TOWARDS SUSTAINABLE AGRI-FOOD SYSTEMS:
BALANCING BETWEEN MARKETS AND SOCIETY

Dealing with corner solutions in multi-crop micro-econometric models: An endogenous switching regime approach with regime fixed costs

Obafèmi Philippe Koutchadé[1], Alain Carpentier[2], Fabienne Femenia[3]

[1] PhD candidate, INRA, UMR SMART-LERECO, Rennes, France

[2] Senior reseacher, INRA, UMR SMART-LERECO, Rennes, France;
Professor, Departemennt of Economics, Agrocampus Ouest, Rennes, France

[3] Reseacher, INRA, UMR SMART-LERECO, Rennes, France;

Contribution presented at the XV EAAE Congress, "Towards Sustainable Agri-food Systems: Balancing Between Markets and Society"

August 29th – September 1st, 2017

Parma, Italy

Abstract

We propose an Endogenous Regime Switching model specifically designed for modelling acreage choices with corner solutions featuring regime fixed costs. Contrary to models based on censored regression systems, this model is fully coherent from an economic point of view, by construction. To illustrate its empirical tractability, we estimate a random parameter version of this model for a panel dataset of French arable crop producers. Our results show that the model fits well the data, the regime fixed cost matter, and that the decision to produce a crop or not play a major role in the acreage responses to economic incentives.

Keywords: corner solutions, endogenous regime switching models, agricultural production choices

1. Introduction

Corner solution problems are pervasive in micro-econometric acreage choice models because farmers rarely produce the same crop set in a considered sample, even in samples considering specialized farms. Agricultural economists usually use two approaches for coping with null crop acreages. First, crops can be aggregated for eliminating or, at least, attenuating the occurrence of null crop acreages. Of course, this approach can lead to substantial information loss. Second, corner solutions can be dealt with by specifying acreage choices as a system of censored regressions (see, e.g., Platoni et al., 2012). However, if censored regression systems explicitly account for null crop acreages from a statistical viewpoint, they cannot consistently represent acreage choices with corner solutions. This point was made, for consumer demand systems, by Arndt et al (1999). More generally, acreage choice models suitably accounting for corner solutions need to be specified as endogenous regime switching (ERS) models, in which production regimes are defined as the subsets of crops with non-null acreages – *i.e.* by the subsets of actually produced crops.

Regime and acreage choice decisions are closely linked since these decisions are taken simultaneously and depend on common drivers. For instance, the choice of the set of crops to be produced depends on the optimal acreages of these crops. Importantly, responses to crop price changes of crop acreage decisions depend on the regime in which these crops are produced. For instance, winter wheat crop acreages cannot respond to corn price changes in regimes where winter wheat is produced whereas corn is not produced. Censored regression systems cannot account for such effects since in these modelling framework farmers' acreage choices are described by a model that is common to all production regimes. In our ERS modelling framework farmers' acreage choices are described by models that are specific to each production regime.

According to our knowledge, micro-econometric ERS models involving multiple corner solutions were defined only for modeling consumer demand systems (see, e.g., Kao et al. 2001) or firm input demand systems (see, e.g., Chakir et al. 2004), following the pioneering works of Wales and Woodland (1980) and of Lee and Pitt (1986). However, these models have rarely been used in practice, probably because their estimation is challenging, and d,espite the development of estimation procedures with simulation methods (see, e.g., Kao et al. 2001).

The main aim of our paper is to propose an ERS model specifically designed for empirically modeling acreage choices with corner solutions. This model is fully coherent from an economic point of view and includes regime fixed costs, which is to our knowledge a unique feature compared to other ERS models with multiple corner solutions found in the economic literature. These regime fixed costs allow accounting for unobserved costs, such as marketing or management costs, which depend on the set of crops grown simultaneously.

The ERS model we propose defines a Nested MultiNomial Logit (NMNL) acreage choice model (Carpentier and Letort, 2014; Koutchadé et al., 2015) for each potential production regime. The regime choice is based on a discrete choice model in which farmers choose the subset of crops they produce by comparing the profit levels associated to each regime. The econometric model derived from this framework is theoretically consistent – in its deterministic and in its random parts – and can be combined with yield supply and variable input demand functions. Furthermore, following Koutchadé et al. (2015), this model accounts for the unobserved heterogeneity in farmers' behaviors through the specification of random parameters. I.e. we assume that most model parameters are farmer specific and estimate their distribution across the farmers' population described by our sample. Given that our model is fully parametric, it can be efficiently estimated within a Maximum Likelihood (ML) estimation framework. The structure of the model and the functional form of its likelihood function actually make the version of the Simulated Expectation-Maximisation (SEM) algorithm developed by Delyon et al. (1999) especially suitable for maximizing the sample likelihood function. Importantly, once their probability distribution has been estimated, each farmer specific parameter can be 'statistically calibrated' for simulation purpose.

We illustrate the empirical tractability of our approach by estimating a seven crops – and ten production regimes – production choice model for a sample of French arable crop producers. The estimated model is then used to simulate the impacts of a crop price change on acreages and illustrate how accounting for endogenous production regime choices and production regime fixed costs can affect the simulation results.

Our results tend to show that our ERS multi-crop model with regime fixed costs perform well according to the standard fit criteria. They also tend to show that the regime fixed costs significantly matter for explaining the production regime choices and that the decision to produce a crop or not plays a major role in the acreage choice responses to economic incentives. In particular, our simulation results show that the acreage choice responses to price changes exhibit threshold effects due to the regime fixed costs.

The rest of the paper is organized as follows. The approach proposed to account for endogenous regime switching and regime fixed costs in the modelling of acreage decisions is presented in the first section. The structure of the econometric model of acreage and production choices is described in the second section. Identification and estimation issues are discussed in the third section. The illustrative estimation results are provided in the fourth section. Finally, we conclude.

2. Endogeneous regime switching acreage choices with regime fixed costs

This section presents the theoretical modelling framework we propose for dealing with corner solutions in micro-econometric acreage choice models. We adopt an ERS approach for the resulting models to be fully consistent from a micro-economic viewpoint. We also allow for regime fixed costs for improving the ability of the resulting models to capture the effects of potentially important drivers of farmers' acreage choices.

Let consider a risk neutral farmer i, who can allocate his fixed cropland to K crops. Let $\mathcal{K} \equiv \{1,...,K\}$ denote the set of crops available to this farmer and let $\mathcal{R} \equiv \{1,...,R\}$ denote the set of production regimes, *i.e.* the set of crop subsets considered by the farmer when choosing the crops he/she will produce.[1] The term $\mathcal{K}^{+}(r)$ denotes the subset of crops produced in regime *r*, with $\mathcal{K}^{+}(1) = \mathcal{K}$ by convention, while the term $\mathcal{K}^{0}(r)$ denotes subset of crops not produced in regime *r*. Finally, let the term $\mathbf{s} \equiv (s_k : k \in \mathcal{K})$ denote an acreage share vector, with $\mathbf{s} \geq \mathbf{0}$ and $\mathbf{s}'\boldsymbol{\iota} = 1$ where $\boldsymbol{\iota}$ is the unitary column vector with dimension *K*, and let the function $\rho(\mathbf{s})$ define the regime of the acreage share vector s.

[1] Some combinations of crops are not possible because of agronomic constraints. $\mathcal{R}$ thus does not contain all the possible subsets of $\mathcal{K}$ but the subsets that can potentially be chosen by farmers.

In period t the considered farmer is assumed to solve the following expected profit maximization problem:

(1) $$\max_{\mathbf{s}} \left\{ \mathbf{s}'\boldsymbol{\pi}_{it}^{e} - C_{it}(\mathbf{s}) - F_{it}\left(\rho(\mathbf{s})\right) \text{ s.t. } \mathbf{s} \geq \mathbf{0} \text{ and } \mathbf{s}'\boldsymbol{\iota} = 1 \right\}$$

where $\boldsymbol{\pi}_{it}^{e} \equiv (\pi_{k,it}^{e} : k \in \mathcal{K})$ is the vector of crop returns expected when choosing **s**, $C_{it}(\mathbf{s})$ is the implicit management cost of acreage **s** and $F_{it}(r)$ is the fixed cost of production regime *r*.

