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

Philippe KOUTCHADE, Fabienne FEMENIA, Alain CARPENTIER

INRA, UMR SMART-LERECO, Rennes

Abstract

Multi-crop micro-economic models aim to describe the production choices of field crop producers that are used for analyzing the effects of agri-environmental policies. These models consist of equation systems describing the yield, variable input quantities and acreages for each crop produced by farmers. Corner solution problems are pervasive in micro-econometric acreage choice models because farmers rarely produce the same crop set in a considered sample. These crop choice problems raise significant modelling issues. The main aim of this paper is to propose an endogenous regime switching model specifically designed for empirically modeling acreage choices with corner solutions. Contrary to models based on censored regression systems that have been used until now, this model is fully coherent from a micro-economic point of view. It also includes regime fixed costs accounting for unobserved costs which only depend on the set of crops grown simultaneously.

We illustrate the empirical tractability of this model by estimating an endogenous regime switching multi-crop model for a panel dataset of 415 French field crop producers followed from 2006 to 2011. The estimated model considers yield supply, variable input demand and acreage share models for 7 crops and considers 8 production regimes. It also accounts for unobserved heterogeneity in farmers' behaviors through the specification of random parameters. We assume that most model – intercept and slope – parameters are farmer specific and estimate their distribution across the farmers' population represented by our sample.

The likelihood function functional form and parameter dimension make the Simulated ML (SML) estimation approach particularly challenging. We consider a multivariate random parameter model with endogenous regimes. This model includes from 10 to 22 interrelated production choices per observation, depending on which of the production regime chosen by the considered farmer in the considered year, and contains 43 random parameters.
While a SML approach would require directly solving a very difficult simulated log-likelihood maximization problem, stochastic versions of Expectation-Maximization (EM) type algorithms of Dempster et al (1977) allow to obtain more easily estimators that are asymptotically equivalent to SML estimators. These algorithms replace a log-likelihood maximization problem by a sequence of simpler ones.
EM type algorithms are particularly well suited for estimating random parameter models and models involving missing variables (the yield and variable input use levels of the crops that are not produced in our case). We used a Stochastic Approximate EM (SAEM) algorithm for its efficient

use of random parameter simulations (Delyon et al 1999). We also relied on the EM algorithm proposed by Ruud (1991) for coping with missing dependent variables in Gaussian Seemingly Unrelated Regression systems as well as on the Laplace approximation proposed by Harding and Hausman (2007) for estimating the production regime choice probabilities of our model.

Our results show that (*a*) our multi-crop micro-econometric model provide a satisfactory fit to our dataset, (*b*) farmers' responses to economic incentives display substantial heterogeneity that cannot be controlled by available variables, (*c*) farmers' acreage choice mechanisms strongly depend on the production regime in which this choice takes place and (*d*) production regime decisions play a major role when considering the aggregated effects of agri-environmental policies.
Our results also demonstrate that SAEM algorithms offer interesting alternatives to SML estimators for estimating relatively large and complicated micro-econometric models.

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

Introduction

Multi-crop micro-economic models aim to describe the production choices of field crop producers that are used for analyzing the effects of agri-environmental policies. These models consist of equation systems describing the yield, variable input quantities and acreages for each crop produced by farmers. Corner solution problems are pervasive in micro-econometric multi-crop models because farmers rarely produce the same crop set in a considered sample, even in samples considering specialized farms. These problems raise significant modelling issues. The numerous null crop acreages observed in farms' level datasets must be modelled as choices of farmers. Yield and input use level are not observed for the crops not produced by the farmers. As these unobserved netput levels might explain why farmers do not produced some crops, endogenous selection issues arise.

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., Lacroix and Thomas 2011, Fezzi and Bateman 2011, 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. Arndt et al (1999) made this point for consumer demand systems. 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 modelling consumer demand systems (see, *e.g.*, Kao et al 2001) or firm input demand systems (see, *e.g.*, Chakir and Thomas 2003), following the pioneering works of Wales and Woodland (1983) and of Lee and Pitt (1986). However, these models have rarely been used in practice, probably because their estimation is challenging, and despite the development of estimation procedures with simulation methods.

The main aim of our paper is to propose an ERS model specifically designed for empirically modeling acreage choices with corner solutions. This model defines farmers' production choice models as resulting from a profit maximization problem. It is fully coherent from a micro-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) 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 Koutchade et al (2018), this model accounts for unobserved heterogeneity in farmers' behaviors through the specification of random parameters. *I.e.* we assume that most model parameters, those defining responses to economic incentives in particular, are farmer specific and estimate their distribution across the farmers' population represented by our sample.

Given that our model is fully parametric, its parameter can be efficiently estimated based on a Maximum Likelihood (ML) estimation framework. The functional form of its likelihood function and the dimension of its parameter vector make the Simulated ML (SML) estimation approach particularly challenging. We consider a multivariate random parameter model with endogenous regimes. Our model considers from 10 to 22 interrelated production choices per observation, depending on which of the 8 production regimes present in our data is chosen by the considered farmer in the considered year, and contains 43 random parameters.

While a SML approach would require directly solving a very difficult simulated log-likelihood maximization problem, stochastic versions of Expectation-Maximization (EM) type algorithms of Dempster et al (1977) allow to obtain more easily estimators that are asymptotically equivalent to SML estimators (see, *e.g.*, MacLachlan and Krishnan 2007, Lavielle 2014). These algorithms basically replace a log-likelihood maximization problem by a sequence of simpler ones. They are particularly well suited for estimating random parameter models and models involving missing variables. We used a Stochastic Approximate Expectation-Maximization (SAEM) algorithm (Delyon et al 1999).

Importantly, once their probability distribution has been estimated, the farmer specific parameter vectors of our model can be 'statistically calibrated' for simulation purpose. The resulting simulation model is composed of a sample of heterogeneous farm model based on a well-defined statistical background: the estimated micro-econometric multi-crop model and the farm sample used for its estimation.

We illustrate the empirical tractability of our approach by estimating a seven crops – and eight production regimes – production choice model for a sample of 415 French arable crop producers observed from 2006 to 2011. 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 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 this article is organized as follows. The approach proposed to account for endogenous regimes and regime fixed costs in the modelling of acreage decisions is presented in the first section. The structure of the micro-econometric multi-crop model considered in our empirical application is described in the second section. Identification and estimation issues are discussed in the third section. Illustrative estimation and simulation results are provided in the fourth section. Finally, we conclude.

Endogenous 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 feasible production regimes, *i.e.* the set of crop subsets considered by the farmer when choosing the crops he/she will produce. Let $\mathcal{K}^{+}(r)$ denote the subset of crops produced in regime *r*, with $\mathcal{K}^{+}(1) = \mathcal{K}$ by convention, and let $\mathcal{K}^{0}(r)$ denote the subset of crops not produced in regime *r*. Finally, let $\mathbf{s} \equiv (s_k : k \in \mathcal{K})$ denote an acreage share vector satisfying $\mathbf{s} \geq \mathbf{0}$ and $\mathbf{s}'\boldsymbol{\iota} = 1$ where $\boldsymbol{\iota}$ is the dimension *K* unitary column vector and let function $\rho(\mathbf{s})$ define the regime of the acreage share vector **s**.

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

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

where $\boldsymbol{\pi}_{it} \equiv (\pi_{k,it} : k \in \mathcal{K})$ is the vector of crop returns expected when choosing **s**, $C_{it}(\mathbf{s})$ is the implicit management cost of acreage **s** and $D_{it}(r)$ 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**.[1] Ignoring the regime fixed costs, the optimal acreage choice is determined by maximizing the sum of the crop expected gross margins $\pi^{e}_{k,it}$ weighted by their acreage shares s_k minus the costs associated to the crop acreage **s**, $C_{it}(\mathbf{s})$. In this model the crop acreage management costs prevent farmers to solely produce the most profitable crop.

The regime fixed cost terms $D_{it}(r)$ introduce discrete elements in farmers' acreage choices with $D_{it} \in \{D_{it}(r) : r \in \mathcal{R}\}$. 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 $D_{it}\left(\rho(\mathbf{s})\right)$ 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

---

[1] Analogous cost functions are used in the Positive Mathematical Programming literature (see, *e.g.*, Heckelei *et al* 2012) and in the multi-crop econometric literature (see, *e.g.*, Carpentier and Letort 2012, 2014).

are expected to be non-decreasing in the number of produced crops.[2]

As commonly done in acreage choice models based on censored regression systems (see, *e.g.*, Fezzi and Bateman 2011, Lacroix and Thomas 2011, Sckokai and Moro 2006, 2009), 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. Accordingly, 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:

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

yielding regime specific optimal acreage shares:

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

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

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

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:

(3) $$r_{it} \equiv \arg\max_{r \in \mathcal{R}} \left\{ \Pi_{it}(r) - D_{it}\left(\rho(\mathbf{s}_{it}(r))\right) \right\}.$$

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:

---

[2] Note however that in specific empirical settings the $D_{it}(r)$ 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 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.

