Using farm cost accounting data to estimate crop rotation effects:
An approach based on a Smooth Mathematical Program with Equilibrium Constraints

Ibirénoyé Sodjahin[1], Alain Carpentier[2], Fabienne Femenia[2], Alexandre Gohin[2]

[1] UMR GAEL, INRAE - Université Grenoble-Alpes
[2] UMR SMART, INRAE - L'Institut Agro Rennes-Angers

Abstract

Crop rotation are key features of environmentally friendly cropping systems. Yet crop sequence acreages, which are required for estimating crop rotation effects, are rarely recorded in farm-level datasets such as the farm accounting panel datasets generally used to estimate economic models of farmers' crop production decisions. We propose here an original estimation procedure aimed at estimating crop rotation effects on yield and chemical input uses, while simultaneously reconstructing farms' unobserved crop sequence acreages. This approach relies on a bi-level optimization problem which is solved by using a Mathematical Program with Equilibrium Constraints (MPEC) approach. Simulated data, based on real observations of farms' production decisions for four crops are used to assess the performance of our approach. The results obtained on these simulated data demonstrate the ability of our proposed approach to simultaneously recover crop sequence acreages and estimate crop rotation effects on crop yield and input uses.

Keywords: crop rotation effects, crop sequence, mathematical programming with equilibrium constraints (MPEC), simulation.

1 Introduction

Agri-environmental policies are now assessed based on environmental criteria and are increasingly designed by referring to environmentally friendly crop production practices. Crop diversification has notably been one of the main objectives of the Common Agricultural Policy (CAP) since the 2013 reform, which introduced a set of crop diversification obligations as eligibility criteria for farmers to receive green direct payments. The new CAP, implemented in 2023, includes similar standards on agricultural crop acreages as part of its cross-compliance greening system and incorporates new conditionality rules to, for instance, encourage farmers to allocate significant share of their land to crops with environmental benefits such as nitrogen-fixing crops. Crop rotations effects thus appear to be key elements to consider in economic models aimed at simulating or evaluating agri-environmental policies. Yet, the use and the effects of crop rotations are poorly documented (Meynard et al., 2013). Farmers' crop sequence acreages are in fact rarely recorded, and crop rotation effects on yields and input uses are mostly measured based on experimental data and only for a few major crop pairs. In the agricultural production economics literature, the first models describing farmers' crop rotation choices were mathematical programming models. Until the mid-1980s, these models adopted a normative view and focused on the optimization of the use of crop rotation effects in farmers' crop sequence acreages choices through the specification of linear programming (LP) problems. The first econometric models aimed at accounting for farmers' use of crop rotation effects were developed based on the dynamic acreage choice model proposed by Eckstein (1984) and have been employed since then (e.g. Orazem and Miranowski (1994); Vitale et al. (2009)). Their estimation requires standard farm accountancy data, but they models rely on very crude assumptions as regards to crop rotation effects. Thomas (2003) explicitly accounts for crop rotation (through nitrogen carry-over effects from previous crops) in a multi-crop model of land allocation and fertilizer application decisions, to derive an optimal nitrogen management. However, because he lacks data on crop sequence acreages, he tests alternative rotations schemes and cannot provide numeric measurements of the effects of each crop pair on nitrogen uses. Hennessy (2006) proposes a theoretical analysis of the effects of crop sequences on yield and optimal input use levels, thereby providing rule of thumb for choosing between monoculture and rotation; his analysis is however limited to two crops: corn and soybean. Estimating crop rotation effects on crop yields and input uses based on farm accountancy data would actually be relatively straightforward if farmers crop sequence acreages were observed in those data. Standard estimation approaches could be employed such as regressing farmers' yields and inputs uses per crop on crop sequence acreages. Unfortunately, accountancy data do not contain such information on farmers' crop sequence acreage choices. There is thus a need to devise an estimation approach allowing to estimate the effects of crop rotations without observing farmers' crop sequence choices, which is our main objective in this paper. Our approach relies on the work of Carpentier and Gohin (2015) who propose a theoretical model of optimal stationary crop sequence choices, based on a dynamic programming framework, which explicitly accounts for the effects of crop rotations on expected yields and variable input uses on farmers' acreage choices. Based on this theoretical framework, we aim at estimating crop rotation effects while simultaneously recovering crop sequence choices that are consistent with the crop acreages observed in the data and with the estimated crop rotation effects. Our estimation problem is in fact closely related to those considered by Rust (1987) for estimating dynamic discrete choice models, or Berry et al. (1995) for estimating differentiated good demand systems, or more recently Su and Judd (2012) who devise an approach which is under some conditions computationally faster than Rust (1987)'s approach. From a practical viewpoint, our estimation approach is based on interrelated structural models: *i* yield and chemical input demand models, and *ii* crop sequence acreage share models. Estimating these models consists of solving, in the parameters describing the effects of crop rotations on crop yields and input uses, a Bi-Level Programming (BLP) problem, which is a particular case of Mathematical Programs with Equilibrium Constraints [1]. The upper level problem consists of optimizing, in these crop rotation parameters, a statistical criterion based on the observed crop yield and chemical input use levels, and on the crop sequence acreages recovered at the lower level. The lower level recovers the crop sequence acreages assuming that they are optimally chosen by farmers. The rest of the paper is organized as follows. The proposed approach to estimate the crop rotation effects is presented in the next section. We then discuss empirical estimation issues and present results obtained on simulated data in order to assess the empirical performances of our approach. Finally, we conclude.

2 Theoretical framework for estimating crop rotation effects

2.1 Definitions and notation

Farm accountancy data generally used to estimate microeconomic models of farmers' production choices provide information on the crop production decisions (yields, input uses, acreage choices, ...) of each sampled farm ($i = 1, ..., N$) during several periods ($t = 1, ..., T_i$).

