Definition and Spatial Patterns of Farmland Quality: The 1992 CAP MacSharry Reforms Revisited

AY Jean-Sauveur
INRA, UMR 1041 CESAER
26 Boulevard du Docteur PetitJean 21079 Dijon Cedex

2èmes journées de recherches en sciences sociales

INRA SFER CIRAD

11 & 12 décembre 2008 – LILLE, France

RÉSUMÉ

La réforme de la PAC de 1992 instaure, pour les exploitations dépassant une certaine taille, l'obligation de retirer de la production (geler) une proportion des terres auparavant cultivées en céréales, oléagineux et protéagineux. Une particularité de ce volet de la réforme tient au fait que le gel, bien qu'obligatoire, peut être mis en lien avec les incitations sous-jacentes en terme de paiements compensatoires. Dans le cadre conceptuel d'un ensemble d'exploitations qui disposent de dotations en terres hétérogènes, cet article propose une modélisation des choix productifs issus de la réforme. L'application de ce modèle aux exploitations spécialisées du bassin parisien français, observées avant et après la réforme nous permet d'aboutir sur trois résultats principaux : (i) la non-neutralité de la réforme sur les choix de production, (ii) une certaine forme d'efficacité des incitations et (iii) l'existence d'une composante spatiale à son application.

ABSTRACT

The objectives of this paper are twofold. First, from a theoretical perspective, a microeconomic model is built to propose an explanation to the fact that European arable producers do not claim arable area payments on all eligible land due to 1992 MacSharry reforms of Common Agricultural Policy (CAP.) We will show this observation can be linked to the heterogeneity of individual land endowments by the way of an economic definition of land quality. Secondly, with an application on crop-specialized French farms observed before and after the reforms, we use this legislative shock to identify spatial distribution of these farmland qualities. This leads us to conclude on three principal results : (i) the reforms are not neutral on production choices (ii) the incentive scheme, based on payments claiming, constitutes a certain form of efficiency (iii) there is a fine spatial structure of reforms' applications.

Je remercie Jérôme Bourdieu, Jean Cavailhès, Cécile Detang-Dessendre, Julie LeGallo, Chantal LeMouël, Murielle Roger et Harris Selod pour leur aide lors de ce travail.

Definition and spatial pattern of farmland quality: The 1992 CAP MacSharry reforms revisited

This article proposes an economic definition of farmland quality, an original modelling of 1992 Common Agricultural Policy reforms and an empirical strategy to identify land quality spatial pattern in the *Bassin Parisien*, a large area around Paris (France.) The need for such a definition for economists stems from the fact that land quality differentials are major causes of land scarcity because all farmland are not equally suited for crop production. Land quality and spatial heterogeneity are also important topics in relation to land allocations among alternative uses (Nelson, 1992), spatial targeting of agricultural policies (Lankoski and Ollikainen, 2003) and environmental considerations (Bockstael, 1996). The concept of land heterogeneity is not new in economics, which is evident from Ricardo's (1815) and von Thünen's (1826) works on intrinsic (fertility) and extrinsic (location) attributes of farmlands, respectively. More recently, some conceptual frameworks to deal with land quality in agriculture have been proposed: Caswell and Zilberman (1986), Lichtenberg (1989), Palmquist (1989), Stavins and Jaffe (1990), Plantinga (1996), Hardie and Parks (1997), Lankoski and Ollikainen (2003). The definition in this paper is directly inspired by these works but a new relation between land quality and spatial heterogeneity enables a conceptual framework that is more general. Another benefit of the present definition is that it enables easier empirical application of the individual data.

Following Stavins and Jaffe (1990) and Stavins (1999), we consider land quality as not directly observable;[1] in turn we suppose that it can be partially deducted from the farmer's behaviour. This 'revealed-preference' approach is based on short run farmer reactions to the MacSharry reforms to the Common Agricultural Policy (CAP.) More precisely, we analyse these reforms as exogenous shocks with important consequences for European farmland allocation: during the 1993 crop year, the first year of the new scheme, 4.7 million hectares of arable land in the European Union (9.6% of the total arable base area) was withdrawn from agricultural production (Roberts et al., 1996). This land use reallocation is a consequence of what is often referred to as compulsory set-aside in general arrangements (Guyomard et al., 1996; Moro and Sckokai, 1999) and was accompagned with area payments to compensate for the substantial reductions in the guaranteed prices of grains and other arable crops (cereals, oilseeds, protein crops.) However, despite the term 'compulsory,' precise review of the reforms shows the fact that set-aside obligations in terms of individual acreages depend on the arable area payments (compensatory payments) claimed. As Colman and Vavra (2002) note for

[1] There are several reasons for the unobservable nature of land quality: unobservable and holistic attributes, simultaneity with land use, unknown aggregation rule, endogeneity due to individual practice, etc.