(4a) $$\mathbf{s}_{it} = \mathbf{s}_{it}(r_{it})$$

and:

(4b) $$\Pi_{it} = \Pi_{it}(r_{it}) .$$

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, wheat acreage is expected to be more responsive to wheat price in farms producing barley than in farms not producing other straw cereals. Acreage choice models based on censored regression systems cannot reproduce such patterns.

The regime fixed cost considered in the maximization problem determining the optimal regime r_{it} is $D_{it}\left(\rho(\mathbf{s}_{it}(r))\right)$ rather than simply $D_{it}(r)$. In effect, 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 acreages of crops of $\mathcal{K}^{+}(r)$ may be null in $\mathbf{s}_{it}(r)$.

The Multinomial Logit (MNL) modelling framework proposed by Carpentier and Letort, (2014) is especially convenient in this context. First, 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 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 Standard MNL function:

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

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

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

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

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

Second, the optimal acreage share of crop *k* in regime *r*, *i.e.* $s_{k,it}(r)$, is ensured to be strictly positive if $k \in \mathcal{K}^+(r)$. This is shown by equation (5b) in the case standard MNL acreage share choice models. Indeed, considering Standard or Nested MNL acreage share choice models ensure that $\mathbf{s}_{it}(r)$ necessarily belongs to regime *r*.

Of course, $s_{k,it}(r)$ is almost null when crop *k* is much less profitable than other crops of regime *r*. 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-negativity constraints which would be involved in the case where the cost function $C_{it}(\mathbf{s})$ was chosen to be quadratic in **s**. These 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.[3] The acreage shares of the least profitable crops only become null at the production regime choice stage, when these crops are excluded from farmers' production plans.[4]

---

[3] 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 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.

[4] 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. 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

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 two 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 Koutchade *et al* (2018) we adopt a random parameter approach for accounting for farmers' and farms' 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 crop k obtained by farmer i in year t is

---

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.

given by:

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

where $x_{k,it}$ denotes the variable input use level. The $\alpha^{x}_{k,i}$ 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 observed farm heterogeneity and climatic conditions. 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 $\alpha^{x}_{k,i}$ random parameter, mainly capture three kinds of effects: those of the natural and material factor endowment of farms (e.g., soil quality, machinery quality), of farmers' practice choices (e.g., crop management practices, cropping systems) and of the skills of farmers. The $\varepsilon^{y}_{k,it}$ and $\varepsilon^{x}_{k,it}$ terms are standard error terms aimed to capture 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) $$y_{k,it} = \beta^{y}_{k,i} + (\boldsymbol{\delta}^{y}_{k,0})'\mathbf{z}^{y}_{k,it} - 1/2 \times \alpha^{x}_{k,i} w^{2}_{k,it} p^{-2}_{k,it} + \varepsilon^{y}_{k,it}$$

and the corresponding yield supply:

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

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,[5] this farmer expects the following return to the variable input:

(8) $$\pi_{k,it} = p_{k,it}\left(\beta^{y}_{k,i} + (\boldsymbol{\delta}^{y}_{k,0})'\mathbf{z}^{y}_{k,it}\right) - w_{k,it}\left(\beta^{x}_{k,i} + (\boldsymbol{\delta}^{x}_{k,0})'\mathbf{z}^{x}_{k,it}\right) + 1/2 \times \alpha^{x}_{k,i} w^{2}_{k,it} p^{-1}_{k,it}$$

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 share 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 corresponding acreage management cost function is given by:

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

[5] As discussed below, this assumption can be relaxed, *e.g.* for accounting for potential correlations between the $\varepsilon^{y}_{k,it}$ and $\varepsilon^{x}_{k,it}$ error terms on the one hand, and the $\varepsilon^{s}_{k,it}$ error terms on the other hand.

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^s, $\alpha_{(g),i}^s$ and $\alpha_{m|(g),i}^s$ terms 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 $\alpha_{m|(g),i}^s \geq \alpha_{(g),i}^s \geq \alpha_i^s > 0$ is sufficient for the cost function $C_{it}(\mathbf{s})$ to be strictly convex in **s**.

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

(10) $$\beta_{k,it}^s \equiv \beta_{k,i}^s + (\boldsymbol{\delta}_{k,0}^s)'\mathbf{z}_{k,it}^s + \varepsilon_{k,it}^s$$

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

These error terms are assumed to be independent from the error terms of the yield and input 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, Koutchade *et al* (2018) 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

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

production regime. It suffices to solve the maximization problem given in equations (3). For instance, eight 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 subgroup *m* of group *g*, we obtain:

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

and:

$$\text{(11b)}\qquad \Pi_{it}(r) = (\alpha^s_i)^{-1}\ln\sum_{h=1}^{G}\left\{\sum_{n=1}^{M(h)}\left(\Upsilon_{(n,h),it}(r)\right)^{\frac{\alpha^s_{(h),i}}{\alpha^s_{n|(h),i}}}\right\}^{\frac{\alpha^s_i}{\alpha^s_{(h),i}}}$$

where:

$$\text{(11c)}\qquad \Upsilon_{(n,h),it}(r) = \sum_{\ell\in\mathcal{K}(n,h)} j_\ell(r)\exp\left(\alpha^s_{n|(h),i}(\pi_{\ell,it}-\beta^s_{\ell,it})\right).$$

At the first level, parameter α^s_i drives the land allocation to crop group acreages. At the second level, parameters $\alpha^s_{(g),i}$ drive the allocation of the crop group acreages to crop subgroup acreages. Finally, parameters $\alpha^s_{n|(g),i}$ drive the allocation of the crop subgroup acreages to crop acreages at the third level.

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 be used for defining a regime choice model according to the choice problem described in equation (4). Let define the regime fixed costs as $D_{it}(r) = d_i(r) - \sigma_i^{-1} e_{it}^r$. The farm specific parameters $d_i(r)$ aim to capture the effects of unobserved factors affecting the regime fixed costs. The error terms e_{it}^r aim to capture the effects of stochastic factors and define the regime choice model as a probabilistic discrete choice model. Scale parameter σ_i determines the extent to which the regime expected profit levels fixed cost included, *i.e.* the $\Pi_{it}(r) - d_i(r)$ terms, explain the production regime choice as regards to the effects of the e_{it}^r idiosyncratic terms. The higher σ_i, the more the 'deterministic' terms $\Pi_{it}(r) - d_i(r)$ impact the observed regime choices in:

(12) $$r_{it} = \arg\max_{r \in \mathcal{R}} \left\{ \Pi_{it}(r) - d_i(r) + \sigma_i^{-1} e_{it}^r \right\} .$$

Regime fixed costs $d_i(r)$ can be specified in different ways. These costs are expected to increase with the number of crops. Transaction costs and labor requirements related to a production regime increase with the number of crops produced in that regime. Indeed, one way to specify $d_i(r)$ is to consider a sum of fixed costs associated to each crop produced in the production regime, with $d_i(r) \equiv \sum_{k \in \mathcal{K}^+(r)} d_{k,i}^c$ where $d_{k,i}^c$ is the fixed costs related to crop *k*. The fixed costs of the crops that are always produced (in the considered dataset) cannot be identified and can be set to zero for normalization purpose. Interestingly, this specification allows computing the fixed costs of regimes which are not observed in the data (provided that these regimes contains the "always produced crops"). This is of particular interest for simulation purposes. This regime fixed cost specification is used in our empirical application.

Yet, farmers may purchase inputs specific to different crops from the same supplier, implying savings in the related transaction costs. Moreover, different crops may generate work peak loads at the same dates, implying that farmers must be on their farm at these dates, whether this is due to a single crop or to several crops. In these cases, the regime fixed costs are sub-additive in the crop fixed costs and need to be directly specified as constant terms on a regime per regime basis, with

$d_i(r) = d_i^r$. This specification allows identification of the regime fixed costs associated to each regime observed in the data (given that the fixed cost of a "benchmark regime" needs to be normalized). But, it does not permit to recover the costs corresponding to regimes which are not observed in the data. This can be an issue for simulations based on the estimated model. For instance, changes in market conditions can foster adoption of new production regimes. Also, this regime fixed cost specification leads to consider estimation of numerous fixed cost parameters. It involves one parameter per regime (minus one) instead of one parameter per crop (minus the number of always produced crops) with the alternative specification.

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 11) and a probabilistic production regime choice model (equation 12). In this section, we briefly present the econometric procedure used to estimate this model.[7]

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

[7] Specification and estimation details are available in supplementary Technical Appendix.

in the population described by our sample. To do so, we rely on a random parameter approach, as proposed in Koutchade *et al* (2018).