The acreage management cost function $C_{it}(\mathbf{s})$ accounts for the crop variable costs not included in the crop gross margins and for the implicit costs related to the constraints on the acreage choices due the limiting quantities of quasi-fixed factors or to agronomic factors. The quasi-fixed factor and agronomic constraints providing motives for diversifying crop acreages, the function $C_{it}(\mathbf{s})$ is assumed to be convex in **s**. In order to ensure that the solution in **s** to problem (1) is unique we strengthen this assumption by assuming that $C_{it}(\mathbf{s})$ is strictly convex in **s**. Such cost functions are used in the Positive Mathematical Programming literature (see, e.g., Howitt, 1995; Heckeleï et al, 2012) and in the multi-crop econometric literature (see, e.g., Carpentier and Letort, 2012, 2014). Ignoring the regime fixed costs, the optimal acreage choice is determined by maximizing the sum of the crop expected gross margins $\pi_{k,it}^{e}$ weighted by their acreage shares s_k minus the costs associated to the crop acreage **s**, $C_{it}(\mathbf{s})$. In this model the management costs of the crop acreage prevent farmers to solely produce the most profitable crop.

The regime fixed cost terms $F_{it}(r)$ introduce discrete elements in farmers' acreage choices with $F_{it} \in \{F_{it}(r) : r \in \mathcal{R}\}$. These terms account for the hidden fixed costs incurred by the farmer for any acreage choice in the regime. They include fixed costs related to the marketing process of the crop products or those incurred when purchasing specific variable inputs, when renting specific machines, when seeking crop specific advises, *etc*. These costs do not depend on the chosen acreage in a given regime, they only depend on the crop set defining this regime.

The smooth acreage management cost function $C_{it}(\mathbf{s})$ and the discontinuous regime fixed cost function $F_{it}\left(\rho(\mathbf{s})\right)$ are expected to impact farmers' crop diversification in opposite directions. While limiting quantities of quasi-fixed factors impose constraints fostering crop diversification, regime fixed costs are expected to foster crop specialization. In particular, the regime fixed costs are expected to be non-decreasing in the number of produced crops.[2]

[2] Note however that in specific empirical settings the $F_{it}(r)$ terms may also capture the effects of exogenous factors preventing farmer *i* to produce specific crops, e.g. due to unsuitable soils or to lacking outlets. In the empirical

It appears convenient, as well as intuitively appealing, to decompose farmers' acreage choice problem into two steps, as commonly done in acreage choice models based on censored regression systems (see, e.g., Skockaï and Moro, 2006, 2009; Lacroix and Thomas, 2011; Fezzi and Bateman, 2011). Namely, we distinguish the production regime choice from the acreage choice while assuming that both choices are linked by the effects of common observed or unobserved drivers.

Our modelling framework is based on a standard backward induction approach according to which farmers choose their production regime after examining their expected profit in each possible production regime. First, the acreage choice problem is solved for each potential regime:

(2) $$\max_{\mathbf{s}} \left\{ \mathbf{s}'\boldsymbol{\pi}_{it}^{e} - C_{it}(\mathbf{s}) \text{ s.t. } \mathbf{s} \geq \mathbf{0}, \mathbf{s}'\boldsymbol{\iota} = 1 \text{ and } s_k = 0 \text{ if } k \in \mathcal{K}^0(r) \right\},$$

yielding regime specific optimal acreage shares:

(3a) $$\mathbf{s}_{it}(r) \equiv \arg\max_{\mathbf{s}} \left\{ \mathbf{s}'\boldsymbol{\pi}_{it}^{e} - C_{it}(\mathbf{s}) \text{ s.t. } \mathbf{s} \geq \mathbf{0}, \mathbf{s}'\boldsymbol{\iota} = 1 \text{ and } s_k = 0 \text{ if } k \in \mathcal{K}^0(r) \right\}$$

and regime specific optimal expected profit – regime fixed excluded – levels:

(3b) $$\Pi_{it}(r) \equiv \max_{\mathbf{s}} \left\{ \mathbf{s}'\boldsymbol{\pi}_{it}^{e} - C_{it}(\mathbf{s}) \text{ s.t. } \mathbf{s} \geq \mathbf{0}, \mathbf{s}'\boldsymbol{\iota} = 1 \text{ and } s_k = 0 \text{ if } k \in \mathcal{K}^0(r) \right\}.$$

for $r \in \mathcal{R}$. Second, the production regime r_{it} is determined by comparing the regime specific expected profit levels while accounting for the production regime fixed costs, i.e. r_{it} is defined as the solution in *r* to a simple maximization problem with:

(4) $$r_{it} \equiv \max_{r} \left\{ \Pi_{it}(r) - F_{it}\left(\rho\left(\mathbf{s}_{it}(r)\right)\right) \text{ s.t. } r \in \mathcal{R} \right\}.$$

The obtained optimal regime r_{it} is assumed to be unique as multiple solutions can only occur in knife-edge cases. Of course, the optimal acreage choice $\mathbf{s}_{it}$ and the expected profit level Π_{it} are obtained by combining equation (4) and equations (3), with $\mathbf{s}_{it} = \mathbf{s}_{it}(r_{it})$ and $\Pi_{it} = \Pi_{it}(r_{it})$.

The regime specific acreage choices $\mathbf{s}_{it}(r)$ are derived from optimization problems that only differ from one regime to the other due to nullity constraints. These constraints are sufficient for these acreage choices to respond significantly differently to economic changes. For instance, the regime *r* acreage choice, $\mathbf{s}_{it}(r)$, doesn't respond to changes in the expected returns of the crops not produced in regime *r*. Similarly, the wheat acreage is expected to be more responsive to the price of wheat in farms producing barley than in farms not producing other straw cereals. Acreage choice models based on censored regression systems cannot reproduce such patterns.

---

application presented in section 4, such features are unlikely to occur. Our sample covers a limited geographical area and we only consider crops which can be profitably produced in this area.

Note that the regime fixed cost considered in the maximization problem (4) determining the optimal regime r_{it} is not simply $F_{it}(r)$ but $F_{it}\left(\rho\left(\mathbf{s}_{it}(r)\right)\right)$. The reason is that the acreage $\mathbf{s}_{it}(r)$ may not belong to regime *r*, depending on the functional form chosen for the cost function $C_{it}(\mathbf{s})$. This acreage is only guaranteed to belong to a regime 'included' in regime *r* in the sense that some crops of $\mathcal{K}^{+}(r)$ may not be produced in the regime $\rho\left(\mathbf{s}_{it}(r)\right)$.

The Multinomial Logit (MNL) modelling framework proposed by Carpentier and Letort (2014) is especially convenient in this context. It is based on functional forms of the acreage management cost function ensuring that the regime specific acreage share $\mathbf{s}_{it}(r)$ and expected profit $\Pi_{it}(r)$ are obtained in analytical closed forms and that are smooth in their parameters. For instance, if we assume that the functional form of the acreage management cost function is given by the 'entropic' function:

(5a) $$C_{it}(\mathbf{s}) = \sum\nolimits_{k\in\mathcal{K}^{+}(r)} s_k \beta_{k,it}^{s} + \alpha_i^{-1} \sum\nolimits_{k\in\mathcal{K}^{+}(r)} s_k \ln s_k \quad \text{with} \quad \alpha_i > 0$$

then the regime specific acreage share vectors $\mathbf{s}_{it}(r)$ are given by:

(5b) $$s_{k,it}(r) = j_k(r) \frac{\exp\left(\alpha_i(\pi_{k,it}^{e} - \beta_{k,it}^{s})\right)}{\sum\nolimits_{\ell\in\mathcal{K}^{+}(r)} \exp\left(\alpha_i(\pi_{\ell,it}^{e} - \beta_{\ell,it}^{s})\right)} \quad \text{with} \quad \begin{cases} j_k(r) = 1 \text{ if } k \in \mathcal{K}^{+}(r) \\ j_k(r) = 0 \text{ if } k \in \mathcal{K}^{0}(r) \end{cases}$$

while the corresponding expected profit levels $\Pi_{it}(r)$ are given by:

(5c) $$\Pi_{it}(r) = \alpha_i^{-1} \ln \sum\nolimits_{\ell\in\mathcal{K}^{+}(r)} \exp\left(\alpha_i(\pi_{\ell,it}^{e} - \beta_{\ell,it}^{s})\right).$$

These specific properties of the MNL modelling framework significantly simplifies the specification of the acreage choice models featuring corner solutions. They basically imply that the production regime choice can be defined as a standard discrete choice, *i.e.* that of the most profitable production regime among a predetermined regime set.