'professional' farms (from general arrangements) in the United Kingdom, arable producers do not often claim arable area payments for all eligible planted crops. In other words, being a professional farm is an endogenous choice, depending on area claims for compensatory payments. Therefore, the majority of theoretical frameworks that consider exogenous set-aside rates or exogenous definitions of general arrangements (Guyomard et al., 1996; Fraser, 1997; Ball et al., 1997; Moro and Sckokai, 1999; Serra et al., 2005) are not truly representative of the microeconomic behaviour of farmers when adverse selection effects are present. Moreover, following Bourgeon et al. (1995), the incentive structure of set-aside can be modelled in relation to economic land quality endowments or (indistinctly) opportunity cost of land diversion.

The originality of our approach is in constructing a model explaining observed set-aside rates of specialized farms and providing an economic interpretation of the findings. It also constitutes a new methodology to describe the spatial heterogeneity of agricultural activity based on short term reactions following an unanticipated event. Such an empirical strategy, as we see below, allows us to control for most of the holism, simultaneity, endogeneity or time dependence problems involved in identifying an economic definition of land quality. Many index of land quality are already produced at different scales (FAO (1991), USDA (1973), etc.) but they all come from observable attributes of land and serve usually as proxy for econometric analysis (Bell et al., 2006). The paper is organized as follows. Section 1 presents the theoretical model and introduces the concept of land quality in relation to economic behaviour. Section 2 introduces the CAP MacSharry reform and analytically determines short run reactions to them. Section 3 describes the empirical methodology to identify a parameter in relation to individual endowments of land qualities. The data are described in Section 4 and Section 5 presents the results. Section 6 concludes.

1. Theoretical model

Consider a farm i producing a unique agricultural good, so that it belongs to a set of others specialized farms I. The production process requires exclusive usage of land that could be allocated to an alternative use. Land endowment[2] is heterogeneous and represented by a scalar measure q for every i's land units. Following Lichtenberg (1989), we firstly call this scalar land quality 'understood as a vector of attributes affecting productivity, p.188' (fertility, water-holding capacity, topography, etc.), in sum, agronomic quality. A cumulative function $G_i(q)$ defined on its one-dimensional support Q_i stands for the differentiation of individual land quality endowments among farms. The set Q_i includes q's continuum, with a lower bound $\underline{q_i}$ and an upper bound $\overline{q_i}$. Based on the classical properties of cumulative functions, farm i has at its disposal $\overline{T_i}G_i(q)ha$ of land of quality at most q and $\overline{T_i}G_i^{'}(q) = \overline{T_i}g_i(q)ha$ of quality q, with $g_i(q)$ being the continuous and differentiable probability density function. To simplify the

[2] Available total acreage (land endowment) is assumed to fixed at surface $\overline{T}_i$ *hectare*.

notations, the subscript i is temporarily dropped out, but we still consider the case of a single farm within the specialized set I.

Two types of inputs are utilized by the farm : quasi-fixed inputs and variable inputs. The quantities of the first (capital, labor, etc.) are chosen at farm level and used on cropland. The quantities of the second (fertilizers, irrigation, etc.) depend on the quality of the land to which they are applied. At the intensive margin (for a land unit), variable profit (equivalent to per-unit of land profit) withdrawn from non-random agricultural activity can be denoted as follows:

$$\pi(p,\mathbf{x},e(q),\mathbf{c}(q),c,q) = pf(\mathbf{x},e(q),q) - \mathbf{c}(q)\mathbf{x}^T - e(q)c \;, \;\; \forall q \in Q \qquad (1)$$

where $f()$ is a neoclassical production function per unity of land, convex and non decreasing. $\mathbf{x}^T$ is the transposition of the quasi-fixed inputs vector per unity of land and the values contained represent the availabilities that can be allocated to the land units according to the quasi-fixed input marginal cost function $\mathbf{c}(q)$. The marginal costs negatively depend on q ($\mathbf{c}^{'}(q) \leqslant 0$) if one considers that land quality positively influences exploitability (based on factors such as degree of difficulty in working it, distance from the main farm holding or parcel's size.) $e(q)$ is the quantity of the variable input,[3] continuous and twice differentiable which could be increasing or decreasing with land quality; c is the associated marginal cost. p is the unit price of the homogeneous output produced and this variable profit function is defined for every land quality q on the farm. For profit maximization at the intensive margin, quasi-fixed inputs can be considered exogenous, consequently we can write the first order conditions for each unity of land. These first order conditions represent the optimal choices for variable input:

$$e^*(q) \text{ such as } pf_e^{'}(\mathbf{x},e^*(q),q) = c \;, \;\; \forall q \in Q \qquad (2)$$

