Estimating an *Ex Ante* Cost Function for Belgian Arable Crop Farms

Kristiana Hansen*† Alexandre Baudry†
Rembert De Blander† Bruno Henry de Frahan†
Philippe Polomé‡

May 1, 2009 Version

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2009 AAEA & ACCI Joint Annual Meeting, Milwaukee, Wisconsin, July 26–29, 2009

*Contact author: `kristiana.hansen@uclouvain.be`

†Université catholique de Louvain, Department of Agricultural Economics (ECRU), Place Croix du Sud 2/15, 1348 Louvain-la-Neuve, Belgium.

‡GATE-Groupe d'Analyse et de Théorie Economique, CNRS UMR 5824-Université Lyon 2, 93, chemin des Mouilles-B.P.167, 69131-ECULLY cedex France.

Estimating an *Ex Ante* Cost Function for Belgian Arable Crop Farms

Kristiana Hansen*† Alexandre Baudry†
Rembert De Blander† Bruno Henry de Frahan†
Philippe Polomé‡

May 1, 2009 Version

Abstract

We estimate a farm-level cost function for Belgian crop farms using FADN data over the study period 1996-2006. We rely on an estimation of farmers' expected yields at the time cropping decisions are made rather than actual yields observed in the FADN data. The use of an *ex ante* cost function improves the cost function estimation. We subsequently suggest how our cost function can be used in simulations to analyze farmer response to changes in output price risk.

1 Introduction

Farmers do not know yields at the time they make their input decisions. Instead, they have expected yields in mind when choosing input quantities, and in particular land allocations. However, expected yields are not observable to the econometrician. Pope and Just (1996) show that the standard practice of taking observed yields as a proxy for expected yields leads to inconsistent estimates of

*Contact author: kristiana.hansen@uclouvain.be

†Université catholique de Louvain, Department of Agricultural Economics (ECRU), Place Croix du Sud 2/15, 1348 Louvain-la-Neuve, Belgium.

‡GATE-Groupe d'Analyse et de Théorie Economique, CNRS UMR 5824-Université Lyon 2, 93, chemin des Mouilles-B.P.167, 69131-ECULLY cedex France.

the cost function parameters. They use duality theory to derive an estimator, which they show to be consistent and asymptotically efficient.

In this paper we propose an alternative estimate of expected yields. We estimate farmers' expected yields at the time cropping decisions are made based upon farmers' input decisions and use this measure of expected yields to estimate a cost function for arable crop farms in the Walloon region of Belgium. Preliminary analysis indicates that use of estimated yields in lieu of observed yields improves the goodness-of-fit of the cost function estimation. We will employ our cost function in simulations to analyze farmers response to changes in output price risk.

The next section of the paper outlines our modeling approach. Section 3 describes the data. Section 4 presents econometric estimation results. Section 5 discusses our plans for simulation. Section 6 concludes.

2 Model Specification

Pope and Just (1996) define an *ex ante* cost function for risk-neutral farmers in which the cost function depends on outputs defined in terms of expected rather than observed yields. They show that if observed outputs are not replaced with expected outputs, it is unlikely that instrumental variable techniques can estimate consistently the coefficients of the cost function. To circumvent this difficulty, Pope and Just (1996) define an *ex ante* cost function for the risk-neutral farmer in which the cost function depends on outputs defined in terms of expected rather than observed yields:

$$\min_x \{wx \mid x \in v(\bar{y})\} = C(\bar{y}, w), \tag{1}$$

where x is a vector of inputs, including land, and w is the corresponding vector of input prices. The *ex ante* cost function C is identical to a conventional cost function except that it is based on the intended output $\bar{y}$ instead of observed output y. The variable v represents the *ex ante* input requirement set.

The cost function is estimated by means of a system of equations defined by Shephard's lemma:

$$C = C(\bar{y}, w), \tag{2}$$

$$x_i = \frac{\partial C(\bar{y}, w)}{\partial w_i} \quad \forall i \tag{3}$$

where the index i refers to inputs.

2.1 Yield estimation

The general goal of the production function estimation is to compute the output levels that the farmer expects when he makes his input decisions. This should enable us to filter out the farmer's effects on the production from the random effect due to weather, pest, or any other uncontrollable cause. Thus, we aim to estimate a regression explaining production only with farmer-controlled variables and to use the residuals of that regression to test their distribution and the correlation between crops. That distribution will be applied below in the simulation.

