Variable Input Allocation: Why Heterogeneity Matters?

Elodie Letort[1], Alain Carpentier[2]

[1] INRA – UMR SMART, Rennes, France
[2] INRA – UMR SMART, Rennes, France

Paper prepared for presentation at the 120th EAAE Seminar "External Cost of Farming Activities: Economic Evaluation, Environmental Repercussions and Regulatory Framework", Chania, Crete, Greece, date as in: September 2 - 4, 2010

Variable Input Allocation: Why Heterogeneity Matters?

Elodie Letort[1] and Alain Carpentier[2]

August 2010

[1]INRA - UMR SMART, Rennes, France
[2]INRA - UMR SMART, Rennes, France.

The allocation of variable inputs among crops is a common problem in applied studies that use farm accountancy data. Standard farm accounting information is typically restricted to aggregate or whole-farm input expenditures; there are usually no details on how these expenditures are split among crops. Most studies employing multi-crop econometric models with land as an allocable fixed input consider generally variable input uses at the farm level (Moore and Negri, 1992). However, the allocation of variable inputs among crops appears to be useful for several objectives, such as to analyze the evolution of gross margins at the crop level, to investigate the empirical validity of a multi-crop econometric model and to provide important information for extension agents or farmer advisors.

A large number of authors have studied on this topic either to provide solutions for allocating input costs between crops and/or activities (Just et al., 1983; Chambers and Just, 1989) or to compute input-output coefficients (Dixon and al., 1984; Hornbaker, Dixon and Sonka, 1989; Peeters and Surry, 1993). Some authors have treated this issue as a necessary step of their analysis (for example, in the evaluation of agro-environmental policies on input use in Lence and Miller, 1998). The most widely used methods to allocate variable input uses to crops are based on regression models or production function models that include constraints on total use variable input (Dixon and al., 1984; Hornbaker and al., 1989; Just and al., 1990). However, the allocation of variable inputs among crops depends on how farmers allocate land among crops, which itself takes into account input uses by crop. Crop input decisions and acreage choices are thus partially simultaneous. Variable input allocation requires the specification of a complete production model, i.e., a description of the land allocation, use of variable inputs and crop yields, to account for the link between acreage and input use choices.

The contribution of this article is threefold. First, we show that the standard regression based approaches for allocating variable input uses to crops are likely to be biased due to the partial simultaneity of the (expected) crop variable input and acreage choices. Second, it proposes a structural econometric multi-crop model for determining the origin of these biases. The structure of the model relies on the timing of farmers' choices. The specified model distinguishes two sorts of error terms, namely: an error term accounting for the heterogeneity of farms and an error term accounting for stochastic events affecting crop production. It provides explicit functional forms for the links between the error terms of the yield supply, input demand allocation and acreage equations. Third, we propose a method based on control functions to eliminate bias associated with the standard regression-based methods. It builds on previous result obtained for the estimation of the so-called correlated random coefficient models (Imbens and Wooldridge, 2007; Wooldridge, 2008) and average treatment effects (Heckman and al., 2003). The empirical implementation of the proposed methods is described in three stages, and an application is presented based on French farm-level data.

This article proceeds as follows. The next section presents a review of the literature on input allocation methods and briefly discusses the endogeneity problems in these standard approaches and the solution adopted in this article, *i.e.*, the control function-based approach. This solution requires an econometric multi-crop (that is, for acreage, yield and input choices) model, which is described in the second section. The third section presents the control function-based approach used to take into account the links between the acreage and the input use choices in the variable input allocation equation. In the fourth section, a general three-stage procedure for implementing the approach and an application based on French farm-level data are presented. The last section of this article provides some concluding remarks.

Literature review

The most common farm data on crop production consist of acreages, yields and prices at the crop level and variable input uses and quasi-fixed factor quantities (that is, measures of labour and capital) at the farm level. Input price indices are generally made available by departments of agriculture at the regional level. Farmer i ($i = 1, ..., N$) produces C crops ($c = 1, ..., C$) to which he allocates his S units of land. In what follows, we suppose one single variable input. X_i denotes the quantity of variable input use at the farm level for farm i. w_i is the input price for farm i. x_{ci} denotes the quantity of variable input uses for crop c per unit of land for farm i. s_{ci} is the acreage share of crop c for farm i. y_{ci} denotes the yield of crop c, and p_{ci} denotes its price for farm i. The input allocation problem consists in recovering input quantities x_{ci} for $c = 1, ..., C$.

Several approaches have been used or proposed to solve this allocation problem. We distinguish two main groups in the literature. The first group includes approaches that solely consider input allocation equation(s) as the one defined below. In these models, input allocations are treated as parameters to be estimated, to use the terminology of Just *et al.* (1990). These are, by far, the most widely used approaches in practice. In the second group, input allocation equations are part of a system of equations that include crop supply and acreage functions or production functions (Chambers and Just, 1989). In what follows, we describe the first group type of approaches, along with their advantages and limitations. These limitations suggest the advantages of using the second type of approach.

Approaches based on single-input allocation equations

Among the available methods for allocating inputs to activities or crops, the most widely used is the regression method that considers variable input allocation x_{ci} as parameters:

$$x_i = \sum_{c=1}^{C} s_{ci} x_{ci} + \eta_i \text{ with } E\left[\eta_i | \mathbf{s}_i\right] = 0 \tag{1}$$

or as parametric functions:

$$x_i = \sum_{c=1}^{C} s_{ci} x_{ci}\left(\mathbf{z}_i; \mathbf{a}\right) + \eta_i \text{ with } E\left[\eta_i | \mathbf{s}_i, \mathbf{z}_i\right] = 0 \tag{2}$$

where $\mathbf{z}_i$ is the vector of exogenous variables such as farm's characteristics and activities, $\mathbf{a}$ the vector of corresponding unknown parameters and $\mathbf{s}_i$ is the vector of acreage shares. Ordinary Least Square (OLS) for a single input model or seemingly unrelated regression (SUR) for a system of input allocation equations provide consistent estimators of x_{ci} and $\mathbf{a}$ under the assumption that the conditional expectation of η_i is zero. [1]

Later, these models have been generalized by adding random terms to the crop input use models to account for the effects of unobserved determinants of input choices. Models (1) and (2) are then respectively written:

$$x_i = \sum_{c=1}^{C} s_{ci} \left[x_{ci} + u_{ci}^x \right] + \eta_i \text{ with } E\left[\eta_i | \mathbf{s}_i\right] = E\left[u_{ci}^x | \mathbf{s}_i\right] = 0 \tag{3}$$

$$x_i = \sum_{c=1}^{C} s_{ci} \left[x_{ci}\left(\mathbf{z}_i; \mathbf{a}\right) + u_{ci}^x \right] + \eta_i \text{ with } E\left[\eta_i | s_i, \mathbf{z}_i\right] = E\left[u_{ci}^x | \mathbf{s}_i, \mathbf{z}_i\right] = 0 \tag{4}$$

where η_i terms include measurement errors or stock variations and the u_{ci}^x terms are defined as the difference between the "true" values of the unobserved input uses and the values that can be "explained" by the variables. Models (3) and (4) are input allocation equations with random parameters. In these models, the error terms, $\sum_{c=1}^{C} s_{ci} u_{ci}^x + \eta_i$ are heteroskedastic, and feasible generalized OLS or SUR estimations will provide efficient estimators of the parameter vector $\mathbf{a}$ under the assumption that the error terms u_{ci}^x and η_i have constant variances and covariances (Dixon, Batte and Sonka, 1984; Hornbaker, Dixon and Sonka, 1989; Dixon and Hornbaker 1992).[2]

The approaches just described are easy to implement and can provide satisfactory results (Just, Zilberman, Hochman and Bar-Shira, 1990). However, the consistency of the regression estimators of $\mathbf{a}$ in the generalized input allocation equation system relies on the assumption that acreage shares $\mathbf{s}_i$ are exogenous with respect to u_{ci}^x, i.e.:

$$E\left[u_{ci}^x | \mathbf{s}_i, \mathbf{z}_i\right] = 0 \tag{5}$$

These conditional mean conditions are unlikely to hold with farm data, for the simple reason that input use x_{ci} partly determines profitability of crop c, which itself is a determinant of crop c acreage. Since x_{ci} are determinants of the acreage choices, any part of x_{ci} is a determinant of the choice of s_{ci}. As a result, the condition:

$$E\left[u_{ci}^{x}|\mathbf{s}_i\right] = 0 \quad (6)$$

holds if and only if $u_{ci}^{x} = 0$, i.e. in the unrealistic case where $\mathbf{z}_i$ are "perfect" control variables for the heterogeneity of x_{ci}. Of course the biases due the endogeneity of $\mathbf{s}_i$ are reduced by the use of "imperfect" control variables. These biases are also likely to be limited if the elements of the x_{ci} vectors represent small amounts when compared to the crop returns.