The second order condition is fulfilled by marginal productivity decrease. For one land unit, the variable input demand is such that marginal cost equals marginal benefit, other things being equal. Under this hypothesis of variable input optimal use, coupled with the knowledge of quasi-fixed inputs marginal cost function and constant return to scale from land, the variable profit is wholly defined by the land quality considered. Even if the effect of land quality is conveyed through different channels (agronomic productivity, fertilizer use, exploitability), this result leads to unequivocal specification of the link between land quality and profitability outcomes. Equation (3) defines what we can call the *economic quality* of a land unit, denoted as θ:

$$\pi(p,\mathbf{x},e^*(q),\mathbf{c}(q),c,q) = \theta \int_{\underline{q}}^{\overline{q}} \pi(p,\mathbf{x},e^*(q),\mathbf{c}(q),c,q)g(q)\,\mathrm{d}q \equiv \theta\overline{\pi}^*(\mathbf{x}) \;, \;\; \forall \theta \in \Theta \qquad (3)$$

This analytical definition allows us to present variable profit as the product combination of two components : the intrinsic attributes of the farmland and the farm's economic structure. Economic land quality is a finite scale-parameter which operates a cardinal classification of

[3] Extending the model to a vector of variable inputs is straightforward, and is achieved by multiplication of the first order conditions for intensive margin.

variable profits from the farm's average potential variable profit. Such a definition, which is conditional on choices made at the intensive margin by rational farms, enables us to distinguish between site-specific components and farm-specific components. The parallel between economic land quality and spatial heterogeneity is thus established. The conditional density function of the parameter is $h(\theta)$, and is differentiable and defined on the continuous set Θ closed and compact. A land unit which is of average economic quality is such that $\theta = 1$.

Such formulation integrates choices at the intensive margin, the next step is the extensive margin resoluteness, i.e. the allocation of land endowments between alternative uses. This involves a trade-off between agricultural activity and *other uses*. Considering positive and constant profit withdrawn from other uses, they could be considered as the stewardship value of the land (recreational activities, amenities, nonproductive woodland, etc. Lichtenberg (2002)) The assumption of independence between other uses and land quality can be took up (Lankoski and Ollikainen, 2003) but it is not essential in a mono-product case. Integrating the variable profits with available land qualities leads the following formulation for the total profit:

$$\Pi = \max_{\mathbf{x},\{L(\theta)\}} \left\{ \overline{T} \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta \overline{\pi}^*(\mathbf{x}) L(\theta) + \pi^a (1 - L(\theta)) \right] h(\theta) \, \mathrm{d}\theta \right\} \tag{4}$$

For an economic land quality, $L(\theta)$ represents the proportion of the acreage allocated to agricultural activity. This program enables a continuum of first order conditions since the farm has a continuum of available land qualities. Considering the (increasing) linear relation between variable profit and economic land quality (3), agricultural use represents a compact subset of Θ (Lichtenberg, 1989). So, we can establish a correspondence between the continuum $\{L(\theta)\}_{\theta\in\Theta}$ and a threshold land quality for all $\theta \in \Theta$:

$$\begin{aligned} L(\theta) &= 1 \Leftrightarrow \theta \geqslant \theta^p \\ L(\theta) &= 0 \Leftrightarrow \theta < \theta^p \quad \textit{(corollary)} \end{aligned} \tag{5}$$

Threshold value is denoted θ^p, an upper land quality is allocated to agricultural activity and a lower one to other use. Consecutively, the extensive margin optimal decision must satisfy the optimal condition as follows:

$$\Pi = \max_{\mathbf{x},\theta^p\in\Theta} \left\{ \overline{T} \int_{\underline{\theta}}^{\theta^p} \pi^a h(\theta) \, \mathrm{d}\theta + \overline{T} \int_{\theta^p}^{\overline{\theta}} \theta \overline{\pi}^*(\mathbf{x}) h(\theta) \, \mathrm{d}\theta \right\} \tag{6}$$

This modelling allows the inclusion of two series of first order conditions for the annual choices at the farm level ($\mathbf{x}$ and $\theta^p \in \Theta$.) For an interior solution, all the optimal amounts of quasi-fixed inputs nullify average marginal profits ($\overline{\pi}^{*'}_{\mathbf{x}}(\mathbf{x}^*) = 0$) and threshold quality equalizes the variable profits withdrawn from alternative uses. A special feature of this model is the equality between threshold value and the quotient of average potential variable profits:

$$\text{FOC} : \frac{\partial \Pi}{\partial \theta^p} = 0 \;\Leftrightarrow\; \theta^{p*} \overline{\pi}^*(\mathbf{x}^*) = \pi^a \;\Leftrightarrow\; \theta^{p*} = \frac{\pi^a}{\overline{\pi}^*(\mathbf{x}^*)} \tag{7}$$

Only a single case of a corner solution is possible, when $\theta^{p*} = \underline{\theta}$ and all the available qualities are allocated to agricultural use. If $\theta^{p*} = \overline{\theta}$, the farm does not have any land in agricultural use and it does not belong to set I. To conclude this section, we can define analytically the individual agricultural acreage:

$$S(p,c) = \overline{T}\left[1 - H\left(\frac{\pi^a}{\overline{\pi}^*(\mathbf{x}^*)}\right)\right] \tag{8}$$

