Working Papers SMART - LERECO

Empirical modeling of production decisions of heterogeneous farmers with random parameter models

Philippe KOUTCHADE, Alain CARPENTIER, Fabienne FEMENIA

Working Paper SMART – LERECO N°15-10

October 2015

UMR INRA-Agrocampus Ouest **SMART** (Structures et Marchés Agricoles, Ressources et Territoires)
UR INRA **LERECO** (Laboratoires d'Etudes et de Recherches en Economie)

Les Working Papers SMART-LERECO ont pour vocation de diffuser les recherches conduites au sein des unités SMART et LERECO dans une forme préliminaire permettant la discussion et avant publication définitive. Selon les cas, il s'agit de travaux qui ont été acceptés ou ont déjà fait l'objet d'une présentation lors d'une conférence scientifique nationale ou internationale, qui ont été soumis pour publication dans une revue académique à comité de lecture, ou encore qui constituent un chapitre d'ouvrage académique. Bien que non revus par les pairs, chaque working paper a fait l'objet d'une relecture interne par un des scientifiques de SMART ou du LERECO et par l'un des deux éditeurs de la série. Les Working Papers SMART-LERECO n'engagent cependant que leurs auteurs.

The SMART-LERECO Working Papers are meant to promote discussion by disseminating the research of the SMART and LERECO members in a preliminary form and before their final publication. They may be papers which have been accepted or already presented in a national or international scientific conference, articles which have been submitted to a peer-reviewed academic journal, or chapters of an academic book. While not peer-reviewed, each of them has been read over by one of the scientists of SMART or LERECO and by one of the two editors of the series. However, the views expressed in the SMART-LERECO Working Papers are solely those of their authors.

Empirical modeling of production decisions of heterogeneous farmers with random parameter models

Philippe KOUTCHADE

INRA, UMR1302 SMART, F-35000 Rennes, France
ARVALIS Institut du Végétal, Boigneville, France

Alain CARPENTIER

INRA, UMR1302 SMART, F-35000 Rennes, France
Agrocampus Ouest, UMR1302 SMART, F-35000 Rennes, France

Fabienne FÉMÉNIA

INRA, UMR1302 SMART, F-35000 Rennes, France

Corresponding author

Fabienne Fémenia
INRA, UMR SMART
4 allée Adolphe Bobierre, CS 61103
35011 Rennes cedex, France
Email:fabienne.femenia@rennes.inra.fr
Téléphone / Phone: +33 (0)2 23 48 56 10
Fax: +33 (0)2 23 48 53 80

Empirical modeling of production decisions of heterogeneous farmers with random parameter models

Abstract

Evidences of the effects of unobserved heterogeneity in micro-econometric models are now pervasive in many applied economics fields. This article investigates this issue for agricultural production choice models. Farms' and farmers' unobserved heterogeneity can be accounted for in micro-econometric agricultural production choice models by relying on available modeling and inference tools. The random parameter (RP) framework allows achieving this goal in a fairly flexible way. This modeling framework has already been successfully used in numerous empirical studies covering many topics. It simply considers RP versions of standard models. Extensions of the Expectation-Maximization algorithms have been specifically developed in the computational statistics literature for estimating RP models. They appear to be well suited for large statistical models such as micro-econometric agricultural production choice models. The estimation of a RP multi-crop econometric model shows that unobserved heterogeneity matters in a sample of French farmers specialized in cash grain production covering a relatively small geographical area. The key parameters of this RP model significantly vary across farms. Simulation results obtained from the estimated RP model confirm that the sampled farmers' choices respond heterogeneously to homogenous economic incentives. Ignoring this heterogeneity impacts both the distribution and the magnitude of the simulated effects.

Keywords: Unobserved heterogeneity, random parameter models, agricultural production choices, policy simulation, SEM algorithms

JEL classifications: Q12, C13, C15

Modélisation empirique des décisions de productions d'agriculteurs hétérogènes à l'aide de modèles à paramètres aléatoires

Résumé

Les preuves d'effets de l'hétérogénéité inobservée sont désormais omniprésentes dans de nombreux champs de l'économie appliquée. Nous nous intéressons à cette question dans le cadre des modèles de choix de production agricole. L'hétérogénéité inobservée des exploitations et des exploitants agricoles peut être prise en compte dans les modèles micro-économétriques de choix de production agricole en utilisant des outils de modélisation et d'inférence disponibles, basés sur des modèles à paramètres aléatoires (PA). Ce type d'approche a déjà été utilisé avec succès dans de nombreuses études empiriques sur différent sujets. Elle consiste simplement à considérer des versions à PA de modèles standards. Des extensions d'algorithmes EM (*Expectation Maximization*) ont été spécifiquement développées dans la littérature statistique pour estimer les modèles à PA et sont tout à fait adaptées aux modèles statistiques de grande dimension comme les modèles micro-économétriques de choix de production. L'estimation d'un modèle économétrique pluri-cultures à PA montre ici que l'hétérogénéité inobservée est vraiment importante dans un échantillon d'agriculteurs français spécialisés en grandes cultures dans une zone géographique relativement restreinte. Les paramètres clés de ce modèle à PA varient significativement d'une exploitation à l'autre. Les résultats de simulation obtenus à partir du modèle à PA estimé confirment que les agriculteurs de l'échantillon répondent de façon hétérogène à des incitations économiques homogènes. Le fait d'ignorer cette hétérogénéité impacte à la fois la distribution et l'ampleur des effets simulés.

Mots-clés : Hétérogénéité inobservée, modèles à paramètres aléatoires, choix de production agricole, simulation de politique, algorithmes SEM

Classifications JEL : Q12, C13, C15

Empirical modeling of production decisions of heterogeneous farmers with random parameter models

1. Introduction

Farmers face different production conditions due to heterogeneous soil quality or usual climatic conditions across space. They also own different machineries and different wealth levels. Farmers also differ because of their various experiences, educational levels or skills, as well as because they may have different objectives with respect to income risk or with respect to the leisure *versus* labor trade-off. These heterogeneity sources are likely to have important impacts on farmers' production choices. But to control for the effects of these heterogeneity sources is difficult in practice. Potential heterogeneity sources are numerous and many of them are not suitably described in the data sets usually used by agricultural economists. Empirical investigators generally rely on a few variables – *e.g.* farms' size and location; farmers' age and education; and, when available, rough soil quality indices – to control for the effects of many heterogeneity sources on farmers' production choices. As a matter of fact, many important heterogeneity sources are unobserved from an empirical modeling perspective.

Evidences of the effects of unobserved heterogeneity in micro-econometric models are now pervasive in many applied economics fields. During the last two decades applied micro-econometricians have developed tools to estimate models explicitly accounting for the effects of unobserved heterogeneity on economic choices. These tools have already been successfully used in several applied economics domains. *E.g.*, following the pioneering work of McFadden and Train (2000) and of Greene and Hensher (2003), econometric discrete choice models accounting for decision makers' preference heterogeneity are now routinely used in applied studies. Empirical studies highlighting the role of unobserved heterogeneity effects in econometric models can be found, *e.g.*, in labor economics (see, *e.g.*, Heckman, 2001; Heckman and Vitlacyl, 2007; Angrist and Pischke, 2009), in empirical industrial organization (see, *e.g.*, Ackerberg *et al.*, 2007) or in trade economics (see, *e.g.*, Keane, 2009; Eaton *et al.*, 2011). Importantly, the effects of unobserved heterogeneity – *e.g.* firms' latent productivity, consumers' preferences or workers' unobserved abilities – are often shown to affect how the modeled choices or outcomes respond to interest variables. Consumers value differently a given good or firms respond heterogeneously to homogenous economic incentives.

Our view is that similar heterogeneity features characterize agricultural production choices. Farms and farmers are heterogeneous and this heterogeneity is likely to affect the way farmers respond

to, *e.g.*, economic incentives. However, the micro-econometric agricultural production choice models found in the literature largely ignore such impacts of unobserved heterogeneity.

The objectives of this article are twofold. First, we show that tools recently developed by micro-econometricians and statisticians allow the specification and estimation of econometric agricultural production choice models accounting for farms' and farmers' unobserved heterogeneity in a fairly flexible way. To this end we simply use random parameter (RP) versions of standard agricultural production choice models. This modeling approach has been used successfully in other applied econometric fields and has several advantages: such models are easy to interpret. Tools developed in the computational statistics literature, the so-called Stochastic Expectation-Maximization (SEM) algorithms, appear to be convenient for estimating RP agricultural production choice models. Estimated RP models can easily be used for constructing simulation models defined as a sample of farm models with "statistically calibrated" farm specific parameters.

Second, we aim at showing that unobserved heterogeneity effects significantly matter in empirical agricultural production choice models. To this end we supplement our factual arguments by presenting estimation and simulation results. These results are obtained from a RP multi-crop econometric model estimated with a panel data set of French cash grain producers. Obtained results demonstrate that unobserved heterogeneity matters for the modeling of micro-economic agricultural production choices, even within a small area. Key parameters of farmers' choice models exhibit significant variability across farmers. We then use the estimated RP model for "statistically calibrating" a multi-crop simulation model based on the considered farm sample. Simulation results based on this model and on counterparts of this model with fixed or partially fixed parameters are finally produced for comparison purpose. They confirm that the heterogeneity of key choice parameters significantly impacts simulation results.

Means commonly employed by agricultural production economists to cope with unobserved heterogeneity of farms and farmers depend on their modeling approaches and purposes.

Mathematical programming models used to analyze agricultural supply responses to economic policies (or other determinants of farmers' choices) are usually built by considering sets of farms, of small regions or of farm-types. A mathematical programming model is calibrated for each element of the considered set of "farms". This disaggregated calibration procedure allows controlling for farms' and farmers' unobserved heterogeneity. Of course the lack of statistical background of the standard calibration procedures is often pointed out as an important limitation of agricultural supply mathematical programming models (Howitt, 1995; Heckeleï and Wolff,

2003; Heckeleï *et al.,* 2012). However, simulations provided by these models appear to be highly valued by decision-makers. These provide disaggregated results with respect to the simulated effects of agricultural policy measures on farmers' choices across more or less large geographical areas.

The ability of micro-econometric models of agricultural production choices to account for farms' and farmers' heterogeneity is much more limited in general. Standard specifications of econometric agricultural production choice models can be defined as a sum of a deterministic part and of a vector of random error terms. In these models, farmers' responses to economic (or other) incentives are governed by the deterministic part – *i.e.* by a few statistically estimated parameters and a few control variables – and the effects of farms' and farmers' unobserved heterogeneity are "pushed" into additively separable error terms. This often leads to simulation results which are unrealistically homogeneous across farms.

We simply consider RP versions[1] of standard production choice models. Hence we use RP models with standard functional forms but with farmer specific parameters. Standard data sets, even panel data sets, do not permit direct estimation of the individual parameters. The objective of the estimation is thus to characterize the probability distribution of the model parameters across the considered farmer population. These estimates are of primary interest. First, they allow characterizing and investigating farmers production choices under fairly general assumptions. Second, estimated RP models can also be used to construct simulation models in which a parameter vector is "statistically calibrated" for each sampled farmer. As such they may provide reliable alternatives to the calibrated mathematical programming models usually used for investigating the effects of agricultural policy instruments.

RP models, or mixed models, are routinely (and successfully) used for discrete choice modeling (see, *e.g.*, Train, 2009), for evaluating public policies (Heckman and Vitlacyl, 2007; Angrist and Pischke, 2009) or for modeling firm choices (see, *e.g.*, Keane, 2009). But, to our knowledge, they have not been used in agricultural production economics yet.[2]

[1] An alternative approach relies on the introduction in the model of latent variables aimed at representing some unobserved heterogeneity factors, such as productivity indices (Eaton *et al.,* 2011) or ability indices (Caponi, 2011).

[2] The modeling framework used by Oude Lansink (1999) also relies on farmer specific parameters. But it significantly differs from ours in two important respects. First, Oude Lansink seeks to estimate the parameters of each sampled farmer whereas we aim at estimating the distribution of these parameters across the farmer population represented by our sample. Second, we adopt the classical inference framework. Oude Lansink takes advantage of the Maximum

Indeed, the practical estimation of RP models is often challenging for three main reasons: (*a*) the integration of the estimation objective function over the joint probability distribution of the random parameters, *i.e.* the mixing distribution of the RP or mixed model, when this distribution is continuous as it is assumed here, (*b*) the complexity of the resulting estimation objective function and/or (*c*) the dimension of the estimation problem. Econometricians mostly use Simulated Maximum Likelihood (SML) estimators[3] for estimating RP models continuous mixing distribution. But the practical implementation of SML estimators is difficult for mixed multi-crop models. The aforementioned problems are magnified because these models consider relatively large sets of interrelated production choices and outcomes.

Stochastic Expectation-Maximization (SEM) algorithms (see, *e.g.*, McLachlan and Krishnan, 2008) appear to be interesting alternative estimation tools for RP multi-crop models. SEM algorithms are extensions of the (deterministic) Expectation-Maximization (EM) algorithms proposed by Dempster *et al.* (1977) and, as such, they are particularly well-suited for maximizing likelihood functions involving missing variables such as random parameters. They are specifically designed for computing estimators (which are asymptotically equivalent to SML estimators) for mixed parametric models with continuous mixing distributions. They are well suited for estimating multi-crop models because they allow, at least to some extent, taking advantage of the specific structure of these models. Also, numerous SEM algorithms have been proposed in the computational statistics literature, offering a rich toolbox for estimating mixed models. The algorithm used for our empirical application was designed by combining the features of SEM algorithms proposed by Delyon *et al.* (1999), by Meng and Rubin (1993), by Caffo *et al.* (2005) and by Train (2008; 2009). This algorithm was designed so as to be relatively easy to code and to monitor.

The general features of RP models are presented in the first section. The second section presents the multi-crop econometric model that we consider in order to investigate the advantages of accounting for unobserved heterogeneity in agricultural production choice models. Estimation issues are discussed in the third section. The estimation results and their interpretations are provided in the fourth section.

Entropy approach for solving ill-conditioned problems: his estimation/calibration problem is under-identified in the classical sense.

[3] Or, but less frequently, Method of Simulated Moments estimators. Such estimators are more often used for models with latent heterogeneity variables (see, *e.g.*, Eaton *et al.*, 2011; Caponi, 2011).

2. Unobserved heterogeneity and random parameter models

This section presents the main features of RP models. It also introduces important elements to be used in the presentation of the estimation issues. The multi-crop model considered in the empirical section serves as an example. Its estimation makes use of a panel data set with observations indexed by $i = 1,...,N$ for farms and by $t = 1,...,T$ for years.[4] We consider short run production choices of farmers – *i.e.* an acreage (share) choice system and a yield supply system – and we take for granted that farmers' choices rely on heterogeneous determinants.

A RP model is composed of two parts. The first part, the "behavioral model", formally describes the process of interest and defines its statistical characteristics conditional on the considered random parameters and on the exogenous variables. Our micro-econometric multi-crop model considers a sequence of acreage share vectors, $\mathbf{s}_i \equiv (\mathbf{s}_{it} : t = 1,...,T)$ where $\mathbf{s}_{it} \equiv (s_{k,it} : k \in \mathcal{K})$, and the corresponding sequence of yield level vectors, $\mathbf{y}_i \equiv (\mathbf{y}_{it} : t = 1,...,T)$ where $\mathbf{y}_{it} \equiv (y_{k,it} : k \in \mathcal{K})$. The term $\mathcal{K}$ denotes the considered crop set with $\mathcal{K} \equiv \{1,...,K\}$. These choices and outcomes are simultaneously modeled, with $\mathbf{c}_i \equiv (\mathbf{c}_{it} : t = 1,...,T)$ where $\mathbf{c}_{it} \equiv (\mathbf{y}_{it}, \mathbf{s}_{it})$. The second part of the model defines the distribution characteristics of the random parameters conditionally on the exogenous variables.

2.1. Behavioral model and "kernel" likelihood function

The equation

$$\mathbf{c}_{it} = \mathbf{r}(\mathbf{z}_{it}, \mathbf{e}_{it}; \boldsymbol{\lambda}, \mathbf{q}_i) \qquad (1)$$

describes the acreage choices and yield levels $\mathbf{c}_{it}$ as a known response function $\mathbf{r}$ to $(\mathbf{z}_{it}, \mathbf{e}_{it})$, the determinants of these production choices and outcomes. The term $\mathbf{z}_{it}$ contains price and climatic variables. It is observed by the econometrician. The term $\mathbf{e}_{it}$ contains the unobserved determinants of $\mathbf{c}_{it}$. The response function $\mathbf{r}$ is parameterized by a fixed parameter vector $\boldsymbol{\lambda}$ and a farmer specific parameter vector $\mathbf{q}_i$. Equation (1) describes how the choices of farmer *i* and their outcomes, $\mathbf{c}_{it}$, are determined by $(\mathbf{z}_{it}, \mathbf{e}_{it})$ up to the characteristics of this farmer and of his farm,

[4] If panel data sets are not necessary for estimating RP models, identification of the probability distribution of the farm specific parameters is greatly facilitated with multiple observations of the sampled farms.

$\mathbf{q}_i$. This equation can be any agricultural production choice model where part of the usual fixed parameters is replaced by farmer specific parameter vector $\mathbf{q}_i$.

In a short run production choice context the random parameter vector $\mathbf{q}_i$ mainly captures the effects of the farms' natural or quasi-fixed factor endowments, of the production technologies used by the farmer and of farmers' characteristics. Our application considers short-run crop production choices and relies on a short panel data set, *i.e.* with $T = 4$. Since farms and farmers' production technology generally evolves slowly over time, we assume that the parameters $\mathbf{q}_i$ of the production choice model are constant during the considered time period.

Importantly, specification of the role of $\mathbf{q}_i$ in the model of $\mathbf{c}_{it}$ depends on how unobserved heterogeneity effects are expected to affect farmers' choices and outcomes. Standard panel data models generally assume that the effects of $\mathbf{q}_i$ and of $\mathbf{e}_{it}$ are additively separable in $\mathbf{r}$.[5] In this case the so-called "individual effect" $\mathbf{q}_i$ does not affect the effect of $\mathbf{z}_{it}$ on $\mathbf{c}_{it}$, implying relatively homogeneous responses of $\mathbf{c}_{it}$ to changes in $\mathbf{z}_{it}$. Keane (2009) highlights a basic trade-off related to this issue. Econometric models with additively separable random terms, *i.e.* error or parameters, are relatively easily estimated without specifying the parametric distribution of the error term. But standard "individual effect" models are unsuitable when the effect of $\mathbf{z}_{it}$ on $\mathbf{c}_{it}$ actually depends on unobserved characteristics of farmer *i* or of its farm.[6] The RP framework allows for interactions between $\mathbf{z}_{it}$ and $\mathbf{q}_i$ in $\mathbf{r}(\mathbf{z}_{it}, \mathbf{e}_{it}; \boldsymbol{\lambda}, \mathbf{q}_i)$. However the estimation of RP models appears to be difficult without parametric assumptions on the probability distribution of the random terms $(\mathbf{e}_{it}, \mathbf{q}_i)$. It remains uneasy even under such assumptions.

Keane (2009) argues that the use of fully parametric mixed models and of relatively involved inference tools is a reasonable price for buying the opportunity to introduce rich unobserved heterogeneity effects in micro-econometric models. Of course, this trade-off is an empirical issue and is likely to significantly depend on the modeled processes. Our viewpoint is that the empirical evidences accumulated in other applied economics fields suggest that it is worth investigating for agricultural production choice modeling. This is the main topic of this article with a specific focus on the effects of unobserved heterogeneity on farmers' responses to economic incentives.

[5] *E.g.* they assume that $\mathbf{r}(\mathbf{z}_{it}, \mathbf{e}_{it}; \boldsymbol{\lambda}, \mathbf{q}_i) = \mathbf{r}_0(\mathbf{z}_{it}; \boldsymbol{\lambda}) + \mathbf{q}_i + \mathbf{e}_{it}$ for some function $\mathbf{r}_0$.

