Modeling Corners, Kinks, and Jumps in Crop Acreage Choices: Impacts of the EU Support to Protein Crops

Obafèmi P Koutchadé, Alain Carpentier, Fabienne Femenia

Modelling corners, kinks and jumps in crop acreage choices: impacts of the UE support to protein crops

Obafèmi P. Koutchadé[1], Alain Carpentier[1], Fabienne Femenia[1]

[1] INRAE, Institut Agro Rennes Angers, UMR SMART, F-35000, Rennes, France

This document is the postprint version of a paper published in the American Journal of Agricultural Economic: Koutchadé, O. P., Carpentier, A., & Femenia, F. (2021). Modeling Corners, Kinks, and Jumps in Crop Acreage Choices: Impacts of the EU Support to Protein Crops. American Journal of Agricultural Economics, 103(4), 1502-1524. https://doi.org/10.1111/ajae.12152

Part of this article is based on work carried out during the PhD thesis of Obafèmi Philippe Koutchadé, which was granted by INRAE and Arvalis. This work was continued and finalized within the framework of the H2020 MIND STEP project which received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 817566.

The authors thank the editor of the American Journal of Agricultural Economics his helpful comments and advices throughout the review process, as well as CER for providing the data, and Valérie Leveau (Arvalis) for her helpful agronomic advices.

Abstract

Null crop acreages raise pervasive issues when modelling acreage choices with farm data. We revisit these issues and emphasize that null acreage choices arise not only due to binding non-negativity constraints but also due to crop production fixed costs. Based on this micro-economic background, we present a micro-econometric multi-crop model that consistently handles null acreage choices and accounts for crop production fixed costs. This multivariate endogenous regime switching model allows for specific crop acreage patterns, such as multiple kinks and jumps in crop acreage responses to economic incentives that are due to changes in produced crop sets. We illustrate the empirical tractability of our modelling framework by estimating a random parameter version of our model on a panel dataset of French farmers. The estimated model is used to simulate the impacts of area-based subsidies on protein peas, which are implemented by the EU for reducing its dependence on imported protein crops. Our results suggest that this subsidy scheme is effective, essentially by leading farmers to incorporate or to keep protein pea in their crop mix.

Modelling corners, kinks and jumps in crop acreage choices: impacts of the UE support to protein crops

Market prices, agricultural policies or climate change impact crop supply through their effects on input uses, yield levels and acreage choices. Starting with the pioneering work of Just *et al* (1983), Chambers and Just (1989) and Chavas and Holt (1990), agricultural production economists have continuously proposed and developed micro-econometric multi-crop (MEMC) models for analyzing and quantifying these effects with farm accountancy data. Currently available MEMC models, however, ignore or poorly describe an important decision of crop producers: the choice to produce a subset of crops among the set of crops they can produce and sell. For instance, many applications of MEMC models with micro data ignore null acreages by relying either on crop aggregation eliminating null crop acreages (*e.g.*, Oude Lansink and Peerlings, 1996; Serra *et al*, 2005; Carpentier and Letort, 2012, Koutchadé *et al*, 2018) or on specific farm samples (*e.g.*, Just *et al*, 1983, 1990; Bayramoglu and Chakir, 2016). Of course, crop aggregation induces information loss while sample selection prevents extrapolation of the estimation results to farmers not producing all the considered crops.

Ignoring or poorly modelling farmers' crop set choices also prevents analyzing current agri-environmental policies. For instance, encouraging the production of protein crops has long been part of the European Union (EU) strategy to reduce its dependence on imported soybean (*e.g.*, Kuhlman *et al*, 2014).[1] As protein crops are minor crops in the EU, this not only requires encouraging farmers already producing these crops to increase their acreages, but also leading new farmers to produce protein crops. This last potential effect of EU policies, which requires suitably accounting for farmers' crop choices, motivated our developing the modelling framework presented in this article.

[1] Furthermore, owing to its fixing atmospheric nitrogen, pea is considered as a crop of particular interest for reducing pollutions due to chemical fertilizers as well as greenhouse gas emissions (*e.g.*, Kuhlman *et al*, 2014).

A few recent MEMC models explicitly account for null crop acreages (*e.g.*, Sckokai and Moro, 2006, 2009; Lacroix and Thomas, 2011; Fezzi and Bateman, 2011; Platoni *et al*, 2012).[2] These models are designed as censored regression (CR) systems. CR-MEMC models are typically built by adding mass points accounting for null crop acreages to standard MEMC models. They rely on reduced form discrete choice models for describing farmers' production regime choices, that is to say the choice of the subset of crops that farmers decide to plant. CR-MEMC models, however, have a main limitation, which comes from their relying on censoring mechanisms for modelling null acreage choices. They describe the acreage choice of a crop that is produced by an acreage choice model that assumes that all crops are produced. But, according standard optimization arguments, the optimal acreage level of a crop that is actually produced structurally depends on the production regime in which the production of this crop takes place.[3]

Admittedly, applications of CR-MEMC models don't focus either on null acreage choices or on crop set choices. By contrast, the main objectives of this article are (*i*) to revisit the null acreage issue in multi-crop models from a theoretical viewpoint, (*ii*) to propose an original MEMC model that accounts for farmers' crop choices in a way that is consistent from an economic viewpoint, together with a suitable estimation approach, and (*iii*) to show, by means of an application focused on protein peas production in France, that considering crop set choices significantly enriches micro-econometric analyses of farmers' crop supply.

Our MEMC modelling framework is based on an expected profit maximization problem considering land as an allocable quasi-fixed input. This problem accounts for the usual crop

[2] The studies by Moore and Negri (1992) and Moore *et al* (1994) being notable, early, exceptions.

[3] For instance, the price of a crop that is not produced cannot impact the acreage levels of the crops that are actually planted. More generally, production regimes largely determine how the acreage levels of the crops that are actually produced respond to price changes. Accordingly, the acreage choices of a crop that is produced should be described by regime specific models.

acreage non-negativity constraints but also for production regime fixed costs, which consist of unobserved costs – such as unmeasured marketing costs or implicit labor and machinery management costs – that depend on the set of crops that are grown but that don't depend on the acreages of these crops. Accordingly, our modelling framework assumes that farmers choose the production regime maximizing their expected profit, regime fixed costs included, together with the acreage, yield and input use of the produced crops.

Based on this micro-economic background, we design our MEMC model as an endogenous regime switching (ERS) multivariate model with multiple regimes. The core of this endogenous regime switching micro-econometric multi-crop (ERS-MEMC) model consists of a probabilistic regime choice model coupled with a set of regime specific Multinomial Logit (MNL) acreage choice models proposed by Carpentier and Letort (2014). Specific properties of MNL acreage choice models actually make them well suited for designing empirically tractable ERS-MEMC models, especially when considering regime specific fixed costs. As proposed by Koutchadé *et al* (2018), our ERS-MEMC model also features random parameters for accounting for farmers' unobserved heterogeneity.

Estimating multivariate ERS models with multiple regimes is challenging because their likelihood function involves integration of expectations over the probability distribution of multivariate latent error terms (*e.g.*, Pudney, 1989). Furthermore, the likelihood function of our ERS-MEMC model needs to be integrated over the probability distribution of its random parameters. Our estimation approach combines tools from the micro-econometrics and computational statistics literatures.

Our empirical application is based on a cost accounting panel dataset describing the production choices of 808 French arable crop producers from 2007 to 2011. Our estimation results demonstrate that our random parameter ERS-MEMC model performs well according to standard fit criteria. These results also show that regime specific fixed costs significantly matter

in farmers' crop choices and that the estimated acreage responses to economic incentives exhibit patterns that cannot be reproduced by CR-MEMC models.

We illustrate the point of considering farmers' crop set choices by using our estimated ERS-MEMC model for assessing the impacts of area-based payments aimed at fostering the production of protein peas in France. Despite decoupling of the Common Agricultural Policy (CAP) payments in 2006, pulse production is eligible to area-based payments in many EU member states, including France. Yet, pulse acreages remain limited in the EU (*e.g.*, Bues *et al*, 2013; Magrini *et al* 2016). Simulation results based on our estimated ERS-MEMC model suggest that both the decision to produce pea and pea acreage levels are responsive to area-based payments. For instance, removing the protein pea area-based payments would have reduced pea acreages by one third in our sample while moderate increases in the pea payments could have doubled pea acreages, mostly by leading farmers to incorporate protein pea in their crop mix at the expense of other rotation entry crops.

Our main contributions in this paper are twofold. First, the ERS-MEMC model presented in this article accounts for null crop acreages and regime specific fixed costs while relying on a consistent micro-economic background and being empirically tractable. Thereby, it provides a coherent response to an issue that is pervasive when analyzing crop production with farm level data.[4] Second, our empirical application shows that suitably modelling corners solutions in acreage choices is crucial when the considered crop set includes minor crops, besides its demonstrating the responsiveness of protein pea acreages to area-based payments in France.

The rest of this article is organized as follows. The approach proposed to account for crop

[4] To our knowledge, previous ERS models with multiple binding non-negativity constraints were considered for demand models, either of consumption goods (*e.g.*, Wales and Woodland, 1983; Lee and Pitt, 1986; Arndt *et al*, 1999; Kao et al, 2001; Millimet and Tchernis, 2008) or of production factors (*e.g.*, Lee and Pitt, 1987; Arndt, 1999: Chakir and Thomas, 2003). None of them consider regime specific fixed costs.

choices in micro-economic models of acreage decisions is presented in the first section. The structure of the corresponding ERS-MEMC model is described in the second section. The main features of our estimation strategy are presented in the third section. The empirical application is presented in the fourth section. Finally, we conclude.

Regime switching in multi-crop acreage models: corners, kinks and jumps

This section presents the theoretical modelling framework we propose for dealing with null crop acreages in micro-econometric acreage choice models. We proceed in two steps. First, we present the micro-economic crop acreage choice model underlying our ERS-MEMC model. Second, we present the functional form of the crop acreage choice models used in our ERS-MEMC model. We focus on the ability of these models to cope with corners, kinks and jumps in farmers' acreage choices.