We would like to estimate crop yield as a function of various crop inputs and farm-specific variables:

$$y_{mft} = \beta'_{mft} x_{mft} + \gamma'_{ft} z_{ft} + \alpha_f + \epsilon_{ft}, \tag{4}$$

The dependent variable of this equation is y_{mft}, yields per hectare for each of m crops on farm f at time t. The vectors x_{mft} and z_{ft} represent crop-specific and farm-specific inputs, respectively. The variable α_f represents unobserved, farm-specific variables on farm f such as farmer ability and soil quality that do not change over time. However, selection bias may be present, as farmers do not grow all crops in all years. It may be the case that a farmer does not grow a crop one year (any year) because he has additional information, not included in the analysis, regarding likely yields or prices. For example, the farmer may know that soil conditions on his farm are not conducive to a particular crop. Alternatively, he may have a low expectation of output price, leading him to switch to a more profitable crop. If the farmer's dichotomous decision whether to grow a crop is linked to his subsequent use of inputs when the crop is grown, the error term ϵft in Equation 4 will be correlated with regressors. Coefficient estimates in the yield equation will consequently be biased.

To address this issue of selection bias, we implement a two-step generalized Heckman procedure, where results from estimation of the dichotomous decision whether to grow crops are subsequently included in the yield estimation. The first step of the estimation captures the farmer's dichotomous decision whether to grow each of the m crops. These crop selection equations should be estimated jointly because crop choice occurs for all activities at the same time. Joint estimation will incorporate into the analysis interactions between the crops resulting from crop rotation considerations. Thus we estimate a multivariate probit system of equations, where each observation ft has a contribution to the likelihood function given by

$$P(d_{mft} \mid R_f, \bar{Z}_f, \bar{Z}_t) = \ln \left(\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi_M(\upsilon, \mu, \Sigma_\upsilon) d\upsilon \right). \tag{5}$$

The variable d_{mft} is a vector of m indicator variables to denote whether crop m is observed. Whether the crop is grown or not depends upon three types of variables. First, the variable R is a region indicator variable equal to 1 if the farm is located in the Limoneuse or Sablo-Limoneuse soil regions of Wallonia, where the high-quality soils are most conducive to arable crop farming. The region dummy is equal to 0 for farms located further south, in the Jurassique, Condroz, Famenne, or Herbagère Liège regions. Second, the vector $\bar{Z}_f$ consists of farm-level explanatory variables: farm size *tha*, a capital variable *capital*, the age *age* and education *edu* of the primary farmer, and a dummy variable *succ* equal to 1 if the farmer has a succession plan and equal to 0 if there is no succession plan or if no response was provided. Finally, the vector $\bar{Z}_t$ contains annual averages of regional crop prices, as farmers likely respond to price signals in deciding whether to grow a crop. The function $\varphi_M(\upsilon, \mu, \Sigma_\upsilon)$ is the multivariate normal probability density function with mean vector μ and covariance matrix Σ_υ.

We would like to include farm-specific dummy variables to capture unobservable, farm-specific characteristics. However, estimating such fixed effects in a probit model leads to inconsistent estimation of the parameters (Wooldridge, 1995). Instead, we include farm-specific and time-specific averages in the model to parse directly these panel effects. [1]

The second step of the generalized Heckman procedure is the following outcome

[1] To the best of our knowledge, this method was first proposed by Mundlak (1978) and Chamberlain (1982). Subsequent examples and further explanation can be found in Chamberlain (1984), Wooldridge (2002), and Verheyden (2008).

equations for each of the m crops:

$$y_{mft} = \delta R_f + \alpha_f' \bar{Z}_f + \gamma_{ft}' Z_{ft} + \beta_{imft}' X_{imft} + \xi_{mft}' Scor_{mft}' + \nu_{mft}. \tag{6}$$

Here, the dependent variable y_{mft} is the estimated yield of crop m on farm f at time t. The vectors R_f and $\bar{Z}_f$ are as described in the system of equations in step 1. The vector Z_{ft} indicates a vector of two farm-specific variables: capital, and farm size. The vector X_{mft} indicates a vector of i crop-specific input variables: seeds, pesticides, fertilizer, and hired services, as well as land allocation in absolute and percentage terms.

The vector $Scor_{mft}$ is scores generated from the step 1 estimation. The variable ν_{mft} is the error term. The final estimates generated by this method are econometrically superior to the original observed yields, as they are unbiased, efficient and consistent. In addition, they are more realistic, as they provide us with the farmer's expectation of yields at the time he makes his planting decisions rather than yields observed at the time of harvest.

2.2 Cost Function Estimation

The cost function that we estimate must conform to certain properties. In particular, the cost function must nondecreasing, concave and continuous in input prices, and nondecreasing in outputs. Further, costs should be strictly greater than zero when input prices and output quantities are greater than zero (Chambers, 1988). We employ the Symmetric Generalized McFadden (SGM) functional form in estimation because it is twice differentiable and places fewer restrictions *a priori* on the unknown cost function before estimation. Further, the global curvature properties can be imposed on the SGM without destroying its second-order flexibility (Diewert and Wales, 1987). The choice of functional form, is discussed in Lau (1986) and Brunke et al. (2009). Kumbhakar (1994), Wieck and Heckelei (2007), and Baudry et al. (2008) also use the SGM for estimation of a multi-output, multi-input farm system.