2. The 1992 MacSharry reforms

These CAP reforms emphasized grains and other arable crops (cereals, oilseeds and protein crops), produced by a subset of the agricultural sector referred to hereafter as the arable sector. They involve significant support price reductions, compensated for by direct payments for all producers per land unit, and a land set-aside scheme compulsory to large producers (Guyomard et al., 1996). More precisely, in the context of France, historical average crop yields and historical average crop areas were set for each *département* (NUTS 3), based on the agricultural administrative statistics for the three marketing years preceding the reforms (1989/90, 1990/91 and 1991/92.) Each French farm in the arable sector can claim compensatory payments up to its historical average crop acreage.[4] The total amount of compensatory payments claimed is verified by the European Community at *département* level. If individual acreages claimed multiplied by historical average yield exceed 92 tons of wheat equivalent (about 20 ha on average), the producer is considered as a professional (from general arrangements) and must set-aside (withdraw from production) at least 15% of his or her claimed acreage.

To incorporate this legislative switching, we start from an assumption made by Ball et al. (1997) that the average land allocation observed before the reforms (historical crop area) is individually optimal in the sense of the above static model. Based on the above notations, individual historical crop area ($S(p,c)$, eq.8) and underlying threshold land quality (θ^{p*}, eq.7) are the exogenous parameters in the reforms. Subsequent to support price reductions, one can observe a substantial reduction in crop price for the 1992/93 marketing year ($p_r < p$.) Assuming that fixed factors are not adjusted in the short term[5] (one year after,) the price cuts modify both the intensive and extensive margin choices. Subscript r represents the value of the parameters for the one year directly following the CAP reforms.

[4] The direct compensatory payments amount (ecu) is equal to the historical acreage claimed for compensation (ha) multiplied by the historical average yield (tons/ha) multiplied by the cropping acreage compensation rate (ecu/ton.) This last rate is identical for all French farms but will be modified in years 1995 and 1997.

[5] Which can be see as lack of reform anticipation, $\mathbf{x}^* = \overline{\mathbf{x}}_r$.

At the intensive margin, optimal variable input use condition (2) coupled with the marginal productivity decrease, demonstrates that for a given q, variable input demand falls off:

$$\frac{c}{p_r} > \frac{c}{p} \Leftrightarrow f_e^{'}(\overline{\mathbf{x}}_r, e_r^*(q), q) > f_e^{'}(\mathbf{x}^*, e^*(q), q) \quad \Leftrightarrow \quad e_r^*(q) < e^*(q) \ \forall q \in Q \tag{9}$$

This result does not mean global variable input reduction in intensification. A price cut decreases the use of variable input for a certain land quality, but the same price cut (and, therefore, paid set-aside) causes agricultural use withdrawal. With temporal inertia on variable input use or increasing variable input application on land quality, average variable input application rates may increase due to price reductions in the agricultural commodity. This mechanism has been termed the *slippage* effect in the literature (Ervin (1988), Fraser (1997), Wu (2000), etc.) The definition is based on observation of a less than proportional production decrease subsequent to an acreage reduction. Hoag et al. (1993) identify three potential causes: (i) less-productive units of land are withdrawn; (ii) less productive farms contribute more to acreage reduction; (iii) inputs are concentrated more on the remaining cultivated land. The above analysis involves the third case but the model below separately integrates all three possibilities.

Price cut also influences the average potential variable profit without acreage payments and the conditional distribution of economic land quality, now noted respectively $\overline{\pi}_r^*(\overline{\mathbf{x}}_r)$ and $h_r(\theta)$. For analytical convenience, consider θ^r such that $h(\theta^{p*}) = h_r(\theta^r)$ which is the lower bound for historically cropped land quality. At the extensive margin, the reforms introduce a third land use : paid set-aside which involves compensatory payments. Moreover, we distinguish between compensated and non compensated cropping acreage, on a payments claiming basis. Without loss of generality, we suppose the non-compensated cropping acreage is at the top of the economic land quality distribution.[6] The farm's problem allows four farmland uses and, therefore, three threshold qualities : θ^r separates stewardship use from set-aside, θ^s separates set-aside from compensated cropping and θ^c separates compensated from non-compensated cropping. Without individual constraints on set-aside location, less favourable (in the economic sense) land units will be set-aside first. By noting respectively τ^c and τ^s the compensatory rates of claimed and set-aside land units,[7] we can express the microeconomic problem as follows:

$$\max_{(\theta^s,\theta^c)\in[\theta^r;\overline{\theta}]} \left\{ \pi^a H_r(\theta^r) + CP(\theta^s,\theta^c) + \overline{\pi}_r^*(\overline{\mathbf{x}}_r) \int_{\theta^s}^{\overline{\theta}} \theta h_r(\theta)\, \mathrm{d}\theta \right\} \tag{10}$$

$$\begin{aligned} s.t. \quad & (1) \quad \overline{T}(H_r(\theta^c) - H_r(\theta^r)) \times RY \geqslant 92 \\ & (2) \quad H_r(\theta^s) - H_r(\theta^r) \geqslant 0,15 \times (H_r(\theta^c) - H_r(\theta^r)) \end{aligned} \tag{11}$$

[6] Based on independence between the compensatory rate and the individual economic land quality.