Crop choices and crop acreages

We assume that farmers can allocate their fixed cropland area to K crops. Accordingly, set $\mathcal{K} = \{1,...,K\}$ denotes the set of crops that any considered farmer can produce and sell. Farmers' problem consists of optimally choosing a crop acreage share vector $\mathbf{s} = (s_k : k \in \mathcal{K})$ satisfying $\mathbf{s} \geq \mathbf{0}$ and $\mathbf{s}'\boldsymbol{\iota} = 1$, term $\boldsymbol{\iota}$ being the dimension K unitary column vector. Set $\mathcal{R} = \{1,...,R\}$ denotes the set of feasible production regimes. A production regime is characterized by the subset of crops with strictly positive acreages. Let set $\mathcal{K}^+(r)$ denote the subset of crops planted in regime r and function $\rho(\mathbf{s})$ defines the regime of the acreage share vector $\mathbf{s}$. We assume that farmers are risk neutral. In year t farmer i is assumed to choose her/his crop acreages by solving the following expected profit maximization problem:

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

Term $\boldsymbol{\pi}_{it} = (\pi_{k,it} : k \in \mathcal{K})$ is the vector of crop returns expected by farmer i when choosing $\mathbf{s}$ in year t. Function $C_{it}(\mathbf{s})$ is the implicit management cost of acreage $\mathbf{s}$ and $D_{it}(r)$ is the fixed cost

of production regime *r* incurred by farmer *i* in *t*. Acreage management costs $C_{it}(\mathbf{s})$ are costs not included in the crop gross margins that vary in **s**. They include unobserved variable input costs. They also account for the implicit costs related to constraints on acreage choices due to limiting quantities of machinery or labor, or to agronomic factors. These constraints providing motives for diversifying crop acreages, 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 function $C_{it}(\mathbf{s})$ is strictly convex in **s**.[5]

Regime fixed cost terms $D_{it}(r)$ introduce discrete elements, and thus salient discontinuities, in farmers' acreage choices. These costs are fixed in the sense that, contrary to acreage management costs, they don't depend on the chosen acreage in a given regime. They only depend on the crop set defining this regime. They account for hidden fixed costs incurred by the farmer for any acreage choice in the considered regime, such as fixed costs related to the marketing process of the crop products or those incurred when purchasing specific variable inputs, renting specific machines, seeking crop specific advises, *etc.* Regime fixed costs may also reflect the effects of farmer specific trade-offs. For instance, farmers may decide not to produce certain crops due to conflicting schedules with their off-farm activities.

Smooth acreage management cost function $C_{it}(\mathbf{s})$ and discrete 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, regime fixed costs are expected to be non-decreasing in the number of produced crops.

Farmers' expected profit maximization problem is solved following a standard backward induction approach according to which farmers choose their production regime after examining

[5] Analogous cost functions are used in the mathematical programming (*e.g.*, Mérel and Howitt, 2014; Heckeleï *et al*, 2012) and econometric literatures (*e.g.*, Heckeleï and Wolff, 2003; Carpentier and Letort, 2012).

their expected profit in each possible production regime. First, the acreage choice problem is solved for each potential regime. This yields the regime specific optimal acreage shares:

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

and the regime specific optimal expected profit levels (regime specific fixed costs excluded):

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

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

(4) $r_{it} = \arg\max_{r \in \mathcal{R}} \left\{ \Pi_{it}^{o}(\boldsymbol{\pi}_{it}; r) - D_{it}\left(\rho(\mathbf{s}_{it}^{o}(\boldsymbol{\pi}_{it}; r)) \right) \right\}$.

Assuming that optimal regime r_{it} is unique, optimal acreage choice $\mathbf{s}_{it}$ is obtained by combining equations (4) and (2), with $\mathbf{s}_{it} = \mathbf{s}_{it}^{o}(\boldsymbol{\pi}_{it}; r_{it})$. Similarly, equations (4) and (3) yield the optimal expected profit level Π_{it}, with $\Pi_{it} = \Pi_{it}^{o}(\boldsymbol{\pi}_{it}; r_{it})$.

Importantly, regime specific acreage choices $\mathbf{s}_{it}^{o}(\boldsymbol{\pi}_{it}; r)$ are derived from optimization problems that differ from one regime to the other due to nullity constraints on crop acreages. These constraints significantly impact how the acreage choices of the produced crops respond to market conditions. For instance, crop acreages are expected to be more responsive to economic incentives in regimes containing numerous crops, substitution opportunities being more limited within regimes containing only a few crops. This dependence of acreage choices on their production regimes implies that regime switches may induce kinks in the responses of acreage choices to economic incentives.

Regime fixed costs may induce discontinuities in the responses of acreage choices to economic incentives at points characterized by regime switches. For instance, regime fixed costs may include crop production fixed costs. As these costs cannot be covered by infinitesimal crop acreages, the acreage of a crop jumps from zero to some non-marginal acreage level as

soon as this crop becomes sufficiently profitable for covering its production fixed costs. More generally, regime fixed costs may add jumps to the kinks in the responses of acreage choices to economic incentives at any point characterized by regime switches, whether these regime switches correspond to binding non-negativity constraints or not.[6]

Regime fixed costs considered in regime choice problem (4) are defined by $D_{it}\left(\rho(\mathbf{s}_{it}^{o}(\boldsymbol{\pi}_{it};r))\right)$ rather than by $D_{it}(r)$. The production regime of $\mathbf{s}_{it}^{o}(\boldsymbol{\pi}_{it};r)$ may not be *r* because non-negativity constraints involved in maximization problem (3) may bind at $\mathbf{s}_{it}^{o}(\boldsymbol{\pi}_{it};r)$. Unfortunately, solutions to optimization problems involving non-negativity constraints are rarely obtained in analytical closed forms and require numerical solution approaches, as in the leading case where cost function $C_{it}(\mathbf{s})$ is quadratic in **s**. Overcoming these issues for empirical purposes requires specific modelling approaches and/or specific functional forms for cost function $C_{it}(\mathbf{s})$.

For instance, choosing $C_{it}(\mathbf{s})$ to be quadratic in **s** would allow obtaining an empirically tractable ERS-MEMC model by adapting the approach proposed by Wales and Woodland (1983).[7] The resulting ERS-MECM model would be based on the Karush-Kuhn-Tucker conditions describing the solutions in **s** to expected profit maximization (1). Similarly, currently available MEMC models, which often feature flexible function forms, could be transformed into empirically tractable ERS-MEMC model following the virtual price approach proposed by Lee and Pitt (1986). Yet, considering regime specific fixed costs in these models would raise particularly challenging specification and estimation issues.

MNL acreage choice models, null acreages and regime specific fixed costs

[6] Acreage choice responses to economic incentives of CR-MEMC models display at most one kink, and potentially a jump, at the point where farmers decide to produce a crop.

[7] Britz and Arata (2019) recently proposed an original estimation approach for quadratic multi-crop models. It allows accounting for null acreages but ignores the self-selection issues induced by farmers' crop set choices.

The ERS-MEMC model considered in this article is derived by choosing the cost function $C_{it}(\mathbf{s})$ to be a member of the family of entropic cost functions that underlies the MNL acreage choice models proposed by Carpentier and Letort (2014). As will be seen below, the main virtue of entropic cost functions lies in the fact that they yield regime specific crop acreage choice models $\mathbf{s}_{it}^{o}(\boldsymbol{\pi}_{it};r)$ and profit functions $\Pi_{it}^{o}(\boldsymbol{\pi}_{it};r)$ in smooth analytical closed forms. In what follows we present (a simplified version of) the acreage choice model that is at the core of the ERS-MEMC model considered in our application.[8] This acreage choice model can be interpreted as an extension of the MNL acreage choice models aimed to suitably account for null acreages and production regime fixed costs.

As discussed in Carpentier and Letort (2014), arable crops can often be grouped according to their competing for the use of quasi-fixed factors or according to their agronomic characteristics. Nested MNL (nMNL) acreage choice models, the specification of which relies on hierarchical partitions of the considered crop set, take advantage of these features of arable crops. The acreage choice model considered in our application is based on a 3 level nMNL acreage choice model. Albeit relatively simple, the specification of this model requires a relatively large amount of notations. We consider here a simpler a 2 level nMNL acreage choice model for demonstrating how a MNL acreage choice model can be extended for accounting for null acreages and production regime fixed costs.[9]

Accordingly, let assume that crop set $\mathcal{K}$ is partitioned into G mutually exclusive groups of crops. Term $\mathcal{G}=\{1,...,G\}$ defines the considered group set and group $h\in\mathcal{G}$ defines the crop subset $\mathcal{K}(h)$. Crops belonging to a same group are assumed to share similar agronomic characteristics and to compete more for farmers' limiting quantities of quasi-fixed factors than

[8] Other advantages and limits of MNL acreage choice models are discussed in Carpentier and Letort (2014).

[9] Extending a 2 level nNML model to a 3 level one significantly increases the flexibility of the considered model while being straightforward from a conceptual viewpoint as shown in Online Supplementary Appendix A1.

they compete with crops of other groups. The corresponding acreage management cost function is given by:

$$(5) \qquad C_{it}(\mathbf{s}) = \sum_{k\in\mathcal{K}} s_k \beta^s_{k,it} + \sum_{h=1}^{G} (\alpha^s_i)^{-1} s_{(h)} \ln s_{(h)} + \sum_{h\in\mathcal{G}} s_{(h)} (\alpha^s_{(h),i})^{-1} \sum_{m\in\mathcal{K}(h)} s_{m|(h)} \ln s_{m|(h)}$$

where $s_{(h)}$ denotes the acreage share of group *g* and $s_{m|(h)}$ that of crop *m* in the acreage of group *h*. Terms α^s_i and $\alpha^s_{(h),i}$ are farm specific parameters determining the flexibility of farmers' acreage choices. The larger they are, the more acreage choices respond to economic incentives (because the less management costs matter). Condition $\alpha^s_{(h),i} \geq \alpha^s_i > 0$ is sufficient for cost function $C_{it}(\mathbf{s})$ to be strictly convex in $\mathbf{s}$.[10] Importantly, function $C_{it}(\mathbf{s})$ is well-defined when $\mathbf{s}$ contains null elements despite its involving elements of $\ln \mathbf{s}$.[11]

Let function $g(k)$ define the group of crop *k* and let function $j_k(r)$ indicate whether crop *k* belongs to regime *r* or not; with $j_k(r) = 1$ if $k \in \mathcal{K}^+(r)$ and $j_k(r) = 0$ otherwise. Based on the entropic cost function (5) and these notations, solving the per regime expected profit maximization problems considered in equations (2) or (3) yields the following regime specific crop acreage share functions

$$(6) \quad s^o_{k,it}(\boldsymbol{\pi}_{it}; r) = \frac{j_k(r) \exp(\pi_{k,it} - \beta^s_{k,it})^{\alpha^s_{(g(k)),i}} \left(\sum_{\ell\in\mathcal{K}(g(\ell))} j_\ell(r) \exp(\pi_{\ell,it} - \beta^s_{\ell,it})^{\alpha^s_{(g(k)),i}} \right)^{\alpha^s_i (\alpha^s_{(g(k)),i})^{-1} - 1}}{\sum_{h\in\mathcal{G}} \left(\sum_{\ell\in\mathcal{K}(h)} j_\ell(r) \exp(\pi_{\ell,it} - \beta^s_{\ell,it})^{\alpha^s_{(h),i}} \right)^{\alpha^s_i (\alpha^s_{(h),i})^{-1}}}$$

and regime specific (indirect expected) profit functions

$$(7) \qquad \Pi^o_{it}(\boldsymbol{\pi}_{it}; r) = (\alpha^s_i)^{-1} \ln \sum_{h\in\mathcal{G}} \left(\sum_{\ell\in\mathcal{K}(h)} j_\ell(r) \exp(\pi_{\ell,it} - \beta^s_{\ell,it})^{\alpha^s_{(h),i}} \right)^{\alpha^s_i (\alpha^s_{(h),i})^{-1}} .$$

Acreage choice models $\mathbf{s}^o_{it}(\boldsymbol{\pi}_{it}; r)$ and profit functions $\Pi^o_{it}(\boldsymbol{\pi}_{it}; r)$ display properties of special interest for designing empirically tractable ERS-MEMC models. First, they are defined in analytical closed forms. Second, they are smooth in their parameters and arguments. Third, it is easily seen that $s^o_{k,it}(\boldsymbol{\pi}_{it}; r) \in (0,1)$ if $k \in \mathcal{K}^+(r)$ while $s^o_{k,it}(\boldsymbol{\pi}_{it}; r) = 0$ otherwise, implying that

[10] Constraint $\alpha^s_{(h),i} = \alpha^s_i$ is imposed if group *h* contains a single crop.