Indeed, the optimal acreage share of crop *k* in regime *r*, i.e. $s_{k,it}(r)$, is ensured to be strictly positive if crop *k* belongs to regime *r*, i.e. if $k \in \mathcal{K}^{+}(r)$, as shown by equation (5b) in the case standard MNL acreage share choice models (ensuring that $\mathbf{s}_{it}(r)$ necessarily belong to regime *r*). Of course, $s_{k,it}(r)$ is almost null when crop *k* is much less profitable than the other crops of the considered regime.[3] This implies that corner solutions are handled in a specific way in the MNL modelling framework: their characterization doesn't rely on the qualification conditions related to the acreage non-

[3] The term $s_{k,it}(r)$ decreases in $\pi_{k,it}^{e} - \beta_{k,it}^{s}$ and tends to 0 as the term $\pi_{k,it}^{e} - \beta_{k,it}^{s}$ goes to $-\infty$. The larger α_i is, the more rapidly $s_{k,it}(r)$ decreases in $\pi_{k,it}^{e} - \beta_{k,it}^{s}$.

negativity constraints which would be involved in the case where the cost function $C_{it}(\mathbf{s})$ was chosen to be quadratic in **s**. The acreage share non-negativity constraints never bind in the MNL framework, they just imply that the optimal acreage shares of the least profitable crops of a given crop set are very small. [4] The acreage shares of the least profitable crops only become null at the production regime choice stage, when these crops are excluded from the produced crop set characterizing the chosen production regime.[5]

3. An ERS micro-econometric multi-crop model with regime fixed costs

This section presents the structure of the ERS micro-econometric multi-crop model considered in the empirical application presented in the next section. This model is composed of three equation subsystems describing the yield supply functions, variable input demand functions and the acreage share choice models of each produced crop on the one hand, and of a probabilistic production regime choice model on the other hand. This micro-econometric multi-crop model can be interpreted as an extension, to an ERS framework with regime fixed costs, of the model proposed by Carpentier and Letort (2014). As in Koutchadé et al. (2015) we adopt a random parameter approach for accounting for farmers' unobserved heterogeneity.

The considered ERS micro-econometric multi-crop model is presented in three steps. First, we present the production choice models defined at the crop level, i.e. the yield supply and variable input demand models. Second, we present the acreage share choice models. Finally, we describe the production regime choice model. This presentation is organized according to the structure of the model: the crop level production choice models are used for defining the acreage share choice models which are themselves used for defining the production regime choice model.

Yield supply and variable input demand models. We assume that farmers produce crop k from a variable input aggregate under a quadratic technological constraint. I.e., we assume that the yield of

---

[4] From a technical viewpoint, this property comes from properties of the entropy function: the term $-s_k \ln s_k$ tends to 0 as s_k decreases to 0 (leading to the continuity extension at 0 of these terms and the related convention stating that $s_k \ln s_k = 0$ if $s_k = 0$) while its derivative in s_k tends to $+\infty$ as s_k decreases to 0.

[5] Indeed, if the acreage management cost function $C_{it}(\mathbf{s})$ were chosen to be quadratic in **s** – as in the usual PMP framework or as in the econometric acreage choice model of Guyomard et al (1996), of Moore and Negri (1992) or of Carpentier and Letort (2012) – farmers' acreage choice problem would be defined as quadratic programming problem. It would then be possible to adapt the approaches developed by Wales and Woodland (1980) or by Lee and Pitt (1986) for modelling acreage choices with corner solutions. Following the primal approach of Wales and Woodland (1980), one would define empirically tractable estimating equations – for recovering the parameters of the cost function – based on the first order conditions of quadratic acreage choice problem, including the qualification conditions related to the acreage non-negativity constraints. However, the resulting modelling framework would ignore production regime fixed costs. To account for regime costs would raise significant difficulties as the per regime optimal expected profit functions could only be obtained numerically and would be characterized by salient discontinuities in the parameters to be estimated.

crop k obtained by farmer i in year t is given by:

(6a) $$y_{k,it} = \beta^y_{k,it} - 1/2 \times \gamma^{-1}_{k,i}(\beta^x_{k,it} - x_{k,it})^2$$

where $x_{k,it}$ denotes the variable input use level. The $\gamma_{k,i}$ parameter is required to be (strictly) positive for the production function to be (strictly) concave in $x_{k,it}$. It determines the extent to which the yield supply and the input demand of crop k respond to the input and crop prices. The terms $\beta^y_{k,it}$ and $\beta^x_{k,it}$ have direct interpretations in the considered yield function. The term $\beta^y_{k,it}$ is the yield level that can be potentially achieved by farmer i in year t while $\beta^x_{k,it}$ is the input quantity required to achieve this potential yield level. These parameters are decomposed as:

(6b) $$\beta^y_{k,it} \equiv \beta^y_{k,i} + (\mathbf{a}^y_{k,0})'\mathbf{z}^y_{k,it} + \varepsilon^y_{k,it} \text{ and } \beta^x_{k,it} \equiv \beta^x_{k,i} + (\mathbf{a}^x_{k,0})'\mathbf{z}^x_{k,it} + \varepsilon^x_{k,it}$$

where the terms $\mathbf{z}^y_{k,it}$ and $\mathbf{z}^x_{k,it}$ are observed variable vector used to control for farm heterogeneity. The $\beta^y_{k,i}$ and $\beta^x_{k,i}$ terms are farmer specific parameters aimed at capturing unobserved heterogeneity across farms and farmers. These terms, as well as the $\gamma_{k,i}$ random parameter, mainly capture two kinds of effects: those of the natural and material factor endowment of farms (e.g. soil quality and machinery quality) and those of the skills of farmers. The $\varepsilon^y_{k,it}$ and $\varepsilon^x_{k,it}$ terms are standard error terms aimed at capturing the effects on production of stochastic events (e.g. climatic conditions, and pest and weed problems). We assume that farmer i is aware of the content of $\varepsilon^x_{k,it}$ when deciding his variable input uses.

Assuming that farmer i maximizes the expected return to variable input uses of each crop, we can easily derive the demand of the variable input for crop k:

(7a) $$y_{k,it} = \beta^y_{k,i} + (\mathbf{a}^y_{k,0})'\mathbf{z}^y_{k,it} - 1/2 \times \gamma_{k,i} w^2_{k,it} p^{-2}_{k,it} + \varepsilon^y_{k,it}$$

and the corresponding yield supply:

(7b) $$x_{k,it} = \beta^x_{k,i} + (\mathbf{a}^x_{k,0})'\mathbf{z}^x_{k,it} - \gamma_{k,i} w_{k,it} p^{-1}_{k,it} + \varepsilon^x_{k,it} \ .$$

The terms $p_{k,it}$ and $w_{k,it}$ respectively denote the expected output and input prices of crop k. Assuming that the expectations of the terms $\varepsilon^y_{k,it}$ and $\varepsilon^x_{k,it}$ of farmer i are null at the beginning of the cropping season, this farmer expects the following return to the variable input:

(8) $$\pi^e_{k,it} = p_{k,it}\left(\beta^y_{k,i} + (\mathbf{a}^y_{k,0})'\mathbf{z}^y_{k,it}\right) - w_{k,it}\left(\beta^x_{k,i} + (\mathbf{a}^x_{k,0})'\mathbf{z}^x_{k,it}\right) + 1/2 \times \gamma_{k,i} w^2_{k,it} p^{-1}_{k,it}$$

for crop k when she/he chooses her/his acreage shares.

