Endogeneity of acreage choices in input allocation equation:

Implied problems and a solution

Alain CARPENTIER [1,2,3] and Elodie LETORT [1,2]

[1] INRA, UMR1302, F-35000 Rennes, France

[2] Agrocampus Ouest, UMR1302, F-35000 Rennes, France

[2] ENSAI, F-35000 Rennes, France

Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

1. 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 *et 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 *et al.*, 1989; Just *et 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 implementation of the control functions approach in three stages. In the fourth section, an application on French farm-level data are proposed. The last section of this paper provides some concluding remarks.

2. 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 labor 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 per unit of land 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.

2.1. 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) $$x_i = \sum_c^C s_{ci} x_{ci} + \eta_i \quad \text{with} \quad E\left[\eta_i \mid \mathbf{s}_i\right] = 0,$$

or as parametric functions:

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

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 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:

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

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

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

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 what 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 s_{ci} are exogenous with respect to u_{ci}^x, *i.e.*:

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

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

hold 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 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.

2.2. 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 studies falling into this category, the problem of variable input allocation is considered as a by product or not considered in further details. The first econometric models designed to model crop acreage decisions explicitly

[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.

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 the variable input allocation by considering a complete model of farmers' choices. Nevertheless 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 interpretations.

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. Variable input uses are usually considered at the farm level in most of these studies employing multi-crop econometric models (Paris, 1989). Using a maximum entropy framework, Lence and Miller (1998) 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. Their approach lies in between the approach of Dixon et al. (1984), Hornbaker et al. (1989) and the approach based on the specification of a complete model of farmers' choices. The approach of Dixon et al. (1984), Hornbaker et al. (1989) does not rely on the modelling of farmers' economic choices. Moreover, they do not consider input uses and acreages (or production levels in Lence and Miller's approach) as (partly) simultaneous choices.

2.3. Outline of the control function approach

The starting point of this research is that the exogeneity conditions $E\left[u_{ci}^{x} \mid \mathbf{s}_i, \mathbf{z}_i\right]$ 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 $s_i u_{ci}^{x} + \eta_i$ contains the endogenous explanatory variables $\mathbf{s}_i$. The use of equation (4) as an estimating equation requires the control of the terms $E\left[u_{ci}^{x} \mid \mathbf{s}_i, \mathbf{z}_i\right]$. 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, 2010; 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 $E\left[u_{ci}^{x} \mid \mathbf{s}_i, \mathbf{z}_i\right]$ terms known functions of $\mathbf{z}_i$, $\mathbf{s}_i$ and of a vector of unknown parameters θ. Let assume also that there exists a consistent estimator of θ, $\hat{\theta}$. The input allocation equation (1) can be transformed as:

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

where $c_c^x(\mathbf{z}_i,\mathbf{s}_i;\theta)$ are the control functions and where the conditional expectation of $E[\omega_i^x \mid \mathbf{z}_i,\mathbf{s}_i]$ is null by construction. Since the $c_c^x(\mathbf{z}_i,\mathbf{s}_i;\hat{\theta})$ terms are consistent estimators of the corresponding $c_c^x(\mathbf{z}_i,\mathbf{s}_i;\theta)$ terms, equation (7) can be used to construct consistent regression based estimators of $\mathbf{a}$. The control function approach basically splits the error terms u_{ci}^x in two terms: the control function $c_c^x(\mathbf{z}_i,\mathbf{s}_i;\theta)=E[u_{ci}^x \mid \mathbf{z}_i,\mathbf{s}_i]$ which "captures" and thus controls the links between u_{ci}^x and the endogenous variable vector $\mathbf{s}_i$; and a "new" error term $u_{ci}^x - c_c^x(\mathbf{z}_i,\mathbf{s}_i;\theta)$. By construction, $\mathbf{s}_i$ is exogenous with respect to the "new" error term. The crucial point is then to define the control functions $c_c^x(\mathbf{z}_i,\mathbf{s}_i;\theta)$ 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:

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

The control functions are determined by the following conditional expectations:

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

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".

