_j + \beta_j \frac{l_j}{L} + \chi_j (x_j)^{\delta_j} \tag{9}$$

$$a_{ij} = \frac{\gamma_{ij}}{\eta_{ij}} - \frac{\chi_j}{\eta_{ij}} (x_j)^{\delta_j} \tag{10}$$

$$\sum_j l_j = L \tag{11}$$

$$r \sum_{j \in ARB} l_j - l_{'OSET'} = 0 \tag{12}$$

3 Sources of data and a priori information

The supply behaviour of the model (8-12) depends on all parameters denoted by lower case Greek letters plus the vector $c = (c_j)$ and the matrix $B = (b_{jk})$. All parameters will not be possible to identify given that the farm model simulation experiments to which the model shall calibrate only contain output prices and production, i.e. lacking technological information about input use per activity and production at different input/output price ratios. In order to identify all parameters, additional information is needed.

Information for the estimation comes from three sources: (*i*) Observations of aggregated supply behaviour of farm models, (*ii*) input allocation to individual activities in the form of a complete and consistent *base year* solution, estimated in another project using data from the *farm accountancy data network* (FADN), and (*iii*) information on second order effects of land share, plant protection and other inputs. The information is matched with the parameters of the model as follows:

To (*i*): The aim of the exercise is to estimate the supply behaviour using observations of prices, subsidies, and production, resulting from the aggregate behaviour of a number of farm level model. To this end, only the matrix B is required.

To (*ii*). We require the model to calibrate perfectly to a complete and consistent historical data set called the *base year*, containing explicit and complete input allocations to activities. That information is sufficient to identify the yield coefficient vector α, the input use parameter matrix γ and the vector c.

To (*iii*). Four types of additional information is required for identifying the parameters η, δ, β, λ and ρ, representing the marginal effect of plant protection on other input use, second-order effects of plant protection use, the marginal effect of land share on yields, and the dual values of land and set-aside constraints. It is planned that those, or equivalent information, are estimated in a previous step, using farm level data from the *farm accountancy data network* (FADN). That estimation is not reported in this article, but simply assumed to be present.

The parameters and the different sources of information utilized are shown in the following table. It has been assumed that the type (*iii*) information from the FADN sample is in the form of point elasticities.

Table 1: Information utilized in estimation (numbers refer to equations)

Param.	Dim.	Source (*i*): Farm models	Source (*ii*): Base year calibration	Source (*iii*): FADN estimation	Eq.
α	$n \times 1$		yield function		9
β	$n \times 1$			$\varepsilon(y_j, l_j/L)$	15
χ	$n \times 1$		first order cond (x)		14
δ	$n \times 1$			$\varepsilon(x_j, v)$	16
γ	$n \times m$		input use function		10
η	$n \times m$			$\varepsilon(a_{ij}, x_j)$	17
c	$n \times 1$		first order cond. (l)		13
B	$n \times n$	simulated sample	second order cond.		18,19
λ	1			scalar	-
ρ	1			scalar	-

4 Estimation

As mentioned, it is the primary purpose of the estimation to fit the model as closely as possible to the aggregate supply behaviour of farm level models, constructed in another research project (SEAMLESS, reference) by spatial disaggregation of the regions in CAPRI into homogeneous units. The farm level models are used in a series of simulations to obtain a set of different price-quantity combinations for each model region. The set of price-quantity combinations could, abusing notation, be referred to as "observations", due to the similarity with a random sample, albeit the disturbance terms in our case will be due to differences in model specification rather than measurement errors.

The key idea of the estimation is that for each price in the set of observations, the model solution should come as close as possible to the observed production. Therefore, one could consider the estimation a bilevel optimisation problem similar to a Stackelberg game. Here, the leader is the economist performing the estimation. He selects his instruments, being the parameters and prices of the programming model, in order to influence the follower, being the profit maximising producer represented by the supply model, to choose activity levels and netputs so that the leader's cost function, being the estimation criterion, is minimised. The leader's constraints are, apart from (*i*) the profit maximising behaviour of the producer, that (*ii*) the base year be reproduced *ex post* and that the *a priori* marginal effects (*iii*) be recovered.

