Estimating crop rotation effects with farm accountancy panel data

A. Carpentier, F. Femenia, A. Gohin, R. Sodjahin

INRA, UMR SMART-LERECO, Rennes

Estimating crop rotation effects with farm accountancy panel data

Abstract

Crop rotations effects are key features of environmentally friendly crop production practices but they are poorly documented. The aim of this article is to present an estimation approach of these effects based on farm accountancy data. Estimating crop rotation effects on yield and variable input use levels from farm accountancy data is a challenging issue since farmers' crop sequence choices are not observed. In response to this data issue, we propose an original approach designed to estimate the crop rotation effects while simultaneously reconstructing farmers' unobserved crop sequences from farmers' observed crop production choices. This estimation approach is based on a well-defined statistical background. It relies on simple crop sequence yield and input use models as well as on an assumption stating that farmers are economically rational when deciding their crop sequence acreages. Our approach also makes use of expert knowledge information on crop rotation effects for 'guiding' the construction of the crop rotation effect estimators. An illustrative application based on French farm accountancy data demonstrate the empirical tractability of the proposed estimation approach and shows that it yields meaningful estimates of crop rotation effects.

Keywords: crop rotation effects, farm accounting data, MPEC problem

JEL codes: Q12, Q15, C61

Estimating crop rotation effects with farm accountancy panel data

Introduction

Crop rotation effects capture the effects of the cropping history of a plot on the current production process of a crop grown on this plot. The cropping history of a plot – *i.e.* the sequence of crop grown on this plot[1] – largely impacts pest and weed populations as well as the structure and nutrient content of soils. As a result, crop rotation effects can significantly impact chemical input uses (pesticides and fertilizers) and labor and fuel uses (*e.g.*, for tillage and mechanical weeding). They can also impact yield levels directly (*e.g.*, by controlling pest population that cannot be controlled by chemical pesticides) or through partial adjustments of input use levels. In other words, the cropping history of a plot impact the potential yield level of the crop grown on this plot and the chemical input quantities required to achieve this potential yield level.

Many effects of crop rotation can be interpreted as eco-systemic services that can be used as substitutes for chemical inputs. Exploiting the agronomic effects of crop rotation is a key strategy used by agricultural scientists, extension agents or farmers for designing chemical input saving agricultural production systems. As a result, investigating and measuring the effects of crop rotations on yield levels and uses of chemical inputs produce essential information for agri-environmental policies aimed to produce food, feed and fuel with limited amounts of chemical inputs.

---

[1] Despite its evoking cycles and repetitions, the expression "crop rotation" is used to denote crop sequences in agricultural science, whether these sequences entail cycles and repetitions or not. Indeed, crop rotations underlie the main dynamic features of the multi-crop production technology.

Crop rotation effects have rarely been quantified despite their potential effects on farmers' production choices, and, when available, crop rotation effect measurements are only for a few pairs of major crops in specific areas (Meynard *et al*, 2013, Carpentier *et al*, 2018). This is probably due to the cost of suitable experiments necessary to collect these data.

Our aim is to propose an original approach to estimate crop rotation effects based on existing datasets, namely on large panel datasets of farm accountancy data with cost accounting. Such datasets describe crop yields, acreages and input uses for a large sample of farms over a few years.[2]

The key issue we face is a lack of information on farmers' crop sequence acreages. If data on farmers' *crop* acreages are common, data on *crop sequence* acreages are seldom. In particular, they are not available in farm accountancy data, making the identification of the economic motivation of crop rotation impossible. In response to this data issue, our approach relies on four main elements: (*i*) statistical models of yield and input use at the crop sequence level; (*ii*) assumptions related to famers' economic rationality regarding their use of crop rotation effects; (*iii*) estimation approaches that allow recovering the unobserved crop sequence acreages while

[2] While experimental data allow estimating crop rotation effects on crop production processes, our estimation approach based on farm accountancy data aims to recover these effects as well as to investigate how farmers make use of them. In particular, the extent to which farmers reduce their chemical input uses thanks to favorable crop rotation effects largely remains a research question. For instance, legumes fix atmospheric nitrogen for themselves while leaving significant nitrogen surpluses in soils for the succeeding crops. Extension agents report that farmers often do not reduce their nitrogen fertilizer use on crops produced after legumes, leading to economically inefficient uses of chemical fertilizers (Carrouée *et al*, 2012).

simultaneously estimating the crop rotation effects underlying these acreages; (*iv*) expert knowledge information – obtained by interviewing agricultural scientists and extension agents – that can be translated into modelling assumptions.

We are interested in first order crop rotation effects, namely on the effects of the preceding crops on the yield and input use levels of the current crops. Cost accounting data report the yields and input use levels for each crop but do not report these data at the crop sequence level. Yet, the observed yield of a given crop is the weighted average of the unobserved per preceding crop yields, the weights being defined by the share of the acreage of the considered crop produced on plots with specific preceding crops. Of course, estimating crop rotation effects based on this linear relationship would be easy if the relevant crop sequence acreages were observed. For instance, simply regressing the observed crop yield and input use levels on the crop sequence acreage share vectors (and, possibly, on heterogeneity control and contextual variables) would yield consistent estimates of crop rotation estimates under suitable exogeneity conditions.

This is where our assumptions related to famers' economic rationality take place. These assumptions enable us to "reconstruct" famers' crop sequence choices as functions of the crop rotation effects of interest along the estimation process of these effects. This follows a simple observation.

A series of crop acreage choices is observed for each sampled farm. Recovering the crop sequence acreages chosen by expected profit maximizing (and forward-looking) farmers from their current and past crop acreages is a standard linear optimal transport problem when crop rotation effects are known. This allows us to devise a crop sequence acreage share "reconstruction process" based on any candidate estimates of the crop rotation effects. This "reconstruction process" can then be combined with standard estimation approaches to define the crop rotation effect estimation problem as a mathematical programming (MPEC) problem.

Our estimation approach builds on the seminal works of Rust (1987, 1988) on the estimation of dynamic discrete choice models, of Berry *et al* (1995, 2004) on that of market shares models with endogenous prices as well as on the subsequent contributions of Su and Judd (2012) and Dubé *et al* (2012).

This paper is organized as follows. The first section presents and discusses our crop sequence yield level and input use models. In the second section, we present a simple approach for computing optimal crop sequences given past crop acreages and targeted current crop acreages. In the third section, we show that crop rotation effects can be estimated with farm accountancy data by solving a MPEC problem. The fourth section discusses the issues raised by the implementation of our estimation approach the solutions that we consider for overcoming the described problems. An application to French data is presented the fifth section to illustrate the empirical tractability of our approach. Finally, we conclude.

1. Yield and input use levels, and crop rotation effects

We consider a panel data set describing the crop level production choices of a large sample of arable crop producers, $i \in \mathcal{I}$ with $\mathcal{I} = \{1,...,N\}$, over a few years, $t \in \{0\} \cup \mathcal{T}$ with $\mathcal{T} = \{1,...,T\}$. The considered crop set is denoted by $\mathcal{K} = \{1,...,K\}$ and farmers use J variable inputs – pesticides, fertilizers, *etc* – to produce each crop. For each (k,i,t) triplet these data report:

- $a_{k,i,t}$: crop *k* acreage share chosen by farmer *i* in year *t*, with $a_{k,i,t} \in [0,1]$ and $\sum_{k \in \mathcal{K}} a_{k,i,t} = 1$,
- $y^0_{k,i,t}$: crop *k* yield level obtained by farmer *i* in year *t*,

- $y_{k,i,t}^{j}$: quantity of input $j \in \{1,...,J\}$ used by farmer *i* in year *t* per unit of acreage of crop *k*,
- $p_{k,i,t}^{0}$: expected selling price of crop *k* of farmer *i* in year *t*, this price anticipation being defined *a priori*,
- $w_{k,i,t}^{j}$: purchase price of input *j* of farmer *i* in year *t*.

Assuming that farmers mostly account for crop rotation effects of order 1, their acreage, yield and input use levels need to be defined at the crop pair level. Let define:

- $s_{mk,i,t}^{*}$ as the acreage share of crop *k* grown on plots with preceding crop *m* by farmer *i* in year *t*, $\sum_{m \in \mathcal{K}} s_{mk,i,t}^{*} = a_{k,i,t}$ and $s_{mk,i,t} = 0$ necessarily holds if $a_{m,i,t-1} = 0$,
- $y_{mk,i,t}^{0,*}$ as the yield level of crop *k* obtained on plots with preceding crop *m* by farmer *i* in year *t*,
- $y_{mk,i,t}^{j,*}$ as the quantity of input *j* used by farmer *i* in year *t* for crop *k* on plots with preceding crop *m*.

Stars are used to indicate unobservable variables.

Netput notations

In order to simplify notations in what follows we define the following compact "netput" notations:

- Crop netput quantity vector: $\mathbf{y}_{k,i,t} = (y_{k,i,t}^{j}, j \in \mathcal{J})$,
- Crop sequence netput quantity vector: $\mathbf{y}_{mk,i,t}^{*} = (y_{mk,i,t}^{j,*}, j \in \mathcal{J})$,
- Netput price vector: $\mathbf{p}_{k,i,t} = (p_{k,i,t}^{0}, -w_{k,i,t}^{1}, ..., -w_{k,i,t}^{J}) = (p_{k,i,t}^{j} : j \in \mathcal{J})$

where $j \in \mathcal{J} \equiv \{0,1,...,J\}$.