3. 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 "concretely" illustrate the basic features of the approach. It combined standard quadratic yield functions with crop acreage (share) functions derived along the line of Heckeleï and Wolff (2003). It is chosen because of its fairly simple interpretation and its flexibility. A specific feature is the structural modeling of error terms of the econometric 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.

3.1. 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:

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

(11) $\quad$ with $\alpha_{ci} = \alpha_{0c} + 0.5\alpha_{1c}s_{ci} + v_{ci}^y$

(12) $\quad$ and $\beta_{ci} = \beta_{0c} + 0.5\beta_{1c}s_{ci} + v_{ci}^x$

where x_{ci} is the quantity of variable input used per hectare by farm i devoted to crop c, α_{ci} and 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 the input requirement are specified as functions of the crop acreage to account for potential scale effects. The estimates of these yield functions can thus be checked with agricultural scientists or extension agents. v_{ci}^y and v_{ci}^x are random terms. These terms are split into two parts for simplifying their interpretation:

(13) $\quad v_{ci}^y = e_{ci}^y + \varepsilon_{ci}^y$ 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 $p_{ci}y_{ci} - w_i x_{ci}$ for each crop c. 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. So we consider that farmers 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, after they have observed stochastic events (as specific climatic conditions or diseases) they can adjust their input uses. Therefore acreages and input uses decisions are partially simultaneous. The maximization of this profit function under technological constraints leads to the following per hectare variable input demand, yield supply and gross margin functions:

(14) $\quad x_{ci} = \beta_{0c} + 0.5\beta_{1c}s_{ci} - \gamma_c \left(w_i / p_{ci}\right) + v_{ci}^x$

(15) $\quad y_{ci} = \alpha_{0c} + 0.5\alpha_{1c}s_{ci} - 0.5\gamma_c \left(w_i / p_{ci}\right)^2 + v_{ci}^y$

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

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 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 analyzing 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. 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.

3.2. Acreage functions

Farmers' acreage choices are modeled 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 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: the scale effects of the crop gross margins $0.5\left(p_{ci}\alpha_{1c} - w_i\beta_{1c}\right)$ and the implicit management cost of the chosen acreage $C(s_c)$. This cost function is used in the positive mathematical programming literature (PMP) (Howitt 1995; Paris and Howitt 1998; Heckeleï and Wolff 2003). It can be interpreted as a reduced form function smoothly approximating the unobserved variable costs associated with a given acreage (energy costs...) and the effects of binding constraints on acreage choices. These constraints are agronomic constraints or constraints associated to limiting 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 is likely to increase due to work peak load or due to machinery overuse, whether machinery is specific or not. Farmers are also subject to agronomic constraints because some crop rotations are ``forbidden'' or impossible due to inconsistencies in planting and harvesting dates. Cultivating a given crop two consecutive years on the same plot may be strongly unwarranted due to dramatic expected pest damages. These crop rotations are thus almost “forbidden” because their opportunity cost is very large in standard price ranges. These “forbidden” or impossible crop rotations determine the bounds imposed to acreage choices in (P)MP models. This 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) and due to the implicit bounds imposed on the acreage choices due to impossible or “forbidden” crop rotations.

This cost function is assumed to have a quadratic form:

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

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

The land use constraint is included into the restricted indirect profit function. All crops are assumed to be cultivated. The crop c is considered as the reference crop. The maximization 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 (and also for the empirical use):

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

(20) with $v_{ci}^{s} = p_{ci} e_{ci}^{y} - p_{Ci} e_{Ci}^{y} - w_i e_{ci}^{x} + w_i e_{ci}^{x} - e_{ci}^{g}$

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_{cm}. 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}.

3.2. "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 is written equal to 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 to allocate variable inputs across crops c and takes part into the econometric model. The complete system is described in appendix A with simple matrix notations. It seems important to define again the error terms of the econometric equation systems. Note that error terms of each equations are now denoted by u_{ci}^{y}, u_{i}^{x} and u_{ci}^{s} and that u_{ci} is similar to v_{ci} except for the input allocation equation.