[1] see Luo et al. (1996) for a thorough description of MPEC problems

Let denote by $\mathscr{K} = \{1, ..., k, ..., K\}$ the set of crops farmers can produce, and by $\mathscr{J} = \{1, ..., j, ..., J\}$, the set of chemical inputs used by farmers. The total arable land area available to farmer i in period t is denoted by A_{it}, and the vector of acreages devoted to each crop by $\mathbf{a}_{it} \equiv (\mathrm{a}_{k,it} : k \in \mathscr{K})$, such that $\sum_{k \in \mathscr{K}} \mathrm{a}_{k,it} = A_{it}$. Obtained yield levels are denoted by $(\mathrm{y}_{k,it} : k \in \mathscr{K})$, and the vector of chemical inputs quantities used for crop k by $\mathbf{x}_{k,it} \equiv (\mathrm{x}_{j,k,it} : j \in \mathscr{J})$. $\mathbf{q}_{k,it} \equiv (\mathrm{p}_{k,it}, -\mathbf{w}_{it})$ is a vector of netput prices, with $\mathrm{p}_{k,it}$ the output price of crop k and $\mathrm{w}_{j,it}$ the price of input j for farmer i in period t. Although we should bear in mind that farmers do not necessarily grow all the crops belonging to the crop set $\mathscr{K}$ every year, we take as a convention that the crops grown in period t are indexed by k while those grown in period $t-1$ are indexed by m. Crop pair (m, k) thus denotes a sequence of crops: crop k is grown in period t on plots where crop m was grown in period $t-1$.

The gross margin of crop k grown after crop m, *i.e.* the gross margin of crop sequence (m, k), is defined as $\pi_{mk,it} = \mathrm{y}_{mk,it}\mathrm{p}_{k,it} - \sum_j \mathrm{x}_{j,mk,it}\mathrm{w}_{j,it}$, with $\mathrm{y}_{mk,it}$ the yield of crop k grown after crop m, and $\mathrm{x}_{j,mk,it}$ the level of chemical input j used for crop k grown after crop m . $\pi_{mk,it}$ provides a measure of the economic benefits of the considered crop sequence. Both $\mathrm{y}_{mk,it}$ and $\mathrm{x}_{j,mk,it}$ are unobserved quantities, because the crop m which precedes k is not recorded in the data.

Let denote $s_{mk,it}$ the acreage of crop k grown, by farmer i in year t, on land with previous crop m . The share of acreage of crop k grown on land with previous crop m by farmer i in year t, is denoted by $\mathrm{z}_{mk,it}$, and writes

$$\mathrm{z}_{mk,it} = \begin{cases} \dfrac{s_{mk,it}}{\mathrm{a}_{k,it}} & \text{if } \mathrm{a}_{k,it} > 0 \\ 0 & \text{if } \mathrm{a}_{k,it} = 0 \end{cases} \quad \text{with} \sum_m \mathrm{z}_{mk,it} = 1$$

We define pre-crop effects as differences with respect to a reference previous crop. For instance, considering crop $r(k)$ as the reference previous crop for crop k, the pre-crop effect of previous crop m on the yield of crop k writes $\theta^{(y)}_{mk,it} = \mathrm{y}_{mk,it} - \mathrm{y}_{r(k)k,it}$, and that of previous crop m on the use of chemical inputs of crop k writes $\theta^{(x)}_{j,mk,it} = \mathrm{x}_{j,mk,it} - \mathrm{x}_{j,r(k)k,it}$.

2.2 Assumptions and models

We assume that farmers choose their crop acreages by considering the effects of previous crops on the production process (yields and input uses) of the current crops. They are thus assumed to choose their crop sequence acreages $s_{mk,it}$, *i.e.* to choose, in year t, the acreage of crop k produced on land where crop m was produced in year $t-1$. Because farmers cannot use more or less acreage of a previous crop than the crop acreages defined by their previous year crop acreage choices, their crop sequence acreage choices are necessarily constrained. These crop rotation constraints state that the demand for land with a given previous crop m equals the past acreage of this crop:

$$(1) \quad \sum_{k \in \mathcal{K}} s_{mk,it} = \mathrm{a}_{m,it-1}$$

We assume that the observed crop yields $\mathrm{y}_{k,it}$, and observed chemical input uses $\mathrm{x}_{j,k,it}$ are respectively equal to the sum of the corresponding crop sequence yields $\mathrm{y}_{mk,it}$ and to the sum of the corresponding crop sequence input uses $\mathrm{x}_{j,mk,it}$, weighted by corresponding crop sequence acreage shares $\mathrm{z}_{mk,it}$:

$$(2) \quad \begin{cases} \mathrm{y}_{k,it} & = \sum_{m \in \mathcal{K}} \mathrm{z}_{mk,it} \mathrm{y}_{mk,it} \\ \mathrm{x}_{j,k,it} & = \sum_{m \in \mathcal{K}} \mathrm{z}_{mk,it} \mathrm{x}_{j,mk,it} \end{cases} \text{ for } k \in \mathcal{K} \text{ and } j \in \mathcal{J}$$

Pre-crop effects on yields and input uses could easily be estimated from equation system (2) if crop sequence acreage shares were observed in farm-level data. This is unfortunately not the case. As a result, modelling assumptions need to be imposed to allow the estimation of pre-crop effects, $\left(\theta^{(\mathrm{y})}_{mk,it}, \theta^{(\mathrm{x})}_{j,mk,it} : j \in \mathcal{J}\right)$, while simultaneously recovering crop sequence acreage shares, $\mathrm{z}_{mk,it}$.

First, we assume that unoserved crop sequence yields and chemical input uses can be decomposed as:

$$(3) \quad \begin{cases} \mathrm{y}_{mk,it} & = \alpha^{(\mathrm{y})}_{k,t,0} + \theta^{(\mathrm{y})}_{mk,0} + e^{(\mathrm{y})}_{k,it} \\ \mathrm{x}_{j,mk,it} & = \alpha^{(\mathrm{x})}_{j,k,t,0} + \theta^{(\mathrm{x})}_{j,mk,0} + e^{(\mathrm{x})}_{j,k,it} \end{cases} \text{ with } \mathbb{E}[e^{(\mathrm{y})}_{k,it}] = \mathbb{E}[e^{(\mathrm{x})}_{j,k,it}] = 0 \text{ for } k \in \mathcal{K} \text{ and } j \in \mathcal{J}$$

