ENDOGENEITY OF ACREAGE CHOICES IN INPUT ALLOCATION EQUATIONS:

IMPLIED PROBLEMS AND A SOLUTION

Elodie Letort, Alain Carpentier

INRA-UMR SMART, Rennes, France.

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

Introduction

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

A large number of authors have been working on this topic, either to provide solutions for allocating input costs between crops 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); or because this was a necessary step of their analysis (for example 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 with constraints on variable input total uses (Dixon and *al.*, 1984; Hornbaker and *al.*, 1989; Just and *al.*, 1990). However allocation of variable inputs among crops depends on how the farmers allocate land among crops, a decision which itself takes into account input uses by crop. Crop input use decisions and acreage choices are partially simultaneous. The underlying idea is that variable input allocation requires the specification of a complete production model, *i.e.* describing land allocation, use of variable inputs and crop yields in order to take into account the link between the acreage and the input use choices.

The contribution of this article is threefold. First, it shows 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 the farmers' choices. The specified model distinguishes two sorts of error terms: the terms accounting for farms' heterogeneity and the terms accounting for the stochastic events affecting crop production. It provides explicit functional forms of the links between the error terms of the yield supply, input demand allocation and acreage equations. Third, it proposes a method based on control functions to eliminate the 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 » (see, *e.g.*, Imbens and Wooldridge, 2007; Wooldridge, 2008) and « average treatment effects » (see, *e.g.*, Heckman and al., 2003). The empirical implementation of the proposed methods is described in three stages and an application is presented on French farm-level data.

This paper proceeds as follows. The next section presents a review of the literature about input allocation method and presents briefly the endogeneity problems in these standard approaches and the solution adopted in this paper, *i.e.* the control function based approach. It requires an econometric multi-crop (for acreage, yield and input choices) model which is described in the second section. The third section presents the control functions 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 on French farm-level data are proposed. The last section of this paper provides some concluding remarks.

1. Literature review

The most common farm data on crop production consist in acreages, yields and prices at the crop level, and variable input uses and quasi-fixed factor quantities (measures of labour and capital) at the farm level. Input price indices are generally made available by the national departments of agriculture at the regional level. Farmer i ($i=1,...,N$) produces C crops ($c=1,...,C$) to which they allocate their 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 for solving this allocation problem. We distinguish two main groups in the literature: the first group includes approaches that consider solely input allocation equation(s) as the one defined above. In these models, input allocations are treated as parameters to be estimated, along the lines of Just, Zilberman, Hochman and Bar-Shira (1990) terminology. These are, by far, the most widely used in practice. In the second group, input allocation equations belong to a system of equations including 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 limits. These limits provide arguments for using the approaches of the second type.

1.1. 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:

(1) $$x_i = \sum_c^C s_{ci} x_{ci} + \eta_i \text{ with } E[\eta_i / \mathbf{s}_i] = 0,$$

or as parametric functions:

(2) $$x_i = \sum_c^C s_{ci} x_{ci}(\mathbf{z}_i; \mathbf{a}) + \eta_i \text{ with } E[\eta_i / \mathbf{s}_i, \mathbf{z}_i] = 0,$$

where $\mathbf{z}_i$ is the vector of exogenous variables such as farm's characteristics and activities, **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 **a** under the assumption that the conditional expectation of η_i is zero. See for example the behavioural model of Just et al. (1990) and the vast majority of the related literature.

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:

(3) $$x_i = \sum_c^C s_{ci} \left[x_{ci} + u_{ci}^x \right] + \eta_i \text{ with } E[\eta_i / \mathbf{s}_i] = E\left[u_{ci}^x / \mathbf{s}_i \right] = 0,$$

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

where η_i terms include measurement errors, 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 what can be « explained » by the $\mathbf{z}_i$ variables. Models (3) and (4) are input allocation equations with random parameters. In these models, the error terms, $\sum_c^C s_{ci} u_{ci}^x + \eta_i$ are heteroskedastic, and feasible generalized OLS or SUR estimations will provide efficient estimators of the parameter vector **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).[1]

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 **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.:

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

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 conditions:

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

---

1 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.

hold if and only if $u_{ci} = 0$, 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 represents 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.

1.2. Approaches based on multicrop econometric models

We discuss here 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 (*e.g.*, Ozarem and Miranowski, 1994), risk aversion (*e.g.*, Coyle, 1992, 1999 ; Chavas and Holt, 1990) and price uncertainty (*e.g.*, Coyle, 1992, 1999 ; Moro and Sckokai, 2006) or models based on plot per plot discrete choice (*e.g.*, Wu and Segerson, 1995) are not considered here. Also, we focus on models in which land is considered as an allocatable fixed input (Shumway, Pope and Nash, 1984), *i.e.*, models designed for analyzing farmers' short run decisions.

In papers falling into this category, the problem of variable input allocation is considered as a by product or not considered in further details. Lence and Miller (1998), in a Maximum Entropy framework, estimate jointly crop production function models and crop input uses. Their use of the flexible maximum entropy estimators enables them to allocate the farm input uses by using a system of production function models (one for each crop) and constraining the crop input uses to sum to the farms' input uses. This approach ensures the consistency of the determined input allocation with a system of production functions. Note also that this approach does not rely on the modelling of farmers' economic choices. In this respect, Lence and Miller's (1998) approach lies in between the « atheoretical » approach of Dixon *et al.* (1984), Hornbaker *et al.* (1989) and the approach based on the specification of a complete model of farmers' choices. They use production functions but they don't use farmers' production choice models. But, as acknowledged by Lence and Miller, their approach as well as the other « atheoretical » approaches share the same drawback: they do not consider input uses and acreages (or production levels in Lence and Miller's approach) as (partly) simultaneous choices.

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 consider variable input allocation issues although these are not the main focus of their studies. They employ an approach similar to the one used here in the sense that they determine the variable input allocation by considering a complete model of farmers' choices. The variable input allocation is merely only a by-product of their modelling exercise. Nevertheless their multicrop econometric model differs from ours in important respects. First, their use of Cobb-Douglas crop yield functions (Just *et al.* 1983) or Translog crop profit (Chambers and Just 1989) functions facilitates their determination of the variable input allocations. Second, their multicrop econometric models is consistent in their deterministic part but they are not consistent in their random parts. Their econometric models are derived from their economic models basically by adding error terms to the deterministic equations derived from the economic model although Just *et al.* (1983) added random terms with structural interpretations. These points are further detailed below. But it is worth noting that the approach considered here builds on the two critics of the previous approaches presented above.