Acreage share choice models. As discussed in Carpentier and Letort (2014), the (Standard MNL) acreage share choice model given in equation (5b) appears to be rather rigid because it treats the different crops symmetrically. Indeed, arable crops can often be grouped according to their competing for the use of quasi-fixed factors or according to their agronomic characteristics. The ERS micro-econometric multi-crop model considered here contains a '3 level-Nested Multinomial Logit" (NMNL) acreage share choice model, which derives from an entropic acreage management cost function as proposed by Carpentier and Letort (2014). In our setting, the crop set $\mathcal{K}$ is partitioned into G mutually exclusive groups of crops, each group $g \in \{1,...,G\}$ being itself partitioned into $M(g)$ subgroups of crops. The m^{th} subgroup of the g^{th} group is defined as the crop subset $\mathcal{K}(m,g)$. Crops (resp. subgroups) belonging to a same subgroup (resp. group) are assumed to have similar agronomic characteristics and to compete more for farmers' limiting quantities of quasi-fixed factors than they compete with crops (resp. subgroups) of other subgroups (resp. of other groups). The three level nested structure of the crop set used in the empirical application is depicted in Figure 1. The corresponding acreage management cost function is given by:

(9)
$$\begin{aligned} C_{it}(\mathbf{s}) = \sum\nolimits_{k\in\mathcal{K}} s_k \beta^s_{k,it} + \sum\nolimits_{g=1}^{G} \alpha_i^{-1} s_{(g)} \ln s_{(g)} + \sum\nolimits_{g=1}^{G} s_{(g)} \alpha_{(g),i}^{-1} \sum\nolimits_{m=1}^{M(g)} s_{m|(g)} \ln s_{m|(g)} \\ + \sum\nolimits_{g=1}^{G} s_{(g)} \sum\nolimits_{m=1}^{M(g)} s_{m|(g)} \alpha_{m|(g),i}^{-1} \sum\nolimits_{\ell\in\mathcal{K}(m,g)} s_{\ell|(m,g)} \ln s_{\ell|(m,g)} \end{aligned}$$

where $s_{(g)}$ denotes the acreage share of group *g*, $s_{m,(g)}$ that of subgroup *m* in group *g*, and $s_{k|(m,g)}$ that of crop *k* in the subgroup *m* of group *g*. The α_i, $\alpha_{(g),i}$ and $\alpha_{m|(g),i}$ are farm specific parameters determining the flexibility of farmers' acreage choices.[6] The larger they are, the more the acreage share choice respond to economic incentives (because the less management costs matter). The condition $\alpha_{m|(g),i} \geq \alpha_{(g),i} \geq \alpha_i > 0$ is sufficient for the cost function $C_{it}(\mathbf{s})$ to be strictly convex in **s**.

The linear terms of the cost function $C_{it}(\mathbf{s})$ are decomposed as:

(10) $$\beta^s_{k,it} \equiv \beta^s_{k,i} + (\mathbf{a}^s_{k,0})'\mathbf{z}^s_{k,it} + \varepsilon^s_{k,it}$$

where $\mathbf{z}^s_{k,it}$ are explanatory variable used to control for observed heterogeneous factors. The $\beta^s_{k,i}$ farm specific factors account for heterogeneity effects unobserved in the data. The error terms $\varepsilon^s_{k,it}$ capture the effects of stochastic variation of the cost due to random events such as climatic events impacting the soil state at planting. The content of these terms are assumed to be known to farmers when they choose their acreages.

These error terms are assumed to be independent from the error terms of the yield and input

[6] We have $\alpha_{g,m,i} = \alpha_{g,i}$ if subgroup *m* contains a single crop. Similarly, we have $\alpha_{g,i} = \alpha_i$ if group *g* contains a single subgroup.

demand functions, $\varepsilon_{k,it}^{y}$ and $\varepsilon_{k,it}^{x}$. I.e., we assume that the potential links between the error terms of the acreage choice model on the one hand and those of the yield supply and input demand functions on the other hand are negligible. To relax this assumption is possible but significantly increases the estimation burden. In a similar context, Koutchadé *et al* (2015) found that the $\varepsilon_{k,it}^{s}$ error terms were not significantly correlated with the error terms $\varepsilon_{\ell,it}^{y}$ and $\varepsilon_{\ell,it}^{x}$.

The optimal acreage share choices of farmers as given by equation (2a) can be derived for any production regime. It suffices to solve the maximization problem given in equations (3). For instance, ten acreage share subsystems are considered in our empirical application, one for each production regime represented in the data. Of course, the functional form of the derived acreage choice function depends on the subset of crops included in the considered regime. Assuming that crop k belongs to the m^{th} subgroup of the g^{th} group, we obtain:

$$\text{(11a)} \quad s_{k,it}(r) = j_k(r)\frac{\exp(\alpha_{g,m,i}\pi_{k,it})\left(\theta_{(m,g),it}(r)\right)^{\alpha_{g,i}\alpha_{g,m,i}^{-1}-1}\left\{\sum_{n=1}^{M(g)}\left(\theta_{(n,g),it}(r)\right)^{\alpha_{(g),i}\alpha_{n|(g),i}^{-1}}\right\}^{\alpha_i\alpha_{g,i}^{-1}-1}}{\sum_{h=1}^{G}\left\{\sum_{n=1}^{M(h)}\left(\theta_{(n,h),it}(r)\right)^{\alpha_{(h),i}\alpha_{n|(h),i}^{-1}}\right\}^{\alpha_i\alpha_{h,i}^{-1}}}$$

and:

$$\text{(11b)} \quad \Pi_{it}(r) = \alpha_i^{-1}\ln\sum_{h=1}^{G}\left\{\sum_{n=1}^{M(h)}\left(\theta_{(n,h),it}(r)\right)^{\alpha_{(h),i}\alpha_{n|(h),i}^{-1}}\right\}^{\alpha_i\alpha_{(h),i}^{-1}}$$

where:

$$\text{(11b)} \quad \begin{cases} j_k(r)=1 \text{ if } k\in\mathcal{K}^{+}(r) \\ j_k(r)=0 \text{ if } k\in\mathcal{K}^{0}(r)\end{cases} \text{ and } \theta_{(n,h),it}(r)=\sum_{\ell\in\mathcal{K}(n,h)} j_\ell(r)\exp\left(\alpha_{n|(h),i}(\pi_{\ell,it}-\beta_{\ell,it}^{s})\right).$$

Production regime choice model. Observing that the regime r optimal acreage choice $\mathbf{s}_{it}(r)$ necessarily belongs to regime r in the MNL case considered here, the regime specific expected profit levels can easily be used for defining a regime choice model according to the choice problem described in equation (4). Let define the regime fixed costs as $F_{it}(r) = f_i^r - \sigma_i^{-1}\varepsilon_{it}^{r}$. The farm specific parameters f_i^r aim at capturing the effects of unobserved factors affecting the regime fixed costs. The error terms ε_{it}^{r} aim at capturing the effects of stochastic factors. They are assumed to be independent from the other elements of the model and to be independently distributed according to a type I extreme value distribution. This implies that the production regime choice is specified as a standard Multinomial Logit discrete choice model.

The probability of the choice of regime *r* – conditionally on σ_i, and on the f_i^r and $\Pi_{it}(r)$ terms for $r \in \mathcal{R}$ – is thus given by:

$$(12) \qquad P[r_{it} = r] = \frac{\exp\left(\sigma_i(\Pi_{it}(r) - f_i^r)\right)}{\sum_{q\in\mathcal{R}} \exp\left(\sigma_i(\Pi_{it}(q) - f_i^q)\right)}.^7$$

The farm specific parameter σ_i is assumed to be positive. This scale parameter allows determining the extent to which the regime specific expected profit levels minus the corresponding fixed cost, i.e. the $\Pi_{it}(r) - f_i^r$ terms, explain the production regime choice as regards to the effects of the ε_{it}^r idiosyncratic terms. The higher σ_i, the more the 'deterministic' terms $\Pi_{it}(r) - f_i^r$ impact the observed regime choices.

3. Parametric specification and estimation procedure

The ERS multi-crop micro-econometric model presented in section 2 is composed of three main parts: a subsystem of yield supply and input demand equations (equations 7), subsystems of acreage share equations (equation 11a) and a probabilistic production regime choice model (equation 12). In this section, we briefly present the econometric procedure used to estimate this model.[8]

Distributional assumptions. Most parameters of the model are farmer specific, which allows accounting for the heterogeneity in the performance levels as well as in the responses to economic incentives of the sampled farmers. Yet, standard data set, even panel data sets, do not permit a direct estimation of each individual parameter: the objective of the estimation here is to characterize the distribution of these parameters in the population described by our sample. To do so, we rely on a random parameter approach, as proposed in Koutchade et al. (2015).

Given the rather complex structure and the size of our model, we adopt a fully parametric framework. Apart from the modelled variables, i.e. the crop yield levels $y_{k,it}$, the crop input use levels $x_{k,it}$, the crop acreage shares $s_{k,it}$ for $k \in \mathcal{K}$, and the production regimes the production regimes r_{it} for $r \in \mathcal{R}$ collected in the vector $\mathbf{c}_{it}$, and the fixed parameters, i.e. the terms $\mathbf{a}_{k,0}^y$, $\mathbf{a}_{k,0}^y$

[7] Note however that the error terms $\varepsilon_{k,it}^s$ of the acreage choice model contained in the expected profit levels $\Pi_{it}(r)$ can only be directly recovered from the data for the crops produced by farmer *i* in year *t*. Indeed, we can recover the vector $\boldsymbol{\varepsilon}_{r,it}^{s,+} = (\varepsilon_{k,it}^s : k \in \mathcal{K}^+(r))$ while the vector $\boldsymbol{\varepsilon}_{r,it}^{s,0} = (\varepsilon_{k,it}^s : k \in \mathcal{K}^0(r))$ cannot. We used Laplace approximations for integrating the expectation of the probability function $P[r_{it} = r]$ over the probability distribution of $\boldsymbol{\varepsilon}_{k,it}^{s,0}$ conditional on $\boldsymbol{\varepsilon}_{k,it}^{s,+}$ (see, e.g., Harding and Hausman, 2007). This expectation is part of the likelihood function of the model.

