WORKING PAPERS

N° 1066

January 2020

"Unintended consequences of environmental policies: the case of set-aside and agricultural intensification"

Raja Chakir and Alban Thomas

Toulouse School of Economics

Unintended consequences of environmental policies: the case of set-aside and agricultural intensification

Raja Chakir* Alban Thomas†

January 20, 2020

Abstract

Set-aside (fallow) policies providing agronomic and ecological benefits have been mainstream practices in European agriculture. Because they can lead to intensification on cultivated land, they may however have mixed environmental effects. To evaluate the indirect impact of a set-aside policy on the environment through crop intensification, we consider two elasticity-based indicators with respect to set-aside subsidy: chemical input demand and intensity of input use. We estimate a structural, multi-output model on a panel of French farmers from 2006 to 2010, accounting for multivariate selection (corner solutions) on crops and land use. We use a parametric and a semi-nonparametric version of a quasi-maximum likelihood (QML) estimator and compare their performances in terms of goodness of fit and parameter efficiency. Results show that a set-aside subsidy can provide farmers with incentives to intensify their production, leading to potential adverse environmental effects that can however be offset by a complementary tax policy instrument.

Keywords: Set-aside, land use, fertilizer and pesticide input demand, corner solutions, semi non-parametric estimation.

JEL Classification: Q12, C33, C34.

*Corresponding author. Economie Publique, INRAE, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France. e-mail: Raja.Chakir@inrae.fr.

†Toulouse School of Economics-Research, INRAE, University of Toulouse, 31000 Toulouse, France. e-mail: Alban.Thomas@inrae.fr

1 Introduction

Set-aside (fallow) schemes involve farmers being paid to take land out of production, and are used widely in the EU and USA as an agricultural policy tool. In many cases, the initial aim of this measure was to reduce excess supply of cereals while lowering the level of public agricultural stocks. It was later complemented by tools to promote the development of non-food crops and maintain good environmental status. In the United States, the Conservation Reserve Program (CRP) was introduced in 1985 as a voluntary set-aside program designed to control crop overproduction, reduce soil erosion, improve water quality, and provide wildlife habitat taking vulnerable agricultural land out of production. In the European Union (EU), compulsory set-aside was one of the most important measures introduced at the time of the 1992 Common Agricultural Policy (CAP) reform. In 2008, the policy package associated with the CAP Health Check abolished set-aside for arable crops, and farmers could continue to set land aside on a voluntary basis while adopting agri-environmental schemes with cross-compliance.

In December 2013, the EU enacted the CAP for the period 2014-2020 where, in addition to the Basic Payment Scheme, each farmer now receives a Green Direct Payment per hectare for respecting specific greening agricultural practices to reduce biodiversity loss and greenhouse gas emissions (European Commission, 2013). Member States are requested to use 30 percent of their national budget in order to contribute to this program. The three greening measures are the following: (i) maintaining permanent grassland; (ii) crop diversification, and (iii) maintaining an Ecological Focus Area (EFA) of at least 5 percent of the arable area of the holding for farms with an area larger than 15 hectares (excluding permanent grassland) - i.e. field margins, hedges, trees, fallow land, landscape features, biotopes, buffer strips, afforested area. This "greening" of the CAP was described as a move back to compulsory set-aside[1]. Indeed, set-aside originally introduced for supply control purposes could have important environmental benefits especially where land was left fallow (Matthews, 2013). However, the literature review by Hauck et al. (2014) shows that the CAP greening measures could have mixed effects on ecosystems services, as they can lead to intensification on cultivated land. The objective of this paper is to shed some light on the impacts of the set-aside policy on the environment, namely on intensification on the cultivated crop area.

By modifying the opportunity cost of farm land, the set-aside policy is expected to modify farmers' production decisions in terms of crop choice and input use. However, the impact of set-aside policies on input use and intensification remains unclear: on the one hand, removing a proportion of land from production might reduce input use and increase extensification (Love and Foster, 1990; Fraser and Stevens, 2008). On the other hand, set-aside might have adverse effects on water and soil quality if input use (fertilizer, pesticide) increases in order to balance the reduction in cultivated land. See Love and Foster (1990); Choi and Helmberger (1993); Rygnestad and Fraser (1996); Ball et al. (1997);

Wu (2000); Fraser and Waschik (2005); Vannini et al. (2008) for more details on slippage effects. In the case of CRP, early analysis found that contracts were targeted to reduce production, rather than achieving environmental benefits (Reichelderfer and Boggess, 1988). Hendricks et al. (2018) show that the government has adjusted CRP acreage over time in response to changes in market conditions but not to environmental impacts. In the case of the new EFA policy implemented in the EU and aimed at preserving biodiversity, the intensity of the impact depends on the type of EFA (fallow, grassland), the specific environmental issue considered (water pollution, soil pollution, biodiversity loss) and site-specific environmental conditions.

Intensification of agrochemical input on crop land is one of the unintended impacts of set-aside policy, as is the slippage effect that was observed in the CRP in the United States (Wu, 2000). These two impacts can be explained by the increased output prices associated with reduced production on set-aside land. In particular, input decisions may be modified following a change in the set-aside subsidy rate, if a) it is more profitable to adapt the structure (distribution) of land to crops (some being more intensive in chemical inputs than others) ; and/or b) even with a distribution of crops unchanged, it pays to intensify production on a narrower proportion of land (by, e.g., raising the application rates of fertilizer and pesticide by the same proportion on all crops). These two changes can be measured by a combination of elasticities: agrochemical input demand and intensity of agrochemical input use, both with respect to the set-aside subsidy rate.

We examine in this paper the impacts of set-aside policy on the environment, namely intensification of chemical inputs (fertilizer and pesticide) on cultivated crop area. To capture these effects, we need to account for changes in the distribution of crops, because fertilizer and pesticide requirements are often heterogeneous across crops, as are their potential environmental impacts. This means that using individual farm data and for each crop is preferable in terms of policy impact evaluation. However, this type of data raises corner solution issues that are critical in empirical work at the individual farm level.

Although estimation of production technology has experienced major advances, there remain some econometric issues. One problem that is relevant specifically to agricultural land use allocation in empirical studies is corner solutions. Corner solutions arise when it is optimal for the farmer not to grow a crop (or combination of crops). Using aggregate data on land allocation, for example at the regional level, we observe positive values for all (region-specific) crops, but this does not imply that all farmers grow all crops. When using individual farm-level data, we need to account for the fact that some farmers may choose not to grow some crops, thus possibly causing selection bias in the parameter estimates. For this reason, models that analyze the effects of agricultural policies should adopt an explicit methodology that accommodates and explains the existence of corner solutions in the context of micro-level farm data.

We apply a dedicated estimation procedure to estimate acreage and agricultural practice responses to the set-aside policy, for a sample of farmers from the *département* of Meuse (eastern France), observed from 2006 to 2010. We estimate a multi-output profit function based on a panel of individual farmers, controlling for multiple selection using both a parametric and a semi-nonparametric quasi-maximum likelihood (QML) estimator. The purpose of this econometric strategy is to check for the robustness and consistency of the parametric estimator by comparing it with a more robust, heteroskedasticity-consistent semi-nonparametric counterpart. Parameters estimated from a structural multicrop system of equations allow us to estimate the change in pesticide and fertilizer demand corresponding to a set-aside policy. We propose two indicators to measure these impacts: chemical inputs (pesticide and fertilizer) demand and intensity elasticities with respect to the set-aside subsidy rate.

Related literature Most of the literature on multi-crop estimations relies on aggregate data or does not deal explicitly with corner solutions that occur in land-use decisions. Guyomard et al. (1996) estimate a quadratic profit function with several crop groups and inputs using French aggregate data but do not discuss the issue of corner solutions in production. Moro and Sckokai (1999) employ a normalized quadratic multi-output profit function to Italian FADN **(Farm Accounting Data Network)** data but do not exploit the panel data structure of their (individual) data set. Moreover, although they recognize the presence of possible sample selection when dealing with multiple crop groups, they do not control for these sample selection effects explicitly or consistently.

The most common technique used to estimate a structural model subject to censored observations is Tobit estimation. This was proposed in the econometric literature by Tobin (1958) and has been widely utilized in the empirical literature on demand estimations. Although the Tobit model is useful, it is an *ad hoc* modification of the regression model, allowing it to be used in cases where there are observations "piled up" at some limiting value (usually zero), and it has no convincing behavioral theory foundation (Pudney, 1989). The standard solution to the problem of a censored dependent variable is to estimate a Tobit model using maximum likelihood (ML) or the (Heckman, 1978) two-step method.

The pioneering works of Wales and Woodland (1983) and Lee and Pitt (1986) offer an economic interpretation of the corner solutions and a direct and appropriate method for specifying the econometric model. They explain that the set of producer choices can be analyzed employing the Kuhn-Tucker conditions associated with the cost minimization program under the usual technical constraints and non-negativity constraints on input demand. The implied fully structural approach in Lee and Pitt (1987) and Lee and Pitt (1986) to estimate demand and take account of corner solutions, are nonlinear simultaneous zero-censored equations models. When the number of equations is large, the subset of decision outcomes in the system likely to occur at kink points increases, requiring multiple integrations for ML estimation.

In response to the issue of dimensionality when estimating demand systems with binding non-negativity constraints, many strategies are adopted in the literature. Alternative estimation methods to the ML procedure include the maximum entropy estimator (Arndt, 1999; Golan et al., 2001), the two-step Tobit system (Perali and Chavas, 2000), generalized method of moments (GMM) techniques (Meyerhoefer et al., 2005). Yen et al. (2003) propose a QML approach which they claim is more efficient for small to medium-sized samples. Comparison with other estimation methods shows that the QML and SML (simulated maximum likelihood) procedures produce remarkably similar demand parameter and elasticity estimates whereas the results of Shonkwiler and Yen (1999)'s two-step estimator differ widely. Yen and Lin (2006) propose a sample-selection alternative with more flexible parameterization (than the Tobit system), using Shonkwiler and Yen (1999) two-step estimator. This approach is used by Lacroix and Thomas (2011) in the multiple selection case. However, the sample selection system presents more prominent computational burdens than the Tobit system since the sample likelihood function contains probability integrals with dimensions as large as the number of selection equations for all sample observations.

Simulation-based estimation methods have been suggested to overcome this problem of high dimensional numerical integration in multivariate limited dependent variables systems. These methods include method of simulated moments, method of simulated maximum likelihood, and method of simulated scores. The simulated ML approach is applied by Kao et al. (2001) and Chakir and Thomas (2003). Another way to overcome the computational issue is to use Bayesian methods. Millimet and Tchernis (2008) use the Gibbs sampling technique with the data augmentation algorithm to solve both the dimensionality and coherency problems[2] while Platoni et al. (2012b) employ the two-step approach proposed by Shonkwiler and Yen (1999) to take account of both censoring and unbalanced panel data structure. Platoni et al. (2012a) account also for the heteroskedastic structure of the error terms in the (second-step) estimation of the expectation-conditional maximization (ECM) model.