Crop netput quantities as weighted sums of crop sequence netput quantities

Importantly, the *crop level* netput vectors $\mathbf{y}_{k,i,t}$ and acreage shares $a_{k,i,t}$ are observed while the *crop sequence level* netput vectors $\mathbf{y}^{*}_{mk,i,t}$ and acreage shares $s^{*}_{mk,i,t}$ are not observed. Nevertheless, the observed crop netput and unobserved crop sequence netput levels are linked, by construction, *via* a simple linear relationship with:

$$y^{j}_{k,i,t} = \sum\nolimits_{m\in\mathcal{K}_{i,t-1}} s^{*}_{mk,i,t} a^{-1}_{k,i,t} y^{j,*}_{mk,i,t} \tag{1}$$

where $\mathcal{K}_{i,t}$ is the subset of crop produced by farmer *i* in year *t*. Term $s_{mk,i,t}a^{-1}_{k,i,t}$ defines the share of crop *k* acreage with preceding crop *m* used by farmer *i* in year *t*. Of course, these links between "what is observed" and "what is unobserved but contains the effects of interest" is the keystone our estimation approach.

Crop and crop sequence netput quantity models

Statistical models of the crop sequence netput quantities need to be specified for enabling us to estimate the effects of crop rotation. In this paper we consider simple (if not simplistic) crop sequence netput quantity models. Let assume that the crop rotation effects can be specified as homogenous – across farms and time – additive effects in the crop sequence netput level models with:

$$y^{j,*}_{mk,i,t} = \alpha^{j}_{k,i,t} + \beta^{j}_{mk,0}\,. \tag{2}$$

Assuming that crop rotation effects are constant across farms and years is admittedly restrictive, excepted in empirical settings considering samples of farms with homogenous productivity levels.[3] We choose this simple modelling framework because it leads to a simple description of

[3] See Carpentier and Letort (2012).

our estimation problem.

Under the homogeneity and linearity assumptions entailed in equation (2) the netput levels are given by:

$$y_{k,i,t}^{j}=\alpha_{k,i,t}^{j}+\sum\nolimits_{m\in\mathcal{K}_{i,t-1}}s_{mk,i,t}^{*}a_{k,i,t}^{-1}\beta_{mk,0}^{j} \tag{3}$$

or, equivalently, by:

$$y_{k,i,t}^{j}=\alpha_{k,i,t}^{j}+(\mathbf{s}_{a,k,i,t}^{*})'\mathbf{R}_{i,t-1}\boldsymbol{\beta}_{k,0}^{j}. \tag{4}$$

Vector $\mathbf{s}_{a,k,i,t}^{*}=(s_{mk,i,t}^{*}a_{k,i,t}^{-1}:m\in\mathcal{K}_{i,t-1})$ describes how crop k acreage is allocated to land areas with specific preceding crops. Vector $\boldsymbol{\beta}_{k,0}^{j}=(\beta_{mk,0}^{j}:m\in\mathcal{K})$ is the parameter vector consisting of the crop rotation effects on netput j for crop k. Matrix $\mathbf{R}_{i,t-1}$ is a selection matrix defined such that $\mathbf{R}_{i,t-1}\boldsymbol{\beta}_{k,0}^{j}=(\beta_{mk,0}^{j}:m\in\mathcal{K}_{i,t-1})$. The netput quantity models described by equations (3) and (4) are linear in the crop rotation effects and, as a result, provide a simple background for estimating these effects in a well-suited statistical framework.

Let finally assume that the "basis netput quantities" $\alpha_{k,i,t}^{j}$ can be modelled as simple linear models with farm and year specific effects:

$$\alpha_{k,i,t}^{j}=\alpha_{k,i}^{j}+\delta_{k,t,0}^{j}+\varepsilon_{k,i,t}^{j}\ \text{ with }\ E[\alpha_{k,i}^{j}]=E[\varepsilon_{k,i,t}^{j}]=0. \tag{5}$$

Control and contextual variable effects could easily be included in this linear set up. In usual farm accountancy panel data sets the cross-sectional dimension N is sufficiently large for obtaining accurate estimates of the year specific parameters $\delta_{k,t,0}^{j}$ while the time dimension T is generally too small for obtaining accurate estimates of the farm specific effects $\alpha_{k,i}^{j}$. Because the elements of $\mathbf{s}_{a,k,i,t}^{*}$ sum to 1 by construction, the models described by equations (3) and (4) are under-identified and normalization conditions need to be imposed to allow the identification of their parameters. Imposing the normalization conditions $T^{-1}\sum_{t=1}^{T}\delta_{k,t,0}^{j}=0$ and $E[\alpha_{k,i}^{j}]=0$ implies that parameter $\beta_{mk,0}^{j}$ defines the expected quantity of netput j characterizing the crop sequence (m,k), with $E[y_{a,k,i,t}^{*}]=\beta_{mk,0}^{j}$. This normalization choice is unusual but it has a major virtue: it doesn't require any normalization condition on the $\beta_{mk,0}^{j}$ parameters.[4]

The assumption set described above implies that the observed netput quantities are modeled as:

$$y_{k,i,t}^{j}=\delta_{k,t,0}^{j}+\alpha_{k,i}^{j}+(\mathbf{s}_{a,k,i,t}^{*})'\mathbf{R}_{i,t-1}\boldsymbol{\beta}_{k,0}^{j}+\varepsilon_{k,i,t}^{j} \text{ with } E[\alpha_{k,i}^{j}]=E[\varepsilon_{k,i,t}^{j}]=0. \tag{6}$$

Let now pile up these models for $j\in\mathcal{J}$:

$$\mathbf{y}_{k,i,t}=\boldsymbol{\delta}_{k,t,0}+\boldsymbol{\alpha}_{k,i}+\mathbf{S}_{a,k,i,t}^{*}\boldsymbol{\beta}_{k,0}+\boldsymbol{\varepsilon}_{k,i,t} \text{ with } E[\boldsymbol{\alpha}_{k,i}]=E[\boldsymbol{\varepsilon}_{k,i,t}]=\mathbf{0} \tag{7}$$

where $\mathbf{S}_{a,k,i,t}^{*}=\mathbf{I}_{J+1}\otimes\left((\mathbf{s}_{a,k,i,t}^{*})'\mathbf{R}_{i,t-1}\right)$, obvious notations being used for vectors $\mathbf{y}_{k,i,t}$, $\boldsymbol{\alpha}_{k,i}$, $\boldsymbol{\delta}_{k,t,0}$, $\boldsymbol{\beta}_{k,0}$ and $\boldsymbol{\varepsilon}_{k,i,t}$. Let then pile up the obtained equation systems for $k\in\mathcal{K}_{i,t}$:

$$\mathbf{y}_{i,t}=\mathbf{Q}_{i,t}^{\delta}\boldsymbol{\delta}_{t,0}+\boldsymbol{\alpha}_{i}+\mathbf{S}_{a,i,t}^{*}\boldsymbol{\beta}_{0}+\boldsymbol{\varepsilon}_{i,t} \text{ with } E[\boldsymbol{\alpha}_{i}]=E[\boldsymbol{\varepsilon}_{i,t}]=\mathbf{0}. \tag{8}$$

Obvious notations are used for vectors $\mathbf{y}_{k,i,t}$, $\boldsymbol{\beta}_{k,0}$ and $\boldsymbol{\varepsilon}_{k,i,t}$, and matrix $\mathbf{S}_{a,i,t}^{*}$. Vector $\boldsymbol{\delta}_{t,0}$ is obtained by piling up vectors $\boldsymbol{\delta}_{k,t,0}$ for $k\in\mathcal{K}$ (not for $k\in\mathcal{K}_{i,t}$), with $\boldsymbol{\delta}_{t,0}=(\boldsymbol{\delta}_{k,t,0}:k\in\mathcal{K})$. Matrix $\mathbf{Q}_{i,t}^{\delta}$ is a selection matrix defined such that $\mathbf{Q}_{i,t}^{\delta}\boldsymbol{\delta}_{t,0}=(\boldsymbol{\delta}_{k,t,0}:k\in\mathcal{K}_{i,t})$. Set $\mathcal{K}_{i}$ is the set of crops produced by farmer i at least once and vector $\boldsymbol{\alpha}_{i}$ is obtained by piling up vectors $\boldsymbol{\alpha}_{k,i}$ for $k\in\mathcal{K}_{i}$, with $\boldsymbol{\alpha}_{i}=(\boldsymbol{\alpha}_{k,i}:k\in\mathcal{K}_{i})$. Matrix $\mathbf{Q}_{i,t}^{\alpha}$ is a selection matrix defined such that

[4] Usual normalization conditions rely on a reference preceding crop: $r(k)$ defines the reference preceding crop for crop k and the normalization constraint $\beta_{r(k)k,0}^{j}=0$ is imposed. A suitable reference crop of a given crop is the most frequent preceding crop of the considered crop. If the choice of the reference crop is obvious for some crops, this choice is problematic for others, unless suitable information on crop sequence acreages are available.

$\mathbf{Q}_{i,t}^{\alpha}\boldsymbol{\alpha}_i = (\boldsymbol{\alpha}_{k,i} : k \in \mathcal{K}_{i,t})$. Obvious notations are used otherwise.

Sampling and exogeneity assumptions

We assume that the "vectors" $\left((\mathbf{y}_{i,t}, \mathbf{S}_{a,i,t}^{*}) : t \in \mathcal{T}\right)$ are independent and identically distributed across the sampled farms. Assuming that (*i*) the crop rotation effects $\mathbf{S}_{a,i,t}^{*}\boldsymbol{\beta}_0$ suitably capture the dynamic features of the multicrop technology and (*ii*) the farm effects $\boldsymbol{\alpha}_i$ suitably capture the persistent features of farmers' choices implies that the error term vector $\boldsymbol{\varepsilon}_{i,t}$ can be assumed serially uncorrelated.[5]

Of course, estimating the crop rotation parameter vector $\boldsymbol{\beta}_0$ would be fairly easy under these modelling assumptions if the crop sequence acreages collected in matrix $\mathbf{S}_{a,i,t}^{*}$ were observed. Suitable estimators of $\boldsymbol{\beta}_0$ depend on what can be assumed regarding the statistical links between $\mathbf{S}_{a,i,t}^{*}$ and $(\boldsymbol{\alpha}_i, \boldsymbol{\varepsilon}_{i,t})$. The most crucial assumptions are related to the links between $\mathbf{S}_{a,i,t}^{*}$ and the error term vectors $\boldsymbol{\varepsilon}_{i,t}$ for $t \in \mathcal{T}$ as standard panel data estimators allows to deal with potential correlation of $\mathbf{S}_{a,i,t}^{*}$ and $\boldsymbol{\alpha}_i$.