[11] The continuous extension property ensures that the domain of function $f(s) = s \ln s$ is $\mathbb{R}^+$ with $f(0) = 0$.

the regime of $\mathbf{s}_{it}^{o}(\boldsymbol{\pi}_{it};r)$ is *r*.

Equation (6) underlies the regime specific acreage choice models of our ERS-MEMC model. The set of equations $s_{k,it} = s_{k,it}^{o}(\boldsymbol{\pi}_{it};r_{it})$ for $k \in \mathcal{K}^{+}(r_{it})$ defines the system of nMNL acreage choice models corresponding to the crops that famer *i* decided to produce in year *t*. Due to the effects of the selection functions $j_k(r_{it})$ featured in functions $s_{k,it}^{o}(\boldsymbol{\pi}_{it};r)$, the functional form of $s_{k,it}^{o}(\boldsymbol{\pi}_{it};r_{it})$ structurally depends on regime r_{it}.

Term $\Pi_{it}^{o}(\boldsymbol{\pi}_{it};r)$ gives the optimal expected profit level achieved by farmer *i* would this farmer choose production regime *r* in year *t*. The choice of regime r_{it} by farmer *i* in year *t* is described by equation (4). Given that the production regime of $\mathbf{s}_{it}^{o}(\boldsymbol{\pi}_{it};r)$ is *r* with MNL acreage choice models, optimal regime choice r_{it} can also be defined as

(8) $\quad r_{it} = \arg\max_{r\in\mathcal{R}}\{\Pi_{it}^{o}(\boldsymbol{\pi}_{it};r) - D_{it}(r)\}$.

Problem (8) shows that farmers' regime choice problem is a basic discrete choice problem. It consists in choosing the profit maximizing (regime fixed costs included) production regime among a finite set of production regimes. As will be shown below, standard probabilistic discrete choice models can be derived from equation (8) by suitably specifying regime specific fixed costs $D_{it}(r)$. The regime choice model of the ERS-MEMC model considered in our application is a random parameter Multinomial Logit discrete choice model (*e.g.*, Train, 2009).

The core model of the ERS-MEMC model considered in our application is the ERS-nMNL acreage choice model composed of the probabilistic discrete regime choice model derived from equation (8) and of the regime specific nMNL acreage choice models given by $s_{k,it} = s_{k,it}^{o}(\boldsymbol{\pi}_{it};r_{it})$ for $k \in \mathcal{K}^{+}(r_{it})$. This ERS-nMNL acreage choice model can be estimated directly when expected crop returns $\boldsymbol{\pi}_{it}$ are observed. In our ERS-MEMC model, terms $\boldsymbol{\pi}_{it}$ are built from elements estimated in crop yield supply and variable input demand models that

complete the core ERS-nMNL acreage choice model.[12]

ERS-MEMC model with regime specific fixed costs: micro-economic structure

This section presents the structure of the ERS-MEMC model considered in the empirical application. This random parameter model is composed, on the one hand, of yield supply equations, variable input demand equations and acreage choice equations for each produced crop, and on the other hand, of a probabilistic production regime choice model. This ERS-MEMC model is an extension to an ERS framework with null acreages and regime fixed costs of the model proposed by Koutchadé *et al* (2018). Farmers' production choice parameters, including those driving their responses to economic incentives, are assumed to be farm specific, which allows accounting for farmers' and farms' unobserved heterogeneity. Accordingly, the main aim of the estimation procedure is to recover their probability distribution across the farmers' population represented by the considered sample. In order to ease its overall analysis, we present the model in its entirety first. Then we discuss its components following its recursive three-part structure.

Our ERS-MEMC model is composed of a system of yield supply and input demand models

$$(9) \qquad \begin{cases} y_{k,it} = \beta^{y}_{k,i} + (\boldsymbol{\delta}^{y}_{k,0})'\mathbf{c}^{y}_{k,it} - 1/2 \times \alpha^{x}_{k,i} w^{2}_{k,it} p^{-2}_{k,it} + \varepsilon^{y}_{k,it} \\ x_{k,it} = \beta^{x}_{k,i} + (\boldsymbol{\delta}^{x}_{k,0})'\mathbf{c}^{x}_{k,it} - \alpha^{x}_{k,i} w_{k,it} p^{-1}_{k,it} + \varepsilon^{x}_{k,it} \end{cases} \quad \text{for} \quad k \in \mathcal{K},$$

of a system of regime specific acreage choice models

$$(10) \qquad s_{k,it} = \frac{v_{k,it}(\pi_{k,it} - \beta^{s}_{k,it}; r_{it}) \left(\sum_{\ell \in \mathcal{K}(g(\ell))} v_{\ell,it}(\pi_{\ell,it} - \beta^{s}_{\ell,it}; r_{it}) \right)^{\alpha^{s}_{i}(\alpha^{s}_{(g(k)),i})^{-1} - 1}}{\sum_{h \in \mathcal{G}} \left(\sum_{\ell \in \mathcal{K}(h)} v_{\ell,it}(\pi_{\ell,it} - \beta^{s}_{\ell,it}; r_{it}) \right)^{\alpha^{s}_{i}(\alpha^{s}_{(h),i})^{-1}}} \quad \text{for} \quad k \in \mathcal{K},$$

and of a probabilistic regime choice model

$$(11) \qquad r_{it} = \arg\max_{r \in \mathcal{R}} \{ \Pi^{o}_{it}(\boldsymbol{\pi}_{it}; r) - \sum_{k \in \mathcal{K}^{+}(r)} \beta^{c}_{k,i} - \sigma^{-1}_{i} e_{r,it} \}$$

where

[12] ERS-MNL acreage choice models display unique features regarding null acreages. Null acreages can only result from the production regime choice in these models. They cannot be standard corner solutions at 0 because non-negativity constraints involved in problems (1) or (2) never bind with MNL acreage choice models.

(12) $v_{k,it}(\pi_{k,it} - \beta^s_{k,it}; r) = j_k(r)\exp\left(\alpha^s_{(g(k)),i}(\pi_{k,it} - \beta^s_{k,it})\right)$ for $k \in \mathcal{K}$ and $r \in \mathcal{R}$,

(13) $\pi_{k,it} = p_{k,it}\left(\beta^y_{k,i} + (\boldsymbol{\delta}^y_{k,0})'\mathbf{c}^{ys}_{k,it}\right) - w_{k,it}\left(\beta^x_{k,i} + (\boldsymbol{\delta}^x_{k,0})'\mathbf{c}^{xs}_{k,it}\right) + 1/2 \times \alpha^x_{k,i} w^2_{k,it} p^{-1}_{k,it}$ for $k \in \mathcal{K}$,

(14) $\beta^s_{k,it} = \beta^s_{k,i} + (\boldsymbol{\delta}^s_{k,0})'\mathbf{c}^s_{k,it} + \varepsilon^s_{k,it}$ for $k \in \mathcal{K}$

and

(15) $\Pi^o_{it}(\boldsymbol{\pi}_{it}; r) = (\alpha^s_i)^{-1} \ln \sum_{h \in \mathcal{G}} \left(\sum_{\ell \in \mathcal{K}(h)} v_{\ell,it}(\pi_{\ell,it} - \beta^s_{\ell,it}; r) \right)^{\alpha^s_i (\alpha^s_{(h),i})^{-1}}$ for $r \in \mathcal{R}$.

Yield supply and variable input demand models, and expected crop returns

Terms $x_{k,it}$ and $y_{k,it}$ respectively denote the variable input use and yield levels of crop *k* for farmer *i* in period *t* and terms $p_{k,it}$ and $w_{k,it}$ corresponding expected output and input prices. The yield supply and variable input demand models (9) are presented in detail in Carpentier and Letort (2014) and in Koutchadé *et al* (2018). We briefly present their background and their main components. Farmers are assumed to produce crops from a variable input aggregate under a quadratic technological constraint. Assuming that they maximize the expected return to variable input uses of each crop allows deriving the demands for the variable input and the yield supply levels described in equation (9). Terms $\mathbf{c}^y_{k,it}$ and $\mathbf{c}^x_{k,it}$ are observed variable vectors used to control for effects of observed heterogeneous factors (farm size and capital endowment) and climatic events. Terms $\beta^y_{k,i}$ and $\beta^x_{k,i}$ are farmer specific parameters aimed at capturing unobserved heterogeneity across farms and farmers. Parameter $\alpha^x_{k,i}$ determines the extent to which the yield supply and the input demand of crop *k* respond to the input and crop prices. It is required to be (strictly) positive for the production function to be (strictly) concave in $x_{k,it}$. Terms $\varepsilon^y_{k,it}$ and $\varepsilon^x_{k,it}$ are standard error terms aimed to capture the effects 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 $x_{k,it}$.

Assuming that the expectations of $\varepsilon^y_{k,it}$ and $\varepsilon^x_{k,it}$ of farmer *i* are null at the beginning of cropping season *t*, farmer *i* expected return of crop *k* is given by $\pi_{k,it}$ when she/he chooses her/his acreage shares. As shown by equation (13), crop returns $\pi_{k,it}$ are obtained from elements that appear in

the crop yield supply and input demand models. Vector $(\mathbf{c}^{ys}_{k,it}, \mathbf{c}^{xs}_{k,it})$ is defined by replacing in vector $(\mathbf{c}^{y}_{k,it}, \mathbf{c}^{x}_{k,it})$ the climatic variables by their expectations.

We recall here that $s_{k,it}$ is null if crop *k* doesn't belong to regime r_{it}. Accordingly, yield level $y_{k,it}$ and variable input use $x_{k,it}$ are not observed if crop *k* doesn't belong to regime r_{it}. Yet, defining these counterfactual variables is necessary. Estimates of the expected returns of the crops not produced in regime r_{it} are needed as inputs of the regime specific profit levels $\Pi^{o}_{it}(\boldsymbol{\pi}_{it}; r)$ that appear in regime choice model (11).

Acreage share choice models

Acreage share equation (10) is obtained from the regime specific acreage share model given in equation (6) by observing that $s_{k,it} = s^{o}_{k,it}(\boldsymbol{\pi}_{it}; r_{it})$. Once again, the functional form of these acreage choice models are presented in details in Carpentier and Letort (2014) and in Koutchadé *et al* (2018). We just focus on the elements that are useful for interpreting our empirical results. Terms α^{s}_{i} and $\alpha^{s}_{(g),i}$ are farm specific parameters determining the flexibility of farmers' acreage choices. Parameter α^{s}_{i} drives the land allocation to crop group acreages while parameters $\alpha^{s}_{(g),i}$ drive the allocation of the crop group acreages to crop acreages. The linear terms of the nMNL cost function given in equation (5), $\beta^{s}_{k,it}$, are defined in equation (14).[13] Terms $\mathbf{c}^{s}_{k,it}$ are explanatory variable vectors used to control for the effects of observed heterogeneous factors (farm size and capital endowment) and climatic events. Terms $\beta^{s}_{k,i}$ are farm specific parameters accounting for unobserved heterogeneity effects. Error terms $\varepsilon^{s}_{k,it}$ capture the effects on management costs of factors such as soil state at planting. Farmers are assumed to know terms $\varepsilon^{s}_{k,it}$ when choosing their acreages.