These approaches based on single input allocation equations suffer from the same limits. Hence, the specification of a complete production model (describing land allocation, use of variable inputs and crop yields) is necessary in order to account for the link between the input uses and acreages choices.

Approaches based on multicrop econometric models

We now discuss models in which input allocation equations are estimated jointly with other equations, such as production technology or models describing acreage choices. Multicrop models dealing with production dynamics (Ozarem and Miranowski, 1994), risk aversion (Coyle, 1992, 1999 ; Chavas and Holt, 1990) and price uncertainty (Coyle, 1992, 1999 ; Moro and Sckokai, 2006) as well as models based on plot per plot discrete choice (Wu and Segerson, 1995) are not considered here. Also, we focus on models in which land is considered an allocatable fixed input (Shumway, Pope and Nash, 1984), i.e., models designed for analyzing the short-run decisions of farmers.

In studies falling into this category, the problem of variable input allocation is either considered a by-product or not considered in further detail. The first econometric models designed to model crop acreage decisions explicitly consider the variable input use allocation problem (Just, Zilberman and Hochman, 1983; Chambers and Just, 1989). Just et al. (1983) and Chambers and Just (1989) also determine variable input allocation by considering a complete model of farmer choices. Nevertheless, their econometric models are basically derived from their economic models by adding error terms to deterministic equations derived from the economic model, although Just et al. (1983) add random terms with interpretations.

Acreage allocation models considered in the 1990s mostly employ the model designed by Moore and Negri (1992) (Moore, Gollehon and Carey, 1994 ; Moore and Dinar, 1995 ; Oude Lansink and Peerlings, 1996; Bel Haj Hassine and Simioni, 2000; Bel, Lacroix, Salani et Thomas, 2006). Moore and Negri's (1992) model is a variant of Chambers and Just's (1989) model for input non-joint multicrop technology. Variable input uses are usually considered at the farm level in most of these studies that employ multi-crop econometric models (Paris, 1989).

Using a maximum entropy framework, Lence and Miller (1998)jointly estimate crop production function models and crop input uses. Their use of flexible maximum entropy estimators enables them to allocate farm input uses by using a system of production function models (that is, one for each crop) and constraining the crop input uses to sum to the input uses of each farm. Their approach lies between the approach of Dixon et al. (1984), Hornbaker et al. (1989) and the approach based on the specification of a complete model of farmer choices. The approach of Dixon et al. (1984), Hornbaker et al. (1989) does not rely on modeling the economic choices of farmers. Moreover, they do not consider input uses and acreages (and production levels in Lence and Miller's approach) as (partially) simultaneous choices.

Outline of the control function approach

The starting point of this research is that the exogeneity conditions $E\left[u_{ci}^{x}|\mathbf{s}_i,\mathbf{z}_i\right]=0$ required for the consistency of the regression based approaches are unlikely to hold in applied work. The argument for this claim is simple. The acreage choices $\mathbf{s}_i$ depend on the relative (marginal) profitability of the crops. This profitability depends on input uses and, consequently, $\mathbf{s}_i$ depends on how x_{ci} affects this profitability. Furthermore, this endogeneity problem cannot be solved by using standard instrumental variable (IV) techniques, because the error term $\mathbf{s}_i u_{ci}^{x}+\eta i$ contains the endogenous explanatory variables $\mathbf{s}_i$. The use of equation (4) as an estimating equation requires the control of the terms $E\left[u_{ci}^{x}|\mathbf{s}_i,\mathbf{z}_i\right]$.

The approach used to control these terms is based on control functions approach. The principle of the control function approach is now standard to account for endogenous sample selection (Heckman, 1974, 1979), correlated fixed effects in panel data models (Chamberlain, 1982) or endogenous explanatory variables in linear (Hausman, 1978) or non-linear models (Smith and Blundell, 1986; Petrin and Train, 2008; see also Imbens and Wooldridge, 2007 for a recent survey).

This section describes briefly the principle of the control function approach. Let us assume that we are able to define the $E\left[u_{ci}^{x}|\mathbf{s}_i,\mathbf{z}_i\right]$ terms as known functions of $\mathbf{z}_i$, $\mathbf{s}_i$ and of a vector of unknown parameters θ. Let us assume also that there exists a consistent estimator of θ, $\hat{\theta}$. The input allocation equation (4) can be defined as:

$$(7)\qquad x_i=\sum_{c=1}^{C}s_{ci}x_{ci}+c_c^x+\omega_i^x \;\text{ with }\; \omega_i^x=\sum_{c=1}^{C}s_{ci}-c_c^x+\eta_i$$

where $c_c^x\left(\mathbf{z}_i,\mathbf{s}_i;\theta\right)$ are the control functions and where the conditional expectation of $E\left[\omega_i^x|\mathbf{z}_i,\mathbf{s}_i\right]$ is null by construction. Since the $c_c^x(\mathbf{z}_i,\mathbf{s}_i;\hat{\theta})$ terms are consistent estimators of the corresponding $c_c^x\left(\mathbf{z}_i,\mathbf{s}_i;\theta\right)$ terms, equation (7) can be used to construct consistent regression based estimators of $\mathbf{a}$. The control function approach basically splits the error term u_{ci}^x in two terms: the control function $c_c^x\left(\mathbf{z}_i,\mathbf{s}_i;\theta\right)=E\left[u_{ci}^x|\mathbf{z}_i,\mathbf{s}_i\right]$ which "captures" and thus controls the links between u_{ci}^{x} and the endogenous variable vector s_i ; and a "new" error term $u_{ci}^{x} - c_c^x(\mathbf{z}_i, \mathbf{s}_i; \theta)$. By construction, s_i is exogenous with respect to the "new" error term. The crucial point is then to define the control functions $c_c^x(\mathbf{z}_i, \mathbf{s}_i; \theta)$ for $c = 1, ..., C$. This requires assumptions about the error terms of the multi-crop econometric model. In the case where the acreage share function model is defined by:

$$s_{ci} = s_{ci}(\mathbf{z}_i; \mathbf{b}) + \omega_{ci}^{s} \text{ with } E[\omega_{ci}^{s} | \mathbf{z}_i] = 0 \tag{8}$$

The control functions are determined by the following conditional expectations:

$$c_c^x(\mathbf{z}_i, \mathbf{s}_i; \theta) = E[u_{ci}^{x} | \mathbf{z}_i, \mathbf{s}_i] = E[u_{ci}^{x} | \mathbf{z}_i, s_{ci}(\mathbf{z}_i; \mathbf{b}) + \omega_{ci}^{s}] = E[u_{ci}^{x} | \mathbf{z}_i, \omega_{ci}^{s}] \tag{9}$$

As a result, it is necessary to define the relationship between the error term vectors u_{ci}^{x} and ω_{ci}^{s}. It is thus necessary to define a "structural" multi-crop econometric model, i.e. a model in which the error terms are specified as unknown determinants of the modelled choices, and not just random terms added to "make statistical noise".

Econometric model specification

Although the proposed approach can be applied with other multi-crop models with some adaptations, a specific multi-crop econometric model is considered to illustrate the basic features of our approach. It combines a standard quadratic yield functions with crop acreage share functions derived according to the specifications of Heckeleï and Wolff (2003). It is chosen because of its fairly simple interpretation and flexibility. A specific feature of our model is its capacity to structurally model econometric error terms (McElroy, 1987). The model is considered in its simplest version, *i.e.*, with constant parameters. In empirical studies, most of the parameters may usefully be defined as parametric functions of the observed exogenous variables to control as much as possible for the heterogeneity of farms and farmers. Finally a single-variable input is considered for simplicity.