Assuming that the crop sequence acreage share matrix $\mathbf{S}_{a,i,t}^{*}$ is strictly exogenous with respect to the error term vectors $\boldsymbol{\varepsilon}_{i,t}$:

$$E[\boldsymbol{\varepsilon}_{i,t} \mid (\mathbf{S}_{a,i,t}^{*} : t \in \mathcal{T}), \boldsymbol{\alpha}_i] = \mathbf{0} \tag{9}$$

[5] Of course, the assumed homogeneity across farms and years question this assumption. Note that the elements of $\boldsymbol{\varepsilon}_{i,t}$ are expected to be strongly correlated, across netputs as well as across crops.

implies that simple linear regression based estimators are consistent for $\boldsymbol{\beta}_0$. Netput quantities $y_{mk,i,t}^{j,*}$ depend on production choices and random events that occur after, sometimes by far, farmers' acreage decision making. This implies that $\boldsymbol{\varepsilon}_{i,t}$ (as well as $\boldsymbol{\delta}_{t,0}$) largely depends on events that are unknown to farmers when their choosing $\mathbf{S}_{a,i,t}^{*}$. This provides arguments for assuming that $\mathbf{S}_{a,i,t}^{*}$ is strictly exogenous with respect to $\boldsymbol{\varepsilon}_{i,t}$ conditionally on $\boldsymbol{\alpha}_i$.[6]

Importantly, we assume here that farm specific effects $\boldsymbol{\alpha}_i$ control for endogeneity biases with respect to $\boldsymbol{\varepsilon}_{i,t}$. But, correlations between elements of $\boldsymbol{\alpha}_i$ and $\mathbf{S}_{a,i,t}^{*}$ cannot be ruled out. For example, efficient farmers endowed with productive land tend to obtain higher sugar beet yields when they grow this crop. They also tend to choose larger sugar beet acreage shares and, mechanically, to have larger crop sequence acreage shares with sugar beet as the preceding crop. Indeed, farmers' crop sequence acreage choices – $s_{mk,i,t}^{*}$ – directly depends on farmers' current and past crop acreage shares – $a_{k,i,t}$ and $a_{m,i,t-1}$ – and, in particular, on their current and past production sets – $\mathcal{K}_{i,t}$ and $\mathcal{K}_{i,t-1}$. Standard "Within" (or farm fixed effects estimators) allow accounting for such correlations. Control function approaches *à la* Mundlak (1978) – *e.g.*, based

[6] Note however that specific random events can significantly impact $\mathbf{S}_{a,i,t}^{*}$ and $\boldsymbol{\varepsilon}_{i,t}$, and generate confounding effects. For instance, extreme climatic conditions at planting can impact both farmers' acreage choices and the obtained crop yield levels (due to unfavorable conditions in the early stages of the crop growth process). When the crop set contains autumn and spring crops (*i.e.*, crops that are sown in autumn and crop that are sown in spring), farmers may find profitable to destroy a crop sown in autumn for replacing it by a spring crop. This occurs when the autumn crop expected yield level is very low due to damages caused by exceptional events (*e.g.*, long frost periods, severe frosts in late winter, heavy winter rainfalls, severe diseases, pest infestations in early spring). With farm samples covering small areas, however, such effects are expected to be common to most farms and thus largely captured by the year specific parameters $\boldsymbol{\delta}_{t,0}$.

on linear projections on crop acreage farm specific means, crop production farm specific frequencies, *etc* – could also be used.

Estimating crop rotation effects with observed crop sequence acreages

Simple suitable estimators of $\boldsymbol{\beta}_0$ can be defined as solutions in $\boldsymbol{\beta}$ to quadratic optimization problems. These estimators are generically defined as FGLS estimators:

$$(\hat{\boldsymbol{\delta}},\hat{\boldsymbol{\beta}}) = \arg\min_{(\boldsymbol{\delta},\boldsymbol{\beta})} \sum\nolimits_{t=1}^{T} \sum\nolimits_{i=1}^{N} \mathbf{u}'_{i,t}(\mathbf{y}_i,\mathbf{S}^*_{a,i},\boldsymbol{\delta},\boldsymbol{\beta})(\mathbf{R}_{i,t}\hat{\boldsymbol{\Omega}}\mathbf{R}'_{i,t})^{-1}\mathbf{u}_{i,t}(\mathbf{y}_i,\mathbf{S}^*_{a,i};\boldsymbol{\delta},\boldsymbol{\beta}) \tag{10}$$

where $\boldsymbol{\delta} = (\boldsymbol{\delta}_t : t \in \mathcal{T})$, $\mathbf{y}_i = (\mathbf{y}_{i,t} : t \in \mathcal{T})$ and $\mathbf{S}^*_{a,i} = (\mathbf{S}^*_{a,i,t} : t \in \mathcal{T})$. Functions $\mathbf{u}_{i,t}(.)$ are residual functions that are linear in $\boldsymbol{\beta}$.[7] Matrix $\hat{\boldsymbol{\Omega}}^{-1}$ is a weighting matrix. $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}_0$ for any matrix $\hat{\boldsymbol{\Omega}}$ as long as this matrix converges to a positive definite matrix.

Matrix $\hat{\boldsymbol{\Omega}}$ is generally designed for estimating an optimal weighting matrix $\boldsymbol{\Omega}_0$ that accounts for the variance matrix structure and heteroskedasticity of the error term vector $\boldsymbol{\varepsilon}_{i,t}$. Matrix $\hat{\boldsymbol{\Omega}}$ can be defined as a consistent estimator of $\boldsymbol{\Omega}_0$ in a final step to be implemented after obtaining a consistent estimate of $\boldsymbol{\beta}_0$ (and for obtaining a more efficient estimate of $\boldsymbol{\beta}_0$). Suitable albeit inconsistent estimates of $\boldsymbol{\Omega}_0$ can be constructed *a priori* based on the netput quantity vectors $\mathbf{y}_{i,t}$.[8]

When the estimation criteria is based on standard fixed effects estimators, the corresponding

[7] These estimators rest on demeaning techniques (*e.g.*, “Within” transformation), simple control functions (*e.g.*, linear control functions à la Mundlak) and linear regression techniques (OLS or FGLS). Including auxiliary parameters in the estimation problem is required in some cases. The presented estimation problem can easily accommodate these cases as long as the extended models are linear in the considered auxiliary parameters.

[8] As well as on the observed control and contextual variable vectors if needed.

estimation problem is simply defined by:

$$\min_{(\boldsymbol{\delta},\boldsymbol{\beta},\boldsymbol{\alpha})}\left\{\begin{array}{c}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{e}'_{i,t}(\boldsymbol{\delta}_t,\boldsymbol{\alpha}_i,\boldsymbol{\beta})(\mathbf{R}_{i,t}\hat{\boldsymbol{\Omega}}\mathbf{R}'_{i,t})^{-1}\mathbf{e}_{i,t}(\boldsymbol{\delta}_t,\boldsymbol{\alpha}_i,\boldsymbol{\beta})\\ \text{s.t.}\\ \mathbf{e}_{i,t}(\boldsymbol{\delta}_t,\boldsymbol{\alpha}_i,\boldsymbol{\beta})=\mathbf{y}_{i,t}-\mathbf{Q}^{\delta}_{i,t}\boldsymbol{\delta}_t-\mathbf{Q}^{\alpha}_{i,t}\boldsymbol{\alpha}_i-\mathbf{S}^{*}_{a,i,t}\boldsymbol{\beta},\ (i,t)\in\mathcal{I}\times\mathcal{T}\\ \sum_{t\in\mathcal{T}}\boldsymbol{\delta}_t=\mathbf{0}\\ \sum_{i\in\mathcal{I}(k)}\boldsymbol{\alpha}_{k,i}=\mathbf{0},\ \ k\in\mathcal{K}\end{array}\right\} \tag{11}$$

where $\boldsymbol{\alpha}=(\boldsymbol{\alpha}_i:i\in\mathcal{I})$ and set $\mathcal{I}(k)$ defines the sub-sample of farmers producing crop k at least once.

2. Farmers' crop sequence acreage share choices

The crop sequence acreage shares collected in $\mathbf{S}^{*}_{a,i,t}$ are not observed in our case. Yet, assuming that farmers are economically rational allows to define $\mathbf{S}^{*}_{a,i,t}$ as a function of $\boldsymbol{\beta}_0$. This function can in turn be used for defining an estimation problem for estimating $\boldsymbol{\beta}_0$.