where the $\left(\alpha^{(\mathrm{y})}_{k,t,0}, \alpha^{(\mathrm{x})}_{j,k,t,0} : j \in \mathcal{J}\right)$ terms are year-specific terms aimed at capturing the effects of economic and weather shocks affecting all farms simultaneously, the $\left(\theta^{(\mathrm{y})}_{mk,0}, \theta^{(\mathrm{x})}_{j,mk,0}\right)$ terms are pre-crop effects assumed to be common to all farms, and $\mathbf{e}_{k,it} = \left(e^{(\mathrm{y})}_{k,it}, e^{(\mathrm{x})}_{j,k,it} : j \in \mathcal{J}\right)$ are random terms. Replacing in equations (2) $\mathrm{y}_{mk,it}$ and $\mathrm{x}_{j,mk,it}$ by their corresponding expression in equations (3) yields the following equations:

$$(4) \quad \begin{cases} \mathrm{y}_{k,it} & = \alpha^{(\mathrm{y})}_{k,t,0} + \mathbf{z}'_{k,it} \theta^{(\mathrm{y})}_{k,0} + e^{(\mathrm{y})}_{k,it} \\ \mathrm{x}_{j,k,it} & = \alpha^{(\mathrm{x})}_{j,k,t,0} + \mathbf{z}'_{k,it} \theta^{(\mathrm{x})}_{j,k,0} + e^{(\mathrm{x})}_{j,k,it} \end{cases}$$

with $\theta^{(\mathrm{x})}_{r(k)k,0} = 0$, $\forall k \in \mathcal{K}$, $\theta^{(\mathrm{x})}_{r(k)k,0} = 0$, $\forall (k,j) \in \mathcal{K} \times \mathcal{J}$, and $\mathbb{E}[e^{(\mathrm{y})}_{k,it}] = \mathbb{E}[e^{(\mathrm{x})}_{j,k,it}] = 0$.

Pre-crop effects $\theta_{k,0} = \left(\theta^{(\mathrm{y})}_{k,0}, \theta^{(\mathrm{x})}_{j,k,0} : j \in \mathcal{J}\right)$ and crop sequence acreage share vectors $\mathbf{z}_{k,it} =$

$\left(z_{mk,it} : m \in \mathcal{K}\right)$ are unobserved.

We assume that error terms $\mathbf{e}_{k,it}$ affecting the levels of yield/chemical inputs uses do not depend on crop sequence acreage shares $\mathbf{z}_{k,it}$, or otherwise stated, that $\mathbf{z}_{k,it}$ is exogenous with respect to the error term $\mathbf{e}_{k,it}$. This assumptions it admittedly restrictive but relaxing it would significantly increase the complexity of an estimation problem that is already challenging. Furthermore, it can be partially justified by the fact that the acreage choice decisions (and thus, crop sequence acreage choice decisions) are made before harvest, and although disturbances affecting input uses can correlate with crop sequence choices, this correlation is likely to be weak as input applications are spread throughout the cropping season.

To circumvent the fact that the crop sequence acreage share vectors $\mathbf{z}_{it} = \left(\mathbf{z}_{k,it} : k \in \mathcal{K}\right)$ are unobserved in our data, we rely on a standard economic rationality assumption. As shown by Carpentier and Gohin (2015), under this rationality assumption, crop sequence acreage shares can be defined as a solution to a linear programming (LP) problem stating that forward-looking farmers choose their optimal crop sequence acreages so as to maximize their expected returns at the farm level under specific constraint on acreages. This allows devising an approach to recover the unobserved crop sequence acreage choices of a farmer based on his observed crop acreage choices $(\mathbf{a}_{it}, \mathbf{a}_{it-1})$, with $\mathbf{a}_{it} = \left(a_{k,it} : k \in \mathcal{K}\right)$ and $\mathbf{a}_{it-1} = \left(a_{m,it-1} : m \in \mathcal{K}\right)$, and given estimates of pre-crop effects $\left(\theta^{(y)}_{k,0}, \theta^{(x)}_{j,k,0} : j \in \mathcal{J}\right)$.

We define the following set of constraints:

Unfeasibility constraints $\mathbf{IMP}_{it}(\hat{\mathbf{z}}_{it}) = 0$: a crop sequence (m, k) is unfeasible if either $a_{k,it} = 0$ or $a_{m,it-1} = 0$. In that case, $z_{mk,it} = 0$.

Acreage share constraints $\mathbf{SHA}_{it}(\hat{\mathbf{z}}_{it}) \geq 0$: the terms $z_{mk,it}$ being defined as the share previous crop m in the acreage of current crop k, the elements of $\hat{\mathbf{z}}_{it}$ must be non-negative: $z_{mk,it} \geq 0 \forall (m, k) \in \mathcal{K} \times \mathcal{K}$, and sum to 1: $\sum_{m\in\mathcal{K}} z_{mk,it} = 1 \forall k \in \mathcal{K}$.

Crop rotation constraints $\mathbf{ROT}_{it}(\hat{\mathbf{z}}_{it}) = 0$: the supply of previous crop acreage, $a_{m,it-1}$ must equal the demand of previous crop m for current crops k, $\sum_{k\in\mathcal{K}} s_{mk,it} = a_{m,it-1}$, in order to ensure a supply-demand equilibrium $\sum_{k\in\mathcal{K}} a_{k,it} z_{mk,it} = a_{m,it-1}$ for $m \in \mathcal{K}$.

Let consider a farmer i who has, in period t, to allocate her/his previous crop acreage $\mathbf{a}_{i,t-1}$ to her/his current crop acreage $\mathbf{a}_{i,t}$. The expected profit of this farmer is given by

$$\Pi_{it} = \sum_{k\in\mathcal{K}} a_{k,it} \sum_{m\in\mathcal{K}} \hat{z}_{mk,it} \pi_{mk,it}$$

where his crop sequence acreage share choices are given by the $\hat{z}_{mk,it}$ terms and the expected profit of a unit of land of crop k with previous crop m is given by

$\pi_{mk,it} = \boldsymbol{\alpha}'_{k,t}\mathbf{q}_{k,it} + \boldsymbol{\theta}'_{mk}\mathbf{q}_{k,it}$ with $\boldsymbol{\alpha}_{k,t} = \left(\alpha^{(y)}_{k,t}, \alpha^{(x)}_{j,k,t} : j \in \mathcal{J}\right)$.