Given the rather complex structure and the dimension of our model, we adopt a fully parametric framework. Apart from

(*a*) the modelled choice production variables, *i.e.* the crop yield levels $y_{k,it}$, the crop input use levels $x_{k,it}$ and the crop acreage shares $s_{k,it}$ – collected in the vector $\mathbf{c}_{k,it}$ – for $k \in \mathcal{K}$, and the production regimes r_{it} for $r \in \mathcal{R}$,

(*b*) the crop prices $p_{k,it}$, variable input prices $w_{k,it}$ and control variable vector $\mathbf{z}^{y}_{k,it}$, $\mathbf{z}^{x}_{k,it}$ and $\mathbf{z}^{s}_{k,it}$ collected in the explanatory variable vector $\mathbf{z}_{it}$ for $k \in \mathcal{K}$

(*c*) the control variable coefficient vectors $\boldsymbol{\delta}^{y}_{k,0}$, $\boldsymbol{\delta}^{x}_{k,0}$ and $\boldsymbol{\delta}^{s}_{k,0}$ for $k \in \mathcal{K}$ collected in the fixed coefficient vector $\boldsymbol{\delta}_{0}$ for $k \in \mathcal{K}$

the considered model contains three main subsets of random elements.

(*d*) The farm specific parameter vectors $\boldsymbol{\gamma}_i$ collect the potential yield parameters $\beta^{y}_{k,i}$, the input requirement parameters $\beta^{x}_{k,i}$, the input use flexibility parameters $\ln \alpha^{x}_{k,i}$ and the cost function linear parameters $\beta^{s}_{k,i}$ for $k \in \mathcal{K}$; the acreage choice flexibility parameters $\ln \alpha^{s}_{i}, \ln \alpha^{s}_{(g),i}$ and $\ln \alpha^{s}_{m|(g),i}$ for $m = 1,...,M(g)$ and $g = 1,...,G$; the regime fixed cost parameters d^{r}_{i} (or $d^{c}_{k,i}$ depending on the specification) for $r \in \mathcal{R}$ and the regime choice scale parameter $\ln \sigma_i$. Vectors $\boldsymbol{\gamma}_i$ are assumed normally and independently distributed across farms, implying that the production choice flexibility parameters, $\alpha^{x}_{k,i}$, $\alpha^{s}_{i}, \alpha^{s}_{(g),i}$ and $\alpha^{s}_{m|(g),i}$, and the scale parameter, σ_i, are assumed log-normally distributed.

(*e*) The yield supply and input demand error term vectors $\boldsymbol{\varepsilon}^{yx}_{k,it}$ contain the error terms $\varepsilon^{y}_{k,it}$ and $\varepsilon^{x}_{k,it}$. They are collected in the vectors $\boldsymbol{\varepsilon}^{yx}_{it}$ for $k \in \mathcal{K}$. The acreage share error term vectors $\boldsymbol{\varepsilon}^{s}_{it}$ collect the error terms $\varepsilon^{s}_{k,it}$ for $k \in \mathcal{K}$. The error term vectors $\boldsymbol{\varepsilon}_{k,it}$ collect the terms $\boldsymbol{\varepsilon}^{yx}_{k,it}$ and $\varepsilon^{s}_{k,it}$ and are collected in the vector $\boldsymbol{\varepsilon}_{it}$ for $k \in \mathcal{K}$. The error term vectors $\boldsymbol{\varepsilon}_{it}$ are assumed normally and independently distributed across farms.

(*f*) The production regime error term vectors $\mathbf{e}_{it}^{\rho}$ containing the error terms e_{it}^{r} for $r \in \mathcal{R}$. The terms e_{it}^{r} are assumed independent across regimes and distributed according to a type I extreme value distribution and years. The terms $\mathbf{e}_{it}^{\rho}$ are assumed independent across farms.

We further assume that the error term vectors $\boldsymbol{\varepsilon}_{it}^{yx}$, $\boldsymbol{\varepsilon}_{it}^{s}$ and $\mathbf{e}_{it}^{\rho}$ 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 $\boldsymbol{\gamma}_i$. 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, Arellano and Bonhomme 2011, 2012).

Importantly, we do not assume that farmers' choices and performances are not significantly impacted by unobserved dynamic features.[8] We assume that these dynamic features are sufficiently stable and persistent to be captured by the random parameters of our model. Indeed, we hypothesize that the random parameters $\boldsymbol{\gamma}_i$ capture the effects on farmers' production choices and performances

[8] Of course, dynamic features of multi-crop production technologies and of farmers' choices are worth investigating. Yet, these research topics are also difficult.

of the stable crop rotation schemes that these farmers' seem to rely on.[9] Our assuming that the error term vectors $\boldsymbol{\varepsilon}_{it}$ and $\mathbf{e}_{it}^{\rho}$ are serially independent across years is mostly based on this hypothesis.

The assumptions imposed on $\mathbf{e}_{it}^{\rho}$ implies that the considered regime choice model is – conditionally on σ_i, $d_i(r)$ and $\Pi_{it}(r)$ for $r \in \mathcal{R}$ – a standard Multinomial Logit probabilistic choice model, with:

(13) $$P[r_{it} \mid \boldsymbol{\gamma}_i, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^{s}] = \frac{\exp\left(\sigma_i(\Pi_{it}(r_{it}) - d_i(r_{it}))\right)}{\sum_{r \in \mathcal{R}} \exp\left(\sigma_i(\Pi_{it}(r) - d_i(r))\right)}.$$

Estimation

The aim of the estimation procedure is to obtain statistical estimates of two parameter sets. Vector $\boldsymbol{\eta}_0$ collects the fixed coefficient vector $\boldsymbol{\delta}_0$ and the elements of variance matrices of the error term vectors $\boldsymbol{\varepsilon}_{it}^{yx}$ and $\boldsymbol{\varepsilon}_{it}^{s}$, *i.e.* the parameters of the kernel probability distribution of our ESR multicrop model. Vector $\boldsymbol{\theta}_0$ collects the mean and variance-covariance parameters of the joint normal probability distribution of the random parameter vector $\boldsymbol{\gamma}_i$, *i.e.* the parameters of the mixing probability distribution of our model. Our model being fully parametric we used a Maximum Likelihood (ML) estimator for estimating $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$.

The observed production choices of farmer *i* in year *t* consists of the production regime r_{it} and of the yield levels, input use levels and acreage shares of the produced crops $\mathbf{c}_{it}^{+} \equiv (\mathbf{c}_{k,it} : k \in \mathcal{K}^{+}(r_{it}))$. 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 $\boldsymbol{\lambda}$. Under the assumptions of the

[9] See, *e.g.*, Koutchadé *et al* (2018) for an empirical analysis providing arguments confirming this hypothesis.

model, the terms $(\mathbf{c}_{it}^{+}, r_{it})$ are independent across years (and farms) conditionally on $(\boldsymbol{\gamma}_i, \mathbf{z}_i)$ where $\mathbf{z}_i \equiv (\mathbf{z}_{it} : t = 1,...,T)$. This implies that the likelihood function at $(\boldsymbol{\theta}, \boldsymbol{\eta})$ of farmer i production choices, $(\mathbf{c}_i^{+}, \mathbf{r}_i) \equiv ((\mathbf{c}_{it}^{+}, r_{it}) : t = 1,...,T)$, conditional on the exogenous variable vector $\mathbf{z}_i$ is given by:

(14) $$\ell_i(\boldsymbol{\theta}, \boldsymbol{\eta}) = \int \prod_{t=1}^{T} g(\mathbf{c}_{it}^{+}, r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\eta}) g(\boldsymbol{\gamma}_i; \boldsymbol{\theta}) d\boldsymbol{\gamma}_i .$$

Terms $g(\boldsymbol{\gamma}_i; \boldsymbol{\theta})$ are defined as probability density functions of multivariate normal random variables.