[8] Specification and estimation details are available from the authors upon request.

and $\mathbf{a}_{k,0}^{s}$ for $k \in \mathcal{K}$ collected in the vector $\mathbf{a}_0$, the considered model contains five subsets of random elements:

The farm specific parameter vectors $\boldsymbol{\delta}_i$ collecting the potential yield parameters $\beta_{k,i}^{y}$, the input requirement parameters $\beta_{k,i}^{x}$, the input use flexibility parameters $\ln\gamma_{k,i}$ and the cost function linear parameters $\beta_{k,i}^{s}$ for $k \in \mathcal{K}$; the acreage choice flexibility parameters $\ln\alpha_i, \ln\alpha_{(g),i}$ and $\ln\alpha_{m|(g),i}$ for $m = 1,...,M(g)$ and $g = 1,...,G$; the regime fixed cost parameters f_i^{r} for $r \in \mathcal{R}$ and the regime choice scale parameter $\ln\sigma_i$. The vectors $\boldsymbol{\delta}_i$ are assumed normally and independently distributed across farms.

The explanatory variable vectors $\mathbf{z}_{it}$ containing the crop prices $p_{k,it}$, the variable input prices $w_{k,it}$ and the control variable vector $\mathbf{z}_{k,it}^{y}$, $\mathbf{z}_{k,it}^{x}$ and $\mathbf{z}_{k,it}^{s}$ for $k \in \mathcal{K}$.

The yield supply and input demand error term vectors $\boldsymbol{\varepsilon}_{it}^{yx}$ containing the error terms $\varepsilon_{k,it}^{y}$ and $\varepsilon_{k,it}^{x}$ for $k \in \mathcal{K}$ and that are assumed normally and independently distributed across farms.

The acreage share error term vectors $\boldsymbol{\varepsilon}_{it}^{s}$ containing the error terms $\varepsilon_{k,it}^{s}$ for $k \in \mathcal{K}$ and that are assumed normally and independently distributed across farms.

The production regime error term vectors $\boldsymbol{\varepsilon}_{it}^{\rho}$ containing the error terms ε_{it}^{r} for $r \in \mathcal{R}$. These terms ε_{it}^{r} are assumed independent across regimes and farms, well as distributed according to a type I extreme value distribution.

We further assume that the error term vectors $\boldsymbol{\varepsilon}_{it}^{yx}$, $\boldsymbol{\varepsilon}_{it}^{s}$ and $\boldsymbol{\varepsilon}_{it}^{\rho}$ are mutually independent, and that the explanatory variables $\mathbf{z}_{it}$ are (*i*) strictly exogenous with respect to these error term vectors and (*ii*) independent of the random parameters $\boldsymbol{\delta}_i$. The vector $\mathbf{z}_{it}$ contains prices, climatic variables and characteristics of the farms' fixed factor endowments. We finally assume that that the error term vectors $\boldsymbol{\varepsilon}_{it}^{yx}$, $\boldsymbol{\varepsilon}_{it}^{s}$ and $\boldsymbol{\varepsilon}_{it}^{\rho}$ are independent across years.

As the explanatory variable vector $\mathbf{z}_{it}$ doesn't contain any lagged endogenous variable, the considered model can be interpreted either as an essentially static model or as a reduced form model as regards the dynamic features of the modelled choices. It is notably difficult to empirically disentangle the effects of farmers' unobserved heterogeneity from those of unobserved persistent dynamic features of the modelled processes (see, e.g., Angrist and Pischke, 2009 or Arellano and Bonhomme, 2012). For instance, the random parameters $\boldsymbol{\delta}_i$ are likely to capture the effects on farmers' production choices and performances of the stable crop rotation schemes that these farmers' seem to rely on.[9] Our assuming that the error term vectors $\boldsymbol{\varepsilon}_{it}^{yx}$, $\boldsymbol{\varepsilon}_{it}^{s}$ and $\boldsymbol{\varepsilon}_{it}^{\rho}$ are serially independent across years is mostly based on this hypothesis. We do not assume that farmers' choices and performances are not significantly impacted by unobserved dynamic features. But we assume that these dynamic features are sufficiently persistent for being mostly captured by the random parameters of our model. Of course, dynamic features of crop production and of farmers' choice are important topics. But these are also difficult ones. Their empirical investigation with models involving corner solutions in acreage choices and regime fixed costs is left for further research.

Estimation. The aim of the estimation procedure is to obtain statistical estimates of two parameter sets: the fixed parameters and the elements of variance matrices of the error term vectors of the model – collected in the vector $\boldsymbol{\theta}_0$ – on the one hand, and the parameters of the probability distribution of its random parameter vector – collected in the vector $\boldsymbol{\eta}_0$. We use a Maximum Likelihood (ML) estimator computed via an Expectation Maximization (EM) algorithm (Dempster et al. (1977) for estimating $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$.

Let the function $g(\mathbf{u} \mid \mathbf{v}; \boldsymbol{\lambda})$ generically define the probability distribution function of the random vector $\mathbf{u}$ conditional on the random vector $\mathbf{v}$ parameterized by $\boldsymbol{\lambda}$. The likelihood function at $(\boldsymbol{\theta}, \boldsymbol{\eta})$ of $\mathbf{c}_i$ conditional on $\mathbf{z}_i$ is given by $\ell_i(\boldsymbol{\theta}) = \int g(\mathbf{c}_i \mid \mathbf{z}_i, \boldsymbol{\delta}_i; \boldsymbol{\theta}) g(\boldsymbol{\delta}_i; \boldsymbol{\eta}) d\boldsymbol{\delta}_i$.[10] This likelihood function can be obtained neither analytically nor numerically. But it can be estimated via simulation simulated methods for computing Simulated ML estimators of $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$. Albeit possible, the empirical implementation of this approach is difficult due to the dimension of our parameter of interest and to the rather complex functional form of the likelihood function $\ell_i(\boldsymbol{\theta})$.[11]

We choose to compute our ML estimator via an EM algorithm because EM algorithms are particularly well suited for estimating models with missing variables such as random parameters. The aim of these algorithms replace a large ML maximization problem by a sequence of simpler

---

[9] See, e.g., Koutchadé et al (2015) for an empirical analysis providing arguments confirming this hypothesis.
[10] The assumed serial independence of the model error terms imply that $g(\mathbf{c}_i \mid \mathbf{z}_i, \boldsymbol{\delta}_i; \boldsymbol{\theta}) = \prod_{t \in \mathcal{H}(i)} g(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \boldsymbol{\delta}_i; \boldsymbol{\theta})$ where $\mathcal{H}(i)$ is the observation history of farm *i*.
[11] In particular, the probability distribution functions $g(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \boldsymbol{\delta}_i; \boldsymbol{\theta})$ cannot be obtained in analytical closed form. These functions contain the probability functions at $\boldsymbol{\theta}$, denoted by, of the regime choice of farmer *i* in year *t* (r_{it}), conditional on the random parameters ($\boldsymbol{\delta}_i$), explanatory variables ($\mathbf{z}_{it}$) and on the acreage shares of the produced crops ($\mathbf{s}_{it}^{+}$), i.e. the terms $g(r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\delta}_i, \mathbf{s}_{it}^{+}; \boldsymbol{\theta})$. These probability functions are computed as expectations over the joint probability distributions function of the terms $\varepsilon_{k,it}^{s}$ for $k \in \mathcal{K}^0(r_{it})$. These error terms are arguments of the regime expected profit levels $\Pi_{it}(r)$ that are themselves arguments of the production regime choice. But they must be integrated out in the model likelihood function because they cannot be recovered from the data. We use Laplace approximations for computing the expectations yielding the terms $g(r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\delta}_i, \mathbf{s}_{it}^{+}; \boldsymbol{\theta})$ (see, e.g., Harding and Hausman, 2007, for an application of this integration approach in a related context).