Using this definition of $\pi_{mk,it}$ allows rewriting the expected profit of farmer i in year t as

$$\Pi_{it} = \underbrace{\sum_{k\in\mathcal{K}} a_{k,it}\left(\boldsymbol{\alpha}'_{k,t}\mathbf{q}_{k,it}\right)}_{\text{does not depend on crop sequence acreage choices}} + \underbrace{\sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{K}} a_{k,it}\hat{z}_{mk,it}\left(\boldsymbol{\theta}'_{mk}\mathbf{q}_{k,it}\right)}_{\text{objective function to maximize to choose crop sequence acreages}}$$

The expected return maximization problem of farmer i in period t, denoted $\mathbf{LP}_{it}$, can thus be defined as follows:

$$\text{(5)} \qquad \text{Problem } \mathbf{LP}_{it} : \max_{\hat{\mathbf{z}}_{it}} \left\{ \begin{array}{l} \sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{K}} \hat{z}_{mk,it}a_{k,it}\left(\boldsymbol{\theta}'_{mk}\mathbf{q}_{k,it}\right) \\ \text{s.t.} \\ \mathbf{SHA}_{it}(\hat{\mathbf{z}}_{it}) \geq 0, \mathbf{ROT}_{it}(\hat{\mathbf{z}}_{it}) = 0, \mathbf{IMP}_{it}(\hat{\mathbf{z}}_{it}) = 0 \end{array} \right\}$$

where $\boldsymbol{\theta}_{mk} = \left(\theta^{(y)}_{mk}, \theta^{(x)}_{j,mk} : j \in \mathcal{J}\right)$. $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_{mk} : (m,k) \in \mathcal{K}^2\right)$.

Yet, defining estimates of $\mathbf{z}_{it}$ as a solution to an optimization problem is appealing mostly from a theoretical viewpoint. In fact, the solutions in $\hat{\mathbf{z}}_{it}$ to problem $\mathbf{LP}_{it}$ may be multiple, implying that the solution to this problem is defined as a solution set $\mathcal{Z}_{it}(\boldsymbol{\theta})$. Condition $\hat{\mathbf{z}}_{it} \in \mathcal{Z}_{it}(\boldsymbol{\theta})$ simply states that the crop sequence acreage $\mathbf{z}_{it}$ is optimal for farmer i in year t given his current and previous crop acreages $(\mathbf{a}_{it}, \mathbf{a}_{it-1})$ and the crop rotation effects measured by $\boldsymbol{\theta}$. Moreover, even in cases where $\mathbf{z}_{it}$ is uniquely defined, it exhibits salient discontinuities in the parameter vector $\boldsymbol{\theta}$. These discontinuities underlie significant empirical issues arising when estimating $\boldsymbol{\theta}$.

Carpentier and Gohin (2014) showed that adding an entropic perturbation term to the objective function of problem $\mathbf{LP}_{it}$ allows alleviating these issues. They showed that the solution in $\hat{\mathbf{z}}_{it}$ to the following perturbed version of $\mathbf{LP}_{it}$ denoted $\mathbf{SmLP}^{\rho}_{it}$ to be particularly convenient from an empirical viewpoint:

$$\text{(6)} \qquad \text{Problem } \mathbf{SmLP}^{\rho}_{it} : \max_{\hat{\mathbf{z}}^{\rho}_{it}} \left\{ \begin{array}{l} \sum_{k\in\mathcal{K}}\sum_{m\in\mathcal{K}} \hat{z}_{mk,it}a_{k,it}\left(\boldsymbol{\theta}'_{mk}\mathbf{q}_{k,it}\right) - \varrho_{it}(\hat{\mathbf{z}}_{it}, \rho) \\ \text{s.t.} \\ \mathbf{SHA}_{it}(\hat{\mathbf{z}}_{it}) \geq 0, \mathbf{ROT}_{it}(\hat{\mathbf{z}}_{it}) = 0, \mathbf{IMP}_{it}(\hat{\mathbf{z}}_{it}) = 0 \end{array} \right\}$$

where

$$\varrho_{it}\left(\hat{\mathbf{z}}_{it},\rho\right)=\rho^{-1}\sum_{k\in\mathcal{K}}\mathrm{a}_{k,it}\sum_{m\in\mathcal{K}}\hat{\mathrm{z}}_{mk,it}\ln\hat{\mathrm{z}}_{mk,it}$$

The solution to this optimization problem, denoted $\mathbf{z}_{it}^{\rho}$, is unique and converges to a point in $\mathcal{Z}_{it}(\theta)$ as the (positive) perturbation parameter ρ grows to infinity, indicating that $\mathbf{z}_{it}^{\rho}$ can be considered as a reliable approximate solution to problem $\mathbf{LP}_{it}$ when ρ is sufficiently large.

Carpentier and Gohin (2014) also showed that $\mathbf{z}_{it}^{\rho}$ can be defined as a particularly well-behaved function of $\left(\theta,\hat{\boldsymbol{\mu}}_{it};\rho\right)$. This function, denoted by $\mathbf{z}_{it}^{0}\left(\theta,\hat{\boldsymbol{\mu}}_{it};\rho\right)$, is defined in analytical closed form and smooth – *i.e.* continuously differentiable at will – in $\left(\theta,\hat{\boldsymbol{\mu}}_{it}\right)$ and writes:

$$\mathrm{z}_{mk,it}^{0}\left(\theta,\hat{\boldsymbol{\mu}}_{it};\rho\right)=\frac{\exp\left[\rho\left(\mathbf{q}_{k,it}^{'}\theta_{mk}-\mu_{m,it}\right)\right]}{\sum_{n\in\mathcal{K}}\exp\left[\rho\left(\mathbf{q}_{k,it}^{'}\theta_{nk}-\mu_{n,it}\right)\right]}$$

with $\hat{\boldsymbol{\mu}}_{it}$, vector of the Lagrange multipliers associated to the crop rotation constraints $\mathbf{ROT}_{it}\left(\hat{\mathbf{z}}_{it}\right)=0$.

This 'smoothness' property is particularly valuable when this function is used to construct objective functions of optimization problem to be solved in $\left(\theta,\hat{\boldsymbol{\mu}}_{it}\right)$ since it allows using standard gradient-based optimization algorithms.