Acreage allocation models considered in the 1990's mostly use the model designed by Moore and Negri (1992) (see *e.g.* Moore, Gollehon and Carey, 1994 ; Moore and Dinar, 1995 ; Guyomard, Baudry and Carpentier, 1996; Oude Lansink and Peerlings, 1996; Bel Haj Hassine and Simioni, 2000; Bel, Lacroix, Salanié 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 where restricted Translog profit functions are replaced by restricted Normalized Quadratic Profit functions at the crop level. Function of profit, production and input demand functions in levels are much easier to use than their counterparts in logarithm because the total profit and the land constraint are defined as sums of the crop profits and the crop acreage levels (or shares). Variable input uses are usually considered at the farm level in most of studies employing multi-crop econometric models (Paris, 1989).

1.3. Outline of the control function approach

The starting point of this research is that the exogeneity conditions $E\left[u_{ci}^{x}/\mathbf{z}_i,\mathbf{s}_i\right]=0$ 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 (VI) techniques, because the error term $s_{ci}u_{ci}^x + \eta_i$ contains the endogenous explanatory variables $\mathbf{s}_i$. The use of equation (1) as an estimating equation requires the control of the terms $E\left[u_{ci}^x / \mathbf{z}_i, \mathbf{s}_i\right]$.

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 assume that the considered model allows to define the $E\left[u_{ci}^x / \mathbf{z}_i, \mathbf{s}_i\right]$ terms known functions of $\mathbf{z}_i$, $\mathbf{s}_i$ and of a vector of unknown parameters $\boldsymbol{\theta}$. Let assume also that there exists a consistent estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$. The input allocation equation (1) can be transformed as:

(7) $$x_i = \sum_c^C s_{ci}\left[\, x_{ci}(\mathbf{z}_i ; \mathbf{a}) + c_c^x(\mathbf{z}_i, \mathbf{s}_i ; \boldsymbol{\theta}) \,\right] + \omega_i^x ,$$

(8) $$\text{with } \omega_i^x = \sum_c^C s_{ci}\left[\, u_{ci}^x(\mathbf{z}_i ; \mathbf{a}) - c_{ci}^x(\mathbf{z}_i, \mathbf{s}_i ; \boldsymbol{\theta}) \,\right] + \eta_i ,$$

where $c_{ci}(\mathbf{z}_i, \mathbf{s}_i ; \boldsymbol{\theta})$ 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 $c_{ci}(\mathbf{z}_i, \mathbf{s}_i ; \hat{\boldsymbol{\theta}})$ terms are consistent estimators of the corresponding $c_{ci}(\mathbf{z}_i, \mathbf{s}_i ; \boldsymbol{\theta})$ terms, equation (7) can be used to construct consistent regression based estimators of **a**. The control function approach basically splits the error terms u_{ci}^x in two terms: the control function $c_{ci}(\mathbf{z}_i, \mathbf{s}_i ; \boldsymbol{\theta}) = E\left[u_{ci}^x / \mathbf{z}_i, \mathbf{s}_i\right]$ which « captures » and thus controls the links between u_{ci}^x and the endogenous variable vector $\mathbf{s}_i$; and a « new » error term $u_{ci}^{x}(\mathbf{z}_i;\mathbf{a})-c_{ci}^{x}(\mathbf{z}_i,\mathbf{s}_i;\boldsymbol{\theta})$. By construction, $\mathbf{s}_i$ is exogenous with respect to the « new » error term.

The crucial point is then to define the control functions $c_{ci}(\mathbf{z}_i,\mathbf{s}_i;\boldsymbol{\theta})$ 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:

(9) $s_{ci}=s_{ci}(\mathbf{z}_i;\mathbf{b})+\omega_{ci}^{s}$ with $E\left[\omega_{ci}^{s}/\mathbf{z}_i\right]=0$,

The control functions are determined by the following conditional expectations:

(10) $c_{ci}^{x}(\mathbf{z}_i,\mathbf{s}_i;\boldsymbol{\theta})=E\left[u_{ci}^{x}/\mathbf{z}_i,\mathbf{s}_i\right]=E\left[u_{ci}^{x}/\mathbf{z}_i,s_{ci}(\mathbf{z}_i;\mathbf{b})+\omega_{ci}^{s}\right]=E\left[u_{ci}^{x}/\mathbf{z}_i,\omega_{ci}^{s}\right]$,

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 ».

2. Econometric model specification

Although the proposed approach can be applied with other multi-crop models (with more or less adaptations), a specific multi-crop econometric model is considered to « concretely » illustrate the basic features of the approach. This model combined standard quadratic yield functions with crop acreage (share) functions derived along the line of Heckeleï and Wolff (2003). The model considered is the one described in chapter two. It is chosen because of its fairly simple interpretation and its flexibility. The error terms of the econometric model are defined as integral parts of this model (see, *e.g.*, McElroy, 1987).

The model is considered in its simplest version, *i.e.* with constant parameters. In empirical work most of the defined parameters may usefully defined as parametric functions of observed exogenous variable to control (as much as possible) for the heterogeneity of the farms and farmers. Finally the single variable input is considered for simplicity.

2.1. Yield 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:

(11) $$y_{ci} = \alpha_{ci} - 0.5\gamma_c^{-1}(\beta_{ci} - x_{ci})^2 \text{ with } \alpha_{ci} = \alpha_{0c} + 0.5\alpha_{1c}s_c + v_{ci}^y \text{ and } \beta_{ki} = \beta_{0k} + 0.5\beta_{1c}s_c + v_{ci}^x,$$

where x_{ci} is the quantity of variable input used per hectare by farm i devoted to crop c, and α_{ci}, γ_c and β_{ci} are parameters to be estimated with $\alpha_{ci} > 0, \gamma_c > 0$ and $\beta_{ci} > 0$. 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 estimates of these yield functions can thus be checked with agricultural scientists or extension agents. The maximum yield and the input requirement terms are specified as functions of the crop acreage to account for potential scale effects.

The v_{ci}^y and v_{ci}^x are random terms. These terms are split into two parts for simplifying their interpretation:

(12) $v_{ci}^{y} = e_{ci}^{y} + \varepsilon_{ci}^{y}$ and $v_{ci}^{x} = e_{ci}^{x} + \varepsilon_{ci}^{x}$.

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

The production of crop c is sold at price p_{ci} and the input is bought at price w_i by the farmer i. These prices are assumed to be known at the beginning of the production process, *i.e.* when acreages are chosen. Farmers are supposed risk-neutral. Farmer i is assumed to choose his input use by maximizing the following gross margins π_{ci} for each crop c :

(13) $p_{ci} y_{ci} - w_i x_{ci}$.