[7] In addition to the claimed acreage compensatory rate (fn. 5, p.6) reforms establish regional historical set-aside yields and set-aside acreage compensation rates to compute the direct contribution of diverted acreage to compensatory payments. Interesting, historical set-aside yields (for all French *départements*) are bigger than the historical regional cropping yields and the compensation rate is almost twice than for acreage claimed : 45 ecu/quintal *versus* 25 ecu/quintal.

$$\text{with} \quad CP(\theta^s, \theta^c) = \tau^c(H_r(\theta^c) - H_r(\theta^s)) + \tau^s(H_r(\theta^s) - H_r(\theta^r)) \tag{12}$$

Equation (12) describes the compensatory payment arithmetic. Constraint (11.1) is the general arrangement condition with *RY* historical regional cropping yield : farms that do not fit with this are not obliged to withdraw any land. The other constraint (11.2) points to the minimal set-aside rate (15%) need to obtain $CP(\theta^s, \theta^c)$. To simplify its resolution, we can express two propositions :

Proposition 1 : *Constraints (11.1) and (11.2) can be reduced to a unique general arrangement participation constraint as follows (proof in annex A.1):*

$$\overline{T}(H_r(\theta^s) - H_r(\theta^r)) \times RY \geqslant 13,8 \tag{13}$$

Proposition 2 : *By noting ψ the difference between compensatory rates ($\tau^s - \tau^c$), we obtain (proof in annex A.2):*

$$\frac{\partial CP(\theta^s)}{\partial \theta^s} = h_r(\theta^s)\,\psi \tag{14}$$

Based on the above propositions, a unique participation constraint on general arrangement describes the incentive scheme of the reforms and θ^c can be taken out of the decision variables. The model then becomes a classical adverse selection model, which does not differ greatly from Bourgeon et al. (1995):

$$\max_{\theta^s \in [\theta^r; \overline{\theta}]} \left\{ \pi^a H_r(\theta^r) + CP(\theta^s) + \overline{\pi}_r^*(\overline{\mathbf{x}}_r) \int_{\theta^s}^{\overline{\theta}} \theta h_r(\theta)\, \mathrm{d}\theta \right\} \tag{15}$$

$$s.t. \qquad \overline{T}(H_r(\theta^s) - H_r(\theta^r)) \times RY \geqslant 13,8 \tag{16}$$

Assume that constraint (16) is verified, the farm develops extensive margin choices to maximize total profit. This amounts to minimizing the set-aside opportunity cost according to economic land quality. In other words, the economic land quality density function is the distribution of marginal opportunity costs to set-aside land. As in the previous section, the farm chooses a threshold quality such that the higher qualities are allocated to agricultural activity and lower ones to set-aside (until θ^r.) Scalar value is as follows:[8]

$$\theta^{s*} = \frac{\psi}{\overline{\pi}_r^*(\overline{\mathbf{x}}_r)} \tag{17}$$

In preparation for the empirical application for microdata and the partial identification of economic farmland quality distribution, participation constraint (16) constitutes a kind of censure. The theoretical set-aside rate $(H_r(\theta^{g*})S(p,c))$ as a part of the historical average crop acreage is not observable if the farm chooses not to participate in general arrangement. The following

[8] The corner solution for total set-aside is excluded because the maximum rate of paid set-aside is 30%. Moreover, we do not observe a density mass of French farms on this rate.

structural empirical model takes account of this special feature of the MacSharry compulsory set-aside, that the theoretical model explain.

3. Empirical model

Application to individual French farms requires the inclusion of subscript i which represents the observational unit. If neither the compensatory payments nor the claimed acreages are included due to lack of data or confidentiality considerations (which is the case here and for a lot of European agricultural data,) only set-aside and crop individual acreages will be available. For $i \in I$, the observed ratio between these two values (the real Set-Aside Rate, SAR_i) can be denoted:

$$SAR_i = \begin{cases} H_{ri}(\theta_i^{s*}) & \text{if} \quad (16)_i \qquad \textit{(participation)} \\ 0 & \text{otherwhise} \quad \textit{(non participation)} \end{cases} \tag{18}$$