(21) $u_{ci}^{y} = v_{ci}^{y} = e_{ci}^{y} + \varepsilon_{ci}^{y}$

(22) $u_{i}^{x} = s_{ci} v_{ci}^{x} + \eta_i = s_{ci}\left[e_{ci}^{x} + \varepsilon_{ci}^{x}\right] + \eta_i$

(23) $u_{ci}^{s} = v_{ci}^{s} = p_{ci} e_{ci}^{y} - p_{Ci} e_{Ci}^{y} - w_i e_{ci}^{x} + w_i e_{ci}^{x} - e_{ci}^{g}$

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. To explain the endogeneity problem, we use simple matrix notations. We consider $\mathbf{e}_i$ the vector of heterogeneity terms such as $\mathbf{e}_i' = \left[\mathbf{e}_i^{y}\ \mathbf{e}_i^{x}\ \mathbf{e}_i^{g}\right]$ and $\boldsymbol{\varepsilon}_i$ the vector of stochastic events terms such as $\boldsymbol{\varepsilon}_i' = \left[\boldsymbol{\varepsilon}_i^{y}\ \boldsymbol{\varepsilon}_i^{x}\ \boldsymbol{\varepsilon}_i^{g}\right]$. The vectors of model error terms are defined by $\mathbf{u}_i^{y} = \left[u_{ci}^{y}\right]$, $\mathbf{u}_i^{x} = \left[u_{i}^{x}\right]$ and $\mathbf{u}_i^{s} = \left[u_{ci}^{s}\right]$. The vector $\mathbf{z}_i$ and $\mathbf{s}_i$ are respectively the vector of exogenous variables (outputs and inputs prices) and the vector of acreage shares. The preceding interpretations of the error terms allow to define the following mean assumptions: $E\left[\mathbf{e}_i^{y} \mid \mathbf{z}_i\right] = 0$, $E\left[\boldsymbol{\varepsilon}_i \mid \mathbf{z}_i\right] = 0$, $E\left[\mathbf{e}_i^{g} \mid \mathbf{z}_i\right] = 0$, $E\left[\eta_i \mid \mathbf{z}_i\right] = 0$ and $E\left[\mathbf{s}_i' \boldsymbol{\varepsilon}_i^{x} \mid \mathbf{z}_i\right] = 0$. This implies that each component of $\mathbf{u}_i$ 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 $E\left[\mathbf{s}_i \mathbf{e}_i^{x} \mid \mathbf{z}_i\right] \neq 0$ or $E\left[\mathbf{e}_i^{x} \mid \mathbf{z}_i, \mathbf{s}_i\right] \neq 0$. 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 β_{0C} can not be identified. Second, the heterogeneity terms $\mathbf{e}_i^{y}$, $\mathbf{e}_i^{x}$ and $\mathbf{e}_i^{g}$ are the "interest error terms" for determining the control functions whereas $\boldsymbol{\varepsilon}_i^{y}$, $\boldsymbol{\varepsilon}_i^{x}$ and η_i can be viewed as "disturbances".

4. Estimation procedure

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.

4.1. The control functions construction

The construction of control function 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)\endnote{Linear projection techniques combined with limited assumptions on the distribution of the heterogeneity terms can be used in some special case (see, e.g., Chamberlain, 1982; Wooldridge, 2004).}. The normal distribution usually appears to be a convenient choice. 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 $\mathbf{u}_i^y$, $\mathbf{u}_i^x$ and $\mathbf{u}_i^s$ are linear transformations of $\mathbf{e}_i$, they are also normally distributed. 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:

$$(24) \qquad V\left[\mathbf{e}_i \mid \mathbf{z}_i\right] = V\left[\mathbf{e}_i\right] = \boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{yy} & \boldsymbol{\Psi}_{yx} & 0 \\ \boldsymbol{\Psi}_{yx} & \boldsymbol{\Psi}_{xx} & 0 \\ 0 & 0 & \boldsymbol{\Psi}_{gg} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Psi}_{yz} & 0 \\ \boldsymbol{\Psi}_{xz} & 0 \\ 0 & \boldsymbol{\Psi}_{gg} \end{pmatrix}$$

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 $\boldsymbol{\Psi}$ (see section on the implementation of the approach).