Production regime choice model

[13] These terms being identified only up to an additive constant term, normalization constraint $\beta^{s}_{\ell,it} = 0$ needs to be imposed for some crop ℓ (winter wheat in our application).

The discrete choice problem described in equation (11) is a probabilistic regime choice model. It relies on regime fixed cost models defined by $D_{it}(r) = d_{r,i} - \sigma_i^{-1} e_{r,it}$ with $d_{r,i} = \sum_{k \in \mathcal{K}^+(r)} \beta_{k,i}^c$. Let us ignore for the moment the second equation. Terms $d_{r,i}$ are farm specific parameters aimed to capture the effects of unobserved factors affecting the regime fixed costs. Error terms $e_{r,it}$ 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 regime expected profit levels $\Pi_{it}^o(\boldsymbol{\pi}_{it}; r) - d_{r,i}$ explain the production regime choice as regards to the effects of the $e_{r,it}$ idiosyncratic terms. The higher parameter σ_i, the more the expected profit levels impact the observed regime choices. Transaction costs and labor requirements related to a given production regime being increasing in the number of crops produced in this regime, regime fixed costs $d_{r,i}$ are expected to increase in the number of produced crops. Term $\beta_{k,i}^c$ being defined as crop *k* fixed cost, equation $d_{r,i} = \sum_{k \in \mathcal{K}^+(r)} \beta_{k,i}^c$ enforces this idea by defining the fixed cost of regime *r* as the sum of the fixed costs of the crops composing this regime. The fixed costs of the crops that are always produced (wheat and barley in our application) can be normalized to zero since they cannot be identified.

The "crop fixed cost" version of the production regime choice model is of special interest for empirical purposes. The number of regimes being generally larger that the number of crops, the "crop fixed cost" version of the model is more parsimonious. Also, estimating regime fixed costs suppose that the considered regimes are sufficiently frequently adopted. Similarly, estimating crop fixed costs suppose that the considered crops are sufficiently frequently, yet not always, produced. The identification conditions of the regime fixed costs are more demanding than those of the crop fixed costs. We use the "crop fixed cost" version of the production regime model in our application for these reasons.

Overall structure of the ERS-MEMC model

The set of dependent variables of our ERS-MEMC model contains the yield levels, input use

levels and acreage shares of each crop that are produced by farmer *i* in year *t*. These are collected in vector $\mathbf{w}_{it}^{+} = (\mathbf{y}_{it}^{+}, \mathbf{x}_{it}^{+}, \mathbf{s}_{it}^{+})$. Production regime r_{it} completes the dependent variable vector of the model. The set of explanatory variables contains crop prices, variable input prices and the control variable vectors used in the crop yield supply, input demand and acreage share equations. These exogenous variables are collected in vector $\mathbf{z}_{it}$, which defines the information set of the model.

The sole fixed parameters appearing in the ERS-MEMC model equations are the coefficients of the control variables. These parameters are collected in vector $\boldsymbol{\delta}_0$. The ERS-MEMC model contains two main subsets of random components: a vector of random parameters and a vector of error terms. Random parameter vector $\boldsymbol{\gamma}_i$ collects the farm specific parameters of the model, with $\boldsymbol{\gamma}_i = (\boldsymbol{\beta}_i, \boldsymbol{\alpha}_i, \sigma_i)$. Vector $\boldsymbol{\beta}_i$ collects the farm specific potential yield parameters ($\beta_{k,i}^{y}$), the input requirement parameters ($\beta_{k,i}^{x}$), the cost function linear parameters ($\beta_{k,i}^{s}$) and the crop fixed costs parameters ($\beta_{k,i}^{c}$). Vector $\boldsymbol{\alpha}_i$ collects the input use flexibility parameters ($\alpha_{k,i}^{x}$) and the acreage choice flexibility parameters (α_i^{s} and $\alpha_{(h),i}^{s}$). Finally, $\boldsymbol{\gamma}_i$ contains the scale parameter, σ_i, of the regime choice model.

In error term vector $\boldsymbol{\varepsilon}_{it} = (\boldsymbol{\varepsilon}_{it}^{yx}, \boldsymbol{\varepsilon}_{it}^{s})$, sub-vector $\boldsymbol{\varepsilon}_{it}^{yx}$ collects the error terms of the crop yield supply and input demand equations while sub-vector $\boldsymbol{\varepsilon}_{it}^{s}$ collects those of the acreage share equations. Finally, vector $\mathbf{e}_{it}$ collects the error terms of the regime choice model.

ERS-MEMC model with regime specific fixed costs: estimation strategy

This section presents the main features of the estimation strategy adopted for estimating the ERS-MEMC model described above. As this model involves multiple endogenous regimes, considers numerous interrelated production choices and features random parameters, we impose parametric distributional assumptions on its random components (*i.e.* error terms and random parameters) for ensuring its empirical tractability. We also impose simplifying assumptions regarding the dynamics of farmers' choices and the multi-crop production

technology. These assumptions are presented and discussed first. Then, we present how the main parameters of interest of our ERS-MEMC model are recovered from the data. Finally, we briefly present the main estimation issues that we face when estimating this model and the approaches we have chosen for overcoming these issues. A detailed description of our estimation procedure is provided in Online Supplementary Appendix A.

Probabilistic assumptions

We assume that terms $(\boldsymbol{\varepsilon}_{is}, \mathbf{e}_{is})$, $\boldsymbol{\gamma}_i$ and $\mathbf{z}_{it}$ are independently distributed for any pair (t, s). This implies that the explanatory variables vector, $\mathbf{z}_{it}$, is assumed to be (*i*) strictly exogenous with respect to the error term vectors and (*ii*) independent of the random parameters $\boldsymbol{\gamma}_i$. We further assume that error term $(\boldsymbol{\varepsilon}_{it}, \mathbf{e}_{it})$ vectors are independently distributed across time. Combined with the fact that vector $\mathbf{z}_{it}$ doesn't contain any lagged endogenous variable, this serial independence assumption implies that our MEMC model can be interpreted as a reduced form model regarding the dynamic features of the modelled choices. Indeed, we hypothesize that random parameters $\boldsymbol{\gamma}_i$ capture the effects on farmers' production choices and performances of the stable crop rotation schemes these farmers rely on. Finally, we assume that the error term vectors $\boldsymbol{\varepsilon}_{it}^{yx}$, $\boldsymbol{\varepsilon}_{it}^{s}$ and $\mathbf{e}_{it}$ are mutually independent.[14]

Distributional assumptions

Random parameter vectors $\boldsymbol{\gamma}_i$ are assumed independent across farms. For sake of simplification, we assume here that these random parameter vectors are normally distributed, with $\boldsymbol{\gamma}_i \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Omega}_0)$. Various transformations of elements of $\boldsymbol{\gamma}_i$ actually allow for other distribution choices for these elements while keeping the multivariate structure of the probability distribution of $\boldsymbol{\gamma}_i$ (*e.g.*, Stanfield *et al*, 1996). For example, considering log-transformations of $\boldsymbol{\alpha}_i$ and σ_i in $\boldsymbol{\gamma}_i$ implies that these random parameters, which are required

[14] Relaxing this independence assumption for $\boldsymbol{\varepsilon}_{it}^{yx}$ and $\boldsymbol{\varepsilon}_{it}^{s}$ significantly increases the estimation burden.

to be positive, are jointly log-normally distributed. We used this log-transformation in the ESR-MEMC model used for our empirical application. Robustness checks demonstrated that other probability distribution choices have a limited impact on the main results.[15] We make the usual assumptions stating that error term vectors $\boldsymbol{\varepsilon}_{it}$ are independent across farms (and years) and normally distributed, with $\boldsymbol{\varepsilon}_{it} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_0)$.

Finally, we assume that the regime choice model error terms $e_{r,it}$ are independent across regimes and distributed according to a type I extreme value distribution for $r \in \mathcal{R}$. This assumption implies that the considered regime choice model is a standard random parameter Multinomial Logit discrete choice model. The corresponding probability function of the observed regime choices is given by:

$$(16) \qquad P(r_{it} \mid \boldsymbol{\varepsilon}^s_{it}, \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\delta}_0) = \frac{\exp\left[\sigma_i \left(\Pi^o_{it}(r_{it}) - \sum_{k \in \mathcal{K}^+(r_{it})} \beta^c_{k,i})\right)\right]}{\sum_{r \in \mathcal{R}} \exp\left[\sigma_i \left(\Pi^o_{it}(r) - \sum_{k \in \mathcal{K}^+(r)} \beta^c_{k,i})\right)\right]}.$$

This probability is defined as a function of $(\boldsymbol{\varepsilon}^s_{it}, \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\delta}_0)$ because regime specific expected profit levels $\Pi^o_{it}(r)$ functionally depend on all elements of vector $(\boldsymbol{\varepsilon}^s_{it}, \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\delta}_0)$, scale parameter σ_i and crop fixed costs $\beta^c_{k,i}$ excepted.

Identification

Under the considered assumptions the probability distribution of farmers' responses to economic incentives, $\boldsymbol{\alpha}_i$, are identified through two main channels. Identification of the probability distribution of variable input use flexibility parameters $\alpha^x_{k,i}$ mostly relies on the variations of the corresponding input to crop price ratios. Identification of the probability distribution of acreage choice flexibility parameters α^s_i and $\alpha^s_{(g),i}$ mainly relies on the variations of the expected crop returns $\boldsymbol{\pi}_{it}$. Importantly, the expected crop returns are defined

[15] In the considered alternative specifications we assumed the $\boldsymbol{\beta}_i$ parameters to be log-normally distributed and/or the $\boldsymbol{\alpha}_i$ parameters to follow a bounded Johnson distribution (*e.g.*, Stanfield *et al*, 1996).

as functions of random parameters (*i.e.*, $\beta^{y}_{k,i}$, $\beta^{x}_{k,i}$ and $\alpha^{x}_{k,i}$) that may be correlated with the acreage choice flexibility parameters (*i.e.*, α^{s}_{i} and $\alpha^{s}_{(g),i}$). The unrestricted variance-covariance matrix of the joint probability distribution of the random parameters $\boldsymbol{\gamma}_i$ takes into account these potential correlations.

Scale parameter σ_i, which is the random coefficient associated to the regime specific expected profit levels $\Pi^{o}_{it}(r)$ in the regime choice model, is mainly identified by the variations in these variables. The joint probability distribution of the identifiable crop fixed costs $\beta^{c}_{k,i}$ vector is mainly identified by the variations in the differences in the regime specific expected profit levels $\Pi^{o}_{it}(r)$ across the production regimes. Potential correlations between, on the one hand, the random parameters that are part of the expected profit levels and, on the other hand, the crop fixed costs and the scale parameter are taken into account in the distribution of $\boldsymbol{\gamma}_i$.