Paper contributions The present paper makes two methodological contributions to the literature. The first contribution concerns the econometric strategy for dealing with multiple selection in a micro-panel sample of farms. Our contribution in this regard is to propose a new consistent and flexible estimator for output and input decisions from individual farm data. So, we go beyond the procedure proposed by Lacroix and Thomas (2011), which is consistent only if the selection equations are uncorrelated. To allow for correlation between all the structural equations in the case of a multivariate Tobit model, we consider a QML approach, based on the joint probability of equation pairs. Yen et al. (2003) and Fezzi and Bateman (2011) use a parametric version of the QML estimator. In our paper, we propose a semi nonparametric version of the QML estimator applied to a multivariate selection model. This allows us to relax distributional and homoskedasticity assumptions, while also avoiding multiple integration. Our estimator is based on easy-to-compute distribution terms with semi nonparametric techniques (splines). To the best of our knowledge, such robust semi nonparametric methods to obtain output supply and input demand estimates from farm-level data have not yet been used in the literature. Because crops are different in terms of input requirements and environmental impacts, accounting for such heterogeneity at the farm level is likely to produce more precise estimates of farm input demands. This will allow us to more precisely estimate the environmental impacts of the set-aside policy in terms of agrochemical input intensification. The second contribution of our paper concerns the *ex ante* environmental assessment of a land set-aside policy when input taxes may be considered, as a policy instrument complementary to a subsidy policy on land set-aside. The first feature of this contribution consists in evaluating the indirect impact of a set-aside policy on crop intensification using two indicators: chemical input demand elasticities and a new indicator namely the input demand intensity elasticities both with respect to the set-aside subsidy rate. Such complementary indicator is better suited, in our opinion, to agricultural settings where input use per unit of land is more relevant than total input demand. The second feature of this contribution is to simulate the impacts, in terms of fertilizer and pesticide demand variation, of a policy aimed at increasing the ecological set-aside area by 5 percent, and to compute the level of the input tax that will be necessary to cope with their increased demand.

The paper is organized as follows. Section 2 presents a production model based on a multi-crop framework, and econometric methods to deal with corner solutions. Parametric and semi non-parametric estimators are discussed in the context of panel data. Section 3 provides both the policy context and the data description. Estimation results are presented in section 4. We evaluate in the same section the potential impacts of a set-aside policy in terms of chemical input use and intensification. We also simulate the impact of a policy consisting of a 5 percent set-aside rate as an agricultural greening measure, and calculate the input-tax on fertilizer and pesticide which would be necessary to cope with increased demand for these inputs. Section 5 concludes the paper.

2 The model and estimation issues

2.1 The model

Consider a farmer using K variable inputs x and a fixed but allocatable factor (land) to produce C different crops, where c is the crop index, $c = 1,...,C$, p_c is the price of crop c; y_c is output level of crop c, w_k is the price of input k; l_c is the land allocated to crop c and L is the total available land ($\sum_{c=1}^{C} l_c = L$). To simplify notation, we consider set-aside as a particular land use. Therefore, it is also indexed by c, although it is not a crop, technically speaking.

Maximizing profit over land and input choices, given the predetermined system of prices, subsidies, and total available land, will produce a unique set of solutions provided regularity conditions are satisfied for the profit function.

Following Guyomard et al. (1996), the multicrop profit function for a joint input technology given the fixed factor allocation (land) may be written as:

$$\Pi(p,w,\tau,L) = max_{y,x,l}\left\{\sum_{c=1}^{C} p_c y_c - \sum_{k=1}^{K} w_k x_k + \sum_{c=1}^{C} \tau_c l_c ; \sum_{c=1}^{C} l_c = L \; ; \; y \leq F(x,L)\right\}, \tag{1}$$

where $y \leq F(x,L)$ represents the technological feasibility set. τ_c is the area-based (per hectare) subsidy rate for crop c.

Under partial decoupling of public payments to agriculture, the equation above has to be modified, to accommodate the situation where only part of area-based payments is crop-specific. Let ϕ, $\phi \in [0,1]$, denote the proportion of area-based subsidies that is decoupled form production, so that $100(1-\phi)$ percent are still "coupled" and depend on the area allocated to production for each crop c.

Note that there is a distinction to be made between farm land that is set aside voluntarily to obtain an eligible payment, and farm land that is not cultivated because of other factors. Such factors include natural events such droughts, severe ground and surface water shortages, and possibly labor shortages that could hamper crop planting. In our setting, as long as land set-aside is subject to a subsidy payment, we assume that all arable land is either cultivated or subject to a set-aside, area-based payment.

Farmer's profit therefore reads

$$\Pi(p,w,\tau,L) = max_{y,x,l}\{\sum_{c=1}^{C} p_c y_c - \sum_{k=1}^{K} w_k x_k + (1-\phi) * \sum_{c=1}^{C} \tau_c l_c + FP ; \sum_{c=1}^{C} l_c = L \; ; \; y \leq F(x,L)\}, \tag{2}$$

where FP denotes a Fixed Payment (on farm, not coupled with production nor land use).

Following Lacroix and Thomas (2011), the normalized quadratic profit function is written as:

$$\begin{aligned}\overline{\Pi} = \alpha_0 + \sum_{c=1}^{C} \alpha_c \overline{p}_c + \sum_{k=1}^{K-1} \beta_k \overline{w}_k + \sum_{c=1}^{C} \gamma_c \overline{\tau}_c + \frac{1}{2}\sum_{c=1}^{C}\sum_{c'=1}^{C} \alpha_{cc'} \overline{p}_c \overline{p}_{c'} + \frac{1}{2}\sum_{k=1}^{K-1}\sum_{k'=1}^{K-1} \beta_{kk'} \overline{w}_k \overline{w}_{k'} + \frac{1}{2}\sum_{c=1}^{C-1}\sum_{c'=1}^{C} \gamma_{cc'} \overline{\tau}_c \overline{\tau}_{c'} \\ + \sum_{k=1}^{K-1}\sum_{c=1}^{C} \delta_{ck}^{pw} \overline{p}_c \overline{w}_k + \sum_{c'=1}^{C}\sum_{c=1}^{C} \delta_{cc'}^{p\tau} \overline{p}_{c'} \overline{\tau}_c + \sum_{k=1}^{K-1}\sum_{c=1}^{C} \delta_{ck}^{w\tau} \overline{w}_k \overline{\tau}_c + \sum_{c=1}^{C} \lambda_c^{pL} \overline{p}_c L + \sum_{c=1}^{C} \lambda_c^{\tau L} \overline{\tau}_c L + \sum_{k=1}^{K-1} \lambda_k^{wL} \overline{w}_k L,\end{aligned} \tag{3}$$

where $\overline{\Pi} = \frac{\Pi}{w_K}, \overline{p}_c = \frac{p_c}{w_K}, \overline{w}_k = \frac{w_k}{w_K}, \overline{\tau}_c = \frac{\tau_c}{w_K}$ indicate respectively normalised profit, normalised output price and normalised subsidy rate, and input price w_K is chosen as numeraire.

Differentiating the profit in (3) with respect to output prices $\overline{p}_c$ yields the output level of crop c (Hotelling Lemma):

$$y_c = \frac{\partial \Pi}{\partial \overline{p}_c} = \alpha_c + \sum_{c'=1}^{C} \alpha_{cc'} \overline{p}_{c'} + \sum_{k=1}^{K-1} \delta_{ck}^{pw} \overline{w}_k + \sum_{c'=1}^{C} \delta_{cc'}^{p\tau} \overline{\tau}_c + \lambda_c^{pL} L, \forall c = 1, ..., C. \tag{4}$$

Differentiating the profit in (3) with respect to input prices $\overline{w}_k$ yields the variable input demand equation (Hotelling Lemma):

$$-x_k = \frac{\partial \Pi}{\partial \overline{w}_k} = \beta_k + \sum_{k'=1}^{K-1} \beta_{kk'} \overline{w}_{k'} + \sum_{c=1}^{C} \delta_{ck}^{pw} \overline{p}_c + \sum_{c=1}^{C} \delta_{cc'}^{w\tau} \overline{\tau}_c + \lambda_k^{wL} L, \forall k = 1,...,K-1. \tag{5}$$

Following Lacroix and Thomas (2011), differentiating profit in (3) with respect to subsidy rates $\overline{\tau}_c$ yields the land allocation equation:

$$(1-\phi) * l_c = \frac{\partial \Pi}{\partial \overline{\tau}_c} = \gamma_c + \sum_{c'=1}^{C} \gamma_{cc'} \overline{p}_{c'} + \sum_{k=1}^{K-1} \delta_{ck}^{w\tau} \overline{w}_k + \sum_{c'=1}^{C} \delta_{cc'}^{p\tau} \overline{\tau}_c + \lambda_c^{\tau L} L, \forall c = 1,...,C. \tag{6}$$

Profit function properties imply that the profit function is (i) non-decreasing in output prices p, non-increasing in input prices w, (ii) homogeneous of degree 1 in prices (p,w), (iii) convex in prices (p,w), and (iv) continuous in prices (p,w). These properties imply some conditions to impose on the parameters. With the normalized form of the profit, the condition of linear homogeneity is automatically satisfied. The convexity conditions imply that the Hessian matrix is symmetric and positive semi-definite. Imposing the convexity restrictions, is equivalent to impose positive semi-definiteness on the matrix of parameters.[3]

Another regularity condition is the land adding-up condition $\sum_{c=1}^{C} l_c = L$, which implies the following conditions on the parameters:

$$\sum_{c=1}^{C} \gamma_{cc'} = \sum_{c=1}^{C} \delta_{ck}^{w\tau} = \sum_{c=1}^{C} \delta_{cc'}^{p\tau} = \sum_{c=1}^{C} \gamma_c = 0; \; \forall k, \forall c, \tag{7}$$

$$\sum_{c=1}^{C} \lambda_c^{\tau L} = 1 - \phi. \tag{8}$$

The model to be estimated consists of the system of equations (3-4-5-6) after imposing convexity restrictions and land adding-up conditions 7-8. We let $s_j(\theta)$ denote the j-th structural equation in the system, depending on exogenous covariates and a vector of parameters θ, where the number of equations depends on the number of inputs, outputs and land use. In order to obtain precise results for policy analysis, we propose in this paper the estimation of this system of equations while explicitly dealing with corner solutions and the panel structure of our data. Contrary to other applications of land-use models in agriculture (Fezzi and Bateman, 2011), the system of output, input and land equations depends only on observed prices and subsidies, and on total crop land. Other applications of land-use models consider input and output equations as an explicit function of land shares for crops. We do not follow this approach here, as the demand for crop land is part of our structural system, which contains only exogenous covariates (more precisely, from the farmer's point of view, assuming total crop land is fixed).

Imposing regularity conditions, as discussed above, is not the only issue when estimating a system of equations derived from profit maximisation with farm-level data. Another concern is the existence of corner solutions, that is, zero land area and output level for some crops, and possibly zero expenditures on some inputs. As farmers rarely consider the same cropping system for every agricultural season because of agronomic and pest-management considerations, all possible crops are not planted every year by a particular farmer, implying that land and production variables are equal to 0 in this case. It is beyond the scope of this paper to provide a structural representation of cropping rotations through a dynamic model, see, e.g., Thomas (2003); Hennessy (2006) and Lacroix and Thomas (2011). Nevertheless, we provide below an original solution to this problem by considering a multivariate selection problem and relaxing some assumptions underlying the usual multivariate Tobit model.

2.2 Corner solutions

Estimating models with multivariate selection often implies a trade-off between computer-intensive numerical procedures and strong distributional assumptions to achieve parameter consistency and efficiency. For example, Yen et al. (2003) discuss solutions based upon the Lee and Pitt (1986) approach, which requires normality and homoskedasticity of structural error terms.

Our strategy is based on a multivariate version of the Tobit model for censored equations and, contrary to Lacroix and Thomas (2011), we consider that prices and subsidies jointly determine the probability of a crop and its associated output level and land use (as would be the case in the original Tobit model). A drawback with the procedure described in Lacroix and Thomas (2011) is that their estimator is consistent only if selection equations are uncorrelated, conditional on a the set of covariates. Although this condition can be tested in practice, this may limit the scope of the method. Furthermore, the correlation pattern between structural and selection equation is also restricted to a linear form, which may depend however on the period, as in Wooldridge (1995). In order to allow for correlation between all structural equations in the case of a multivariate Tobit model, a possibility is to consider the Quasi-Maximum Likelihood (QML) approach, based on the joint probability of pairs of equations.[4] This is a simple alternative to multiple integration, which consists in exploring potential correlation between all pairs of structural equations.

Let $s_{ij}(\theta)$ denote observation i of the j-th structural equation, evaluated at parameter vector θ. The residual of equation j is denoted $h_{ij}(\theta) = Z_{ij} - s_{ij}(\theta)$, where Z_{ij} is the dependent variable in equation j.