problems. EM algorithms iterate an Expectation (E) step and a Maximization (M) step until numerical convergence. In our case, the E step consisting of computing the expectation of the probability distribution functions $g(\mathbf{c}_i \mid \mathbf{z}_i, \boldsymbol{\delta}_i; \boldsymbol{\theta})$ and $g(\boldsymbol{\delta}_i; \boldsymbol{\eta})$ conditional on the observed choices $\mathbf{c}_i$ and explanatory variables $\mathbf{z}_i$ according to the probability distribution functions obtained from the preceding iteration. Denoting the last available estimates of $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ by $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$, this step aims at obtaining the 'modified' log-likelihood functions given by $\hat{\ell}_i(\boldsymbol{\theta}) = \int g(\mathbf{c}_i \mid \mathbf{z}_i, \boldsymbol{\delta}; \boldsymbol{\theta}) g(\boldsymbol{\delta} \mid \mathbf{c}_i, \mathbf{z}_i; \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) d\boldsymbol{\delta}$ and $\hat{\ell}_i(\boldsymbol{\eta}) = \int g(\boldsymbol{\delta} \mid \boldsymbol{\eta}) g(\boldsymbol{\delta} \mid \mathbf{c}_i, \mathbf{z}_i; \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) d\boldsymbol{\delta}$. The M step then consists of maximizing in the interest parameters $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ the sample modified log-likelihood functions obtained from the E step. I.e. it consists of solving the problems $\max_{\boldsymbol{\theta}} N^{-1} \sum_{i=1}^{N} \ln \hat{\ell}_i(\boldsymbol{\theta})$ and $\max_{\boldsymbol{\eta}} N^{-1} \sum_{i=1}^{N} \ln \hat{\ell}_i(\boldsymbol{\eta})$. This is where EM algorithms take advantage of the likelihood functions of random parameter models in general. Our model being composed of four equation sub-systems, the structure of the terms $\hat{\ell}_i(\boldsymbol{\theta})$ allows for further splitting the maximization problem $\max_{\boldsymbol{\theta}} N^{-1} \sum_{i=1}^{N} \ln \hat{\ell}_i(\boldsymbol{\theta})$ into a few simpler maximization problems.

However, the 'modified' likelihood functions $\hat{\ell}_i(\boldsymbol{\theta})$ and $\hat{\ell}_i(\boldsymbol{\eta})$ can be integrated neither analytically nor numerically. Stochastic versions of the EM algorithm – i.e. the so-called SEM or Monte Carlo EM (MCEM) algorithms – have been proposed in the computational statistics literature for combining the advantages of the EM algorithm and simulation methods (see, e.g., McLachlan and Krishnan, 2008). In this study, we rely on the SEM algorithm proposed by Delyon et al. (1999) combined with Importance Sampling simulation methods proposed in this context by (Caffo et al., 2005) for estimating the 'modified' likelihood functions $\hat{\ell}_i(\boldsymbol{\theta})$ and $\hat{\ell}_i(\boldsymbol{\eta})$.

Calibration of the simulation model. The estimated ERS multi-crop micro-econometric model can be used for "statistically calibrating" its random parameters for each farm of our sample and thus for obtaining a simulation model consisting of a sample of farm specific "calibrated" models (see, e.g., Koutchade et al., 2015). The underlying idea of this procedure is to use the estimated distribution of the random parameters and farmers' observed choices compute estimates of the farm specific parameters according to a "Tell me what you did, I will tell you who you are" logic.

Interestingly, the Expectation step of the SEM algorithm we use relies on computations closely related to this calibration procedure since both rely on the probability distributions of the random parameters $\boldsymbol{\delta}_i$ conditional on the observed choices $\mathbf{c}_i$ and explanatory variables $\mathbf{z}_i$. In this study, the specific parameter $\boldsymbol{\delta}_i$ of farm *i* is calibrated as the mode – i.e. according to a ML 'calibration' criterion – of its simulated probability distribution conditional on $(\mathbf{c}_i, \mathbf{z}_i)$, i.e. on what is known about farm *i* in the data. Also, this 'statistical calibration' procedure and its counterpart in the SEM algorithm allow for calibrating the random parameters corresponding to crops that have not been grown by the considered farmer or corresponding to regime fixed costs for regimes that have not been chosen by the considered farmer.

3. Empirical application

Data. The model is estimated on an unbalanced panel data set containing 2871 observations of 778 French grain crop producers in the North and North-East of France, over the years 2006 to 2011. This sample has been extracted from data provided by an accounting agency located in the French territorial division *La Marne*. It contains detailed information about crop production for each farm (acreages, yields, input uses and crop prices at the farm gate). We consider seven crops (or crop aggregates): sugar beet, alfalfa, peas, rapeseed, winter crops (wheat and barley), corn and spring barley, which represent more than 80% of the total acreage in the considered area.[12]

The variable input aggregate account for the use of fertilizers, pesticides and seeds. The corresponding price index is computed as a standard Tornqvist index. When a farmer doesn't produce a crop the corresponding output and input prices are unobserved. These missing prices were estimated by the yearly average of the corresponding observed prices. All prices are deflated by the hired production services price index (base 1 in 2006) obtained from the French department of Agriculture. This aggregated price index mainly depends on the price indices of machinery, fuel and hired labor, the main inputs involved in the implicit acreage management cost function.

Figure 1 depicts the three levels nesting structure that we adopt for the seven crops. In a first level we distinguish a cereal group composed of winter crops, corn and spring barley, and a group of crops that are generally planted at the head of rotation: corn, alfalfa, peas and rapeseed. This structure is intended to reflect the basic rotation scheme of grain producers in France. In a second level, the cereal group is split into two subgroups: winter cereals on the one hand and other cereals on the other hand, in order to account for the differences in planting seasons between those cereals. The 'head of rotation' group is split into an 'oilseeds and protein crops' subgroup and a subgroup

[12] The EU sugar beet subsidy scheme requires limited adjustments in our application. This scheme scheme is based on sugar beet production quotas – held by farmers on a historical basis – with subsidized prices. Yet, the actual sugar beet production largely exceeds the subsidized quota for all sugar beet producers of our sample. This suggests that these farmers choose their sugar beet acreages, yield levels and input use levels according to the off-quota sugar beet prices.

including only sugar beet (the only root crop considered here).

Table 1 provides descriptive statistics concerning the production regimes observed in the data. Based on these seven crops, 127 regimes, could theoretically be grown by farmers. Only 10 of them are actually observed in our sample. All farmers grow winter crops and at least two additional crops. The most frequent regimes in the sample (regimes 2, 3 and 4) actually include five or six crops. It is interesting to note that most farmers adopt different production regimes over the 5 years of our sample: only 11 out of 778 farmers have not changed their production regime during the period. The average gross margins associated to each regime are reported in the last column of Table 1. An interesting feature appears here: the most frequently chosen regimes are not the ones that lead on average to the highest gross margin. For instance, regime 8 – which excludes peas, rapeseed and corn – is characterized by the highest observed gross margin, but has been adopted in only 2.8% of the observations. This comes to illustrate the fact that farmers' choices of production regime are driven by other factors than gross returns, such as the acreage management and regime fixed costs represented in our model.

Estimation results. The parameter estimates of our model are not reported here due to space limitation: we only provide some insights of these results.[13] The expectations of all the random parameters are precisely estimated, their values lie in reasonable ranges and they have expected signs for all crops and crop groups, with notably expected ranges of the acreage flexibility parameters ($\alpha_{g,m,i} \geq \alpha_{g,i} \geq \alpha_i > 0$). The variance parameters of the random parameter distributions are all statistically different from zero. This indicates that the technical and behavioral parameters of our model significantly vary across farms, despite the fact that we control for observed factors characterizing farm heterogeneity (land and capital endowments and climate). Finally, the mean estimated value of the scale parameter σ_i in the regime choice model (13) being relatively large (3.40), the profit and regime fixed costs in the regime choice appear to be significant drivers of the regime choices.

Estimated criteria tend to show that the proposed model offers a satisfactory fit to our data, with a better fit to the major crops than to the ones less frequently produced or with smaller acreages (peas, corn and, to a lesser extent, alfalfa). Importantly, our investigations on this issue tend to demonstrate that our results are robust to various distributional assumptions related to the model random parameters. Random parameter variations account for a significant part of the observed variations in farmers' choices. But, variations in the model error terms account for a comparable part. Even if

[13] Detailed estimation results are available from the authors upon request.

crop production is known to be significantly affect by numerous random events, this indicates that there is still room for further improving the explanatory power of our model. We pursue our current research in that direction, by looking for additional control variables in particular.