In many agricultural systems, climatic conditions such as drought can entail unforeseen expenses, causing costs incurred by farmers over a growing season to be stochastic. In such cases, a deterministic cost function specification such as the SGM would be inappropriate. However, in our application of arable crop agriculture in Northwestern Europe, farmers have little room to react to yield variability caused by weather or pests. The costs of growing crops depend only on variables under the control of the farmer; once input expenses have been committed, they cannot be significantly modified. The cost function thus depends only on the *expected* output; a non-stochastic specification such as the SGM is consequently appropriate.

The system of equations defined in equations 2 and 3 is assigned a functional form:

$$C_{ft} = \sum_m \phi_m \hat{Y}_{mft} \left(\frac{\sum_i \sum_j e_{ij} W_{irt} W_{jrt}}{2(\sum_i \theta_i W_{irt})} \right) + \sum_i c_i W_{irt} + \left(\sum_i b_{it} W_{irt} \right) t \sum_m \phi_m \hat{Y}_{mft}$$
$$+ \sum_i \sum_m c_{im} W_{irt} Y_{mft} + \sum_i \theta_i W_{irt} \left(\sum_m \sum_n g_{mn} \hat{Y}_{mft} \hat{Y}_{nft} \right) + \mu_{ft}, \tag{7}$$

where C_{ft} is the total production cost of farm f at time t, and W_{irt} is the regional Törnqvist price index on input i at time t,. The variable $\hat{Y}_{mft}$ is the expected farm-level production of output m, $\hat{y}_{mft} ha_{mft}$. The indices i and j represent inputs, and

m and n outputs. The parameters ϕ_m and θ_i are calculated indices on $\hat{Y}_{mft}$ and W_irt respectively (Diewert and Wales, 1987).

By Shephard's lemma, we obtain the derived input demands that complete the system of equations to be estimated:

$$\begin{aligned}\frac{\partial C_{ft}}{\partial W_{irt}} = X_{ift} = \sum_m \phi_m \hat{Y}_{mft} \left(\frac{\sum_i e_{ij} W_{irt}}{\sum_i \theta_i W_{irt}} - \frac{\theta_i}{2} \left(\frac{\sum_i \sum_j e_{ij} W_{irt} W_{jrt}}{(\sum_i \theta_i W_{irt})^2} \right) \right) \\ + c_i + tb_{it} \sum_m \phi_m \hat{Y}_{mft} + \sum_m c_{im} \hat{Y}_{mft} + \theta_i \left(\sum_m \sum_n g_{mn} Y_{mft} Y_{nft} \right) + \varepsilon_{ift}\end{aligned} \tag{8}$$

The variables $c_i, b_{it}, c_{im}, e_{ij}$, and g_{mn} in Equations 7 and 8 are the parameters to be estimated. The parameter c_i is a farm-specific fixed effect. The parameters e_{ij} are elements of the input price matrix E. The parameters g_{mn} are elements of the output matrix G. We impose symmetry conditions in estimation so that $e_{ij} = e_{ji}$ for all i, j and $g_{mn} = g_{nm}$ for all m, n. Further, the adding up constraints are imposed so that $\sum_j e_{ij} = 0$ for all i. In order ensure that the cost function is well-behaved, we impose global concavity on input prices (the matrix E must be negative semi-definite) and global convexity on outputs (the matrix G must be positive semi-definite).

We estimate this system of equations using the generalized method of moments. Imposing concavity by Cholesky decomposition creates difficulties with convergence. We thus implement the Cholesky-Lau decomposition, the method proposed by Diewert and Wales (1988) and applied by Moschini (1998) and Moro, Nardella, and Sckokai (2005), to restrict the rank of the Cholesky matrix to the point where estimation converges.

Table 1: Farms by Region

	Condroz	Famenne	Herbagère	Jurassique	Limoneuse	Sablo-Limoneuse	Walloon Totals
# of farms	13	1	2	1	48	8	73
Other cereals	69	10	5	5	177	47	226
Chicory			2		153	36	278
Potatoes	21				123	34	178
Sugar beets	80	11	8		345	78	522
Winter wheat	81	10	8		356	77	532
	251	31	21	5	977	225	1736

3 Data

We consider a panel of 73 Walloon crop farms from the European Farm Accountancy Data Network (FADN) database observed during a period of 11 years, from 1996 to 2006. Farms are included in the crop farm sample if they receive most of their gross standard margin from field crop activities. Table 1 indicates that the farms are primarily located in the Limoneuse and Sablo-Limoneuse regions of Wallonia and that not all farms grow all crops every year. The unbalanced nature of the panel data is one reason for the modeling specification described below.