The solution to problem $\mathbf{SmLP}_{it}^{\rho}$ is $\mathbf{z}_{it}^{\rho}=\mathbf{z}_{it}^{0}\left(\theta,\boldsymbol{\mu}_{it}^{\rho};\rho\right)$ with $\boldsymbol{\mu}_{it}^{\rho}$ the optimal value of the Lagrange multiplier vector associated to the crop rotation constraints. This result has a nice economic interpretation. Indeed, since these constraints are equilibrium constraints, ensuring the equality between the supply and demand of land devoted to the different previous crops, $\boldsymbol{\mu}_{it}^{\rho}$ can be interpreted as a vector of shadow prices of land with specific previous crops.

3 Smooth MPEC estimation approach

Our setup provides two main elements for devising an estimation approach of the effects of pre-crop on yields and input uses:

(a) statistical models describing how the observed crop netput quantities $\left(\mathrm{y}_{k,it},\mathrm{x}_{j,k,it}:j\in\mathcal{J}\right)$ depend on farmers' unobserved crop sequence acreage share choices $\mathbf{z}_{k,it}$, and on the crop rotation parameters θ_0 to be estimated;

(b) procedures to obtain crop sequence acreage share choices $\mathbf{z}_{it}$ consistent with the estimated values of the crop rotation parameters θ_0 and the observed crop acreage choices $(\mathbf{a}_{it},\mathbf{a}_{it-1})$, based the SmLP problem;

Our general estimation approach combine these components in the following theoretical bi-level programming estimation problem:

$$\text{Problem } \mathbf{U}: \min_{\theta} \left\{ \begin{array}{l} \sum_{k\in\mathcal{K}} \sum_{(i,t)\in S} \mathrm{e}^{(\mathrm{y})}_{k,it}\left(\mathbf{z}_{k,it};\theta^{(\mathrm{y})}_{k}\right)^2 + \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \sum_{(i,t)\in S} \mathrm{e}^{(\mathrm{x})}_{j,k,it}\left(\mathbf{z}_{k,it};\theta^{(\mathrm{x})}_{j,k}\right)^2 \\ \text{s.t.} \\ \mathrm{y}_{k,it} = \alpha^{(\mathrm{y})}_{k,t} + \mathbf{z}'_{k,it}\theta^{(\mathrm{y})}_{k} + \mathrm{e}^{(\mathrm{y})}_{k,it} \ \forall\ k \in \mathcal{K} \\ \mathrm{x}_{j,k,it} = \alpha^{(\mathrm{x})}_{j,k,t} + \mathbf{z}'_{k,it}\theta^{(\mathrm{x})}_{j,k} + \mathrm{e}^{(\mathrm{x})}_{j,k,it} \ \forall\ (k,j) \in \mathcal{K}\times\mathcal{J} \\ \mathbf{z}_{it} \in \mathcal{Z}_{it}(\theta), (i,t) \in S \end{array} \right\}$$

$$\text{Problem } \mathbf{L}_{it}: \mathcal{Z}_{it}(\theta) = \arg\max_{\hat{\mathbf{z}}_{it}} \left\{ \begin{array}{l} \sum_{k\in\mathcal{K}} \sum_{m\in\mathcal{K}} \hat{\mathrm{z}}_{mk,it}\mathrm{a}_{k,it}\left(\theta'_{mk}\mathbf{q}_{k,it}\right) \\ \text{s.t.} \\ \mathbf{SHA}_{it}(\hat{\mathbf{z}}_{it}) \geq 0, \mathbf{ROT}_{it}(\hat{\mathbf{z}}_{it}) = 0, \mathbf{IMP}_{it}(\hat{\mathbf{z}}_{it}) = 0 \end{array} \right\} \forall (i,t) \in S$$

Problem **U** is the upper level (master) problem. It seeks to minimize the sum of squared residuals of the considered set of netput quantity models with respect to the crop rotation parameter vector θ. This simple OLS criterion yields a consistent estimator of θ_0 under the assumptions underlying our netput quantity models. More sophisticated estimators could be used, such as SUR-type estimators but the related efficiency gain is likely to be limited.

Problems $\mathbf{L}_{it}$ are the lower level (subordinate) problems. They aim to deliver 'estimates' of the crop sequence acreage vectors $\mathbf{z}_{it}$ that are used as explanatory variable vectors in the upper level problem.

BLP problems cannot be solved directly according to their theoretical formulation. A BLP problem is generally solved by transforming it into a standard 'one-level' constrained optimization problem, according to the so-called 'mathematical program with equilibrium constraints' (MPEC) approach (Luo et al., 1996). In this approach, the lower level problems $\mathbf{LP}_{it}$ are replaced by the optimality conditions characterizing their solutions. These KKT conditions are the 'equilibrium constraints' involved in the resulting MPEC problem. In our case, adopting the MPEC approach consists of solving the following problem:

$$
\min_{\theta,\boldsymbol{\mu},\boldsymbol{\lambda}}
\begin{cases}
\sum_{k\in\mathcal{K}}\left(\sum_{(i,t)\in S}\hat{\epsilon}^{OLS}_{k,it}\left(\boldsymbol{\lambda}^{(y)}_{k},\theta^{(y)}_{k};\mathbf{z}^{\rho}_{it}\left(\theta,\boldsymbol{\mu}_{it};\rho\right)\right)^{2}\right)+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\left(\sum_{(i,t)\in S}\hat{\epsilon}^{OLS}_{j,k,it}\left(\boldsymbol{\lambda}^{(y)}_{j,k},\theta^{(y)}_{j,k};\mathbf{z}^{\rho}_{it}\left(\theta,\boldsymbol{\mu}_{it};\rho\right)\right)^{2}\right)\\
\text{s.t.}\\
\mathrm{y}_{k,it}=\alpha^{(y)}_{k,t}+\mathbf{z}'_{k,it}\theta^{(y)}_{k}+\epsilon^{OLS}_{k,it}\ \forall k\in\mathcal{K}\\
\mathrm{x}_{j,k,it}=\alpha^{(x)}_{j,k,t}+\mathbf{z}'_{k,it}\theta^{(x)}_{j,k}+\epsilon^{OLS}_{j,k,it}\ \forall(k,j)\in\mathcal{K}\times\mathcal{J}\\
\mathrm{z}^{\rho}_{mk,it}\left(\theta,\boldsymbol{\mu}_{it};\rho\right)=\dfrac{\exp\left[\rho\left(\mathbf{q}'_{k,it}\theta_{mk}-\mu_{m,it}\right)\right]}{\sum_{n\in\mathcal{K}}\exp\left[\rho\left(\mathbf{q}'_{k,it}\theta_{nk}-\mu_{n,it}\right)\right]}\ \forall(m,k)\in\mathcal{K}\times\mathcal{K},\forall \mathrm{a}_{m,it-1}>0,\mathrm{a}_{k,it}>0\\
\sum_{k\in\mathcal{K}}\mathrm{a}_{k,it}\mathrm{z}^{\rho}_{mk,it}\left(\theta,\boldsymbol{\mu}_{it};\rho\right)=\mathrm{a}_{m,it-1}\forall m\in\mathcal{K}\text{ and }\mathrm{a}_{m,it-1}>0\\
\theta^{(y)}_{r(k)k}=\theta^{(x)}_{j,r(k)k}=0\text{and }\mu_{r(k),it}=0\ \forall(k,j,i,t)\in\mathcal{K}\times\mathcal{J}\times S\\
\mathrm{z}^{\rho}_{mk,it}\left(\theta,\boldsymbol{\mu}_{it};\rho\right)=\theta^{(y)}_{mk}=\theta^{(x)}_{j,mk}=0\text{ if }(m,k)\text{ is unfeasible }\forall(j,i,t)\in\mathcal{J}\times S
\end{cases}
$$

Note that the acreage share constraints ($\mathbf{SHA}_{it}(\hat{\mathbf{z}}_{it})\geq 0$) are automatically fulfilled due to the shape of the expression of crop sequence acreage shares.

This problem is to be solved in the crop rotation parameter vector θ, crop sequence acreages $\hat{\mathbf{z}}=(\hat{\mathbf{z}}_{it})$ and crop rotation constraint Lagrange multiplier vectors $\hat{\boldsymbol{\mu}}=\left(\hat{\boldsymbol{\mu}}_{it}:(i,t)\in S\right)$ and $\hat{\boldsymbol{\lambda}}=\left(\hat{\boldsymbol{\lambda}}_{it}:(i,t)\in S\right)$, the Lagrange multiplier vector associated to the acreage share constraints ($\mathbf{SHA}_{it}(\hat{\mathbf{z}}_{it})\geq 0$).

Standard (gradient-based) algorithms perform well for solving this smooth nonlinear optimization problem.

The next step is to assess through simulations the empirical performance of our proposed smooth MPEC estimator. We consider two cases regarding the rationality assumption. In the first case, we consider that crop sequence acreages are entirely determined by crop rotation effects, and thus by the crop sequence gross margin. In a second case, we add some disturbances to the crop sequence gross margin to acknowledge that farmers' crop sequence acreage choices may be driven by motives other than profit maximization, *e.g.*, by environmental concerns.

4 Empirical performance of the smooth MPEC estimator

4.1 Simulating the data

In order to assess the empirical performance of this estimator, we proceed as follow. First, we create a set of simulated data in three steps:

i) We act as if we observed the crop rotation effects, our parameters of interest, so we set the population parameters of crop rotation effects;

ii) Then we obtain the crop sequence acreage share consistent with the population parameters by solving the $\mathbf{LP}_{it}$ problem;

iii) We simulate the observed netputs (yield and input uses) from the population parameters defined in step *i*) and the crop sequence acreages obtained in *ii*), and adding a noise with zero mean drawn from the normal distribution;

iv) Second, we run the smooth MPEC procedure by using the observed crop acreages, the simulated yield and chemical input uses of step *iii*), leading to estimates of the crop rotation effects and the crop sequence acreage shares.

The steps *iii*) and *iv*) are repeated B times, producing B estimates of crop rotation effects and crop sequence acreage shares as well. We can average over the B estimates and compare these averages with the corresponding values in the population parameters. Associated standard deviations are computed as well. If the average of the B estimates are close enough to the values of the population parameters, then our smooth MPEC procedure yields an unbiased estimate of the crop rotation effects.

In what follows, we restrict the crop set to four main crops:

$\mathcal{K} = \{$wheat, barley, rapeseed, sugar beet$\}$. We consider in our simulation a sample of observed data on crop acreages vector $\mathbf{a}_{it}$, output and input prices vector $\mathbf{q}_{k,it}$, to get a data structure on acreages and prices that has a meaning in the economic sense, and close to what is observed. This sample consists of 3685 observations describing the production choices of 722 farms.

We set the following values of all the pairs (m,k) for $\theta_{mk,0}$

Table 1: Population parameters for crop rotation effects on yield

		Current crops			
		Wheat	**Barley**	**Rapeseed**	**Sugar beet**
Prev. crop.	**Wheat**	-0.50	0.00	0.00	0.00
	Barley	-0.80	-0.70	0.15	0.10
	Rapeseed	0.00	0.50	0.05	-0.20
	Sugar beet	-0.10	0.10	0.08	-0.30

Note. Rapeseed is the reference previous crop for wheat, and wheat is the reference previous crop for barley, rapeseed and sugar beet.

Table 2: Population parameters for crop rotation effects on input uses

		Current crops			
		Wheat	**Barley**	**Rapeseed**	**Sugar beet**
		Fertilizer use			
Previous crops	**Wheat**	0.40	0.00	0.00	0.00
	Barley	0.30	0.10	0.10	-0.10
	Rapeseed	0.00	-0.30	-0.05	-0.20
	Sugar beet	0.10	-0.40	-0.10	0.10
		Pesticide use			
	Wheat	0.30	0.00	0.00	0.00
	Barley	0.20	0.10	-0.20	0.10
	Rapeseed	0.00	-0.20	0.15	0.20
	Sugar beet	-0.15	-0.30	0.10	0.10

Note. Rapeseed is the reference previous crop for wheat, and wheat is the reference previous crop for barley, rapeseed and sugar beet.