Simulation results. The structure of the proposed ERS multi-crop micro-econometric model allows for investigating the relative importance of the main drivers of the production regime choices. For that purpose we consider the simulation model obtained from the estimated one by calibrating the farm specific parameters for each farm of our sample. Then we use this simulation model for investigating the prediction power of three elements of the regime choice models: the weighted sum of the expected crop gross returns $\mathbf{s}_{it}(r)'\boldsymbol{\pi}_{it}^{e}$, the acreage management costs $C_{it}\left(\mathbf{s}_{it}(r)\right)$ and the regime fixed costs f_i^r for $r \in \mathcal{R}$. We simulate the regime choices according to each of these elements as well as to combinations of these elements, and then confront them, on average, with the observed regime choices. Taken together these simulation results confirm that the regime fixed costs matter, but mainly in combination with the other drivers of the regime choice model. Maximization of the gross margins, of the acreage management costs or of the regime fixed cost alone leads to very poor predictions of regime choices. Considering pairs of these choice criteria only slightly improve the predictions, while considering together these three criteria unsurprisingly provides predicted choices very close, on average, to the observed ones.

The estimated regime costs tend to slightly decrease as the crop number decrease from 7 to 5 crops. They don’t follow any clear pattern below 5 crops. Moreover, the fact that these costs appear to significantly vary across farms complicates their analysis and interpretation. This point clearly deserves further research. To be completed: these fixed costs are highlighted in the article.

To illustrate the relevance of the approach we propose to deal with corner in acreage choices, we simulate the impacts of changes in expected crop prices on acreage choices. Acreage price elasticities play a crucial role in this type of exercise. Yet these elasticities account both for the impact of crop prices on acreages within any given regime and for the switch in production regimes induced by crop price changes. These two effects can be distinguished by generalizing to multiple regimes the decomposition originally proposed by McDonald and Moffit (1980) in the case of a Tobit model. The average acreage own price elasticities of our farm sample are reported in Table 2. They have expected signs and, because of the crop disaggregation level of our data, are larger than those commonly found in the literature. The decomposition of these elasticities shows that a large part of the price effects on acreages can be due to the inclusion or not of these crops in the production regimes chosen by farmers. For crops like corn or peas with small overall acreage shares, changes in the production regimes account for about one third of the estimated price elasticities. However, changes in the production regimes can also be substantial for frequently produced crops. For instance, they account for a quarter of the sugar beet acreage own price elasticities.

The impact of the production regime choice is further highlighted by simulating the effects of increases in the price of peas on the acreages of the crop. Owing to their fixing atmospheric nitrogen for themselves as well as for their following crops in the rotation this crop is often considered as 'diversification crop' of significant interest. Yet, protein peas acreages have declined over the last decade in the considered area mostly because of lacking profitability, as regards to that of the other rotation heads in particular.[14] The simulated impacts of increases in the price of peas on acreages are reported on Figure 2. According to our results, a 50% increase in the price of peas would increase the average peas acreage share by 1.7%, from 1.1% to 2.8%. These additional peas acreages would mainly replace those of other rotation heads: the average combined acreage share of rapeseed, alfalfa and sugar beet would decrease by around 1.2% while that of cereals would only decrease by around 0.5%. This illustrates the interests in considering the crop – agronomic and management – characteristics when specifying the acreage management cost function. This also suggests that the increase in the rapeseed price due to the EU support to bio-fuels has played significant role in the decrease in the peas acreages in the considered area.

Interestingly, about two thirds of the increase in the peas acreage would be due to new producers. This also explains another feature of our simulation results. The simulated increases in the peas acreage in not linear in the price of peas: in particular, the increase in the peas acreages is more pronounced above the 20% price increase than below. Figure 3 shows that the adoption rates of the production regimes including protein peas, regime 2 in particular, have similar patterns. This is partly explained by the threshold effects generated by the production regime fixed costs.

4. Concluding remarks

The main aims of this article are twofold. First, it presents an original modelling framework for dealing with corner solutions in multi-crop micro-econometric models. This framework is based on the ERS approach, implying that it is fully consistent from an economic viewpoint. It also explicitly considers regime fixed costs. These features make the proposed ERS multi-crop micro-econometric models suitable for analysing, and to some extent for disentangling, the effects of the main drivers of farmers' acreage choices at disaggregation levels at which the corner solution issue is pervasive.

---

[14] In other parts of France the extension of a soil infection (*aphanomyces*) severely impacts peas yields and explains the decrease peas acreages. The diversified cropping systems used in *La Marne* seems to limit this extension.

For instance, our estimation and simulation results and the structure of the considered model tend to demonstrate the expected crop returns are not the sole significant drivers of farmers' acreage choices, at least in the short run.

Second, the application presented in this article illustrates the empirical tractability of ERS models, of random parameter ERS models in particular, for investigating farmers' production choices. Of course, to estimate such models raises challenging issues. But, this is also necessary for estimating structured micro-econometric models suitably accounting for important features characterizing micro-economic agricultural production data, among which significant unobserved heterogeneity. In particular, to estimate such models enables analysts to calibrate simulation models consisting of samples of farm specific models. According to our experience, ML estimators computed with SEM or MCEM algorithms combining simulation methods appear to be interesting alternatives to Simulated ML estimators for relatively large systems of interrelated equations such as the random parameter ERS models considered in our empirical application.

Of course, significant specification and estimation issues remain to be addressed for ERS multi-crop micro-econometric models such as ours to meet the needs of the agricultural production economist community. However, as fostering crop diversification tend to become an important agri-environmental objective in many countries, including those of the European Union, the modelling framework proposed in this article can be seen as a first step in the right direction.

5. References

Angrist, J. D., & Pischke, J. S. (2009). Instrumental variables in action: sometimes you get what you need. Mostly harmless econometrics: an empiricist's companion, 113-220.

Arellano, M., & Bonhomme, S. (2012). Identifying distributional characteristics in random coefficients panel data models. The Review of Economic Studies, 79(3), 987-1020.

Arndt, C., Liu, S. and Preckel P.V. (1999). On Dual Approaches to Demand Systems Estimation in the Presence of Binding Quantity Constraints. *Applied Economics*, 31(8): 999-1008.

Carpentier, A. and Letort, E. (2014). Multicrop models with MultiNomial Logit acreage shares. *Environmental and Resource Economics*, 59(4): 537-559.

Chakir, R., Bousquet, A. and Ladoux, N. (2004). Modeling Corner Solutions with Panel Data: Application to the Industrial Energy Demand in France. *Empirical Economics*, 29(1): 193-208.

Delyon, B., Lavielle, M. and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, 27(1): 94-128.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1): 1-38.

Fezzi, C. and Bateman, I.J. (2011). Structural agricultural land use modelling for spatial agro-environmental policy analysis. American Journal of Agricultural Economics, 93(4): 1168-1188.

Guyomard, H., Baudry, M., & Carpentier, A. (1996). Estimating crop supply response in the presence of farm programmes: application to the CAP. European Review of Agricultural Economics, 23(4), 401-420.

Harding, M.C. and Hausman, J. (2007). Using Laplace approximation to estimate the random coefficients Logit model by nonlinear least squares. *International Economic Review*, 48(4): 1311-1328.

Kao, C., Lee, L.-F. and Pitt, M.M. (2001). SML Estimation of the Linear Expenditure System with Binding Non-negativity Constraints. *Annals of Economics and Finance*, 2: 203-223.

Koutchadé, P., Carpentier, A. and Femenia, F. (2015). *Empirical modeling of production decisions of heterogeneous farmers with random parameter models*. SMART-LERECO Working Paper WP15-10

Lacroix, A. and Thomas, A. (2011). Estimating the Environmental Impact of Land and Production Decisions with Multivariate Selection Rules. *American Journal of Agricultural Economics*, 93(3): 780-798.

Lee, L.-F. and Pitt, M.M. (1986). Microeconometric demand systems with binding nonnegativity constraints: the dual approach. *Econometrica*, 54(5): 1237-1242.

McLachlan G. and T. Krishnan, 2008. The EM algorithm and extensions. 2nd Ed. Wiley Ed

Moore, M. R., & Negri, D. H. (1992). A multicrop production model of irrigated agriculture, applied to water allocation policy of the Bureau of Reclamation. Journal of Agricultural and Resource Economics, 29-43.

Platoni, S., Sckokai P. and Moro, D. 2012. Panel Data Estimation Techniques and Farm-level Data Models. *American Journal of Agricultural Economics,* 94(4): 1202-1217.