The primary outputs of Walloon crop farms are listed in Table 2, along with information on yields and land allocations on FADN farms. We aggregate cereals other than winter wheat into a single crop category in order to reduce the number of outputs to a manageable number. The aggregation is done using a Törnqvist index on per-hectare revenues, or yield-in-revenue. Processing plants pay farmers for sugar beets based on sugar content. Yields for sugar beets are consequently estimated after first adjusting for sugar content. Energy crops, flax, maize, and rapeseed have been excluded from the analysis due to the insufficient number of observations. Dried peas, green beans, green peas, and flax are generally under contract with a buyer, in which case the crop area is known but the yield and

Table 2: Crop Model Outputs

Activity	By crop Mean ha per farm*	Mean % of farm	Mean yields	By category Mean ha per farm*	Mean % of farm	Mean yields	# of farm-year obsvns
Chicory	8.10	10%	43.35	8.10	10%	43.35	226
Spring barley	9.06	9%	5.73	11.35	14%	7.66	278
Spring oats	4.34	6%	5.67				
Spelt	7.61	8%	7.52				
Winter barley	9.82	12%	7.89				
Other potatoes	10.63	14%	42.41	10.63	14%	42.41	178
Sugarbeets	13.92	18%	71.64	13.92	18%	71.64	522
Winter wheat	27.21	34%	8.50	27.21	34%	8.50	532
# of obsvns							1736

Notes: "Mean Hectares" and "Mean % of Farm" do not include farm-year observations in which the crop is not grown.

the value are not. Potatoes are often under contract with a buyer. Contract observations have been removed from the data.

Inputs, summarized in Table 3, fall into three categories. First are crop-specific inputs for which the FADN reports expenditures per crop. These inputs are chemical fertilizers, pesticides, seeds, and hired services. Regional price indices exist for the first three. Hired services are highly crop-specific, e.g. transport services for sugar beet or mowing services for cereals. Thus no price index exists. Hired services may also be a substitute for capital.

Second are farm-specific inputs, for which expenditures are reported in the FADN only at the farm level. The three unallocated inputs included in the analysis, building, machines, and energy, are complementary and can be lumped together in a composite input called capital. Three unallocated inputs reported in the FADN data are not used in the analysis. Family labor is observed with errors (among others: no accounting of hours, family members used as straw persons for tax reasons) and is therefore dropped. Organic fertilizer is measured with error

Table 3: Crop Model Inputs

Variable	Contains	Quantity	Price
Crop-specific input category	Fertilizer Pesticides Seeds Third party services	Total costs divided by prices (FADN)	Regional Price Indices
Farm-specific input category	Building Electricity Machinery	na	Farm expenditures (FADN)
Land	Land	FADN	Land lease rate (FADN)

and therefore also dropped . Hired labor and salaried labor are rather negligible expenses compared to other items and are also dropped. Generally speaking, farm-specific input quantities are poorly measured in the FADN data; several farms often report identical the same figures, indicating that values are assigned rather than actually measured.

The third input is land. We use lease prices to represent land. Average land price per district is available from the Belgian National Institute of Statistics (NIS) through 2005. However, as land sale prices are not available for 2006, we use farm-specific lease data from the FADN survey. This introduces a downward bias in land prices, as official lease prices are generally lower than prices actually paid.

4 Yield and Cost Function Estimation Results

Results of this yield estimation procedure are found in Table 4. The variables $psau_m^2$, pes_m^2, sem_m^2, eng_m^2, and hir_m^2 have been removed from several equations due to multicollinearity. Many of the explanatory variables included in the yield estimation are significantly different than zero. In particular, all of the crop-specific variables that are significantly different than zero indicate that increased input use leads to increased yields at a decreasing rate, with the sole exception of hired labor hir_m in the other cereals equation. Further, the score variables are largely significant, indicating that the binary decision whether to grow a crop in the first place is in fact linked to the subsequent decision of input use levels, and should be incorporated into the analysis.[2]

We measure the performance of this two-step generalized Heckman model relative to two more simplistic models that do not address the selection bias issue. Column 1 of Table 5 presents information on the number of significant coefficients and R^2 values where each of the five output yield equations is estimated independently. Column 2 presents the same information for a system where the panel nature of the data is taken into account through the addition of farm-level fixed effects and year dummies, though the equations are once again estimated independently of one another. The third column presents the results of the two-step method propsoed in Equations 5 and 6. The two-step model provides estimation results with a better goodness-of-fit than the other two models, both with respect to the percentage of significant coefficients and the R^2 values of the equations. Further, a likelihood ratio test comparing the two-step model both to a constant-only model and a model where the multivariate probit model from Step 2 is estimated without including the scores from Step 1 both indicate that the two-step proce-

[2]Farm-year observations where the farm grows nothing (i.e., farm joins the sample late or leaves the sample early) are not included in the analysis because when these observations are included, multivariate probit estimation of all five crops does not converge. This is a reasonable simplification, as the selection problem we believe to be an issue is crop rotation rather than bias resulting from farms leaving the sample.

dure improves the explanatory value of the model. It is also worth noting that the coefficients on crop-specific inputs in the models from Columns 1 and 2 of Table 5 (not reported in this paper) had more coefficients with counter-intuitive signs than is the case in Table 4 further evidence that cross-crop interactions are important to address in a model of this type.