Given the structure of the model and the assumptions on its error terms, the probability density function conditional on $(\mathbf{z}_{it}, \boldsymbol{\gamma}_i)$ of the observed choices $(\mathbf{c}_{it}^{+}, r_{it})$ of farmer i in year t can be decomposed as:

(15) $$g(\mathbf{c}_{it}^{+}, r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\eta}) = g(r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i, \mathbf{s}_{it}^{+}; \boldsymbol{\eta}) g(\mathbf{c}_{it}^{+} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\eta})$$

where $\mathbf{s}_{it}^{+} \equiv (s_{k,it} : k \in \mathcal{K}^{+}(r_{it}))$ denotes produced crop acreage share vector. The term $g(\mathbf{c}_{it}^{+} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\eta})$ can be written in analytical closed form (Koutchadé *et al*, 2018) but the regime choice probability function $g(r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i, \mathbf{s}_{it}^{+}; \boldsymbol{\eta})$ cannot. This function is defined as the integral:

(16) $$g(r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i, \mathbf{s}_{it}^{+}; \boldsymbol{\eta}) = \int P[r_{it} \mid \boldsymbol{\gamma}_i, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^{s}] g(\boldsymbol{\varepsilon}_{it}^{s,0}; \boldsymbol{\eta}) d\boldsymbol{\varepsilon}_{it}^{s,0}$$

of the (Multinomial Logit) probability function $P[r_{it} \mid \boldsymbol{\gamma}_i, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^{s}]$ over the probability distribution of $\boldsymbol{\varepsilon}_{it}^{s,0} \equiv (\varepsilon_{k,it}^{s} : k \in \mathcal{K}^{0}(r_{it}))$. This sub-vector of $\boldsymbol{\varepsilon}_{it}^{s}$ collects the error terms of the implicit acreage cost function related to the non produced crops. These error terms are parts of the regime expected profit levels. As a result, they partly drive farmers' production regime choices. Indeed, the dependence of production regime and acreage choices on $\boldsymbol{\varepsilon}_{it}^{s}$ implies that the regime choices are endogenous from the econometric viewpoint.

The terms $\boldsymbol{\varepsilon}_{it}^{s,0}$ must be considered as missing variables in the estimation process because they cannot be recovered by combining the model and data. Building on the work of Harding and

Hausman (2007), we used Laplace approximates of the terms $g(r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i, \mathbf{s}_{it}^{+}; \boldsymbol{\eta})$ for computing the likelihood function of our model.[10]

Even when considering that the terms $g(r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i, \mathbf{s}_{it}^{+}; \boldsymbol{\eta})$ can suitably be integrated, the likelihood function $\ell_i(\boldsymbol{\eta}, \boldsymbol{\theta})$ can be obtained neither analytically nor numerically due to its integration over the probability distribution of the random parameters $\boldsymbol{\gamma}_i$. It can be integrated *via* direct simulation methods for computing Simulated ML (SML) estimators of $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$. Practically implementing this approach is difficult in our setting. The dimension of our parameter of interest and the rather complex functional form of the simulated version of likelihood function $\ell_i(\boldsymbol{\eta}, \boldsymbol{\theta})$ render the computation of SML estimators of $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ particularly challenging.

We choose to compute our ML estimator *via* a Stochastic Approximate Expectation-Maximization (SAEM) algorithm. SAEM algorithms were proposed by Delyon *et al* (1999) as extensions of the Generalized Expectation-Maximization (GEM) algorithms originally proposed by Dempster *et al* (1977). EM type algorithms replace a large ML maximization problem by a sequence of simpler problems. The SAEM algorithm used for obtaining the results presented in the next section is described in a Technical Appendix that can be obtained from the authors. The sketch of this algorithm is briefly described in rest of this section.

EM type algorithms are constructed based on the expectation conditional on the "observed data" of the "complete data" sample log-likelihood function of the considered model.

The complete – observed and unobserved modelled variables – data of farmer *i* consist of the

[10] The Laplace approximation approach consists of replacing function $P[r_{it} \mid \boldsymbol{\gamma}_i, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^{s,+}, \boldsymbol{\varepsilon}_{it}^{s,0}]$ by a second order Taylor expansion of this function in $\boldsymbol{\varepsilon}_{it}^{s,0}$ around an optimally chosen value of $\boldsymbol{\varepsilon}_{it}^{s,0}$.

observed production choices $(\mathbf{c}_i^+,\mathbf{r}_i)$ and of the random parameter $\boldsymbol{\gamma}_i$. The "complete data" sample log-likelihood function of the considered panel dataset is given by $\sum_{i=1}^{N}\ln\ell_i^C(\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\gamma}_i)$ where $\ln\ell_i^C(\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\gamma}_i)\equiv\sum_{t=1}^{T}\ln g(\mathbf{c}_{it}^+,r_{it}\,|\,\mathbf{z}_{it},\boldsymbol{\gamma}_i;\boldsymbol{\eta})+\ln g(\boldsymbol{\gamma}_i;\boldsymbol{\theta})$ for our model. The "observed data" related to farmer i, $\boldsymbol{\kappa}_i$, consist of the production choices $(\mathbf{c}_i^+,\mathbf{r}_i)$ on the one hand, and of the conditions of these choices and the characteristics of this farmer, $\mathbf{z}_i$, on the other hand. Let function $Q(\boldsymbol{\theta},\boldsymbol{\eta})$ define the expectation of $\sum_{i=1}^{N}\ln\ell_i^C(\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\gamma}_i)$ conditional on $(\boldsymbol{\kappa}_i:i=1,...,N)$ assuming that the density of $\boldsymbol{\gamma}_i$ conditional on $\boldsymbol{\kappa}_i$ is given by $g(\boldsymbol{\gamma}_i\,|\,\boldsymbol{\kappa}_i;\boldsymbol{\theta},\boldsymbol{\eta})$. We have:

(17a) $$Q(\boldsymbol{\theta},\boldsymbol{\eta})=\sum_{i=1}^{N}E_{(\boldsymbol{\theta},\boldsymbol{\eta})}[\ln\ell_i^C(\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\gamma}_i)\,|\,\boldsymbol{\kappa}_i]$$

where:

(17b) $$E_{(\boldsymbol{\theta},\boldsymbol{\eta})}[\ln\ell_i^C(\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\gamma}_i)\,|\,\boldsymbol{\kappa}_i]\equiv\int\ln\ell_i^C(\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\gamma}_i)g(\boldsymbol{\gamma}_i\,|\,\boldsymbol{\kappa}_i;\boldsymbol{\theta},\boldsymbol{\eta})d\boldsymbol{\gamma}_i\,.$$

The definition of $Q(\boldsymbol{\theta},\boldsymbol{\eta})$ ensures that its maximum in $(\boldsymbol{\theta},\boldsymbol{\eta})$ is the ML estimator of $(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)$. EM type algorithms are shown to converge to a maximum of $Q(\boldsymbol{\theta},\boldsymbol{\eta})$ under mild assumptions (see, *e.g.*, Wu 2003, Lavielle 2014).

SAEM algorithms iterate three steps until numerical convergence: a Simulation (S) step, an Approximation (A) step and a Maximization (M) step. These algorithms aim to progressively construct a stochastic approximation of $Q(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)$. Let assume that the estimate of $(\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)$ obtained at the end of iteration n is given by $(\hat{\boldsymbol{\theta}}_{(n)},\hat{\boldsymbol{\eta}}_{(n)})$.

At iteration $n+1$ the S step consists of integrating with simulation methods the terms $E_{(\boldsymbol{\theta},\boldsymbol{\eta})}[\ln\ell_i^C(\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\gamma}_i)\,|\,\boldsymbol{\kappa}_i]$ for $i=1,...N$. Building on the work of Caffo *et al* (2005), we use an Importance Sampling approach. The terms $E_{(\boldsymbol{\theta},\boldsymbol{\eta})}[\ln\ell_i^C(\boldsymbol{\theta},\boldsymbol{\eta},\boldsymbol{\gamma}_i)\,|\,\boldsymbol{\kappa}_i]$ are approximated by the weighted sums $\sum_{j=1}^{J_{(n+1)}}\tilde{\omega}_{i,(n+1)}^j\ln\ell_i^C(\boldsymbol{\theta},\boldsymbol{\eta},\tilde{\boldsymbol{\gamma}}_{i,(n+1)}^j)$ where the $\tilde{\boldsymbol{\gamma}}_{i,(n+1)}^j$ terms are independent random draws from the proposal density function $g(\boldsymbol{\gamma}_i;\hat{\boldsymbol{\theta}}_{(n)})$ and the $\tilde{\omega}_{i,(n+1)}^j$ terms are the corresponding normalized importance weights for $j=1,...,J_{(n+1)}$, with:

(18) $$\tilde{\omega}^j_{i,(n+1)} \equiv \frac{\prod_{t=1}^{T} g(\mathbf{c}^+_{it}, r_{it} \mid \mathbf{z}_{it}, \tilde{\boldsymbol{\gamma}}^j_{i,(n+1)}; \hat{\boldsymbol{\eta}}_{(n)})}{\sum_{j=1}^{J_{(n+1)}} \prod_{t=1}^{T} g(\mathbf{c}^+_{it}, r_{it} \mid \mathbf{z}_{it}, \tilde{\boldsymbol{\gamma}}^j_{i,(n+1)}; \hat{\boldsymbol{\eta}}_{(n)})}.$$