Yields and input demand functions

The yield y_{ci} of each crop c ($c = 1, , C$) for farm i ($i = 1, , N$) is assumed to be a quadratic function of the single-variable input for simplicity. This function represents the short-term "agronomic" yield function and is defined as:

$$y_{ci} = \alpha_{ci} - 0.5\gamma_c^{-1} \left(\beta_{ci} - x_{ci}\right)^2 \tag{10}$$

$$\text{with} \quad \alpha_{ci} = \alpha_{0c} + \alpha_{1c} s_{ci} + v_{ci}^y \tag{11}$$

$$\text{and} \quad \beta_{ci} = \beta_{0c} + \beta_{1c} s_{ci} + v_{ci}^x \tag{12}$$

where x_{ci} is the quantity of variable input used per hectare by farm i devoted to crop c, and α_{ci}, β_{ci} and γ_c are parameters to be estimated, with $\alpha_{ci} > 0$, $\beta_{ci} > 0$ and $\gamma_c > 0$. This alternative specification of the standard quadratic function is also used by Pope and Just (2003), albeit for other purposes. The yield function is strictly concave if $\gamma_c > 0$. Under this assumption, the term α_{ci} can be interpreted as the maximum yield of crop c for farm i. The variable input quantity required for achieving this maximum yield is given by β_{ci}. The maximum yield and input requirements are specified as functions of the crop acreage to account for potential scale effects. The estimates of these yield functions can thus be checked by agricultural scientists. v_{ci}^y and v_{ci}^x are random terms. These terms are split into two parts to simplify interpretation:

$$v_{ci}^y = e_{ci}^y + \epsilon_{ci}^y \quad \text{and} \quad v_{ci}^x = e_{ci}^x + \epsilon_{ci}^x \tag{13}$$

The terms e_{ci}^y and e_{ci}^x are denoted as heterogeneity terms. They represent effects on the yield of crop c from factors that are known to farmer i at the time he chooses his acreages (*e.g.*, rotation effects, soil quality and quasi-fixed input availabilities). These terms are closely related to the so-called fixed effects discussed in the panel data econometrics literature (Griliches and Mairesse, 1995), but they may not be permanent in context of the current framework. They are considered as random because they are unknown to the econometrician. The terms ϵ_{ci}^{y} and ϵ_{ci}^{x} are denoted as stochastic events. They represent effects on the yield of crop c from factors that are unknown to farmer i at the time he chooses his acreages (*e.g.*, climatic conditions and pest infestations). These factors are considered random because they vary across farms and years, and are unknown to the econometrician. Their expectations are normalized to be zero.

The production of crop c is sold at price p_{ci}, and the input is bought at price w_i by farmer i. These prices are assumed to be known at the beginning of the production process, *i.e.*, when acreages are chosen. Farmers are assumed to be risk-neutral. Farmer i is also assumed to choose his input use by maximizing following gross margins $\pi_{ci} = p_{ci}y_{ci} - w_i x_{ci}$ for each crop c. Variable input and target yield choices are set based on output and input prices and adjusted to specific production conditions after a farmer has observed ϵ_{ci}^{y} and ϵ_{ci}^{x}. We thus allow farmers to make production decisions in two step. First, at the beginning of the production process, they choose acreages and input uses for each crop. Second, during the production process, they can adjust their input uses after they have observed stochastic events (such as specific climatic conditions or diseases). Therefore, acreages and input uses decisions are partially simultaneous. The maximisation of this profit function under technological constraints leads to the following per hectare variable input demand, yield supply and gross margin functions:

$$(14) \qquad x_{ci} = \beta_{ci} - \gamma_c\,(w_i/p_{ci}) + v_{ci}^x$$

$$(15) \qquad y_{ci} = \alpha_{ci} - 0.5\gamma_c\,(w_i/p_{ci})^2 + v_{ci}^y$$

$$(16) \qquad \pi_{ci}^e = p_{ci}\alpha_{0c} - w_i\beta_{0c} + 0.5s_{ci}(p_{ci}\alpha_{1c} - w_i\beta_{1c}) + 0.5\gamma_c p_{ci}\,(w_i/p_{ci})^2 + p_{ci}e_{ci}^y - w_i e_{ci}^x$$

Consequently, v_{ci}^x can be interpreted as the effects production conditions that can be corrected by variable input uses, while v_{ci}^y represents the effects of production conditions already under way. The quadratic yield has a main practical advantage, as it provides yield supply and variable input demand functions with additive error terms. This feature appears to be very useful for analysing the error term structure of the econometric model; see McElroy (1987) and Pope and Just (2003) in other contexts. Distinguishing heterogeneity effects and stochastic events in the yield function allows us to determine the gross margins of the crops as expected by farmers at the time they choose their acreages. The gross margin expectations of farmers cannot depend on the ϵ_{ci}^y and ϵ_{ci}^x terms because these terms are unknown when farmers choose their acreages.

Acreage functions

The acreage choices of farmers are modeled within the framework developed by Heckeleï and Wolff (2003). This framework is simple and flexible and, it draws on both the econometric literature and the positive mathematical programming (PMP) literature on production choice modeling. Farmer i is assumed to allocate his total land quantity S_i by maximizing the following indirect restricted profit function:

$$(17) \qquad \sum_{c=1}^{C} s_{ci}\pi_{ci}^e(s_c) - C(s_c)$$

where s_{ci} is the acreage share devoted to crop c by farmer i. This restricted profit function is strictly concave in $\mathbf{s}$. According to this model, farmers have two motives for crop diversification, namely, the scale effects of the crop gross margins $0.5(p_{ci}\alpha_{1c} - w_i\beta_{1c})$ and the implicit management cost of the chosen acreage $C(s_c)$. This cost function is used in the literature on positive mathematical programming (Howitt 1995; Paris and Howitt 1998; Heckeleï and Wolff 2003). It can be interpreted as a reduced form function that smoothly approximate the unobserved variable costs associated with a given acreage (such as energy costs) and the effects of binding constraints on acreage choices. These constraints are agronomic constraints or constraints associated with limiting the quantities of quasi-fixed inputs. Quasi-fixed inputs such as labour or machinery are limiting in the sense that their costs per unit of land devoted to a given crop are likely to increase due to peak work loads or machinery overuse, whether machinery is crop specific or not. Farmers are also subject to agronomic constraints because some crop rotations are "forbidden" or impossible due to issues with planting and harvesting dates. Cultivating a given crop for two consecutive years on the same plot may be strongly unwarranted due to dramatic expected pest damages. These crop rotations are thus highly restricted because their opportunity cost is very large within standard price ranges, and thus these crop rotations determine the bounds imposed on acreage choices in PMP models. The implicit cost function $C(s_c)$ is assumed to be non-decreasing and quasi-convex in acreages to reflect the constraints due to the limiting quantities of quasi-fixed factors (other than land) as well as due to the implicit bounds imposed on the acreage choices due to crop rotations. This cost function is assumed to have a quadratic form:

$$C(s_c) = a_i + \sum_{c=1}^{C} g_{ci}s_{ci} + 0.5\sum_{c=1}^{C}\sum_{m=1}^{C} g_{cm}s_{ci}s_{mi} \text{ with } g_{ci} = g_{0c} + e_{ci}^{g} \tag{18}$$

where a_i, g_{ci} and g_{cm} are parameters to be estimated. The term a_i is a constant. The fixed cost g_{ci} per unit of land of crop c for farmer i is split into two parts, where g_{0c} is a parameter and e_{ci}^{g} is a random term accounting for the cost heterogeneity term known to farmer i but unknown to the econometrician. If the matrix $\mathbf{G} = [g_{cm}, c, m = 1, ...C]$ is definite positive, then the cost function $C(s_c)$ is strictly convex in acreages.

The land use constraint is included in the restricted indirect profit function. All crops are assumed to be cultivated. Crop C is considered the reference crop. The maximisation of this restricted indirect profit function leads to $C-1$ acreage functions. These acreage functions have a closer form but we use first-order conditions to simplify notations here and in the empirical application below:

$$\sum_{m=1}^{C-1} Q_{mi}s_{mi} + (\pi_{ci} - \pi_{Ci}) - (g_{ci} - g_{Ci}) - F_{ci} = v_{ci}^{s} \tag{19}$$

$$\text{with } \; v_{ci}^{s} = p_{ci}e_{ci}^{y} - p_{Ci}e_{Ci}^{y} - w_i e_{ci}^{x} + w_i e_{Ci}^{x} - e_{ci}^{g} \tag{20}$$

where $m = 1, ..., C-1$. The terms Q_{mi} and F_{ci} depend on output and input prices, scale effects α_{Ci} and quadratic costs terms g_{cmi}. They are described more precisely in Appendix A. These acreage functions have two interesting features. First, they have additive error terms. Second, these errors terms contain the heterogeneity parameters of the input demand and yield supply functions e_{ci}^{y} and e_{ci}^{x}.