Variable input and "target" yields choices are assumed based on output and input prices and adjusted to specific production condition, *i.e.* after farmer has observed ε_{ci}^{y} and ε_{ci}^{x}. The maximisation of this profit function under technological constraints leads to the following per hectare variable input demand and supply functions:

(14) $x_{ci} = \beta_{0c} - \gamma_c (w_i / p_{ci}) + v_{ci}^x$

(15) $y_{ci} = \alpha_{0c} - 0.5\gamma_c (w_i / p_{ci})^2 + v_{ci}^y$.

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 fully undergone production conditions. The quadratic yield have a main practical advantage: they provide 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, *e.g.*, McElroy, 1987, and Pope and Just, 2003, in other contexts). Distinguishing the heterogeneity effects and the stochastic events in the yield function allows to determine the gross margins of the crops as they are expected by the farmers at the time they choose their acreages:

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

The farmers' gross margin expectations can not depend on the ε_{ci}^y and ε_{ci}^x terms because these terms are unknown when farmers choose their acreages.

The system of yield supply, input demand and expected gross margin functions can be defined with simple matrix notations (see Appendix A):

(17) $\mathbf{y}_i = \mathbf{a}_0^y + 0.5\mathbf{B}_0^y \mathbf{s}_i + \mathbf{v}_i^y$

(18) $\mathbf{x}_i = \mathbf{a}_0^x + 0.5\mathbf{B}_0^x \mathbf{s}_i + \mathbf{v}_i^x$

(19) $\pi_i^e = \mathbf{M}_i \left[\mathbf{a}_0 + 0.5\mathbf{B}_0 \mathbf{s}_i + \mathbf{e}_i^z \right]$

where $\mathbf{a}_0 \equiv (\mathbf{a}_0^y, \mathbf{a}_0^x)$, $\mathbf{B}_0 \equiv (\mathbf{B}_0^y, \mathbf{B}_0^x)$ and $\mathbf{e}_i^z \equiv (\mathbf{e}_i^y, \mathbf{e}_i^x)$. The terms indexed by 0 are parameters to be estimated. Parameters $\mathbf{a}_0^y$ and $\mathbf{a}_0^x$ can be defined as functions of the matrix $\mathbf{z}_i$, which denotes the output and input prices. Parameters $\mathbf{B}_0^y$ et $\mathbf{B}_0^x$ are matrix of scale effects. Matrix $\mathbf{M}_i$ is defined such as $\mathbf{M}_i \mathbf{e}_i^z = \mathbf{P}_i{}' \mathbf{e}_i^y - w_i \mathbf{e}_i^x$ with $\mathbf{P}_i$ the matrix of outputs prices.

2.2. Acreage functions

Farmers' acreage choices are modelled within the framework developed by Heckeleï and Wolff (2003). This framework is simple, flexible and links the econometric and mathematical programming literature on production choice modelling. Farmer *i* is assumed to allocate his total land quantity S_i by maximizing the following indirect restricted profit function:

(20) $$\sum_{c=1}^{C} s_{ci}\pi_{ci}^{e} - C(\mathbf{s}_i),$$

where $\mathbf{s}_i$ is the vector of acreage share for farmer *i*. According to this model, farmers have two motives for crop diversification: the scale effects of the crop gross margins $0.5\left(p_{ki}\alpha_{1k} - w_i\beta_{1k}\right)s_{ki}$ and the implicit management cost of the chosen acreage $C(s_k)$. This cost function is increasing and quasi-convex in $\mathbf{s}_i$. It is used in the mathematical programming literature. It can be interpreted as a reduced form function smoothly approximating the unmeasured and implicit costs associated with a given land allocation and farm specific constraints. This cost function is assumed to have this form:

(21) $$C(\mathbf{s}_i) = a_i + \sum_{c=1}^{C} s_{ci} g_{ci} + 0.5 \sum_{c=1}^{C} \sum_{m=1}^{C} g_{cm} s_{ci} s_{mi} \quad \text{with} \quad g_{ci} = g_{0c} + e_{ci}^{g},$$

where a_i , g_{ki} and g_{kmi} are parameters to be estimated. The “fixed” cost g_{ki} per unit of land of crop *k* of farmer *i* is split into two parts g_{0k} a parameter and e^g_{ki} a random term accounting for the cost heterogeneity term known to farmer *i* but unknown to the econometrician.

The restricted indirect profit function is rewritten in including the land use constraint $1 = \sum_{k=1}^{K} s_{ki}$ and with matrix notations for simplicity:

(22) $$\mathbf{f}_{1i} + \mathbf{s}_i^-{}' \left[\Delta' \mathbf{M}_i \left[\mathbf{a}_0 + \mathbf{e}_i^z \right] - \left[\mathbf{g}_0 + \mathbf{e}_i^g \right] - \mathbf{F}_{1i} \right] + 0.5 \mathbf{s}_i^-{}' \mathbf{Q}_0 \, \mathbf{s}_i^-$$

where $\mathbf{s}^-$ is the acreage vector of dimension $C-1$ and the matrix Δ is a differentiation matrix such that $\Delta' q$ is equal to a column vector of dimension $C-1$ and its elements are defined as $q_c - q_1$ for $c = 1,...,C$. The crop 1 is the “reference” crop. The vector $\mathbf{g}_0$ denotes the g_{0c} parameters for $c = 2,...,C$ and the vector $\mathbf{e}_i^g$ contains error terms e^g_{ci} for $c = 2,...,C$. The term $\mathbf{Q}_0$ is defined by $\mathbf{A}(\mathbf{b}_0) - \Delta' \mathbf{G}_0 \Delta$. The matrix $\mathbf{G}_0$ contains parameters g_{cm} and the matrix $\mathbf{A}(\mathbf{b}_0)$ is a function of parameters β_{1c}. The term $\mathbf{F}_{1i}$ contains the effects on the acreage choices of the gross margins scale effects. All these terms are defined in the appendix A. The restricted indirect profit function is strictly concave in *s* if the quadratic form $\mathbf{s}_i^-{}' \mathbf{Q}_0 \, \mathbf{s}_i^-$ is semi-definite negative.