We further explore whether *ex ante* yields improve our understanding of the farming system we model by estimating four different specifications of the cost function. The specifications vary by whether the farm-level production variable included in the cost function estimation is based upon observed yields y_{mft} or estimated yields $\hat{y}_{mft}$. They also vary by the degree of input aggregation. The farm-level input specifications aggregate all crop-level input expenditures into a single crop-level input category, so that there are only three inputs to be estimated.This represents the degree of detail on input expenditures that is available for farm-level analysis using the FADN data in most European countries. However, the Belgian FADN does report inputs at the crop level. Thus we also run several specifications using all the information on input expenditures available in the FADN data for Belgium with crop-level inputs broken out into seven input categories. The latter level of input specificity may provide a better estimation. However, the former may be more tractable.

A brief comparison of the estimation results for the four specifications is located in Table 6. The models with estimated yields $\hat{y}_{mft}$ perform better than those without, measured in terms of the total sum of squared errors and the root mean squared error in each system of equations and the average R^2 and adjusted R^2 values of the equations. Table 6 also reports the percentage of significant coefficients in each system. The results regarding significant coefficients are mixed.

Table 4: Yield Estimation Results

	Chicory	Other cereals	Potatoes	Sugar beets	Winter wheat
ha_m	0.107	-1.660	0.050	-1.578***	-0.024***
	(0.472)	(0.320)	(0.480)	(0.000)	(0.003)
$psau_m$	-8.785	128.574	178.077***	19.795	2.763***
	(0.582)	(0.688)	(0.000)	(0.143)	(0.000)
$psau_m^2$		-882.965	-440.581***		
		(0.249)	(0.000)		
eng_m	0.042***	0.386***	0.023***	-0.000	0.008***
	(0.001)	(0.001)	(0.000)	(0.627)	(0.000)
eng_m^2	-0.000***				
	(0.010)				
pes_m	0.009**	0.382***	0.017***	-0.000	0.007***
	(0.039)	(0.002)	(0.000)	(0.877)	(0.000)
sem_m	0.189***	0.153	0.003*	0.003***	0.007***
	(0.000)	(0.495)	(0.073)	(0.000)	(0.000)
hir_m	0.005	-0.620***	0.003	0.001**	0.001
	(0.128)	(0.000)	(0.560)	(0.010)	(0.391)
hir_m^2		0.000	0.000	-0.000***	-0.000**
		(0.818)	(0.243)	(0.009)	(0.028)
tha	-0.003	-0.500	-0.003	0.161***	0.012***
	(0.830)	(0.105)	(0.760)	(0.000)	(0.000)
tha_i	0.053***	2.378***	-0.010	0.098***	0.005*
	(0.000)	(0.000)	(0.440)	(0.009)	(0.087)
$capital_i$	0.008	-1.055***	0.058***	-0.185***	-0.007***
	(0.488)	(0.000)	(0.000)	(0.000)	(0.004)
age_i	-0.052	10.427***	0.079***	0.438***	-0.001
	(0.140)	(0.000)	(0.007)	(0.000)	(0.890)
$succ_i$	4.416***	-28.085	0.097	5.204**	0.018
	(0.000)	(0.246)	(0.911)	(0.015)	(0.927)
edu_i	-0.617**	25.003***	0.944***	-1.940***	-0.003
	(0.011)	(0.000)	(0.000)	(0.000)	(0.937)
$Scor_{CER}$	-1.194**	561.272***	-1.069***	-3.837***	-0.006
	(0.012)	(0.000)	(0.004)	(0.000)	(0.928)
$Scor_{CHI}$	5.708***	-185.985***	0.786**	0.659	0.084
	(0.000)	(0.000)	(0.013)	(0.367)	(0.214)
$Scor_{POT}$	-0.007	-118.260***	6.062***	1.612**	-0.091
	(0.986)	(0.000)	(0.000)	(0.035)	(0.190)
$Scor_{SUG}$	2.274***	-58.900***	2.097***	31.705***	0.078
	(0.005)	(0.001)	(0.001)	(0.000)	(0.590)
$Scor_{WHE}$	0.793	104.272***	-1.404**	-3.198**	1.904***
	(0.289)	(0.000)	(0.048)	(0.036)	(0.000)
$WALOSL$	3.536***	-371.256***	2.906***	2.319	0.168
	(0.001)	(0.000)	(0.000)	(0.138)	(0.228)
$Constant$	-0.682	154.031***	-6.640***	29.775***	3.506***
	(0.749)	(0.001)	(0.000)	(0.000)	(0.000)
Observations	548	548	548	548	548
R^2	0.934	0.934	0.947	0.606	0.643