Dynamic acreage choices and farmers' economic rationality

Let assume that farmer i expected profit (or utility) function is defined in year t by:

$$\Pi_{i,t}(\mathbf{s}_{i,t};\boldsymbol{\pi}_{i,t})=\sum_{m\in\mathcal{K}}\sum_{k\in\mathcal{K}}s_{mk,i,t}\pi_{mk,i,t}-C_{i,t}(\mathbf{a}_{i,t}) \tag{12}$$

if he chooses crop sequences acreages defined by the crop sequence acreage share vector $\mathbf{s}_{i,t}=\left(s_{mk,i,t}:(m,k)\in\mathcal{K}_{i,t-1}\times\mathcal{K}_{i,t}\right)$. Crop acreage share vector $\mathbf{a}_{i,t}=(a_{k,i,t}:k\in\mathcal{K}_{i,t})$ is given by $a_{k,i,t}=\sum_{m\in\mathcal{K}}s_{mk,i,t}$ and function $C_{i,t}(\mathbf{a})$ defines an acreage cost management function or/and a profit risk premium.[9]

Term $\pi_{mk,i,t}$ denotes the expected return of crop sequence (m,k). It is defined as farmers'

[9] See, *e.g.*, Carpentier and Letort (2012, 2014) and Carpentier and Gohin (2014, 2015).

expectation of the obtained crop sequence returns $\mathbf{p}'_{k,i,t}\mathbf{y}^{*}_{mk,i,t}$ when they choose their crop and crop sequence acreages. Given the properties of crop sequence netput quantities models described above, crop sequence expected returns can be parameterized as:

$$\pi_{mk,i,t}(\boldsymbol{\beta}) = \mathbf{p}'_{k,i,t}\boldsymbol{\alpha}_{k,i} + \mathbf{p}'_{k,i,t}\boldsymbol{\beta}_{mk} \tag{13}$$

for any candidate estimate of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}$. We assume here that farmers' expectation of the year specific effects $\delta^{j}_{k,t,0}$ is given by $T^{-1}\sum_{t\in\mathcal{T}}\delta^{j}_{k,t,0}$, theses term being null by the chosen normalization conditions. This assumption is admittedly restrictive when T is small.

Recovering crop sequence acreages from crop acreages series

Assuming that farmer i is forward-looking and accounts for crop rotation effects implies that his general optimization problem under uncertainty implicitly contains the following optimization problem, conditional on his choice of acreage shares $\mathbf{a}_{i,t}$ [10] (Carpentier *et al*, 2018; Fémenia *et al*, 2018):[11]

$$\text{Problem } LP_{i,t}: \max_{\mathbf{s}_{i,t}\geq\mathbf{0}} \left\{ \begin{array}{c} \sum_{m\in\mathcal{K}_{i,t-1}}\sum_{k\in\mathcal{K}_{i,t}} s_{mk,i,t}\pi_{mk,i,t}(\boldsymbol{\beta}) \\ \text{s.t.} \\ \sum_{m\in\mathcal{K}_{i,t-1}} s_{mk,i,t} = a_{k,i,t},\ k\in\mathcal{K}_{i,t} \\ \sum_{k\in\mathcal{K}_{i,t}} s_{mk,i,t} = a_{m,i,t-1},\ m\in\mathcal{K}_{i,t-1} \end{array} \right\}. \tag{14}$$

Nullity constraints $s_{mk,i,t}=0$ are implicitly imposed in Problem $LP_{i,t}$ for $(m,k)\notin\mathcal{K}_{i,t-1}\times\mathcal{K}_{i,t}$. Maximization Problem $LP_{i,t}$ has the structure of a standard linear optimal transport problem. A solution in $\mathbf{s}_{i,t}$ to this problem defines the expected profit maximizing allocations of the crop acreage chosen in year t, $a_{k,i,t}$ for $k\in\mathcal{K}_{i,t}$, to the available preceding crop acreages, $a_{m,i,t-1}$ for

[10] Acreage shares are indeed exogenous in this sub-problem, which explains why the acreage cost management function $C_{i,t}(\mathbf{a})$ does not appear here.

[11] See also Rockafellar (1999) for a broader perspective from the stochastic programming viewpoint.

$m \in \mathcal{K}_{i,t-1}$ (*i.e.*, those chosen in year $t-1$). The constraint set, namely the crop rotation constraints $\sum_{m \in \mathcal{K}_{i,t-1}} s_{mk,i,t} = a_{k,i,t}$ and $\sum_{k \in \mathcal{K}_{i,t}} s_{mk,i,t} = a_{m,i,t-1}$ for $(m,k) \in \mathcal{K}_{i,t-1} \times \mathcal{K}_{i,t}$, ensures that the crop sequence acreage share vector $\mathbf{s}_{i,t}$ is feasible given the crop acreages of farmer *i* in years $t-1$ and *t*. Note that equality $\sum_{k \in \mathcal{K}_{i,t}} a_{k,i,t} = \sum_{m \in \mathcal{K}_{i,t-1}} a_{m,i,t-1} = 1$ implies that a crop rotation constraint is redundant.

Assuming that the constraints $\sum_{m \in \mathcal{K}_{i,t-1}} s_{mk,i,t} = a_{k,i,t}$ are enforced for $k \in \mathcal{K}_{i,t}$, the sum $\sum_{k \in \mathcal{K}_{i,t}} s_{mk,i,t}$ defines the (derived) demand of land with preceding crop *m,* while $a_{m,i,t-1}$ defines the corresponding land supply, on farm *i* in year *t*. Similarly, assuming that the constraints $\sum_{k \in \mathcal{K}_{i,t}} s_{mk,i,t} = a_{m,i,t-1}$ are enforced for $m \in \mathcal{K}_{i,t-1}$, sum $\sum_{m \in \mathcal{K}_{i,t-1}} s_{mk,i,t}$ defines the supply of land devoted to crop *k* while $a_{k,i,t}$ defines the corresponding land demand.

Entropic perturbations for "smoothing" solutions to linear programming problems

Problem $LP_{i,t}$ is easy to solve and allows to determine (estimates of) the unobserved crop sequence acreage vectors $\mathbf{s}_{i,t}^{*}$ as "functions" of the crop rotation effect vector $\boldsymbol{\beta}_0$. Nevertheless, this problem is a linear programming problem, implying potentially multiple solutions in $\mathbf{s}_{i,t}$ and severe discontinuities of the corresponding solution set with respect to $\boldsymbol{\beta}$. A simple solution to these issues consists of "perturbing" Problem $LP_{i,t}$ with an entropic perturbation (Fang *et al*, 1997).

Let define the entropic perturbation

$$\kappa^{\rho}(\mathbf{s}_{i,t}) = \rho^{-1} \sum_{m \in \mathcal{K}_{i,t-1}} \sum_{k \in \mathcal{K}_{i,t}} s_{mk,i,t} \ln s_{mk,i,t} \text{ with } \rho > 0 \tag{15}$$

to define the following perturbed version of Problem $LP_{i,t}$:

$$\text{Problem } SLP_{i,t}^{\rho}:\ \max_{\mathbf{s}_{i,t}\geq\mathbf{0}}\left\{\begin{array}{c}\sum_{m\in\mathcal{K}_{i,t-1}}\sum_{k\in\mathcal{K}_{i,t}} s_{mk,i,t}\pi_{mk,i,t}(\boldsymbol{\beta})-\kappa^{\rho}(\mathbf{s}_{i,t})\\ \text{s.t.}\\ \sum_{m\in\mathcal{K}_{i,t-1}} s_{mk,i,t}=a_{k,i,t},\ k\in\mathcal{K}_{i,t}\\ \sum_{k\in\mathcal{K}_{i,t}} s_{mk,i,t}=a_{m,i,t-1},\ m\in\mathcal{K}_{i,t-1}\end{array}\right\}. \tag{16}$$

Suitably perturbed problems have two main advantages. First, the perturbation term $\kappa^{\rho}(\mathbf{s}_{i,t})$ being strictly convex in $\mathbf{s}_{i,t}$, it has a unique solution in $\mathbf{s}_{i,t}$. Second, the perturbation term $\kappa^{\rho}(\mathbf{s}_{i,t})$ converging uniformly to 0 as ρ goes to infinity when $\mathbf{s}_{i,t}$ satisfies the constraints of Problem $SLP_{i,t}^{\rho}$, the solution in **s** to Problem $SLP_{i,t}^{\rho}$ converges to a solution in **s** to Problem $LP_{i,t}$ as ρ goes to infinity.

Entropic perturbations in linear programming problems have two main advantages. First, the solution in $\mathbf{s}_{i,t}$ to Problem $SLP_{i,t}^{\rho}$, defined by the solution function $\mathbf{s}_{i,t}^{\rho*}(\boldsymbol{\beta})$, can be defined in analytical closed forms up to Lagrange multiplier vectors that are fairly easy to compute. Second, solution function $\mathbf{s}_{i,t}^{\rho*}(\boldsymbol{\beta})$ is particularly "smooth" in its parameters: it is continuously differentiable "at will" in $\boldsymbol{\beta}$.