This looks like a non-randomly selected sample problem (Heckman, 1979). Individual selection equation $(16)_i$ represents the trade off between average potential variable profit and potential compensatory payments. To simplify the empirical model, we can rewrite the participation constraint as:

$$(16)_i \;\Rightarrow\; (RY_{dep}S_i(p,c))^{\alpha} \geqslant \overline{\pi}^*_{ri}(\overline{\mathbf{x}}_{ri}) \;, \;\; \forall i \in I \tag{19}$$

The left term represents the potential compensatory payment and the right term the average potential variable profit. Subscript dep denotes that regional yields are observed by *département*, and an unknown parameter α is included to compensate for the simplification of the reciprocal function $H_r^{-1}()$. Estimation of the structural empirical model proceeds with two additional functional specifications : average potential variable profit and economic land quality distribution. Under the previous assumption of the lack of anticipation of reforms, pre-reform fixed input quantities per historical cropping acreage can be treated as good exogenous covariates (because predeterminate) to explain post-reform average potential variable profit.[9] Remember that the average potential variable profit is the variable profit component which depends on the farm economic structure (3):

$$\ln\left(\overline{\pi}^*_{ri}(\overline{\mathbf{x}}_{ri})\right) = \beta_1 \ln(k_i) + \beta_2 \ln(fl_i) + \beta_3 \ln(pl_i) + \mathbf{z}_i^T \beta_4 + \varepsilon_i \;, \;\; \forall i \in I \tag{20}$$

Such a formulation for average variable profit follows Cobb-Douglas production functions with constant input elasticities : the βs. Important in the French arable sector, family labour (fl_i) is separately specified from salaried labour (pl_i.) Capital intensity per ha is denoted k_i, the vector $\mathbf{z}_i$ represents pre-reforms control covariates and $(\varepsilon_i)_{i\in I}$ is an independent and identically-distributed zero mean random variable for omitted individual heterogeneity (e.g. managerial practices.) Likewise, consider that individual distributions of economic land qualities could be fitted by standard exponential distributions of parameters λ_i. For all values of farm i's θ,

[9] Variable input quantities were not available in our database, thus, they do not appear here.

the post-reform cumulative distribution function is given by $H_{ri}(\theta) = 1 - \exp(-\lambda_i \theta)$. Such a specification allows us to define individual land heterogeneity using a unique parameter. It could be seen as restrictive (monotone density function) but its benefit is the possibility to define the proximity between farms in term of land quality : the more λ_i is close to λ_j, the more that land endowments between i and j are similar. With a first order Taylor expansion of the exponential function, we obtain from (18) and (19):

$$SAR_i = \begin{cases} \lambda_i \left(\psi_{dep} \, / \, \overline{\pi}^*_{ri}(\overline{\mathbf{x}}_{ri}) \right) & \text{if} \quad (RY_{dep} S_i(p,c))^{\alpha} - \overline{\pi}^*_{ri}(\overline{\mathbf{x}}_{ri}) \geqslant 0 \\ 0 & \text{otherwhise} \end{cases} \tag{21}$$

As in Heckman (1979), we can view this empirical structure as a specification error and estimates it using a two-step procedure. By substituting (20) into the logarithm transformation of the set-aside rate (21), we obtain a structural reduced form for farms from arable crop sector within the general arrangements. This subset of I (noted I') groups together all farms that participates in the set aside scheme and obtain compensatory payments for acreage over 92 tons of wheat equivalent:

$$\ln(SAR_i) = \beta_0 \ln(\psi_{dep}) - \beta_1 \ln(k_i) - \beta_2 \ln(fl_i) - \beta_3 \ln(pl_i) - \mathbf{z}_i^T \beta_4 - \sigma_{\varepsilon_1 \varepsilon_2} \frac{\phi_i}{\Phi_i} + \ln(\lambda_i) - \varepsilon_i \tag{22}$$

The coefficient β_0 is introduced to permit more flexibility in the specification. This Heckit reduced form is conditional on the first stage estimation of selection probability using probit on all $i \in I$. The Inverse Mills ratio can then be considered as an additional covariate to correct for sample selection under Heckman's hypothesis. Indeed, the naive estimation of the individual set-aside rate for the professional arable sector only (without taking into account selection) produces the residual vector $(\varepsilon_{2i})_{i \in I'}$ which has following property:

$$\mathbb{E}(\varepsilon_{i2} | (19)) = \sigma_{\varepsilon_1 \varepsilon_2} \frac{\phi(\mathbf{w}_i^T \delta)}{\Phi(\mathbf{w}_i^T \delta)} \equiv \sigma_{\varepsilon_1 \varepsilon_2} \frac{\phi_i}{\Phi_i} \quad \text{with} \quad \sigma_{\varepsilon_1 \varepsilon_2} = cov(\varepsilon_1, \varepsilon_2) \tag{23}$$