All crops are assumed to be cultivated. The maximisation in $\mathbf{s}^-$ of this restricted indirect profit function leads to the closed form of the acreage functions:

(23) $$\mathbf{s}_i^- = -\mathbf{Q}_0^{-1} \left[\Delta' \mathbf{M}_i \, \mathbf{a}_0 - \mathbf{g}_0 - \mathbf{F}_{1i} \right] + \mathbf{v}_i^s \text{ with } \mathbf{v}_i^s = -\mathbf{Q}_0^{-1} \left[\Delta' \mathbf{M}_i \, \mathbf{e}_i^z - \mathbf{e}_i^g \right]$$

These closer forms show that the 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 $\mathbf{e}_i^x$ and $\mathbf{e}_i^y$.

2.3. "Complete" multi-crop econometric model

The multi-crop econometric model is composed of three subsets of equations, yield equations, acreage equations and an input allocation equation. The total variable input X is define as the sum of the acreage share devoted to each crop *c* multiplied by the per hectare variable input quantity used for each crop k : $X_i = \sum_{c=1}^{C} s_{ci} x_{ci}$. This input variable allocation equation takes part into the econometric model, which is defined as:

$$(24) \qquad \begin{cases} \mathbf{y}_i = \mathbf{a}_0^y + 0.5\,\mathbf{B}_0^y \mathbf{s}_i + \boldsymbol{\omega}_i^y \\ X_i = \mathbf{s}_i\,'\left[\mathbf{a}_0^x + 0.5\,\mathbf{B}_0^x \mathbf{s}_i\right] + \boldsymbol{\omega}_i^x \\ \mathbf{s}_i^- = -\mathbf{Q}_0^{-1}\left[\Delta'\mathbf{M}_i\,\mathbf{a}_0 - \mathbf{g}_0 - \mathbf{F}_{1i}\right] + \boldsymbol{\omega}_i^s \end{cases}$$

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

$$(25) \qquad \begin{cases} \boldsymbol{\omega}_i^y = \mathbf{v}_i^y = \mathbf{e}_i^y + \boldsymbol{\varepsilon}_i^y \\ \boldsymbol{\omega}_i^x = \mathbf{s}_i\,'\mathbf{e}_i^x + \mathbf{s}\,'\boldsymbol{\varepsilon}_i^x + \eta_i \\ \boldsymbol{\omega}_i^s = \mathbf{v}_i^s = -\mathbf{Q}_0^{-1}\left[\Delta'\mathbf{M}_i\,\mathbf{e}_i^z - \mathbf{e}_i^g\right] \end{cases}$$

with $\Delta'\mathbf{M}_i\mathbf{e}_i^z - \mathbf{e}_i^g = \boldsymbol{\omega}_i^\pi$. An error term η_i is added in the input allocation equation and represents the effects of measurement errors due, *e.g.*, to stock variations. We denote $\mathbf{z}_i$ the output and input prices for farmer *i*.

The preceding interpretations of the error terms allow to define the following mean assumptions: $E\left[\boldsymbol{\omega}_i^y / \mathbf{z}_i\right] = 0$, $E\left[\boldsymbol{\varepsilon}_i / \mathbf{z}_i\right] = 0$, $E\left[\mathbf{e}_i^g / \mathbf{z}_i\right] = 0$, $E\left[\eta_i / \mathbf{z}_i\right] = 0$ et $E\left[\mathbf{s}_i\,'\boldsymbol{\varepsilon}_i^x / \mathbf{z}_i\right] = 0$.

This implies that each component of $\boldsymbol{\omega}_i$ has a null expectation conditionally on prices excepted the $\mathbf{s}_i'\mathbf{e}_i^x$ term in the input allocation equation. $\mathbf{s}_i$ is an endogenous explanatory variable but this is a standard problem that can be worked out with standard instrumental variable techniques. The main problem is that $E\left[\mathbf{s}_i'\mathbf{e}_i^x / \mathbf{z}_i\right] \neq 0$ or $E\left[\mathbf{e}_i^x / \mathbf{z}_i, \mathbf{s}_i\right] \neq 0$. These terms need thus to be determined. Before proceeding to the determination of the control functions two remarks are in order. First, the yield supply and the acreage choice functions identify almost the entire set of parameters. Only the term a_{0K}^x can not be identified. Second, the heterogeneity terms $\mathbf{e}_i \equiv (\mathbf{e}_i^y, \mathbf{e}_i^x, \mathbf{e}_i^g)$ are the « interest error terms » for determining the control functions while $\boldsymbol{\varepsilon}_i$ and η_i can be viewed as « disturbances ».

3. Control function approach

The econometric model considered is fully consistent, *i.e.* consistent with respect to its deterministic parts and with respect to its error terms. It provides thus explicit forms of the relationship between the error term vectors of the yield supply, input demand allocation and acreage equations. The main problem is the link between the acreage and the input use choices in the variable input allocation equation. The control function idea is to determine explicitly this link and its estimator, and integrate this term in the fully 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 $E\left[\mathbf{e}_i^x / \mathbf{z}_i, \mathbf{s}_i\right]$ terms. Another approach would be based on the functional form of the $E\left[\mathbf{s}_i'\mathbf{e}_i^x / \mathbf{z}_i\right]$ terms. This second approach relies on less restrictive assumptions but requires more involved computations. Wooldridge (2008) distinguishes both approaches, denoting the functional form of $E\left[\mathbf{e}_i^x / \mathbf{z}_i, \mathbf{s}_i\right]$ by the usual term « control function » and denoting the functional form of $E\left[\mathbf{s}_i '\mathbf{e}_i^x / \mathbf{z}_i\right]$ by the term « correction function ».

The construction of control functions relies on two main approaches: the use of distributional assumptions with respect to the error terms and/or the use of the linear projection techniques (see, e.g., Chamberlain, 1982; Wooldridge, 2004). It is shown that distributional assumptions are generally necessary to define control functions for the general multi-crop econometric model (see, *e.g.*, 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.