Based on the Lagrangian function:

$$\begin{aligned}\mathcal{L}_{i,t}(\mathbf{s}_{i,t},\boldsymbol{\lambda}_{i,t},\boldsymbol{\mu}_{i,t};\boldsymbol{\beta}) &= \sum_{m\in\mathcal{K}_{i,t-1}}\sum_{k\in\mathcal{K}_{i,t}} s_{mk,i,t}\pi_{mk,i,t}(\boldsymbol{\beta})-\kappa^{\rho}(\mathbf{s}_{i,t})\\ &\quad+\sum_{k\in\mathcal{K}_{i,t}}\lambda_k\left(a_{k,i,t}-\sum_{m\in\mathcal{K}_{i,t-1}} s_{mk,i,t}\right)+\sum_{m\in\mathcal{K}_{i,t-1}}\mu_m\left(a_{m,i,t-1}-\sum_{k\in\mathcal{K}_{i,t}} s_{mk,i,t}\right)\end{aligned}$$

(17)

$\mathbf{s}_{i,t}^{\rho*}(\boldsymbol{\beta})$ can be characterized by:

$$s_{mk,i,t}^{\rho*}(\boldsymbol{\beta})=s_{mk,i,t}^{\rho}\left(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t}^{\rho*}(\boldsymbol{\beta})\right)\ \text{for } (m,k)\in\mathcal{K}_{i,t-1}\times\mathcal{K}_{i,t} \tag{18}$$

where:

$$s_{mk,i,t}^{\rho}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t}) = a_{m,i,t-1}\frac{\exp(\pi_{mk,i,t}(\boldsymbol{\beta})-\lambda_{k,i,t})^{\rho}}{\sum_{\ell\in\mathcal{K}_{i,t}}\exp(\pi_{m\ell,i,t}(\boldsymbol{\beta})-\lambda_{\ell,i,t})^{\rho}} \tag{19}$$

and $\boldsymbol{\lambda}_{i,t}^{\rho*}(\boldsymbol{\beta})$ solves in $\boldsymbol{\lambda}_{i,t}$ the following crop rotation constraint subsystem:

$$\left\{\sum_{m\in\mathcal{K}_{i,t-1}} s_{mk,i,t}^{\rho}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t}) = a_{k,i,t},\ k\in\mathcal{K}_{i,t}\right. . \tag{20}$$

The second term defining $s_{mk,i,t}^{\rho}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})$ is a Logit acreage share function corresponding to the share of preceding crop *m* allocated to current crop *k*, $a_{m,i,t-1}^{-1}s_{mk,i,t}$. The redundancy of crop rotation constraint implies that an element of Lagrange multipliers $\boldsymbol{\lambda}_{i,t}$ needs to be fixed by a normalization condition[12].

The crop sequence acreage share functions $s_{mk,i,t}^{\rho}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})$ being continuously differentiable "at will" in $(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})$, the Implicit Function Theorem and the crop rotation constraints characterizing $\boldsymbol{\lambda}_{i,t}^{\rho*}(\boldsymbol{\beta})$ allow to compute the derivatives in $\boldsymbol{\beta}$ of $\boldsymbol{\lambda}_{i,t}^{\rho*}(\boldsymbol{\beta})$ and $\mathbf{s}_{i,t}^{\rho*}(\boldsymbol{\beta})$. These are useful for obtaining comparative statics results, for implementing gradient based optimization algorithms (Rust's NFX algorithm in particular) or for deriving the asymptotic distribution of the estimators of $\boldsymbol{\beta}_0$ presented in the next section.

3. Estimating crop rotation effects as solutions to a "smooth" MPEC problem

We can now combine the elements presented above for devising estimators of the crop rotation parameters $\boldsymbol{\beta}_0$ as solutions to well-defined MPEC problems. Vectors $\boldsymbol{\lambda}_i$ is obtained by piling up vectors $\boldsymbol{\lambda}_{i,t}$ for $t\in\mathcal{T}$ and vector $\boldsymbol{\lambda}$ is obtained by piling up vectors $\boldsymbol{\lambda}_i$ for $i\in I$. Let define the matrix functions $\mathbf{S}_{a,i}^{\rho}(\boldsymbol{\beta},\boldsymbol{\lambda}_i)$ from the $a_{k,i,t}^{-1}s_{mk,i,t}^{\rho}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})$ functions in the same way as matrices

[12] In our empirical application, the Lagrange multiplier associated to wheat, which is always produced by farmers, is fixed to zero.

$\mathbf{S}^{*}_{a,i}$ are defined from the $a^{-1}_{k,i,t}s^{*}_{mk,i,t}$ terms. A relevant estimator of $\boldsymbol{\beta}_0$ can be defined as the solution in $\boldsymbol{\beta}$ to the following MPEC problem:

$$\min_{(\boldsymbol{\delta},\boldsymbol{\beta},\boldsymbol{\lambda})}\left\{\begin{array}{c}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{u}'_{i,t}\left(\overline{\mathbf{y}}_i,\mathbf{S}^{\rho}_{a,i}(\boldsymbol{\beta},\boldsymbol{\lambda}_i);\boldsymbol{\delta},\boldsymbol{\beta}\right)(\mathbf{R}_{i,t}\hat{\boldsymbol{\Omega}}\mathbf{R}'_{i,t})^{-1}\mathbf{u}_{i,t}\left(\overline{\mathbf{y}}_i,\mathbf{S}^{\rho}_{a,i}(\boldsymbol{\beta},\boldsymbol{\lambda}_i);\boldsymbol{\delta},\boldsymbol{\beta}\right)\\ \text{s.t.}\\ \sum_{m\in\mathcal{K}_{i,t-1}}s^{\rho}_{mk,i,t}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})=a_{k,i,t},\quad k\in\mathcal{K}^{-r}_{i,t},\ (i,t)\in I\times\mathcal{T}\\ \lambda_{r,i,t}=0,\ k\in\mathcal{K}^{-r}_{i,t},\ \ (i,t)\in I\times\mathcal{T}\\ \sum_{t\in\mathcal{T}}\boldsymbol{\delta}_t=0\end{array}\right\}. \quad (21)$$

This MPEC problem involves a standard FLGS estimation problem and "smooth" equilibrium constraints. Crop subset $\mathcal{K}^{-r}_{i,t}$ is defined as the set of crops produced by farmer *i* in year *t*, crop *r* excepted. Conditions $k\in\mathcal{K}^{-r}_{i,t}$ and $\lambda^{r}_{i,t}=0$ indicate that crop *r* is used as the reference crop in the "reconstruction process" of the crop sequence acreage shares.

When the estimation criteria is based on standard fixed effects estimators, the corresponding MPEC problem is simply defined by:

$$\min_{(\boldsymbol{\delta},\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\alpha})}\left\{\begin{array}{c}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{e}'_{i,t}(\boldsymbol{\delta}_t,\boldsymbol{\alpha}_i,\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})(\mathbf{R}_{i,t}\hat{\boldsymbol{\Omega}}\mathbf{R}'_{i,t})^{-1}\mathbf{e}_{i,t}(\boldsymbol{\delta}_t,\boldsymbol{\alpha}_i,\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})\\ \text{s.t.}\\ \mathbf{e}_{i,t}(\boldsymbol{\delta}_t,\boldsymbol{\alpha}_i,\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})=\overline{\mathbf{y}}_{i,t}-\mathbf{Q}^{\delta}_{i,t}\boldsymbol{\delta}_t-\mathbf{Q}^{\alpha}_{i,t}\boldsymbol{\alpha}_i-\mathbf{S}^{\rho}_{a,i,t}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})\boldsymbol{\beta},\ \ (i,t)\in I\times\mathcal{T}\\ \sum_{m\in\mathcal{K}_{i,t-1}}s^{\rho}_{mk,i,t}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})=a_{k,i,t},\ \ k\in\mathcal{K}^{-r}_{i,t},\ \ (i,t)\in I\times\mathcal{T}\\ \lambda_{r,i,t}=0,\ \ k\in\mathcal{K}^{-r}_{i,t},\ \ (i,t)\in I\times\mathcal{T}\\ \sum_{t\in\mathcal{T}}\boldsymbol{\delta}_t=\mathbf{0}\\ \sum_{i\in I(k)}\boldsymbol{\alpha}_{k,i}=\mathbf{0},\quad k\in\mathcal{K}\end{array}\right\}. \quad (22)$$

Solving such MPEC problems can rely either on constrained optimization algorithms or on Rust's NFX algorithm (Su and Judd, 2012). Given the structure of the considered MPEC problems – with numerous small implicit markets of differentiated land, one for each farm/year pair – the use of constrained optimization algorithms is warranted as long as these make use of inversion techniques specifically designed for sparse matrices (Su and Judd, 2012; Dubé *et al*, 2012).

4. Estimation issues and proposed solutions

As estimating crop rotation effects requires sufficient variations in the crop sequence acreages, the considered estimation problem is only meaningful with sufficiently large farm samples. Furthermore, the MPEC problem that we consider is a highly "non quadratic" constrained optimization problem with a relatively large vector of parameter of interest. This raises obvious computing time and practical operational research issues.[13] Then, besides these purely technical issues, our estimation approach raises different identification issues that we propose to address as follows.

Crop rotation effects versus crop production risks

Crop production processes are affected by numerous significant random events such as climatic conditions or pest and weed infestations. Also, technical efficiency levels significantly vary across farms whether due to farmers' skills or to specific constraints impacting farmers' choice sets. The magnitude of these unobserved effects is likely to exceed those of crop rotations. This implies that crop rotation effects need to be "extracted" from relatively noisy data. If panel data appear to be necessary, their information content may not suffice.

Still, we expect to be able to identify crop rotation effects because the stochastic events affecting yields with the highest magnitude are either climatic events (*e.g.*, drought, rainy springs) or driven by climatic events (*e.g*, disease outbreaks). Our data covering a small area,

[13] From a purely technical perspective, our main concerns are due to overflow issues related to the computation of the Logit acreage share functions $s^{\rho}_{mk,i,t}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t})$ when the perturbation parameter ρ is large. We proceed by starting with small ρ values that are gradually increased.

the effects of the main random events on crop yields or chemical input uses are expected to be largely captured by year effects.

Information set

The "exogenous" variations used to estimate the crop rotation parameters come from the variations in farmers' crop sequence acreages induced by variations in farmers' crop acreage choices and crop prices. These are, by construction, the main drivers of the optimal crop sequence acreage shares $s_{mk,i,t}^{\rho*}(\boldsymbol{\beta})$. Yet, while crop price substantially vary across years, input prices show very little variability. Production sets and crop acreages vary substantially across farms, even when considering specialized cropping systems and limited areas as in our case. Part of this between-farms variation in crop acreages is, however, "lost" for identifying the crop rotation effects in netput quantity models featuring correlated farm effects. Variations across years are more limited although they appear to be significant. For instance, less than 3% of the sampled farms stick to the same production set along the 6 years covered by our data.