A common approach to the solution of bilevel programs is to replace the follower's problem by it's Kuhn-Tucker (KT) conditions and if necessary the second order optimality conditions. When the bilevel program is an estimation it would mean to minimise some estimation criterion subject to KT, second order and other constraints. In general, such a programming problem is difficult to solve due to the complementary slackness conditions contained in the Kuhn-Tucker (KT) conditions when inequality constraints are present. However, in the formulation in equations (8-12), there are only equality constraints, so the estimation can be formulated as an ordinary equality constrained non-linear programming problem.

We start by deriving the first and second order conditions for optimal activity levels (13) and plant protection use (14). The second order conditions for optimal activity levels are stated further below (19).

$$\left(\alpha_j + 2\beta_j \frac{l_{tj}}{L} + \chi_j (x_{tj})^{\delta_j}\right) p_{tj} + s_{tj} - x_{tj} v_t - \sum_i \left(\frac{\gamma_{ij}}{\eta_{ij}} - \frac{\chi_j}{\eta_{ij}} (x_{tj})^{\delta_j}\right) w_{ti} \\ - c_j - \sum_k l_{tk} b_{jk} - \lambda_t - \rho_t r I_{ARB}(j) + \rho_t I_{\{OSET\}}(j) = 0 \tag{13}$$

$$\delta_j \chi_j (x_{tj})^{\delta_j - 1} \left(p_{tj} - \sum_i \frac{w_i}{\eta_{ij}} \right) - v_t = 0 \tag{14}$$

The parameters of the model (8-12) can be computed in three steps:

<u>Step 1</u>: Use the elasticities $\varepsilon(y_j, l_j/L)$ and $\varepsilon(x_j, v)$ evaluated at base year values of variables to determine β_j and δ_j. Explicit functional forms for the elasticities are derived by differentiation of the yield function (9) to obtain

$$\varepsilon(y_j, \frac{l_j}{L}) = \frac{\partial y_j}{\partial (l_j/L)} \frac{l_j^*/L}{y_j^*} = \frac{\beta_j l_j^*}{y_j^* L} \Leftrightarrow \beta_j = \frac{\varepsilon(y_j, l_j/L) y_j^* L}{l_j^*}, \tag{15}$$

and reformulating the first order conditions with respect to x_j to obtain

$$\varepsilon(x_j, v) = \frac{\partial \ln x_j}{\partial \ln v} = \frac{-1}{1-\delta_j} \Leftrightarrow \delta_j = \frac{1}{\varepsilon(x_j, v)} + 1. \tag{16}$$

<u>Step 2</u>: Determine the values of α, χ, γ and η using four equations given by the requirements that the base year uses of inputs and plant protection are optimal plus the condition that the elasticities of input use to plant protection, $\varepsilon(a_{ij}, x_j)$, match the *a priori* values. The four equations necessary to identify the four parameters are the first order conditions for optimal use of plant protection (14), the yield function (9), the Leontieff input requirement equation (10) and the following definition of the aforementioned elasticity, all evaluated at base year values l^*, x^*, a^* and y^*.

$$\varepsilon(a_{ij}, x_j) = \frac{\delta_j \chi_j}{\eta_{ij} a_{ij}^*} x_j^* \tag{17}$$

Step 3: Now only c, B and the dual values λ and ρ remain to be determined. This is done by minimising the sum of squared deviations (18) from the set of simulation experiment outcomes of the farm level models, outcomes indexed by t, subject to that for each t, the estimated l and x satisfy the Kuhn-Tucker and second order conditions for optimality, and that for the base year price vectors (p, w and v) precisely the base year variable levels (stars as superscripts) result. I.e. solve

$$\min_{c,B} \sum_{tj} \left(l_{tj}^s - l_{tj} \right)^2 \tag{18}$$

subject to first order conditions for the observations, (eq. 13, 14), the constraints of the final model (9-12), and the following second order condition for activity levels, where negative semidefiniteness of the Hessian matrix H is ensured by a Cholesky factorization of B into the square of an upper triangular matrix u,

$$\frac{\partial^2 L}{\partial l_j \partial l_k} = \frac{2\beta_j p_j}{L} I_j(k) - b_{jk} = H \text{ is negative semidefinite for all } p > 0,$$

$$\Rightarrow B = u'u \text{, given L > 0 and } \beta < 0 \tag{19}$$

and subject to base year calibration conditions for activity levels, i.e. equation 13 evaluated at the base year variable levels.