The A step consists of (implicitly) updating the stochastic approximation of $Q(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ based on $(\hat{\boldsymbol{\theta}}_{(n)}, \hat{\boldsymbol{\eta}}_{(n)})$, on the approximation used at iteration *n*, $\tilde{Q}_{(n)}(\boldsymbol{\theta}, \boldsymbol{\eta})$, and on the outcome of the S step. The updated approximate value of $Q(\boldsymbol{\theta}, \boldsymbol{\eta})$, $\tilde{Q}_{(n+1)}(\boldsymbol{\theta}, \boldsymbol{\eta})$, is obtained by using the recursive formula $\tilde{Q}_{(n+1)}(\boldsymbol{\theta}, \boldsymbol{\eta}) = (1-\lambda_{(n+1)})\tilde{Q}_{(n)}(\boldsymbol{\theta}, \boldsymbol{\eta}) + \lambda_{(n+1)} \sum_{i=1}^{N} \sum_{j=1}^{J_{(n+1)}} \tilde{\omega}^j_{i,(n+1)} \ln \ell^C_i(\boldsymbol{\theta}, \boldsymbol{\eta}, \tilde{\boldsymbol{\gamma}}^j_{i,(n+1)})$ with $\lambda_{(n+1)} \in [0,1]$.[11]

The M step consists of updating the estimate of $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ by computing $(\hat{\boldsymbol{\theta}}_{(n+1)}, \hat{\boldsymbol{\eta}}_{(n+1)})$. This updated estimate is defined $(\hat{\boldsymbol{\theta}}_{(n+1)}, \hat{\boldsymbol{\eta}}_{(n+1)}) \equiv \arg\max_{(\boldsymbol{\theta},\boldsymbol{\eta})} \tilde{Q}_{(n+1)}(\boldsymbol{\theta}, \boldsymbol{\eta})$ or as any value of $(\boldsymbol{\theta}, \boldsymbol{\eta})$ such that condition $\tilde{Q}_{(n+1)}(\hat{\boldsymbol{\theta}}_{(n+1)}, \hat{\boldsymbol{\eta}}_{(n+1)}) > \tilde{Q}_{(n+1)}(\hat{\boldsymbol{\theta}}_{(n)}, \hat{\boldsymbol{\eta}}_{(n)})$ holds.

The main advantage of SAEM algorithms for maximizing the log-likelihood function of random parameters models lies in the fact that the objective function of the M step can be decomposed as as:

(19a) $$\tilde{Q}_{(n+1)}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \tilde{Q}_{\theta,(n+1)}(\boldsymbol{\theta}) + \tilde{Q}_{\eta,(n+1)}(\boldsymbol{\eta})$$

where:

(19b) $$\tilde{Q}_{\theta,(n+1)}(\boldsymbol{\theta}) = (1-\lambda_{(n+1)})\tilde{Q}_{(n)}(\boldsymbol{\theta}) + \lambda_{(n+1)} \sum_{i=1}^{N} \sum_{j=1}^{J_{(n+1)}} \tilde{\omega}^j_{i,(n+1)} \ln g(\tilde{\boldsymbol{\gamma}}^j_{i,(n+1)}; \boldsymbol{\theta})$$

and:

(19c) $$\begin{aligned} \tilde{Q}_{\eta,(n+1)}(\boldsymbol{\eta}) &= (1-\lambda_{(n+1)})\tilde{Q}_{\eta,(n+1)}(\boldsymbol{\eta}) \\ &\quad + \lambda_{(n+1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{\tilde{\omega}^j_{i,(n+1)}}^{J_{(n+1)}} \tilde{\omega}^j_{i,(n+1)} \ln g(\mathbf{c}^+_{it}, r_{it} \mid \mathbf{z}_{it}, \tilde{\boldsymbol{\gamma}}^j_{i,(n+1)}; \boldsymbol{\eta}) \end{aligned}.$$

In our case, the term $\hat{\boldsymbol{\theta}}_{(n+1)} \equiv \arg\max_{\boldsymbol{\theta}} \tilde{Q}_{\theta,(n+1)}(\boldsymbol{\theta})$ can be obtained in analytical closed form based on the so-called "sufficient statistic approach" (Delyon *et al* 1999, Lavielle 2014) since $g(\boldsymbol{\gamma}_i; \boldsymbol{\theta})$ is the probability density function of a multivariate normal random variable.

---

[11] The terms $\lambda_{(n)}$ must be designed such that $\{\lambda_{(n)}\}$ is a decreasing sequence satisfying $\lambda_{(1)} = 1$, $\sum_{n=1}^{+\infty} \lambda_{(n)} = +\infty$ and $\sum_{n=1}^{+\infty} (\lambda_{(n)})^2 < +\infty$.

Maximizing $\tilde{Q}_{\eta,(n+1)}(\boldsymbol{\eta})$ in $\boldsymbol{\eta}$ is much more difficult due to the functional form of the kernel log-likelihood function $\ln g(\mathbf{c}_{it}^{+}, r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\eta})$. The assumptions underlying our model yields that this this function can be decomposed as $\ln g(r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i, \mathbf{s}_{it}^{+}; \boldsymbol{\eta}) + \ln g(\mathbf{c}_{it}^{+} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\eta})$, implying that function $\tilde{Q}_{\eta,(n+1)}(\boldsymbol{\eta})$ can be rewritten as:

(20a) $$\tilde{Q}_{\eta,(n+1)}(\boldsymbol{\eta}) = \tilde{Q}_{(\eta,c),(n+1)}(\boldsymbol{\eta}) + \tilde{Q}_{(\eta,r),(n+1)}(\boldsymbol{\eta})$$

where:

(20b) $$\begin{aligned}\tilde{Q}_{(\eta,c),(n+1)}(\boldsymbol{\eta}) &= (1-\lambda_{(n+1)})\tilde{Q}_{(\eta,c),(n)}(\boldsymbol{\eta}) \\ &\quad + \lambda_{(n+1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J_{(n+1)}} \tilde{\omega}_{i,(n+1)}^{j} \ln g(\mathbf{c}_{it}^{+} \mid \mathbf{z}_{it}, \tilde{\boldsymbol{\gamma}}_{i,(n+1)}^{j}; \boldsymbol{\eta})\end{aligned}$$

and:

(20c) $$\begin{aligned}\tilde{Q}_{(\eta,c),(n+1)}(\boldsymbol{\eta}) &= (1-\lambda_{(n+1)})\tilde{Q}_{(\eta,c),(n)}(\boldsymbol{\eta}) \\ &\quad + \lambda_{(n+1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J_{(n+1)}} \tilde{\omega}_{i,(n+1)}^{j} g(r_{it} \mid \mathbf{z}_{it}, \tilde{\boldsymbol{\gamma}}_{i,(n+1)}^{j}, \mathbf{s}_{it}^{+}; \boldsymbol{\eta})\end{aligned}.$$

The terms $\ln g(\mathbf{c}_{it}^{+} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\eta})$ are defined – up to an additive term that doesn't depend on $\boldsymbol{\eta}$ – as the log-likelihood functions at $\boldsymbol{\eta}$ of a Gaussian Seemingly Unrelated (linear) Regression (SUR) system with dependent variables missing at random (conditionally on $(\mathbf{z}_{it}, \boldsymbol{\gamma}_i)$). We used an iteration of the EM algorithm proposed by Ruud (1991) for computing ML estimator of the parameters of such models for defining $\hat{\boldsymbol{\eta}}_{(n+1)}$ such that that condition $\tilde{Q}_{(\eta,c),(n+1)}(\hat{\boldsymbol{\eta}}_{(n+1)}) > \tilde{Q}_{(\eta,c),(n+1)}(\hat{\boldsymbol{\eta}}_{(n)})$ holds. The "sufficient statistic approach" allows to obtain $\hat{\boldsymbol{\eta}}_{(n+1)}$ in analytical closed form solution.