We use expert knowledge information to determine suitable constraints aimed at initiating and guiding the MPEC problem solving process.[14] For instance, experts identified "strongly unwarranted" crop sequences. These were used for reducing the feasible crop sequence sets that were *a priori* defined by the $\mathcal{K}_{i,t-1} \times \mathcal{K}_{i,t}$ sets. We also gathered rankings of the expected crop rotation effects from consulted experts. These rankings concern the effects of the preceding crops on the netput quantities of a given crop. These are included as inequality constraints involving parameters β_{mk}^{j} for $m \in \mathcal{K}_{i,t-1}$ in the MPEC problems.

[14] As well as for checking the "agronomic" pertinence of our results.

Impacts of long run crop rotation effects

In this study, we are interested in short crop rotation effects, namely on the effects of preceding crops on netput quantities of current crops. Yet, long run crop rotation effects may also impact farmers' crop input uses and yield levels. We propose to control for these effects by using a measure of the diversification of farms' acreage defined as an average entropy index of acreages at the farm level:

$$d_{i,t} = t^{-1}\sum\nolimits_{\tau=1}^{t} h_{i,\tau-1} \text{ with } h_{i,t} = -\sum\nolimits_{k\in\mathcal{K}} a_{k,i,t} \ln a_{k,i,t} \tag{23}$$

with $a_{k,i,t} \ln a_{k,i,t} = 0$ if $a_{k,i,t} = 0$, by continuity extension of the entropy function. Acreage entropy $h_{i,t}$ increases in the number of grown crops and as the distribution of the crop acreages comes close to a uniform distribution.

Reference crop sequence acreages and cross-entropic perturbations

The European Commission requires farmers to report their crop acreages each year in order to receive EU subsidies. These mandatory reports are collected in the series of yearly Integrated Administration and Control System (IACS) datasets that are, at least partly, made available for research purposes. These data allow to recover, under reasonable assumptions, approximate farm crop sequence acreages (*e.g.*, Levavasseur and Martin, 2015; Levavasseur *et al*, 2015).

Although the IACS datasets necessarily contain information on the farms of our sample, our accountancy data and the IACS data cannot be matched directly because suitable identification keys are lacking.[15] Yet, the IACS data can be used to produce relevant information on crop sequence acreages in the area covered by our data. We consider here an approach where these data are used neither to replace nor to accurately estimate the crop sequence acreages that are unobserved in our data, but to determine "reference" crop sequence acreages.

[15] Mostly due to confidentiality issues.

More specifically, we intend to define a "synthetic" farm/year pair for each of the farm/year pair of our dataset based on the observed crop acreages reported in both data sets. The observed crop sequence acreages of the "synthetic" IACS farm/year pairs can then be introduced as "reference" crop sequence acreages in the entropic perturbation terms of the crop sequence acreage reconstruction problems, $\kappa^{\rho}(\mathbf{s}_{i,t})$. We expect these acreages to be sufficiently informative to "guide" the estimation process of the crop rotation effects.

Let $\hat{s}_{mk,i,t}$ define the crop sequence acreage share of the crop sequence (*m*,*k*) of the "synthetic" IACS farm/year pair matched with the year *t* observation of farm *i* of our sample. The crop sequence acreage share vector $\hat{\mathbf{s}}_{i,t}$, obtained by piling up terms $\hat{s}_{mk,i,t}$ for $(m,k)\in\mathcal{K}_{i,t-1}\times\mathcal{K}_{i,t}$, can be used for two purposes. In both cases, the crop sequence acreage share vector of synthetic IACS observation (*i*,*t*), $\hat{\mathbf{s}}_{i,t}$, is used as a, possibly inaccurate, estimate of the unobserved crop sequence acreage share vector of observation (*i*,*t*), $\mathbf{s}_{i,t}^{*}$.[16] First, $\hat{\mathbf{s}}_{i,t}$ can be used for defining the feasible crop sequence set underlying $\mathbf{s}_{i,t}^{*}$ or, at least, for investigating the content of this set. Second, $\hat{\mathbf{s}}_{i,t}$ can be used for defining "cross-entropic" perturbation terms, with:

$$\kappa^{\rho}(\mathbf{s}_{i,t}\,|\,\hat{\mathbf{s}}_{mk,i,t})=\rho^{-1}\sum\nolimits_{m\in\mathcal{K}_{i,t-1}}\sum\nolimits_{k\in\mathcal{K}_{i,t}}s_{mk,i,t}(\ln s_{mk,i,t}-\ln\hat{s}_{mk,i,t})\,. \tag{24}$$

[16] Note that the IACS crop sequence acreages $\hat{\mathbf{s}}_{i,t}$ and the reconstructed crop sequence acreages $\mathbf{s}_{i,t}^{\rho*}(\hat{\boldsymbol{\beta}})$ both measure with errors the unobserved farm crop sequence acreage acreages $\mathbf{s}_{i,t}^{*}$. In particular, the fact that $\mathbf{s}_{i,t}^{\rho*}(\hat{\boldsymbol{\beta}})$ is a proxy of $\mathbf{s}_{i,t}^{*}$ is ignored in our estimation approach. Double proxy techniques could be used for using $\hat{\mathbf{s}}_{i,t}$ and $\mathbf{s}_{i,t}^{\rho*}(\hat{\boldsymbol{\beta}})$ terms as proxies of $\mathbf{s}_{i,t}^{*}$ while accounting for their measurement errors (and, thus, for controlling attenuation estimation biases), at least under the assumption stating that the considered measurement errors are centered and exogenous.

Provided that terms $\hat{\mathbf{s}}_{i,t}$ and $\mathbf{s}_{i,t}$ have the properties of probability distributions on $\mathcal{K}_{i,t-1}\times\mathcal{K}_{i,t}$, function $\rho\kappa^{\rho}(\mathbf{s}_{i,t}\,|\,\hat{\mathbf{s}}_{mk,i,t})$ can be interpreted as Kullback-Leibler's divergence of $\mathbf{s}_{i,t}$ from $\hat{\mathbf{s}}_{i,t}$.

Adding $-\kappa^{\rho}(\mathbf{s}_{i,t}\,|\,\hat{\mathbf{s}}_{mk,i,t})$ to the objective function of Problem $LP_{i,t}$ consists of including a penalty term that decreases from 0 as $\mathbf{s}_{i,t}$ moves away from $\hat{\mathbf{s}}_{i,t}$.[17] Using these "cross-entropic" perturbation terms in the MPEC problem aims to keep, at least in the early stages of the search process, the "reconstructed" crop sequence acreages within reasonable ranges according to the information provided by the IACS data. These ranges are "centered" at $\hat{\mathbf{s}}_{i,t}$ and their "width" is defined by the perturbation parameter value, ρ.

The corresponding optimal crop sequence acreage shares is given by simple expressions, with:

$$s^{\rho}_{mk,i,t}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t}\,|\,\hat{\mathbf{s}}_{i,t})=a_{m,i,t-1}\frac{\hat{s}_{mk,i,t}\exp(\pi_{mk,i,t}(\boldsymbol{\beta})-\lambda_{k,i,t})^{\rho}}{\sum_{\ell\in\mathcal{K}_{i,t}}\hat{s}_{m\ell,i,t}\exp(\pi_{m\ell,i,t}(\boldsymbol{\beta})-\lambda_{\ell,i,t})^{\rho}}. \tag{25}$$

The functional form of these optimal crop sequence acreage shares are based on the effects of the economic incentives described by the Logit terms, $\exp(\pi_{m\ell,i,t}(\boldsymbol{\beta})-\lambda_{\ell,i,t})^{\rho}$ for $\ell\in\mathcal{K}_{i,t}$, and on the "reference" crop acreage shares, $\hat{s}_{m\ell,i,t}$ for $\ell\in\mathcal{K}_{i,t}$.[18] The lower the value of the perturbation parameter ρ, the more the optimal acreage shares $s^{\rho}_{mk,i,t}(\boldsymbol{\beta},\boldsymbol{\lambda}_{i,t}\,|\,\hat{\mathbf{s}}_{i,t})$ stick to the "reference" acreage shares $\hat{s}_{mk,i,t}$. The larger ρ, the more economic incentives impact

---

[17] The "reference" crop sequence acreage shares implicitly used by the standard entropic perturbation are given by $\hat{s}_{mk,i,t}=K^{-1}_{i,t}K^{-1}_{i,t-1}$ for $(m,k)\in\mathcal{K}_{i,t-1}\times\mathcal{K}_{i,t}$.

[18] Equation $\sum_{\ell\in\mathcal{K}_{i,t}}\hat{s}_{m\ell,i,t}=a_{m,i,t-1}$ holds by construction. Carpentier and Letort (2014) also considered "reference" crop acreages are also used by for defining specific – entropy based – parametric functional forms of the implicit acreage management cost function considered in equation (13).

$s^{\rho}_{mk,i,t}(\boldsymbol{\beta}, \boldsymbol{\lambda}_{i,t} \mid \hat{\mathbf{s}}_{i,t})$ and the more these optimal crop sequence acreage shares can move away from the corresponding "reference" acreage shares.

5. Illustrative application

Farm accounting data

To illustrate the empirical tractability of our approach, we use an unbalanced panel dataset describing the production choices of 378 French arable crop producers located in the Marne *département*[19] over the years 2008 to 2014, each farm being observed from 4 to 7 years. This sample is extracted from a dataset obtained from CDER, an accountancy firm, and provides detailed information on crop production for each farm: acreages, yields, crop prices at the farm gate and cost accounting (*i.e.*, variable input uses at the crop level). We consider two aggregated chemical inputs: fertilizers and pesticides, for which price indices are constructed based on indices made available at the regional level by the French Department of Agriculture.