Sckokai, P. and Moro, D. (2009). Modelling the impact of the CAP Single Farm Payment on farm investment and output *European Review of Agricultural Economics*, 36(3): 395-423.

Sckokaï, P. and Moro, D. (2006). Modelling the Reforms of the Common Agricultural Policy for Arable Crops under Uncertainty. *American Journal of Agricultural Economics*, 88(1): 43-56.

Wales, T.J. and Woodland, A.D. (1983). Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints. *Journal of Econometrics*, 21: 263-285.

Wooldridge, J.M. (2009). *Econometric analysis of cross section and panel data. Second edition*. Cambridge, MA: The MIT Press.

6. Tables and Figures

Table 1: Descriptive statistics of the production regimes represented in the data

Regime number	Crops produced in the regime							Regime frequency	Average gross margin (€/ha)[d]
	Winter crops	Corn	Spring Barley	Sugar beet	Alfalfa	Peas	Rape-seed		
1								5.3%	767
2								16.6%	797
3								10.9%	851
4								43.8%	884
5								4.4%	868
6								4.3%	719
7								6.6%	870
8								2.8%	997
9								2.8%	765
10								2.5%	648
Average acreage share[a]	38.6%	2.6%	18.7%	14.7%	8.9%	1.1%	15.4%		
Production frequency	100%	29%	97%	90%	79%	22%	98%		

a Standard deviation in parentheses. b Winter wheat. c. Off-quota price. d. Sugar beet subsidies excluded.

Table 2: Own price elasticities of acreages computed at sample average

	Winter cereals	**Corn**	**Spring Barley**	**Sugar beet**	**Alfalfa**	**Protein peas**	**Rapeseed**
Overall elasticitiy	0.51	3.84	0.82	2.72	0.95	1.50	1.06
Part of the elasticity due to:							
Acreage changes within regime	0.50	2.94	0.75	2.19	0.78	0.91	1.00
Change in production regime	0.00	0.90	0.07	0.53	0.17	0.59	0.06

Figure 1: Nesting structure of the acreage choice model

Groups	**Cereals**			**Rotation heads**			
Subgroups	***Winter cereals***	***Spring cereals***		***Oil and protein crops***			***Root crops***
Crops	Winter cereals*	Spring barley	Corn	Rapeseed	Protein peas	Alfalfa	Sugar beet

* Winter wheat (mostly) and winter barley

Figure 2: Simulated impacts of changes in the price of peas on acreage shares

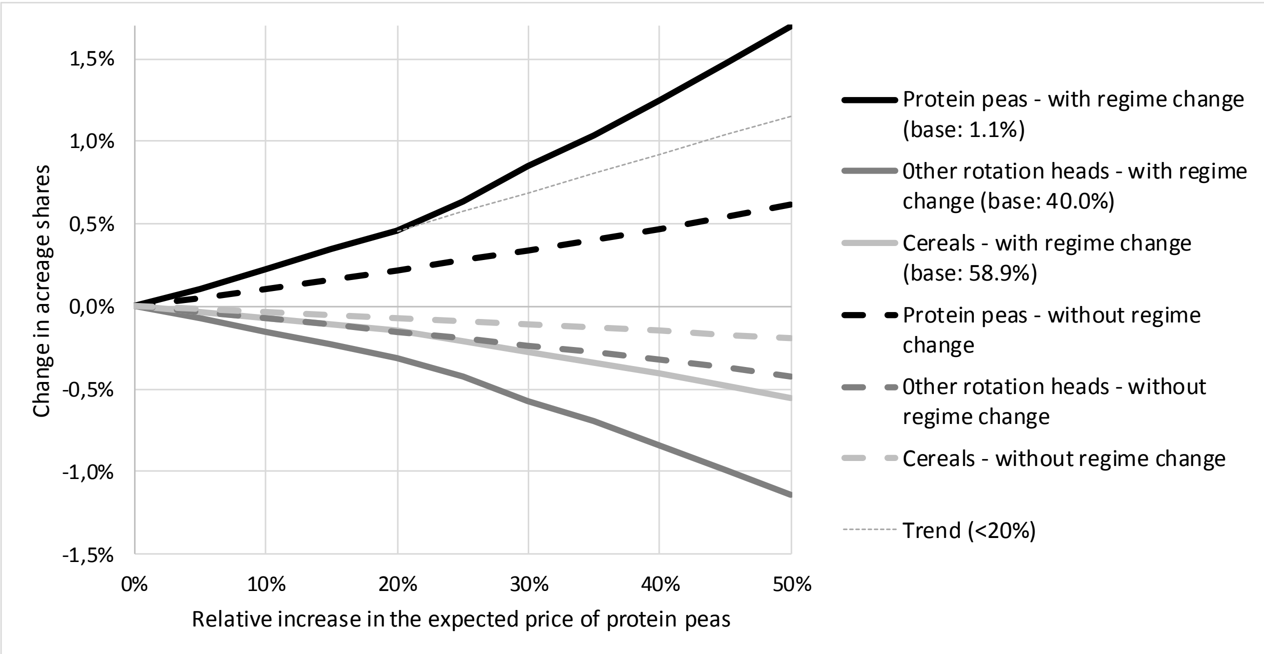
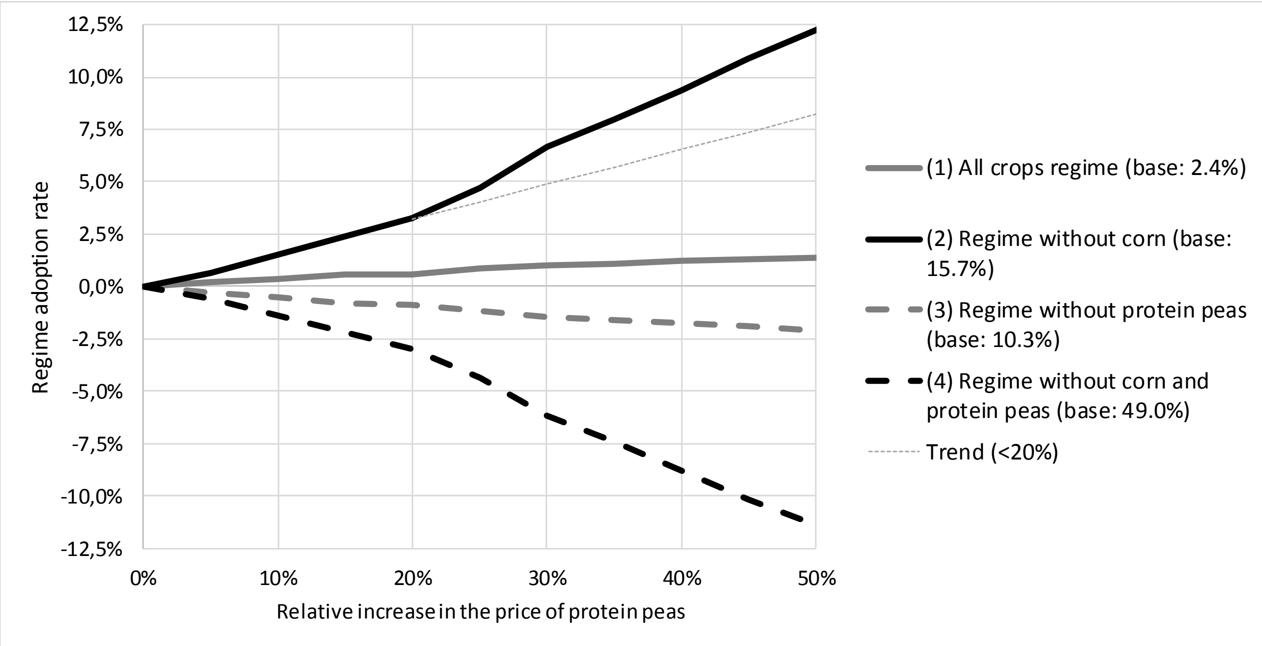


Figure 3: Simulated impacts of changes in the price of peas on regime adoption rates


12,5%
10,0%
7,5%
5,0%
2,5%
0,0%
-2,5%
-5,0%
-7,5%
-10,0%
-12,5%
Regime adoption rate
0%
10%
20%
30%
40%
50%
Relative increase in the price of protein peas
(1) All crops regime (base: 2.4%)
(2) Regime without corn (base: 15.7%)
(3) Regime without protein peas (base: 10.3%)
(4) Regime without corn and protein peas (base: 49.0%)
Trend (<20%)