[6] *I.e.* when $\frac{\partial}{\partial \mathbf{z}} \mathbf{r}(\mathbf{z}_{it}, \mathbf{e}_{it}; \boldsymbol{\lambda}, \mathbf{q}_i)$ actually depends on $\mathbf{q}_i$.

Equation (1) is completed by statistical assumptions in order to define the "behavioral model". It is assumed here that $(\mathbf{z}_{it},\mathbf{q}_i)$, and $\mathbf{e}_{it}$ are independent. This assumption is restrictive but it is standard, at least for reduced form models where $\mathbf{z}_{it}$ describes external factors – such as market prices or climatic events – affecting $\mathbf{c}_{it}$. The random parameter $\mathbf{q}_i$ is assumed to capture the persistent unobserved characteristics of farm/farmer *i* affecting the choice and outcome variable $\mathbf{c}_{it}$. The term $\mathbf{e}_{it}$ is assumed to represent the effects of idiosyncratic shocks on $\mathbf{c}_{it}$, *i.e.* $\mathbf{e}_{it}$ basically is a "standard" error term.

We also assume that the probability density function of $\mathbf{e}_{it}$ is known and parameterized by $\boldsymbol{\psi}$, a parameter vector to be estimated. The probability density function of $\mathbf{e}_{it}$ and equation (1) allow computing $f(\mathbf{c}_{it}\,|\,\mathbf{z}_{it},\mathbf{q}_i;\boldsymbol{\mu})$, the probability density function of $\mathbf{c}_{it}$ conditional on $(\mathbf{z}_{it},\mathbf{q}_i)$ parameterized by $\boldsymbol{\mu}\equiv(\boldsymbol{\lambda},\boldsymbol{\psi})$. The term $f(\mathbf{c}_{it}\,|\,\mathbf{z}_{it},\mathbf{q}_i;\boldsymbol{\mu})$ defines the "kernel"[7] likelihood function at $\boldsymbol{\mu}$ of the parametric RP model of $\mathbf{c}_{it}$.

Equation (1) and the independence assumptions described above define a "behavioral model" which can be used with cross-section data. With panel data additional assumptions are required in order to describe the potential dynamic features of the considered choices. In the simplest case, $\mathbf{z}_{it}$ and $\mathbf{e}_{is}$ are independent conditionally on $\mathbf{q}_i$ for any pair of years (s,t) and the $\mathbf{e}_{it}$ terms are independent across *t*. These assumptions imply that equation (1) describes a static process repeated for $t=1,...,T$.

These assumptions are assumed to hold in our empirical application which deals with cash crop short run production choices and outcomes. The dynamic features of cash crop production are mostly due to crop rotations. Such dynamic effects can be suitably approximated by farm specific parameters such as $\mathbf{q}_i$ when farmers base their production choices on stable rotation schemes. Short run production choices are repeated each year and follow the same process as long as the production technology, crop rotation schemes included, and the quasi-fixed factor endowment do not change. Under the assumptions given above, the joint density of the vector $\mathbf{c}_i$ conditional on $(\mathbf{q}_i,\mathbf{z}_i)$, where $\mathbf{z}_i\equiv(\mathbf{z}_{i1},...,\mathbf{z}_{iT})$, is given by:

$$f(\mathbf{c}_i\,|\,\mathbf{z}_i,\mathbf{q}_i;\boldsymbol{\mu})=\prod_{t=1}^{T} f(\mathbf{c}_{it}\,|\,\mathbf{z}_{it},\mathbf{q}_i;\boldsymbol{\mu} \qquad (2)$$

[7] According to Train's (2008) terminology

Of course, farmers' choices and outcomes are statistically linked across time due to their relying on the same parameter vector $\mathbf{q}_i$. But these choices and outcomes are assumed to be independent across time conditionally on $\mathbf{q}_i$.

2.2. Mixing probability distribution

The second part of a parametric RP model describes the probability distribution of the farmers' specific parameters $\mathbf{q}_i$ conditionally on the observed variables $\mathbf{z}_i$. It is assumed here that $\mathbf{z}_i$ and $\mathbf{q}_i$ are independent. The exogenous variable vector $\mathbf{z}_{it}$ contains exogenous determinants of the production choices and outcomes, *i.e*. prices and climatic conditions, which mostly vary across time. Let $h(\mathbf{q}_i;\boldsymbol{\eta})$ the probability density function of $\mathbf{q}_i$. This function is defined up to the parameter vector $\boldsymbol{\eta}$ to be estimated. The probability distribution of $\mathbf{q}_i$ describes the distribution of $\mathbf{q}_i$ across the considered famers' population. Statistical estimates of $\boldsymbol{\eta}$ allow the investigation of the distribution of the random parameter $\mathbf{q}_i$. These estimates can be used to test the empirical relevance of the RP specification by checking whether $\mathbf{q}_i$ actually exhibits statistically significant variability or not. Estimates of $\boldsymbol{\eta}$ can also be used to interpret the empirical content of the $\mathbf{q}_i$ terms. *E.g*., investigation of the statistical relations among the elements of $\mathbf{q}_i$ may point out some sources of unobserved heterogeneity.

Of course the choice of the "mixing" probability distribution function, *i.e*. the parametric family of the probability distribution of $\mathbf{q}_i$, is crucial to suitably capture the unobserved heterogeneity effects in the considered model. Being related to unobserved variables, this choice basically is an empirical issue. It is usually based on trials with different parametric models.[8]

[8] Using flexible parametric models, *e.g.* finite discrete mixtures of Gaussian models, or non parametric models appears to be difficult in practice. Such models can only be used when the dimension of $\mathbf{q}_i$ is very small and with very large samples.

2.3. Likelihood functions and "statistical calibration" of individual parameters

The probability density function of the dependent variable $\mathbf{c}_i$ conditional on its observed determinants $\mathbf{z}_i$ defines the likelihood function on which the ML estimation framework is based. For RP models, this probability density function is defined as the mean of $f(\mathbf{c}_i \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\mu})$, the assumed "kernel" probability distribution function, over the distribution of the random parameter $\mathbf{q}_i$, the assumed "mixing" distribution:

$$f(\mathbf{c}_i \mid \mathbf{z}_i; \boldsymbol{\theta}) = \int f(\mathbf{c}_i \mid \mathbf{z}_i, \mathbf{q}; \boldsymbol{\mu}) h(\mathbf{q}; \boldsymbol{\eta}) d \qquad (3)$$

The term $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \boldsymbol{\eta})$ is the parameter vector of the considered parametric RP model. The probability density functions $f(\mathbf{c}_i \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\mu})$ and $h(\mathbf{q}_i; \boldsymbol{\eta})$ are defined by the model and generally have simple analytical forms. But the integral in equation (3) can rarely be solved neither analytically nor numerically when $h(\mathbf{q}; \boldsymbol{\eta})$ is continuous, as it is assumed here. Simulation methods need to be used for approximating the RP model individual likelihood functions $f(\mathbf{c}_i \mid \mathbf{z}_i; \boldsymbol{\theta})$ and this explains why the estimation of mixed model is often difficult.

An estimated RP model can also be used for constructing simulation models defined as a sample of heterogeneous farm models, by using the response model (1) and by computing an estimate $\mathbf{q}_i$ for each farm of the sample used for estimating the considered RP model. These estimates of the $\mathbf{q}_i$ terms can be based on simple "statistical calibration" procedures based on a well defined and coherent statistical background: the RP model of $\mathbf{c}_i$ and the statistical estimate of $\boldsymbol{\theta}$ obtained from this model and the data. Such procedures require either to generate random draws from the probability distribution of $\mathbf{q}_i$ conditional on $(\mathbf{c}_i, \mathbf{z}_i)$ or to estimate its density function.

The marginal probability distribution of $\mathbf{q}_i$, characterized by the density function $h(\mathbf{q}_i; \boldsymbol{\eta})$, is the *ex ante* or *prior* distribution of the random parameter. It describes the distribution of $\mathbf{q}_i$ in the considered farmer population. The probability distribution of $\mathbf{q}_i$ conditional on $(\mathbf{c}_i, \mathbf{z}_i)$ is designated as an *ex post* or *a posteriori* density probability distribution. Its probability density function, denoted here as $h(\mathbf{q}_i \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta})$, sums up what is known about $\mathbf{q}_i$ for farmer *i*, according to a simple "Tell me what you do, I'll know you who you are" logic. Variables $(\mathbf{c}_i, \mathbf{z}_i)$ is the information directly brought by the data on farmer *i* and the functional form of $h(\mathbf{q}_i \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta})$

sums up the information brought by the considered RP model. To estimate $h(\mathbf{q}_i \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta})$ just requires to consistently estimate $\boldsymbol{\theta}$ and standard simulation methods. *E.g.*, application of Bayes' rule yields:

$$h(\mathbf{q}_i \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}) = \omega(\mathbf{c}_i, \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\theta}) h(\mathbf{q}_i; \boldsymbol{\eta}) \text{ where } \omega(\mathbf{c}_i, \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\theta}) \equiv \frac{f(\mathbf{c}_i \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\mu})}{f(\mathbf{c}_i \mid \mathbf{z}_i; \boldsymbol{\theta})} \quad (4)$$

As discussed above the term $f(\mathbf{c}_i \mid \mathbf{z}_i; \boldsymbol{\theta})$ can be estimated, with simulation methods if needed, when a consistent estimate of $\boldsymbol{\theta}$ is available.

The term $E[\mathbf{q}_i \mid \mathbf{c}_i, \mathbf{z}_i]$ is the best predictor of $\mathbf{q}_i$ conditional on $(\mathbf{c}_i, \mathbf{z}_i)$ according to the minimum squared prediction error criterion.[9] Provided that

$$E[\mathbf{q}_i \mid \mathbf{c}_i, \mathbf{z}_i] \equiv \int \mathbf{q} h(\mathbf{q} \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}) d\mathbf{q} \quad (5)$$

$E[\mathbf{q}_i \mid \mathbf{c}_i, \mathbf{z}_i]$ can be estimated when a consistent estimate of $\boldsymbol{\theta}$ is available. *E.g.*, draws from $h(\mathbf{q}_i \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta})$ can be obtained from a Metropolis-Hastings simulator (see, *e.g.*, Train, 2009). But this approach is time consuming in our case. Equation (5) suggests estimating $E[\mathbf{q}_i \mid \mathbf{c}_i, \mathbf{z}_i]$ by using an importance sampling approach with $h(\mathbf{q}_i; \boldsymbol{\eta})$ as the proposal density function. *E.g.*, the integral in equation (5) and the terms $\omega(\mathbf{c}_i, \mathbf{z}_i, \mathbf{q}; \boldsymbol{\theta})$ involved in equations (4) can be computed with simulation methods if needed.

An estimate of $E[\mathbf{q}_i \mid \mathbf{c}_i, \mathbf{z}_i]$ provides a prediction of $\mathbf{q}_i$ with a consistent statistical background. One can thus build a simulation model with heterogeneous farms and farmers from the considered sample and the estimated model. It suffices to "statistically calibrate" $\mathbf{q}_i$ for each sampled farm as shown above.[10]

3. Random parameter multi-crop model

This section presents the model to be used in the empirical application. This model is a RP version of an econometric multi-crop model proposed by Carpentier and Letort (2014). It combines a

[9] The mode of $h(\mathbf{q}_i \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta})$ may also be used for estimating $\mathbf{q}_i$, according to the maximum likelihood criterion.

[10] We keep on using the term "calibration" here because the considered estimate of $\mathbf{q}_i$ rely on a limited amount information specific to farm *i*, *i.e.* it only relies on $(\mathbf{z}_i, \mathbf{c}_i)$

Nested MultiNomial Logit (MNL) acreage share model with a system of (dual) quadratic yield supply and input demand functions. It was chosen due to its parameter parsimony and to the easy interpretation of its parameters, as well as because its functional form is particularly smooth. Further details on this multi-crop model and on its theoretical background are provided in Appendix A.

In this article, this model is mainly used as a simple statistical model structured by a few micro-economic assumptions related to farmers' production choices and outcomes. In this respect, we depart from its presentation in Carpentier and Letort (2014) which was focused on its structural interpretation and on its theoretical consistency. As will be discussed below, farmers' acreage choices depend on many factors. These factors cannot be included in a single econometric model and, maybe more importantly, are likely to significantly depend on the farm and on the farmer itself. These observations, which we take for granted, deeply impact our modeling approach, in particular our interpretations of our acreage choice model and of the related estimation results.

3.1. Yield supply and acreage choice models

The considered model is defined according to the assumption that farmers maximize their expected profit in two steps. First they maximize the expected return to variable input uses for each crop under the assumption that this return doesn't depend on the crop acreages. This step provides to the farmer an expected gross margin for each crop. It also provides systems of variable input demand and yield supply functions to the econometrician. Second, farmers allocate their cropland to the different crops according to their specific objectives and constraints. It is expected that farmers' acreage of a crop increases in its own expected gross margin and that crop diversification motives generally prevent farmers to allocate their entire cropland to the most profitable crop.

The following RP multi-crop model satisfies the aforementioned assumptions (provided that some parametric restrictions hold). It is composed of a sub-system of yield supply equations

$$y_{k,it} = \beta^{y}_{k,i} + \mathbf{d}'_{k,it}\boldsymbol{\delta}^{y}_{k} - 1/2 \times \gamma_k w^{2}_{it} p^{-2}_{k,it} + e^{y}_{k,it} \text{ for } k \in \mathcal{K} \quad (6a)$$

and of a sub-system of acreage share models

$$s_{k,it} = \frac{\exp(\rho_{g,i}\pi_{k,it})}{\sum_{\ell\in\mathcal{K}_g}\exp(\rho_{g,i}\pi_{\ell,it})} \times \frac{\exp\left(\alpha_i\rho_{g,i}^{-1}\ln\sum_{\ell\in\mathcal{K}_g}\exp(\rho_{g,i}\pi_{\ell,it})\right)}{\sum_{h\in\mathcal{G}}\exp\left(\alpha_i\rho_{h,i}^{-1}\ln\sum_{\ell\in\mathcal{K}_h}\exp(\rho_{h,i}\pi_{\ell,it})\right)} \quad \text{for } k\in\mathcal{K}_g \text{ and } g\in\mathcal{G} \tag{6b}$$

where

$$\pi_{k,it} = p_{k,it}\beta_{k,i}^{y} + 1/2\times\gamma_k w_{it}^2 p_{k,it}^{-1} - (\beta_{k,i}^{s} + e_{k,it}^{s}) \quad \text{for } k\in\mathcal{K} \tag{6c}$$

Equation (6a) defines the yield of crop *k*, $y_{k,it}$, as a function of the anticipated[11] price of crop *k*, $p_{k,it}$, of the price of an aggregate variable input, w_{it}, and of an error term, $e_{k,it}^{y}$. This yield supply function also depends on crop specific climatic effects through $\mathbf{d}_{k,it}$, a variable vector aggregating monthly climatic variables specific to crop *k*. Given that the fixed curvature parameter γ_k needs to be positive for the yield function to be well behaved, the farmer specific parameter $\beta_{k,i}^{y}$ can be interpreted as the maximum expected yield of crop *k* on farm *i*. This random parameter depends on the natural endowment of the farm, on the production technology used by the farmer as well as on his technical skills. The error term $e_{k,it}^{y}$ captures the effects of random events, *e.g.* sanitary conditions. It is assumed that farmer *i* considers the elements of $\mathbf{e}_{it}^{y}\equiv(e_{k,it}^{y}:k\in\mathcal{K})$ and $\mathbf{d}_{it}\equiv(\mathbf{d}_{k,it}:k\in\mathcal{K})$ as centered random variables when he chooses his acreage.[12]

The yield supply functions presented in equation (6a) are obtained as the solutions to the maximization of the expected gross margin of crop *k* under the assumptions, among others, that the yield function of crop *k* is quadratic in the aggregate variable input level. The congruent variable input demand functions are not considered in the estimated multi-crop model due to data constraints. The term $\pi_{k,it}$ given by equation (6c) is to be interpreted as a profitability measure of crop *k* defined at the land unit level. It includes the expected gross margin of crop *k* as it is defined by the maximization problem of the expected gross margin of crop *k*. The farmer specific parameter $\beta_{k,i}^{s}$ is part of the production cost of crop *k*. It includes part of the expected costs of the

[11] Farmers' price anticipations are assumed to be naïve in the empirical application.

[12] The climatic variables $\mathbf{d}_{k,it}$ are demeaned at the farm level

aggregate variable input as well as unobserved production costs. The error term $e^s_{k,it}$ captures the effects of random (*e.g.* climatic and/or sanitary) events affecting the crop planting costs.

The technical elements related to the definitions of the yield supply functions (6a) and of the expected crop margins (6c) are given in Appendix A.

The acreage share model described in equation (6b) is a two-stage Nested MNL acreage share model. It relies on a partition of the crop set $\mathcal{K} \equiv \{1,...,K\}$ into *G* mutually exclusive crop groups $\mathcal{K}_g$ for $g \in \mathcal{G} \equiv \{1,...,G\}$. Defining the acreage share of group g as $\bar{s}_{g,it} \equiv \sum_{\ell \in \mathcal{K}_g} s_{\ell,it}$ and the acreage share of crop *k* within group g as $s_{k,it} = s_{k|g,it}\bar{s}_{g,it}$ for $k \in \mathcal{K}_g$, it is easily seen that

$$s_{k|g,it} \equiv \frac{\exp(\rho_{g,i}\pi_{k,it})}{\sum_{\ell \in \mathcal{K}_g} \exp(\rho_{g,i}\pi_{\ell,it})} \tag{7a}$$

defines a model for the acreage share of crop *k* in nest g while

$$\bar{s}_{g,it} \equiv \frac{\exp\left(\alpha_i \rho_{g,i}^{-1} \ln \sum_{\ell \in \mathcal{K}_g} \exp(\rho_{g,i}\pi_{\ell,it})\right)}{\sum_{h \in \mathcal{G}} \exp\left(\alpha_i \rho_{h,i}^{-1} \ln \sum_{\ell \in \mathcal{K}_h} \exp(\rho_{h,i}\pi_{\ell,it})\right)} \tag{7b}$$

defines a model for the acreage share of nest g in total cropland. As will be discussed below, the definition of this partition may have different underpinnings.

3.2. Unobserved heterogeneity in acreage choices

The main feature of the acreage share model (6b) lies in the fact that it is a function of the crop expected gross margin vector $\boldsymbol{\pi}_{it} \equiv (\pi_{k,it} : k \in \mathcal{K})$ parameterized by farmer specific parameters, α_i and $\boldsymbol{\rho}_i \equiv (\rho_{g,i} : g \in \mathcal{G})$. According to equation (6c) the centered error term vector $\mathbf{e}^s_{it} \equiv (e^s_{k,it} : k \in \mathcal{K})$ and the farm specific parameter vector $\boldsymbol{\beta}^s_i \equiv (\beta^s_{k,i} : k \in \mathcal{K})$ basically play the roles of unobserved crop specific production costs in $\boldsymbol{\pi}_{it}$. The random parameter $\boldsymbol{\beta}^s_i$ plays in the Nested MNL acreage share model the role played by additively separable individual (random or fixed) effects in linear panel data models. As a result, it plays a crucial role for the prediction performances of the acreage share model (6b), just as the yield potential level vector $\boldsymbol{\beta}^y_i \equiv (\beta^y_{k,i} : k \in \mathcal{K})$ terms play a crucial role for the prediction performances of the yield supply model (6a). Similarly, $\mathbf{e}^s_{it}$ basically is the error term vector of the acreage share model. Note also

that the elements of $\boldsymbol{\beta}_i^s$ and of $\mathbf{e}_{it}^s$ are only defined up to an additive term. The terms $\beta_{K,i}^s$ and $e_{K,it}^s$ are normalized at 0.