"Complete" multi-crop econometric model

The multi-crop econometric model is composed of three subsets of equations, that is, yield equations, acreage equations and an input allocation equation. The total variable input X_i is written as the sum of the acreage share devoted to each crop c multiplied by the per hectare variable input quantity used for each crop c : $X_i = \sum_{c=1}^{C} s_{ci}x_{ci}$. This equation allows us to allocate variable inputs across crops c as part of the econometric model. The complete system is described in Appendix A using simple matrix notation.

It seems pertinent to again define the error terms of the econometric systems of equations.

Note that the error terms of each equation are now denoted by u^y_{ci}, u^x_i and u^s_{ci} and that u_{ci} is similar to v_{ci} except for the input allocation equation.

$$u^y_{ci} = v^y_{ci} = e^y_{ci} + \epsilon^y_{ci} \tag{21}$$

$$u^x_i = s_{ci} v^x_{ci} + \eta_i = s_{ci}[e^x_{ci} + \epsilon^x_{ci}] + \eta_i \tag{22}$$

$$u^s_{ci} = v^s_{ci} = p_{ci} e^y_{ci} - p_{Ci} e^y_{Ci} - w_i e^x_{ci} + w_i e^x_{Ci} - e^g_{ci} \tag{23}$$

An error term η_i is introduced in the input allocation equation and represents the effects of measurement errors due, e.g., to stock variations. To explain the endogeneity problem, we use simple matrix notation. We consider the vector of heterogeneity terms $\mathbf{e}_i$ as $\mathbf{e}'_i = [\mathbf{e}^y_i \quad \mathbf{e}^x_i \quad \mathbf{e}^g_i]$ and the vector of stochastic events terms $\boldsymbol{\epsilon}_i$ as $\boldsymbol{\epsilon}'_i = [\boldsymbol{\epsilon}^y_i \quad \boldsymbol{\epsilon}^x_i]$. The vectors for model error terms are defined by $\mathbf{u}^y_i = [u^y_{ci}]$, $\mathbf{u}^x_i = [u^x_i]$ and $\mathbf{u}^s_i = [u^s_{ci}]$. The vectors $\mathbf{z}_i$ and $\mathbf{s}_i$ are the vector of exogenous variables (that is, outputs and inputs prices) and the vector of acreage shares, respectively. The interpretations of the error terms discussed above allows us to define the conditional mean assumptions such that: $E[\mathbf{e}^y_i|\mathbf{z}_i] = 0$, $E[\boldsymbol{\epsilon}_i|\mathbf{z}_i] = 0$, $E[\mathbf{e}^g_i|\mathbf{z}_i] = 0$, $E[\eta_i|\mathbf{z}_i] = 0$ and $E[\mathbf{s}'_i\boldsymbol{\epsilon}^x_i|\mathbf{z}_i] = 0$. This implies that each component of $\mathbf{u}_i$ has a null expectation conditional on prices, except for the $\mathbf{s}_i\mathbf{e}^x_i$ term in the input allocation equation. Although $\mathbf{s}_i$ is an endogenous explanatory variable, this is a standard problem that can be adressed with standard instrumental variable techniques. The main problem is that $E[\mathbf{s}_i\mathbf{e}^x_i|\mathbf{z}_i] \neq 0$ or $E[\mathbf{e}^x_i|\mathbf{z}_i, \mathbf{s}_i] \neq 0$. These terms thus must be determined. Before proceeding to the determination of the control functions, two remarks are in order. First, the yield supply and acreage choice functions identify almost the entire set of parameters. Only β_{0C} cannot be identified. Second, the heterogeneity terms $\mathbf{e}^y_i$, $\mathbf{e}^x_i$ and $\mathbf{e}^g_i$ are the error terms of interest for determining the control functions whereas $\boldsymbol{\epsilon}^y_{ci}$, $\boldsymbol{\epsilon}^x_i$ and η_i can be viewed as nuisance parameters.

Control function approach

Our econometric model is structural, *i.e.* it provides explicit forms for the relationship between the error term vectors of the yield supply, input demand allocation and acreage equations. The main problem involves linking the acreage and the input use choices in the variable input allocation equation. The control function idea is to explicitly determine this link and its associated estimator to integrate this term in the full multi-crop econometric model.

Different approaches based on control functions

Two types of approach can be used. The one considered here is conditional on $\mathbf{s}_i$ and is based on the functional form of the $E[\mathbf{e}_i^x|\mathbf{z}_i, \mathbf{s}_i]$ terms. Another approach would be based on the functional form of the $E[\mathbf{s}_i\mathbf{e}_i^x|\mathbf{z}_i]$ terms. This second approach relies on less restrictive assumptions but requires more involved computations. Wooldridge (2008) distinguishes both approaches, denoting the functional form of $E[\mathbf{e}_i^x|\mathbf{z}_i, \mathbf{s}_i]$ by the usual term "control function" and denoting the functional form of $E[\mathbf{s}_i\mathbf{e}_i^x|\mathbf{z}_i]$ by the term "correction function".

The construction of control functions relies on some assumptions. First, it is shown that distributional assumptions are generally necessary to define control functions for the general multi-crop econometric model (Imbens and Wooldridge, 2007). The normal distribution usually appears to be a convenient choice. However linear projection techniques combined with limited assumptions on the distribution of the heterogeneity terms can be used in some special cases (Chamberlain, 1982; Wooldridge, 2004).

Then both types of approach rely on the additional conditional mean and homoskedasticity assumptions: $E[\mathbf{e}_i|\mathbf{z}_i] = 0$, $V[\mathbf{e}_i|\mathbf{z}_i] = \mathbf{\Psi}$ and $E[\boldsymbol{\epsilon}_i|\mathbf{z}_i, \mathbf{e}_i] = 0$. It is further assumed that $\mathbf{e}_i^x$, $\mathbf{e}_i^y$ and $\mathbf{e}_i^g$ are not correlated. This assumption is not necessary but it simplifies the approach and may appear empirically reasonable. As a result, the variance-covariance matrix of $\mathbf{e}_i$ has the following structure:

$$(24)\ V[\mathbf{e}_i|\mathbf{z}_i] = V[\mathbf{e}_i] = \mathbf{\Psi} = \begin{pmatrix} \mathbf{\Psi}_{yy} & \mathbf{\Psi}_{yx} & 0 \\ \mathbf{\Psi}_{yx} & \mathbf{\Psi}_{xx} & 0 \\ 0 & 0 & \mathbf{\Psi}_{gg} \end{pmatrix} = \begin{pmatrix} \mathbf{\Psi}_{yz} & 0 \\ \mathbf{\Psi}_{xz} & 0 \\ 0 & \mathbf{\Psi}_{gg} \end{pmatrix} \text{ with } \mathbf{e}_i = \begin{pmatrix} \mathbf{e}_i^y \\ \mathbf{e}_i^x \\ \mathbf{e}_i^g \end{pmatrix}$$

The main implications of these additional assumptions for the control function purpose concern the conditional variance-covariance structure of the error terms of the econometric model. In fact, these assumptions allow to determine moment conditions that can be used to define regression estimators of the useful parts of the variance-covariance matrix $\mathbf{\Psi}$(see section on the implementation of the approach).

Control functions under normality assumptions

Determining control functions requires additional assumptions with respect to either the structure of the model, or the distribution of the $\mathbf{e}_i$ terms. Distributional assumptions are the most frequent basis for determining control functions (Imbens and Wooldridge, 2007). It is assumed that $\mathbf{e}_i$ is jointly normal conditional on $\mathbf{z}_i$, *i.e.* its entire distribution is characterized by its null conditional mean and its conditional variance-covariance matrix $\mathbf{\Psi}$. Since all the considered error terms of the model $\mathbf{u}_i^y$, $\mathbf{u}_i^x$ and $\mathbf{u}_i^s$ are linear transformations of $\mathbf{e}_i$, they are also normally distributed.