The control functions defined here seek to solve two problems: 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:

$$(25) \qquad E\left[\mathbf{u}_{ci}^x \mid \mathbf{s}_i, \mathbf{z}_i\right] = \mathbf{s}_i' E\left[\mathbf{e}_{ci}^x \mid \mathbf{s}_i, \mathbf{z}_i\right]$$

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

$$(26) \qquad E\left[\mathbf{e}_{ci}^x \mid \mathbf{s}_i, \mathbf{z}_i\right] = \boldsymbol{\Psi}_{xz}\mathbf{C}_i^x\mathbf{u}_i^s \text{ and } E\left[\mathbf{e}_{ci}^y \mid \mathbf{s}_i, \mathbf{z}_i\right] = \boldsymbol{\Psi}_{yz}\mathbf{C}_i^y\mathbf{u}_i^s$$

where $\mathbf{C}_i = \left[\mathbf{C}_i^y \ \mathbf{C}_i^x\right]$ 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.

4.2. A three-stage procedure

The control function approach is implemented in three stages. 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 identifiable parameters, *i.e.* all parameters except β_C. This system is a simultaneous equation system and uses the SUR estimator. This stage allows to obtain $\mathbf{u}_i^s$ and to proceed to the next stage.

In the second stage, estimators of the first stage are assumed to be available for constructing a consistent estimator of a useful part of the variance-covariance matrix $\mathbf{\Psi}$. 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 SUR system. This stage allows 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 interests parameters are estimated and auxiliary parameters, *i.e.* $\mathbf{\Psi}_{yz}$ and $\mathbf{\Psi}_{xz}$. This econometric model is not a standard non linear SUR system for two reasons. First, the different equations of the system share many parameters and the corresponding SUR estimators are generally non consistent. Second, the input allocation equation use $\mathbf{s}_i$ as a regressor, whereas $\mathbf{s}_i^-$ is the dependant variable of the acreage equations. Thus we use the generalized method of moments (GMM) to construct consistent estimators.

A few remarks are in order for the implementation of this approach. First, 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 $\mathbf{\Psi}_{yz} = \mathbf{\Psi}_{xz} = 0$. This is a test of the interest 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.

5. Empirical application

5.1. The data

An illustrative application of the approach is provided by an empirical analysis of farm production data from the region *Meuse* in France. This data set consists of an unbalanced panel of farms covering a period between 1955 and 2007. It contains approximately 4,000 observations. It has the advantage to have detailed data on input allocations between crops that is useful to examine the performance of our model. This database provides farm data on variable input expenditures for each crop, output quantities and prices, subsidies and acreage for each crop. Three crop groups are considered, such as wheat, other cereals (mainly barley and maize) and oilseeds and protein crops (mainly rapeseed). The different variable inputs (fertilizer, pesticides and seeds) are aggregated into a single variable input for simplicity. The corresponding price index is obtained from Eurostat. All economics quantities are defined in € in units of 2000.

5.2. Parameter estimates

The following section reports on the results of the estimation of the multi-crop econometric model. It is composed of a yield supply equation for each crop, an input allocation equation and an acreage share equations for wheat and for the other cereals. The oilseeds are considered as the crop reference. It is estimated using the three-stage procedure described in the last section. Some variables are integrated into the model to control for technical change and the heterogeneity of farms. Parameters α_1 and β_1 are assumed to be null for simplicity. The parameter α_0, which is interpreted as the maximum yield of crop c in the yield supply equations, is defined as a function of spatial and regional dummies. The parameter β_0, which is interpreted as the variable input quantity required for achieving the maximum yield in the input allocation equation, depends only on regional dummies[3]. In the acreage choices equations, the parameter of fixed costs g_0 is defined as a function of labor variable because it is interpreted as the fixed costs associated with limiting quantities of quasi-fixed inputs such as labor or machinery.
Table 1 contains the parameter estimates α_0, β_0 and γ of the yield supply and input demand equations, and the parameter estimates g and q_1, q_2 and q_3, the elements of the matrix $\mathbf{Q}$ of the acreage equations. The estimates of parameters associated to regional and spatial dummies are not reported in the table to save space. The fit of the model is correct. The R² criterion ranges from 0.28 to 0.40 for yield supply and input demand equations[4]. Almost all coefficients are different from zero at high significance levels.