In our case, the terms $g(r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}_i, \mathbf{s}_{it}^{+}; \boldsymbol{\eta})$ are complex functions of the parameter vector $\boldsymbol{\eta}$. These regime choice probability functions are ignored in the M step of our SAEM algorithm, implying that they impact the computation of the ML estimator of $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0)$ solely by impacting the value of the importance weights $\tilde{\omega}_{i,(n+1)}^{j}$.[12]

---

[12] The M step of our SAEM (type) algorithm doesn't ensure that condition $\tilde{Q}_{(n+1)}(\hat{\boldsymbol{\theta}}_{(n+1)}, \hat{\boldsymbol{\eta}}_{(n+1)}) > \tilde{Q}_{(n+1)}(\hat{\boldsymbol{\theta}}_{(n)}, \hat{\boldsymbol{\eta}}_{(n)})$ holds. It only ensures that condition $\tilde{Q}_{\theta,(n+1)}(\hat{\boldsymbol{\theta}}_{(n+1)}) + \tilde{Q}_{(\eta,c),(n+1)}(\hat{\boldsymbol{\eta}}_{(n+1)}) > \tilde{Q}_{\theta,(n+1)}(\hat{\boldsymbol{\theta}}_{(n)}) + \tilde{Q}_{(\eta,c),(n+1)}(\hat{\boldsymbol{\eta}}_{(n)})$ holds. We devised a simple heuristic for coping with cases where $(\hat{\boldsymbol{\theta}}_{(n+1)}, \hat{\boldsymbol{\eta}}_{(n+1)})$ doesn't succeed in increasing $\tilde{Q}_{(n+1)}(\boldsymbol{\theta}, \boldsymbol{\eta})$ from $\tilde{Q}_{(n+1)}(\hat{\boldsymbol{\theta}}_{(n)}, \hat{\boldsymbol{\eta}}_{(n)})$. This heuristic was rarely activated when running our SAEM algorithm, probably because the value

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. 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 SAEM algorithm we use relies on computations closely related to this calibration procedure since both rely on the probability distributions of the random parameters $\boldsymbol{\gamma}_i$ conditional on the observed choices $\mathbf{c}_i$ and explanatory variables $\mathbf{z}_i$. In this study, the specific parameter $\boldsymbol{\gamma}_i$ 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 SAEM 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.

Empirical application

---

of $\boldsymbol{\theta}$ plays a much crucial role in the value of the likelihood function of the model than $\boldsymbol{\eta}$ does. Indeed, our results demonstrate that the random parameters (the probability distribution of which is determined by the value of $\boldsymbol{\theta}_0$) play a dominating role in our model. Vector $\boldsymbol{\eta}$ contains the coefficients of the control variables and the elements of variance-covariance matrix of the model. Control variables play a minor role in our estimated model and the error term variance-covariance matrix is basically adjusted to the value of the other parameters of the model.

Data and empirical specification of the model

The model is estimated on an unbalanced panel data set containing 2276 observations of 415 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, mostly, and barley), corn and spring barley, which represent more than 80% of the total acreage in the considered area.[13]

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. The climatic variables are provided at the municipality level by Météo France, the French national meteorological service.

Figure 1 depicts the three levels nesting structure that we adopt for the seven crops. In a first level

[13] The EU sugar beet subsidy scheme requires limited adjustments in our application. This 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.

we distinguish a cereal group composed of winter cereals, corn and spring barley, and a group of crops that are generally planted at the head of rotation: corn, alfalfa, peas and rapeseed.

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

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 including only sugar beet (the only root crop considered here). The winter cereal aggregate is used as the benchmark crop. Based on these seven crops, 127 regimes, could theoretically be grown by farmers. The eight most frequently used were considered for selecting our sample. All farmers grow winter cereals, (spring) barley, (winter) rapeseed and most of them (91.7%) grow at least two additional crops. The most frequent regimes in the sample (regimes 2, 3 and 4) include five or six crops. Table 1 provides descriptive statistics concerning the production regimes observed in the data. Most farmers adopt different production regimes over the 5 years of our sample: only 8 out of 415 farmers have not changed their production regime. 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 2 – which excludes corn – is characterized by the highest

Table 1. Descriptive statistics

Regime number	Crops produced in the regime							Regime frequency	Average gross margin (€/ha)[d]
	Winter cereals[b]	*Corn*	*Spring Barley*	*Sugarbeet*	*Alfalfa*	*Peas*	*Rapeseed*		
1	■	■	■	■	■	■	■	6.6%	953
2	■		■	■	■	■	■	21.5%	1014
3	■	■	■	■	■		■	11.8%	930
4	■		■	■	■		■	48.6%	1007
5	■	■	■	■			■	2.8%	989
6	■	■	■				■	2.5%	825
7	■		■	■			■	4.9%	970
8	■		■				■	1.3%	738
Production frequency	100%	24%	100%	96%	88%	28%	100%		
Average acreage share[a]	0.38 (0.09)	0.02 (0.05)	0.18 (0.07)	0.15 (0.06)	0.10 (0.05)	0.02 (0.03)	0.16 (0.06)		
Average acreage share[a] if produced[a]	0.38 (0.09)	0.08 (0.07)	0.18 (0.07)	0.15 (0.06)	0.11 (0.04)	0.06 (0.03)	0.16 (0.06)		
Average gross margin (€/ha)[a]	843 (327)	872 (449)	756 (287)	1789[d] (379)	562 (286)	663 (269)	843 (311)		
Average yield (t/ha)[a]	8.58[b] (0.88)	9.23 (1.73)	6.82 (1.21)	95.19 (13.01)	12.62 (1.92)	4.72 (1.28)	3.89 (0.64)		
Average price (€/t)[a]	149[b] (31)	131 (34)	155 (35)	25[c] (3)	72 (15)	198 (25)	323 (64)		
Average fertilization and crop protection costs[a]	431 (91)	308 (74)	294 (70)	547 (126)	350 (125)	246 (66)	415 (83)		

a Empirical standard deviation in parentheses, b. Winter wheat (mostly) and barley, c. Off-quota price, d. Sugar beet subsidies excluded.

observed gross margin, but has been adopted in only 21.5% 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.

Regarding regime fixed costs, we rely on the second type of specification presented in section 6.2.3 and assume that regime fixed costs are equal to a sum of fixed costs associated to each crop produced in the regime, with $d_i(r) \equiv \sum_{k \in \mathcal{K}^+(r)} d^c_{k,i}$. As winter cereals, spring barley and rapeseed are always produced in the considered sample, the corresponding fixed costs are set to zero.

Estimation results

The parameter estimates of the yield, input demand, acreage shares and regime choice equations are reported in Tables 2 to 5.

Table 2. Selected Parameter Estimates of Yield Supply Models[a]

		Winter cereals	*Corn*	*Spring Barley*	*Sugar beet*	*Alfalfa*	*Peas*	*Rapeseed*
Error term $\varepsilon^y_{k,it}$								
	Std dev	0.66*	1.83*	0.95*	9.70*	2.96*	1.72*	0.49*
		(0.02)	(0.07)	(0.02)	(0.02)	(0.02)	(0.03)	(0.016)
Potential yield $\beta^y_{k,i}$								
	Mean	8.71*	9.06*	6.81*	95.60*	12.23*	4.15*	4.04*
		(0.02)	(0.04)	(0.02)	(0.32)	(0.04)	(0.03)	(0.01)
	Std dev	0.26*	0.65*	0.33*	5.7*	0.69*	0.51*	0.24*
		(0.01)	(0.03)	(0.01)	(0.17)	(0.02)	(0.02)	(0.01)
Input use flexibility $\alpha^x_{k,i}$								
	Mean	0.43*	0.08*	0.30*	0.49*	0.25*	0.33*	0.79*
		(0.01)	(0.00)	(0.00)	(0.03)	(0.00)	(0.01)	(0.02)
	Std dev	0.13*	0.09*	0.05*	0.58*	0.02	0.18*	0.31*
		(0.01)	(0.04)	(0.00)	(0.06)	(0.03)	(0.01)	(0.01)

a. Estimated standard errors of the ML estimator are in parentheses. Note: Asterisk (*) denotes a statistically significant parameter at the 5% level.

As shown in Table 2, the expectations of random parameters representing potential yields, $\beta^{y}_{k,i}$, are precisely estimated for all crops and their values lie in reasonable ranges regarding the average yields observed in the sample (Table 1). More importantly, the variances of their distributions are also significantly different from zero for all crops. These parameters thus significantly vary across farms, despite the fact that we control for observed factors characterizing farm heterogeneity (land and capital endowments and climatic conditions). This comes to illustrate the importance of unobserved farm heterogeneity and the relevance of the use of a random parameter approach that allows dealing with this heterogeneity.

Table 3. Selected Parameter Estimates of the Input Demand Models[a]

	Winter cereals	*Corn*	*Spring barley*	*Sugar beet*	*Alfalfa*	*Peas*	*Rapeseed*
Error term $\varepsilon^{x}_{k,it}$							
Standard deviation	0.52* (0.01)	0.59* (0.02)	0.41* (0.01)	0.84* (0.02)	0.88* (0.02)	0.60* (0.02)	0.58* (0.01)
Input requirement $\beta^{x}_{k,i}$							
Mean	4.36* (0.02)	2.57* (0.02)	2.92* (0.01)	5.44* (0.03)	3.15* (0.03)	2.29* (0.02)	4.44* (0.02)
Standard deviation	0.37* (0.02)	0.33* (0.01)	0.24* (0.01)	0.54* (0.02)	0.44* (0.01)	0.37* (0.01)	0.41* (0.02)
Input use flexibility $\alpha^{x}_{k,i}$							
Mean	0.43* (0.01)	0.08* (0.00)	0.30* (0.00)	0.49* (0.03)	0.25* (0.00)	0.33* (0.01)	0.79* (0.02)
Standard deviation	0.13* (0.01)	0.09* (0.04)	0.05* (0.00)	0.58* (0.06)	0.02 (0.03)	0.18* (0.01)	0.31* (0.01)

a. Estimated standard errors of the ML estimator are in parentheses. Note: Asterisk (*) denotes a statistically significant parameter at the 5% level.