Both types of approach rely on the additional conditional mean and homoskedasticity assumptions: $E\left[\mathbf{e}_i / \mathbf{z}_i\right] = 0$, $V\left[\mathbf{e}_i / \mathbf{z}_i\right] = \mathbf{\Psi}$ and $E\left[\boldsymbol{\varepsilon}_i / \mathbf{z}_i, \mathbf{e}_i\right] = 0$. 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:

$$(26) \qquad \mathbf{\Psi} = \begin{bmatrix} \mathbf{\Psi}_{zz} & 0 \\ 0 & \mathbf{\Psi}_{gg} \end{bmatrix} = \begin{bmatrix} \mathbf{\Psi}_{yz} & 0 \\ \mathbf{\Psi}_{xz} & 0 \\ 0 & \mathbf{\Psi}_{gg} \end{bmatrix} = \begin{bmatrix} \mathbf{\Psi}_{yy} & \mathbf{\Psi}_{yx} & 0 \\ \mathbf{\Psi}_{xy} & \mathbf{\Psi}_{xx} & 0 \\ 0 & 0 & \mathbf{\Psi}_{gg} \end{bmatrix}$$

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:

(27a) $V\left[\boldsymbol{\omega}_i^{\pi}\boldsymbol{\omega}_i^{\pi}{}'/\mathbf{z}_i\right]=\Delta'\mathbf{M}_i\boldsymbol{\Psi}_{zz}\mathbf{M}_i{}'\Delta+\boldsymbol{\Psi}_{gg}$

(27b) $V\left[\boldsymbol{\omega}_i^{y}\boldsymbol{\omega}_i^{\pi}{}'/\mathbf{z}_i\right]=-\boldsymbol{\Psi}_{yz}\mathbf{M}_i{}'\Delta$

(27c) $V\left[\boldsymbol{\omega}_i^{x}\boldsymbol{\omega}_i^{\pi}{}'/\mathbf{z}_i\right]=-\boldsymbol{\Psi}_{xz}\mathbf{M}_i{}'\Delta$

with $\boldsymbol{\omega}_i^{\pi}=-\mathbf{Q}_0\,\boldsymbol{\omega}_i^{s}$. These moment conditions can be used to define regression estimators of the useful parts of the variance-covariance matrix $\boldsymbol{\Psi}$ (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 (see, *e.g.*, 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 $\boldsymbol{\Psi}$. Since all the considered error terms of the model $\boldsymbol{\omega}_i^{y}$, $\boldsymbol{\omega}_i^{\pi}$, $\boldsymbol{\omega}_i^{x}$ and $\boldsymbol{\omega}_i^{s}$ are linear transformations of $\mathbf{e}_i$, they are also normally distributed.

The control functions defined here seek to solve two problems: the non null expectation of $\mathbf{s}_i{}'\mathbf{e}_i^{x}$ 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 $\boldsymbol{\omega}_i^{x}$ 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:

(28) $E\left[\boldsymbol{\omega}_i^x / \mathbf{z}_i, \mathbf{s}_i\right] = \mathbf{s}_i ' E\left[\mathbf{e}_i^x / \mathbf{z}_i, \mathbf{s}_i\right]$.

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

(29) $E\left[\mathbf{e}_i^x / \mathbf{z}_i, \mathbf{s}_i\right] = \boldsymbol{\Psi}_{xz} \mathbf{C}_i\left(\boldsymbol{\Psi}\right) \boldsymbol{\omega}_i^\pi$ and $E\left[\mathbf{e}_i^y / \mathbf{z}_i, \mathbf{s}_i\right] = \boldsymbol{\Psi}_{yz} \mathbf{C}_i\left(\boldsymbol{\Psi}\right) \boldsymbol{\omega}_i^\pi$

where $\mathbf{C}_i = -\mathbf{M}_i ' \Delta\left(\Delta ' \mathbf{M}_i \boldsymbol{\Psi}_{zz} \mathbf{M}_i ' \Delta + \boldsymbol{\Psi}_{gg}\right)^{-1}$. These functions can be used as control functions in the yield supply and input demand allocation equations.

4. Implementation: a three-stage procedure

This section considers the implementation of the control function approach in the general case. It presents a simple three-stage inference procedure. 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.

4.1 A three-stage inference procedure

In the first stage the equation system composed of the yield supply and acreage choice equations is estimating. The objective is to construct a consistent estimator of all identifiable parameters $\boldsymbol{\theta}$, *i.e.* all the parameters except B_{01}^x. This system is a simultaneous equation system due to the endogeneity of the acreage choices:

(30) $$\begin{cases} \mathbf{y}_i = \mathbf{a}_0^y + 0.5\,\mathbf{B}_0^y \mathbf{s}_i + \boldsymbol{\omega}_i^y \\ \mathbf{s}_i^- = -\mathbf{Q}_0^{-1}\left[\Delta ' \mathbf{M}_i \mathbf{a}_0 - \mathbf{g}_0 - \mathbf{F}_{1i}\right] + \boldsymbol{\omega}_i^s \end{cases}$$

with $\boldsymbol{\omega}_i^s = -\mathbf{Q}_0^{-1}\,\boldsymbol{\omega}_i^\pi$. The estimation of this equation system used the three-stage least squares (3SLS) or the generalized method of moments (GMM) estimator. Valid instrumental variables are construct for the elements of the system depending on $\mathbf{s}_i$. The efficient instruments of this system are functions of the expectation of $\mathbf{s}_i$ conditional on $\mathbf{z}_i$. These "predictors" $\hat{\mathbf{s}}_i\left(\mathbf{z}_i\right)$ of $\mathbf{s}_i$ are defined according to another simple acreage model.

In the second stage, these estimators $\hat{\boldsymbol{\theta}}$ are assumed to be available for constructing a consistent estimator of an useful part of the variance-covariance matrix $\boldsymbol{\Psi}$. This stage is similar to the second stage of the construction of a standard GLS estimator. It relies on the second order moment conditions and uses a linear in its parameters SUR system:

$$
(31) \qquad \begin{cases} \boldsymbol{\omega}_i^\pi(\hat{\boldsymbol{\theta}})\ \boldsymbol{\omega}_i^\pi(\hat{\boldsymbol{\theta}})' = \Delta'\mathbf{M}_i\boldsymbol{\Psi}_{zz}\mathbf{M}_i{}'\Delta + \boldsymbol{\Psi}_{gg} + \boldsymbol{\xi}_i^{ss} \\ \boldsymbol{\omega}_i^\pi(\hat{\boldsymbol{\theta}})\ \boldsymbol{\omega}_i^y(\hat{\boldsymbol{\theta}})' = \boldsymbol{\Psi}_{yz}\mathbf{M}_i{}'\Delta + \boldsymbol{\xi}_i^{ys} \end{cases}
$$

with $E\left[\boldsymbol{\xi}_i^{ss}/\Delta\mathbf{M}_i^{'}\right]=0$ and $E\left[\boldsymbol{\xi}_i^{ys}/\Delta\mathbf{M}_i^{'}\right]=0$. The estimates of variance-covariance matrix of the error terms $\boldsymbol{\omega}_i^\pi\left(\hat{\boldsymbol{\theta}}\right)$ are used to construct the control functions $\boldsymbol{\Psi}_{yz}\mathbf{C}_i(\hat{\boldsymbol{\Psi}})\boldsymbol{\omega}_i^\pi(\hat{\boldsymbol{\theta}})$ and $\boldsymbol{\Psi}_{xz}\mathbf{C}_i(\hat{\boldsymbol{\Psi}})\boldsymbol{\omega}_i^\pi(\hat{\boldsymbol{\theta}})$.