Notes: p values in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 5: Comparison of Yield Estimation Results

	Ordinary lease squares	Fixed effects with year indicator variables	Multivariate probit for sample selection bias
% of significant coefficients	60%	30%	62%
# of significant coefficients	5.400	7.400	13.800
R^2 values by equation			
Chicory	0.179	0.362	0.934
Other cereals	0.224	0.284	0.934
Potatoes	0.299	0.184	0.947
Sugarbeet	0.189	0.509	0.606
Winter wheat	0.259	0.359	0.643
Average R^2	0.230	0.340	0.813
Likelihood ratio tests			
With constant-only model			5,552.960
With no-score model			973.730

Notes: Coefficients are considered significant if they are significant at the 5% level.

However, on balance, it appears that the estimation using *ex ante* yields and crop-level inputs is preferred. We consequently use the *ex ante* cost function, estimated with expected yields and crop-specific inputs, in the policy simulations below.

As the cost function is at the farm-level, farm-specific variable input price elasticities can be calculated. Table 7 reports sample means and standard deviations. All own-price elasticities are negative, as concavity of the input matrix E has been imposed. The own-price elasticity for $Othercereals$ variable inputs is elastic; all the other own-price elasticities are inelastic. Most elasticities indicate substitutability between inputs. The effects are all small, however, and often not significantly different than zero.

Table 6: Comparison of Cost Estimation Results

Equation	SSE	RMSE	R^2	Adjusted R^2	% of significant coefficients, all equations
Observed yields, farm-level inputs					
Cost	94,747.23	14.60	-3.31	-3.44	
Crop-specific inputs	41,265.23	9.57	0.57	0.57	
Farm-specific inputs	3,663.19	2.85	-0.58	-0.61	
Land	409.87	0.95	0.27	0.26	
Total/Average	140,085.52	27.98	-0.76	-0.80	58%
Observed yields, crop-level inputs					
Cost	16,079.55	6.13	0.20	0.16	
Wheat inputs	2,849.23	2.53	0.13	0.12	
Other cereal inputs	353.69	0.89	0.66	0.65	
Potato inputs	3,801.37	2.92	0.39	0.38	
Chicory inputs	641.35	1.20	0.72	0.72	
Sugar beet inputs	1,397.95	1.77	0.42	0.41	
Farm-specific inputs	2,195.02	2.22	0.04	0.03	
Land	193.71	0.66	0.65	0.65	
Total/Average	27,511.87	18.32	0.40	0.39	37%
Estimated yields, farm-level inputs					
Cost	53,070.88	10.93	-1.42	-1.48	
Crop-specific inputs	63,807.96	11.90	0.34	0.33	
Farm-specific inputs	2,062.00	2.14	0.11	0.10	
Land	220.09	0.70	0.61	0.60	
Total/Average	119,160.92	25.67	-0.09	-0.11	69%
Estimated yields, crop-level inputs					
Cost	13,660.86	5.65	0.32	0.28	
Wheat inputs	1,895.96	2.06	0.42	0.42	
Other cereal inputs	599.70	1.16	0.42	0.41	
Potato inputs	3,548.75	2.82	0.43	0.42	
Chicory inputs	513.65	1.07	0.78	0.77	
Sugar beet inputs	2,319.85	2.28	0.04	0.03	
Farm-specific inputs	2,675.16	2.45	-0.17	-0.18	
Land	164.43	0.61	0.71	0.70	
Total/Average	25,378.37	18.11	0.37	0.36	56%

Note: Totals of sum of squared errors and root mean squared errors are provided for each system of equations. R^2 and Adjusted R^2 values are equation averages.

Table 7: Price elasticities of variable input demands for the Specification with *Ex Ante* Yields and Crop-Specific Inputs

Value (Std Dev.)	Wheat inputs	Other cereal inputs	Potato inputs	Chicory inputs	Sugar beet inputs	Farm-level inputs	Land
Wheat inputs	-0.029 (0.042)	0.139 (0.304)	0.015 (0.016)	0.000 (0.001)	-0.150 (0.371)	0.022 (0.052)	0.004 (0.007)
Other cereal inputs	0.561 (0.208)	-3.693 (2.216)	-0.403 (0.101)	0.095 (0.049)	3.020 (1.854)	0.475 (0.553)	-0.055 (1.052)
Potato inputs	0.008 (0.023)	-0.055 (0.240)	-0.007 (0.012)	0.002 (0.008)	0.043 (0.192)	0.010 (0.045)	-0.002 (0.007)
Chicory inputs	-0.001 (0.001)	0.035 (0.082)	0.006 (0.003)	-0.005 (0.007)	-0.007 (0.006)	-0.030 (0.059)	0.002 (0.004)
Sugar beet inputs	-0.160 (0.221)	0.798 (1.606)	0.082 (0.082)	-0.004 (0.010)	-0.840 (1.955)	0.101 (0.210)	-0.011 (0.042)
Farm-level inputs	0.029 (0.034)	0.157 (1.184)	0.026 (0.017)	-0.028 (0.042)	0.127 (0.263)	-0.312 (0.593)	0.001 (0.563)
Land	0.000 (0.006)	0.000 (0.564)	0.000 (0.003)	0.000 (0.003)	0.000 (0.041)	0.000 (0.130)	-0.048 (0.487)