The sampled farms mostly produce crops selected from an eight crops set: wheat, barley, grain rapeseed, maize, protein pea, alfalfa, sugar beet, and potato.[20]

Table 1. Summary statistics, 2008 – 2014 averages

	wheat	barley	rape-seed	maize	protein pea	alfalfa	sugar beet	potato
Average yield (t/ha)	8.56	7.09	3.88	9.17	4.81	12.19	93.3	50.8
Average fertilizer use (€2010/ha)	237	183	234	195	84	267	295	340
Average pesticide use (€2010/ha)	181	106	206	107	153	60	262	667
Average gross margin (€/ha)	1353	1173	1383	1173	1021	1019	2434	5455
Average acreage share when produced	0.411	0.163	0.160	0.119	0.063	0.108	0.143	0.096
Average acreage share	0.411	0.147	0.150	0.052	0.023	0.075	0.116	0.026
Production frequency (%)	100%	90%	94%	44%	36%	69%	81%	27%

The standard summary statistics displayed in Table 1 reveal that the sampled farmers' always

[19] A *département* is a relatively small French territorial division (8162 km^2 for the Marne *département*).
[20] These selection criteria are not as stringent as they may appear. During the considered period the acreages of the considered crops cover more than 90% of the arable land devoted to crops in the considered area.

produce wheat and almost always produce barley, rapeseed and, albeit to a lesser extent, sugar beet. Around 69% produce alfalfa thanks to the outlet provided by local factories that dry alfalfa for feed. Less than half of the sampled farmers produce peas, maize and potatoes.

The Marne *département* is among the most productive areas for arable crops in France, as shown by the average yield levels of wheat, nearly 9 t/ha, and of sugar beet, more than 90 t/ha, over the 2008-2014 period. Sugar beet is the most intensive crop in pesticide and fertilizer uses. Rapeseed is the second most intensive.

Potato is the most profitable crop, with an average gross margin around 5500€/ha, followed by sugar beet, with an average gross margin around 2400 €/ha. Peas and alfalfa are the least profitable ones, with gross margins around 1000 €/ha. These gross margin differences are important for crop sequence choices. If two crops compete for a given preceding crop because it is well suited for these crops, then the land area with these preceding crops is likely to be devoted to the crop with the largest gross margin.

Expert knowledge information

As explained in section 5, expert knowledge information are introduced as additional constraints in the MPEC problem in order to reduce the number of parameters to be estimated and the solution process.

These information have been collected from interviews with agricultural scientists. They take the form of strongly unwarranted crop sequences, which come to reduce the set of feasible crop sequences, and of rankings of the impacts of crop sequences on yields and input uses, which are introduced as inequality constraints in the MPEC problem.

Tables 2a-c provide an overview of these constraints. The grey cells in each table correspond to the crop sequences that, according to the experts, are (almost) never adopted in practice: for instance growing rapeseed after rapeseed or after corn. The figures reported in

Table 2a correspond to the ranking of the expected effects of each preceding crops on the current crops yields, 1 corresponding to the highest potential expected yield. Following these rankings, constraints on the β^{o}_{mk} parameters are thus introduced in our MPEC problem to ensure that, for instance, the yield of wheat grown after peas or alfalfa (rank 1) is at least as high as the yield of wheat after grown rapeseed (rank 2), which itself is at least as high as the yield of wheat after sugar beet or potato (rank 3), etc. Similarly, the figures reported in Table 2b (resp. Table 2c) correspond to the ranking the expected effects of preceding crops on fertilizer (resp. pesticide) uses and are used to define constraints on the β^{j}_{mk} parameters.

Table 2a. Rankings of expected yields based on expert knowledge

		Current crops							
		wheat	**barley**	**rape-seed**	**maize**	**protein pea**	**alfalfa**	**sugar beet**	**potato**
Preceding crops	**wheat**	5	2	2	1	1	1	1	1
	barley	6	3	2	1	1	2	1	1
	rapeseed	2	1		1	2		2	3
	maize	4	1		1	2		2	3
	protein pea	1	3	1	1				3
	alfalfa	1	3		1		1	3	3
	sugar beet	3	2		2	2		3	3
	potato	3	2					2	2

Interpretation: 1 = highest expect yield; 2 = second highest expected yield; etc

Table 2b. Rankings of expected fertilizer use based on expert knowledge

		Current crops							
		wheat	**barley**	**rape-seed**	**maize**	**protein pea**	**alfalfa**	**sugar beet**	**potato**
Preceding crops	**wheat**	1	1	1	1	1	1	1	3
	barley	1	1	1	1	1	1	1	3
	rapeseed	2	2		2	1		1	1
	maize	1	1		1	1		1	1
	protein pea	3	3	2	3				1
	alfalfa	3	3		3		2	2	1
	sugar beet	2	2		2	1		1	1
	potato	2	2					1	2

Interpretation: 1 = highest expect fertilizer use; 2 = second highest expected fertilizer use; etc

Table 2c. Rankings of expected pesticide use based on expert knowledge

		Current crops							
		wheat	**barley**	**rape-seed**	**maize**	**protein pea**	**alfalfa**	**sugar beet**	**potato**
Preceding crops	**wheat**	1	1	1	2	1	2	1	3
	barley	1	1	2	2	1	2	1	3
	rapeseed	2	2		2	1		1	1
	maize	2	1		1	1		1	1
	protein pea	2	2	2	2				1
	alfalfa	2	2		2		1	1	1
	sugar beet	2	2		2	1		1	1
	potato	2	2					1	2

Interpretation: 1 = highest expect pesticide use; 2 = second highest expected pesticide use; etc

IACS data

As explained above, IACS data are used to define farm/year "synthetic" crop sequence acreages, $\hat{s}_{mk,i,t}$, that are introduced as references in the entropic perturbation terms of the MPEC problem, as shown in equation (24), in order to guide the estimation process.

To define these synthetic crop sequence acreages, we start by computing average crop sequence acreages from the IACS data collected in the Marne *département* between 2008 and

2014[21]. These averages are reported in Table 3 and are in fact quite close to what can generally be observed at the plot level. For instance, 42% of wheat appears to be grown after rapeseed, 63% of barley after wheat and 62% of rapeseed after barley. Yet, the wheat-barley-rapeseed crop sequence is a standard crop rotation scheme in France (Meynard *et al* 2013). We can also notice that two third of alfalfa is grown after alfalfa, which is not surprising since alfalfa is a perennial crop which is generally grown for two or three consecutive years.

The average values in Table 3 cannot be used directly as synthetic crop sequence acreages at the farm/year level, notably because feasible crop sequence sets, $\mathcal{K}_{i,t-1} \times \mathcal{K}_{i,t}$, vary from one farm to another and from one year to another. We use instead the crop sequences with the shortest Euclidean distance to IACS averages in the sets of feasible crop sequences for each farm/year:

$$\hat{\mathbf{s}}_{i,t} = \arg\min \left\{ \begin{array}{c} \left(\hat{\mathbf{s}}_{i,t} - \mathbf{s}_{i,t}^{IACS}\right)' \left(\hat{\mathbf{s}}_{i,t} - \mathbf{s}_{i,t}^{IACS}\right) \\ \text{s.t.} \\ \sum_{m \in \mathcal{K}_{i,t-1}} \hat{s}_{mk,i,t} = a_{k,i,t},\ k \in \mathcal{K}_{i,t} \\ \sum_{k \in \mathcal{K}_{i,t}} \hat{s}_{mk,i,t} = a_{m,i,t-1},\ m \in \mathcal{K}_{i,t-1} \end{array} \right\},\quad (i,t) \in \mathcal{I} \times \mathcal{T}. \tag{26}$$

with $\mathbf{s}_{i,t}^{IACS}$ the IACS crop sequence averages.

Table 3. Average crop sequence acreage shares in the Marne *département* extracted from IACS data

		Current crops							
		wheat	**barley**	**rape-seed**	**maize**	**protein pea**	**alfalfa**	**sugar beet**	**potato**
Preceding crops	**wheat**	0.12	0.63	0.38	0.29	0.20	0.04	0.56	0.44
	barley	0.04	0.18	0.62	0.23	0.42	0.29	0.38	0.28
	rapeseed	0.42	0.02		0.01	0.01		0.01	0.02
	maize	0.08	0.05		0.45	0.03		0.01	0.02
	protein pea	0.07	0.01	0.01	0.01				0.01
	alfalfa	0.09	0.01		0.01		0.67	0.01	0.01
	sugar beet	0.16	0.10		0.01	0.03		0.03	0.05
	potato	0.03	0.01					0.01	0.17

[21] The RPG Explorer tool (Levavasseur et al., 2015) has been used to extract crop sequences from raw IACS data

Results

Table 4 reports the average estimated crop sequence acreage shares. These figures are rather similar to the IACS averages in Table 3, except that higher shares of crops are grown after wheat here. This is in fact not surprising given that wheat acreage shares are particularly large in our sample (see Table 1).

We can also notice from Table 4 that some crop sequences, that were not *a priori* identified as unwarranted or infeasible, appear to be rarely (if not never) used is our results: estimated acreage shares of barley after pea, pea after rapeseed or maize, sugar beet after alfalfa and potato after rapeseed, pea or alfalfa, are very close to zero. As a matter of fact, the crop rotation effects associated to these sequences cannot be identified (they are marked by a '-' in Tables 5a-c).