The parameter estimates of the input demand equations, reported in Table 3, confirm this result: the probability distribution of their farm specific parameters is precisely estimated and displays significant heterogeneity. This is true for the random intercepts $\beta^{x}_{k,i}$ (the input use requirement) but also for the random slope parameters, $\alpha^{x}_{k,i}$, which represents the response of farmers to change in netput prices.

Turning to the parameter estimates of the acreage share equations in Table 4, again, the expectations and variance of random parameters are precisely estimated. Ranges of expectations of the acreage flexibility parameters are theoretically consistent. Conditions $\alpha^{s}_{m|(g),i} \geq \alpha^{s}_{(g),i} \geq \alpha^{s}_{i} > 0$ hold "on average". These are sufficient conditions for the acreage model to be well-behaved.

Table 4. Selected Parameter Estimates of the Acreage Share Models[a]

Crop level random terms	*Winter cereals*	*Corn*	*Spring barley*	*Sugar beet*	*Alfalfa*	*Peas*	*Rape--seed*
Error term $\varepsilon^{s}_{k,it}$							
Standard deviation	0	11.12* (0.38)	9.91* (0.19)	6.25* (0.13)	6.77* (0.15)	8.56* (0.28)	7.09* (0.16)
Acreage share shifters $\beta^{s}_{k,i}$							
Mean	0	17.41* (0.73)	13.88* (0.37)	24.51* (0.23)	11.15* (0.24)	18.78* (0.37)	11.07* (0.24)
Standard deviation	0	3.92* (0.02)	4.19* (0.02)	3.96* (0.03)	2.70* (0.06)	2.62* (0.01)	2.20* (0.01)

Acreage choice flexibility parameters	***Level 1*** α^{s}_{i}	***Level 2 (groups)*** $\alpha^{s}_{(g),i}$		***Level 3 (subgroups)*** $\alpha^{s}_{n|(g),i}$	
		Cereals	**Rotation heads**	**Spring cereals**	**Oil and protein crops**
	Cereals *vs* rotation heads	Spring cereals *vs* winter cereals	Sugar beet *vs* oil and protein crops	Corn *vs* spring barley	Rapeseed *vs* protein pea *vs* alfalfa
Mean	0.046* (0.001)	0.053* (0.001)	0.073* (0.001)	0.530* (0.029)	0.11* (0.002)
Standard deviation	0.015* (0.001)	0.013* (0.001)	0.025* (0.001)	0.640* (0.029)	0.020* (0.002)

a. Estimated standard errors of the ML estimator are in parentheses. Note: Asterisk (*) denotes a statistically significant parameter at the 5% level.

Finally, as shown in Table 5, the regime costs associated to crops, $d^{c}_{k,i}$, and the scale parameter, σ_i, of the regime choice equation are significantly estimated and heterogeneous across the sample. The mean value of the scale parameter, 1.80, is relatively large which reflects a relative importance of the deterministic part of the model. Regime profit and fixed cost levels in the regime choice appear to be significant drivers of the regime choices. The estimated mean fixed cost associated to alfalfa is negative. Two main reasons might explain this result. First, alfalfa is planted for at least two years. This crop requires farmers' intervention mostly at planting and harvesting. In the *Marne* region, the alfalfa downstream (dehydration) industry generally takes on harvest operations, which comes to decrease farmers' workload significantly. Second, being a legume alfalfa exhibits good agronomic properties, especially when used as a previous crop for cereals. The crop fixed cost estimates should however be considered cautiously given their high variability across farms.

Table 5. Parameter Estimates of Regime Choice Models

	Crop fixed costs $d_{k,i}^{c}$							Scale parameter σ_i
	Winter cereals	***Corn***	***Spring barley***	***Sugar beet***	***Alfalfa***	***Peas***	***Rape--seed***	
Mean[a]	0	3.80* (0.24)	0	0.30* (0.12)	-4.70* (0.28)	1.30* (0.04)	0	1.80* (0.07)
Std dev[a]	0	4.16* (0.10)	0	2.22* (0.05)	4.40* (0.11)	0.67* (0.01)	0	1.40* (0.07)

a. Estimated standard deviation of the estimator in parentheses. Note: Asterisk (*) denotes a statistically non null parameter at the 5% level.

Once we have estimated the parameters characterizing the distribution of the random parameters $\boldsymbol{\gamma}_i$, we can "statistically calibrate" those parameters for each farmer in our sample and thus obtain a set of farmer specific models. We used the expectation of $\boldsymbol{\gamma}_i$ conditional on $\boldsymbol{\kappa}_i$ for estimating the specific value of $\boldsymbol{\gamma}_i$ for each farmer in our sample. One interesting feature is that this procedure also allows us to calibrate the parameters of the yield, input demand and acreage equations corresponding to crops that have not been grown by the considered farmer as well as farmer specific regime fixed costs for regime that have never been chosen by the considered farmer.

Table 6. Fitting criteria (*Sim-R²*)

	Winter cereals	***Corn***	***Spring barley***	***Sugar beet***	***Alfalfa***	***Peas***	***Rape--seed***
Yield supply models	0.37	0.24	0.35	0.42	0.28	0.39	0.45
Input demand models	0.44	0.30	0.40	0.34	0.30	0.43	0.40
Acreage share models		0.57	0.34	0.83	0.70	0.53	0.41

The estimated farmer specific models allow us to compute some fitting criteria, Sim-R^2, which are reported in Table 6. The Sim-R^2 criterion measures the quality of the prediction of the observed choices of farmers by the estimated models. Its construction is analogous to that of the R^2 criterion of the standard linear regression model: for a given choice variable and a given model, the Sim-R^2 criterion is defined as the ratio of the empirical variance of the prediction of this variable to the empirical variance of the observed variable. These estimated criteria tend to show that the proposed model offers a satisfactory fit to our data.[14] Best fits are obtained (*a*) for crops with large acreage shares and (*b*) for acreage share models. Investigating the components of the Sim-R^2 criteria shows that their values are largely determined by estimates of the additively separable farm-specific terms $\beta_{k,i}^{y}$, $\beta_{k,i}^{x}$ and $\beta_{k,i}^{s}$ of our equation systems. These terms basically anchor the predicted levels of farmers' choices at their individual mean levels while the exogenous variables have limited explanatory power.

Using the estimated farmer specific models to predict the regimes choices observed in our data, we find our model to exhibit a relatively good predictive power with 72.4% of regime choices correctly predicted.

Importantly, our investigations on this issue tend to demonstrate that our results are robust to various distributional assumptions related to the model random parameters.

Simulation results

The structure of the proposed ERS multi-crop micro-econometric model allows to investigate the relative importance of the main drivers of production regime choices. For that purpose, we

[14] Much better fit levels are obtained for crop supply, acreage and input demand model defined at the farm level, mostly due to the explanatory power of the cropland area variable.

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 d_i^r for $r \in \mathcal{R}$. We simulate the regime choices according to each of these elements as well as combinations of these elements, and then confront them, on average, with the observed regime choices. Taken together these simulation results confirm that regime fixed costs matter, but mainly in combination with the other drivers of the regime choice model. The maximization of gross margins, or the minimization of acreage management costs or 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.

To illustrate the relevance of the approach we propose to deal with corner solutions 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 effect of the price of crop *k* on its expected acreage share for farmer i in year *t* can be decomposed as:

$$(21) \qquad \tfrac{\partial}{\partial p_{k,it}} E[s_{k,it} \mid \boldsymbol{\delta}_i, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^s] = \sum_{r \in \mathcal{R}} \left\{ \begin{array}{l} \tfrac{\partial}{\partial p_{k,it}} P[r_{it} = r \mid \boldsymbol{\delta}_i, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^s] \times E[s_{k,it} \mid \boldsymbol{\delta}_i, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^s, r_{it} = r] \\ + P[r_{it} = r \mid \boldsymbol{\delta}_i, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^s] \times \tfrac{\partial}{\partial p_{k,it}} E[s_{k,it} \mid \boldsymbol{\delta}_i, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^s, r_{it} = r] \end{array} \right\}.$$

The average acreage own price elasticities of our farm sample are reported in Table 7. 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 11% of the sugar beet acreage own price elasticities.