In order to account for the fact that farmers crop sequence acreage choices might be driven by factors other than profit maximization, we consider simulating a stochastic profit as follows: $\tilde{\pi}_{mk,it} = \theta'_{mk,0}\mathbf{q}_{k,it} + \omega_{mk,it}$ with $\omega_{mk,it} \sim \mathcal{N}\left(0, \sigma^2_{\omega_{mk}}\right)$ where $\sigma^2_{\omega_{mk}}$ is chosen such that $\sigma^2_{\omega_{mk}} = \text{var}\left(\omega_{mk,it}\right) = \text{var}\left(\tilde{\pi}_{mk,it}\right)/2$

Table (3) displays the average of the solution to the $\mathbf{LP}_{it}$ problem, *i.e.*, $\text{z}_{mk,0} = (\text{Card}(S))^{-1} \sum_{(i,t)\in S} \hat{\text{z}}_{mk,it}$ where $\text{Card}(S)$ is the number of elements in S.

Table 3: Average optimal crop sequence acreage shares

		Current crops			
		Wheat	**Barley**	**Rapeseed**	**Sugar beet**
Prev. crop.	**Wheat**	0.24	0.45	0.43	0.47
	Barley	0.08	0.03	0.44	0.51
	Rapeseed	0.44	0.38	0.05	0.02
	Sugar beet	0.24	0.15	0.08	0.00

We now generate the yield and chemical input use data from these models:

$$\begin{cases} \text{y}^{\text{sim}}_{k,it} & = \bar{y}_k + \mathbf{z}'_{k,it}\theta^{(\text{y})}_{k,0} + \upsilon^{(\text{y})}_{k,it} \\ \text{x}^{\text{sim}}_{j,k,it} & = \bar{x}_{j,k} + \mathbf{z}'_{k,it}\theta^{(\text{x})}_{j,k,0} + \upsilon^{(\text{x})}_{j,k,it} \end{cases} \text{with} \begin{cases} \upsilon^{(\text{y})}_{k,it} & \sim \mathcal{N}\left(0, \sigma^2_{\upsilon_k}\right) \\ \upsilon^{(\text{x})}_{j,k,it} & \sim \mathcal{N}\left(0, \sigma^2_{\upsilon_{j,k}}\right) \end{cases} \forall (k,j,i,t) \in \mathcal{K} \times \mathcal{J} \times S$$

where $\sigma^2_{\upsilon_k}$ and $\sigma^2_{\upsilon_{j,k}}$ are chosen such that the noise explains a large share of the variations of the simulated yield/chemical input uses:

$\sigma^2_{\upsilon_k} = \text{var}\left(\text{y}^{\text{sim}}_{k,it}\right) = \sigma^2_{\upsilon_{j,k}} = \text{var}\left(\text{x}^{\text{sim}}_{j,k,it}\right) \approx 0.95 \; \forall (k,j) \in \mathcal{K} \times \mathcal{J}$

We use the data simulated above to estimate the crop rotation effects. We provide suitable starting values for the smooth MPEC problem by estimating successively the problems $\mathbf{L}_{it}$ and $\mathbf{U}$. The value of the entropic perturbation parameter used for the estimation is set to $\rho = 0.1$; the model leads to an overflow for values of ρ higher than 0.1

The simulation of the netputs and the simultaneous estimation of the crop rotation effects/crop sequence acreage shares are repeated $B = 100$ times, leading to a sequence of B estimates of $\theta_{mk,0}$ denoted $\left(\hat{\theta}_{mk,b} : b = 1,...,B\right)$. We are interested in the mean and standard deviation of our estimates:

$$\hat{\theta}_{mk,B} = B^{-1} \sum_{b=1}^{B} \hat{\theta}_{mk,b} \text{ and } \hat{\boldsymbol{\sigma}}_{mk,B} = \left(B^{-1} \sum_{b=1}^{B} \hat{\theta}^{2}_{mk,b} - \hat{\theta}^{2}_{mk,B} \right)^{1/2}$$

The average of crop sequence acreage shares are obtained by computing the quantities:

$$\hat{z}^{\rho}_{mk,it,B} = B^{-1} \sum_{b=1}^{B} \hat{z}^{\rho}_{mk,it,b} \text{ and } \hat{z}^{\rho}_{mk,B} = (\text{Card}(S))^{-1} \sum_{(i,t)\in S} \hat{z}^{\rho}_{mk,it,B}$$

We present in Tables (4-6) the obtained results.

Table 4: Average of crop rotation effects on yield levels

		Current crops							
		Wheat		Barley		Rapeseed		Sugar beet	
		$\hat{\theta}_{mk,B}$	$\theta_{mk,0}$	$\hat{\theta}_{mk,B}$	$\theta_{mk,0}$	$\hat{\theta}_{mk,B}$	$\theta_{mk,0}$	$\hat{\theta}_{mk,B}$	$\theta_{mk,0}$
Previous crops	**Wheat**	-0.39 (0.05)	-0.50	0.00 (-)	0.00	0.00 (-)	0.00	0.00 (-)	0.00
	Barley	-0.72 (0.12)	-0.80	-0.57 (0.31)	-0.70	0.08 (0.01)	0.15	0.12 (0.02)	0.10
	Rapeseed	0.00 (-)	0.00	0.36 (0.06)	0.50	0.00 (0.26)	0.05	0.09 (1.41)	-0.20
	Sugar beet	-0.09 (0.07)	-0.10	0.08 (0.27)	0.10	0.04 (0.23)	0.08	-0.05 (0.72)	-0.30

Standard deviations, $\hat{\boldsymbol{\sigma}}_{mk,B}$ are displayed in parentheses below the corresponding average estimates, $\hat{\theta}_{mk,B}$.
Rapeseed is the reference previous crop for wheat, and wheat is the reference previous crop for barley, rapeseed and sugar beet.