The control functions defined in this article seek to solve two problems, namely, the non null expectation of $\mathbf{s}_i'\mathbf{e}_i$ and the endogeneity of $\mathbf{s}_i$ in the input allocation (and yield supply) equation(s). To solve the second problem, one needs to determine the expectation of $\mathbf{u}_i^x$ conditional on $\mathbf{z}_i$ and $\mathbf{s}_i$. The properties of the conditional expectation operator and the additivity of the error terms of the acreage equations allow to show that:

$$(25) \qquad E[\mathbf{u}_i^x|\mathbf{s}_i, \mathbf{z}_i] = \mathbf{s}_i' E[\mathbf{e}_i^x|\mathbf{s}_i, \mathbf{z}_i]$$

The conditioning properties of normally distributed vectors and the zero conditional mean of $\mathbf{e}_i^x$, $\mathbf{u}_i^y$, $\mathbf{e}_i^y$ and $\mathbf{u}_i^s$ allow then to show that:

$$E[\mathbf{e}_i^x|\mathbf{s}_i,\mathbf{z}_i] = \boldsymbol{\Psi}_{xz}\mathbf{C}_i^x\mathbf{u}_i^s \quad \text{and} \quad E[\mathbf{e}_i^y|\mathbf{s}_i,\mathbf{z}_i] = \boldsymbol{\Psi}_{yz}\mathbf{C}_i^y\mathbf{u}_i^s \tag{26}$$

where $\mathbf{C}_i = [\mathbf{C}_i^y \quad \mathbf{C}_i^x]$ depends on output and input prices and a part of the variance-covariance matrix of $\mathbf{e}_i$. This matrix is presented in Appendix B. Under the joint normality assumption, the form of $\mathbf{C}_i$ is known thanks to our structural econometric model and thanks to the error term structure defined previously. It is then possible to integrate these control functions in the yield supply and input demand allocation equations to capture the correlation between heterogeneity error terms and acreages.

Empirical application

This section considers the implementation of the control function approach for general case. It presents a simple three-stage inference procedure. Details of this procedure are presented in Annex C. This brief description of the procedure mainly focuses on identification and consistency issues and ignores efficiency issues. A simple empirical application based on French farm-level data is then presented to illustrate the control function approach.

A three-stage procedure

In the first stage, the system composed of the yield supply and acreage choice equations is estimated in order to construct a consistent estimator of identifiable parameters, *i.e.* all parameters except β_{0C}. This system is a simultaneous equation system due to the endogeneity of the acreage choices. The estimation of this system of equations used the Generalized Method of Moments (GMM) estimator. An estimated instrument for $\mathbf{s}_i$ is necessary in yield supply equations.[3] This stage allows to obtain $\mathbf{u}_i^s$ and to proceed to the next stage.

In the second stage, estimators from the first stage are used to construct a consistent estimator of a part of the variance-covariance matrix $\boldsymbol{\Psi}$. This stage is similar to the second stage of the construction of the standard Generalized Least Square (GLS) estimator. It relies on the second-order moment conditions and uses a SUR system linear in its parameters. This stage allows us to obtain an estimate of $\mathbf{C}_i$.

The third stage of the procedure considers the estimation of the complete system composed of yield supply, input allocation and acreage choice equations. Control functions are integrated in the yield supply and input allocation equations. All parameters of interest are estimated, including auxiliary parameters such as $\boldsymbol{\Psi}_{yz}$ and $\boldsymbol{\Psi}_{xz}$. This econometric model is not a standard non-linear SUR system because the different equations of the system share many parameters. The corresponding SUR estimators are generally non consistent. Thus, we use the GMM to construct a consistent estimator.

This approach can be interpreted as a generalized version of the augmented regression technique used to control for the endogeneity of explanatory variables in models linear in their explanatory variables. The augmented regression test can be used to test the endogeneity of $\mathbf{s}_i$ in the yield supply and input demand allocation equations. The null hypothesis is then $\boldsymbol{\Psi}_{yz} = \boldsymbol{\Psi}_{xz} = 0$. This is a test of the approach proposed in this study. If the null hypothesis is rejected then acreages are endogenous in the yield supply and input demand equations.

The data

The three-stage procedure is applied to a sample of French grain crop producer over 1988-2006 using a rotating panel data sample from the French Farm Accountancy Data Network (FADN). It contains approximately 4,000 observations. The available information includes acreage, yield and price for each crop and variable input expenditures at the farm level. Six different crop group are considered, such as wheat, other cereals (mainly barley and corn), oilseeds (mainly rapeseed) and protein crops (mainly peas), sugar beets, potatoes and miscellaneous crops, and fodder crops. Acreages for sugar beets, potatoes and miscellaneous crops, and fodder crops are considered as exogenous, because most of them are contract crops.[4] The different variable inputs (*i.e.*, fertilizers, pesticides, energy and seeds) are aggregated into a single-variable input for simplicity. The corresponding price index is obtained from French agricultural statistics. All economic quantities are defined in € in units of 2000.

The system is composed of three yield supply equations (for wheat, other cereals and oilseeds and protein crops), one input allocation equation (for agregated input) and two acreage choice equations (for wheat and other cereals). Oilseeds and protein crops are the reference crop. In the input demand equation, variable inputs are allocated between all crops, *i.e.*, wheat, other cereals, oilseeds and protein crops, sugar beets, potatoes and miscellaneous crops, and fodder crops.

Some variables were introduced into the model to control for technical change and the heterogeneity of farms. The parameter α_{0c}, which is interpreted as the maximum yield of crop c in the yield supply equations, is defined as a function of several exogenous variables, including: a quadratic trend, regional dummies, the lagged acreage shares of sugar beets and potatoes, to account for the beneficial effects of the induced crop rotations, the acreage share of cereal except corn (of the total acreage of cereals except wheat) and the acreage share of protein crops (of the total acreage of oilseeds and protein crops). The parameter β_{0c}, which is interpreted as the variable input quantity required for achieving the maximum yield in the input allocation equation, depends only on a trend. In the acreage choice equations, the parameter of fixed costs g_{0c} for crop c is defined as a function of physical capital and labor variables because it is interpreted as the fixed costs associated with limiting quantities of quasi-fixed inputs such as labor or machinery.

Main results

The multi-crop econometric model is estimated following the three-stage procedure described in the last section. It is denoted by model 1. Results are presented in Appendix C. Table

1 presents the estimates for yield supply, input demand and acreage share parameters. The fit of the model is correct given that we use data at the farm level. The R^2 criterion ranges from 0.16 to 0.31 for yield equations and 0.12 to 0.22 for acreage choice equations; and it equals 0.33 for the input allocation equation. Almost 90% of parameters are significantly different from zero at the 10% confidence level (or less).

The price effects correspond to parameters γ_c for each crop c associated with the price ratio. These parameters are significantly positive for all crops, implying concavity for the yield functions. The parameter α_{0c} represents the maximum yield of crop c. It is estimated at €9.45 per are for wheat, €9.98 per are for other cereals and €8.63 per are for oilseeds and protein crops. These parameters are defined as functions of exogenous variables. The effects of these heterogeneity control variables are as expected. The past acreages of sugar beets and potatoes have positive effects on cereals yield. These effects are consistent with the known beneficial effects of root crops at the beginning of the crop rotation sequence. The variable corresponding to agregate "others cereals" has the expected sign. This means that corn has a more important yield than other cereals. The paramater α_{1c}, which is associated with the acreage of the crop c in the yield supply equations, can be interpreted as a scale effect. It is significant and negative for other cereals, oilseeds and protein crops, which confirms that the yield of these crops decreases with the land allocated to these crops.

The parameter β_{0c} represents the variable input quantity required to achieve maximum yield for crop c. It is estimated at €7.62 per are for wheat, €7.04 per are for other cereals and €5.29 per are for oilseeds and protein crops. These parameters are defined as a function of a trend. In this context, we assume for simplicity that $\beta_{1c} = 0$. Parameters associated with sugar beets, potatoes and fodder crops in the input demand equations represents the quantity of variable inputs allocated to these crops. We observe that farmers use important quantities of variable inputs to grow sugar beets and potatoes. These quantites are estimated at €7.40 per are for sugar beets and €14.27 per are for potatoes.

In the acreage choice equations, there are two sets of parameters to estimate, namely, fixed costs g_{0c} and terms Q_c. The paramater g_{0c}, which represents the fixed costs of crop c cannot be identified by the complete system. Thus, we estimate only the difference of fixed costs $g_{0c} - g_{0C}$ between a crop c and a reference crop C. Consequently, the estimated parameter must be interpreted with caution. The difference in fixed costs between wheat and oilseeds and protein crops is €-1.98 per are[5], and the difference between the other cereals and oilseeds and protein crops is €-0.35 per are. This means that wheat and other cereals require less fixed costs than oilseeds and protein crops. This can be explained by the fact that French farmers specialize in wheat, and therefore, crop management is more expensive for farmers who cultivate oilseeds and protein crops. These differences in fixed costs are defined as a function of physical capital and labor variables. We observe that these variables have a significant and negative effects that confirms the previous result. The term Q_c accounts for the motives behind crop diversification for farmers. In fact, it depends on the parameters α_{1c}, which represents scale effects, and on the parameters of the quadratic cost function g_{cm}. All these parameters are not identifiable, and so we estimate only a paramater for each crop c. These estimated parameters imply concavity in the restricted profit function without imposing constraints. These results globally indicate that this model provides sensible results with respect to price effects and heterogeneity control variable effects.