The third stage of the procedure considers the estimation of the interest parameters $\mathbf{a}_0, \mathbf{B}_0, \mathbf{g}_0, \mathbf{G}_0$ and the auxiliary parameters $\boldsymbol{\Psi}_{yz}$ and $\boldsymbol{\Psi}_{xz}$. The corresponding estimating equation system uses the control functions:

$$
(32) \qquad \begin{cases} \mathbf{y}_i = \mathbf{a}_0^y + 0.5\,\mathbf{b}_0^y\mathbf{s}_i + \boldsymbol{\Psi}_{yz}\mathbf{C}_i^y(\hat{\boldsymbol{\Psi}})\boldsymbol{\omega}_i^\pi(\hat{\boldsymbol{\theta}}) + \boldsymbol{\mu}_i^y \\ X_i = \mathbf{s}_i{}'\left[\mathbf{a}_0^x\right] + \mathbf{s}_i{}'\boldsymbol{\Psi}_{xz}\mathbf{C}_i^x(\hat{\boldsymbol{\Psi}})\boldsymbol{\omega}_i^\pi(\hat{\boldsymbol{\theta}}) + \boldsymbol{\mu}_i^x \\ \mathbf{s}_i^- = -\mathbf{Q}_0{}^{-1}\left[\Delta'\mathbf{M}_i\,\mathbf{a}_0 - \mathbf{g}_0 - \mathbf{F}_{1i}\right] + \boldsymbol{\omega}_i^s \end{cases}
$$

with $E\left[\boldsymbol{\mu}_i^y / \mathbf{s}_i, \mathbf{z}_i\right]=0$, $E\left[\boldsymbol{\mu}_i^x / \mathbf{s}_i, \mathbf{z}_i\right]=0$ and $E\left[\boldsymbol{\omega}_i^y / \mathbf{z}_i\right]=0$. This econometric model is not a standard non linear SUR system for two reasons. First, the yield supply and input allocation equations use $\mathbf{s}_i$ as a regressor whereas $\mathbf{s}_i^-$ is the dependant variable of the acreage equations. Second, the different equations of the system share many parameters. The corresponding SUR estimators are generally non consistent. It is however possible to construct consistent GMM estimator.

A few remarks are in order for the implementation of this final stage. First, sometimes it is impossible or inconvenient to use equations highly non-linear in its parameters. In this case first order conditions equations can replace the acreage equations:

(33) $$\mathbf{Q}_0\,\mathbf{s}_i^- + \left[\Delta'\mathbf{M}\,\mathbf{a}_0 - \mathbf{g}_0 - \mathbf{F}_{1i}\right] + \boldsymbol{\omega}_i^{\pi} = 0$$

One disadvantage of specifying equations in general form (and not in normalized form) is that there are no actual values associated with the equation, so the R^2 statistic cannot be computed. Second, the estimator $\hat{\boldsymbol{\theta}}$ in the first stage provides a useful set of starting values for the empirical implementation of the GMM estimator of the parameters in the third stage. Third, this approach can be interpreted as a generalized version of the "augmented regression" technique controlling 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 the input demand allocation equations. The null hypothesis is then $\boldsymbol{\Psi}_{yz} = \boldsymbol{\Psi}_{xz} = 0$. This is a test of the interest of the approach proposed in this study.

4.2 An empirical application

The three-stage procedure is applied to the French grain crop producer over 1988-2006 using rotating panel data sample of the French Farm Accountancy Data Network (FADN). It contains approximately 6000 observations. The information available is acreage, yield and price for each crop, and variable input expenditures at the farm level. Six different crops group are considered: 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 of sugar beets, potatoes and miscellaneous crops, and fodder crops were considered as exogenous since most of them are contract crops[2]. The different variable inputs (fertilizers, pesticides, energy, 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 € of 2000.

Some variable were introduced in the yield and input use equations to account for technical changes and farms' heterogeneity. In particular, a « production potential index » is included to control for farm heterogeneity. This index is defined by $q_{it} = \left(y_{1i,t-1} - y_{1i,t-1}^{Med} \right) / \left(y_{1i,t-1}^{Max} - y_{1i,t-1}^{Min} \right)$, where $y_{1i,t-1}^{Med}$, $y_{1i,t-1}^{Max}$ and $y_{1i,t-1}^{Min}$ denote, respectively, the median, 99% quantile and 1% quantile of the yield of wheat in the sample year *t-1*. It is based on wheat yields due to the specialization of the sampled farms, and it is defined on a year per year basis to control for year specific conditions. While this index mostly accounts for persistent production conditions, farmers' choices and yields also depend on crop rotation effects. The lagged acreage shares of root crops are introduced to account for the beneficial effects of the induced crop rotations. Since considered crops are aggregated, the acreage share of cereals expect corn (in the total acreage of cereals except wheat) and the acreage share of protein crops (in the total acreage of oilseeds and protein crops) are also introduced.

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

The multi-crop econometric model is estimated following the three-stage procedure described in the last section. Table 1 presents the estimates of yield supply, input demand and acreage shares functions parameters and table 2 presents the price elasticities. These results show that the considered econometric model provides satisfactory econometric modelling frameworks. First, all necessary conditions are respected. The yield functions are concave because γ_k are superior to zero. The quadratic form $\mathbf{s}_i^{-}\,'\mathbf{Q}_0\,\mathbf{s}_i^{-}$ is semi-definite negative because q_1 and q_2 are inferior to zero and the determinant of Q_0 is strictly positive, implying the concavity of the profit function. And the scale effects α_{1k} in the yield functions are negative. This implies that the model is well-behaved. Second, the fit of the model is correct given that data used are at farm level. The R^2 criteria lie between .31 and .42 for yield and input use functions. Almost 95% of the parameter estimates are statistically different from 0 at 5% confidence levels. Third, results are consistent with agronomic principles and the variables used to control farm heterogeneity have expected effects on yield and input requirement. The production potential index has positive effects on yield supply and on input demand. This last result described an intensification process. The lagged acreage shares of potatoes and sugar beets have a positive effect on yield for all the considered crops. The aggregate composition variables show that cereals require less variable input than corn and protein crops less than oilseeds. Fourth, price elasticities are comparable with other studies analysing crop supply response in France and in the European Union. These elasticities are evaluated at the average of the sample. Yields are price inelastic with estimated response in the 0.24-0.36 range. Guyomard and al. (1996) estimated from French aggregated data yield response in the 0.2-0.4 range for cereals and oilseeds. Input use is also quite inelastic with respect to its own-price and crop prices. In acreage share functions, all own-prices elasticities are positive, with other cereals being the most price elastic (1.195). This result is also conforming to other estimations on French data (1.27 for corn estimated by Guyomard and *al.* (1996), and 0.922 for barley used in MECOP, a model of the European Union's producing sector of cereals, oilseeds and protein crops (2001)).