Note: Sample means of farm-specific elasticities are presented, with standard deviations in parentheses.

5 Profit Function Simulation

Before implementing policy simulations, it is first necessary to verify that the model replicates cropping patterns observed in the base year. This validation of the model is done at the farm-level. The farmer chooses expected production levels$\hat{Q}_{mfs}$ so as to maximize profits,

$$\max_{\hat{Q}_{mft}}\{EU[\sum_m P_{mft}\hat{Q}_{mft} + S_{mft} - TC_{ft}(\hat{Q}_{mft}, w_i) - \epsilon_{ft}]\} \tag{9}$$

subject to the following constraints:

$$\hat{Q}_{mfs} = \hat{y}_{mft} l_{mfs} \quad \text{Link between land and production through } \textit{ex ante} \text{ yields}$$

$$\hat{Q}_{mft} = \hat{Q}_{mfs}\left[\rho_{mf}\right] \quad \text{for all crops } m \text{ except sugar beet under quota}$$

$$\hat{Q}_{SBA,ft} = \hat{Q}_{SBA,fs}. \quad \text{In-quota sugar beet constraint}$$

The first of these constraints provides the link between production, $\hat{Q}_{mfs}$, and land allocated to each crop in the simulation year, ℓ_{mfs}, based upon the *ex ante* yields $\hat{y}_{mft}$, for chicory, potatoes, sugar beets, and wheat estimated above, and *ex ante* yield-in-revenue for the aggregated crop category other cereals, $\hat{y}_{CER,fs}$. The price for other cereals is consequently $P_{CERmft} = 1$. The index s is a time index indicating a variable whose level is determined by the model in simulation.

The second constraint generates a shadow value ρ_{mft} on production which will be used in the simulation below to calibrate the model. The calibration term is a penalty function which represents hidden costs not taken directly into account in the econometric estimation of the cost function (Howitt, 1995; Heckelei and Wolff, 2003; Henry de Frahan et al., 2007). Without this term, the model produces more of each output than is observed in the base year. Sugar beet production in Belgium is under a quota system. The third constraint limits sugar beet production on each farm to that observed in the base year, $\hat{Q}_{SBA,ft}$.[3]

$$\max_{\hat{Q}_{mfs}}\{EU[\sum_{f\in r}\sum_{m} P_{mft}\hat{Q}_{mfs} + S_{mft} - C_{ft}(\hat{Q}_{mfs}, w_i) - \epsilon_{ft}]\} - \sum_{m}\rho_{mf}\hat{Q}_{mfs} \quad (10)$$

[3]In-quota sugar beets $\hat{y}_{SBA,ft}$ and out-of-quota sugar beets $\hat{y}_{SBC,ft}$ are treated as separate outputs in the profit simulation, as the price that farmers receive for in-quota sugar beets is significantly higher than the world price, at which out-of-quota sugar beets are sold.

subject to the following constraints:

$$\begin{aligned}
\hat{Q}_{mfs} =& \hat{y}_{mft}\ell_{mfs} && \text{Link between land and production through } \textit{ex ante} \text{ yields} \\
\hat{Q}_{SBA,fs} \leq& Q_{SBA,ft} && \text{In-quota sugar beet constraint} \\
\sum_{f\in r}\ell_{mft} \leq& \sum_{f\in r}\sum_{m}\ell_{mfs} && \text{Regional land constraint}
\end{aligned}$$

The final term of the profit function is the linear penalty function constructed using the shadow values on the production constraint from Equation 9. The third constraint is a regional land market, specifying that all the land allocated to crops in an agricultural region cannot exceed the total land available in the region in the base year, $\sum_{f\in r}\sum_{m}\ell_{mft}$. Farms are, however, able to re-allocate land among themselves, through regional land markets.