Equations (6b) and (7) show that $\rho_{g,i}$ mainly governs the land allocation process within crop group *g* while the allocation process between crop groups is mainly governed by α_i. The marginal effect of the expected gross margin of crop $\ell \in \mathcal{K}_h$, $\pi_{\ell,it}$, on the (log of) acreage share of crop $k \in \mathcal{K}_g$, $s_{k,it}$, *i.e.*

$$\frac{\partial}{\partial \pi_{\ell,it}} \ln s_{k,it} = \begin{cases} \rho_{g,i} - (\rho_{g,i} - \alpha_i) s_{k|g,it} - \alpha_i s_{k,it} & \text{if } \ell = k \\ -(\rho_{g,i} - \alpha_i) s_{\ell|g,it} - \alpha_i s_{\ell,it} & \text{if } h = g \text{ and } \ell \neq k \\ -\alpha_i s_{\ell,it} & \text{if } h \neq g \end{cases} \tag{8}$$

indicates that the responsiveness of the acreage choices to changes in the crop gross margins increases with the parameters α_i and $\boldsymbol{\rho}_i$ (as well as with the relevant crop acreage levels). These marginal effects also show that this acreage share model is well-behaved if $\rho_{g,i} \geq \alpha_i > 0$ for $g \in \mathcal{G}$, this condition ensuring that $\frac{\partial}{\partial \pi_{k,i}} s_{k,i} > 0$.

Finally, a result of Carpentier and Letort (2014) appears to be useful for presenting our interpretation of the RP Nested MNL acreage share model (6b). This result states that $\mathbf{s}_{it}$ can be viewed as the solution in $\mathbf{s} \equiv (s_k : k \in \mathcal{K})$ to the following maximization problem

$$\max_{\mathbf{s}} \left\{ \mathbf{s}'\boldsymbol{\pi}_{it} - C(\mathbf{s}; \vartheta_i, \mathbf{e}_{it}^c) \text{ s.t. } \mathbf{s} \geq \mathbf{0} \text{ and } \mathbf{s}'\boldsymbol{\iota} = 1 \right\} \tag{9a}$$

where

$$\begin{aligned} C(\mathbf{s}; \vartheta_i, \mathbf{e}_{it}^c) = & \sum_{g \in \mathcal{G}} \sum_{k \in \mathcal{K}_g} s_k (\beta_{k,i}^c + e_{k,it}^c) \\ & + \alpha_i^{-1} \sum_{g \in \mathcal{G}} (1 - \alpha_i \rho_{g,i}^{-1}) \bar{s}_g \ln \bar{s}_g + \sum_{g \in \mathcal{G}} \rho_{g,i}^{-1} \sum_{k \in \mathcal{K}_g} s_k \ln s_k \end{aligned} \tag{9b}$$

The term $\boldsymbol{\iota}$ is a conformable unitary vector, the vector $\vartheta_i \equiv (\alpha_i, \boldsymbol{\rho}_i, \boldsymbol{\beta}_i^c)$ collects the elements of the farmer specific random parameters with $\boldsymbol{\beta}_i^c \equiv (\beta_{k,i}^c : k \in \mathcal{K})$, and $\mathbf{e}_{it}^c \equiv (e_{k,it}^c : k \in \mathcal{K})$ collects the acreage share model error terms. The multi-crop model (6) imposes the normalization constraints $\boldsymbol{\beta}_i^c = \mathbf{0}$ and $\mathbf{e}_{it}^c = \mathbf{0}$ because these terms are included in the corresponding of the expected gross margin vector $\boldsymbol{\pi}_{it}$, *i.e.* in $\boldsymbol{\beta}_i^s$ and $\mathbf{e}_{it}^s$.

According to this result, the acreage share vector $\mathbf{s}_{it}$ maximizes in **s** the expected gross revenue of the farm, $\mathbf{s}'\boldsymbol{\pi}_{it}$, minus an entropic cost/penalty term $C(\mathbf{s};\vartheta_i,\mathbf{e}_{it}^c)$ under the total land use constraint, $\mathbf{s}'\boldsymbol{\iota}=1$. $C(\mathbf{s};\vartheta_i,\mathbf{e}_{it}^c)$ is strictly convex in **s** on its admissible set if $\rho_{g,i}\geq\alpha_i>0$ for $g\in\mathcal{G}$. Under this condition $C(\mathbf{s};\vartheta_i,\mathbf{e}_{it}^c)$ can be interpreted as a function of the crop acreage **s** summing up the effects of the acreage diversification motives of farmer *i*. Carpentier and Letort (2014) interpret $C(\mathbf{s};\vartheta_i,\mathbf{e}_{it}^c)$ as an implicit management cost function, and define it as the sum of the unobserved costs and the shadow costs related to binding constraints due to limiting quasi-fixed factor quantities or to bio-physical factors.[13] Since these endowments are highly heterogeneous across farms, the empirical specification of this cost function needs to be farmer specific as much as possible, as it is the case here. All parameters of $C(\mathbf{s};\vartheta_i,\mathbf{e}_{it}^c)$ are assumed to be farmer specific.

In empirical applications however, the parameter ϑ_i of the acreage share model may also capture the effects of other diversification motives of crop acreages. *E.g.*, it may partly capture the effects of risk spreading motives (see, *e.g.*, Chavas and Holt, 1990) in which case the term $C(\mathbf{s};\vartheta_i,\mathbf{e}_{it}^c)$ includes elements of a profit risk premium. This provides further arguments for its specification based on farm specific parameters. Farmers may have heterogeneous attitudes toward risk, financial constraints or personal wealth levels. This also suggests that the empirical estimate of the acreage share model (6b) are to be interpreted as a reduced form model capturing various acreage diversification motives while accounting for the variability of the strength of these motives among the considered farmers' population.

Many multi-crop models proposed in the literature use more flexible functional forms than the one considered here (see, *e.g.*, Chambers and Just, 1989; Oude Lansink and Peerlings, 1996; Moro and Skockai, 2013). As far as short run micro-economic choices are concerned, our viewpoint is that it may be more important to account for unobserved heterogeneity than to use a highly flexible functional form. Roughly speaking, if heterogeneity really matters it may be preferable to use a first order approximation for each sampled farm rather than to use a second order approximation for the whole sample.

[13] $C(\mathbf{s};\vartheta_i,\mathbf{e}_{it}^c)$ also admits a unique minimum in **s** (in its admissible set) and can also be interpreted as a penalty function for deviations from some benchmark/target acreage share vector. Carpentier and Letort (2014) interpret this "minimum cost" acreage share vector as the one that for which the quasi-fixed factor endowment of the farm is best suited.

Note finally that the above discussion suggests that the crop groups are to be defined so that the crops of a given group compete more in the land allocation process than they compete with crops of other groups, their contributions to the farm expected profit apart. Crops compete in the cropland allocation process when they are substitutes for specific purposes of interest to farmers or when they compete for the use of quasi-fixed inputs. *E.g.*, crops may be grouped because they generate similar rotation effects, because they require similar previous crops, because they have highly correlated output prices or yields, or because they have similar cropping schedules.

3.3. Distributional assumptions and likelihood functions

System level notations are required to present the distributional assumptions defining the model considered in the empirical application, *i.e.* $\boldsymbol{\delta} \equiv (\boldsymbol{\delta}_k : k \in \mathcal{K})$, $\boldsymbol{\gamma} \equiv (\gamma_k : k \in \mathcal{K})$ and $\mathbf{p}_{it} \equiv (p_{k,it} : k \in \mathcal{K})$. In order to relate the estimated multi-crop model to the generic model considered in the preceding section, we also define the farmer production choice and outcome vector $\mathbf{c}_{it} \equiv (\mathbf{y}_{it}, \mathbf{s}_{it})$, the exogenous variable vector $\mathbf{z}_{it} \equiv (\mathbf{p}_{it}, w_{it}, \mathbf{d}_{it})$ and the error term vector $\mathbf{e}_{it} \equiv (\mathbf{e}_{it}^y, \mathbf{e}_{it}^s)$. The response function $\mathbf{r}$ considered in the preceding section is given by equations (6). It is parameterized by the random parameter vector $\mathbf{q}_i \equiv (\ln \boldsymbol{\beta}_i^y, \ln \alpha_i, \ln \boldsymbol{\rho}_i, \boldsymbol{\beta}_i^s)$. The counterpart of $\mathbf{r}$ in the considered multi-crop model is also parameterized by the fixed parameter vector $\boldsymbol{\lambda} \equiv (\boldsymbol{\gamma}, \boldsymbol{\delta})$.

Following the arguments presented in the preceding section, the terms $\mathbf{z}_{it}$, $\mathbf{e}_{is}$ and $\mathbf{q}_i$ are assumed to be mutually independent, and $\mathbf{e}_{it}$ and $\mathbf{e}_{is}$ are assumed to be independent for $t \neq s$. Provided that the stochastic events affecting the crop production process are unknown at the time acreage choices are made, it can be assumed that the terms $\mathbf{e}_{it}^y$ and $\mathbf{e}_{it}^s$ are also independent. This implies that $f(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \mathbf{q}_i; \boldsymbol{\mu})$ can be decomposed as:

$$f(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \mathbf{q}_i; \boldsymbol{\mu}) = f(\mathbf{y}_{it} \mid \mathbf{z}_{it}, \mathbf{q}_i; \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\Psi}^y) f(\mathbf{s}_{it} \mid \mathbf{z}_{it}, \mathbf{q}_i; \boldsymbol{\gamma}, \boldsymbol{\Psi}^s) \tag{10}$$

As is standard for error terms, $\mathbf{e}_{it}^y$ and $\mathbf{e}_{it}^s$ are assumed to be normal with $\mathbf{e}_{it}^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}^y)$ and $\mathbf{e}_{it}^s \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}^s)$. The mixing distribution of the model is also assumed to be normal with $\mathbf{q}_i \sim \mathcal{N}(\boldsymbol{\tau}, \boldsymbol{\Omega})$. Due to the log transformation of $\boldsymbol{\beta}_i^y$, α_i and $\boldsymbol{\rho}_i$ in $\mathbf{q}_i$, these terms are indeed assumed to be jointly log-normal. This ensures their strict positivity.

Once again, in order to relate the multi-crop model considered here to the more general framework of the previous section, the fixed parameter vector $\boldsymbol{\lambda}$ and the distinct elements of the covariance matrices $\boldsymbol{\Psi}^y$ and $\boldsymbol{\Psi}^s$, *i.e.* $\boldsymbol{\psi}^y$ and $\boldsymbol{\psi}^s$ with $\boldsymbol{\psi} \equiv (\boldsymbol{\psi}^y, \boldsymbol{\psi}^s)$, are collected in $\boldsymbol{\mu}$. The distinct elements of $\boldsymbol{\tau}$ and $\boldsymbol{\Omega}$ are collected in $\boldsymbol{\eta}$. Finally, $\boldsymbol{\theta} \equiv (\boldsymbol{\mu}, \boldsymbol{\eta})$ defines the "full" parameter vector to be estimated.

The mixing probability distribution density function, $h(\mathbf{q}_i, \boldsymbol{\eta})$, is simply obtained by using the density of normal random vectors. The "kernel" likelihood function of observation (i,t) at $\boldsymbol{\theta}$, $f(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \mathbf{q}_i, \boldsymbol{\mu})$, is then obtained by inverting the function $\mathbf{r}$, by applying equation (2) and by using the probability density function of normal random vectors. Let $\varphi(\mathbf{u}; \mathbf{B})$ denote the probability density function of $\mathcal{N}(\mathbf{0}, \mathbf{B})$ at $\mathbf{u}$. The density of $\mathbf{y}_{it}$ conditional on $(\mathbf{z}_i, \mathbf{q}_i)$ is given by:

$$f(\mathbf{y}_{it} \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\Psi}^y) = \varphi\left(\mathbf{e}_i^y(\mathbf{s}_{it}, \mathbf{z}_{it}; \boldsymbol{\gamma}, \boldsymbol{\delta}); \boldsymbol{\Psi}^y\right) \tag{11a}$$

and that of $\mathbf{s}_{it}$ conditional on $(\mathbf{z}_i, \mathbf{q}_i)$ is given by:

$$f(\mathbf{s}_{it} \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\gamma}, \boldsymbol{\Psi}^s) = \alpha_i^{1-G}\left(\prod_{g \in \mathcal{G}} \rho_{g,i}^{1-K_g}\right)\left(\prod_{k \in \mathcal{K}} s_{k,it}^{-1}\right) \times \varphi\left(\mathbf{e}_i^s(\mathbf{s}_{it}, \mathbf{z}_{it}; \boldsymbol{\gamma}); \boldsymbol{\Psi}^y\right) \tag{11b}$$

Finally, the random parameter vector density is given by:

$$h(\mathbf{q}_i; \boldsymbol{\eta}) = \varphi(\mathbf{q}_i - \boldsymbol{\tau}; \boldsymbol{\Omega}) \tag{11c}$$

The derivation of $f(\mathbf{s}_{it} \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\gamma}, \boldsymbol{\Psi}^s)$ is presented in Appendix B. It makes use of Berry's (1994) device for inverting Nested Logit functions. It is finally assumed that the $(\mathbf{c}_i, \mathbf{z}_i)$ variables are independent across farms in addition to be equi-distributed.

4. Estimation issues

From a theoretical viewpoint, the parameters of our fully parametric RP multi-crop model can be efficiently estimated according to the ML principle. But the ML estimator of $\boldsymbol{\theta}$ is practically "infeasible" for this model. The individual likelihood functions, *i.e.* the $f(\mathbf{c}_i \mid \mathbf{z}_i; \boldsymbol{\theta})$ terms given in equation (3), must be integrated with simulation methods, implying that the estimators of $\boldsymbol{\theta}$ must be simulated counterparts of the standard ML estimator. The $f(\mathbf{c}_i \mid \mathbf{z}_i; \boldsymbol{\theta})$ term can be estimated by

generating *S* independent (pseudo-)random draws from $h(\mathbf{q};\boldsymbol{\eta})$, the $\tilde{\mathbf{q}}_{i,s}(\boldsymbol{\eta})$ terms for $s=1,...,S$. The empirical mean $S^{-1}\sum_{s=1}^{S} f(\mathbf{c}_i \mid \mathbf{z}_i,\tilde{\mathbf{q}}_{i,s}(\boldsymbol{\eta});\boldsymbol{\mu})$ is an unbiased simulator of $f(\mathbf{c}_i \mid \mathbf{z}_i;\boldsymbol{\theta})$.

While econometricians usually employ Simulated ML (SML) estimators in this context, statisticians usually prefer to rely on Stochastic Expectation-Maximization (SEM) algorithms for computing asymptotically equivalent estimators (Jank and Booth, 2003). The asymptotic properties, as *S* and *N* grows to infinity with *S* rising faster than $N^{1/2}$, of the SML and SEM estimators are those of the "infeasible" ML estimator of $\boldsymbol{\theta}$ (Jank and Booth, 2003). This section presents the main features of our estimation strategy and of its practical implementation.

4.1. EM algorithms

The EM algorithm is particularly well suited for computing ML estimators in cases where the model of interest involves hidden variables such as random parameters. It consists in iterating two steps, the Expectation step (E step) and the Maximization step (M step), until numerical convergence. It basically replaces a large ML problem by a sequence of simpler maximization problems. In our case the EM algorithm involves the probability density function of the "complete" dependent variable vector $(\mathbf{c}_i,\mathbf{q}_i)$ conditional on the exogenous variable $\mathbf{z}_i$, $\kappa(\mathbf{c}_i,\mathbf{q}_i \mid \mathbf{z}_i;\boldsymbol{\theta})$. Application of Bayes' law simply yields

$$\kappa(\mathbf{c}_i,\mathbf{q}_i \mid \mathbf{z}_i;\boldsymbol{\theta}) = f(\mathbf{c}_i \mid \mathbf{z}_i,\mathbf{q}_i;\boldsymbol{\mu})h(\mathbf{q}_i;\boldsymbol{\eta}) \tag{12}$$

At iteration *n*, provided that $\boldsymbol{\theta}_{n-1}$ is the value of $\boldsymbol{\theta}$ obtained at the end of iteration $n-1$, the EM algorithm iterates the following steps until numerical convergence:

E step. Compute

$$Q_N(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{n-1}) \equiv \sum_{i=1}^{N} E[\ln \kappa(\mathbf{c}_i,\mathbf{q}_i \mid \mathbf{z}_i;\boldsymbol{\theta}) \mid \mathbf{z}_i,\mathbf{c}_i;\boldsymbol{\theta}_{n-1}] \tag{13a}$$

where

$$E[\ln \kappa(\mathbf{c}_i,\mathbf{q}_i \mid \mathbf{z}_i;\boldsymbol{\theta}) \mid \mathbf{z}_i,\mathbf{c}_i;\boldsymbol{\theta}_{n-1}] \equiv \int \ln \kappa(\mathbf{c}_i,\mathbf{q} \mid \mathbf{z}_i;\boldsymbol{\theta})h(\mathbf{q} \mid \mathbf{z}_i,\mathbf{c}_i;\boldsymbol{\theta}_{n-1})d\mathbf{q} \text{ for } i=1,...N \tag{13b}$$

M step. Update of the value of $\boldsymbol{\theta}$ by finding

$$\boldsymbol{\theta}_n \text{ such that } Q_N(\boldsymbol{\theta}_n \mid \boldsymbol{\theta}_{n-1}) > Q_N(\boldsymbol{\theta}_{n-1} \mid \boldsymbol{\theta}_{n-1}), \text{ if possible} \tag{14a}$$

or by computing

$$\boldsymbol{\theta}_n \equiv \arg\max_{\boldsymbol{\theta}} Q_N(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{n-1}) \tag{14b}$$

We ignore the integration problem involved in equation (13b) for the moment. The E step thus consists in integrating the "complete data" individual log-likelihood functions $\ln \kappa(\mathbf{c}_i, \mathbf{q}_i \mid \mathbf{z}_i; \boldsymbol{\theta})$ over the *ex post* density function $h(\mathbf{q} \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}_{n-1})$ of the random parameters for obtaining $Q_N(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{n-1})$. Equation (14b) defines a standard EM algorithm while condition (14a) defines a Generalized EM algorithm (Dempster *et al.*, 1977). In some cases to define $\boldsymbol{\theta}_n$ by condition (14a) is much less computationally demanding than to maximize $Q_N(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{n-1})$ in $\boldsymbol{\theta}$.

EM algorithms allow taking advantage of the mixed structure of RP models. Equation (13) is specific to models involving hidden variables such as random parameters. In our case it is used to split the M step into two simpler problems, *i.e.* to find $\boldsymbol{\mu}_n$ such that

$$Q_N^c(\boldsymbol{\mu}_n \mid \boldsymbol{\theta}_{n-1}) > Q_N^c(\boldsymbol{\mu}_{n-1} \mid \boldsymbol{\theta}_{n-1}) \tag{15a}$$

where $Q_N^c(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{n-1}) \equiv \sum_{i=1}^{N} E[\ln f(\mathbf{c}_i \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\mu}) \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}_{n-1}]$ and to obtain

$$\boldsymbol{\eta}_n \equiv \arg\max_{\boldsymbol{\eta}} Q_N^q(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1}) \tag{15b}$$

where $Q_N^q(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1}) \equiv \sum_{i=1}^{N} E[\ln h(\mathbf{q}_i; \boldsymbol{\eta}) \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}_{n-1}]$ and $\boldsymbol{\theta}_n \equiv (\boldsymbol{\mu}_n, \boldsymbol{\eta}_n)$. The parameters of the "behavioral model" on the one hand, and those of the "mixing" probability distribution model on the other hand are separately updated in this M step.