The bivariate likelihood for observation i is given by

$$\mathscr{L}_i = \prod_{j=2}^{N-1} \prod_{k=1}^{j-1} \mathscr{L}_{ijk}, \tag{9}$$

where N is the total number of observations, and

$$\mathscr{L}_{ijk} = \left\{F_2(h_{ij}, h_{ik})\right\}^{1(Z_{ij}=0, Z_{ik}=0)} \times \left\{ f_2(h_{ij}, h_{ik})\right\}^{1(Z_{ij}>0, Z_{ik}>0)} \tag{10}$$

$$\times \left\{F(h_{ik}|h_{ij}) \times f(h_{ij})\right\}^{1(Z_{ij}>0, Z_{ik}=0)} \times \left\{F(h_{ij}|h_{ik}) \times f(h_{ik})\right\}^{1(Z_{ij}=0, Z_{ik}>0)}.$$

Denote $h^*_{ij}(\theta) = [Z_{ij} - s_{ij}(\theta)]/\sigma_j$ the standardized residual of equation j, where Z_{ij} is the dependent variable and σ_j the standard deviation of the residual in equation j. F is the cumulative distribution function and f is the density function of $h^*_{ij}(\theta)$.

Under the normality assumption, the individual contribution to the likelihood becomes

$$\mathscr{L}_{ijk} = \left\{\Psi(h^*_{ij}, h^*_{ik}, \rho_{jk})\right\}^{1(Z_{ij}=0, Z_{ik}=0)} \times \left\{\psi(h^*_{ij}, h^*_{ik}, \rho_{jk})/(\sigma_j\sigma_k)\right\}^{1(Z_{ij}>0, Z_{ik}>0)} \tag{11}$$

$$\times \left\{\frac{\phi(h^*_{ij})}{\sigma_j}\Phi\left[\frac{h^*_{ik} - \rho_{jk}h^*_{ij}}{\sqrt{1-\rho^2_{jk}}}\right]\right\}^{1(Z_{ij}>0, Z_{ik}=0)} \times \left\{\frac{\phi(h^*_{ik})}{\sigma_k}\Phi\left[\frac{h^*_{ij} - \rho_{jk}h^*_{ik}}{\sqrt{1-\rho^2_{jk}}}\right]\right\}^{1(Z_{ij}=0, Z_{ik}>0)},$$

where $\Psi(.,.,.)$ and $\phi(.,.,.)$ are the bivariate cumulative density and density functions respectively; $\phi(.)$ and $\Phi(.)$ respectively denote the univariate density and cumulative density functions of the standard Normal distribution. Maximizing (9) with respect to θ and the variance-covariance matrix yields consistent Quasi-Maximum Likelihood (QML) estimates.

2.3 The semi nonparametric estimator

As discussed by Perali and Chavas (2000), the parametric QML estimator of the multivariate Tobit above (Yen et al., 2003) is not robust to deviations from normality and homoskedasticity assumptions. A straightforward way to check for consistency is to compare estimates with a more flexible estimation method.[5] A natural way to do this is to replace standard Normal density and cumulative density functions in the parametric QML of Yen et al. (2003) by their nonparametric counterparts or by semi non parametric approximations. The distribution of equation residuals is then left unrestricted and structural parameters θ can be estimated jointly with nonparametric or semi nonparametric estimation of the univariate and bivariate distributions. This may however imply computer-intensive numerical methods if the sample size is large.

Inspecting the form of the QML estimator (11), we see that four functions need to be estimated: a) the bivariate cumulative density function, $F_2(h_{ij}, h_{ik})$; b) the bivariate density function, $f_2(h_{ij}, h_{ik})$; c) the univariate conditional cumulative density function, $F(h_{ij}|h_{ik})$; d) the univariate density function, $f(h_{ik})$.

We consider here the semi nonparametric estimator as a consistent benchmark for the parametric QML estimator, and not the nonparametric version, because the latter is highly computer-intensive and has a lower convergence rate, requiring significantly larger samples. One possible approach for semi nonparametric models with selection is to follow Gallant and Nychka (1987) and replace unknown joint distributions by a series approximation. As discussed in Schwiebert (2013), in some series expansion-based approximation of multi-dimensional density function (as in Gallant and Nychka (1987)), the number of terms grows with the sample size. An alternative is to set a fixed number of series terms, with the advantage that the estimator asymptotic distribution is known (follows from the parametric ML framework). Conditions for consistency and asymptotic normality of sieve ML estimation have been provided by Chen et al. (2006) and Chen (2007). The bivariate density function can be computed from the derivative of an approximation to the bivariate cumulative distribution function, obtained from a tensor-product spline, such tensor-product spline approximations being easy to compute from a B-spline representation.

In our multivariate framework, semi nonparametric estimation consists of replacing the nonparametric estimation of the density and cumulative density functions of the model's error terms by parametric approximations to their empirical counterparts. The principle of *sieve estimation* is to replace the infinite-dimensional space associated with the non-parametric functions by a flexible parametric one. Consistency and asymptotic normality of the sieve spline-based estimator are discussed in Chen (2007). Technical details are presented in the appendix.

An interesting aspect of B-splines is that they can be constructed under some shape and smoothness restrictions that are easy to impose, as well as their derivatives of any order, which provide us with a natural procedure for estimating probability and cumulative density functions:

- Compute the empirical cdfs in the univariate and bivariate cases, for each selection regime;
- Approximate the univariate and bivariate cdfs for every selection regime by B-splines, imposing smoothness and non-decreasing approximations over the $[0,1]$ interval;
- Compute the derivative of the spline approximation to obtain the pdf;
- Combine joint and marginal distributions to obtain the conditional cdfs for each selection regime.

2.4 Panel data

The above model can be adapted to panel data with the total number of observations $N = \sum_{i=1}^{n} T_i$, where n is the number of individuals (cross-sectional units), and T_i denotes observations for cross sectional unit i. In other words, unbalanced panels can be accommodated for, when T_i is different across cross sections.

To control for unobserved individual heterogeneity, possibly correlated with explanatory variables, we may consider a fixed-effect approach to the production model. However, as the above model is non linear, within-type estimators would not be consistent with a fixed number of time periods. We choose to control for such unobserved heterogeneity by implementing the Mundlak method (see Wooldridge (1995)). Assume

$$h_{ijt}(\theta) = Z_{ijt} - s_{ijt}(\theta) - \eta_{ij}, \quad i = 1, 2, \ldots, N,$$

where η_{ij} is the individual effect in equation j, possibly correlated with explanatory variables in $s_{ijt}(\theta)$. Consider first the balanced panel data case, with $T_i = T$, $\forall i$. We further assume

$$\eta_{ij} = X_{ij1}\gamma_{j1} + \cdots + X_{ijT}\gamma_{jT} + v_{ij},$$

where X_{ijt} is the K-vector of explanatory variables in structural equation $s_{ij}(\theta)$, such that $E(v_{ij}|X_{ijt}) = 0$. Under this conditional moment condition, η_{ij} can be replaced in the structural equation by its projection on explanatory variables. The Mundlak approach corresponds to the special case of Chamberlain (1982) with $\gamma_{j1} = \cdots = \gamma_{jT} = \bar{\gamma}_j/T$, $\forall j$, and is preferred in practice if the number of explanatory variables is large, because it inflates by $K \times T$ the number of parameters. Such conditioning is typically designed for discrete-choice models (Probit, Logit) and is naturally adapted to the Tobit framework. Because unobserved heterogeneity in structural and selection equations are likely to be correlated with explanatory variables, we do not consider random-effects estimation but only fixed-effects estimation, to control such possible source of bias in parameter estimates.

The Mundlak approach for fixed effects is easily adapted to the unbalanced panel-data case, with

$$\eta_{ij} = \left(\frac{1}{T_i} \sum_{t=1}^{T_i} X_{ijt} \right) \gamma_j + v_{ij},$$

where T_i is the individual-specific number of periods. With large N, we can consistently estimate parameters γ_j even when the panel data set is severely unbalanced.

3 Policy context and data

The study is conducted on a sample of French farmers from the *département* of Meuse. The data were provided by the *Centre d'Economie Rurale de La Meuse*, an agricultural extension service which provides farmers with assistance for accounts auditing. Our sample is an unbalanced panel observed between 2006 and 2010, with a total of 2014 observations.

3.1 CAP and set aside policy context

An interesting feature of our data and the period considered is that, during that period, France operated a decoupling scheme with a hybrid status. Following the European CAP reform in 2003 (Luxembourg Agreement, see Britz et al. (2006)) decoupling was introduced with environmental cross-compliance among other measures, aimed at making farmers' production decisions more market-oriented (Moro and Sckokai, 2013). Member States retained some flexibility in the choice between partial and full decoupling. For example, Spain, France and Portugal opted to maintain the maximum permitted amount of coupled payments (25%), in both the livestock and arable crop sectors. In the case of France, the 2008 CAP health check provided for an "*à la carte*" selection of the tools, allowing voluntary implementation to start by 2010 and end by 2012 at the latest (Boulanger, 2010). In terms of model notation of subsection 2.1, the proportion of area-based subsidies that is decoupled from production, ϕ, is equal to 0.75.

Compulsory set-aside was one of the most important measures introduced in the European Union (EU) at the time of the 1992 reform of the Common Agricultural Policy (CAP), which introduced a new support system for producers of cereals, oilseed and protein crops. Farmers with production greater than 92 tons were eligible for set-aside payment. In order to alleviate farmers' revenue decrease due to the compulsory set-aside, farmers were allowed to cultivate energy crops (diester from rapeseed in our data) on set-aside land without losing the subsidy (Rozakis and Sourie, 2005). Among the many changes to set-aside rules during the period 1993-2007, the major ones concerned the adjustments of the rate of compulsory set-aside, the introduction of voluntary set-aside against payment, and the possibility of a fixed instead of a rotational set-aside, which was the only form available at the outset. In 2008, the policy package associated with the Health Check of the CAP abolished both the energy crop scheme and the compulsory set-aside scheme. However, farmers could continue to set land aside on a voluntary basis, while adopting agri-environmental schemes with cross-compliance. The eligibility requirement for payment is in that case that at least 5 percent of land are under ecological focus area. Moreover, eligibility conditions remained the same regarding crop area: farms are eligible for area-based payment only if their arable area for cereals, oilseed and protein crops is greater than 0.3 hectares, under European directive CE/1973/2004 (October 29, 2004). In terms of maximum voluntary set aside area that is eligible for CAP payment, the ceiling rate since 2008 is $1/9^{\text{th}}$ for arable crops. It can be extended to 25 percent for arable crops (cereals, oilseed and protein crops) for energy, chemical use or animal feed. In practice, for rapeseed production, farmers were eligible for area-based payment under CAP after the 1992 reform, while voluntary set-aside over and above the compulsory rate was allowed. After 2008, rapeseed for diester production was also possible on set-aside areas, with the same subsidy rate as for non-set aside area, i.e., a different subsidy rate from the "agronomic" set-aside rate. In this case, farmers have to show evidence of a farming contract with energy or industrial buyers.

3.2 Sample description

The average total farm area is 197.9 ha, of which 28.1 ha of arable land (standard deviation of 18.02 ha) and 74.9 for permanent grazing and temporary land for pasture. Average production cost is 263,777.82 Euros / year and total profit per farm is 56,783.74 Euros per year, about 285.18 Euros per ha (standard deviation of 598.82). These statistics of dispersion indicate that farm diversity is limited, as far as size and economic performance are concerned. In terms of spatial location, farms are widespread over the whole *département* Meuse, as can be seen from Figure 1, which represents the spatial distribution of sample farms in the *cantons* (an administrative unit, between the *commune* and the *département*). The total area covered by the sample is relatively limited (about 6200 squared kilometers, about 2400 sq. miles), so that differences in climate and soil characteristics are fairly limited as well. Moreover, as fixed effects procedures are employed in the estimation, farm-specific or site-specific non time-varying characteristics will be filtered out from the model.