Table 3 presents parameter estimates associated to the control functions. These parameters correspond to the elements of the variance-covariance matrix of model's residuals $\mathbf{\Psi}_{yz}$ and $\mathbf{\Psi}_{xz}$. Variance of yield and input use equations residuals for each crops are positive. Almost 70% of these parameter estimates are statistically different from 0 at 10% confidence levels. The standard test of the hypothesis $\mathbf{\Psi}_{yz} = \mathbf{\Psi}_{xz} = 0$ is a test of the endogeneity of acreage in the yield supply and the input demand allocation equations. A Wald test is conducted to test the null hypothesis that all equal to zero. The null hypothesis is rejected, so it confirms the acreage endogeneity problem and comforts the use of the control function approach.

Conclusion

The contribution of this research is threefold. First it shows that the standard regression based approaches for allocating variable input uses to crops are potentially biased to the (partial) simultaneity of the (expected) crop variable input and acreage choices. Second, it proposes a structural econometric multi-crop model, *i.e.* a model which is consistent in its deterministic and random parts, for determining the origin of these biases and providing potential solutions. Third, it proposes different approaches based on the use of control functions to eliminate these biases. The interest of the empirical application is twofold. First, it shows that the econometric multi-crop model used is well-behaved and provide interesting results. Second, it confirms the acreage endogeneity in yield and input allocation equations and thus, shows the usefulness of the proposed approach.

The proposed approach is described within the context of crop production but could be applied in other contexts where inputs need to be allocated to activities. It could also be applied by using other structural econometrics models with an explicit specification of (deterministic and random) links between production, input uses and activity level choices. Note however that the error term additivity plays a crucial role in the proposed approach.

The proposed approach has potentially three main drawbacks. First, as it is « fully » structural it is thus subject to specification biases. A potential useful extension would replace the structural activity choice model by a more flexible model of the expected gross margin. The second drawback is linked to the first: the econometric model used cannot account for corner solutions of activity choices. This is a potentially important weakness of this framework, particularly in the crop production context. But, the specification of a fully structural model for activity choices with corner solutions is an involved exercise. This highlights the usefulness of « acceptable approximations » to replace a fully structural framework. Third, the identification of the control functions relies on models of the square and cross products of the crop and input prices. As a result, the empirical identification of these functions requires price data at the farm level of good quality.

References

Carpentier, A., et R. D. Weaver. 1996. "Intertemporal and Interfirm Heterogeneity: Implications for Pesticide Productivity." *Canadian Journal of Agricultural Economics* 44:219-236.

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

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.

Guyomard, H., Baudry, M. et A. Carpentier. 1996. “Estimating Crop Supply Response in the Presence of Farm Programmes: Application to the CAP.” *European Review of Agricultural Economics* 23:401-420.

Heckeleï, T. et 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.

Howitt, E. 1995. “Positive Mathematical Programming.” *American Journal of Agricultural Economics* 77:329-342.

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

Just, R.E., et R.D. Pope. 2001. “The Agricultural Producer: Theory and Statistical Measurement.” In B.L. Gardner and G.C. Rausser, ed. *Handbook of Agricultural Economics,* Vol. 1A. Amsterdam, Elsevier.

Kao, C., L.-F. Lee and M.M. Pitt, 2001. "Simulated Maximum Likelihood Estimation of the Linear Expenditure System with Binding Non-negativity Constraints." *Annals of Economics and Finance* 2:215-55.

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. et 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'économie et sociologie rurales* 53:50-73.

Ozarem, P.F. et 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.

Petrin, A. et K. E. Train. 2008. "A Control Function Approach to Endogeneity in Consumer Choice Model." *Forthcoming, Journal of Marketing Research*.

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

Röhm, O., et S. Dabbert. 2003. Integrating Agri-Environmental Programs into Regional Production Models: An Extension of Positive Mathematical Programming, *American Journal of Agricultural Economics* 85:254-265.

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.

Wooldridge, J.M. 2002. *The Econometrics of Cross-Sections and Panel Data.* The MIT Press: Cambridge MA.

Wooldridge, J.M. 1996. "Estimating systems of equations with different instruments for different equations." *Journal of Econometrics* 74:387-405.

Table 1. Estimates of the Yield, Input Demand and Acreage Shares Equations.

Explanatory Variable	Wheat		Other cereals		Oilseeds protein crops	
Yield supply						
Price effects (γ)	3.48	(0.32)	1.99	(0.18)	2.60	(1.22)
Constant	3.69	(0.20)	2.11	(0.29)	2.47	(0.15)
Production index	-0.07	(0.22)	-0.15	(0.21)	4.26	(0.26)
Average potential yield	9.72	(1.06)	8.74	(1.00)	7.75	(1.12)
Constant (α_0)	8.50	(0.07)	8.24	(0.10)	6.58	(0.07)
Trend	0.14	(0.01)	0.12	(0.01)	0.18	(0.01)
Production index	2.33	(0.09)	1.99	(0.11)	3.19	(0.12)
Root crop acreage	1.30	(0.14)	1.87	(0.19)	0.75	(0.22)
Aggregate composition	-		-1.18	(0.23)	1.20	(0.13)
Irrigation	0.41	(0.10)	1.64	(0.09)	-0.08	(0.06)
Scale effects (α_1)	-1.45	(0.23)	-0.93	(0.28)	-5.66	(0.36)
R-square	*0.42*		*0.35*		*0.31*	
Input demand						
Average input requirement (β)	6.63	(0.56)	5.97	(0.62)	6.85	(1.88)
Constant	6.54	(0.14)	6.25	(0.21)	5.85	(0.17)
Trend	0.01	(0.02)	0.14	(0.02)	-0.11	(0.02)
Production index	0.94	(0.21)	0.44	(0.22)	3.79	(0.26)
Root crop acreage	-2.31	(0.75)	1.29	(0.79)	2.66	(0.94)
Aggregate composition	-		-2.83	(0.30)	-4.08	(0.30)
Irrigation	-0.49	(0.18)	0.84	(0.11)	0.57	(0.15)
Sugar beets			9.19	(1.43)		
Potatoes			12.64	(2.07)		
Fodder crops			2.86	(0.80)		
R-square			*0.36*			
Acreage shares						
Fixed costs (g_0)	-		-2.97	(0.24)	-3.33	(0.52)
Fixed costs (g_1)	-		-0.54	(0.47)	-7.27	(0.89)
Matrix Q_0 elements	-		-7.99	(0.91)	-6.49	(0.57)
	-		-6.49	(0.57)	-31.90	(3.39)

Note: Standard errors are in parentheses.