Table 1. Estimates for yield supply, input demand and acreage share equations

	Wheat	**Other cereals**	**Oilseeds**
Price effects γ	334.82***	296.68***	149.63
Potential yield β_0	768.47***	793.56***	575.91***
Optimal input use α_0	633.35***	627.55***	513.91***
Fixed costs g	-59.42***	-100.71***	
Element q_1		-219.91***	
Element q_2		-149.85***	
Element q_3		-129.67***	
Test Hansen	0.92	-	-
Test Wald	<0.05	-	-
R² yield	0.35	0.28	0.37
R² input use	0.40	-	-

The price effects correspond to parameters γ_c for each crop c associated with the price ratio. These parameters are significantly positive for all cereals, implying concavity for the yield functions. It is positive but not significantly different from zero for oilseeds. The estimated

[3] The base year is 2006 and the base region includes some *cantons* with the highest yield.
[4] We have not the R² criterions for acreage equations because we have estimated first order conditions. Nevertheless it is possible to calculate them.

parameters of the matrix $\mathbf{Q}$ imply concavity in prices of the restricted profit function without imposing constraints. Necessary and sufficient conditions are that the principal minors of the matrix $\mathbf{Q}$ alternate in sign, starting with negative value. These conditions are fulfilled because $q_1 < 0$ and $q_1 q_2 - q_3 q_3 > 0$. The validity of the control functions are tested using a Wald test. The null hypothesis is that elements of the variance-covariance $\mathbf{\Psi}$ are not jointly significantly different from zero. The t-test is … with p-value <0.001. Thus, the null hypothesis is rejected. This test shows evidence of an acreage endogeniety problem in the yield and input equations. A test of overidentifying restrictions is also realised to verify the validity of our instruments and the validity of our model specification. The Hansen test does not reject the joint validity of our instruments. The other estimates parameters are globally consistent with our expectations. The maximum yield is estimated at 768€ per hectare for wheat, 794€ per hectare for other cereals and 576€ per hectare for oilseeds. The variable input quantity required to achieve these maximum yield is estimated at 633€per hectare for wheat, 627€ per hectare for other cereals and 514€ per hectare for oilseeds. In order to underline the validity of the estimation results, elasticities of input demand with respect to input prices and output prices are derived. Only own price elasticities are presented in table 2.

Table 2. Own price elasticities of input demand[5]

	Input prices	**Output prices**
Wheat	- 0.43 (0.13)	0.85 (0.33)
Other cereals	- 0.49 (0.14)	1.01 (0.42)
Oilseeds	- 0.22 (0.08)	0.30 (0.09)

They are found to be in a reasonable range. All own input prices elasticities are negative and all own output prices elasticities are positive. A unit increase of price input has a lower effect on input demand for oilseeds. Similarly, the response of input demand with respect to an increase of output prices is lower for oilseeds.

4.2. Statistical comparison with alternative method

The objective of the paper is to determine an approach to allocate variable input among crops, thus an examination of the predictions seems to be useful. These predictions are compared with predictions obtained with two others models. The first model consists on the complete system with a variable input demand equation for each crop. This model, denoted by model BASE, is estimated with the SUR method. The second model is the model generally used in the literature. The only difference between this model and ours is that they do not take account the acreage endogeneity. They replace the acreage variable in the input allocation equation by the reduced form of the acreage function. They have no longer endogeneity problem and they can estimate this system with regression techniques. This model is denoted by model RT.

Table 3 presents the average estimated input use for each crop obtained with the three models. The model BASE provides the best average prediction for the all outputs. This is an expected

[5] They are calculated at the sample mean. Approximated standard errors are in parenthesis.

result. What is more interesting is the comparison between the two others models. The allocation in mean of input use among the crops is better with the model CF for the three crops. The model RT tends to overestimate the input quantities used for other cereals and underestimate the input quantities used for wheat and oilseeds.