5 Explorative implementation and numerical results

A didactic-size instance of the estimation problem was implemented using the general algebraic modelling system (GAMS). The equation system in step 2 and the optimisation problem in step 3 were solved using the non-linear programming solver CONOPT with GAMS.

Data was invented for the estimation in the following manner: A smaller version of CAPRI was constructed, using only seven production activities and the inputs plant protection and “rest” for clarity. The balanced base year dataset, to which the model must calibrate exactly, was arbitrarily chosen and is shown in table 2. One can see in the table data implies a set-aside rate of 12.5%, and that it was assumed that at the aggregate level, some voluntary set-aside also takes place. Assumed dual values were ρ = -10 EUR for the set-aside constraint and λ = 100 EUR for the land constraint. In the real application, the dual values would results from the simulation experiments and thus be different for each t.

Table 2: Invented base year solution

	CERE	OILS	POTA	FODD	OSET	VSET	FALL	W
Y	8	5	50	10	0	0	0	
P	100	80	50	50	0	0	0	
X	1	0.8	2	0.1	0	0	0	100
REST	1	2	3	0.5	0.1	0.1	0.1	100
S	300	300		300	300	300		
L	60	10	10	5	10	1	4	

Source: Invented base year data for fictive region used for testing estimation method.

The three marginal effects mentioned that are used in order to identify the parameters of the yield and input demand functions were assumed to be $\varepsilon(a_{ij},x_j) = 0.2$, $\varepsilon(x_j,v) = -2.0$ and $\varepsilon(y_j,l_j) = -0.1$.

The farm models that are to deliver simulation experiments, finally, were simulated using a different model in order to obtain a similar data-set with specification errors. That substitute model was formulated using a function of the form

$$l_j = \theta_j \prod_k p_k^{\mu_{jk}} \quad \forall j\,,$$

with the elasticity matrix μ_{jk} as in table 3, and the vector θ chosen to calibrate the model to the base year dataset. That model was used to simulate outcomes of the farm models by drawing 50 output price vectors p_t (with t the index for simulation number) from the uniform distribution ranging +/- 30% around the base year output price vector. Trials with different sample sizes indicated that the simulation results are fairly robust already at a lower number of draws.

Table 3: Elasticity matrix used in simulation experiments

	CERE	OILS	POTA	FODD
CERE	0.800	-0.100	-0.100	-0.050
OILS	-1.500	2.000	-0.100	-0.100
POTA	-1.500	-0.500	2.000	-0.100
FODD	-2.000	-1.000	-0.500	2.000
OSET	0.100	0.100	-0.900	-0.500
VSET	-2.000	-0.500	-0.400	-0.100
FALL	-1.750	-1.125	-0.275	0.025

Source: Invented data used for testing purposes.

Since neither this model nor the micro-economic model in CAPRI contains any motivation for farmers to keep fallow land or voluntary set-aside at non-zero dual values, and since the model used for simulation experiments does not deliver any observations of dual values, additional assumptions are required in order to identify the supply parameters of the activities that have no physical yield (voluntary and obligatory set-aside and fallow land). In the estimation, this was done by simply fixing the diagonal elements of B for those three activities to some feasible value. We used

$$b_{jj} = \frac{1}{0.5} \frac{GM_j - \lambda - \rho r I_{ARB}(j) + \rho I_{\{OSET\}}(j) + 500 I_{\{FALL\}}(j)}{l_j^*}$$

where GM is gross margin in the base year, λ dual value of land, ρ dual value of set-aside constraint, and I index functions. With more sophisticated simulation experiments this fixing of B-elements would not be necessary. Since it means that the supply *behaviour* of those activities are estimated only using cross effects (non-diagonal B), we do not expect any good fit for them.