Table 2. Estimates average price elasticities.

	Wheat	Other cereals	Oilseeds protein crops	Input
Yield supply functions				
Wheat	0.366	-	-	-0.366
Other cereals	-	0.241	-	-0.241
Oilseeds, protein crops	-	-	0.352	-0.352
Input demand functions				
Wheat	0.894	-	-	-0.894
Other cereals	-	0.460	-	-0.460
Oilseeds, protein crops	-	-	0.536	-0.536
Acreage share functions				
Wheat	1.062	-0.951	-0.047	0.077
Other cereals	-1.004	1.195	-0.207	-0.057
Oilseeds, protein crops	-0.058	-0.244	0.254	-0.020

Table 3. Estimates of variance-covariance matrix.

	Wheat		Other cereals		Oilseeds protein crops	
Yield (ω_{km}^{yy})						
Wheat	0.27	(0.04)	0.30	(0.04)	0.23	(0.04)
Other cereals	0.30	(0.04)	0.33	(0.05)	0.45	(0.04)
Oilseeds protein crops	0.23	(0.04)	0.23	(0.05)	0.44	(0.08)
Input use × Yield (ω_{km}^{yx})						
Wheat	0.09	(0.05)	0.09	(0.05)	0.10	(0.05)
Other cereals	0.11	(0.05)	0.11	(0.06)	0.12	(0.05)
Oilseeds protein crops	0.01	(0.06)	-0.04	(0.07)	0.23	(0.08)
Input use ($\omega_{km}^{xx} - \omega_{ww}^{xx}$)						
Wheat	0	(0.00)	0.03	(0.01)	-0.07	(0.07)
Other cereals	0.03	(0.01)	0.02	(0.03)	-0.22	(0.08)
Oilseeds protein crops	-0.07	(0.07)	-0.22	(0.08)	0.63	(0.14)

Note: Standard errors are in parentheses.

Appendix A – Matrix Notations.

$$\mathbf{a}_0^y = \begin{bmatrix} \alpha_{01} - 0.5\,\gamma_1 (w/p_1)^2 \\ \vdots \\ \alpha_{0K} - 0.5\,\gamma_K (w/p_K)^2 \end{bmatrix},\ \mathbf{a}_0^x = \begin{bmatrix} \beta_{01} - \gamma_1 (w/p_1) \\ \vdots \\ \beta_{0K} - \gamma_K (w/p_K) \end{bmatrix}.$$

$$\mathbf{B}_0^y = \begin{bmatrix} \alpha_{11} & & 0 \\ & \ddots & \\ 0 & & \alpha_{1K} \end{bmatrix},\ \mathbf{B}_0^x = \begin{bmatrix} \beta_{11} & & 0 \\ & \ddots & \\ 0 & & \beta_{1K} \end{bmatrix}.$$

$$\mathbf{M} = \begin{bmatrix} p_1 & & 0 & -w & & 0 \\ & \ddots & & & \ddots & \\ 0 & & p_K & 0 & & -w \end{bmatrix} \text{ and } \Delta'\mathbf{M} = \begin{bmatrix} -p_1 & p_2 & & 0 & w & -w & & 0 \\ \vdots & & \ddots & & \vdots & & \ddots & \\ -p_1 & 0 & & p_K & w & 0 & & -w \end{bmatrix}.$$

$$\mathbf{g}_0 = \begin{bmatrix} g_{02} \\ \vdots \\ g_{0K} \end{bmatrix},\ \mathbf{G}_0 = \begin{bmatrix} g_{11} & \cdots & g_{1K} \\ \vdots & \ddots & \vdots \\ g_{1K} & \cdots & g_{KK} \end{bmatrix}.$$

$$\mathbf{A}(\mathbf{b}_0) = \begin{bmatrix} (p_1\alpha_{11} - w\beta_{11}) + (p_2\alpha_{12} - w\beta_{12}) & \cdots & (p_1\alpha_{11} - w\beta_{11}) \\ \vdots & \ddots & \vdots \\ (p_1\alpha_{11} - w\beta_{11}) & \cdots & (p_K\alpha_{1K} - w\beta_{1K}) + (p_K\alpha_{1K} - w\beta_{1K}) \end{bmatrix}$$

$$\Delta'\mathbf{G}_0\Delta = \begin{bmatrix} g_{22} + g_{11} - 2g_{12} & \cdots & g_{2K} + g_{11} - g_{12} - g_{1K} \\ \vdots & \ddots & \vdots \\ g_{2K} + g_{11} - g_{12} - g_{1K} & \cdots & g_{KK} + g_{11} - 2g_{1K} \end{bmatrix} \text{ et } \mathbf{Q}_0 = \begin{bmatrix} q_1 & q_3 \\ q_3 & q_2 \end{bmatrix}$$

$$f_1 = s_e\,\pi_1^e + s_e^{\,2}\left[0.5(p_1\alpha_{11} - w\beta_{11}) - 0.5g_{11}\right] - g$$

$F_{1c} = s_e\left[(g_{1c} - g_{11}) + (p_1\alpha_{11} - w\beta_{11})\right]$ with s_e the acreage share of the endogenous crops. .

$$\mathbf{\Psi}_{yy} = \begin{bmatrix} \omega_{11}^y \cdots \omega_{1K}^y \\ \vdots \ \ddots \ \vdots \\ \omega_{1K}^y \cdots \omega_{KK}^y \end{bmatrix},\ \mathbf{\Psi}_{yx} = \begin{bmatrix} \omega_{11}^{yx} \cdots \omega_{1K}^{yx} \\ \vdots \ \ddots \ \vdots \\ \omega_{K1}^{yx} \cdots \omega_{KK}^{yx} \end{bmatrix} \text{ et } \mathbf{\Psi}_{gg} = \begin{bmatrix} \omega_{11}^g \cdots \omega_{1K}^g \\ \vdots \ \ddots \ \vdots \\ \omega_{1K}^g \cdots \omega_{KK}^g \end{bmatrix}.$$