4.2. Stochastic EM algorithms

The EM algorithm described above would lead to the ML estimator of $\boldsymbol{\theta}$. The SEM algorithms were proposed to extend the use of the EM algorithms in cases where the E step requires integration by simulation methods. In our application the expectations in equation (15) were integrated with an (self-normalized) Importance Sampling (IS) simulator employed in a similar context by Train (2008; 2009). This IS simulator uses $h(\mathbf{q}; \boldsymbol{\eta}_{n-1})$ as the proposal density function

for estimating expectations of functions of $\mathbf{q}_i$ conditional on $(\mathbf{z}_i, \mathbf{c}_i)$ parameterized by $\boldsymbol{\theta}_{n-1}$.[14,15] It was used for approximating the expectations of the E step, *i.e.* $Q_N^c(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{n-1})$ was approximated by:

$$\tilde{Q}_{N,S_n}^c(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{n-1}) \equiv \sum_{i=1}^{N} S_n^{-1} \sum_{s=1}^{S_n} \sum_{t=1}^{T} \tilde{\omega}_{i,s}(\boldsymbol{\theta}_{n-1}) \ln f\left(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \tilde{\mathbf{q}}_{i,s}(\boldsymbol{\eta}_{n-1}); \boldsymbol{\mu}\right) \tag{16a}$$

while $Q_N^q(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1})$ was approximated by:

$$\tilde{Q}_{N,S_n}^q(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1}) \equiv \sum_{i=1}^{N} S_n^{-1} \sum_{s=1}^{S_n} \tilde{\omega}_{i,s}(\boldsymbol{\theta}_{n-1}) \ln h\left(\tilde{\mathbf{q}}_{i,s}(\boldsymbol{\eta}_{n-1}); \boldsymbol{\eta}\right) \tag{16b}$$

Where

$$\tilde{\omega}_{i,s}(\boldsymbol{\theta}) \equiv \frac{\prod_{t=1}^{T} f\left(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \tilde{\mathbf{q}}_{i,s}(\boldsymbol{\eta}); \boldsymbol{\mu}\right)}{S_n^{-1} \sum_{s=1}^{S_n} \prod_{t=1}^{T} f\left(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \tilde{\mathbf{q}}_{i,s}(\boldsymbol{\eta}); \boldsymbol{\mu}\right)} \tag{16c}$$

The objective function of problem (16a) takes advantage of the panel structure of the data. The objective functions $\tilde{Q}_{N,S_n}^c(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{n-1})$ and $\tilde{Q}_{N,S_n}^q(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1})$ can be interpreted as the log-likelihood functions of standard models. They are simply weighted by the $\tilde{\omega}_{i,s}(\boldsymbol{\theta}_{n-1})$ terms and involve simulated pseudo-observations, the $(\mathbf{c}_i, \mathbf{z}_i, \tilde{\mathbf{q}}_{i,s}(\boldsymbol{\eta}_{n-1}))$ vectors for $i = 1,...,N$ and $s = 1,...,S_n$. To solve in $\boldsymbol{\eta}$ and in $\boldsymbol{\mu}$ the simulated counterparts of problems (15) appeared to be much easier than to directly maximize in $\boldsymbol{\theta}$ the corresponding sample simulated log-likelihood function

$$\tilde{L}_{N,S}(\boldsymbol{\theta}) \equiv \sum_{i=1}^{N} \ln\left(S^{-1} \sum_{s=1}^{S}\left(\prod_{t=1}^{T} f\left(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \tilde{\mathbf{q}}_{i,s}(\boldsymbol{\eta}); \boldsymbol{\mu}\right)\right)\right) \tag{17}$$

The dimension of $\boldsymbol{\theta}$ is quite large in our case and the functional form of $\tilde{L}_{N,S}(\boldsymbol{\theta})$ makes it difficult to split its maximization problem into smaller optimization problems. The Newton-type algorithms usually employed for maximizing the simulated likelihood function in problem (17)

[14] It can be seen as a direct application of equation (4). It was used to estimate the conditional expectations in equations (14)–(15) as well as to calibrate the farmer's specific parameters in our empirical application.

[15] This proposal density is unlikely to be very efficient when the random parameters strongly impact the modeled choice and outcomes. Moreover the use of normal density functions as proposal density is usually unwarranted owing to their light tails. But this proposal density has two main advantages. First, its use is simple. Second, it allows interpreting the obtained estimator of $\boldsymbol{\theta}$ as a Method of Simulated Score estimator which is closely linked to the usual SML estimator (Train, 2008). Statisticians rarely refer to SML estimators. They often approximate the $E[\ln \kappa(\mathbf{c}_i, \mathbf{q}_i \mid \mathbf{z}_i; \boldsymbol{\theta}) \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}_{n-1}]$ terms by directly drawing from the *ex post* density $h(\mathbf{q} \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}_{n-1})$, *e.g.* by using Metropolis-Hasting simulators.

only have local convergence properties. This makes the definition of "good" starting values a crucial and difficult issue.

Deterministic EM algorithms increase the sample log-likelihood at each iteration, implying that they generally lead to a (local) maximum of the considered likelihood function. SEM algorithms do not necessarily monotonically increase the simulated sample log-likelihood due to the simulation noise, but they are expected to do so when S is "large enough".[16]

4.3. SAEM algorithms and ECM algorithms

Numerous extensions of the basic EM algorithms initially proposed by Dempster, Laird and Rubin (1977) are now available in the computational statistics literature. This offers to applied statisticians/econometricians a rich toolbox for designing SEM algorithms specifically adapted to their empirical problems.[17] The SEM algorithm we used for the empirical application was designed so as to simplify the M step as much as possible.

This algorithm is a Stochastic Approximation EM (SAEM) algorithm, the SAEM algorithms consisting in a class of SEM algorithms proposed by Delyon *et al.* (1999) with two main advantages. First, SAEM algorithms are numerically stable despite their requiring integrations relying on simulation methods at each of their iterations. Second, they allow using simplified versions of the M step. In order to compute $\boldsymbol{\mu}_n$ we combined simplifications of the M step proposed by Meng and Rubin (1993) on the one hand and and by Caffo *et al.* (2005) on the other hand. These allow taking advantage of the factorization given in equation (10).

We finally end up with a SEM algorithm for which the elements of $\boldsymbol{\eta}_n$ are computed as weighted empirical means and covariances whereas the elements of $\boldsymbol{\mu}_n$ are defined as weighted empirical covariances or as weighted Feasible Generalized Least Squares estimators. A detailed presentation of this algorithm is provided in Appendix C. The particular design of the SEM algorithm we use was mainly based on practical considerations. Other SEM algorithms may be more efficient from

[16] The main drawback of EM algorithms is their linear convergence rate. They are slower than Newton-type algorithms which enjoy a quadratic convergence rate. However, this comparison only holds locally, *i.e.* within a neighborhood of the likelihood function. Indeed, EM algorithms are known to quickly converge to a neighborhood of the likelihood function but to slowly converge within this neighborhood. This led Ruud (1991) to suggest designing hybrid algorithms with an EM algorithm starting the maximization process and a Newton-type algorithm terminating it.

[17] As well as to their preferences.

a numerical viewpoint. But this algorithm is relatively easy to code, has good theoretical properties and seems to perform well in practice, at least as far as our limited experience proves this.

4.4. Identification issues

Both sub-systems could be estimated separately. The yield supply equation sub-system (6a) would identify the price parameter vector $\boldsymbol{\gamma}$, the climatic effect vector $\boldsymbol{\delta}$ and the parameters of probability distribution of the yield potential levels $\boldsymbol{\beta}_i^y$. The acreage share choices (6b)–(6c) would identify $\boldsymbol{\gamma}$ and the parameters of probability distribution of $\mathbf{q}_i$. However, this identification would be empirically difficult as it would mainly rely on the exogenous variations of functions of the price variable vector $(\mathbf{p}_{it}, w_{it})$, *i.e.* on the variations of the $p_{k,it}$, $p_{K,it}$, $w_{it}^2 p_{k,it}^{-1}$ and $w_{it}^2 p_{K,it}^{-1}$ terms, as well as on the parametric assumptions related to the "kernel" and "mixing" probability distributions of the considered model.

When both sub-systems are considered simultaneously, the probability distribution of $\boldsymbol{\beta}_i^y$ and the price parameter vector $\boldsymbol{\gamma}$ is still mainly identified by the yield supply equation sub-system (6a), although the acreage share model (6b)–(6c) also contributes to this identification. The probability distribution of the farm specific parameters of the acreage management cost function $(\alpha_i, \boldsymbol{\rho}_i, \boldsymbol{\beta}_i^s)$ is mostly identified by the variations of $p_{\ell,it} \times (\beta_{\ell,i} + 1/2 \times \gamma_\ell w_{it}^2 p_{\ell,it}^{-2})$ in the crop gross margins given in equation (6c), *i.e.* by the variations of the price variable vector $(\mathbf{p}_{it}, w_{it})$ and by the implicit variations of $\boldsymbol{\beta}_i^y$ as they are identified by the yield supply sub-system (7a). In this case, the model parameters identified by the yield supply sub-system are mostly "supplied" to the acreage share choice sub-system.

Of course, whether both sub-systems are considered simultaneously or not, the choice of the parametric model of the "mixing" probability distribution, *i.e.* the choice of $h(\mathbf{q}_i; \boldsymbol{\eta})$, also contributes to identify the model parameters. This distributional assumption implicitly constrains the functional form of the correlation between the elements of $\mathbf{q}_i$. Note also that the crop gross margins $\boldsymbol{\pi}_{it}$ are correlated with the random parameters $(\alpha_i, \boldsymbol{\rho}_i, \boldsymbol{\beta}_i^s)$ of the acreage share model, *i.e.* they are endogenous with respect to these random terms. The choice of $h(\mathbf{q}_i; \boldsymbol{\eta})$ implicitly imposes a functional form for this correlation and, as a result, allows controlling for this correlation.

5. Empirical application

As an illustrative application of the approach proposed in this article to account for farm heterogeneity, we use a set of French data to estimate the multi-crop model presented in the second section. These estimation results allow an investigation of the distribution of the random parameters of the model, which comes to illustrate the importance of unobserved heterogeneity in farmers' production choices. We also perform a "statistical calibration" of the model parameters for each sampled farmer in order (*i*) to evaluate the performances of the estimated model and (*ii*) to reveal some potential determinants of the heterogeneity in farmers' behaviors. We then perform some simulations in order to study the impacts and potential implications of the modeling of heterogeneous behaviors on simulation results.

5.1. Data

Our main objective in this empirical application being to investigate the importance of unobserved heterogeneity effects, our data set has been selected so as to contain farms with relatively homogenous production choices. More precisely, we focus on farms highly specialized in grain production: the data set used to estimate our model is a panel data sample of 370 observations of French grain crop producers in the large (geological) Paris basin over the years 2004 to 2007, obtained from the Farm Accountancy Data Network (FADN). It provides detailed information on crop production for each farm: acreage, yield and price at the farm gate. The aggregated input price index is made available at the regional level by the French Department of Agriculture. The climatic variables are provided at the municipality level by Meteo France, the French national meteorological service.

In our application, yield levels and acreage share choices are considered for three (aggregated) crops: soft wheat (crop 1), other cereals (mainly barley and corn, crop 2) and, oilseeds (mainly rapeseed) and protein crops (mainly peas) (crop 0). Crop aggregates are based on agronomic considerations. The basic rotation scheme of the French grain producers is a sequence with three crops as: rootcrops (*e.g.*, potato or sugar beet) or protein crop or oilseed (*e.g.*, rapeseed or sunflower) – winter wheat – secondary cereal (*e.g.*, barley or wheat). Rootcrops require good quality soils which are found in the north of France. Sunflower is grown in the south of France while rapeseed, the other main oilseed crop is grown in the north half of France (our region of interest). Sugar beet and potato acreages were considered exogenous due to production quotas for sugar beet and production contracts for potatoes.

The considered sample only includes observations with strictly positive acreages. This selection rule doesn't lead to significant attrition thanks to the crop aggregation procedure (less than 25% of the observations have been excluded from the sample). Our sample covers the French regions specialized in grain production, with the notable exception of the south-west of France where corn monoculture is the dominant cropping system. The 105 farms of our sample are observed for 2 to 4 years. We assume that farms' attrition is exogenous because the French FADN is constructed as a rotating panel seeking to collect data for 4 years for each sampled farm. Such an exogenous attrition is easily accommodated in our modeling framework. Farms' likelihood functions are computed according to the observed sequences.

5.2. Estimation Results

Our estimations were conducted by using the SAS® software (IML procedure). The recursive step of simulation of SAEM algorithm was implemented using 1000 draws. The algorithm converged without difficulties after 244 iterations. Results were not significantly affected by the use of alternative starting values or by the use of larger numbers of draws.

The estimation results are reported in Table 1 and Table 2. Most parameters, especially the expectations and covariances of the random parameters, $\mathbf{q}_i$, and the variance matrices of the error terms, $\mathbf{e}_{it}$, appear to be precisely estimated. The fixed parameters representing the price effects, $\boldsymbol{\gamma}$, also appear to be relatively precisely estimated.

The probability distribution of the yield equations random parameters, $\boldsymbol{\beta}_i^y$, which represent potential crop yields on each farm, is precisely estimated and the parameter estimates lie in reasonable ranges. This was expected since the yield equation system basically is a regression equation system with individual random terms (Biorn, 2004). The estimated variances and covariances of the $\boldsymbol{\beta}_i^y$ parameters show that these parameters significantly vary across farms while being strongly positively correlated to each other. Yield potentials vary across regions and good growing conditions for a grain crop are also good for the others. The variance of the parameters representing potential yields $\beta_{k,i}^y$ is higher or close to that of error terms $e_{k,it}^y$ for wheat and other cereals, but the variance of $e_{k,it}^y$ is twice that of $\beta_{k,i}^y$ in the oilseeds case. This may reflect at least two points: first, a large part of the heterogeneity in cereals, and notably wheat, yields is due to differences in unobservable characteristics of each farm or farmer; second, given that rapeseed is by far the most important oilseed in northern France, these results may be due to the fact that the

rapeseed yield is more risky than the cereal yield, mostly due to its sensitivity to climatic conditions as well as to bugs and diseases.

Table 1: Selected parameter estimates, yield supply equations

	γ_k	$E[\beta^y_{k,i}]$	$Cov[\beta^y_{k,i}, \beta^y_{\ell,i}]$			$V[e^y_{k,it}]$
			Winter wheat ($\ell=1$)	Other cereals ($\ell=2$)	Oilseeds ($\ell=3$)	
Winter wheat	0.637	8.354	0.914	0.747	0.530	0.480
($k=1$)	(0.279)	(0.230)	(0.383)	(0.325)	(0.230)	(0.072)
Other cereals	0.808	8.363	0.747	1.010	0.493	0.988
($k=2$)	(0.292)	(0.301)	(0.325)	(0.491)	(0.205)	(0.156)
Oilseeds	0.994	6.255	0.530	0.493	0.420	0.714
($k=3$)	(0.291)	(0.208)	(0.230)	(0.205)	(0.236)	(0.120)

Note: standard errors are in parentheses

Table 2: Selected parameter estimates, acreage share equations

		Expectation		**Covariances with**		
		$\ln \alpha_i$	$\ln \rho_i$	$\ln \beta^y_{1,i}$	$\ln \beta^y_{2,i}$	$\ln \beta^y_{3,i}$
				Winter wheat	Other cereals	Oilseeds
$\ln \alpha_i$	-2.434	0.177	0.139	0.005	0.011	0.007
	(0.130)	(0.070)	(0.078)	(0.011)	(0.014)	(0.012)
$\ln \rho_i$	-2.179	0.139	0.301	-0.013	-0.002	-0.004
	(0.165)	(0.078)	(0.117)	(0.016)	(0.019)	(0.018)

Note: standard errors are in parentheses

The acreage share equation parameter estimates representing the flexibility of acreage adjustment between cereals and oilseeds, α_i, and between wheat and other cereals, ρ_i, are reported in Table 2. These estimates also range in reasonable ranges. Importantly, the estimate of the mean of ρ_i is

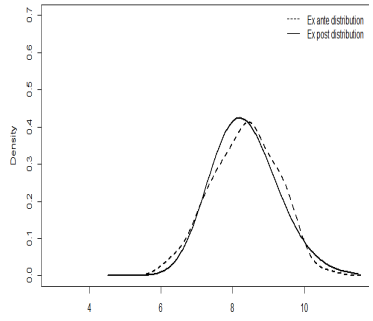
higher than that of α_i. ρ_i being larger than α_i is a sufficient condition for the entropic acreage management cost function, lying at the root of the Nested MNL acreage share function, to be convex. According to the estimates of their respective variances, the α_i and ρ_i parameters significantly vary across farms. This result is important for simulation studies because these parameters largely determine acreage price elasticities in MNL acreage share models. The higher are α_i and ρ_i, the more acreages adjust in response to economic incentives.

5.3. Statistical calibration of individual parameters

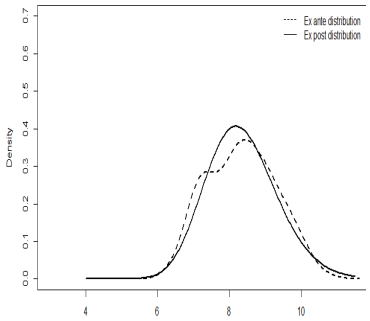
As explained in the first section, the estimated parametric model allows a statistical calibration of the (random) parameters, $\mathbf{q}_i$, for each farm/farmer of the sample. Once the *ex ante* distribution of $\mathbf{q}_i$ in the population has been estimated we "statistically calibrate" the specific parameters for each individual i based on the *ex post* density of $\mathbf{q}_i$. The *ex post* and *ex ante* density of the random parameters $\boldsymbol{\beta}_i^y$, α_i and ρ_i are represented on Figure 1. The two distributions are almost superimposed for all parameters, which reflects a good specification of our model. We can also notice that the distributions of the yield parameters, $\boldsymbol{\beta}_i^y$, appear to be more spread for other cereals than for the two other crops, reflecting a higher heterogeneity of yields between farms for that crop. That might be due to the fact that "other cereals" is an aggregate of various crops (mainly corn and barley), whereas "wheat" is a single crop and "oilseeds" is essentially composed of rapeseed in our sample. The probability distributions of α_i and ρ_i reflects the fact that α_i parameters generally take lower values than ρ_i parameters (this is actually the case for 76% of the farms/farmers, the remaining 24% individuals having ρ_i values almost equal to α_i values).[18] Figure 2 reports the calibrated values of the $\boldsymbol{\beta}_i^y$, α_i and ρ_i parameters together with their confidence intervals for each farm/farmer of our sample. These graphs show that confidence intervals of parameters do not overlap for all individuals, demonstrating that these parameters do actually take different values from one individual to another. This comes to illustrate the heterogeneity, across farms, in potential yields and in the way farmers are able to adjust their acreages in response to economic incentives.

[18] Note that the inequality constraints $\alpha_i < \rho_i$ can be enforced for $i = 1,...,N$ by a suitable parameterization of the acreage share model.