We plan to simulate crop insurance for wheat in order to test the performance of the *ex ante* yields in simulation. Prices and yields in the FADN data are farm-specific; the crop insurance premia and trigger will consequently also be farm-specific. Farms will be assumed to exhibit constant relative risk aversion. Under the current French insurance system, a farmer contracts for insurance at the farm-average price. If yields fall below a pre-specified level, the farmer receives an insurance payment. This will be the crop insurance instrument we will model initially. We might expect to see that simulations performed with *ex ante* yields predict greater value to crop insurance than simulations performed with observed yields. This is because using observed yields may give the impression to the econometrician that farmers predict their production levels more accurately than they really do.

6 Conclusion/Discussion of Results

As demonstrated by Pope and Just (1996), the *ex ante* cost function is preferred on theoretical grounds to the cost function based on observed yields. Here, we demonstrate a relatively straightforward variation of their method in the setting of Belgian crop farms. Further, we have begun to construct an analysis of farmer response to increased price risk resulting from trade liberalization. Although the construction of this particular policy simulation is still at an intermediate level of development, the potential for using the model to perform policy simulations has been proven. This model has also already been used to test farmer response to changes in voluntary and mandatory agri-environmental policies implemented under the Common Agricultural Policy (Baudry et al., 2009).

We believe that the *ex ante* cost function is theoretically superior to a cost function based on observed yields. Next, we intend to compare the results of simulation based on the *ex ante* cost function to simulations based on the observed yield cost function, to see whether there are also improvements at the simulation level in terms of model performance.

References

Baudry, A., R. de Blander, C. Delaunoy, K. Hansen, B. Henry de Frahan, A. le Roi, S. Paque, P. Polomé, G. van Impe, and T. Walot. 2009. "Système Intégré d'Aide à la Décision Publique Agricole et Environnementale en Région wallonne ADAGE-RW." Unpublished, Université catholique de Louvain, Groupe Interuniversitaire de Recherches en Ecologie Appliquée (GIREA-ECOL) and Unité d'économie rurale (ECRU).

Baudry, A., B.H. de Frahan, P. Polomé, and R. Howitt. 2008. "Dairy Farms without Quotas: Simulations Based on a Multi-output Multi-input Cost Function." Unpublished, ECRU Working Paper 2008.

Brunke, H., B. Henry de Frahan, Y. Surry, and D. Desbois. 2009. "A Literature Review on Cost of Production Studies in Agriculture." Unpublished, Document prepared for Farm Accountancy Cost Estimation and Policy Analysis of European Agriculture, FACEPA Deliverable No. 1.2-May 2009.

Chamberlain, G. 1982. "Multivariate Regression Models for Panel Data." *Journal of Econometrics* 18:5–46.

—. 1984. "Panel Data." In Z. Griliches and M. D. Intriligator, eds. *Handbook of Econometrics*. Amsterdam, Holland: North Holland, pp. 1247–1318.

Chambers, R.G. 1988. *Applied Production Analysis: A Dual Approach*. Cambridge University Press, Cambridge.

Diewert, W.E., and T.J. Wales. 1987. "Flexible Functional Forms and Global Curvature Conditions." *Econometrica* 55:43–68.

Heckelei, T., and H. Wolff. 2003. "Estimation of constrained optimisation models for agricultural supply analysis based on generalised Maximum Entropy." *European Review of Agricultural Economics* 30:27–50, 1.

Henry de Frahan, B., J. Buysse, P. Polome, B. Fernagut, O. Harmignie, L. Lauwers, G.V. Huylenbroeck, and J.V. Meensel. 2007. "Positive Mathematical Programming for Agricultural and Environmental Policy Analysis: Review and Practice." In A. Weintraub, C. Romero, T. Bjrndal, R. Epstein, and J. Mirand, eds. *Handbook of Operations Research in Natural Resource Economics*. Boston, MA: Kluwer Publishing.

Howitt, R.E. 1995. "Positive mathematical programming." *American Journal of Agricultural Economics* 77:329–342.

Kumbhakar, S.C. 1994. "A Multiproduct Symmetric Generalized McFadden Cost Fucntion." *Journal of Productivity Analysis* 5:349–357.

Lau, L.J. 1986. "Functional Forms in Econometric Model Building." In Z. Griliches and M. D. Intriligator, eds. *Handbook of Econometrics*. Amsterdam, Holland: Elsevier Science Publishers, pp. 1515–1566.

Mundlak, Y. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46:69–85.

Pope, R.D., and R.E. Just. 1996. "Empirical Implementation of Ex Ante Cost Functions." *Journal of Econometrics* 72:231–249.

Verheyden, B. 2008. "Essays on the Economics and Econometrics of Education." PhD dissertation, Facults universitaires Notre Dame de la Paix Namur.

Wieck, C., and T. Heckelei. 2007. "Determinants, Differentiation, and Development of Short-term Marginal Costs in Dairy Production: an Empirical Analysis for Selected Regions of the EU." *Agricultural Economics* 36:201–218.

Wooldridge, J. 1995. "Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions." *Journal of Econometrics* 68:115–132.

Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge MA.