[INSERT FIGURE 1 ABOUT HERE]

Concerning crops and inputs, we use the major cropping systems in the *département* of Meuse and select wheat, barley, rapeseed and diester, the last of which is used for biofuel production. Because wheat is in a vast majority of cases associated with barley, and farm-gate prices of both crops evolve in parallel, we combine them to form a composite cereal output. Land decisions are associated with these crops (cereals, rapeseed and diester) as well as voluntary land set-aside, as discussed above. We consider only three inputs: seed, fertilizer and pesticide, which are considered the most crop-specific and therefore whose demand is more likely to be influenced by a change in cropping pattern (for our set of arable crops, as opposed to, e.g, labor and fuel).[6] To obtain farm-specific input prices, we proceed as follows. As database records contain only input expenditures and not input physical quantities nor unit prices, we use official yearly statistics on agricultural input price indexes at the département level. We then convert to farm-specific price indexes using the Tornqvist formula, with 2004 as the baseline year.[7] We then check for possible multicollinearity in input and output prices, with the condition number, computed every year and over the whole sample. The condition number is always less than 15, confirming that multicollinearity in input prices is limited. Crop outputs are in tons, fertilizer input is in kg, and pesticide input is computed from pesticide expenditure divided by its price index.

All prices and unit subsidies are normalized to the unit cost of seed so that in the estimation we consider only the other two inputs. Our sample covers the pre- and the post-2008 period, where compulsory set-aside was abandoned and industrial crops such as biofuels were allowed on set-aside land with the same subsidy rate as other arable crops (different from the agronomic set-aside subsidy rate). This concerns, in our case, rapeseed that can be produced for energy use or for the agrofood industry, with the same area-based subsidy rate after 2008. Our data at the farm and crop level allow us to account for the different cases, i.e., area under rapeseed for industrial use has a different area-based unit payment before and after 2008, while it remains at the same rate for rapeseed sold for agrofood industry. Because changes in use-specific area-based payments over the period are fully accounted for at the farm level when constructing our payment τ_c and area l_c variables, we assume that the model parameters will not depend on such policy changes. Therefore, we do not include in the model a dummy variable for the pre- and post-2008 period, as we assume parameters are neither period- nor policy-dependent. Moreover, as mentioned in the data subsection above , the eligibility condition for payments under CAP was a minimum arable area of 0.3 ha. There are no such farmers in our sample, so that all were eligible and actually received some form of area-based CAP payments every year from 2006 to 2010.

Table 1 presents descriptive statistics for the sample. Almost all farmers grow cereals (wheat and barley) while the respective proportions of positive land shares for rapeseed, diester and land set aside are 83%, 31% and 70% of the full sample. These descriptive statistics need to be interpreted with caution since some farmers may not grow a particular crop over the whole period.

[INSERT TABLE 1 ABOUT HERE]

3.3 Intensity of input indicators

As discussed above, one reason for using detailed farm-level and crop-level production data is that input use (and therefore, environmental indicators) may differ across cropping systems, implying that farmer decisions including corner solutions should be modeled explicitly. Using the sample of farmers, we compute average environmental indicators for various cropping systems: cereals only, cereals-rapeseed, etc. II_k denotes the intensity of input k (k=fertilizer, pesticide) demand on cultivated land, defined as

$$II_k = \frac{input\ k\ demand}{cropland} = \frac{x_k}{(l_{cereal} + l_{rapeseed} + l_{diester})} = \frac{x_k}{L - l_{setaside}}. \tag{12}$$

Descriptive statistics for our environmental indicators are presented in Table 2. Means and standard deviations are computed for a given combination of crops (e.g. cereals-diester) over all corresponding observations across farmers and years. Crop combinations with less than 10 observations are discarded. Table 2 confirms that input use is strongly heterogeneous across cropping systems. This results in very different average environmental indicators, and implies that accounting for farmer decisions over a whole cropping system (i.e. including decisions leading to corner solutions for some crops) is preferable in terms of policy impact evaluation. Cereal-only cropping systems (C) have a lower fertilizer and pesticide input use (x_F and x_P), and a lower input intensity indicator, except for fertilizer where II_F is slightly higher for cereals only than for cereal and rapeseed ($C+R$ only). The cereal-rapeseed ($C+R$)

crop combination is associated with a fertilizer input intensity very close to cereal alone (C), while pesticide input intensity is much higher in the ($C+R$) system than in C. The difference in input intensity is more pronounced, especially for fertilizer, when diester is included in the crop combination ($C+D$ or $C+R+D$). Interestingly, cropping systems that involve diester ($C+R+D$) have a lower fertilizer use intensity than $C+D$ only, but a higher intensity of pesticide use (4.72 compared with 1.82, 4.52 and 4.13).

[INSERT TABLE 2 ABOUT HERE]

4 Econometric estimation and simulation results

We apply in this paper a parametric QML (PQML) as well as a semi nonparametric QML (SNPQML) estimator to a multivariate selection model, both with a fixed effects assumption.[8] Given the large number of estimated parameters and for reasons of space, we report only elasticity estimates for both PQML and SNPQML computed at the sample mean in Tables 4 and 5 respectively. All standard errors are estimated using a heteroskedasticity-robust (Huber-White) procedure.

Before interpreting parameter estimates, we compare the goodness of fit of the two estimated models. We examine in particular the behavior of the semi nonparametric estimator SNPQML in terms of predictive power, both on continuous (structural equations) and on discrete outcomes (selection of observations, positive or equal to zero). The objective is to examine the possible gains in flexibility brought about by SNPQML, compared with the parametric version PQML, which relies on a normality assumption.

4.1 Comparison of alternative methods

To evaluate the relative performance of the parametric and semi nonparametric QML estimators on our panel data sample, we compute goodness-of-fit measures for the continuous and discrete parts of the model. We first compute R^2s on the full sample, for each structural equation. These R^2s are computed with respect to the observed, continuous explanatory variables, without accounting for the contribution of estimated fixed effects. Because such goodness-of-fit measures are difficult to interpret when the proportion of zero observations is large, we also produce R^2s for each equation on the subset of positive observations only. Columns 2 and 3 of Table 3 present computed R^2s for the parametric and semi nonparametric QML estimators. We note that the difference between R^2s on the full sample or the sub-sample of positive observations is noticeable mostly when the proportion of censored observations is higher (case of q rapeseed, q diester, l rapeseed, l diester and l setaside), with the R^2s on the full sample being lower than the restricted version in 12 cases out of 27. Interestingly, the model fits data better with parametric PQML than with semi non parametric SNPQML on the full sample (6 cases out of 9), but this is the opposite on the restricted samples. The semi non parametric estimator performs better than PQML on sub-samples with positive observations only.

To check for serial correlation in residuals estimated from PQML and SNPQML, we compute the Heteroskedasticity-Robust test statistic (HR) proposed by Born and Breitung (2016) in the context of a panel data model with fixed effects. For fixed T and when N tends to infinity, the text statistic HR has a standard normal limiting distribution under the null of no serial correlation. As discussed in Born and Breitung (2016), rejecting the null hypothesis of no serial correlation provides evidence that a static model may not be appropriate and may be replaced by a dynamic panel data model. Test results reported in Table 3 indicate that serial correlation is not present in 15 cases out of 18, with a 5 percent significance level, so that the static specification is valid in a majority of structural equations. Moreover, in a static panel data model with fixed effects, serial correlation can be accommodated with the estimation of a robust variance-covariance matrix, as in the case here, to produce consistent standard errors.

Turning now to the goodness-of-fit measures for discrete outcomes (last three columns of Table 3), it is more interesting to focus on the equations with a significant proposition of zero observations (q rapeseed, q diester, l rapeseed, l diester and l setaside), because measures such as percentage of correct predictions are difficult to interpret on a (very) small number of false predictions (q cereal, x fertilizer, x pesticide and l cereal have proportions of censored observations less than 1 percent on a sample of size 2014). For these five equations, the proportion of correct predictions (positive and negative outcomes) is fairly similar for PQML and SNPQML, so that the gain associated with the latter is only minor. When inspecting the proportion of false positive predictions for negative outcomes or the opposite (i.e., predicting a zero observation as a positive one), the number of cases where SNPQML fares better is fairly similar to cases where PQML performs better on such criterion. Out of the five equations with a significant proportion of zero observations, SNPQML has a lower proportion of 1s predicted as 0s in three cases of of five, while SNPQML has a higher proportion of 0s predicted as 1s in three cases of of five. However, the difference in such goodness-of-fit measures is significantly larger when SNPQML performs better than PQML.

[INSERT TABLE 3 ABOUT HERE]

4.2 Econometric estimation results

[INSERT TABLE 4 ABOUT HERE]

As shown in Table 4, set-aside area is significantly related to its unit set-aside subsidy (elasticity=0.1479). The results also show that an increase in this subsidy rate implies an increase in output and planted area of rapeseed, as well as a minor increase in the cereal output and land area. This means that, when the set-aside unit subsidy rate increases, farmers tend to intensify their production for these crops as they increase in parallel their set-aside area. This is confirmed by the elasticities of fertilizer and pesticide with respect to set-aside subsidy, which are positive and significant and equal to 0.0589 and 0.029 respectively. In the case of diester, the results show that an increase in the set-aside subsidy implies a reduction of both output and area of this crop, and this substitution effect is stronger than the increase in cereal and rapeseed output and land use discussed above. Fertilizer demand is increasing with cereal and rapeseed prices (elasticities of 0.0117 and 0.0555 respectively) and is decreasing with the price of diester (elasticity of -0.1261). Fertilizer demand is also increasing with rapeseed subsidy (0.0350) and decreasing with diester subsidy (elasticity of -0.1200), and does not vary significantly with the price of cereals (elasticity of -0.0010). In the case of pesticide, demand for this input is increasing with the diesters price (elasticity of 0.0958) and decreasing with the prices of cereals and rapeseed (elasticities of -0.0120 and -0.0548 respectively). Pesticide demand is also increasing with the subsidy of diester and decreasing with the unit subsidy of cereals and rapeseed (elasticities of 0.0498 and -0.0316 respectively).

Results of the estimated elasticities for the semi non-parametric estimator with fixed effects are reported in Table 5. The magnitude of elasticities is similar to the estimates obtained using the PQML model in Table 4. There is one noticeable difference, namely the impact of the set-aside unit subsidy rate is less significant in the output and land equations than in the parametric estimation, although elasticities remain fairly low. Moreover, the own-price elasticity of fertilizer demand is no longer significant. Recall that Table 4 presents minus the elasticities of fertilizer and pesticide demand with respect to set-aside subsidy; most of them are not significant, although their values are close to those estimated using the parametric model. The elasticities of fertilizer and pesticide demand with respect to set-aside subsidy are not significant, even though their values are very close to those estimated with the parametric model.

To summarize our results in terms of comparison of the PQML and SNPQML estimators, the semi nonparametric QML is less efficient than the parametric version even if the parameter estimates are fairly close. However, the semi nonparametric estimator performs at least as well in terms of predicting output, input and land use decisions, as well as in predicting the probability of corner solutions (discrete outcomes).

[INSERT TABLE 5 ABOUT HERE]

4.3 Simulation of environmental impacts

The purpose here is not to provide an assessment of direct environmental consequences of policy implementation. Rather, we insist on the importance of an ex ante evaluation of land-use policies such as set aside and greening of the Common Agricultural Policy, in a context of increasing "social demand" for more sustainable agricultural practices. In the broader context of the CAP reform and more generally, of agricultural policies embedding environmental objectives, our simulation aims at documenting the implementation issues of tax policies on nonpoint source emissions such as from pesticide and fertilizer inputs. In the context of regulating nonpoint source pollution from agricultural sources, indirect tax-subsidy policy is sometimes advocated, such as subsidizing less input-intensive crops or alternative land use (e.g., land set aside).