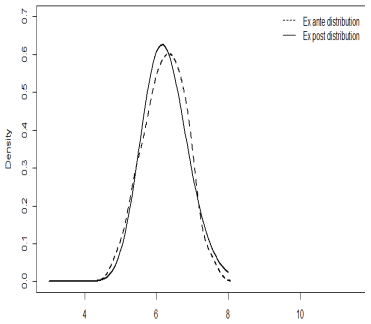
Figure 1: *Ex post* and *ex ante* probability distributions of the random parameters



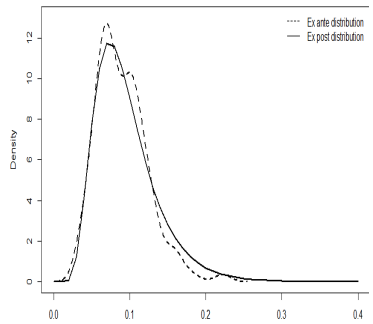
$\beta_{1,i}^{y}$, wheat



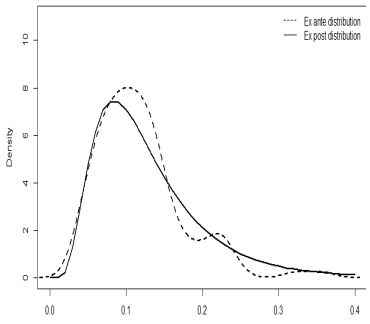
$\beta_{2,i}^{y}$, other cereals



$\beta_{3,i}^{y}$, oilseeds



α_i



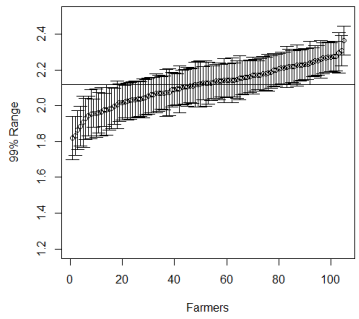
ρ_i

Ex post distribution

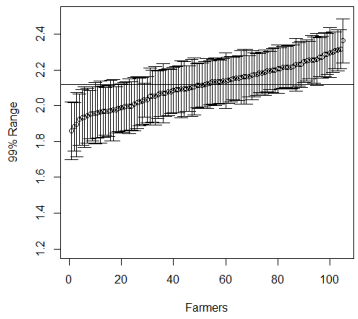
Ex ante distribution

Figure 2: Calibrated values and confidence intervals of individual parameters

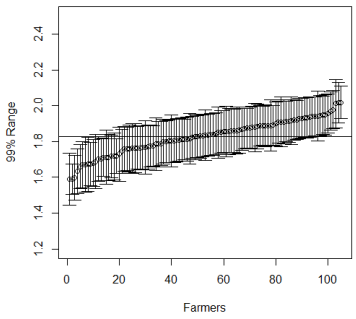
$\ln \beta_{1,i}^{y}$, wheat



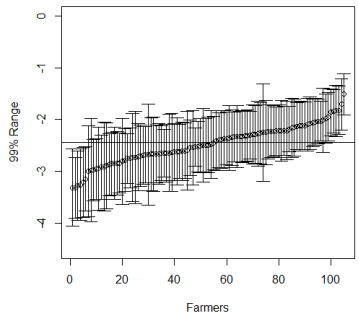
$\ln \beta_{2,i}^{y}$, other cereals



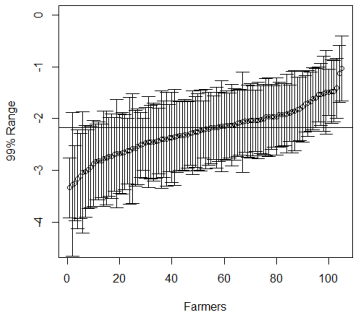
$\ln \beta_{3,i}^{y}$, oilseeds



$\ln \alpha_i$



$\ln \rho_i$



Having calibrated individual parameters for each farm/farmer, *i.e.* having computed an estimate of $E[\mathbf{q}_i \mid \mathbf{z}_i, \mathbf{s}_i]$ for $i = 1, ..., N$, we are able to compute the individual yields and acreages predicted by the NMNL model. These predictions are then used to compute "pseudo R^2" criteria corresponding to the share of the variance of interest variables predicted by the model, and to compare the observed values of these variables to their predicted values. The fitting criteria of the model are reported in Table 3. The model proves to fit well the data, especially for wheat and other cereals with "pseudo R^2" close to 70%.

Table 3: Fitting Criteria of the Model: Pseudo R^2

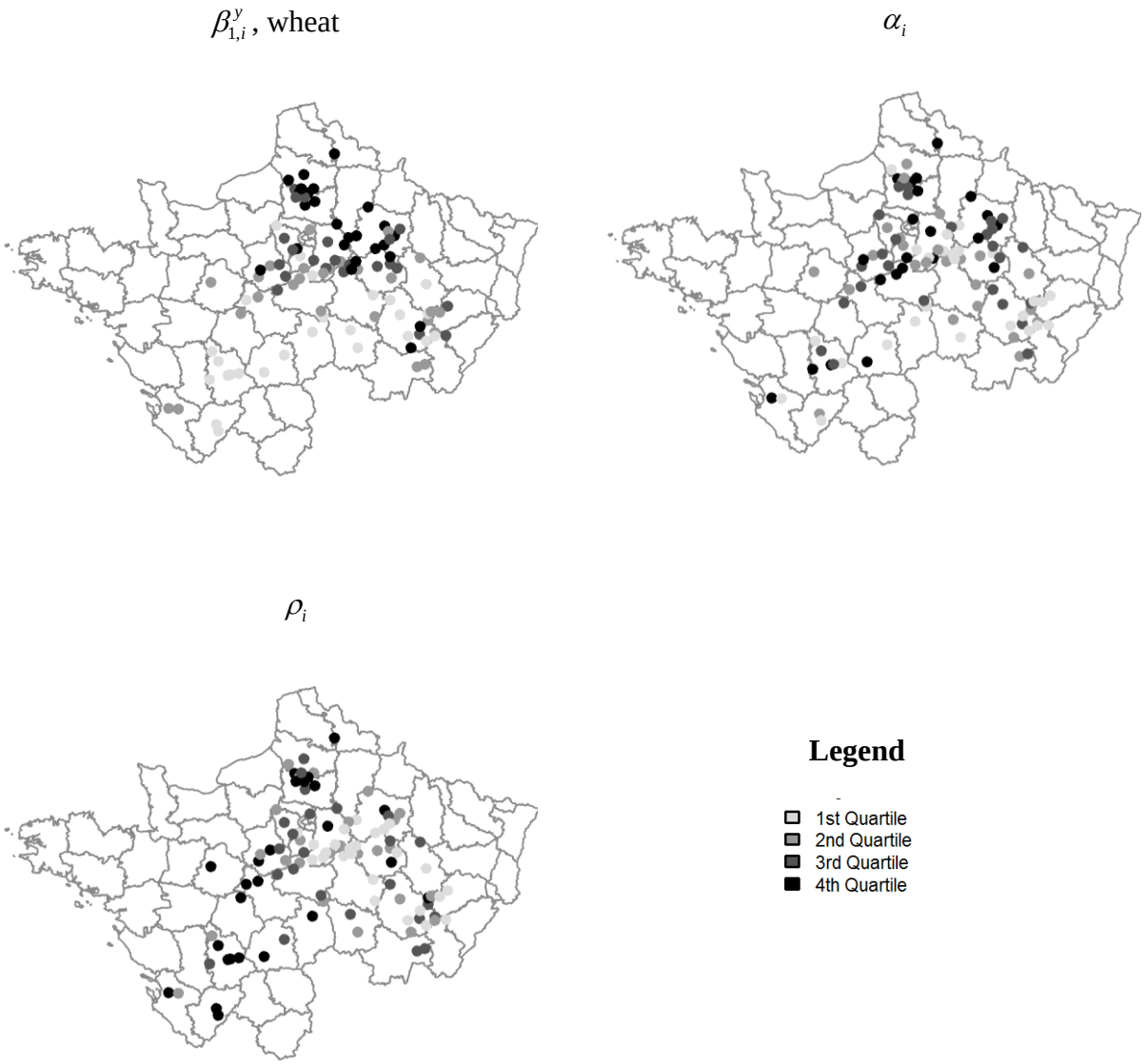
	Pseudo R^2	
	Yields	Acreage shares
	$y_{k,it}$	$s_{k,it}$
Winter wheat ($k=1$)	.79	.79
Other cereals ($k=2$)	.68	.77
Oilseeds ($k=3$)	.61	.43

Up to this point, our estimation results have shown that farmers' behaviors do actually rely on heterogeneous factors. It thus seems crucial to account for heterogeneity in micro-econometric production choice models. If the sources of this heterogeneity were known to econometricians, they could be controlled for through, *e.g.* the use of control variables.[19] However, if some of them are identifiable, heterogeneity sources are multiple and most of them can certainly not be reduced to farm/farmers' observable characteristics.

This point is illustrated on Figure 3 and Table 4. Maps reported on Figure 3 show the calibrated values of three parameters: $\beta_{k,i}^{y}$ for wheat, α_i and ρ_i for each farm of our sample. The top left map clearly shows that the distribution of potential wheat yields exhibits a spatial pattern, the highest yields being located in the North of France. This is in total accordance with what is known about the different agronomic potentials of French regions. Introducing spatial farm characteristics in the model could thus help accounting for some heterogeneity. Farms' localization is however not the only source of heterogeneity in agricultural production choices. This is reflected by the two other maps on Figure 3: the distributions of the α_i and ρ_i parameters across space are different from that of the $\beta_{k,i}^{y}$ parameters. No specific spatial pattern seems to emerge from these maps.

[19] Of course the use of control variables is allowed in our modeling approach. But it is omitted for simplicity as well as for investigating the potential of random parameter models.

Figure 3: Distribution of selected random parameters across the population sample



In a further attempt to qualify the potential sources of farmers' behavior heterogeneity, we computed the correlations between the individual parameters and some observable farms/farmers characteristics considered as exogenous in the model: the amount of farm capital, the root crops acreage and the age of farmer.[20] These correlations are reported in Table 4.

[20] Other variables such as the number of labor hours or the total acreage of the farm have been tested but none of them were significantly correlated to any of the individual parameters.

Table 4: Correlations between random parameters and farmers' characteristics

	$\beta^{y}_{1,i}$ Winter wheat	$\beta^{y}_{2,i}$ Other Cereals	$\beta^{y}_{3,i}$ Oilseeds	α_i	ρ_i
Farm capital	0.15	0.11	0.23	0.16	0.00
	(0.12)	(0.24)	(0.02)	(0.10)	(0.99)
Root crop acreage	0.29	0.19	0.27	0.41	-0.02
	(<0.01)	(0.05)	(<0.01)	(<0.01)	(0.83)
Farmer's age	0.28	0.17	0.24	0.17	-0.06
	(<0.01)	(0.09)	(0.01)	(0.07)	(0.57)

Note: Student's Test p-values are in parentheses

Farm capital is positively and significantly correlated with the $\beta^{y}_{k,i}$ parameter for oilseed, the α_i parameter, and to a lesser extent with the $\beta^{y}_{k,i}$ parameter for wheat and other cereals. This may reflect the fact that farms endowed with more capital are the more productive ones and also own enough machinery to easily adjust their acreages. Two alternative explanations can lie at the root of the positive and significant correlations between root crop acreage and the $\beta^{y}_{k,i}$ parameters. On the one hand, root crops are good preceding crops for wheat and other cereals, implying that root crop acreages may positively impact the cash grain yields through crop rotation effects. On the other hand root crops require high soil quality, implying that large root crop acreages may also indicate soil quality. Root crops can be used as an alternative to oilseeds as previous crops for wheat and other cereals and thus relax some constraints on acreage adjustments which translates into a positive correlation with the α_i parameters. The positive and significant correlations between farmers' age and potential yields might be due to the role played by experience in farmers' skills and abilities, or by generational differences in the intensity of input, notably pesticides, use. All the aforementioned exogenous variables could thus help controlling for part of farm heterogeneity in our production choice model. However, none of the correlations presented in Table 4 is strong enough to conclude that using these control variables would be sufficient to capture all the sources of heterogeneity.

5.4. Simulation Results

Our "statistically calibrated" model is now used to conduct some simulations aimed at illustrating the potential impacts that accounting for farm heterogeneity can have on such model outcomes. We simulate the impacts of changes in crop prices that roughly correspond to the changes that have been observed in France since 2007, namely a 20% increase in wheat and other cereal prices and a 50% increase in oilseeds prices.

Table 5 reports the distribution characteristics of the elasticities of acreages to changes in crops prices in our sample. These elasticities are key parameters determining farmers' responses to price shocks. The calibrated elasticities lie in a reasonable range and reflect the higher flexibility of acreage adjustments within the cereal nest: wheat (respectively other cereals) acreage responds more to a change in other cereals (respectively wheat) price than to a change in oilseed price. Furthermore, the reported quantile values reflect a significant dispersion of elasticities within our sample. One can thus expect each farmer to react differently to the price changes we simulate here, which is not surprising given the variances of the model random parameters.

Table 5: Characteristics of the Distribution of the Acreage Shares Price Elasticities

	Average	Quantiles				
		Q5	Q25	Q50	Q75	Q95
Wheat acreage						
Wheat price	0.42	0.24	0.31	0.39	0.49	0.73
Other cereals price	-0.25	-0.59	-0.31	-0.20	-0.15	-0.11
Oilseeds price	-0.13	-0.21	-0.15	-0.12	-0.10	-0.07
Other cereals acreage						
Wheat price	-0.49	-1.23	-0.67	-0.37	-0.23	-0.14
Other cereals price	0.61	0.20	0.33	0.48	0.80	1.31
Oilseeds price	-0.13	-0.21	-0.15	-0.12	-0.10	-0.07
Oilseeds acreage						
Wheat price	-0.34	-0.79	-0.42	-0.28	-0.20	-0.12
Other cereals price	-0.21	-0.56	-0.26	-0.15	-0.09	-0.05
Oilseeds price	0.46	0.16	0.29	0.40	0.58	0.87

The first column of Table 6 reports the effects on acreages of the changes in crop prices simulated using our RP model. The relative increase in oilseeds price compared to wheat and other cereals

prices lead farmers to reallocate part of their land to this now more profitable crop: among the 46,475 ha devoted to crops in our sample, 766 ha of wheat and 563 ha of other cereals acreages are reallocated to oilseeds which acreage thus increases by 1,330 ha. This represents average variations of 2.1 ha, 1.5 ha and 3.6 ha for respectively wheat, other cereals and oilseeds acreages. However, these variations significantly vary from one farm to another: the increases in oilseeds acreage notably vary between 0.5 ha and 14.5 ha in absolute term, and between 3.3% and 54.4% of initial oilseeds acreages, depending on the farm. These contrasting results come to illustrate the heterogeneity in farmers' response to economic incentives.

Table 6: Simulated Impacts on Acreages of the Price Shock

	RP model	**Fixed parameter model**	**Fixed/RP model**
Wheat Acreage			
Total change (ha)	-766 (-3.7%)	-299 (-1.4%)	-821 (-4.0%)
Average change (ha)	-2.1 (-3.7%)	-0.8 (-1.5%)	-2.1 (-3.9%)
Max change (ha)	+1.9 (+1.9%)	+0.6 (+1.0%)	+1.6 (+2.6%)
Min change (ha)	-11.6 (-17.9%)	-3.5 (-3.9%)	-12.3 (-11.3%)
Other cereals Acreage			
Total change (ha)	-563 (-4.1%)	-59 (-0.4%)	-606 (-4.4%)
Average change (ha)	-1.5 (-4.4%)	-0.2 (-0.4%)	-1.6 (-4.6%)
Max change (ha)	+2.0 (+10.4%)	+1.0 (+4.6%)	+0.5 (+1.0%)
Min change (ha)	-8.4 (-18.8%)	-2.5 (-3.9%)	-7.8 (-12.6%)
Oilseeds acreage			
Total change (ha)	+1330 (+11.1%)	+358 (+3.0%)	+1427 (+11.9%)
Average change (ha)	+3.6 (+13.9%)	+1.0 (+3.1%)	+3.9 (+12.8%)
Max change (ha)	+14.5 (+54.4%)	+3.3 (+6.1%)	+14.4 (+28.3%)
Min change (ha)	+0.5 (+3.3%)	<0.1 (+1.3%)	+0.2 (+4.5%)

Note: Numbers in parentheses correspond to percent changes compared to initial acreages

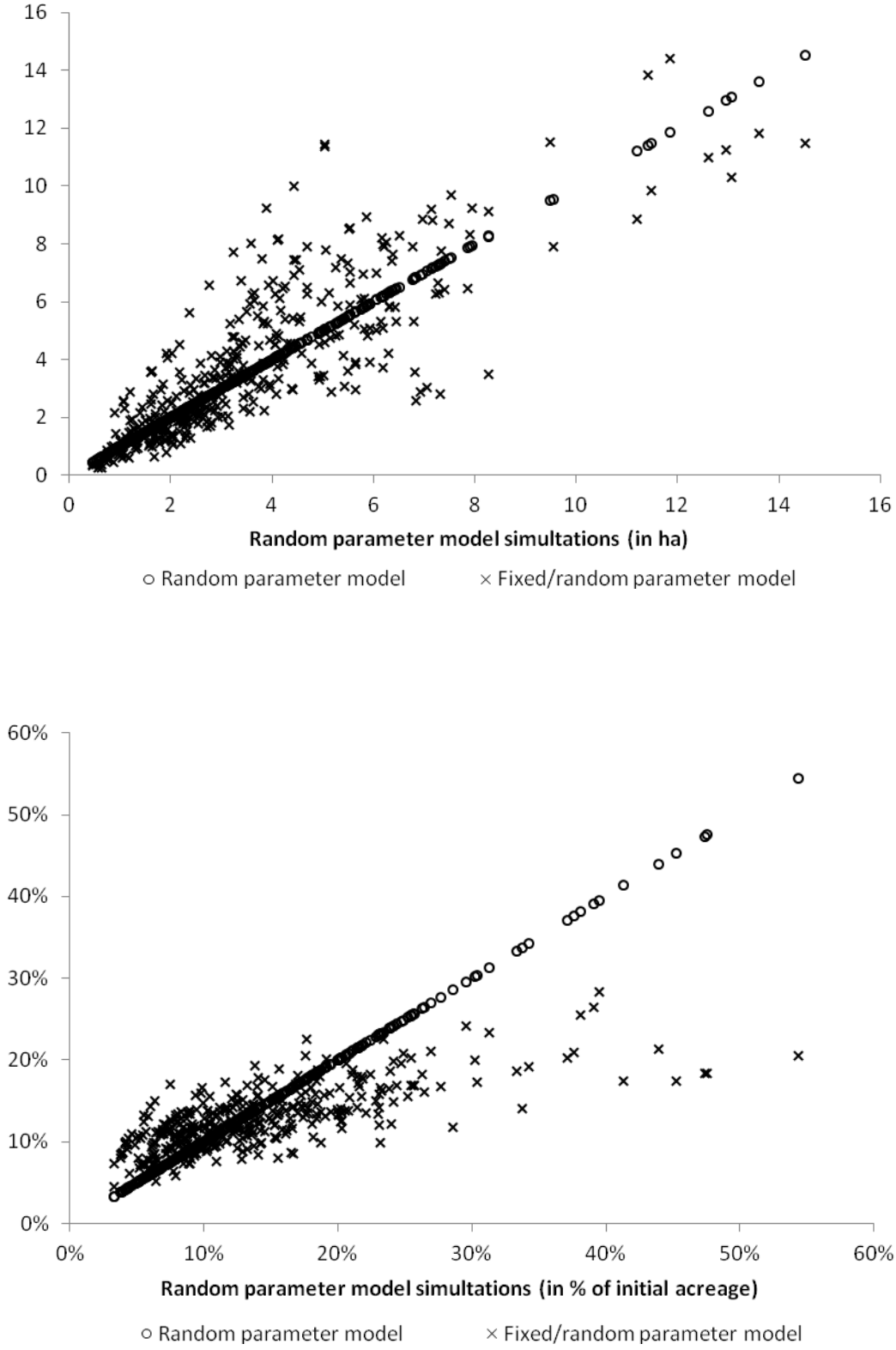
In order to further assess the potential impacts of the approach proposed in this article to account for heterogeneity, two alternative versions of the Nested MNL model have been estimated and used to simulate the effects of the same price shock. In the first model, all parameters are considered as fixed. This model, which will subsequently be referred to as the fixed parameter (FP) model, is estimated using a standard ML approach. In the second model, individual effects

are introduced by considering the constant term of the yield and acreage equations, $\boldsymbol{\beta}_i^y$ and $\boldsymbol{\beta}_i^s$, as random, the α and ρ coefficients being fixed. This model, which will subsequently be referred to as "fixed/random parameter (F/RP) model", is estimated using the SAEM algorithm. The estimation results of these two models are not presented here due to space limitation but are available upon request. Two main elements come out of these results. (*i*) In the F/RP model, the estimated values of α and ρ are respectively equal to 0.08 and 0.09 and are thus relatively close to their estimated means in the RP model. This is not the case with the FP model: the estimated value of α equals 0.02 and that of ρ equals 0.06. (*ii*) The log likelihoods of the F/RP and FP models respectively equal –953.1 and –855.9, compared to –723.4 for the RP model. The likelihood ratio test thus clearly indicates that the RP model significantly better fits the data than its constrained counterparts.