The conditional expected operator is denoted $\mathbb{E}(|)$, Φ and ϕ are respectively the cumulative and density functions of centered reduced normal law and ε_1 is the probit residual vector, assuming they are normally distributed with normalized variance. Vector δ contains structural parameters which are first step covariates.[10] Then, equation (22) can be estimated if we consider $\ln(\lambda_i)$ is observable. Instead of proxies for the distribution of economic land qualities using an index of productivity or some observed attributes of the land (organic composition, water-holding capacity, etc.), we consider in a first step land quality distributions to be unobservable (fn 2 p.2) and formulate the following assumption of idiosyncratic random distributions:

$$u_i = ln(\lambda_i) - \varepsilon_i \quad \text{with} \quad u_i \; iid \; (0, \sigma_u^2) \tag{24}$$

[10] First step probit covariates of farm i could be obtained by substituting (20) into the logarithm transformation of (19) : $\mathbf{w}_i = \{\ln(k_i) \;\; \ln(fl_i) \;\; \ln(pl_i) \;\; \mathbf{z}_i\}$. We put another set of associated coefficients (δ) in place of βs ((20) and (22)) because of the likely gap between theory and data due to simplifications.

We can now estimate convergent estimators of the model's parameters, in particular $\widehat{\sigma_{\varepsilon_1\varepsilon_2}}$, the vector $\hat{\mathbf{u}}$ and its two first moments. To ensure orthogonality of the residuals and covariates in the regression function,[11] we also use an instrumental approach - Two Stage Least Squares (2SLS). Following the methodology presented by Lee et al. (1980) and its practical application by Wooldridge (2002), we obtain what is probably a better estimation. The results are described below.

As we define in the theoretical model, economic land qualities are strongly related to location and the land's site-specific attributes. More precisely, economic land quality can be considered as the spatial component of variable profit, i.e. a black box that contains all the profit determinants that do not depend directly on farm structure (agronomic quality but also parcel's size, distance to the market, vertical integration, Marshallian externalities, climate, etc.) So, a spatial structure to $\hat{\mathbf{u}}$ would probably arise via the spatial pattern of the λs (24.) This positive spatial autocorrelation of the errors does not induce inconsistent parameter estimations but biases the standard errors downward (Cliff and Ord, 1973). The previous hypothesis (24) accounts for a first step of identification and will be released. A statistical test of the null hypothesis of the residuals with no spatial autocorrelation is conducted for the instrumented models above using Moran's I (Anselin and Kelejian, 1997). Interestingly, geocoded data permits spatial interpolation of the residuals by a kind of kernel smoothing methodology (Cressie, 1993). This exploratory method constitutes a tool to obtain the spatial empirical distribution of the $\hat{\mathbf{u}}$s and must be seen as an example of the empirical consequences of the economic definition of farmland quality presented in this paper. Consider $\Omega = \{\omega_1, ..., \omega_n\}$ as finite set of discrete locations which operate a partition of the land area of farms I. Space units might be heterogeneous, some might not contain any observation, the equation for statistical interpolation is as follows:

$$\hat{u}(\omega_j)|\kappa, b = \frac{1}{n}\sum_{l=1}^{n} \kappa_{jl}(b) \times u(\omega_l) \ , \ \forall \omega_j \in \Omega \tag{25}$$

The scalar κ_{jl} is an *a priori* spatial weight between ω_j and ω_l (from a kernel function), b is the intensity of smoothing (bandwidth) and $u(\omega_l)$ is the value of the residual vector at location ω_l. By substituting u_i's definition (24) in (25) and applying the strong law of large numbers for weighted sums (Chow and Lai, 1973), we eliminate the independent and identically-distributed zero mean random variable ε_i. Two more conditions are necessary for this simplification: sufficient observations for the law of large numbers to hold (there is a consensus on at least 30) and a weight structure that is square summable ($\sum_{l\in\Omega} \kappa_{jl}(b) < \infty$.) So, we have:

$$\hat{u}(\omega_j)|\kappa, b = \frac{1}{n}\sum_{l=1}^{n} \kappa_{jl}(b) \times \ln\left(\lambda(\omega_l)\right) \ , \ \forall \omega_j \in \Omega \tag{26}$$

[11] With standard Ordinary Least Squares (OLS,) covariates may be correlated with residuals if any initial assumption is not correct: for an example see fn.3 p.3.

The functional form of the kernel is said to be less of a determinant for the results than bandwidth value b (Cressie, 1993). As in other non parametric approaches, bandwidth determination follows the classical bias/variance trade off; optimal bandwidth and kernel specification will be discussed in the results section. Formulation (26) permits an economic interpretation of the spatial interpolation of the residuals based on their correspondence with the spatial interpolation of the $\lambda(\omega_l)$s. In compliance with the theoretical model, the spatial autocorrelation test (Moran's I) tells us if there is spatial dependence that is not taken account of in the aspatial modelling. This can be seen as a test for the land quality effect (more two locations are near together, more the land qualities are close to) even though it is not an exclusive interpretation (one apply other explanations for spatial dependence and others omitted variables.) Then, the higher the $\hat{u}(\omega_j)$ value, the poorer is the economic quality of the farmland at location ω_j 'relatively to others.' This means that spatial interpolation can be interpreted as the spatial distribution of individual endowments in economic land qualities. Unfortunately, to our knowledge, there is no way to reject the alternative interpretations, which it reinforces the exploratory nature of this last analysis.