To address the possibility that subsidizing set-aside could worsen the environmental effects associated with chemical inputs, we propose two environmental indicators, which can be linked to a policy instrument such as a set-aside subsidy.[9] Note that, because we use only production data, these indicators will be proxies for actual environmental impacts, and will tend to measure environmental "pressure" from production rather than an actual impact on the ecosystem. The first indicator we compute is the elasticity of input demand (fertilizer, pesticide) with respect to the set-aside subsidy rate, which measures the sensitivity of farm-level input demand to a change in the unit set-aside subsidy rate all else remaining equal. Assuming total farm land is constant, this indicator is relevant at the farm level and depends indirectly on land set-aside and crop decisions. The drawback of using this indicator is that the intensive margin (i.e., increasing input intensity per unit of land) is relevant only if computed for the cropped area (Green et al., 2005). Therefore, we consider a second indicator based on intensity of input use per unit of cultivated land, which allows us to measure the intensification effect of changes in the set-aside subsidy.

We discuss below the measurement of the second sensitivity indicator, and the results of the calculations for both indicators: input demand elasticity and input intensity elasticity with respect to the set-aside subsidy rate. In subsection 4.3.2, we evaluate environmental impacts of a set-aside policy, obtained as proxies from pesticide and fertilizer demand and input intensity elasticities with respect to land set-aside subsidy. We simulate the impact of a policy consisting of a 5 percent set-aside rate as an agricultural greening measure. We measure the impacts of such policy on our two indicators and calculate the input-tax on fertilizer and pesticide that would be necessary to cope with fertilizer and pesticide demand increase induced by the greening policy.

4.3.1 Chemical input demand and intensity elasticities

The impact on the environment of a set-aside policy will depend on changes to land use and cropping practices, and the farm level environmental conditions. In the absence of detailed environmental data, indicators of environmental pressure based on production and land use data can be used as proxies for environmental impact. Because the purpose here is to examine farmers' reactions to set-aside policies in terms of agricultural production intensification, we consider two different elasticities as indicators. The first is elasticity of chemical input demand (fertilizer, pesticide) at farm level with respect to the set-aside subsidy, irrespective of land use and, in particular, crop distribution. This indicator is useful for environmental policies targeting input sales when there is no or limited information on crop distribution and farm land area. It is therefore less costly for the environmental regulator to implement, and presumably is less distorting for cropping decisions since it does not account for set-aside land. The second indicator is defined as the elasticity of chemical input quantity per unit of cultivated land with respect to the set-aside subsidy. This indicator accounts explicitly for land set aside by considering chemical input use intensity by unit of cultivated land. It accounts for farmers' decisions about land set-aside following a change to the subsidy but is not dependent on crop distribution (only on total cultivated area). The second indicator, namely elasticity of chemical input intensity with respect to the set-aside subsidy, is calculated as follows:

$$\varepsilon_{II_k\overline{\tau}_s} = \frac{\partial II_k}{\partial \overline{\tau}_s} \times \frac{\overline{\tau}_s}{II_k} = \frac{\partial(\frac{x_k}{L-l_{setaside}})}{\partial \overline{\tau}_s} \times \frac{\overline{\tau}_s}{(\frac{x_k}{L-l_{setaside}})} = \frac{\partial x_k}{\partial \overline{\tau}_s} \times \frac{\overline{\tau}_s}{x_k} - \frac{\partial(L-l_{setaside})}{\partial \overline{\tau}_s} \times \frac{\overline{\tau}_s}{L-l_{setaside}}$$

$$= \frac{\partial x_k}{\partial \overline{\tau}_s} \times \frac{\overline{\tau}_s}{x_k} + \frac{\partial(l_{setaside})}{\partial \overline{\tau}_s} \times \frac{\overline{\tau}_s}{L-l_{setaside}} = \varepsilon_{x_k\overline{\tau}_s} + \varepsilon_{l_{setaside}\overline{\tau}_s} * \frac{l_{setaside}}{L-l_{setaside}}. \quad (13)$$

The first term $\varepsilon_{x_k\overline{\tau}_s}$ is the demand elasticity of input k with respect to the set-aside subsidy, and measures how a variation of 1 percent in the set-aside subsidy affects the percentage demand for input k. The second term $\varepsilon_{l_{setaside}\overline{\tau}_s}$, is the elasticity of set-aside area with respect to the set-aside subsidy which measures how a variation of 1 percent in the set-aside subsidy affects the percentage of the area set aside. A positive $\varepsilon_{II\overline{\tau}_s}$ means that an increase of 1 percent in the set-aside subsidy will increase input demand k by unit of cropped land, implying that an intensification effect is observed.

[INSERT TABLE 6 ABOUT HERE]

Table 6 summarizes the elasticities of fertilizer and pesticide demand with respect to set-aside subsidy for PQML and SNPQML specifications, and the elasticity of set-aside area with respect to set-aside subsidy. It also presents the results for input intensity elasticity of fertilizers $\varepsilon_{II_f\overline{\tau}_s}$ and pesticide $\varepsilon_{II_p\overline{\tau}_s}$ with respect to the set-aside subsidy.

There are several conclusions that emerge from these results regarding the environmental impact of a set-aside policy. First, all else being equal, a set-aside subsidy has a positive impact on farm-level fertilizer and pesticide demand: an increase in the set-aside subsidy by 1 percent implies an increase by 0.0385 percent in fertilizer demand and 0.0239 percent in pesticide demand, respectively. Second, comparing results of the parametric and semi nonparametric models shows that the estimated elasticities are very close, although not in the case of the standard error estimates. Third, input demand intensity elasticities with respect to a set-aside subsidy are positive and significant for both fertilizer and pesticide demand in the PQML model. This means that, when the set-aside subsidy increases, farmers tend to increase both their set-aside area $l_{set-aside}$ and their input use (fertilizer and pesticide); to compensate for the loss due to a reduced crop area, farmers intensify their production by increasing their chemical input demand per hectare of crop area. In our case, this means that increasing the set-aside subsidy could have a negative impact on the environment (in terms, for example, of water contamination and biodiversity loss).

Comparing the input elasticities for fertilizer and pesticide, the value of $\varepsilon_{II_f \overline{\tau}_s}$ is always higher than $\varepsilon_{II_p \overline{\tau}_s}$. Fertilizer is usually considered a risk-increasing input since it increases the expected crop yield and its variance. In contrast, pesticide is generally considered a risk-reducing input since its main purpose is to control for pest damage to crops which decreases the variance in crop yield.

4.3.2 Set-aside policy simulation

According to the new rules following the current CAP reform 2014-2020, farmers are required to implement greening measures or lose up to 30 percent of their basic payment scheme payment. The greening rules cover three areas: crop diversification, ecological focus areas, and non-intensification measures to maintain permanent grassland. Farmers must ensure that 5 percent of their total land is set aside as an ecological focus area. However, they are free to choose how to meet this requirement from among a list of agricultural practices or systems which include land left fallow, buffer strips, "catch and cover crops" used to manage soil fertility and quality, and nitrogen fixing crops such as legumes and hedgerows. Unless exempted, farmers with more than 15 hectares of arable land must set aside as EFA, 5 percent of their total arable land.

As the results of our elasticity estimation show, an increase in the set-aside subsidy could imply an increased demand for fertilizer and pesticide inputs. This means that a set aside policy introduced as an EFA in order to preserve biodiversity, could have some potential adverse environmental impacts due to intensification at the farm level. We use our estimates to simulate the impacts of a public policy which imposes a 5-percent increase in the set-aside area, on demand for fertilizer and pesticide. To do this, we use our elasticity estimates of set-aside area with respect to the set-aside subsidy to calculate the subsidy increase required to achieve the 5 percent increase in the set-aside area.

Let us start with the elasticity of set-aside area, $\varepsilon_{l_s\overline{\tau}_s}$, with respect to set-aside subsidy, τ_s on crop s:

$$\varepsilon_{l_s\overline{\tau}_s} = \frac{\partial l_s}{\partial \overline{\tau}_s} \times \frac{\overline{\tau}_s}{l_s} = \frac{\partial l_s}{l_s} / \frac{\partial \overline{\tau}_s}{\overline{\tau}_s}. \tag{14}$$

If we assume that $\frac{\partial l_s}{l_s} = 0.05$ (5 percent), we can calculate the corresponding (equivalent) variation in the set-aside subsidy as

$$\frac{\partial \overline{\tau}_s}{\overline{\tau}_s} = \frac{0.05}{\varepsilon_{l_s\overline{\tau}_s}}.$$

We then use this variation of the set-aside subsidy above and the fertilizer and pesticide demand elasticities with respect to the set-aside subsidy ($\varepsilon_{f\overline{\tau}_s}$ and $\varepsilon_{p\overline{\tau}_s}$ respectively) to calculate the corresponding fertilizer and pesticide demand variations, denoted $\dfrac{\partial x_f}{x_f}$ and $\dfrac{\partial x_P}{x_P}$ respectively. We finally use these input demand variations and fertilizer and pesticide own-price elasticities ($\varepsilon_{x_f\overline{w}_f}$ and $\varepsilon_{x_p\overline{w}_p}$ respectively) to calculate the corresponding fertilizer and pesticide price variations ($\dfrac{\partial w_f}{w_f}$ and $\dfrac{\partial w_p}{w_p}$).

For this simulation, we use elasticities calculated from the parametric PQML model with fixed effects (Table 4). This choice is motivated by more efficient estimates (with respect to the semi nonparametric model, as discussed above), with most elasticities significant at the 1 percent level. The results of this exercise are summarized in Table 7. Confidence intervals obtained from robust standard errors of parameter estimates are presented in brackets for the key expressions in this table.

[INSERT TABLE 7 ABOUT HERE]

Our results show that, in order to obtain an increase in the set-aside area by 5 percent, we need to increase the set-aside subsidy rate by 33.81 percent. Using fertilizer and pesticide elasticities with respect to set-aside subsidy, our simulations show that an increase of the set aside subsidy by 33.81 percent implies an increase in fertilizer demand ranging from 1.87 to 2.11 percent, and for pesticide, an increase in demand ranging from 0.81 to 1.15 percent (columns 3 and 4 of Table 7). This could have potentially adverse effects on the environment including, e.g., nitrogen runoff and ground water pollution. Using own price elasticities of fertilizer and pesticide demand, we can calculate the tax level necessary to offset such increase in fertilizer and pesticide demand. Our simulations show that such tax rates would range from 36.9 to 41.63 percent for fertilizer, and from 2.18 to 3.12 percent for pesticide (changes in input prices from last two columns of Table 7). In line with most empirical papers dealing with elasticities of input use in agriculture (see,e.g., (Boecker and Finger, 2017)), our results show that it requires a substantial tax level on pesticide and fertilizer to yield a significant reduction in input use: a tax on fertilizer between 37 and 42 percent to offset an increase in demand between 1.87 to 2.11 percent; and a tax on pesticide use between 2.18 and 3.12 percent to offset an increase in demand between 0.81 and 1.15 percent.

5 Conclusion

CAP reforms to land set aside policies imply changes to crop choices and production practices whose effects and intensity depend on various factors. Set-aside policies may increase crop yield since farmers tend to use low-yield soils to meet set-aside requirements. As a consequence, average cultivated land quality increases, which implies an increase in aggregate crop yield per hectare. Yield increases may be obtainable also because intensification, which could have some adverse impacts on the environment and would conflict with the initial objectives of the policy. We evaluate the potential environmental effect of a set-aside policy such as that implemented in the European Union in 2003, based on changes to agricultural practices and land use for a sample of French farmers.

To obtain consistent estimates of fertilizer and pesticide input demand in the case of multiple crops, we estimated a parametric QML estimator, to deal with corner solutions in a system of equations for multi-output production and land allocation. To check for possible bias due to departure from normality and homoskedasticity assumptions, we compared our parametric estimator to a semi nonparametric version, which allows distributional assumptions and homoskedasticity to be relaxed while avoiding multiple integration. Since numerical values of the parameter estimates are similar for both estimators, we used parametric QML estimates because of their greater efficiency. These estimates were used to evaluate the environmental impacts of the European land set-aside policy for a panel of French farmers in the Meuse *Département* observed between 2006 and 2010. We calculated elasticities with respect to the set-aside subsidy for two indicators: pesticide and fertilizer demand, and pesticide and fertilizer demand intensity per unit of cultivated land. Our results show that a set-aside policy could provide farmers with incentives to intensify production on the remaining land, potentially leading to adverse environmental effects in terms of biodiversity loss and water pollution. Estimates of fertilizer and pesticide input demand were found to be low but significant, and higher for fertilizer than for pesticide. These estimates are numerically close to the estimated elasticity of fertilizer and pesticide input intensity with respect to a set-aside subsidy.