The results are evaluated using two criteria. Firstly, the share of explained variance for activity levels is computed for each activity (index omitted) as

$$R^2 = \frac{SSR}{SST} \text{ with } SST = \sum_t \left(l_t^s - \bar{l}_t^s\right)^2 \text{ , } SSE = \sum_t \left(l_t^s - l_t\right)^2$$

and $SSR = SST - SSE$. The superscript s denotes simulation experiment data and the bar means sample mean. The measure R^2 computed this way shows how big a share of the variances of the activity levels in the simulation experiments can be explained using the calibrated CAPRI model. It is always less than unity and it is desirable that it is high, with unity meaning a perfect reproduction of the farm model results. The results are shown in table 4.

Table 4: R-squared measures for activity levels, sample size 50

	n	E	SSE	SST	SSR	R2
CERE	50	59.37	99.84	2989.60	2889.76	0.967
OILS	50	10.29	134.66	1014.53	879.87	0.867
POTA	50	10.75	415.93	1031.72	615.79	0.597
FODD	50	6.05	126.38	424.31	297.93	0.702
OSET	50	10.95	163.23	175.94	12.71	0.072
VSET	50	1.24	4.16	8.70	4.53	0.521
FALL	50	5.04	55.06	156.29	101.24	0.648

Source: Results of explorative estimation.

The activities for which there is a yield, i.e. the first four lines, have good or acceptable fits, which were fairly stable at different sample sizes. Also the last two activities have acceptable fits, though that should depend more on luck when considering the fixed diagonal elements of B for those activities, and they were less robust for smaller samples. The estimator only had the non-diagonal elements of B and the vector c to adjust, strongly constraining the range of possible behaviour. The bad fit of obligatory set-aside reflects the fact that the model used in the simulation experiments did not have a set-aside constraint and the elasticities were not chosen to properly simulate one. That problem should disappear with “real” farm models.

Another way of evaluating the same estimation results is by using the calibrated model in new simulation experiments to compute the point elasticities of activity levels to output prices in the base year, to compare with the elasticities used in the farm models in table 3. This was done by increasing each price once by 10% and solving the model. Table 5 shows the resulting matrix of elasticities.

Table 5: Elasticities of activity levels to output prices of estimated model.

	CERE	OILS	POTA	FODD
CERE	0.724	-0.142	-0.160	-0.097
OILS	-1.692	1.705	-0.340	-0.169
POTA	-0.317	-0.057	1.919	-0.068
FODD	-1.908	-0.280	-0.675	1.504
OSET	0.379	0.122	-0.186	-0.107
VSET	-3.764	-1.160	-0.214	0.414
FALL	-3.455	-1.652	-0.184	0.323

Source: Own calculations.

The own-price elasticities are generally lower those in table 3. That may depend on the lack of set-aside constraint in the farm-level model—the invented elasticity matrix did not properly represent this. Even so, it is the impression of the authors that the elasticities match surprisingly well, having in mind that the CAPRI supply model has a mathematical specification differing from that of the farm level models and only a quadratic function available for fitting the behaviour of the model.

6 Discussion

This first explorative application shows that the method is workable, though requiring some rather strong assumptions. With more complex farm level models for the simulation experiments, the assumptions regarding dual values of land and set-aside constraints can be relaxed. The need to fix the diagonal elements of the yield-less activities is a more serious problem, pointing at the lack of a microeconomic explanation of the co-existence of slack activities and dual values in CAPRI. More work is required in order to come up with a scientific procedure for identifying the behavioural parameters for those activities.

Further work will also be conducted in order to extend the method to work alternatively with a time series of data sets of the same type as the base year data, for all regions of the model. That means working with measurement errors in addition to the specification errors, but on the other hand with variation also of input prices and input use.

References

Britz, W. (ed) (2005). CAPRI Modelling System Documentation (online). http://www.agp.uni-bonn.de/agpo/rsrch/capri/capri-documentation.pdf

GAMS, see www.gams.com

Heckelei, T., and H. Wolff (2003). Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximim entropy. *European Review of Agricultural Economics*, 30: 27-50.

Howitt. R.E. (1995). Positive Mathematical Programming. *American Journal of Agricultural Economics*, 77(2): 329-42

SEAMLESS. See http://www.seamless-ip.org/