Estimation issues and procedure

The considered ERS-MEMC model being fully parametric, we use a Maximum Likelihood (ML) estimator for efficiently estimating its parameters. These parameters are collected in vector $\boldsymbol{\theta}_0 = (\boldsymbol{\delta}_0, \boldsymbol{\Psi}_0, \boldsymbol{\mu}_0, \boldsymbol{\Omega}_0)$. Contribution of farmer *i* to the likelihood function of the model corresponds to the probability density function of her/his sequence of production choices conditional on the set of exogenous variables characterizing this choice sequence. Let function $f(\mathbf{u} \mid \mathbf{v}; \boldsymbol{\eta})$ generically denote the probability distribution function of $\mathbf{u}_{it}$ conditional on $\mathbf{v}_{it} = \mathbf{v}$ at $\mathbf{u}_{it} = \mathbf{u}$ parameterized by $\boldsymbol{\eta}$, and let function $\varphi(\mathbf{u}; \boldsymbol{\Omega})$ denote the density function of $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ at $\mathbf{u}$. Given the probabilistic assumptions defining the considered random parameter ERS-MEMC model, contribution of farmer *i* to the likelihood function at $\boldsymbol{\theta}$ is given by:

$$(17) \qquad \ell_i(\boldsymbol{\theta}) = \int \left(\prod_{t=1}^{T} f(\mathbf{w}^{+}_{it}, r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi}) \right) \varphi(\boldsymbol{\gamma} - \boldsymbol{\mu}; \boldsymbol{\Omega}) d\boldsymbol{\gamma} .$$

Likelihood function $\ell_i(\boldsymbol{\theta})$ can be obtained neither analytically nor numerically due to its integration over the probability distribution of random parameter $\boldsymbol{\gamma}_i$. Micro-econometricians generally solve this problem by integrating $\ell_i(\boldsymbol{\theta})$ *via* direct simulation methods for computing

simulated ML (SML) estimators of $\boldsymbol{\theta}_0$ (*e.g.*, Keane, 2009). Yet, implementing this approach is particularly challenging here due to the dimension of $\boldsymbol{\theta}_0$ and the complexity of the probability distribution function $f(\mathbf{w}_{it}^{+}, r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi})$. Instead, we compute the ML estimator of $\boldsymbol{\theta}_0$ by devising a Stochastic Approximate Expectation-Maximization (SAEM) algorithm. SAEM algorithms are extensions of the Expectation-Maximization algorithms proposed by Dempster, Laird and Rubin (1977). They were proposed by Delyon *et al* (1999) for computing ML estimators of models featuring continuous random parameters. They appear to use simulations more efficiently than competing alternatives (*e.g.*, McLachlan and Krishnan, 2007; Lavielle, 2014), which is a particularly relevant property when considering large samples, large multivariate models and/or large random parameter vectors.

Decomposing probability distribution function $f(\mathbf{w}_{it}^{+}, r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi})$ demonstrates that the issues related to its functional form are twofold:

(18) $$f(\mathbf{w}_{it}^{+}, r_{it} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi}) = f(r_{it} \mid \mathbf{s}_{it}^{+}, \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi}) f(\mathbf{w}_{it}^{+} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi}).$$

Function $f(\mathbf{w}_{it}^{+} \mid \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi})$ is the likelihood of crop level choice vector $\mathbf{w}_{it}^{+}$ conditional on $(\mathbf{z}_{it}, \boldsymbol{\gamma}_i = \boldsymbol{\gamma})$. Its functional form is that of a Gaussian Seemingly Unrelated Regression (SUR) system with missing observations, the missing observations being the yield levels, input uses and acreages of the crops that are not produced. Based on Ruud's (1991) insights we devised an EM type approach for updating the estimates of $(\boldsymbol{\delta}_0, \boldsymbol{\Psi}_0)$ within our SAEM algorithm.

Function $f(r_{it} \mid \mathbf{s}_{it}^{+}, \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi})$ is the probability function of regime r_{it} conditional on $(\mathbf{s}_{it}^{+}, \mathbf{z}_{it}, \boldsymbol{\gamma}_i = \boldsymbol{\gamma})$. Given the structure of our MEMC model, this probability function can be defined as a function of the error terms of the acreage share equations. Let vector $\boldsymbol{\varepsilon}_{it}^{s,+}$ collect the error terms of the acreage share models of the crops produced in regime r_{it} and vector $\boldsymbol{\varepsilon}_{it}^{s,0}$ collect those of the crops that are not produced in this regime. The residual term of the acreage share model, denoted here by $\hat{\boldsymbol{\varepsilon}}_{it}^{s,+}(\boldsymbol{\gamma}, \boldsymbol{\delta})$, provides an estimate of vector $\boldsymbol{\varepsilon}_{it}^{s,+}$. Vector $\boldsymbol{\varepsilon}_{it}^{s,0}$ must be considered as a missing variable because it cannot be recovered by combining the model and

the data. The structure of our MEMC model and equation (11) imply that:

(19) $$f(r_{it} \mid \mathbf{s}_{it}^{+}, \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi}) = \int P(r_{it} \mid \boldsymbol{\gamma}, \mathbf{z}_{it}, \boldsymbol{\varepsilon}_{it}^{s}; \boldsymbol{\delta}) f(\boldsymbol{\varepsilon}_{it}^{s,0} \mid \hat{\boldsymbol{\varepsilon}}_{it}^{s,+}(\boldsymbol{\gamma}, \boldsymbol{\delta}); \boldsymbol{\delta}, \boldsymbol{\Psi}) d\boldsymbol{\varepsilon}_{it}^{s,0}$$

where $f(\boldsymbol{\varepsilon}_{it}^{s,0} \mid \hat{\boldsymbol{\varepsilon}}_{it}^{s,+}(\boldsymbol{\gamma}, \boldsymbol{\delta}); \boldsymbol{\delta}, \boldsymbol{\Psi})$ denotes the density of $\boldsymbol{\varepsilon}_{it}^{s,0}$ conditional on $\boldsymbol{\varepsilon}_{it}^{s,+} = \hat{\boldsymbol{\varepsilon}}_{it}^{s,+}(\boldsymbol{\gamma}, \boldsymbol{\delta})$ and probability function $P(r_{it} \mid \boldsymbol{\varepsilon}_{it}^{s}, \mathbf{z}_{it}, \boldsymbol{\gamma}_i; \boldsymbol{\delta}_0)$ is given in equation (16). Following Harding and Hausman (2007), we use Laplace approximates of regime choice probability functions $f(r_{it} \mid \mathbf{s}_{it}^{+}, \mathbf{z}_{it}, \boldsymbol{\gamma}; \boldsymbol{\delta}, \boldsymbol{\Psi})$ in order to reduce the computing cost of our ML estimator.

Empirical application: impacts of area-based subsidies on protein peas in France

This section presents an empirical application focused on protein pea production in France. The objectives of this application are twofold. First, it aims at illustrating the empirical tractability of our modelling approach as well as to demonstrate the role of crop set choices in analyzes of farmers' crop production choices. Second, production of protein crops has benefited from public support for a long time in the EU in order to reduce the European dependence on imported soybean meal (*e.g.*, Kuhlman *et al* 2014; Magrini *et al*, 2016). For instance, area-based payments were implemented for promoting protein pea production in France. To our knowledge, the efficiency of this support scheme has not been assessed until now. We show how the estimated version of our ERS-MEMC model can be used for assessing the effects of the support scheme to protein pea implemented in France.

We present our sample and the estimated ERS-MEMC model first. Then, we present results illustrating how this model works. Finally, we use the simulation model calibrated at the farm level obtained from our estimated ERS-MEMC model for analyzing the public support to protein pea production in France.

Data and model specification details

The ERS-MEMC model presented above is estimated on an unbalanced panel data set containing 3,453 observations of 808 French grain crop producers in the North and North-East

of France, over the years 2007 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: sugar beet, alfalfa, protein pea, rapeseed, winter wheat, corn and spring barley, which represent more than 80% of the total acreage in the considered area.

The variable input aggregate accounts for the use of fertilizers, pesticides and seeds. Our data report the expenditures of these inputs at the farm level. The corresponding aggregate price index is computed as a standard Tornqvist index from the fertilizer, pesticide and seed price indices issued by the French Department of Agriculture at the country level. When a farmer doesn't produce a crop, the corresponding output price and input price index are unobserved. These missing prices are approximated by the yearly average of the corresponding observed prices. All prices are deflated by the hired production services price index (base 1 in 2007) obtained from the French department of Agriculture. This aggregated price index mainly depends on the price indices of machinery, fuel and hired labor. These are the main inputs implicitly involved in the acreage management cost function. Climatic variables are provided at the municipality level by *Météo France*, the French national meteorological service. Farmers' crop price expectations are defined by the corresponding lagged prices, according to a naïve anticipation scheme.[16]

Figure 1 depicts the three levels nesting structure adopted for the considered seven crop set. In a first level we distinguish a cereal group composed of wheat, corn and barley, and a group of rotation entry crops composed of sugar beet, alfalfa, peas and rapeseed. This structure is intended to reflect the basic rotation scheme of grain and industrial crop producers in France.

[16] Robustness checks demonstrate that anticipation scheme choices mostly impact estimates of the probability distribution of input use flexibility parameters $\alpha_{k,i}^{x}$, with very limited effects on our main results.

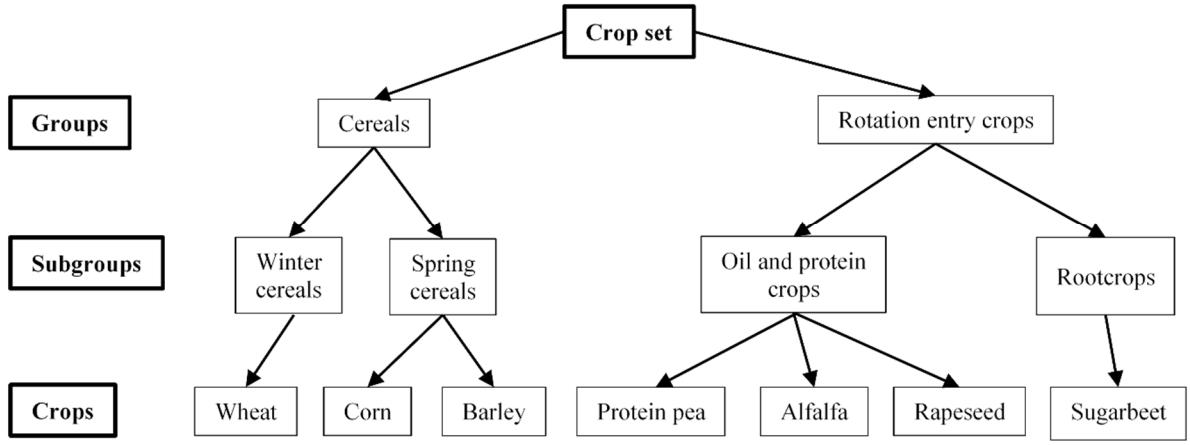


Figure 1. Nesting structure of the acreage choice model

In a second level, the cereal group is split into two subgroups: winter cereals on the one hand and spring cereals on the other hand, in order to account for the differences in planting seasons. The 'rotation entry crop' group is split into an 'oilseeds and protein crops' subgroup and a subgroup including only sugar beet (the only root crop considered here). Wheat, which is the only winter cereal, is used as the benchmark crop. The 15 most frequently observed regimes, out of 78 regimes present in the original dataset, were considered for selecting our estimation sample, which corresponds to 87% of observations included in the original dataset.