Table 4. Average estimated crop sequence acreage shares

		Current crops							
		wheat	**barley**	**rape-seed**	**maize**	**protein pea**	**alfalfa**	**sugar beet**	**potato**
Preceding crops	**wheat**	0.23	0.59	0.52	0.49	0.64	0.18	0.61	0.24
	barley	0.02	0.09	0.45	0.06	0.35	0.23	0.24	0.50
	rapeseed	0.35	0.05		0.02	0.00		0.02	0.00
	maize	0.05	0.05		0.41	0.00		0.01	0.01
	protein pea	0.04	0.00	0.04	0.01				0.00
	alfalfa	0.08	0.01		0.01		0.60	0.00	0.00
	sugar beet	0.18	0.18		0.01	0.01		0.12	0.07
	potato	0.04	0.04					0.00	0.18

The estimated effects of crop sequences on yields are reported in Table 5a. These effects have been computed as percentage change in yields compared to a reference preceding crop for each crop. Rapeseed here is considered as the reference crop for wheat and wheat as the reference crop for all other crops.

The ranking of these effects was generally expected. For instance, we find that growing cereals after cereals leads to lower yields than growing cereals after other crops: growing wheat after barley or wheat leads to a 3% decrease in yield compared to growing wheat after rapeseed, and growing barley after rapeseed, maize sugar beet or potato leads to a 6% increase in yields compared to growing barley after wheat or barley. Our results also reflect the fact that wheat is considered to be a good previous crop for sugar beet and that peas and alfalfa are good preceding crops for wheat.

Table 5a. Average estimated impacts of crop sequences on yields: percentage changes compared to 'reference' preceding crops

		Current crops							
		wheat	**barley**	**rape-seed**	**maize**	**protein pea**	**alfalfa**	**sugar beet**	**potato**
Preceding crops	**wheat**	-0.03	ref	ref	ref	ref	ref	ref	ref
	barley	-0.03	0.00	0.00	-0.21	-0.02	-0.12	0.00	0.10
	rapeseed	ref	0.06		-0.21	-		-0.06	-
	maize	-0.03	0.06		-0.10	-		-0.06	0.00
	protein pea	0.00	-	0.00	-0.19				-
	alfalfa	0.14	0.00		-0.23		-0.12	-	-
	sugar beet	-0.03	0.06		-0.24	-0.45		0.00	0.00
	potato	-0.03	0.06					-	0.00

Turning to the estimated impacts of crop sequences on fertilizer uses reported in Table 5b, we find that growing crops after peas or alfalfa leads to the lowest levels of nitrogen use, which is consistent with the fact that these two crops are legumes and thus tend to increase the nitrogen content in soils. Similarly, rootcrops like sugar beet and potatoes have positive impacts on soil structure which reduces the needs for nitrogen for the crops grown after them. On the contrary cereals like wheat and barley tend to absorb nitrogen in soil and thus generally require more use of fertilizer for the crops grown after them. These results thus seem consistent from an agronomic viewpoint.

Results on pesticide uses are more difficult to interpret, for instance, growing a crop after the same crop was expected to increase pesticide use because of the increased risk of pest infestations it generates. If our results conform to this idea for wheat and maize, this is not the case for sugar beet.

More importantly, the magnitude of some the estimated effects of crop rotation on pesticide uses is surprisingly high, particularly those regarding the use of pesticides for potato. This probably reflects an identification issue related to a lack of observations for these crop sequences. Potato is indeed the crop with the lowest acreages in our sample (see Table 1) and is estimated to be essentially grown after wheat or barley, leaving few observations for the other preceding crops. The same type of remarks applies for the sugar beet – peas crop sequence, for which the estimated effects on pesticide uses, but also on fertilizer uses are extremely larges.

Table 5b. Average estimated impacts of crop sequences on fertilizer use: percentage changes compared to 'reference' preceding crops

		Current crops							
		wheat	barley	rape-seed	maize	protein pea	alfalfa	sugar beet	potato
Preceding crops	wheat	0.00	ref	ref	ref	ref	ref	ref	ref
	barley	0.03	0.00	0.01	-0.03	0.41	-0.15	-0.06	-0.04
	rapeseed	ref	0.06		-0.10	-		-0.09	-
	maize	0.00	0.06		-0.10	-		-0.34	0.02
	protein pea	-0.08	-	-0.10	-0.10				-
	alfalfa	-0.08	0.00		-0.87		-0.15	-	-
	sugar beet	-0.08	0.06		-0.10	12.16		-0.11	0.02
	potato	-0.08	0.06					-	0.02

Table 5c. Average estimated impacts of crop sequences on pesticide use: percentage changes compared to 'reference' preceding crops

		Current crops							
		wheat	**barley**	**rape-seed**	**maize**	**protein pea**	**alfalfa**	**sugar beet**	**potato**
Preceding crops	**wheat**	0.00	ref	ref	ref	ref	ref	ref	ref
	barley	0.00	-0.01	-0.04	0.06	0.07	-0.11	-0.15	3.34
	rapeseed	ref	-0.12		-0.46	-		-0.05	-
	maize	-0.11	-0.01		0.06	-		-0.48	11.41
	protein pea	-0.02	-	-0.06	-0.66				-
	alfalfa	0.00	-0.01		0.06		0.00	-	-
	sugar beet	0.00	-0.11		0.06	1.26		-0.19	3.34
	potato	0.00	-0.01					-	3.34

Conclusion

Crop rotations effects are key features of environmentally friendly crop production practices but their effects have rarely been quantified, probably due to the cost of suitable experiments necessary to collect these data. We propose here an original approach to estimate these effects based on farm accounting data. This approach allows to simultaneously estimate crop sequence acreage shares, which are unobserved in accounting data, and the effects of crop rotations on yields, fertilizer and pesticide uses. This approach relies on simple statistical models of yield and input uses integrated in a smooth MPEC problem. We also make use of expert knowledge information to reduce the size of this problem and guide its solving.

Our first results appear to be promising, although we still face some identification issues, especially regarding the effects of crop rotation on pesticide uses. These identification issues appear to be related to some crop sequences that are found to be rarely adopted by farmers. They may however also be due to the fact that crop rotations have so far been essentially used by farmers as a mean of managing the structure and nutrient content of soils, and not for controlling the populations of pest and weed, and thus play a more significant role in fertilizers than in pesticides uses decisions.

The information contained in IACS data has however not been fully exploited yet. Indeed, approximate crop sequence acreages at the farm level recovered from these data have been used as synthetic reference crop acreages, but farm level data couldn't be used directly because of a lack of suitable identification keys to match these data with our farm accounting data. An indirect matching approach, based on crop acreage variables common to both datasets, could in fact be used instead. In case of perfect matching for a sufficient number of farms, IACS data on crop sequence acreages could then be used to compare the outcome of our approach with and without observed crop sequence acreages. And, even in case of an imperfect matching, we could benefit from better references of crop sequences.

References

Berry, S., J. Levinsohn and A. Pakes (2004). Differentiated products demand systems from combination of micro and macro data: the new car market. *Journal of Political Economy* 112(1):68–105.

Berry, S., J. Levinsohn and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Carpentier, A., F. Féménia, A. Gohin and J. Helming (2018). Final report of the specific contract 154208.X29 “Lot 2: *Feasibility study on modelling crop rotation effects*” of the AgEconEurope framework contract 154208-2014-A08-NL.

Carpentier, A. and A. Gohin (2015). On the economic theory of crop rotations: value of the crop rotation effects and implications on acreage choice modeling. *Working Paper SMART – LERECO* N°15-04.

Carpentier, A. and A. Gohin (2014). Dynamic acreage choices. Selected paper, 14th *Triennal Congress of the European Association of Agricultural Economists*. Ljubjana, Slovenia, August 2014.

Carpentier, A. and A. Gohin (2011). Investigating acreage choices accounting for crop rotation effects and constraints. *Annual Meeting of the Agricultural and Applied Economics Association*, Pittsburgh, PA, July 2011.

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

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

Carrouée, B., A. Schneider A., F. Flénet, M.-H. Jeuffroy et T. Nemecek. (2012). Introduction du pois protéagineux dans des rotations à base de céréales à paille et colza : impacts sur les performances économiques et environnementales. *Innovations Agronomiques* 25:125-142.

Dubé, J.-P., J. T. Fox and C.-L. Su (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica* 80(5):2231-2267.

Fang S.-C., J. R. Rajasekera and H.-S. J. Tsao (1997). *Entropy optimization and mathematical programming*. Kluwer Academic Publishers.

Féménia, F., A. Carpentier, A. Gohin and I. R. Sodjahin (2018). Estimating crop rotation effects with farm accountancy data. *Triennal Conference of the International Association of Agricultural Economics*, Vancouver, BC, July/August.

Levavasseur, F. and P Martin (2015). *Analyse du Registre Parcellaire Graphique avec RPG Explorer. Exemples d'application sur des Aires d'Alimentation de Captage*. Rapport du projet EMADEC pour l'ONEMA. AgroParisTech et INRA. 127p.

Levavasseur F., P. Martin and O. Scheurer (2015). *RPG Explorer Version 1.8. Notice d'utilisation*. AgroParisTech, INRA et Institut Polytechnique Lasalle. 113p.

Meynard, J.-M., A. Messéan, A. Charlier, F. Charrier, M. Farès, M. Le Bail and M.-B. Magrini (2013). *Freins et leviers à la diversification des cultures. Etude au niveau des exploitations agricoles et des filières*. Rapport d'Etude INRA, 226p.

Mundlak, Y. R. (1978). On the Pooling of Time Series and Cross Section data. *Econometrica* 46(1): 69-85.

Rockafellar R. T. (1999). Duality and optimality in multistage stochastic programming. *Annals of Operations Research* 85:1-19.

Rust, J. (1988). Maximum likelihood estimation of discrete control processes. *SIAM Journal on Control and Optimization* 26(5):1006-1023.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica* 55:999-1033.

Su, C.-L. and K. Judd (2012). Constrained Optimization Approaches to Estimation of Structural Models. *Econometrica* 80:2213–2230.