Table 3. Average of predicted and observed input use[6]

	Observed	Model BASE	Model RT	Model CF
Wheat	345.11 (76.5)	344.49 (31.9)	327.69 (46.5)	335.73 (49.6)
Other cereals	313.66 (76.5)	312.91 (36.9)	351.58 (39.0)	327.31 (57.2)
Oilseeds	390.78 (95)	390.12 (40.9)	366.97 (48.1)	385.64 (85.1)

Some figures have been realized to compare observed input use with predicted input use by the three models. These figures show that input use predictions are quite similar between model RT and model CF for cereals. On the other side, our model predicts much better quantities input used for oilseeds. All results about the estimation of the three models are available on request.

Conclusion

This paper highlights two mains 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 that few variables are generally available to control farmers' heterogeneity. The test build and realized in the application confirms this intuition. The comparison of models with and without control functions shows the usefulness to consider acreage as endogenous. There are some differences in input allocations in average between the models. These differences are show graphically for oilseeds.
Second we suggest that a structural econometric model is necessary to account for the bias associated to the acreage endogeneity. In this paper, we propose a structural econometric multicrop model for explicitly determining the origin of the bias and providing potential solutions to allocate inputs among crops. This model is composed of yield supply, input demand and acreage choices equations. We consider land as an input fixed but allouable as in an important 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 links between yield, input uses and acreage choices. The structural modelling of error terms and especially the error term additivity play 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. It could be also applied in other contexts where inputs need to be allocated to activities.
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

[6] Standard errors are in parentheses.

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 do not account for corner solutions of activity choices. This is a potentially important weakness of this framework, particularly in the crop production context. The specification of a fully structural model for activity choices with corner solutions is possible but difficult to implement. 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

Antle, J.M. 1983. “Sequential Decision Making in Production Models.” *American Journal of Agricultural Economics* 65:282-290.

Antle, J.M. and S.A. Hatchett. (1986). “Dynamic Input Decisions in Econometric Production Models.” *American Journal of Agricultural Economics* 68:939-949.

Blackorby, C., D. Primont and R. R. Russell. 1978. *Duality, Separability, and Functional Structure: Theory and Economic Implications.* Elsevier, North-Holland:New-York.

Blundell, R. and S. Bond. 2000. “GMM Estimation with Persiste nt Panel Data: An Application to Production Functions,” *Econometric Reviews* 19, 321-340.

Carpentier, A., and R. D. Weaver. 1997. “Damage Control Productivity: Why specification matters.” *American Journal of Agricultural Economics* 52:11-22.

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

Chamberlain, G. 1987. ”Asymptotic Efficiency in Estimation with Conditional Moment Restrictions”, *Journal of Econometrics* 34, 305–334.

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

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 Statistics* 2: 360-366.

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.

Heckeleï, 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. 2008. “Econometric causality.” *Cemmap working paper* CWP1/08.

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.

Heckman, J. J. and E. Vytlacil. 1998. “Instrumental variables methods for the correlated random coefficient model.” *Journal of Human Resources* 33, 974-987.

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.

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.

Lence, S. H. and D. J. Miller. “Recovering Output-Specific Inputs from Aggregate Input Data: A Generalized Cross-Entropy Approach.” *American Journal of Agricultural Economics* 80:852-867.

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 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.

Newey, W., and D. McFadden. 1994. “Large Sample Estimation and Hypothesis Testing”, Chapter 36 Handbook of Econometrics, Vol 4, McFadden and Engle, (eds.), Elsevier, North Holland.

Paris, Q. 1988. “A Sure Bet on Symmetry.” *American Journal of Agricultural Economics* 71:344-351.

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.

Reiss, P. C. and F. A. Wolak. 2007. Structural Econometric Modeling: Rationales and Examples from Industrial Organization." Chapter 64 Handbook of Econometrics, in: J.J. Heckman and E.E. Leamer (eds), Handbook of Econometrics, Vol 6, Elsevier, North Holland.

Rivers, D. and Q. H. Vuong. 1988. “Limited information estimators and exogeneity tests for simultaneous probit models.” *Journal of Econometrics* 39, 347-366.

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.

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.