We aim to show that there is a problem of acreage endogeneity that may have consequences on our results, especially with respect to variable input allocation. Parameters of the control functions $\mathbf{\Psi}_{xz}$ and $\mathbf{\Psi}_{yz}$, which are estimated using the complete model, are elements of the matrice variance-covariance $\mathbf{\Psi}$ of heterogeneity error terms $\mathbf{e}_i$. More than 55% of these parameters are significantly different from zero based on t-tests. We also wanted to test for joint significance using a Wald test. The null hypothesis is $\mathbf{\Psi}_{xz} = \mathbf{\Psi}_{yz} = 0$. The t-test statistic is about 100.83 with p-value < 0.001. Thus, we reject the null hypothesis and conclude that these parameters are jointly statistically different from zero. This test confirms that there is an acreage endogeneity problem in the yield and input equations.

In addition to these tests, we estimate another model (model 2) with the same structure using the same data without the control functions. Table 2 presents the estimated parameters from the model 2, and Tables 3 and 4 report its main differences as compared to the previous model (model 1), especially in terms of input allocation. There are two important differences between models related to the acreage endogeneity.

The first important difference is that the R^2 criteria of yield equations for wheat and other cereals are better under model 1. These criteria are 0.21 and 0.14 for wheat and other cereals, respectively, in model 2 and 0.31 and 0.28 in model 1. This means that there are non-observed factors that influence both acreage choice and cereals yield. Control functions allow us to take into account the effects of these factors on yield and thus to improve the fit of the model. These effects are non-observed by the econometrician but are known by the farmer at the moment he decides on land use. One factor may be, for example, the quality of land in a farm. This variable is rarely available to the econometrician, and it influences both the choices of crops and their yield.

The second important result is that variable input allocations are different from model 1 to model 2. Variable input allocations among crops are presented in Table 3. Under model 2, the quantity of allocated variable inputs is €4.29 per are for wheat, €4.80 per are for other cereals and €1.91 per are for oilseeds and protein crops. Under model 1, the quantity of allocated variable inputs is €4.47 per are for wheat, €4.59 per are for other cereals and €1.77 per are for oilseeds and protein crops. These average differences between the two models are not statistically significant. But there are more interesting results if we analyze differences at the farmer level. We note that allocations of inputs for oilseeds and protein crops are overestimated by more than 9% and up to 29% for half of the sample farms, which represents a difference between €0.18 and €0.32 per are. This result shows that there are non-observed factors that influence both the choice of acreage and input uses. These factors tend to overestimate input quantities applied to oilseeds and protein crops and affect only a portion of farmers.

To better understand this last effect, we observe specific characteristics of these farmers.

Table 4 reports some interesting results. We construct several samples of farmers based on the difference in input allocations for oilseeds and protein crops between models. We have four samples of farmers. For example, in the first sample, the difference between the two models in input quantity allocated to oilseeds and protein crops is less than €0.1 per are. This sample is composed of 600 farmers. We then calculate the average of several variables for each sample to explain the difference between these input allocations. It seems that farmers with the most biased allocations (*i.e.*, more than >€2 per are) are those who receive the highest subsidies per are for oilseeds and protein crops. Farmers in the first sample, who show no differences across the two models, have an average of €1.02 per are subsidies for oilseeds and protein crops. However, farmers in the last sample show a difference greater than €2 per are and have an average of €4.71 per are subsidies for the same crops. Moreover, we observe that the more farmers get subsidies for oilseeds and protein crops the more they cultivate protein crops at the expense of oilseeds crops. The acreage share of protein crops in the total of oilseeds and protein crops ranges from 22% for farmers in the first sample to 48% for farmers in the fourth sample, whereas the share of oilseeds and protein crops in total area is constant, as it ranges between 20% and 24%. Only the total subsidies perceived by farmers for oilseeds and protein are available, but we can suppose that these subsidies support the culture of protein crops. Since protein crops require much less variable inputs than oilseeds, subsidies also have effects on input uses. Control functions allow us to capture these effects and thus to better allocate variable inputs among crops. It is interesting to note that the built samples correspond to specific periods. In the first sample, the estimated input quantity for oilseeds and protein crops are similar in model 1 and model 2. Farmers in this sample receive the lowest subsidies; this sample corresponds to the year 2006-2007. However, the estimated input quantity for oilseeds and protein crops are very different from one model to another in the fourth sample. Farmers in this sample receive higher subisidies; this sample corresponds to the years 1995-1999. These results show that it is important to consider acreage endogenous in a production choice model and confirms the usefulness of the proposed approach.

Conclusion

In this article, we present an approach to allocate variable inputs among crops. This approach has potentially two main drawbacks. First, the econometric model used here does not account for corner solutions of activity choices. This is a potentially important weakness in this framework, particularly in the crop production context. Nevetheless the specification of a fully structural model for activity choices with corner solutions is possible but more difficult to implement. Second, the identification of the control functions relies on models based on squares and cross-products of the crop and input prices. As a result, the empirical identification of these functions requires good-quality price data at the farm level.

Nevertheless this article highlights two important points about variable input allocation among crops.

First, we show that it is important to consider acreage endogeneity to allocate variable inputs among crops. The standard regression-based approaches for allocating variable input uses to crops are potentially biased due to the partial simultaneity of the expected crop variable input and acreage choices. This bias is even more important given that few variables are generally available to control for heterogeneity among farmers. The test built and applied on our data confirms this intuition. The comparison of models with and without control functions shows the usefulness of considering acreage endogenous. Differences in input allocations between the models are not significant in average. By analyzing input allocations at the farmer level, we observe differences across two proposed models, especially with respect to the quantity of input allocated to oilseeds and protein crops.

Second, we suggest that a structural econometric model is necessary to account for the bias associated with acreage endogeneity. In this article, we propose a structural econometric multi-crop model to explicitly determine the origin of bias and provide potential solutions to allocate inputs among crops. This model is composed of yield supply, input demand and acreage choices equations. We consider land an input fixed and allouable as in the case in the main literature on production choices model (Chambers and Just 1989; Moore and Negri 1992 and many others). The main feature of our model is that it allows an explicit specification of the links between yield, input uses and acreage choices. The structural modeling of error terms, especially error term additivity, plays a crucial role in the proposed approach. This approach could be applied by using other structural econometrics models with an explicit specification of these deterministic and random links between choices production. This model could also be applied in other contexts in which inputs must be allocated to activities.

References

Bhattarcharyya, A., E. Parker and K. Rafie (1994), "An Examination of the Effect of Ownership on the Relative Efficiency of Public and Private Water Utilities", Land Economics, 70, 197-209.

Bel, F., Lacroix, A., Salani, F. and A. Thomas. 2006. "Evaluating the Impact of CAP Reforms on Land Use and the Environment: a Two-Step Estimation with Multiple Selection Rules and Panel Data", Working paper.

Bel Haj Hassine and M. Simioni. 2000. "Estimation of Two-Stage Models of Multicrop Production: With and Application to Irrigated Water Allocation in Tunisian Agriculture" Rgion et Dveloppement,12, 2000, p. 121-141.

Chamberlain, G. 1982. "Multivariate Regression models for panel data" Journal of Econometrics 18:5-46.

Chambers, R. et R. Just. 1989. "Estimating Multiouput Technologies" American Journal of Agricultural Economics 71:980-95.

Chavas, J.-P. et T. Holt. 1990. "Acreage Decisions under Risk: The Case of Corn and Soybeans" American Journal of Agricultural Economics 72:529-538.

Coyle, B. T. 1992. "Risk Aversion and Price Risk in Duality Models of Production: A Linear Mean-Variance Approach" American Journal of Agricultural Economics 74:849-859.

Coyle, B. T. 1999. "Risk Aversion and Price Risk in Duality Models of Production: A Mean-Variance Approach" American Journal of Agricultural Economics 81:553-567.