The impacts of price changes on acreages simulated with these two models are reported in the second and third columns of Table 6. The overall impacts on acreages are clearly underestimated with the FP model: the changes in wheat, other cereals and oilseeds acreages are respectively equal to –299 ha, –59 ha and +358 ha, these effects are thus 60% to 90% lower than those simulated with the RP model. This can certainly essentially be attributed to the lower estimated values of α and ρ, which determine the magnitude of acreages responses to economic incentives. Differences are less obvious between the RP and the F/RP models. However, a closer look at the results shows that the impacts simulated with the two models differ significantly. This is illustrated on Figure 4 where we have reported the individual impacts on oilseeds acreage simulated with the F/RP model with respect to those simulated with the RP model. The simulated impacts are represented in absolute term (changes in ha) on the upper part graph and in relative term (percentage change compared to initial acreages) on the lower part graph. While the absolute changes in acreage simulated with the F/RP model appear to be relatively homogeneously spread around those simulated with the RP model, the picture is different when it comes to relative changes in acreage: the highest impacts on oilseeds acreages tend to be underestimated by the F/RP model and the smallest impacts tend to be overestimated. The reason is that, in absolute terms, the impacts on farms with relatively small initial oilseed acreage are higher when simulated with the RP model, which translates in higher percentage changes. On the contrary, the F/RP model tends to simulate larger impacts, in absolute term, for farms with higher initial oilseed acreages. So, even if the impacts simulated with the two models are similar on average, their distribution across farms varies substantially. These results tend to show that partially or totally

ignoring unobserved heterogeneity of farms or/and of farmers in agricultural production choice model may induce misleading interpretations and/or policy recommendations.

Figure 4: Simulated impacts of the price shock on oilseeds acreages, in %



6. Concluding remarks

Taking for granted that many unobserved heterogeneous factors can impact farmers' production decisions, the contribution of this article is threefold.

First, we show that to consider RP versions of standard models of agricultural production choices model allows accounting for this unobserved heterogeneity in a fairly flexible way.

Second, this article shows that estimators and optimization procedures found in the applied statistics and computational statistics literatures prove to be especially well suited for estimating RP models of agricultural production choices. In particular, estimators computed with suitably designed SEM algorithms appear to be interesting alternatives to the estimators derived from the SML framework. SEM algorithms are useful when estimating mixed models composed of sub-models mostly connected because of their depending on random parameters.

Third, our illustrative application tends to show that unobserved heterogeneity really matters in micro-econometric agricultural production models. Key parameters of the model exhibit significant variability across farmers. We also show how RP models can be used to "statistically calibrate" a simulation model based on a sample of heterogeneous farms and use this "calibrated" model to simulate the impact of crop price changes on acreages. This allows us to further illustrate the potential role of heterogeneity in micro econometric production choice models. In particular, our results show that ignoring the unobserved heterogeneity effects on farmers' choices can lead to misleading simulation results and, as a result, to incorrect policy recommendations.

Of course, our empirical framework has many limitations calling for improvement and further research.

The limited size of the sample we consider is an issue. This sample was selected so as to only contain farms highly specialized in grain production, *i.e*. with homogeneous production choices, in order to investigate the importance of unobserved heterogeneity effects.

To consider crop aggregates is not satisfactory, especially when specific features of the agricultural production technology, *e.g.*, crop rotation effects, are to be considered. The use of disaggregated crop sets poses additional challenges. Even farms mostly specialized in cash grain crops have quite different crop sets. Specification and estimation of multi-crop models with corner solutions remain open research questions in the agricultural production economic literature. Our crop aggregates were mainly built to avoid the occurrence of null acreages.

Our application only considers three crops or crop aggregates. In the RP multicrop model presented here the number of parameters to be estimated quadratically increases in the crop

number, due to covariances of the random parameters. This calls for an alternative approach to – or at least for an adaptation of – the "full RP" approach adopted here.

Our considering a small time period and our focusing on short run choices lies at the root of our modeling the random parameter according to a unique, and stable across time, probability distribution. To consider longer time periods raises further issues. In particular, farmers' technology choices and quasi-fixed factor endowments evolve over time, implying that the probability distribution of the random parameters cannot be assumed to be constant along the long time period in this case.

References

Ackerberg, D., Benkard, L., Berry, S.T., Pakes, A. (2007). Econometric Tools for Analyzing Market Outcomes. In Heckman, J.J., and E.E. Leamer (eds), *Handbook of Econometrics*, Vol 6A. Amsterdam: Elsevier.

Angrist, J.D., Pischke, J.-S. (2009). *Mostly harmless econometrics. An empiricist's companion*. Princeton: Princeton University Press.

Berry, S.T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics,* 25:242-262.

Biorn, E. (2004). Regression systems for unbalanced panel data: a stepwise maximum likelihood procedure. *Journal of Econometrics,* 122(2):281-291.

Booth, A. L., Francesconi, M., Frank, J. (2001). *Temporary jobs: who gets them, what are they worth, and do they lead anywhere?* ISER Working Paper University of Essex 00-13.

Booth, J.G., Hobert, J.P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society*, Series B 61(1):265-285.

Caffo, B.S., Jank, W., Jones, G.L. (2005). Ascent-based Monte Carlo Expectation–Maximization. *Journal of the Royal Statistical Society, Series B* 67(2):235-251.

Caponi, V. (2011). Intergenerational transmission of abilities and self selection of Mexican immigrants. *International Economic Review,* 52(2):523-547.

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

Carpentier, A., Letort, E. (2012). Accounting for heterogeneity in multi-crop micro-econometric models: Implications for variable input demand modelling. *American Journal of Agricultural Economics*, 94(1):209-224.

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

Chavas, J.-P., Holt, M.T. (1990). Acreage Decisions under Risk: The Case of Corn and Soybeans. *American Journal of Agricultural Economics,* 72(2):529-538.

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

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

Eaton, J., Kortum, S., Kramarz, F. (2011). An Anatomy of International Trade: Evidence From French Firms. *Econometrica,* 79(5):1453-1498.

Greene, W. H., Hensher, D.A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological,* 37(8):681-698.

Gu, M.G., Zhu, H.T. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *Journal of the Royal Statistical Society,* Series B 63(2):339-355.

Heckeleï, T., Britz, W., Zhang, Y. (2012). Positive mathematical programming approaches. Recent developments in literature and applied modelling. *Bio-based and Applied Economics,* 1(1):109-124.

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

Heckman, J.J., Vytlacil, E.J. (2007). Econometric evaluation of social programs, part I: causal models, structural models and econometric policy evaluation." In Heckman, J.J., and E.E. Leamer (eds), Chapter 70, *Handbook of Econometrics*, Volume 6B. Amsterdam: Elsevier.

Heckman, J.J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy,* 109(4):673-748.

Howitt, E. (1995). Positive Mathematical Programming. *American Journal of Agricultural Economics,* 77(2):329-342.

Jank, W. (2006). Implementing and diagnosing the stochastic approximation EM algorithm. *Journal of Computational and Graphical Statistics,* 15(4):803-829.

Jank, W., Booth, J. (2003). Efficiency of Monte Carlo EM and Simulated Maximum Likelihood in Two-Stage Hierarchical Models. *Journal of Computational and Graphical Statistics,* 12(1):214-229.

Keane, M.P. (2009). Simulated Maximum likelihood estimation based on FOCs. *International Economic Review,* 50(2):627-675.

Levine, R.A., Casella, G. (2001). Implementations of the Monte Carlo EM Algorithm. *Journal of Computational and Graphical Statistics,* 10(3):422-439.

McFadden, D., Train, K. (2000). Mixed MNL models for discrete response. *Journal of applied Econometrics,* 15(5):447–470.

McLachlan, G., Krishnan, T. (2008). *The EM algorithm and extensions, 2nd edition*. Wiley Edition.

Meng, X.L., Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika,* 80(2):267-278.

Moro, D., Sckokai, P. (2013). The impact of decoupled payments on farm choices: Conceptual and methodological challenges. *Food Policy,* 41(1):28-38.

Oude Lansink, A.G. (1999). Generalized maximum entropy estimation and heterogeneous technologies. *European Review of Agricultural Economics,* 26(1):101-115.

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

Polyak, B.T., Juditski A.B. (1992). Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4):838-855.

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

Train, K. (2009). *Discrete Choice Methods with Simulation, 2nd ed*. Cambridge: University Press.

Train, K. (2008). EM algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling,* 1(1):40-69.

Wei, G.C., Tanner M.A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association,* 85(411):699-704.

Technical Appendices

Appendix A. Multicrop model details

A.1. Yield supply and expected gross margin functions

The yield supply function of crop *k* given in equation (6a) is obtained by maximizing in the aggregate variable input level, $x_{k,it}$, the expected margin of crop *k* under the assumptions that the yield function is quadratic in the aggregate variable input level:

$$y_{k,it} = \beta^{y}_{k,i} + \mathbf{d}'_{k,it}\boldsymbol{\delta}^{y}_{k} + e^{y}_{k,it} - 1/2 \times \gamma_k^{-1}(\beta^{x}_{k,i} + \mathbf{d}'_{k,it}\boldsymbol{\delta}^{x}_{k} + e^{x}_{k,it} - x_{k,it})^2 \qquad \text{(A.1)}$$

and that the random terms $e^{x}_{k,it}$ and the climatic events $\mathbf{d}_{k,it}$ are observed when the input level is decided.[21] This yield function is parameterized by fixed parameters, the curvature parameter γ_k and the climatic event coefficients $\boldsymbol{\delta}_k \equiv (\boldsymbol{\delta}^{y}_{k}, \boldsymbol{\delta}^{x}_{k})$, and a random parameter $\beta^{y}_{k,i}$. It depends on the effects of random events represented by the centered error terms $\mathbf{e}_{k,it} \equiv (e^{y}_{k,it}, e^{x}_{k,it})$. The term $\beta^{y}_{k,i} + \mathbf{d}'_{k,it}\boldsymbol{\delta}^{y}_{k} + e^{y}_{k,it}$ can be interpreted as the maximum yield level achievable by farmer *i* in year *t* while the term $\beta^{x}_{k,i} + \mathbf{d}'_{k,it}\boldsymbol{\delta}^{x}_{k} + e^{x}_{k,it}$ can be interpreted as the input use level required to achieve this maximum yield level.

Assuming that the aggregated variable input price w_{it} is known at the beginning of the production process and that the price expectation for the output of crop *k* of farmer *i* in year *t* is given by $p_{k,it}$, the optimal input level of farmer *i* in *t* on crop *k* is obtained as:

$$x_{k,it} = \beta^{x}_{k,i} + \mathbf{d}'_{k,it}\boldsymbol{\delta}^{x}_{k} - \gamma_k w_{it} p^{-1}_{k,it} + e^{x}_{k,it} \text{ for } k \in \mathcal{K} \qquad \text{(A.2)}$$

the corresponding yield supply function as:

$$y_{k,it} = \beta^{y}_{k,i} + \mathbf{d}'_{k,it}\boldsymbol{\delta}^{y}_{k} - 1/2 \times \gamma_k (w_{it} p^{-1}_{k,it})^2 + e^{y}_{k,it} \text{ for } k \in \mathcal{K} \qquad \text{(A.3)}$$

and the corresponding gross margin as:

$$\tilde{\pi}_{k,it} = p_{k,it}(\beta^{y}_{k,i} + \mathbf{d}'_{k,it}\boldsymbol{\delta}^{y}_{k} + e^{y}_{k,it}) + 1/2\gamma_{k,0} w^2_{it} p^{-1}_{k,it} - w_{it}(\beta^{x}_{k,i} + \mathbf{d}'_{k,it}\boldsymbol{\delta}^{x}_{k} + e^{x}_{k,it}) \text{ for } k \in \mathcal{K} \qquad \text{(A.4)}$$

[21] Whether the random event effects $e^{y}_{k,it}$ are observed or not doesn't matter. These effects are forgone by the considered (risk neutral) farmer.

It is assumed that the crop gross margin levels considered by farmer *i* when choosing his acreages are the expectations of those given in equation (A.4) conditionally on the information set available at this time. Assuming that farmer *i* knows the parameters of his crop *k* production technologies $\boldsymbol{\beta}_{k,i} \equiv (\beta^y_{k,i}, \beta^x_{k,i})$, $\boldsymbol{\delta}_k \equiv (\boldsymbol{\delta}^y_k, \boldsymbol{\delta}^x_k)$ and γ_k, and given his price expectations in year *t*, his expectation of $\tilde{\pi}_{k,it}$ depends on his expectations of the random terms $\mathbf{e}_{k,it}$ and $\mathbf{d}_{k,it}$ affecting crop *k* production process. It is assumed that the terms $\mathbf{e}_{k,it}$ and $\mathbf{d}_{k,it}$ are centered from the viewpoint of farmer *i* when he chooses his acreage. This assumption yields the following expected crop *k* gross margin for farmer *i* in year *t*:

$$\bar{\pi}_{k,it} = p_{k,it}\beta^y_{k,i} - w_{it}\beta^x_{k,i} + 1/2\gamma_{k,0}w^2_{it}p^{-1}_{k,it} \text{ for } k \in \mathcal{K} \tag{A.5}$$

It also implies that the random parameters $\boldsymbol{\beta}_{k,i}$ capture the effects of farm/farmer specific production conditions and/or technologies and requires the climatic variables to be centered for each farm (this is achieved in our case at the municipality level). According to this assumption, $\beta^y_{k,i}$ is farmer *i* expectation of the maximum achievable yield of crop *k* and farmer *i*, and the term $\mathbf{d}'_{k,it}\boldsymbol{\delta}^y_k + e^y_{k,it}$ captures the effects of random events on this maximum achievable yield. The terms $\beta^x_{k,i}$ and $\mathbf{d}'_{k,it}\boldsymbol{\delta}^x_k + e^x_{k,it}$ can be interpreted accordingly on "the input side".

A.2. Two-levels Nested MNL acreage share model

The crop acreage share model given in equations (6b)–(6c) is derived from the following two-levels Nested MNL acreage share choice model:

$$s_{k,it} = \frac{\exp\left(\rho_{g,i}(\bar{\pi}_{k,it} - \beta^c_{k,i} - e^c_{k,it})\right)}{\sum_{\ell \in \mathcal{K}_g}\exp\left(\rho_{g,i}(\bar{\pi}_{\ell,it} - \beta^c_{\ell,i} - e^c_{\ell,it})\right)} \times \frac{\exp\left(\alpha_i\rho^{-1}_{g,i}\ln\sum_{\ell \in \mathcal{K}_g}\exp\left(\rho_{g,i}(\bar{\pi}_{\ell,it} - \beta^c_{\ell,i} - e^c_{\ell,it})\right)\right)}{\sum_{h \in \mathcal{G}}\exp\left(\alpha_i\rho^{-1}_{h,i}\ln\sum_{\ell \in \mathcal{K}_h}\exp\left(\rho_{h,i}(\bar{\pi}_{\ell,it} - \beta^c_{\ell,i} - e^c_{\ell,it})\right)\right)} \tag{A.6}$$

for $k \in \mathcal{K}_g$ and $g \in \mathcal{G}$. This model can be used to define crop group acreage share models and crop acreage share models within a given crop group, *i.e.*

$$s_{k,it} = s_{k|g,it}\bar{s}_{g,it} \text{ for } k \in \mathcal{K}_g \text{ and } g = 1,...,G \tag{A.7}$$

Where

$$\overline{s}_{g,it} \equiv \sum\nolimits_{\ell \in \mathcal{K}_g} s_{\ell,it} \tag{A.8}$$

and

$$s_{k|g,it} \equiv s_{k,it}(\overline{s}_{g,it})^{-1} \text{ if } k \in \mathcal{K}_g \tag{A.9}$$

Using equation (A.7) it is easily seen that

$$s_{k|g,it} \equiv \frac{\exp\left(\rho_{g,i}(\overline{\pi}_{k,it} - \beta^c_{k,i} - e^c_{k,it})\right)}{\sum_{\ell \in \mathcal{K}_g} \exp\left(\rho_{g,i}(\overline{\pi}_{\ell,it} - \beta^c_{\ell,i} - e^c_{\ell,it})\right)} \tag{A.10}$$

defines the optimal acreage share of crop *k* in nest *g* while

$$\overline{s}_{g,it} \equiv \frac{\exp\left(\alpha_i \rho^{-1}_{g,i} \ln \sum_{\ell \in \mathcal{K}_g} \exp\left(\rho_{g,i}(\overline{\pi}_{\ell,it} - \beta^c_{\ell,i} - e^c_{\ell,it})\right)\right)}{\sum_{h \in \mathcal{G}} \exp\left(\alpha_i \rho^{-1}_{h,i} \ln \sum_{\ell \in \mathcal{K}_h} \exp\left(\rho_{h,i}(\overline{\pi}_{\ell,it} - \beta^c_{\ell,i} - e^c_{\ell,it})\right)\right)} \tag{A.11}$$

defines the optimal acreage share of nest *g* in total cropland.

Carpentier and Letort (2014) show that $\mathbf{s}_{it} \equiv (s_{k,it} : k \in \mathcal{K})$ can be defined as the solution in $\mathbf{s} \equiv (s_k : k \in \mathcal{K})$ to the following maximization problem

$$\max_{\mathbf{s}} \left\{ \mathbf{s}'\boldsymbol{\pi}_{it} - C(\mathbf{s}; \vartheta_i, \mathbf{e}^c_{it}) \text{ s.t. } \mathbf{s} \geq \mathbf{0} \text{ and } \mathbf{s}'\boldsymbol{\iota} = 1 \right\} \tag{A.12}$$

where

$$\begin{aligned} C(\mathbf{s}; \vartheta_i, \mathbf{e}^c_{it}) = & \sum\nolimits_{g \in \mathcal{G}} \sum\nolimits_{k \in \mathcal{K}_g} s_k(\beta^c_{k,i} + e^c_{k,it}) \\ & + \alpha_i^{-1} \sum\nolimits_{g \in \mathcal{G}} (1 - \alpha_i \rho^{-1}_{g,i}) \overline{s}_g \ln \overline{s}_g + \sum\nolimits_{g \in \mathcal{G}} \rho^{-1}_{g,i} \sum\nolimits_{k \in \mathcal{K}_g} s_k \ln s_k \end{aligned} \tag{A.13}$$

The term $\boldsymbol{\iota}$ is a conformable unitary vector, the vector $\vartheta_i \equiv (\alpha_i, \boldsymbol{\rho}_i, \boldsymbol{\beta}^c_i)$ collects the elements of the farmer specific random parameters with $\boldsymbol{\beta}^c_i \equiv (\beta^c_{k,i} : k \in \mathcal{K})$, and $\mathbf{e}^c_{it} \equiv (e^c_{k,it} : k \in \mathcal{K})$ collects the acreage share model error terms. The term $C(\mathbf{s}; \vartheta_i, \mathbf{e}^c_{it})$ is strictly convex in **s** if and only if $\rho_{g,i} \geq \alpha_i > 0$ for $g \in \mathcal{G}$, in which case it can be interpreted as a diversification motive from the optimal crop acreage.