Table 7. Average own price elasticities of crop acreages

	Winter cereals	*Corn*	*Spring barley*	*Sugar beet*	*Alfalfa*	*Protein peas*	*Rape- -seed*
Average crop acreage own price elasticities	0.33	4.26	0.44	1.39	0.74	1.22	0.76
Due to changes in acreages within production regimes	0.33 (100%)	2.33 (55%)	0.43 (98%)	1.24 (89%)	0.60 (81%)	0.71 (58%)	0.75 (99%)
Due to changes in production regimes	0.00 (0%)	1.93 (45%)	0.01 (2%)	0.15 (11%)	0.14 (19%)	0.51 (42%)	0.01 (1%)

Observing how crop acreage elasticities within production regimes vary across regimes allows to illustrate features distinguishing endogenous regime models from censored regression models. Table 8 reports the estimated per regime means of own price crop acreage elasticities. These estimates display significant differences across production regimes because crop acreage choice mechanisms depend on the considered production regime. In particular, these estimates show that crop acreage own price elasticities grows with the number of crops characterizing the considered production regime. The higher the crop number, the more farmers can make use of crop acreage substitution possibilities. For instance, the more the considered regime contains crop rotation heads, the more rapeseed acreage choices are responsive to rapeseed price. Crop acreage models based on censored regression models cannot represent such patterns. They account for crop regimes but impose the same crop acreage model across production regime.

Table 8. Per regime average own price crop acreage elasticities

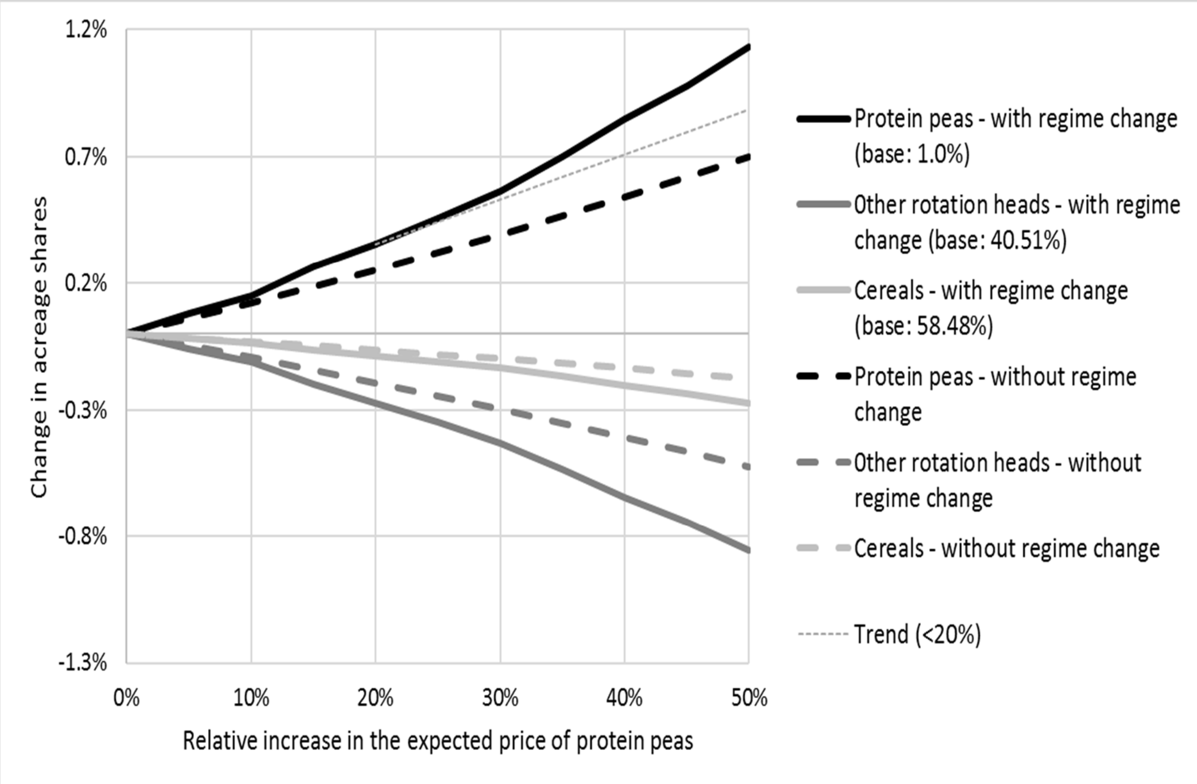
Regime			Crops produced in the regime						
Number	*Frequency*	*Crop number*	*Winter cereals*	*Corn*	*Spring barley*	*Sugar beet*	*Alfalfa*	*Peas*	*Rape- -seed*
1	6.6%	7	0.33	0.95	0.92	1.19	0.62	0.68	0.84
2	21.5%	6	0.31		0.32	1.17	0.61	0.67	0.82
3	11.8%		0.32	0.95	0.92	1.16	0.57		0.75
4	48.6%	5	0.30		0.32	1.14	0.56		0.74
5	2.8%		0.31	0.95	0.90	1.10			0.44
6	2.5%	4	0.29	0.95	0.90				0.35
7	4.9%		0.29		0.31	1.10			0.43
8	1.3%	3	0.27		0.30				0.30

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 its fixing atmospheric nitrogen for themselves as well as for their following crops in the rotation this crop is often considered as 'diversification crop' of particular 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.[15] 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 1.2%, from 1.0% to 2.2%. 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 0.85% while that of cereals would only decrease by around 0.3%. 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.

[15] 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.

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



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. This is partly explained by the threshold effects generated by the production regime fixed costs.

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 analyzing, 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. 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 stochastic versions of the EM algorithms 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.

References

Angrist, J.D., and Pischke, J.-S. (2008). Mostly harmless econometrics: An empiricist's companion (Princeton University Press).

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

Arellano, M., and Bonhomme, S. (2011). Nonlinear panel data analysis. Annual Review of Economics 3, 395–424.

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, 999–1008.

Caffo, B.S., Jank, W., and Jones, G.L. (2005). Ascent-based Monte Carlo expectation–maximization. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 235–251.

Carpentier, A., and Letort, E. (2014). Multicrop production models with Multinomial Logit acreage shares. Environmental and Resource Economics 59, 537–559.

Carpentier, A., and Letort, E. (2012). Accounting for heterogeneity in multicrop micro-econometric models: implications for variable input demand modeling. American Journal of Agricultural Economics 94, 209–224.

Carpentier, A., Gohin, A., Sckokai, P., and Thomas, A. (2015). Economic modelling of agricultural production: past advances and new challenges. Revue d'Études En Agriculture et Environnement 96, 131–165.

Chakir, R., and Thomas, A. (2003). Simulated maximum likelihood estimation of demand systems with corner solutions and panel data application to industrial energy demand. Revue d'Economie Politique 113(6), 773–799.

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

(Methodological) 1–38.

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

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

Harding, M.C., and Hausman, J. (2007). Using a Laplace approximation to estimate the random coefficients logit model by nonlinear least squares. International Economic Review 48, 1311–1328.

Heckelei, T., Britz, W., and Zhang, Y. (2012). Positive mathematical programming approaches–recent developments in literature and applied modelling. Bio-Based and Applied Economics 1, 109–124.

Kao, C., Lee, L., and Pitt, M.M. (2001). Simulated maximum likelihood estimation of the linear expenditure system with binding non-negativity constraints. Annals of Economics and Finance 2, 203–223.

Keane, M.P. (2009). Simulated maximum likelihood estimation based on first-order conditions. International Economic Review 50, 627–675.

Koutchadé, O.P., Carpentier, A., and Femenia, F. (2018). Modeling heterogeneous farm responses to EU biofuel support with a random parameter multicrop model. American Journal of Agricultural Economics (forthcoming)

Lacroix, A., and Thomas, A. (2011). Estimating the environmental impact of land and production decisions with multivariate selection rules and panel data. American Journal of Agricultural Economics 93, 784–802.

Lavielle, M. (2014). Mixed effects models for the population approach: models, tasks, methods and tools (CRC press).

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

McDonald, J.F., and Moffitt, R.A. (1980). The uses of Tobit analysis. The Review of Economics and Statistics 62(2), 318–321.

McLachlan, G., and Krishnan, T. (2007). The EM algorithm and extensions (John Wiley & Sons).

Moore, M.R., and 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 17(1), 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, 1202–1217.

Ruud, P.A. (1991). Extensions of estimation methods using the EM algorithm. Journal of Econometrics 49, 305–341.

Sckokai, P., and Moro, D. (2006). Modeling the reforms of the common agricultural policy for arable crops under uncertainty. American Journal of Agricultural Economics 88, 43–56.

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, 395–423.

Wales, T.J., and Woodland, A.D. (1983). Estimation of consumer demand systems with binding non-negativity constraints. Journal of Econometrics 21, 263–285.

Wei, G.C., and Tanner, M.A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. Journal of the American Statistical Association 85, 699–704.

Wu, C.J. (1983). On the convergence properties of the EM algorithm. The Annals of Statistics 11(1), 95–103.