Dixon, B. L., M. T. Batte and S. T. Sonka. 1984. "Random Coefficients Estimation of Average Total Product Costs For Multiproduct Firms" Journal of Business Econonomics and Statististics 2: 360-366.

Dixon, B. and R.H. Hornbaker. 1992. "Estimating the Technology Coefficients in Linear

Programming Models" American Journal of Agricultural Economics 74:1029-1038.

Griliches, Z. and J. Mairesse. 1999 "Production Functions: the Search for Identification" In S. Strom (ed.), Essays in Honour of Ragnar Frisch, Econometric Society Monograph Series, Cambridge University Press, Cambridge.

Hausman, J. (1978) "Specification tests in econometrics," Econometrica, 46, 1251-1271.

Heckele, 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.

Heckman, J. J. 1979. "Sample selection bias as a specification error" Econometrica 47, 153-161.

Heckman, J. J. 1974. "Shadow prices, market wages, and labor supply" Econometrica 42, 679-693.

Heckman, J. J., J. L. Tobias, and E. Vytlacil. 2003. "Simple estimators for treatment parameters in a latent-variable framework" Review of Economics and Statistics 85, 748-755.

Hornbaker, R. H., B. L. Dixon and S. T. Sonka. 1989. "Estimating Production Activity Costs for Multioutput Firms with a Random Coefficient Regression Model" American Journal of Agricultural Economics 71:167-177.

Imbens, G. and J. M. Wooldridge. 2007. "Control Functions and Related Methods" Lecture notes NBER Summer Institute '07, What's new in econometrics?, NBER.

Just, R. E., Zilberman, D., Hochman, E., and Z. Bar-Shira. 1990. "Input Allocation in Multicrop Systems" American Journal of Agricultural Economics 72:200-209.

Just, R.E., Zilberman, D., et E. Hochman. 1983. "Estimation of Multicrop Production Functions" American Journal of Agricultural Economics 65:770-780.

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

McElroy, M. B. 1987. "Additive General Error Models for Production, Cost, and Derived Demand or Share systems" Journal of Political Economy 95:737-757.

Moore, M.R. and A. Dinar. 1995. "Water and Land as Quantity-Rationed Inputs in California Agriculture: Empirical Tests and Water Policy Implications" Land Economics 74(4):445-461.

Moore, M.R., Gollehon, N.R. and M.B. Carey. 1994. "Multicrop Production Decisions in Western Irrigation Agriculture: The Role of Water Price" American Journal of Agricultural Economics 11:143-158.

Moore, M. R., and D. H. Negri. 1992. "A Multicrop Production Model of Irrigated Agriculture, Applied to Water Allocation Policy of the Bureau of Reclamation" Journal of Agricultural and Resource Economics 17:29-43.

Moro, D. et P. Sckokai. 1999. "Modelling the CAP Arable Crop Regime in Italy: Degree of Decoupling and Impact of Agenda 2000" Cahiers d'conomie et sociologie rurales 53:50-73.

Oude Lansink, A. and J. Peerlings. 1996. "Farm-specific Impacts of Quantitative Restrictions of N-Fertiliser Use in Dutch Arable Farming" Journal of Agricultural economics 52:38-52.

Ozarem, P.F. and J.A. Miranowski. 1994. "A Dynamic Model of Acreage Allocation with General and Crop-Specific Soil Capital" American Journal of Agricultural Economics 76:385-395.

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

Peeters, L. et Y. Surry.1993. "Estimating Feed Utilisation Matrices Using a Cost Function Approach" Agricultural Economics 9:109-126.

Petrin, A. and K. E. Train. 2008. “A control function approach to endogeneity in consumer choice model” Forthcoming, Journal of Marketing Research.

Pope, R. D. and R. E. Just. 2003. “Distinguishing Errors in Measurement from Errors in Optimization” American Journal of Agricultural Economics 85:348-358.

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

Smith, R. J. and R. W. Blundell. 1986. “An exogeneity test for a simultaneous equation tobit model with an application to labor supply” Econometrica 54, 679-686.

Wooldridge, J. M., 2008. “Instrumental Variables Estimation of the Average Treatment Effect in Correlated Random Coefficient Models,” in Advances in Econometrics, Volume 21, Millimet D., J. Smith, and E. Vytlacil (eds.), 93-117. Amsterdam: Elsevier.

Wooldridge, J. M. 2004. “Estimating Average Partial Effects under Conditional Moment Independence Assumptions” The Institute for Fiscal Studies, Dpt of Economics, UCL. Cemmap, WP CWP03/04.

Wu, J. et K. Segerson. 1995. “The Impact of Policies and Land Characteristics on Potential Groundwater Pollution in Wisconsin” American Journal of Agricultural Economics 77:1033-47.

Appendix A

The complete model

We can define the complete model with simple matrix notations:

$$
\begin{cases}
\mathbf{y}_i & = \mathbf{a}_0^y + \mathbf{B}_0^y \mathbf{s}_i + \mathbf{u}_i^y \\
X_i & = \mathbf{s}_i' [\mathbf{a}_0^x + \mathbf{B}_0^x \mathbf{s}_i] + u_i^x \\
\mathbf{Q}_0 & \mathbf{s}_i^- + [\Delta' \mathbf{M}_i \mathbf{a}_0 - \mathbf{g}_0 - \mathbf{F}_{Ci}] = \mathbf{u}_i^s
\end{cases}
$$

The error terms of the econometric equation systems are then provided by:

$$
\begin{cases}
\mathbf{u}_i^y & = \mathbf{e}_i^y + \boldsymbol{\epsilon}_i^y \\
u_i^x & = \mathbf{s}_i [\mathbf{e}_i^x + \boldsymbol{\epsilon}_i^x] + \eta_i \\
\mathbf{u}_i^s & = \Delta' \mathbf{M}_i \mathbf{e}_i - \mathbf{e}_i^g
\end{cases}
$$

Matrix notations

$$\mathbf{a}_0^y = \begin{pmatrix} \alpha_{01} - 0.5\gamma_1(w/p_1)^2 \\ \vdots \\ \alpha_{0C} - 0.5\gamma_C(w/p_C)^2 \end{pmatrix} \text{ and } \mathbf{a}_0^x = \begin{pmatrix} \beta_{01} - \gamma_1(w/p_1) \\ \vdots \\ \beta_{0C} - \gamma_C(w/p_C) \end{pmatrix}$$

$$\mathbf{B}_0^y = \begin{pmatrix} \alpha_{11} & & 0 \\ & \ddots & \\ 0 & & \alpha_{1C} \end{pmatrix} \text{ and } \mathbf{B}_0^x = \begin{pmatrix} \beta_{11} & & 0 \\ & \ddots & \\ 0 & & \beta_{1C} \end{pmatrix}$$

$$\mathbf{M}_i = \begin{pmatrix} p_1 & & 0 & -w & & 0 \\ & \ddots & & & \ddots & \\ 0 & & p_C & 0 & & -w \end{pmatrix} \text{ and } \Delta'\mathbf{M}_i = \begin{pmatrix} p_1 & & 0 & -p_C & -w & & 0 & w \\ & \ddots & & \vdots & & \ddots & & \vdots \\ 0 & & p_{C-1} & -p_C & 0 & & -w & w \end{pmatrix}$$

$$\mathbf{Q}_0 = \begin{pmatrix} (p_1\alpha_{11} - w\beta_{11}) + (p_C\alpha_{1C} - w\beta_{1C}) & & (p_C\alpha_{1C} - w\beta_{1C}) \\ & \ddots & \\ (p_C\alpha_{1C} - w\beta_{1C}) & & (p_{C-1}\alpha_{1C-1} - w\beta_{1C-1}) + (p_C\alpha_{1C} - w\beta_{1C}) \end{pmatrix}$$

$$- \begin{pmatrix} g_{11} + g_{CC} - 2g_{1C} & & g_{1C-1} + g_{CC} - g_{1C} - g_{CC-1} \\ & \ddots & \\ g_{1C-1} + g_{CC} - g_{1C} - g_{CC-1} & & g_{C-1C-1} + g_{CC} - 2g_{C-1C} \end{pmatrix}$$

$$\mathbf{F}_{Ci} = \begin{pmatrix} 0.5[p_C\alpha_{1C} - w\beta_{1C}] + g_{1C} + g_{CC} \\ \vdots \\ 0.5[p_C\alpha_{1C} - w\beta_{1C}] + g_{C-1C} + g_{CC} \end{pmatrix} \text{ and } \mathbf{g}_0 = \begin{pmatrix} g_{01} - g_{0C} \\ \vdots \\ g_{0C-1} - g_{0C} \end{pmatrix}$$