It is then easy to derive the effects of a change in the expected gross margin of crop ℓ on the acreage share of crop *k*. Provided that $k \in \mathcal{K}_g$ and $\ell \in \mathcal{K}_h$, we simply have:

$$\frac{\partial}{\partial \pi_{\ell,i}} \ln s_{k,i} = \begin{cases} \rho_{g,i} - (\rho_{g,i} - \alpha_i) s_{k|g,it} - \alpha_i s_{k,it} & \text{if } \ell = k \\ -(\rho_{g,i} - \alpha_i) s_{\ell|g,it} - \alpha_i s_{\ell,it} & \text{if } h = g \text{ and } \ell \neq k \\ -\alpha_i s_{\ell,it} & \text{if } h \neq g \end{cases} \tag{A.14}$$

with:

$$\frac{\partial}{\partial \pi_{\ell,i}} \ln s_{k|g,it} = \begin{cases} \rho_{g,i} - \rho_{g,i} s_{k|g,it} & \text{if } \ell = k \\ -\rho_{g,i} s_{\ell|g,it} & \text{if } h = g \text{ and } \ell \neq k \\ 0 & \text{if } h \neq g \end{cases} \tag{A.15}$$

and:

$$\frac{\partial}{\partial \pi_{\ell,i}} \ln \bar{s}_{g,it} = \begin{cases} \alpha_i s_{\ell|g,it} - \alpha_i s_{\ell,it} & \text{if } h = g \\ -\alpha_i s_{\ell,it} & \text{if } h \neq g \end{cases} \tag{A.16}$$

Equations (A.14)–(A.16) illustrate the crucial role of the curvature parameters, α_i and $\rho_{g,i}$ for $g \in \mathcal{G}$ of the implicit management cost function in the extent to which optimal acreage choices respond to changes in economic incentives. It is also easily checked that $\frac{\partial}{\partial \pi_{k,i}} \ln s_{k,i} > 0$ while $\frac{\partial}{\partial \pi_{\ell,i}} \ln s_{k,i} < 0$ if $\ell \neq k$ when $\rho_{g,i} \geq \alpha_i > 0$ for $g \in \mathcal{G}$.

A.3. Estimated multi-crop model

Input demand equation (A.2) is not included in the estimated multi-crop models and the expected gross margin used in the acreage share model (6b) is not that given by equation (A.5), for two reasons. Input use levels are not observed at the crop level in our data set. They are only recorded at the farm level, unfortunately. This aggregation problem can be overcome by defining an input use allocation equation as in Carpentier and Letort (2012). However, this option would have increased significantly the complexity of the considered multi-crop model and of its estimation. Second, due to insufficient variation of the aggregate input prices in our data set as well as due to our not modeling input demands, it is difficult to separately identify the probability distributions of the parameters $\beta^x_{k,i}$ and $\beta^c_{k,i}$ empirically.

This explains why the expected gross margin used in the acreage share model (6b) is not that given by equation (A.5). The term $\beta^s_{k,i}$ in equation (6c) is given by:

$$\beta^{s}_{k,i} \simeq \beta^{c}_{k,i} + w_{it}\beta^{x}_{k,i} \tag{A.17}$$

while

$$e^{s}_{k,it} = e^{c}_{k,it} \tag{A.18}$$

for $k \in \mathcal{K}$. This implies that the crop profitability measures provided in equation (6c) are obtained as

$$\begin{aligned} \pi_{k,it} &\equiv p_{k,it}\beta^{y}_{k,i} + 1/2\gamma_{k,0}w^{2}_{it}p^{-1}_{k,it} - (\beta^{s}_{k,i} + e^{s}_{k,it}) \\ &\simeq p_{k,it}\beta^{y}_{k,i} - w_{it}\beta^{x}_{k,i} + 1/2\gamma_{k,0}w^{2}_{it}p^{-1}_{k,it} - (\beta^{c}_{k,i} + e^{c}_{k,it}) \quad \text{for } k \in \mathcal{K} \\ &= \bar{\pi}_{k,it} - (\beta^{c}_{k,i} + e^{c}_{k,it}) \end{aligned} \tag{A.19}$$

Appendix B. Likelihood function

The inverse function of **r** is required to determine the likelihood function of the considered model. The elements of $\mathbf{e}_{it}^{y}$ can easily be recovered from equation (6a) while the elements of $\mathbf{e}_{it}^{s}$ can be obtained by application of Berry's (1994) device from equation (6b). With $k \in \mathcal{K}_g$ we have:

$$\begin{aligned} e_{k,it}^{s} = \alpha_i^{-1} \times \left(\ln s_{k,it} - \ln s_{K,it} - (1 - \alpha_i \rho_{g,i}^{-1}) \times (\ln s_{k,it} - \ln \bar{s}_{g,it}) \right) + \beta_{k,i}^{s} \\ - p_{k,it} \left(\beta_{k,i}^{y} + 1/2 \times \gamma_k (w_{it} p_{k,it}^{-1})^2 \right) + p_{K,it} \left(\beta_{K,i}^{y} + 1/2 \times \gamma_K (w_{it} p_{K,it}^{-1})^2 \right) \end{aligned} \tag{B.1}$$

with $e_{K,it}^{s} \equiv 0$. It is assumed here that group *G* contains a single crop, crop *K*. This simplifies equation (B.1) but this is not necessary for inverting in $\mathbf{e}_{it}^{s}$ the part of the response function **r** related to $\mathbf{s}_{it}$, *i.e.* for obtaining the function $\mathbf{e}_i^{s}(\mathbf{s}_{it}, \mathbf{z}_{it}; \boldsymbol{\gamma}, \boldsymbol{\delta})$ satisfying $\mathbf{e}_i^{s}(\mathbf{s}_{it}, \mathbf{z}_{it}; \boldsymbol{\gamma}, \boldsymbol{\delta}) = \mathbf{e}_{it}^{s}$.

The density of $\mathbf{c}_i$ conditional on $(\mathbf{z}_i, \mathbf{q}_i)$ can then be obtained by applying equations (2) and (10) and by using the density of normal random vectors. Let $\varphi(\mathbf{u}; \mathbf{B})$ denote the probability density function of $\mathcal{N}(\mathbf{0}, \mathbf{B})$ at **u**. The density of $\mathbf{y}_{it}$ conditional on $(\mathbf{z}_i, \mathbf{q}_i)$ is given by:

$$f(\mathbf{y}_{it} \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\Psi}^{y}) = \varphi\left(\mathbf{e}_i^{y}(\mathbf{s}_{it}, \mathbf{z}_{it}; \boldsymbol{\gamma}, \boldsymbol{\delta}); \boldsymbol{\Psi}^{y} \right) \tag{B.2}$$

and that of $\mathbf{s}_{it}$ conditional on $(\mathbf{z}_i, \mathbf{q}_i)$ is given by:

$$f(\mathbf{s}_{it} \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\gamma}, \boldsymbol{\Psi}^{s}) = \alpha_i^{1-G} \left(\prod\nolimits_{g \in \mathcal{G}} \rho_{g,i}^{1-K_g} \right) \left(\prod\nolimits_{k \in \mathcal{K}} s_{k,it}^{-1} \right) \times \varphi\left(\mathbf{e}_i^{s}(\mathbf{s}_{it}, \mathbf{z}_{it}; \boldsymbol{\gamma}); \boldsymbol{\Psi}^{y} \right) \tag{B.3}$$

The derivation of $f(\mathbf{s}_{it} \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\gamma}, \boldsymbol{\Psi}^{s})$ relies on the usual transformation formula of the probability density function:

$$f(\mathbf{s}_{it} \mid \mathbf{z}_i, \mathbf{q}_i; \boldsymbol{\gamma}, \boldsymbol{\Psi}^{s}) = \left| \det \tfrac{\partial}{\partial \mathbf{s}_{it}'} \mathbf{e}_i^{s}(\mathbf{s}_{it}, \mathbf{z}_{it}; \boldsymbol{\gamma}) \right| \times \varphi\left(\mathbf{e}_i^{s}(\mathbf{s}_{it}, \mathbf{z}_{it}; \boldsymbol{\gamma}); \boldsymbol{\Psi}^{y} \right) \tag{B.4}$$

provided that $s_{K,it} = 1 - \sum_{k=1}^{K} s_{k,it}$ (and $e_{K,it}^{s} \equiv 0$). The proof of

$$\det \tfrac{\partial}{\partial \mathbf{s}_{it}'} \mathbf{e}_i^{s}(\mathbf{s}_{it}, \mathbf{z}_{it}; \boldsymbol{\gamma}) = \alpha_i^{1-G} \left(\prod\nolimits_{g \in \mathcal{G}} \rho_{g,i}^{1-K_g} \right) \left(\prod\nolimits_{k \in \mathcal{K}} s_{k,it}^{-1} \right) \tag{B.5}$$

makes use of the identities

$$\det(\mathbf{A} + \gamma \times \mathbf{u}') = (1 + \gamma \times \boldsymbol{\iota}' \mathbf{A}^{-1} \boldsymbol{\iota}) \times \det \mathbf{A} \tag{B.6}$$

and

$$(\mathbf{A}+\gamma\times\mathbf{u}')^{-1}=\mathbf{A}^{-1}-\gamma\times\mathbf{A}^{-1}\mathbf{u}'\mathbf{A}^{-1}\times(1+\gamma\times\boldsymbol{\iota}'\mathbf{A}^{-1}\boldsymbol{\iota})^{-1} \tag{B.7}$$

where $\mathbf{A}$ is a non-singular symmetric matrix and $\boldsymbol{\iota}$ is conformable unitary vector.

Finally, the random parameter vector density is simply given by:

$$h(\mathbf{q}_i;\boldsymbol{\eta})=\varphi(\mathbf{q}_i-\boldsymbol{\tau};\boldsymbol{\Omega}) \tag{B.8}$$

Appendix C. Algorithms

This appendix presents the algorithm we used in three steps, after a brief presentation of the self-normalized IS simulator used here, as well as of its interests and limits. First, we restate the basic SEM algorithm using the IS simulator. Second, we show how the ideas underlying the ECM algorithms proposed by Meng and Rubin (1993) on the one hand and the Ascent-based Monte Carlo EM algorithms proposed by Caffo *et al.* (2005) can be used for designing an "Ascent-based SECM" algorithm simplifying the computations involved in the SEM algorithm. Finally we present the SAEM and its interests and show how the "Ascent-based SECM" algorithm can easily be adapted into an "Ascent-based SAECM" algorithm.

C.1. IS simulator

The IS simulator used here was employed by, *e.g.* Train (2008; 2009), for implementing SEM algorithms. It allows integrating by simulation method an expectation over the probability distribution of $\mathbf{q}_i$ conditional on $(\mathbf{c}_i, \mathbf{z}_i)$ by using draws for the marginal probability distribution of $\mathbf{q}_i$. It allows approximating

$$E[\ln \kappa(\mathbf{c}_i, \mathbf{q}_i \mid \mathbf{z}_i; \boldsymbol{\theta}) \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}_{n-1}] \equiv \int \ln \kappa(\mathbf{c}_i, \mathbf{q} \mid \mathbf{z}_i; \boldsymbol{\theta}) h(\mathbf{q} \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}_{n-1}) d\mathbf{q} \tag{C.1}$$

by

$$S_n^{-1} \sum_{s=1}^{S_n} \tilde{\omega}_{i,s,n-1} \sum_{t=1}^{T} \ln \kappa(\mathbf{c}_{it}, \tilde{\mathbf{q}}_{i,s,n-1} \mid \mathbf{z}_{it}; \boldsymbol{\theta}) \tag{C.2}$$

where the

$$\tilde{\mathbf{q}}_{i,s,n-1} \equiv \tilde{\mathbf{q}}_{i,s}(\boldsymbol{\eta}_{n-1}) \tag{C.3}$$

terms are independent random draws from $h(\mathbf{q}; \boldsymbol{\eta}_{n-1})$ for $s = 1, ..., S_n$ and $i = 1, ..., N$, and

(C.4)

$$\tilde{\omega}_{i,s,n-1} \equiv \tilde{\omega}_{i,s}(\boldsymbol{\theta}_{n-1}) \equiv \frac{f(\mathbf{c}_i \mid \mathbf{z}_i, \tilde{\mathbf{q}}_{i,s,n-1}; \boldsymbol{\mu}_{n-1})}{S_n^{-1} \sum_{s=1}^{S_n} f(\mathbf{c}_i \mid \mathbf{z}_i, \tilde{\mathbf{q}}_{i,s,n-1}; \boldsymbol{\mu}_{n-1})} \tag{C.4}$$

for $s = 1, ..., S_n$ and $i = 1, ..., N$.

This (self-normalized) IS simulator is an importance sampling simulator with $h(\mathbf{q}_i; \boldsymbol{\eta})$ as the proposal probability density function. This proposal probability density function is inefficient but

this simulator is fairly easy to code – (quasi-)random draws from $h(\mathbf{q}_i;\boldsymbol{\eta})$ are easily obtained and the $f(\mathbf{c}_i \mid \mathbf{z}_i, \tilde{\mathbf{q}}_{i,s,n-1}; \boldsymbol{\mu}_{n-1})$ terms need to be evaluated for computing the approximated objective function – and allows, to some extent, using large random draw numbers. Expectations such as $E[\ln \kappa(\mathbf{c}_i, \mathbf{q}_i \mid \mathbf{z}_i; \boldsymbol{\theta}) \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}_{n-1}]$ can be integrated by using draws from $h(\mathbf{q}_i \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta})$. But such draws are more difficult to obtain. *E.g.*, it is always possible to obtain Metropolis-Hastings (quasi-)random draws from $h(\mathbf{q}_i \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta})$. This simulation technique consists in a rather long process to be repeated at each iteration of the SEM algorithm. It can be time consuming (Levine and Casella, 2001).

The use of normal density functions as proposal density is usually unwarranted due to their light tails. But this proposal density has two main advantages. First, its use is simple. Second, it allows interpreting the obtained estimator of $\boldsymbol{\theta}$ as a Method of Simulated Score estimator which is closely linked to the usual SML estimator (Train, 2008; 2009).

The algorithms to be defined rely on the simulated versions of the conditional expectation of the sample log-likelihood function $N^{-1}\sum_{i=1}^{N} E[\ln \kappa(\mathbf{c}_i, \mathbf{q}_i \mid \mathbf{z}_i; \boldsymbol{\theta}) \mid \mathbf{z}_i, \mathbf{c}_i; \boldsymbol{\theta}_{n-1}]$ of the complete data vector:

$$\tilde{Q}_{N,S_n}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{n-1}) \equiv N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n} \tilde{\omega}_{i,s,n-1}\sum_{t=1}^{T} \ln \kappa(\mathbf{c}_{it}, \tilde{\mathbf{q}}_{i,s,n-1} \mid \mathbf{z}_{it}; \boldsymbol{\theta}) \tag{C.5}$$

and, using $\ln \kappa(\mathbf{c}_{it}, \mathbf{q} \mid \mathbf{z}_{it}; \boldsymbol{\theta}) = \ln f(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \mathbf{q}; \boldsymbol{\mu}) + \ln h(\mathbf{q}; \boldsymbol{\eta})$, its decomposition given by

$$\tilde{Q}_{N,n}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{n-1}) = \tilde{Q}^{\mathbf{c}}_{N,n}(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{n-1}) + \tilde{Q}^{\mathbf{q}}_{N,n}(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1}) \tag{C.6}$$

with

$$\tilde{Q}^{c}_{N,S_n}(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{n-1}) \equiv N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n} \tilde{\omega}_{i,s,n-1}\sum_{t=1}^{T} \ln f(\mathbf{c}_{it} \mid \mathbf{z}_{it}, \tilde{\mathbf{q}}_{i,s,n-1}; \boldsymbol{\mu}) \tag{C.7}$$

and:

$$\tilde{Q}^{s}_{N,S_n}(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1}) \equiv N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n} \tilde{\omega}_{i,s,n-1} \ln h(\tilde{\mathbf{q}}_{i,s,n-1}; \boldsymbol{\eta}) \tag{C.8}$$

C.2. SEM algorithm

The APIS Simulator allows designing a simple SEM algorithm. The "recursive estimator" proposed by Train (2008; 2009), as well as its extensions, are computed by using such SEM algorithms. The SEM algorithm iterates a sequence composed of a SE step and of a M step.

SE step. Integration of the conditional expectations

Obtain independent (pseudo-)random draws $\tilde{\mathbf{q}}_{i,s,n-1}$ from $h(\mathbf{q};\boldsymbol{\eta}_{n-1})$ and compute the weight terms $\tilde{\omega}_{i,s,n-1}$ for $s=1,...,S_n$ and $i=1,...,N$.

M step. Update of the value of $\boldsymbol{\theta} \equiv (\boldsymbol{\mu},\boldsymbol{\eta})$

Compute:

$$\boldsymbol{\theta}_n \equiv \arg\max_{\boldsymbol{\theta}} \tilde{Q}_{N,n}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{n-1}) \tag{C.9}$$

or, equivalently:

$$\boldsymbol{\mu}_n \equiv \arg\max_{\boldsymbol{\mu}} \tilde{Q}^c_{N,S_n}(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{n-1}) \tag{C.10a}$$

and

$$\boldsymbol{\eta}_n \equiv \arg\max_{\boldsymbol{\eta}} \tilde{Q}^c_{N,S_n}(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1}) \tag{C.10b}$$

The decomposition of the maximization problem of $\tilde{Q}_{N,S_n}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{n-1})$ in $\boldsymbol{\theta}$ into the maximization problems of $\tilde{Q}^s_{N,S_n}(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{n-1})$ in $\boldsymbol{\mu}$ and of $\tilde{Q}^q_{N,S_n}(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1})$ in $\boldsymbol{\eta}$ illustrates the main interest in using EM algorithms for estimating random parameter models. This decomposition is specific to models involving latent/hidden variables such as random parameter models.

C.3. "Ascent-based SCME" algorithm

With $\ln h(\mathbf{q};\boldsymbol{\eta}) = \ln h(\mathbf{q};\boldsymbol{\tau},\boldsymbol{\Omega})$ we observe that $\tilde{Q}^q_{N,S_n}(\boldsymbol{\eta} \mid \boldsymbol{\theta}_{n-1}) = \tilde{Q}^q_{N,S_n}(\boldsymbol{\tau},\boldsymbol{\Omega} \mid \boldsymbol{\theta}_{n-1})$ is easily maximized in $(\boldsymbol{\tau},\boldsymbol{\Omega})$. It is the weighted log-likelihood function of a multivariate normal variable. Further decomposing $\ln f(\mathbf{c}_{it} \mid \mathbf{z}_{it},\mathbf{q};\boldsymbol{\mu})$, and thus $\tilde{Q}^c_{N,S_n}(\boldsymbol{\mu} \mid \boldsymbol{\theta}_{n-1})$, suggests further simplification for the M step. With

$$\ln f(\mathbf{c}_{it} \mid \mathbf{z}_{it},\mathbf{q};\boldsymbol{\mu}) = \ln f(\mathbf{y}_{it} \mid \mathbf{z}_{it},\mathbf{q};\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\Psi}^y) + \ln f(\mathbf{s}_{it} \mid \mathbf{z}_{it},\mathbf{q};\boldsymbol{\gamma},\boldsymbol{\Psi}^s) \tag{C.11}$$

the terms

$$\tilde{Q}^y_{N,S_n}(\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\Psi}^y \mid \boldsymbol{\theta}_{n-1}) \equiv N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n} \tilde{\omega}_{i,s,n-1}\sum_{t=1}^{T} \ln f(\mathbf{y}_{it} \mid \mathbf{z}_{it},\tilde{\mathbf{q}}_{i,s,n-1};\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\Psi}^y) \tag{C.12}$$

corresponding to the yield supply function sub-system, and

$$\tilde{Q}^{s}_{N,S_n}(\boldsymbol{\gamma},\boldsymbol{\Psi}^{s}\,|\,\boldsymbol{\theta}_{n-1}) \equiv N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n}\tilde{\omega}_{i,s,n-1}\sum_{t=1}^{T}\ln f(\mathbf{s}_{it}\,|\,\mathbf{z}_{it},\tilde{\mathbf{q}}_{i,s,n-1};\boldsymbol{\gamma},\boldsymbol{\Psi}^{s}) \qquad \text{(C.13)}$$

corresponding to the acreage choice sub-system, are weighted log-likelihood function of multivariate Gaussian linear (in $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$) regression models. Since these terms both depend on $\boldsymbol{\gamma}$, the maximization of

$$\tilde{Q}^{c}_{N,S_n}(\boldsymbol{\mu}\,|\,\boldsymbol{\theta}_{n-1}) = \tilde{Q}^{y}_{N,S_n}(\boldsymbol{\delta},\boldsymbol{\gamma},\boldsymbol{\Psi}^{y}\,|\,\boldsymbol{\theta}_{n-1}) + \tilde{Q}^{s}_{N,S_n}(\boldsymbol{\gamma},\boldsymbol{\Psi}^{s}\,|\,\boldsymbol{\theta}_{n-1}) \qquad \text{(C.14)}$$

in $\boldsymbol{\mu} \equiv (\boldsymbol{\delta},\boldsymbol{\gamma},\boldsymbol{\Psi}^{y},\boldsymbol{\Psi}^{s})$ cannot be split into two simpler maximization problems corresponding to the yield supply function and acreage choices sub-system. Even though, to maximize $\tilde{Q}^{y}_{N,S_n}(\boldsymbol{\delta},\boldsymbol{\gamma},\boldsymbol{\Psi}^{y}\,|\,\boldsymbol{\theta}_{n-1})$ in $(\boldsymbol{\delta},\boldsymbol{\gamma},\boldsymbol{\Psi}^{y})$ or to maximize $\tilde{Q}^{s}_{N,S_n}(\boldsymbol{\gamma},\boldsymbol{\Psi}^{s}\,|\,\boldsymbol{\theta}_{n-1})$ in $(\boldsymbol{\gamma},\boldsymbol{\Psi}^{s})$ requires the use of nonlinear optimization algorithms. Of course a simple Gauss-Seidel algorithm, the so-called Iterative Feasible Generalized Least Squares, can be used here.