All farmers grow wheat and barley and most of them (95.9%) grow at least two additional crops. The most frequent regimes (regimes 2, 3 and 5) 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 6 years of our sample: only 8 out of 808 farmers have not changed their production regime. The average gross margins associated to each regime are reported in the last column of table 1. Interestingly, the most frequent regimes are not the ones that yield the highest average gross margin per hectare. For instance, regime 13 is characterized by the highest observed gross margin on average, but has been adopted in only 2.4% of the observations. This comes to illustrate the fact that farmers' production choices are

driven by factors other than gross returns, such as acreage management and regime fixed costs.

Our data configuration illustrates an important advantage of specifying regime fixed cost as the sum of the fixed costs of the crops produced in the considered regime. According to table 1, the less frequently produced crop (*i.e.*, protein peas) is produced in at least 28% of our observations while 7 production regimes out of 15 are adopted in less than 3% of our observations, regime 15 being adopted in less than 0.4% of our observations. The probability distribution of regime specific fixed costs could not be estimated accurately with our dataset on a pure per regime basis whereas that of crop fixed costs can be.

Our ERS-MEMC model describes 20, actual or counterfactual, production choices. It features 80 exogenous variables, 37 farm specific random parameters and 20 error terms.[17] Its 786 parameters are estimated based on the 52,611 production choices reported in our dataset. Our estimation results are comprehensively presented in Online Supplementary Appendix B. We just discuss their main features in what follows.

[17] Excluding those of the regime choice model, their probability distribution being known by assumption.

Table 1. Descriptive Statistics

Regime number	Regime crop number	Average crops acreage shares per regime							Regime frequency	Average gross margin (€/ha)[b]
		Winter wheat	*Corn*	*Barley*	*Sugar beet*	*Alfalfa*	*Protein pea*	*Rapeseed*		
1	7	0.32	0.07	0.21	0.12	0.09	0.06	0.13	04.3%	998
2	6	0.33		0.20	0.15	0.11	0.06	0.15	13.8%	1071
3		0.32	0.07	0.23	0.14	0.10		0.14	09.0%	980
4		0.40	0.09	0.19	0.11		0.07	0.14	02.4%	1009
5	5	0.32		0.25	0.17	0.11		0.15	37.7%	1058
6		0.40	0.13	0.20	0.11			0.17	05.4%	1023
7		0.40	0.17	0.18			0.07	0.17	02.6%	940
8		0.40		0.20	0.14		0.07	0.19	02.6%	1077
9		0.30		0.23	0.22	0.15	0.10		00.8%	1069
10	4	0.50	0.14	0.14				0.22	06.3%	881
11		0.38		0.28	0.15			0.19	06.6%	1051
12		0.42		0.26			0.09	0.24	01.6%	873
13		0.30		0.28	0.26	0.16			02.4%	1150
14	3	0.44		0.26				0.30	04.1%	841
15		0.50		0.28	0.22				00.4%	1115
Production frequency		100%	30.0%	100%	85.4%	67.9%	28.1%	96.4%		
Average acreage share[a]		0.35 (0.09)	0.03 (0.07)	0.23 (0.09)	0.13 (0.08)	0.08 (0.06)	0.02 (0.03)	0.16 (0.07)		
Average acreage share if produced[a]		0.35 (0.09)	0.11 (0.09)	0.23 (0.09)	0.15 (0.07)	0.11 (0.04)	0.06 (0.03)	0.17 (0.07)		
Average gross margin (€/ha)[a,b]		944 (327)	957 (452)	985 (321)	1747[d] (393)	553 (286)	823 (275)	891 (271)		
Average yield (t/ha)[a]		8.55 (1.02)	9.31 (1.78)	8.46 (0.96)	93.77 (13.80)	12.59 (2.08)	4.70 (1.26)	3.82 (0.67)		
Average price (€/t)[a]		155 (31)	133 (34)	147 (35)	25[c] (3)	72 (15)	196 (25)	321 (64)		
Average production costs[a]		438 (97)	308 (76)	316 (80)	545 (130)	350 (129)	241 (74)	407 (88)		

A Empirical standard deviation in parentheses, B Area-based coupled payments are included in crop gross margins (see Table 2).

Estimation results

The parameter estimates of the yield, input demand, acreage shares and regime choice equations are reported in tables 2 to 4. As shown in table 2, the means of the random parameters representing potential yields, $\beta^{y}_{k,i}$, are precisely estimated (table 1). More importantly, the variances of these parameters are also statistically 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), which illustrates the importance of unobserved farm heterogeneity in our sample. The parameter estimates of the input demand equations, also reported in table 2, 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 effects of changes in netput prices on crop input uses and target yield levels.

Turning to the acreage share equations in table 3, again, the expectations and variances of the random parameters are precisely estimated. They show significant heterogeneity in acreage choice patterns across farms. Also, conditions $\alpha^{s}_{k|(h),i} \geq \alpha^{s}_{(h),i} \geq \alpha^{s}_{i} > 0$ hold on average, indicating that the estimated acreage model is well-behaved. Parameters $\alpha^{s}_{k|(h),i}$ define the flexibility of the crop acreage choices within the crop subgroups considered in the 3 level nMNL acreage choice model that is part of the estimated ERS-MEMC model.

Finally, as shown in table 4, the probability distribution of crop fixed costs $\beta^{c}_{k,i}$ and scale parameter σ_{i} are relatively precisely estimated in the regime choice model. The estimated variability across crops and farms of the crop fixed costs confirms their being significant drivers of farmers' regime choices. It is worth noting that the estimated mean fixed cost of alfalfa is significantly lower than those associated to other crops. Two main reasons might explain this result. First, alfalfa is planted for at least two years and requires farmers' intervention mostly at

planting and harvesting. Second, in the *Marne* region, the alfalfa downstream (dehydration) industry generally takes on harvest operations, which comes to decrease farmers' workload significantly.

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

	Winter wheat	*Corn*	*Barley*	*Sugar beet*	*Alfalfa*	*Protein pea*	*Rape-seed*
Yield supply model							
Error term $\varepsilon_{k,it}^{y}$							
Std dev	0.85*	1.76*	0.79*	12.15*	2.08*	1.17*	0.57*
	(0.01)	(0.04)	(0.01)	(0.15)	(0.02)	(0.03)	(0.01)
Potential yield $\beta_{k,i}^{y}$							
Mean	8.68*	8.89*	8.48*	93.61*	13.08*	4.50*	3.91*
	(0.01)	(0.03)	(0.01)	(0.25)	(0.02)	(0.02)	(0.01)
Std dev	0.33*	0.60*	0.34*	6.32*	0.39*	0.40*	0.17*
	(0.05)	(0.10)	(0.05)	(0.78)	(0.08)	(0.06)	(0.02)
Input demand model							
Error term $\varepsilon_{k,it}^{x}$							
Standard deviation	0.62*	0.57*	0.54*	0.93*	0.97*	0.63*	0.65*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Input requirement $\beta_{k,i}^{x}$							
Mean	4.54*	3.30*	3.03*	6.11*	3.83*	2.74*	4.73*
	(0.02)	(0.02)	(0.01)	(0.03)	(0.04)	(0.02)	(0.02)
Standard deviation	0.35*	0.36*	0.26*	0.63*	0.85*	0.41*	0.39*
	(0.06)	(0.05)	(0.03)	(0.08)	(0.12)	(0.06)	(0.05)
Input use flexibility $\alpha_{k,i}^{x}$							
Mean	0.65*	0.49*	0.19*	0.24*	0.51*	0.85*	1.67*
	(0.01)	(0.03)	(0.01)	(0.00)	(0.084)	(0.02)	(0.02)
Std dev	0.15*	0.34*	0.07*	0.06*	0.69*	0.29*	0.38*
	(0.03)	(0.08)	(0.01)	(0.01)	(0.28)	(0.06)	(0.06)

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

Table 3. Selected Parameter Estimates of the Acreage Share Models[a]

Crop level random terms	*Winter wheat*	*Corn*	*Barley*	*Sugar beet*	*Alfalfa*	*Protein pea*	*Rape-seed*
Error term $\varepsilon^{s}_{k,it}$							
Standard deviation	0	6.41* (0.16)	6.14* (0.09)	4.86* (0.07)	5.22* (0.08)	6.37* (0.17)	5.53* (0.06)
Acreage share shifters $\beta^{s}_{k,i}$							
Mean	0	8.66* (0.11)	3.58* (0.07)	18.14* (0.12)	4.83* (0.09)	11.79* (0.07)	5.73* (0.10)
Standard deviation	0	2.28* (0.31)	1.57* (0.27)	2.91* (0.38)	2.03* (0.27)	1.56* (0.26)	2.38* (0.35)

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 entry crops**	**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	0.058* (0.001)	0.088* (0.002)	0.120* (0.004)	0.611* (0.068)	0.180* (0.003)
Standard deviation	0.018* (0.003)	0.032* (0.006)	0.057* (0.010)	0.569* (0.190)	0.066* (0.010)

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

Table 4. Parameter Estimates of Regime Choice Models

	Crop fixed costs $\beta^{c}_{k,i}$						**Scale parameter** σ_i
	Winter wheat	***Corn***	***Barley***	***Sugar beet***	***Alfalfa***	***Peas***	***Rape-seed***	
Mean[a]	0	2.66* (0.16)	0	0.85* (0.22)	-6.67* (0.35)	1.33* (0.03)	-2.73* (0.11)	1.01* (0.04)
Std dev[a]	0	4.00* (0.62)	0	5.18* (0.74)	8.80* (1.46)	0.71* (0.11)	2.52* (0.40)	0.47* (0.10)

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 farm specific "calibrated" models to be used for simulation purposes (Koutchadé *et al*, 2018). In this study, the specific parameter $\boldsymbol{\gamma}_i$ of farm *i* is calibrated as the mode of its (simulated) probability distribution conditional on what is known about farm *i* in

the data, that is to say following a ML criteria conditional on $(\mathbf{w}_{it}^{+}, r_{it}, \mathbf{z}_{it})$ for $t = 1,...,T$.

The estimated farmer specific models allows computing fitting criteria, Sim-R², which are reported in table 5. Sim-R² criteria measure the quality of the prediction of the observed choices of farmers by the farm specific "calibrated" models. 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. The Sim-R² criteria tend to show that the proposed model offers a satisfactory fit to our data. Using the estimated farmer specific models to predict the regime choices observed in our data, we find our model to exhibit a relatively good predictive power with 73% of regime choices correctly predicted.