Table 5: Average of crop rotation effects on chemical input use levels

		Current crops							
		Wheat		Barley		Rapeseed		Sugar beet	
		$\hat{\theta}_{mk,B}$	$\theta_{mk,0}$	$\hat{\theta}_{mk,B}$	$\theta_{mk,0}$	$\hat{\theta}_{mk,B}$	$\theta_{mk,0}$	$\hat{\theta}_{mk,B}$	$\theta_{mk,0}$
		Fertilizer use levels							
Previous crops	**Wheat**	0.31 (0.05)	0.40	0.00 (-)	0.00	0.00 (-)	0.00	0.00 (-)	0.00
	Barley	0.23 (0.07)	0.30	0.01 (0.15)	0.10	0.11 (0.01)	0.10	-0.09 (0.02)	-0.10
	Rapeseed	0.00 (-)	0.00	-0.27 (0.04)	-0.30	-0.06 (0.62)	-0.05	-0.90 (6.68)	-0.20
	Sugar beet	0.06 (0.06)	0.10	-0.35 (0.18)	-0.40	-0.16 (1.04)	-0.10	0.12 (3.94)	0.10
		Pesticide use levels							
	Wheat	0.19 (0.05)	0.30	0.00 (-)	0.00	0.00 (-)	0.00	0.00 (-)	0.00
	Barley	0.13 (0.05)	0.20	-0.02 (0.14)	0.10	-0.19 (0.03)	-0.20	0.05 (0.01)	0.10
	Rapeseed	0.00 (-)	0.00	-0.19 (0.04)	-0.20	0.18 (0.68)	0.15	2.02 (13.32)	0.20
	Sugar beet	-0.17 (0.05)	-0.15	-0.28 (0.14)	-0.30	0.49 (1.84)	0.10	0.36 (5.38)	0.10

Standard deviations, $\hat{\boldsymbol{\sigma}}_{mk,B}$ are displayed in parentheses below the corresponding average estimates, $\hat{\theta}_{mk,B}$.
Rapeseed is the reference previous crop for wheat, and wheat is the reference previous crop for barley, rapeseed and sugar beet.

The average of the estimated crop rotation parameters are quite close to their population value, except for some crop sequences. The standard errors, in parentheses, show that most coefficients are estimated with a high level of accuracy. In fact, as evidenced in Table (6), the estimated crop rotation parameters that appear to be quite far from their "true" values and are estimated with less accuracy are those that correspond to very small crop sequence acreage shares.

Table 6: Average of crop sequence acreage shares

		Current crops							
		Wheat		**Barley**		**Rapeseed**		**Sugar beet**	
		$\hat{z}^{p}_{mk,B}$	$z_{mk,0}$	$\hat{z}^{p}_{mk,B}$	$z_{mk,0}$	$\hat{z}^{p}_{mk,B}$	$z_{mk,0}$	$\hat{z}^{p}_{mk,B}$	$z_{mk,0}$
Prev. crops	**Wheat**	0.23	0.24	0.37	0.45	0.49	0.43	0.45	0.47
	Barley	0.06	0.08	0.02	0.03	0.48	0.44	0.53	0.51
	Rapeseed	0.41	0.44	0.49	0.38	0.02	0.05	0.01	0.02
	Sugar beet	0.31	0.24	0.12	0.15	0.01	0.08	0.01	0.00

5 Conclusion

We consider a BLP problem designed to estimate pre crop effects on yield and chemical input uses while simultaneously reconstructing farmers' unobserved crop sequence acreages from farmers' observed (current and previous) crop acreages. Our estimation approach is based on a well-defined statistical procedure and relies on crop sequence yield and chemical input use models, as well as on an assumption stating that farmers are economically rational when deciding their crop sequence acreages. We assess the empirical performance of our approach, and it shows its ability to reconstruct the crop sequence acreage shares and to recover the crop rotation effects. We consider two situations in the farmers' rationality assumption: a situation in which the reconstruction of the crop sequence acreage is entirely determined by the total crop sequence gross margin, and a situation in which some noise is added to this gross margin. Both situations lead to satisfactory estimates good estimations of crop sequence acreage shares and crop rotation effects. This suggests that the estimation approach we propose can uncover crop rotation effects, at least if observed crop acreages are a major determinant of the crop sequence acreages choices in combination with the maximization of the gross margin.

The main weakness of this paper is that farmers' choices of crop acreages (and thus, of crop sequence acreages) do not allow us to estimate the crop rotation effects associated with all the crop sequence pairs; their rationality leads them to choices that downplay the crop sequence acreage shares of less profitable crop sequences.

References

Berry, S., Levinsohn, J. and Pakes, A. (1995), 'Automobile Prices in Market Equilibrium', *Econometrica* **63**(4), 841–890.

Carpentier, A. and Gohin, A. (2014), Accounting for crop rotations in acreage choice modeling: a tractable modeling framework, 2014 International Congress, August 26 – 29, 2014, Ljubljana, Slovenia 182799, European Association of Agricultural Economists.

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

Eckstein, Z. (1984), 'A Rational Expectations Model of Agricultural Supply', *Journal of Political Economy* **92**(1), 1–19.

Hennessy, D. A. (2006), 'On Monoculture and the Structure of Crop Rotations', *American Journal of Agricultural Economics* **88**(4), 900–914.

Luo, Z. Q., Pang, J. S. and Ralph, D. (1996), *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press.

Meynard, J.-M., Messean, A., Charlier, A., Charrier, F., Fares, M., Le Bail, M. and Magrini, M.-B. (2013), Freins et leviers à la diversification des cultures: Etude au niveau des exploitations agricoles et des filières, Rapport d'étude 10-18, INRA.

Orazem, P. F. and Miranowski, J. A. (1994), 'A Dynamic Model of Acreage Allocation with General and Crop-Specific Soil Capital', *American Journal of Agricultural Economics* **76**(3), 385–395.

Rust, J. (1987), 'Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher', *Econometrica* **55**(5), 999–1033.

Su, C.-L. and Judd, K. L. (2012), 'Constrained Optimization Approaches to Estimation of Structural Models', *Econometrica* **80**(5), 2213–2230.

Thomas, A. (2003), 'A dynamic model of on-farm integrated nitrogen management', *European Review of Agricultural Economics* **30**(4), 439–460.

Vitale, J. D., Djourra, H. and Sidibé, A. (2009), 'Estimating the supply response of cotton and cereal crops in smallholder production systems: recent evidence from Mali', *Agricultural Economics* **40**(5), 519–533.