The ECM algorithms proposed by Meng and Rubin (1993) allow simplifying M steps. It is possible to replace a M step, *e.g.* the direct maximization $\tilde{Q}^{c}_{N,S_n}(\boldsymbol{\mu}\,|\,\boldsymbol{\theta}_{n-1})$ in $\boldsymbol{\mu}$, by a sequence of simpler maximization problems, *i.e.* a sequence of Conditional M(CM) steps. This sequence of CM steps sequentially updates the value of $\boldsymbol{\mu}$ according to a predetermined partition of this parameter vector. The objective of these CM steps is to update the value of $\boldsymbol{\mu}$, not by maximizing $\tilde{Q}^{c}_{N,S_n}(\boldsymbol{\mu}\,|\,\boldsymbol{\theta}_{n-1})$ in $\boldsymbol{\mu}$, but by simply computing a value of $\boldsymbol{\mu}$ such that $\tilde{Q}^{c}_{N,S_n}(\boldsymbol{\mu}\,|\,\boldsymbol{\theta}_{n-1}) > \tilde{Q}^{c}_{N,S_n}(\boldsymbol{\mu}_{n-1}\,|\,\boldsymbol{\theta}_{n-1})$, if possible.[22] We used the following SECM algorithm because it only involves very simple CM steps:

SE step. Integration of the conditional expectations

Obtain independent random draws $\tilde{\mathbf{q}}_{i,s,n-1}$ from $h(\mathbf{q};\boldsymbol{\eta}_{n-1})$ and compute the weight terms $\tilde{\omega}_{i,s,n-1}$ for $s = 1,...,S_n$ and $i = 1,...,N$.

CM step. Conditional update of the value of $\boldsymbol{\theta} \equiv (\boldsymbol{\mu},\boldsymbol{\eta})$

[22] In their seminal article, Dempster, Laird and Rubin (1997) also considered this extension of the standard M step to define an extension of the standard EM algorithm which they designated as the Generalized EM (GEM) algorithm.

Compute:

$$\boldsymbol{\tau}_n \equiv \arg\max_{\boldsymbol{\tau}} \tilde{Q}^{q}_{N,S_n}(\boldsymbol{\tau}, \boldsymbol{\Omega}_{n-1} \mid \boldsymbol{\theta}_{n-1}) \tag{C.15a}$$

$$\boldsymbol{\Omega}_n \equiv \arg\max_{\boldsymbol{\Omega}} \tilde{Q}^{q}_{N,S_n}(\boldsymbol{\tau}_n, \boldsymbol{\Omega} \mid \boldsymbol{\theta}_{n-1}) \tag{C.15b}$$

$$\begin{aligned} \boldsymbol{\Psi}^{s}_n &\equiv \arg\max_{\boldsymbol{\Psi}^s} \tilde{Q}^{s}_{N,S_n}(\boldsymbol{\gamma}_{n-1}, \boldsymbol{\Psi}^s \mid \boldsymbol{\theta}_{n-1}) \\ &= \arg\max_{\boldsymbol{\Psi}^s} \tilde{Q}^{c}_{N,S_n}(\boldsymbol{\gamma}_{n-1}, \boldsymbol{\delta}_{n-1}, \boldsymbol{\Psi}^{y}_{n-1}, \boldsymbol{\Psi}^s \mid \boldsymbol{\theta}_{n-1}) \end{aligned} \tag{C.15c}$$

$$\begin{aligned} \boldsymbol{\Psi}^{y}_n &\equiv \arg\max_{\boldsymbol{\Psi}^y} \tilde{Q}^{y}_{N,S_n}(\boldsymbol{\gamma}_{n-1}, \boldsymbol{\delta}_{n-1}, \boldsymbol{\Psi}^y \mid \boldsymbol{\theta}_{n-1}) \\ &= \arg\max_{\boldsymbol{\Psi}^y} \tilde{Q}^{c}_{N,S_n}(\boldsymbol{\gamma}_{n-1}, \boldsymbol{\delta}_{n-1},, \boldsymbol{\Psi}^y, \boldsymbol{\Psi}^{s}_{n-1} \mid \boldsymbol{\theta}_{n-1}) \end{aligned} \tag{C.15d}$$

and:

$$(\boldsymbol{\gamma}_n, \boldsymbol{\delta}_n) \equiv \begin{cases} (\boldsymbol{\gamma}_*, \boldsymbol{\delta}_*) \text{ if } \tilde{Q}^{c}_{N,S_n}(\boldsymbol{\gamma}_*, \boldsymbol{\delta}_*, \boldsymbol{\Psi}^{y}_n, \boldsymbol{\Psi}^{s}_n \mid \boldsymbol{\theta}_{n-1}) > \tilde{Q}^{c}_{N,S_n}(\boldsymbol{\gamma}_{n-1}, \boldsymbol{\delta}_{n-1}, \boldsymbol{\Psi}^{y}_n, \boldsymbol{\Psi}^{s}_n \mid \boldsymbol{\theta}_{n-1}) \\ (\boldsymbol{\gamma}_{n-1}, \boldsymbol{\delta}_{n-1}) \text{ otherwise} \end{cases} \tag{C.15e}$$

where:

$$(\boldsymbol{\gamma}_*, \boldsymbol{\delta}_*) \equiv \arg\max_{(\boldsymbol{\gamma},\boldsymbol{\delta})} \tilde{Q}^{y}_{N,S_n}(\boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\Psi}^{y}_n \mid \boldsymbol{\theta}_{n-1}).^{23} \tag{C.15f}$$

The idea underlying the definition of $(\boldsymbol{\gamma}_n, \boldsymbol{\delta}_n)$ in equation (C.15e) is that of the Ascent-based Monte Carlo EM algorithms of Caffo *et al.* (2005). This simplified version of the "ascent-based" device of Caffo, Jank and Jones ensures that $\tilde{Q}^{c}_{N,S_n}(\boldsymbol{\mu}_n \mid \boldsymbol{\theta}_{n-1}) > \tilde{Q}^{c}_{N,S_n}(\boldsymbol{\mu}_{n-1} \mid \boldsymbol{\theta}_{n-1})$.

In our case, the terms $\boldsymbol{\tau}_n$, $\boldsymbol{\Omega}_n$, $\boldsymbol{\Psi}^{y}_n$, $\boldsymbol{\Psi}^{s}_n$ and $(\boldsymbol{\gamma}_n, \boldsymbol{\delta}_n)$ have analytical closed form solutions:

$$\boldsymbol{\tau}_n = N^{-1} \sum_{i=1}^{N} S_n^{-1} \sum_{s=1}^{S_n} \tilde{\omega}_{i,s,n-1} \tilde{\mathbf{q}}_{i,s,n-1} \tag{C.16a}$$

$$\boldsymbol{\Omega}_n = N^{-1} \sum_{i=1}^{N} S_n^{-1} \sum_{s=1}^{S_n} \tilde{\omega}_{i,s,n-1} \tilde{\mathbf{q}}_{i,s,n-1} \tilde{\mathbf{q}}'_{i,s,n-1} - \boldsymbol{\tau}_n \boldsymbol{\tau}'_n \tag{C.16b}$$

$$\boldsymbol{\Psi}^{s}_n = N^{-1} \sum_{i=1}^{N} S_n^{-1} \sum_{s=1}^{S_n} \tilde{\omega}_{i,s,n-1} \sum_{t=1}^{T} \tilde{\mathbf{e}}^{s}_{it,s,n-1} (\tilde{\mathbf{e}}^{s}_{it,s,n-1})' \tag{C.16c}$$

[23] The update of $(\boldsymbol{\gamma}, \boldsymbol{\delta})$ given in equations (CB.15e) and (C15f) is used to overcome the maximization of $\tilde{Q}^{\mathbf{c}}_{N,n}(\boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\Lambda}_n, \boldsymbol{\Psi}_n \mid \boldsymbol{\theta}_{n-1})$ in $(\boldsymbol{\gamma}, \boldsymbol{\delta})$. The solution in $(\boldsymbol{\gamma}, \boldsymbol{\delta})$ to this maximization problem has an analytical closed form solution in the case considered here, *i.e.* for the "random parameter" model. But the counterparts of the function $\tilde{Q}^{\mathbf{s}}_{N,n}(\boldsymbol{\gamma}, \boldsymbol{\Psi} \mid \boldsymbol{\theta}_{n-1})$ for the "fixed parameters" and the "fixed/random parameters" models need to be numerically maximized in their fixed parameters.

$$\boldsymbol{\Psi}_n^y = N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n}\tilde{\omega}_{i,s,n-1}\sum_{t=1}^{T}\tilde{\mathbf{e}}_{it,s,n-1}^y(\tilde{\mathbf{e}}_{it,s,n-1}^y)' \tag{C.16d}$$

and:

$$(\boldsymbol{\gamma}_*,\boldsymbol{\delta}_*) \equiv \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{Z}_{it}\boldsymbol{\Lambda}_n^{-1}\mathbf{Z}_{it}'\right)^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n}\tilde{\omega}_{i,s,n-1}\sum_{t=1}^{T}\mathbf{Z}_{it}\boldsymbol{\Lambda}_n^{-1}(\mathbf{y}_{it}-\tilde{\boldsymbol{\beta}}_{i,s,n-1}) \tag{C.16e}$$

with obvious notations.

C.4. "Ascent-based SACME" algorithm

When new random terms $\tilde{\mathbf{q}}_{i,s,n-1}$ are drawn at each iteration, the numerical convergence of SECM algorithm may be difficult due to the simulation noise. Delyon *et al.* (1999) proposed the SAEM algorithms in order to attenuate this problem. In the SAEM algorithms the M step (or the sequence of CM steps) is modified in order to "smooth" out the objective functions considered along the SE(C)M iteration process. *E.g.*, the objective function of the M step of the SEM algorithm given above, *i.e.* $\tilde{Q}_{N,S_n}(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}_{n-1})$, is replaced by:

$$\tilde{P}_{N,S_n}(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}_{n-1}) \equiv (1-\vartheta_n)\times\tilde{P}_{N,S_{n-1}}(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}_{n-2}) + \vartheta_n\times\tilde{Q}_{N,S_n}(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}_{n-1}) \tag{C.17a}$$

or, equivalently, by

$$\tilde{P}_{N,S_n}(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}_{n-1}) = \sum_{j=1}^{n-1}\left(\prod_{m=j+1}^{n}(1-\vartheta_m)\right)\times\vartheta_j\times\tilde{Q}_{N,S_j}(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}_{j-1}) + \vartheta_n\times\tilde{Q}_{N,S_n}(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}_{n-1}) \tag{C.17b}$$

where ϑ_n is a decreasing sequence of positive step size such that:

$$\vartheta_1 = 1,\ \sum_{n=1}^{+\infty}\vartheta_n = +\infty \text{ and } \sum_{n=1}^{+\infty}(\vartheta_n)^2 < +\infty \tag{C.18}$$

The "SACME" algorithm we used for computing the estimators of our random parameter multicrop model is defined by:

SE step. Integration of the conditional expectations

Obtain independent random draws $\tilde{\mathbf{q}}_{i,s,n-1}$ from $h(\mathbf{q};\boldsymbol{\eta}_{n-1})$ and compute the weight terms $\tilde{\omega}_{i,s,n-1}$ for $s = 1,...,S_n$ and $i = 1,...,N$.

CM step. Conditional update of the value of $\boldsymbol{\theta} \equiv (\boldsymbol{\mu},\boldsymbol{\eta})$

Compute:

$$\boldsymbol{\tau}_n = (1-\vartheta_n)\times\boldsymbol{\tau}_{n-1} + \vartheta_n \times N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n}\tilde{\omega}_{i,s,n-1}\tilde{\mathbf{q}}_{i,s,n-1} \tag{C.19a}$$

$$\boldsymbol{\Omega}_n = (1-\vartheta_n)\times(\boldsymbol{\Omega}_{n-1}+\boldsymbol{\tau}_n\boldsymbol{\tau}_n') + \vartheta_n \times N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n}\tilde{\omega}_{i,s,n-1}\tilde{\mathbf{q}}_{i,s,n-1}\tilde{\mathbf{q}}_{i,s,n-1}' - \boldsymbol{\tau}_n\boldsymbol{\tau}_n' \tag{C.19b}$$

$$\boldsymbol{\Psi}_n^s = (1-\vartheta_n)\times\boldsymbol{\Psi}_{n-1}^s + \vartheta_n \times N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n}\tilde{\omega}_{i,s,n-1}\sum_{t=1}^{T}\tilde{\mathbf{e}}_{it,s,n-1}^s(\tilde{\mathbf{e}}_{it,s,n-1}^s)' \tag{C.19c}$$

$$\boldsymbol{\Psi}_n^y = (1-\vartheta_n)\times\boldsymbol{\Psi}_{n-1}^y + \vartheta_n \times N^{-1}\sum_{i=1}^{N} S_n^{-1}\sum_{s=1}^{S_n}\tilde{\omega}_{i,s,n-1}\sum_{t=1}^{T}\tilde{\mathbf{e}}_{it,s,n-1}^y(\tilde{\mathbf{e}}_{it,s,n-1}^y)' \tag{C.19d}$$

and:

$$(\boldsymbol{\gamma}_*, \boldsymbol{\delta}_*) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{Z}_{it}\boldsymbol{\Lambda}_n^{-1}\mathbf{Z}_{it}'\right)^{-1} \times \sum_{j=1}^{n}\phi_{j,n}\sum_{i=1}^{N} S_j^{-1}\sum_{s=1}^{S_j}\tilde{\omega}_{i,s,j-1}\sum_{t=1}^{T}\mathbf{Z}_{it}\boldsymbol{\Lambda}_n^{-1}(\mathbf{y}_{it}-\tilde{\boldsymbol{\beta}}_{i,s,j-1}) \tag{C.19e}$$

where

$$\phi_{j,n} \equiv \left(\prod_{m=j+1}^{n}(1-\vartheta_m)\right)\times\vartheta_j \text{ for } j=1,...,n-1 \text{ and } \phi_{n,n}\equiv\vartheta_n. \tag{C.20}$$

C.5. Monitoring and stopping rule of the algorithm

The calibration of the sequence steps ϑ_n and a suitable stopping rule for the "SACME" algorithm are essential criteria for its convergence. SAEM algorithms are shown to theoretically converge if ϑ_n is a positive sequence steps satisfying conditions (C.18). We used a standard decreasing sequence of positive step sizes (see, *e.g.*, Jank, 2006 ; Polyak and Juditski, 1992) :

$$\vartheta_1 = 1,\ \vartheta_n = n^{-\nu} \text{ with } \nu\in(1/2,1] \tag{C.21}$$

We retained $\nu = 0.7$ after several trials.

We also used a standard stopping rule (Booth and Hobert, 1999; Booth *et al.*, 2001) based on the relative changes in the values of the estimated parameters from an iteration to the next one. The algorithm stops when the following condition:

$$\max_j \left(\frac{\left| \theta_{j,n} - \theta_{j,n-1} \right|}{\left| \theta_{j,n} \right| + \sigma_1} \right) < \sigma_2 \tag{C.22}$$

holds for three consecutive iterations for chosen positive values of the convergence parameters σ_1 and σ_2. Several iterations need to be considered due to the simulation noise generated by the random draws of the $\mathbf{q}_i$ terms at each iteration (see, *e.g.*, Booth and Hobert, 1999). We set up $\sigma_1 = .01$ and $\sigma_2 = .001$. Because condition (C.22) may hold for $\boldsymbol{\theta}_{n-2}$, $\boldsymbol{\theta}_{n-1}$ and $\boldsymbol{\theta}_n$ even if these parameter values do not (approximately) achieve the maximum of the considered likelihood function, we checked that the scores were null and that the Hessian matrix was negative definite at the estimated value of $\boldsymbol{\theta}$ (Gu and Zhu, 2001).

Les **Working Papers SMART – LERECO** sont produits par l'UMR SMART et l'UR LERECO

- **UMR SMART**

L'Unité Mixte de Recherche (UMR 1302) *Structures et Marchés Agricoles, Ressources et Territoires* comprend l'unité de recherche d'Economie et Sociologie Rurales de l'INRA de Rennes et les membres de l'UP Rennes du département d'Economie Gestion Société d'Agrocampus Ouest.

Adresse :
UMR SMART - INRA, 4 allée Bobierre, CS 61103, 35011 Rennes cedex
UMR SMART - Agrocampus, 65 rue de Saint Brieuc, CS 84215, 35042 Rennes cedex

- **LERECO**

Unité de Recherche *Laboratoire d'Etudes et de Recherches en Economie*
Adresse :
LERECO, INRA, Rue de la Géraudière, BP 71627 44316 Nantes Cedex 03

Site internet commun : http://www.rennes.inra.fr/smart

Liste complète des Working Papers SMART – LERECO :
http://www.rennes.inra.fr/smart/Working-Papers-Smart-Lereco
http://ideas.repec.org/s/rae/wpaper.html

The **Working Papers SMART – LERECO** are produced by UMR SMART and UR LERECO

- **UMR SMART**

The « Mixed Unit of Research » (UMR1302) *Structures and Markets in Agriculture, Resources and Territories*, is composed of the research unit of Rural Economics and Sociology of INRA Rennes and of the members of the Agrocampus Ouest's Department of Economics Management Society who are located in Rennes.

Address:
UMR SMART - INRA, 4 allée Bobierre, CS 61103, 35011 Rennes cedex, France
UMR SMART - Agrocampus, 65 rue de Saint Brieuc, CS 84215, 35042 Rennes cedex, France

- **LERECO**

Research Unit *Economic Studies and Research Lab*
Address:
LERECO, INRA, Rue de la Géraudière, BP 71627 44316 Nantes Cedex 03, France

Common website: http://www.rennes.inra.fr/smart_eng/

Full list of the Working Papers SMART – LERECO:
http://www.rennes.inra.fr/smart_eng/Working-Papers-Smart-Lereco
http://ideas.repec.org/s/rae/wpaper.html

Contact

Working Papers SMART – LERECO
INRA, UMR SMART
4 allée Adolphe Bobierre, CS 61103
35011 Rennes cedex, France
Email : smart_lereco_wp@rennes.inra.fr

2015

Working Papers SMART – LERECO

UMR INRA-Agrocampus Ouest **SMART** (Structures et Marchés Agricoles, Ressources et Territoires)

UR INRA **LERECO** (Laboratoire d'Etudes et de Recherches en Economie)

Rennes, France