Table 5. Fitting Criteria (*Sim-R²*)

	Winter wheat	***Corn***	***Barley***	***Sugar beet***	***Alfalfa***	***Peas***	***Rape-seed***
Yield supply models	0.37	0.27	0.39	0.34	0.13	0.35	0.33
Input demand models	0.51	0.40	0.47	0.50	0.35	0.46	0.51
Acreage share models		0.94	0.81	0.86	0.81	0.90	0.77

Price elasticities of crop acreages

Table 6 reports the estimated means of own price crop acreage elasticities per regime. These elasticities are larger than those commonly found in the literature, probably due to the disaggregation level of our data. More importantly, observing how crop regime specific acreage elasticities within production regimes vary across regimes allows us to illustrate the main features distinguishing ERS-MEMC models from their CR-MEMC counterparts. These estimates display significant differences across production regimes. In particular, crop acreage own price elasticities generally grow with the number of crops produced in the considered production regime. The higher the number of crops, the more farmers can make use of crop acreage substitution opportunities. We can also notice that, the more the considered regime

contains rotation starting crops, the more rapeseed acreage choices are responsive to rapeseed price. This elasticity range, on average, from 0.43, when rapeseed is the only rotation starting crop in the regime (as in regimes 10 and 14), to 1.23, in regimes with 4 rotation entry crops (as in regimes 1 and 2). Similarly, barley acreages are much more responsive to changes in barley price in regimes including corn (regimes 1, 3 and 4 for instance) than in regimes without corn (regimes 2, 5 and 8 for instance). In fact, corn and barley are the only spring cereals in the considered crop set.

Table 6. Per Regime Average Own Price Crop Acreage Elasticities

Regime			Per regime average own price crop acreage elasticities						
Number	***Frequency***	***Crop number***	***Winter wheat***	***Corn***	***Barley***	***Sugar beet***	***Alfalfa***	***Protein pea***	***Rape-seed***
1	04.3%	7	0.56	5.10	1.22	1.93	0.93	1.16	1.23
2	13.8%		0.54		0.66	1.92	0.93	1.16	1.23
3	09.0%	6	0.56	5.10	1.22	1.89	0.88		1.22
4	02.4%		0.55	5.10	1.22	1.84		1.08	0.97
5	37.7%		0.54		0.65	1.89	0.88		1.19
6	05.4%		0.55	5.10	1.21	1.78			0.70
7	02.6%	5	0.53	5.10	1.21			1.03	0.80
8	02.6%		0.54		0.65	1.84		1.08	0.97
9	00.8%		0.53		0.64	1.78	0.71	1.02	
10	06.3%		0.52	5.10	1.20				0.43
11	06.6%	4	0.53		0.64	1.77			0.70
12	01.6%		0.51		0.63			1.03	0.79
13	02.4%		0.52		0.64	1.69	0.49		
14	04.1%	3	0.50		0.63				0.43
15	00.4%		0.49		0.62	0.89			

Adapting the decomposition proposed by McDonald and Moffit (1980) for standard Tobit models to our ERS-MEMC model allows computing elasticities accounting for the impact of crop prices both on acreages within any given regime and on switches in production regimes.[18] The average acreage own price elasticities in our farm sample and their components are reported in table 7. A significant part of the price effects on acreages can be due to the inclusion or not of crops in the production regimes chosen by farmers. For protein peas notably, which represent

[18] The elasticity formulas are provided in Online Supplementary Appendix A4.

a small part of acreages in the considered area, changes in production regimes account for about one fifth of the estimated price elasticities. However, changes in the production regimes can also have significant effects 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 wheat	***Corn***	***Barley***	***Sugar beet***	***Alfalfa***	***Protein pea***	***Rape-seed***
Average crop acreage own price elasticities[a]	0.55 (0.22)	5.32 (3.90)	0.77 (0.44)	2.08 (1.33)	1.08 (0.65)	1.38 (0.56)	1.02 (0.45)
Due to changes in acreages within production regimes[b]	0.55 (100%)	5.10 (96%)	0.76 (99%)	1.85 (89%)	0.87 (81%)	1.11 (80%)	0.96 (94%)
Due to changes in production regimes[b]	0.00 (0%)	0.22 (4%)	0.02 (1%)	0.23 (11%)	0.21 (19%)	0.27 (20%)	0.06 (6%)

A Sample standard deviations in parentheses. B Shares in global elasticities in parenthesis.

Estimated effects of the support of pea production on observed pea acreages in France

Our sample data covers a period where two distinct crop payments schemes were implemented in France (see table 8). Between 2007 and 2009, part of the payments were still coupled to production for cereals and oilseeds crops. These payments amounted to around 100€/ha for wheat, barley, corn and rapeseed. At that time, protein pea benefited from the same 100€/ha payment plus a 50€/ha premium. From 2010, the decoupling of CAP payments was completely achieved in France, implying the removal of the area-based payment for cereals and oilseeds crop. The protein pea premium was however maintained at 155€/ha in 2010 and was increased up to 183€/ha in 2011.

Table 9 reports simulation results aimed at assessing the effects of the area-based payments specifically aimed at promoting protein pea production in France, that is to say 50€/ha from 2007 to 2009 and 163€/ha on average from 2010 to 2011. These results show that the specific support to protein pea had a significant impact on pea acreages in France. This support increased

the average pea acreage share by 0.3% (*i.e.*, 25% of the actual average pea acreage) from 2007 to 2009 and by 1.3% (*i.e.*, 42% of the actual average pea acreage) in 2010 and 2011.

Table 8. Average annual crop area-based payments (€/ha), 2007 -- 2011

Year	*Winter wheat*	*Corn*	*Barley*	*Sugar beet*	*Alfalfa*	*Protein pea*	*Rape-seed*
2007	100	99	100	0	0	155	94
2008	97	96	98	0	0	155	103
2009	94	91	93	0	0	147	120
2010	0	0	0	0	0	151	0
2011	0	0	0	0	0	183	0

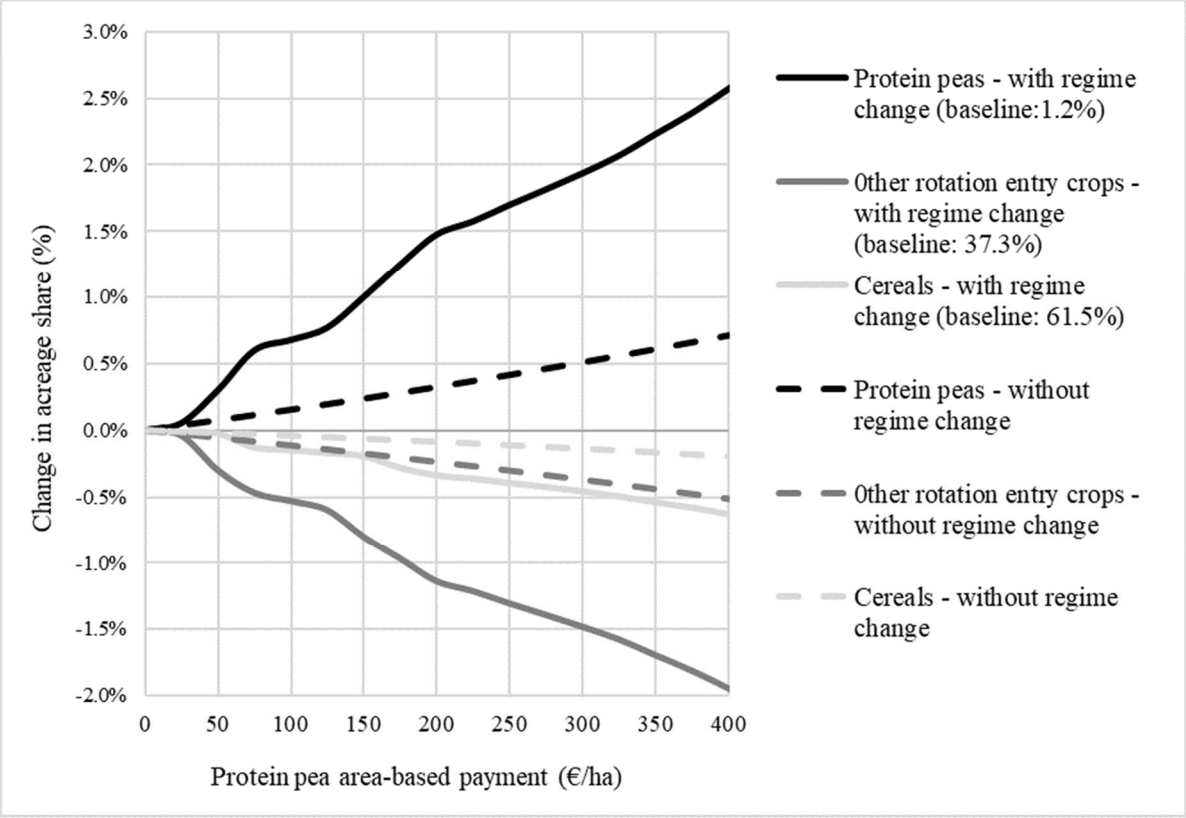
Table 9. Simulated effects of the removal of the protein pea specific area-based payments, 2007–2009 and 2010 – 2011

	Actual situation		**Simulated situation: protein pea area-based specific payment removed**		**Estimated effects of the protein pea area-based specific payment**	
	'07 – '09	**'10 – '11**	**'07 – '09**	**'10 – '11**	**'07 – '09**	**'10 – '11**
Average pea area-based specific payment	50€/ha	163€/ha			+50€/ha	+163€/ha
Acreage shares						
Protein pea	1.2%	3.1%	0.9%	1.8%	+0.3%	+1.3%
Other rotation heads	37.0%	37.2%	36.5%	37.5%	–0.2%	–1.0%
Cereals	61.8%	60.4%	61.9%	60.7%	–0.1%	–0.3%

Estimated effects of hypothetical support levels of pea production on crop acreages in France

Our analysis of the effects of the past support scheme to protein pea production suggests that reasonable increases in the protein pea area-based payments could achieve significant increases in pea production in France. We investigate this point by simulating the effects on farmers' crop acreage choices of protein pea area-based subsidies ranging from 25€/ha to 400€/ha, starting from the situation "with protein pea area-based specific payments removed" as it is estimated above (see table 9). This "free of support to pea production" baseline situation makes it easier to analyze the effects of the considered subsidy scheme. In order to assess the effects of farmers' production regime choices, we present two sets of simulation results. The first one assesses the

effects of area-based subsidies on protein peas by allowing changes in both crop acreage levels and production regimes (represented in solid lines on figure 2). The second set of results allows changes in crop acreage levels only, holding fixed farmers' production regimes (represented in dotted lines on figure 2). Contrasting these results highlights the impacts production regime switches on our simulation outcomes.



According to our results, a 300€/ha subsidy would increase the average pea acreage share by 1.9%, from 1.2%, the average pea acreage share in our baseline scenario, to 3.1%, which represents a significant increase as pea acreages are more than doubled. These additional pea acreages would mainly replace those of other rotation starting crops. The combined average acreage share of rapeseed, alfalfa and sugar beet would decrease by 1.5% while that of cereals would only decrease by 0.5%. This illustrates the point in considering crop – agronomic and management – characteristics when specifying the acreage management cost function.

Comparing these results to those simulated with a model not allowing for changes in crop production regimes shows that about 75% of the increase in the pea acreage would be due to new producers. The simulation results without regime changes actually show that a 0.5% increase in pea acreages would be achieved with a 300€/ha subsidy. Production regime changes also explain another feature of our simulation results. The simulated average pea acreage is increasing in the pea area-based payment but displays changing increasing rates. This pattern is due to the kinks and jumps, which are smoothed by the averaging process, in farmers' pea acreage choices induced by regime switches.