Environmental effects are likely to be more harmful if the input demand intensity indicator increases for a reduced (cultivated) area. We would point also to the fact that this indicator (II) measures only the potential environmental impacts of the policy, and not the actual impacts which are likely to depend on plot-level factors (soil type, slope, climate, distance to surface or groundwater, etc.).

Finally, we used our estimated elasticities to simulate the impacts of a 5 percent increase in the set-aside area. Our results reveal an increase in fertilizer and pesticide demand ranging from 1.87 to 2.11 percent and from 0.81 to 1.15 percent respectively. This chemical input intensification could have potentially adverse environmental effects. Taxes ranging from 36.9 to 41.63 percent on fertilizer and from 2.18 to 3.12 percent on pesticide would be necessary to offset these increases in fertilizer and pesticide demand.

Taxing fertilizer and pesticide use in agriculture to correct for market failures (externalities including water contamination and human poisoning) is generally considered a cost-effective policy in theory (as opposed to command-and-control policies). However, it is also a nonpoint source pollution issue, implying that second-best outcomes only may be achieved, using, e.g., indirect taxation (pesticide and fertilizer sales and not actual application rates, output level, land use, etc.) Revenues streams generated by an input tax can be earmarked to subsidize more sustainable agricultural practices. This implies that in principle, considering a policy consisting of complementing an input tax with a subsidy on set aside would correspond to the earmarking strategy above. However, the objective of the tax simulation considered here is to offset the increase in input demand following intensification, to illustrate the magnitude of the tax on fertilizer and pesticide as equivalent policies. In other words, the level of the input tax required to offset the negative consequences of a set-aside policy can also be interpreted as the tax level that would be necessary to reduce fertilizer and pesticide use by the same amount (as the increase in demand in the first place). We do not discuss the relative advantages of tax vs. area-based subsidies as policy instruments, which can be found, e.g., in (Sterner, 2003). Alternative instruments can be considered to correct for market failures associated with agriculture pollution (in this case, damages from non source pollution that are not internalized by farmers, as no market for rights exists regarding pollution externalities), each involving different characteristics in terms of cost-effectiveness, environmental efficiency, distortions due to revenue redistribution and acceptability (or participation performance) of policy instruments.

This analysis could be extended in several directions. First, the model could be improved by incorporating other policy instruments, for example, in the case of the recent European Common Agricultural Policy reform, the number of crops in rotation and the proportion of grassland area. Another extension would involve linking our production model to observed farm level environmental variables, such as water quality and biodiversity.

Regarding econometric considerations, several extensions could also be considered, starting with the correction of possible errors-in-variables (EIV) bias when deflating the price system by the price of seed as numeraire, as proposed by (**?**). Their estimation methods (IV and GMM) are different from ours, but the quasi-maximum likehood framework can accommodate a specific treatment of EIV, provided minimum knowledge on the source of measurement errors in unit prices is available. Furthermore, our estimation results are conditioned on our choice of a numeraire (seed price in our application) in the system of structural equations, although this choice does not modify the structure of the problem. To obtain results invariant to the choice of numeraire, an extension would be to consider a symmetric normalized quadratic profit function, as in (Kohli, 1993), however with an additional computational cost.[10] This is left for future research.

References

Arndt, C. (1999). Demand for herbicide in corn: An entropy approach using micro-level data. *Journal of Agricultural and Resource Economics* 24: 204–21.

Ball, V. E., Bureau, J.-C., Eakin, K. and Somwaru, A. (1997). Cap reform: modelling supply response subject to the land set-aside. *Agricultural Economics* 17: 277–288.

Boecker, T. and Finger, R. (2017). A meta-analysis on the elasticity of demand for pesticides. *Journal of Agricultural Economics* 68(2): 518–533.

Born, B. and Breitung, J. (2016). Testing for serial correlation in fixed-effects panel data models. *Econometric Reviews* 35: 1290–1316.

Boulanger, P. (2010). Distribution of agricultural support: selected French evidences. In *OECD Workshop on the Disaggregated Impacts of CAP Reform, March*, 10–11.

Britz, W., Heckelei, T. and Pérez, I. (2006). Effects of decoupling on land use: an eu wide, regionally differentiated analysis. *Agrarwirtschaft* 55: 215.

Chakir, R. and Thomas, A. (2003). Simulated maximum likelihood estimation of firm demand systems with corner solutions and panel data. *Revue d'Economie Politique* 6: 773–997.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics* 18: 5–46.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Heckman, J. and Leamer, E. (eds), *Handbook of Econometrics*. London: Elsevier, *6*, chap. 76.

Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association* 101: 1228–1240.

Choi, J.-S. and Helmberger, P. G. (1993). How sensitive are crop yields to price changes and farm programs? *Journal of Agricultural and Applied Economics* 25: 237–244.

European Commission (2013). Overview of cap reform 2014-2020. *Agricultural Policy Perspectives Brief* : 5–12.

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

Fraser, I. and Stevens, C. (2008). Nitrogen deposition and loss of biological diversity: Agricultural land retirement as a policy response. *Land Use Policy* 25: 455–463.

Fraser, I. and Waschik, R. (2005). Agricultural land retirement and slippage: lessons from an australian case study. *Land Economics* 81: 206–226.

Gallant, A. and Nychka, D. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica* 55: 363–390.

Golan, A., Perloff, J. and Shen, E. (2001). Estimating a demand system with nonnegativity constraints: Mexican meat demand. *The Review of Economics and Statistics* 83(3): 541–550.

Green, R. E., Cornell, S. J., Scharlemann, J. P. and Balmford, A. (2005). Farming and the fate of wild nature. *Science* 307: 550–555.

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

Hauck, J., Schleyer, C., Winkler, K. J. and Maes, J. (2014). Shades of greening: reviewing the impact of the new eu agricultural policy on ecosystem services. *Change and Adaptation in Socio-Ecological Systems* 1.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46 (6): 931–959.

Hendricks, N. P., Er, E. et al. (2018). Changes in cropland area in the united states and the role of crp. *Food Policy* 75: 15–23.

Hennessy, D. A. (2006). On monoculture and the structure of crop rotations. *American Journal of Agricultural Economics* 88: 900–914.

Kao, C., Lee, L.-F. and Pitt, M. (2001). Simulated maximum likelihood estimation of linear expenditure system with binding non-negativity constraints. *Annals of Economics and Finance* 2: 215–235.

Kohli, U. (1993). A symmetric normalized quadratic gnp function and the us demand for imports and supply of exports. *International Economic Review* 34: 243–255.

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

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

Lee, L.-F. and Pitt, M. M. (1986). Microeconometric demand system with binding nonnegativity constraints: The dual approach. *Econometrica* Vol. 54, No. 5.: 1237–1242.

Love, H. and Foster, W. (1990). Commodity program slippage rates. *Western Journal of Agricultural Economics* 15(2): 272–281.

Matthews, A. (2013). Greening agricultural payments in the eu' s common agricultural policy. *Bio-based and Applied Economics* 2: 1–27.

Meyerhoefer, C. D., Ranney, C. K. and Sahn, D. E. (2005). Consistent estimation of censored demand systems using panel data. *American Journal of Agricultural Economics* 87: 660–672.

Micchelli, C., Rivlin, T. and Winograd, S. (1976). The optimal recovery of smooth functions. *Numerische Mathematik* 26: 279–285.

Millimet, D. L. and Tchernis, R. (2008). Estimating high-dimensional demand systems in the presence of many binding non-negativity constraints. *Journal of Econometrics* 147: 384–395.

Moro, D. and Sckokai, P. (1999). Modelling the cap arable crop regime in italy: Degree of decoupling and impact of agenda 2000. *Cahiers d'Economie et Sociologie Rurales* 0: 49–73.

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

Paudel, K. P. and McIntosh, C. S. (2007). Numeraire choice in agricultural supply analysis. *Applied Economics* 37: 1209–1214.

Perali, F. and Chavas, J.-P. (2000). Estimation of censored demand equations from large cross-section data. *American Journal of Agricultural Economics* Volume 82, Issue 4: 1022–1037.

Platoni, S., Sckokai, P. and Moro, D. (2012a). A note on two-way ECM estimation of SUR systems on unbalanced panel data. *Econometric Reviews* 31: 119–141.

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

Pudney, S. (1989). *Modelling individual choice: The econometrics of corners, kinks and holes*. Basil Blackwell, Oxford.

Reichelderfer, K. and Boggess, W. G. (1988). Government decision making and program performance: The case of the conservation reserve program. *American Journal of Agricultural Economics* 70: 1–11, doi:10.2307/1241970.

Rozakis, S. and Sourie, J.-C. (2005). Micro-economic modelling of biofuel system in france to determine tax exemption policy under uncertainty. *Energy policy* 33: 171–182.

Rygnestad, H. and Fraser, R. (1996). Land heterogeneity and the effectiveness of cap set-aside. *Journal of Agricultural Economics* 47: 255–260.

Schwiebert, J. (2013). Sieve maximum likelihood estimation of a copula-based sample selection model, university of Hannover. Working paper.

Shonkwiler, J. and Yen, S. (1999). Two-step estimation of a censored system of equations. *American Journal of Agricultural Economics* 81(4): 972–82.

Sterner, T. (2003). *Policy instruments for environmental and natural resource management*. Resources for the Future, Washington, DC.

Thomas, A. (2003). A dynamic model of on-farm integrated nitrogen management. *European review of agricultural economics* 30: 439–460.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* 26: 24–36.

Vannini, L., Gentile, E., Bruni, M., Loi, A., Aragrande, M., Theuvsen, L., Haman, K., Venturi, G., Monti, A., Canali, M. and Bernini, C. (2008). Evaluation of the Set Aside Measure 2000 to 2006: Final Report. Tech. rep.

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

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics* 68: 115 – 132, doi:http://dx.doi.org/10.1016/0304-4076(94)01645-G.

Wu, J. (2000). Slippage effects of the conservation reserve program. *American Journal of Agricultural Economics* 82(4): 979–992.

Yen, S. T. and Lin, B.-H. (2006). A sample selection approach to censored demand systems. *American Journal of Agricultural Economics* 88: 742–749.

Yen, S. T., Lin, B.-H. and Smallwood, D. M. (2003). Quasi-and simulated-likelihood approaches to censored demand systems: food consumption by food stamp recipients in the united states. *American Journal of Agricultural Economics* 85: 458–478.

Appendix

The semi nonparametric estimator

Denote $\hat{\gamma}_N = (\hat{\theta}_N, \hat{f}_N)$ the estimate of the true parameter $\gamma_0 = (\theta_0, f_0) \in \Gamma = \mathscr{T} \times \mathscr{F}$, defined as

$$\hat{\gamma}_N = \text{argmax}_{(\theta,f)\in\Gamma} \ln L(\gamma;Z) = \frac{1}{N}\sum_{i=1}^{N} l(\gamma, f, Z_i), \tag{15}$$

where $l(.)$ is the likelihood of observation i and Z_i is the vector of observed variables, $Z_i = (y_i, x_i)$.

We follow Chen (2007) in making the following assumptions:

Assumption 1. $Z_i = \{y_i, x_i\}_{i=1}^N$ is i.i.d.

Assumption 2. h_i and x_i are independent.

Assumption 3. $E \ln L(\gamma;Z)$ is continuous at γ_0 and $E \ln L(\gamma_0;Z) > -\infty$.