Appendix B

The three-stage procedure

In the first stage the equation system composed of the yield supply and acreage choice equations is estimated using the GMM estimator. The objective is to construct a consistent estimator of all identifiable parameters θ.

$$\begin{cases} \mathbf{y}_i & = \mathbf{a}_0^y + \mathbf{B}_0^y \mathbf{s}_i + \mathbf{u}_i^y \\ \mathbf{Q}_0 & \mathbf{s}_i^- + [\Delta' \mathbf{M}_i \mathbf{a}_0 - \mathbf{g}_0 - \mathbf{F}_{Ci}] = \mathbf{u}_i^s \end{cases}$$

In the second stage, these estimators $\hat{\theta}$ are assumed to be available to construct a consistent estimator of a useful part of the variance-covariance matrix $\mathbf{\Psi}$. This stage relies on the second order moment conditions and uses a SUR system linear in its parameters:

$$\begin{cases} \mathbf{u}_i^s(\hat{\boldsymbol{\theta}})\mathbf{u}_i^s(\hat{\boldsymbol{\theta}}) & = \Delta' \mathbf{M}_i \mathbf{\Psi}_{zz} \mathbf{M}_i' \Delta + \mathbf{\Psi}_{gg} + \boldsymbol{\zeta}_i^{ss} \\ \mathbf{u}_i^y(\hat{\boldsymbol{\theta}})\mathbf{u}_i^s(\hat{\boldsymbol{\theta}}) & = \mathbf{\Psi}_{yz} \mathbf{M}_i' \Delta + \boldsymbol{\zeta}_i^{ys} \end{cases}$$

with $E[\boldsymbol{\zeta}_i^{ss}|\Delta \mathbf{M}_i'] = 0$ and $E[\boldsymbol{\zeta}_i^{ys}|\Delta \mathbf{M}_i'] = 0$. The estimates of variance-covariance matrix of the error terms $\mathbf{u}_i^s(\hat{\boldsymbol{\theta}})$ are used to construct the control functions which have the following form $\mathbf{\Psi}_{zz}\mathbf{C}_i^z(\hat{\mathbf{\Psi}})\mathbf{u}_i^s(\hat{\boldsymbol{\theta}})$ with $\mathbf{C}_i = \mathbf{M}_i' \Delta \hat{\mathbf{\Psi}}_s^{-1}$. The third stage of the procedure considers the estimation of the complete system using the GMM estimator:

$$\begin{cases} \mathbf{y}_i & = \mathbf{a}_0^y + \mathbf{B}_0^y \mathbf{s}_i + \mathbf{\Psi}_{yz} \mathbf{C}_i^y(\hat{\mathbf{\Psi}})\mathbf{u}_i^s(\hat{\boldsymbol{\theta}}) + \boldsymbol{\mu}_i^y \\ X_i & = \mathbf{s}_i'[\mathbf{a}_0^x] + \mathbf{\Psi}_{xz} \mathbf{C}_i^x(\hat{\mathbf{\Psi}})\mathbf{u}_i^s(\hat{\theta}) + \mu_i^x \\ \mathbf{Q}_0 & \mathbf{s}_i^- + [\Delta' \mathbf{M}_i \mathbf{a}_0 - \mathbf{g}_0 - \mathbf{F}_{Ci}] = \mathbf{u}_i^s \end{cases}$$

with $E[\boldsymbol{\mu}_i^y|\mathbf{s}_i, \mathbf{z}_i] = 0$, $E[\mu_i^x|\mathbf{s}_i, \mathbf{z}_i] = 0$ and $E[\mathbf{u}_i^s|\mathbf{z}_i] = 0$.

Appendix C

Table 1. Estimates for yield supply, input demand and acreage shares equations (model 1)

	wheat	others cereals	oilseeds/protein crops
The R^2 criterion			
Yield supply equations	0.31	0.28	0.16
Input demand equation	0.33	-	-
Acreage shares equations	0.12	0.22	-
Yield supply equations			
Price effects γ_c	3.32***	2.45***	3.76***
Average potential yield α_{0c}	9.45***	9.98***	8.63***
Constant	8.44***	9.71***	9.84***
Trend	0.35***	0.33***	0.04
Squared trend	−0.02***	−0.01***	−0.007***
Sugar beets past acreage	4.24***	3.28***	−2.24***
Potatoes past acreage	1.78***	1.02*	−0.03
Composition aggregate	−	−1.31***	0.06
Acreage share α_{1c}	0.14	−3.45***	−3.95***
Input demand equation			
Required input β_{0c}	7.62***	7.04***	5.29***
Constant	7.33***	7.09***	4.30***
Trend	0.04*	−0.007	0.14***
Sugar beets		7.40***	
Potatoes		14.27***	
Fodder crops		2.51***	
Acreage shares equations			
Fixed costs $g_{0c} - g_{0C}$	−1.98***	−0.35***	-
Constant	−1.98***	−3.28***	-
Capital	< -0.001***	< -0.001***	-
Labor	−0.23***	−0.09	-
Diversification terms Q_c	−20.87***	−32.22***	−16.47***

Note: *, **, *** denote parameter estimates statistically different from 0 at respectively 10%, 5% and 1% confidence levels.

Table 2. Estimates for yield supply, input demand and acreage shares equations (model 2)

	wheat	others cereals	oilseeds/protein crops
The R^2 criterion			
Yield supply equations	0.21	0.14	0.15
Input demand equation	0.33	-	-
Acreage shares equations	0.18	0.22	-
Yield supply equations			
Price effects γ_c	3.12***	2.54***	3.72***
Average potential yield α_{0c}	11.08***	10.47***	8.43***
Constant	10.36***	10.13***	9.31***
Trend	0.30***	0.34***	0.06*
Squared trend	-0.01***	-0.01***	-0.01*
Sugar beets past acreage	5.47***	3.20***	-1.25
Potatoes past acreage	1.91***	0.75	0.11
Composition aggregate	–	-1.39***	0.05
Acreage share α_{1c}	-3.03***	-5.90***	-3.15***
Input demand equation			
Required input β_{0c}	7.26***	7.34***	5.39***
Constant	6.88***	7.36***	4.61***
Trend	0.05*	-0.003	0.11***
Sugar beets		7.73***	
Potatoes		14.34***	
Fodder crops		2.58***	
Acreage shares equations			
Fixed costs $g_{0c} - g_{0C}$	-0.31	-3.33***	-
Constant	0.40	-2.78***	-
Capital	< -0.001***	< -0.001***	-
Labor	-0.27***	-0.11***	-
Diversification terms Q_c	-24.32***	-30.62***	-16.03***

Note: *, **, *** denote parameter estimates statistically different from 0 at respectively 10%, 5% and 1% confidence levels.

Table 3. Average allocations of variable inputs between crops (in € per are)

	wheat	others cereals	oilseeds/protein crops
model 1	4.47 (0.32)	4.59 (0.41)	1.77 (0.51)
model 2	4.29 (0.25)	4.80 (0.41)	1.91 (0.44)

Note : Standard errors are in parentheses.

Table 4. Statistics according the level of differences in input quantities for OPC

	Number of observations	Subsidies for OPC	Acreage share of protein crops in OPC	Period
$<$ €0.1 per are	600	€1.02 per are	22%	2006-2007
€0.1 to €1 per are	1172	€3.27 per are	31%	2003-2006
€1 to €2 per are	1526	€3.54 per are	31%	1999-2002
$>$ €2 per are	1655	€4.71 per are	48%	1995-1999

Note : OPC corresponds to oilseeds and protein crops.

Notes

[1]See for example the behavioural model of Just et al. (1990) and the vast majority of the related literature.

[2]Surry and Peeters (2001) consider a similar equation system but exploit the flexibility of the Maximum Entropy (GME) statistical framework to compute crop input use estimates per farm. The ME framework also permits to easily impose positivity constraints on the input allocation and to make use of information provided by extension services.

[3]Acreage shares are regressed on all exogenous explanatory variables of the model. The predictions of these acreage shares are then used to construct the instruments for the s_i terms in yield supply equations.

[4]All farmers of the sample cultivate wheat, other cereals, and oilseeds and/or protein crop.

[5]one are=one hundred square metres