Concluding remarks

The main methodological objective of this article is to present an original modelling framework for dealing with null acreages in MEMC models. This framework is fully consistent from an economic viewpoint and explicitly considers regime fixed costs. These features make the ERS-MEMC model proposed in this article suitable for analyzing and, to some extent, disentangling, the effects of the main drivers of farmers' acreage choices at disaggregation levels at which issues raised by null acreages are pervasive. Our estimation and simulation results notably tend to demonstrate that accounting for production regime choices deeply impact farmers' choices regarding minor crops. Our results also show that crop acreage choices involving minor crops display patterns that cannot be accounted for by the CR-MEMC models currently used for handling null acreage choices.

Our application demonstrates the empirical tractability of the random parameter ERS-MEMC models presented here. Of course, estimating such models raises challenging issues. But, this is also necessary for estimating micro-econometric models suitably accounting for important features characterizing micro-economic agricultural production data, among which frequent null crop acreages and significant, if not massive, unobserved heterogeneity. According to our experience, ML estimators computed with stochastic versions of EM

algorithms, SAEM algorithms in particular, appear to be interesting alternatives to Simulated ML estimators for relatively large systems of interrelated equations such as our random parameter ERS-MEMC model.

This article proposes solutions to methodological issues that could be used for improving micro-econometric analyzes of policies impacting crop acreage choices. The ERS-MEMC models considered in this article not only allow us to disentangle intensive and extensive margin effects, they also allow investigation of crop choice effects. This unique feature is of special interest for analyzing samples containing both specialized and diversified farms as well as for analyzing the effects of policy instruments impacting or targeting farmers' crop set choices. For instance, owing to its positive agronomic effects, crop diversification is a key feature of environmentally friendly crop production systems (*e.g.*, Matson *et al*, 1997; Tilman *et al*, 2002; Lin, 2011). Given current trends in the EU agri-environmental policy, this calls for research on original policy instruments, besides subsidy schemes targeting crops of special agronomic interest, aimed at fostering the adoption of diversified acreages by farmers. The overall structure of our model makes it relevant in this context.

Of course, the ERS-MEMC model presented in this article relies on restrictive assumptions. Even if this is not specific to our MEMC model, our assuming that farmers are risk neutral is admittedly questionable. Adapting our ERS-MEMC model for accounting for the effects of farmers' risk preferences while maintaining its empirical tractability constitutes a research topic of interest. Yet, this topic is also challenging. The empirical tractability of our ERS-MEMC model relies on properties that are specific to the MNL crop acreage share models while the theoretical background of these models strongly relies on farmers' risk neutrality assumption.

Admittedly, CR-MEMC models appear to be less restrictive in this respect, which is of interest when the limitations of these models can be overcome. For instance, CR-MEMC models may be used as reduced form approximations of farmers' choices when the production

regimes of most sampled farms are close to the "all crops" regime, which is the regime on which their specification is primarily based. In such cases, CR-MEMC models enable considering a wide variety of functional forms. For instance, Sckokai and Moro (2006) and Platoni *et al* (2012) consider CR-MEMC models relaxing farmers' risk neutrality assumption.

Also, the ERS-MEMC models presented here rely on restrictive assumptions regarding the dynamic features of multi-crop technologies and farmers' choice process. More work is needed to represent the agronomic effects of crop rotation and crop diversification in our models. Yet, this remark holds for the MEMC models proposed in the literature.

Estimation cost appears to be among the limitations of our modelling framework that need to be addressed first. Significant computing and coding costs make applied research work, such as specification search, tedious and time consuming. Relatively slight modifications of its specification could, however, significantly reduce the estimation burden of the ERS-MEMC model presented in this article. For instance, in our application the covariance parameters of the random parameter vector γ_i account for two-thirds of the estimated parameters. Yet, our results suggest that many elements of γ_i are linked by a few farmer specific "latent productivity factors". Taking advantage of this latent productivity factor structure could significantly reduce the dimension of the estimation problem. Also, estimating the regime choice probability function of our ERS-MEMC model is both tedious and time consuming. Slight alterations of the specification of this model could significantly reduce the cost of its estimation in this regard.

References

Arndt, C. 1999. "Demand for herbicide in corn: an entropy approach using micro-level data." *Journal of Agricultural and Resource Economics* 24(1):204–221.

Arndt, C., Liu, S., and P.V. Preckel. 1999. "On dual approaches to demand systems estimation in the presence of binding quantity constraints." *Applied Economics* 31(8):999–1008.

Britz, W. and L. Arata. 2019. Econometric mathematical programming: an application to the estimation of costs and risk preferences at farm level. Forthcoming in *Agricultural Economics*.

Bayramoglu, B., and R. Chakir. 2016. "The impact of high crop prices on the use of agro-chemical inputs in France: A structural econometric analysis." *Land Use Policy* 55:204–211.

Bues, A., Preissel, S., Reckling, M., Zander, P., Kuhlman, T., Topp, K., C. Watson, K. Lindström F.L. Stoddard and D. Murphy-Bokern. 2013. The environmental role of protein crops in the new common agricultural policy. European Parliament No. 2012-067.

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

Carpentier, A., and E. Letort. 2012. "Accounting for heterogeneity in multicrop micro-econometric models: implications for variable input demand modeling." *American Journal of Agricultural Economics* 94(1):209–224.

Chakir, R., and A. Thomas. 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.

Chambers, R., and R. Just. 1989. "Estimating Multiouput Technologies." *American Journal of Agricultural Economics* 71(4):980-95.

Chavas, J.-P., and M.T. Holt. 1990. "Acreage decisions under risk: the case of corn and

soybeans." *American Journal of Agricultural Economics* 72(3):529–538.

Delyon, B., Lavielle, M., and E. Moulines. 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 I.J. Bateman. 2011. "Structural agricultural land use modeling for spatial agro-environmental policy analysis." *American Journal of Agricultural Economics* 93(4):1168–1188.

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

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

Heckeleï, T., and H. Wolff. 2003. "Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximum entropy." *European Review of Agricultural Economics* 30(1):27-50.

Just, R.E., Zilberman, D., and E. Hochman. 1983. "Estimation of Multicrop Production Functions." *American Journal of Agricultural Economics* 65(4):770-780.

Just, R.E., Zilberman, D., Hochman, E., and Z. Bar-Shira. 1990. "Input Allocation in Multicrop Systems." *American Journal of Agricultural Economics* 72(1):200-209.

Kao, C., Lee, L., and M.M. Pitt. 2001. "Simulated maximum likelihood estimation of the linear expenditure system with binding non-negativity constraints." *Annals of Economics and Finance* 2(1):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 F. Femenia. 2018. “Modeling heterogeneous farm responses to EU biofuel support with a random parameter multi-crop model.” *American Journal of Agricultural Economics* 100(2):434-455.

Kuhlman, T., Helming, J., and D. Oudendag. 2014. Policy impacts on legume-based agriculture at EU level. Paper prepared for presentation at the EAAE Congress, August 26 to 29 2014, Ljubljana, Slovenia

Lacroix, A., and A. Thomas. 2011. “Estimating the environmental impact of land and production decisions with multivariate selection rules and panel data.” *American Journal of Agricultural Economics* 93(3):784–802.

Lavielle, M. 2014. *Mixed effects models for the population approach: models, tasks, methods and tools*. New York: Chapman and Hall/CRC.

Lee, L.F., and M.M. Pitt. 1987. “Microeconometric models of rationing, imperfect markets, and non-negativity constraints.” *Journal of Econometrics* 36(1-2):89–110.

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

Lin, B.B. 2011. “Crop Diversification: Adaptative Management for Environmental Change.” *BioScience* 61(3):183–193.

Magrini, M.-B., M. Anton, C. Cholez, G. Corre-Hellou, G. Duc, M.-H. Jeuffroy, J.-M. Meynard, E. Pelzer, A.-S. Voisin and S. Walrand. 2016. Why are grain-legumes rarely present in cropping systems despite their environmental and nutritional benefits? Analyzing lock-in in the French agrifood system. *Ecological Economics* 126: 152–162.

Matson, P.A., Parton, W.J., Power, A.G., and M.J. Swift. 1997. “Agricultural Intensification and Ecosystem Properties.” *Science* 277: 504–508.

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

McLachlan, G., and T. Krishnan. 2007. *The EM algorithm and extensions*, 2nd. ed. New York: John Wiley & Sons.

Mérel, P., and R. Howitt. 2014. “Theory and application of positive mathematical programming in agriculture and the environment.” *Annual Review of Resources Economics* 6(1):451-470.

Millimet, D.L., and R. Tchernis. 2008. “Estimating high-dimensional demand systems in the presence of many binding non-negativity constraints.” *Journal of Econometrics* 147(2): 384–385.

Moore, M., Gollehon, N. R., and M. B. Carey. 1994. “Multicrop Production Decisions in Western Irrigated Agriculture: The Role of Water Price.” *American Journal of Agricultural Economics* 76(4):385-395.

Moore, M.R., and D.H. Negri. 1992. “A multicrop production model of irrigated agriculture, applied to water allocation policy of the Bureau of Reclamation.” *Journal of Agricultural and Resource Economics* 17(1):29–43.

Oude Lansink, A., and J. Peerlings. 1996. “Modelling the new EU cereals and oilseeds regime in the Netherlands.” *European Review of Agricultural Economics* 23(2):161-178.

Platoni, S., Sckokai, P., and D. Moro. 2012. “Panel data estimation techniques and farm-level data models.” *American Journal of Agricultural Economics* 94(5):1202–1217.

Pudney, S. 1989. *Modelling Individual Choice: the Econometrics of Corners, Kinks and Holes.* Oxford: Basil Blackwell.

Ruud, P.A. 1991. “Extensions of estimation methods using the EM algorithm.” *Journal of Econometrics* 49(3):305–341.

Serra, T., Zilberman, D., Goodwin, B.K., and K. Hyvonen. 2005. “Replacement of agricultural

price supports by area payments in the European Union and the effects on pesticide use." *American Journal of Agricultural Economics* 87(4):870–884.

Sckokai, P., and D. Moro. 2006. "Modeling the reforms of the common agricultural policy for arable crops under uncertainty." *American Journal of Agricultural Economics* 88(1):43–56.

Sckokai, P., and D. Moro. 2009. "Modelling the impact of the CAP Single Farm Payment on farm investment and output." *European Review of Agricultural Economics* 36(2): 395–423.

Stanfield, P.M., Wilson, J.R., Mirka, G.A., Glasscock, N.F., Psihogios, J.P., and J.R. Davis. 1996. "Multivariate input modeling with Johnson distributions." *Proceedings of the 28th conference on Winter simulation, IEEE Computer Society*, pp. 1457–1464.

Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R. and S. Polasky. 2002. "Agricultural sustainability and intensive production practices." *Nature* 418:671–677.

Wales, T.J., and , A.D. Woodland. 1983. "Estimation of consumer demand systems with binding non-negativity constraints." *Journal of Econometrics* 21(3):263–285.