Assumption 4. For all $\varepsilon > 0$, $E \ln L(\gamma_0, Z) > \sup_{\theta\in\Theta; d(\gamma,\gamma_0)\geq\varepsilon} E \ln L(\gamma;Z)$.

Assumption 5. $\Gamma_k \subseteq \Gamma_{k+1} \subseteq \Gamma$ for all $k \geq 1$, and for any $\gamma \in \Gamma$ there exists a sequence $\pi_k\gamma_0 \in \Gamma_k$ such that $d(\gamma_0, \pi_k\gamma_0) \to 0$ as $k \to \infty$.

Assumption 6. For each $k \geq 1$, $\ln L(\gamma;Z)$ is a measurable function of Z for all $\gamma \in \Gamma_k$, and for any Z, $\ln L(\gamma;Z)$ is upper semicontinuous on Γ_k for metric $d(.,.)$.

Assumption 7. The sieve spaces Γ_k are compact under metric $d(.,.)$.

Assumption 8. For all $k \geq 1$, $\text{plim}_{N\to\infty} \sup_{\gamma\in\Gamma_K} |\ln L(\gamma) - E \ln L(\gamma)| = 0$.

Under assumptions 1-8, Chen (2007) shows that $d(\hat{\gamma}, \gamma_0) = o_P(1)$.

In practice, consistency and efficiency properties of semi-nonparametric estimators depend on the type of sieve chosen. In particular, a sieve based on Hermite polynomials is preferred if density functions are approximated under thin tails or non-compact support (the cdf being obtained by integration). In our case, we do not follow such route and start directly from the approximation of the cdf on a compact support, the density being computed from the derivation with respect to its arguments. This allows us to use simpler forms of sieve, as the linear B-spline.

In practice, we construct a series of splines as continuous piecewise polynomial approximations of pdfs and cdfs. A spline is defined as follows. Consider a sample of observation points $x_0, \ldots, x_N$ over the interval $[a,b]$ such that $a = x_0 < x_1 < \cdots < x_N = b$. A function $S(x)$ that satisfies

(C1) for each interval (x_i, x_{i+1}), $\frac{d^k S(x)}{dx} = 0$,

(C2) $S(x) \in C^{k-2}[a,b]$,

is defined as a spline function of degree $k-1$ for discrete points called knots. It is a polynomial of degree $k-1$ that is defined specifically for each interval (x_i, x_{i+1}), with derivatives up to degree $k-2$ that are continuous over $[a,b]$. There are several ways to represent splines functions (e.g., by the truncated power function), but the B-spline representation is usually preferred as it is more stable numerically in most applications.

Consider now a series of knots $\{t_r\}$ such that

$$t_{-k+1} \leq t_{-k+2} \leq \cdots \leq t_0 = x_0 < t_1 = x_1 < t_N = x_N \leq t_{N+1} \leq t_{N+2} \leq \cdots \leq t_{N+K+1},$$

and the function $g_k(t,x)$ as

$$g_k(t,x) = \begin{cases} (t-x)^{k-1} & \text{if } t \geq x, \\ 0 & \text{if } t < x. \end{cases}$$

The normalized basis spline (or simply B-spline) of degree $k-1$ is then defined as

$$N_{j,k}(x) = \left(t_{j+k} - t_j\right) g_k\left[t_j, t_{j+1}, \ldots, t_{j+k}, x\right],$$

and is such that $N_{j,k}(x) = 0$ when $x \leq t_j$ or when $x \geq t_{j+k}$, and when $j \leq i-k$ or when $j \geq i+1$. It can be shown that any spline function $S(x)$ satisfying conditions (C1) and (C2) can be represented by

$$S(x) = \sum_{j=-k+1}^{N-1} c_j N_{j,k}(x),$$

where the full expression of $N_{j,k}(x)$ is

$$N_{j,k}(x) = \left(t_{j+k} - t_j\right) = \sum_{r=j}^{j+k} \frac{(t_r - x)^{k-1}}{(t_r - t_j)\ldots(t_r - t_{r-1})\ldots(t_r - t_{j+k})},$$

and c_j, $j = -k+1, \ldots, N-1$ are constant parameters. The function value, its derivatives and integrals can all be computed from non-zero elements of $N_{j,k}(x)$, which are calculated recursively, using

$$N_{r,s}(x) = \frac{x - t_r}{t_{r+s-1-t_r}} N_{r,s-1}(x) + \frac{t_{r+s} - x}{t_{r+s} - t_{r+1}} N_{r+1,s-1}(x),$$

where $N_{r,1}(x) = 1$ if $r = i$ and 0 otherwise.

The bivariate spline and B-spline functions are defined as extensions to the above. For example, the bivariate spline function $S(x,y)$ must satisfy

(C3) for each open region $R_{i,j} = \left\{(x,y)|x_i < x < x_{i+1}, y_j < y < y_{j+1}\right\}$, $\frac{d^k S(x,y)}{dx} = \frac{d^k S(x,y)}{dy} = 0$,

(C4) $S(x,y) \in C^{k-2,k-2}[R]$,

where R is a closed region defined as $R = \{(x,y)|a \leq x \leq b, c \leq y \leq d\}$ on the plane $(x-y)$.

We first construct a "not-a-knot" spline knot sequence as an even partition of the support of random error terms, which is appropriate for interpolation of data by splines of order k. The vector t contains the knot sequence in its first $N+k$ positions. If k is even and we assume that the entries in the input vector x are increasing, then t is returned as

$$\begin{cases} t_i = & x_1 & \text{for } i = 1, \ldots, k, \\ t_i = & x_{i-k/2} & \text{for } i = k+1, \ldots, N, \\ t_i = & x_N + \varepsilon & \text{for } i = N+1, \ldots, N+k, \end{cases}$$

where ε is a small positive constant.

If k is odd, then t is returned as

$$\begin{cases} t_i = & x_1 & \text{for } i = 1, ...k, \\ t_i = & \left(x_{i-(k-1)/2} + x_{i-1-(k-1)/2}\right), & \text{for } i = k+1, ..., N, \\ t_i = & x_N + \varepsilon & \text{for } i = N+1, ..., N+k. \end{cases}$$

We choose the order of the spline $k = 4$. We also experimented with the algorithm of Micchelli et al. (1976) to compute the optimal knot sequence of the spline, based on the number of observations and the order of the spline. This algorithm minimizes the constant in the error term:

$$||f - s|| \leq c||f^{(k)}||, \tag{16}$$

where f is the function to evaluate and s is the spline interpolant of f at point x. Results were not significantly different from the ones obtained with the simpler "not-a-knot" rule above.

Table 1: Descriptive statistics

Variable	Mean	Std. dev.	Proportion > 0
Cereal output (tons)	532.3851	342.0180	0.9985
Rapeseed output (tons)	88.4597	78.2572	0.8332
Diester output (tons)	11.0806	20.9001	0.3133
Fertilizer input (kg)	590.4984	200.7343	0.9911
Pesticide input	163.5981	118.3055	0.9916
Cereal area (ha)	19.7412	12.2209	0.9995
Rapeseed area (ha)	6.4915	5.4881	0.8342
Diester area (ha)	0.8473	1.5709	0.3153
Voluntary land set-aside (ha)	1.0280	1.6751	0.7066
Cereal price	1.0207	0.3177	
Rapeseed price	2.2857	0.5206	
Diester price	1.2061	0.9199	
Fertilizer price	0.2756	0.1274	
Pesticide price	0.9552	0.0306	
Cereal subsidy	5.4009	2.7268	
Rapeseed subsidy	5.9490	1.3463	
Diester subsidy	5.1049	2.7187	
Land set-aside subsidy	8.6846	2.8985	

Notes. 2014 observations. Price and subsidy variables (in Euro) are normalized by the unit price of seed. Crop outputs are in tons, fertilizer input is in kg, and pesticide input is computed from pesticide expenditure divided by its price index. Cereal is the combination of wheat and barley crops.

Table 2: Environmental indicators, by crop combination

		Indicator			
Cropping system	N	x_F	x_P	II_F	II_P
C only	312	270.77	34.52	17.48	1.82
		(118.13)	(35.40)	(7.04)	(1.06)
$C+R$ only	1067	585.08	188.13	17.01	4.52
		(111.64)	(114.21)	(6.43)	(1.27)
$C+D$ only	21	516.06	100.86	24.61	4.13
		(122.38)	(74.33)	(7.05)	(1.64)
$C+R+D$	610	767.75	189.44	22.79	4.72
		(136.25)	(110.05)	(8.41)	(1.37)

Notes. N denotes the number of observations. x_F, x_P, II_F and II_P respectively denote input demand for fertilizer and pesticide, and intensity of input use for fertilizer and pesticide. Cropping systems are defined as combinations of cereals (C), rapeseed (R) and diester (D). Standard deviations are in parentheses.

Table 3: Goodness-of-fit measures for parametric and semi nonparametric estimators

Structural equation	Estimation method	R^2 (sample)	R^2 (obs.>0)	HR test statistic	Correct predictions (percent)	Predicted 1s as 0s (percent of 1s)	Predicted 0s as 1s (percent of 0s)
q cereal	PQML	0.9157	0.9192	-0.90 (0.36)	99.85	0.00	0.15
	SNPQML	0.9102	0.9144	0.92 (0.35)	99.85	0.00	0.15
q rapeseed	PQML	0.7661	0.7369	0.46 (0.65)	83.32	0.00	100.00
	SNPQML	0.7544	0.7346	0.30 (0.76)	83.12	5.18	75.30
q diester	PQML	0.5093	0.3193	0.78 (0.43)	79.69	55.63	4.19
	SNPQML	0.4101	0.4243	1.61 (0.10)	82.22	40.25	7.52
x ferti.	PQML	0.5225	0.5445	0.61 (0.54)	99.26	0.25	55.6
	SNPQML	0.4147	0.4331	0.71 (0.47)	99.60	0.20	22.22
x pest.	PQML	0.8304	0.8376	2.47 (0.01)	99.16	0.00	100.0
	SNPQML	0.8309	0.8383	2.72 (0.00)	99.16	0.00	100.0
l cereal	PQML	0.9617	0.9607	0.25 (0.80)	99.95	0.00	100.0
	SNPQML	0.9617	0.9616	1.31 (0.19)	99.95	0.00	100.0
l rapeseed	PQML	0.8195	0.8001	1.00 (0.31)	83.42	0.00	100.0
	SNPQML	0.8192	0.8075	0.83 (0.40)	83.42	4.88	75.45
l diester	PQML	0.5288	0.3445	0.34 (0.73)	79.15	55.12	5.08
	SNPQML	0.5123	0.4332	1.98 (0.05)	80.73	49.76	5.22
l setaside	PQML	0.1232	0.1337	-2.13 (0.03)	72.54	19.82	45.85
	SNPQML	0.1248	0.1066	-1.75 (0.08)	77.26	6.61	61.59

Notes. 2014 observations. PQML and SNPQML respectively denote Parametric and Semi Nonparametric Quasi-Maximum Likelihood estimation. HR test statistic (p-value in parentheses) is the Heteroskedasticity-Robust test for serial correlation in panel data of Born and Breitung (2016). The proportion of correct prediction is the sum of sensitivity (correctly predicted positive outcomes) and specificity (correctly predicted negative outcomes) measures. The proportion of positive outcomes (resp., negative) wrongly predicted as negative (resp., positive) outcomes is computed with respect to the number of positive (resp., negative) outcomes. The last three columns are goodness-of-fit measures designed for discrete outcomes computed form estimated probabilities.

Notes

[1]See :`http://capreform.eu/assessment-of-the-commission's-proposal-for-an-obligatory-set-aside-programme/`

[2]Coherency conditions in a structural econometric modeling framework are conditions that ensure that model is correctly specified from the theoretical point of view. For example, **?** show a negative semi-definite parameter matrix is necessary and sufficient for the translog cost function to be concave. In most econometric models estimating a cost function imposing the semi-negativity of the matrix of the parameters ensure that the coherency conditions to be respected.

[3]Imposing positive semi-definiteness on a matrix B is equivalent to writing $B = AA'$ where A is a lower triangular matrix of the same dimension as B.

[4]A parametric version of this estimator has been proposed by Yen et al. (2003) and applied to land use decisions by Fezzi and Bateman (2011).

Table 4: Parametric QML elasticity estimates with fixed effects

	p cereal	p rapeseed	p diester	ferti. w	pest. w	τ cereal	τ rapeseed	τ diester	τ set-aside
p cereal	0.0096***	0.0353***	-0.0707***	-0.0117***	0.0120***	0.0116***	0.0239***	-0.0790***	0.0259***
	(0.0009)	(0.0015)	(0.0035)	(0.0008)	(0.0006)	(0.0012)	(0.0011)	(0.0037)	(0.0013)
p rapeseed		0.1418***	-0.3020***	-0.0555***	0.0548***	0.0280***	0.0883***	-0.3356***	0.1249***
		(0.0042)	(0.0048)	(0.0010)	(0.0014)	(0.0034)	(0.0009)	(0.0059)	(0.0020)
p diester			0.6714***	0.1261***	-0.0958***	-0.0293***	-0.1760***	0.7460***	-0.2987***
			(0.0023)	(0.0028)	(0.0034)	(0.0058)	(0.0006)	(0.0014)	(0.0046)
ferti. w				0.0507***	-0.0394 ***	0.0010	-0.0350***	0.1200***	-0.0589***
				(0.0013)	(0.0026)	(0.0014)	(0.0012)	(0.0030)	(0.0018)
pest. w					0.3700***	0.0498***	0.0316***	-0.1071***	-0.0290***
					(0.0032)	(0.0037)	(0.0037)	(0.0020)	(0.0026)
τ cereal						0.0537***	0.0022	-0.0448***	-0.0104***
						(0.0017)	(0.0027)	(0.0044)	(0.0015)
τ rapeseed							0.1475***	-0.2180***	0.0638***
							(0.0011)	(0.0014)	(0.0011)
τ diester								0.9046***	-0.3314***
								0.0018	(0.0054)
τ set-aside									0.1479***
									(0.0044)

Notes. 2014 observations. p, w and τ respectively denote output prices, input prices and area-based subsidy rates, and ferti. and pest. denote fertilizer and pesticide inputs. Robust standard errors are in parentheses. Dependent variables are in rows, price/subsidy in columns. Elasticities of inputs are computed with the minus sign. *, ** and *** respectively denote parameter significance at the 10, 5 and 1 percent level.

Table 5: Semi nonparametric QML elasticity estimates with fixed effects

	p cereal	p rapeseed	p diester	ferti. w	pest. w	τ cereal	τ rapeseed	τ diester	τ set-aside
p cereal	0.0100***	0.0371***	-0.0760***	-0.0118	0.0128	0.0132	0.0235	-0.0773***	0.0290**
	(0.0011)	(0.0046)	(0.0118)	(0.0128)	(0.0105)	(0.0101)	(0.0183)	(0.0163)	(0.0103)
p rapeseed		0.1542***	-0.3288***	-0.0630	0.0598	0.0267	0.0937	-0.3406***	0.1376***
		(0.0402)	(0.0633)	(0.0563)	(0.0563)	(0.0527)	(0.0987)	(0.0491)	(0.0320)
p diester			0.7126***	0.1403	-0.1024	-0.0341	-0.1962	0.7444***	-0.3077**
			(0.1891)	(0.0908)	(0.0656)	(0.1093)	(0.1638)	(0.1957)	(0.1556)
ferti. w				0.0516	-0.0355	0.0085	-0.0403	0.1354	-0.0655**
				(0.0563)	(0.0604)	(0.0193)	(0.0776)	(0.1119)	(0.0292)
pest. w					0.3746***	0.0552	0.0361	-0.1090*	-0.0298
					(0.1110)	(0.0557)	(0.0967)	(0.0574)	(0.0771)
τ cereal						0.0717	-0.0075	-0.0343	-0.0059
						(0.0452)	(0.0223)	(0.1430)	(0.0369)
τ rapeseed							0.1404	-0.2210	0.0790*
							(0.1057)	(0.1693)	(0.0421)
τ diester								0.8452***	-0.3266*
								(0.1060)	(0.1367)
τ set-aside									0.1412**
									(0.0954)

Notes. 2014 observations. p, w and τ respectively denote output prices, input prices and area-based subsidy rates, and ferti. and pest. denote fertilizer and pesticide inputs. Robust standard errors are in parentheses. Dependent variables are in rows, price/subsidy in columns. Elasticities of inputs are computed with the minus sign. *, ** and *** respectively denote parameter significance at the 10, 5 and 1 percent level.

Table 6: Elasticities of fertilizer and pesticide demand and intensity with respect to unit set-aside subsidy

	$\varepsilon_{x_{fertilizer}\overline{\tau}_s}$	$\varepsilon_{x_{pesticide}\overline{\tau}_s}$	$\varepsilon_{l_{setaside}\overline{\tau}_s}$	$\varepsilon_{II_f\overline{\tau}_s}$	$\varepsilon_{II_p\overline{\tau}_s}$
PQML	0.0589***	0.0290***	0.1479***	0.0645***	0.0346***
	(0.0018)	(0.0026)	(0.0044)	(0.0018)	(0.0026)
SNPQML	0.0655**	0.0298	0.1412	0.0709**	0.0352
	(0.0292)	(0.0771)	(0.0954)	(0.0294)	(0.0772)

Notes. $\varepsilon_{x_{fertilizer}\overline{\tau}_s}$ and $\varepsilon_{x_{pesticide}\overline{\tau}_s}$ denote respecrively the elasticity of fertilizer and pesticide demand with respect to set-aside subsidy. $\varepsilon_{l_{setaside}\overline{\tau}_s}$ denote the elasticity of set-aside area with respect to the seubsidy set-aside. $\varepsilon_{II_f\overline{\tau}_s}$ and $\varepsilon_{II_p\overline{\tau}_s}$ denote respectively the elasticity of fertilizer and pesticide intensity with respect to the set-aside subsidy. PQML and SNPQML respectively denote Parametric and Semi Nonparametric Quasi-Maximum Likelhood estimation. Robust standard errors are in parentheses (computed with the Delta method for input intensity elasticities). *, ** and *** respectively denote parameter significance at the 10, 5 and 1 percent level.

[5]This more flexible estimator is in general less efficient if the restrictions implied by the less flexible estimator are valid, such as, e.g., normality or homoskedasticity.

[6]In our sample, fertilizer and pesticide account on average for respectively 9.49 and 8.25 percent of operating costs, compared with 9.06 percent for wage labor, 4.18 percent for fuel and 3.73 percent for seed.

[7]The Tornqvist price index of input j for farm i and year t is computed as follows: $\log p_{ijt} = (w_{ijt} + w_{j0})\log(p_{jt}/p_{j0})$, where w_{ijt} and w_{j0} respectively denote the cost share of input j for farm i at year t and for baseline year 0 ; p_{jt} and p_{j0} denote the unit input price of input j for year t and for baseline period 0, respectively.

[8]As noted above, to avoid possible bias due to unobserved, farm-specific individual effects, we do not consider the random-effects version of the PQML model.

[9] We make the distinction between the "environmental indicators" introduced above, which are defined in levels and ratios, and the "sensitivity indicators" that are computed from them and are unit-free.

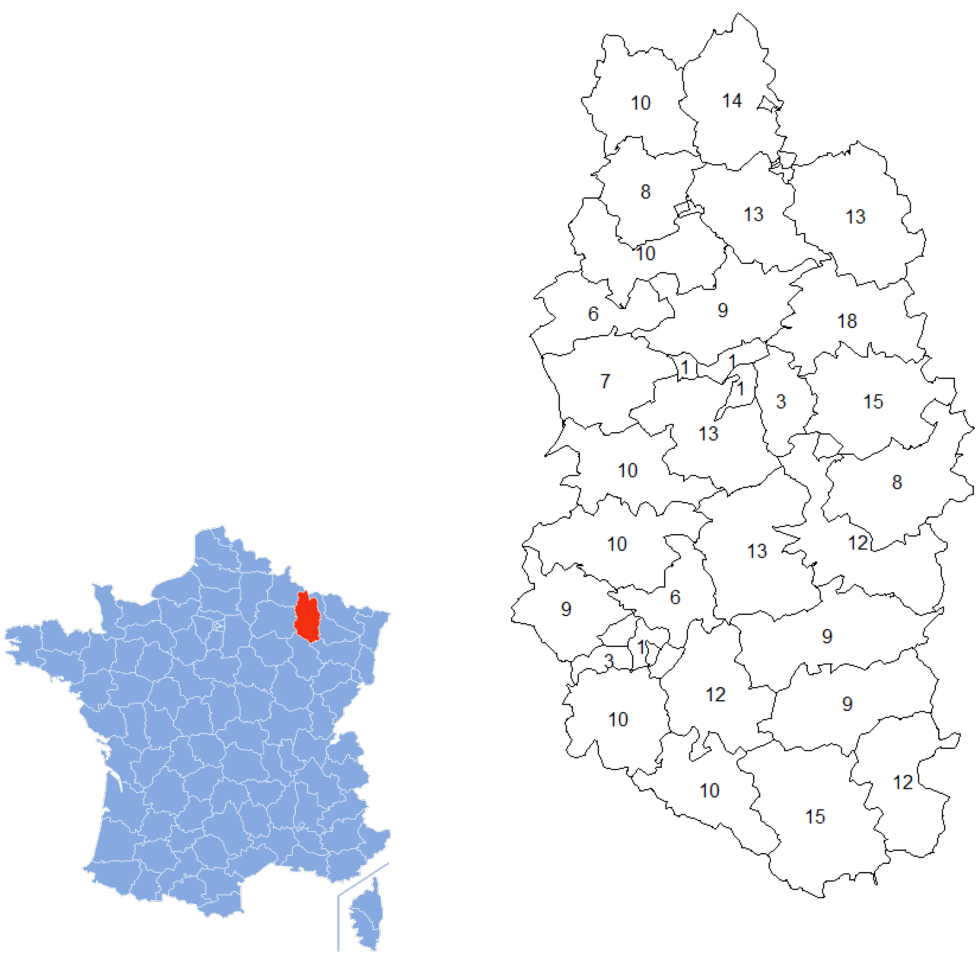
[10](Paudel and McIntosh, 2007) show that estimation results do not always vary significantly with the choice of the numeraire.

Table 7: Simulation results of a 5 percent increase in set-aside area

	(1) $\frac{\partial \overline{\tau}_s}{\overline{\tau}_s}$	(2) $\frac{\partial x_f}{x_f}$	(3) $\frac{\partial x_P}{x_P}$	(4) $\frac{\partial w_f}{w_f}$	(5) $\frac{\partial w_p}{w_p}$
PQML	0.3381	0.0199	0.0098	0.3927	0.0265
		[0.0187;0.0211]	[0.0081;0.0115]	[0.3690;0.4163]	[0.0218;0.0312]
SNPQML	0.3541	0.0232	0.0106	0.4495	0.0282
		[0.0029;0.0435]	[-0.0430;0.0641]	[0.0567;0.8423]	[-0.1147;0.1710]

Notes. $\frac{\partial \overline{\tau}_s}{\overline{\tau}_s}$ denote the variation of set-aside subsidy in percent. $\frac{\partial x_f}{x_f}$ and $\frac{\partial x_P}{x_P}$ denote respectively the fertilizer and pesticide demand variation in percent, $\frac{\partial w_f}{w_f}$ and $\frac{\partial w_p}{w_p}$ denote respectively the price of fertilizer and pesticide variation in percent. 95 percent confidence intervals obtained from standard errors of parameter estimates are presented in brackets. PQML and SNPQML respectively denote Parametric and Semi Nonparametric Quasi-Maximum Likelihood estimation.

Figure 1: Location of sample farms in *cantons* of the Meuse *département*



Note. Numbers for each *canton* represent the total number of farms in the sample.