4. Data

We observe the economic structures and acreages for the same 8,028 specialized farms in the arable sector for 1990 and 1993. The data come from a French agricultural ministerial survey, the '*structure des exploitations agricoles*' which was a stratified face-to-face survey which assuring representativity in terms of size, economic orientation and *départements*. Farms not involved in growing grains, oilseeds or protein crops were dropped because the MacSharry reforms detailed above do not apply to them. Final sample size is geographically limited to the *Bassin Parisien* which is generally considered a relatively homogeneous plain of some $104,422km^2$ around Paris (see figure 2.) In 1990, farms were bigger than in the rest of France (average Total Farmland Acreage, TFA hereafter, was about $90ha$ compared to a national average of about $42ha$.) Paris and its immediate surroundings were excluded to keep to 18 *départements*, containing 7,750 *communes*, the spatial scale of farm geocodage. These farms have an average specialization rate of 70% with a coefficient of variation of almost 0.37. Specialization rate is the quotient of individual crop acreage and TFA, its high value and low variance being in accordance with the asumption of a set I of specialized farms.

From point of view of the theoretical model, the 1990 survey represents the pre-reform economic situation from which we compute individual historical crop acreages. Data from 1993 are seen as representing the short term consequences of CAP reforms. Determining the historical theoretical production of surveyed farms (historical acreage time historical regional yields) and comparing them to 92 tons define whether a farm can qualify for the general arrangements.

Taken together with observed set-aside acreages, we can establish a typology of the farms as follows:

[Table 1 about here.]

The above cross-classification determines whether a farm participates to general arrangement, based on the database variables. Participating farms are presented in columns (d) and (e) and represent around 70% of the sample. If we compute the percentage of potential professional farms that claim compensatory payments on less acreage than is cultivated, we obtain 32.6% which is similar to what Colman and Vavra (2002) found for the United Kingdom.[12] Plotting the observed frequencies of set-aside rates according to the classification in Table 1 leads to the depiction in Figure 1 :

[Figure 1 about here.]

The participation probability is increasing with farm size because potential compensatory payments are computed on bigger historical base acreages. Nevertheless, 537 farms (c) with sufficient arable land to be classified in general arrangement do not set-aside land and thus are foregoing the compensatory payments for more than 92 tons of wheat equivalent. Once participation is decided, the set-aside rate seems to be correlated less with farm size and theoretical production than participation choice : columns (d) and (e) are similar in terms of economic dimension. The above model proposes an explanation for this. For farms that participate, the set-aside rate comes from the local slope of variable profit (intensive margin), which is not correlated directly to farm size but with input intensities or the economic land quality distribution.

Table 2 presents summary statistics of variables used to estimate the model in the previous section. It identifies four types of variables based on their functions in the econometric strategy: s represents structural variables which derive directly from the theoretical model, c represents the control covariates in the vector $\mathbf{z}_i$, some of which (ec) are endogenous and i represents the instruments used.

[Table 2 about here.]

As in the previous section, SAR_i is the individual Set-Aside Rate ; $RY_{dep}S_i(p,c)$ and ψ_{dep} do not change, they are respectively in tons and ecu. k_i is the weighted[13] sum of farm machinery, fl_i is the sum of family labour in a standardized annual working unit and pl_i is the equivalent for salaried labour: the wage earners. We use as covariates the variation in TFA between the two periods ($(\Delta TFA)/TFA$) and specialization degree in 1990 ($GOAC/TFA$.) The data show that the hypothesis of fixed land endowments used to simplify the theoretical model is not verified:

[12] =(537+1,483)/(537+1,483+4,165); Colman and Vavra (2002) find 33.3%, table 3 p. 294.

[13] According to their respective horsepowers.

on average, farms extend to almost 15% of their initial TFA between 1990 and 1993. Moreover, we do not take into account cross-effects between farmland uses because of the monoproduct model: even if more than 25% of the sample has less than 5% of non-GOAC use, almost 25% has only 50% of GOAC and 25% have livestock. The problem with these control covariates is the simultaneity with the set-aside choice which is both a cause and a consequence of acreage variations and farm specialization. The instrumentation strategy is described more precisely in the next section. We drop the two extreme centiles to obtain a final sample of 7,640 farms of which 5,524 participate (72%.)

5. Results and discussion

Estimation of the sample selection model (21) based on assumption (24) and endogeneity of some control variables is conducted following Wooldridge (2002)'s procedure (17.2 p.568-570.) The joint presence of sample selection and endogenous covariates requires an additional assumption which looks like a rank condition in classic estimations using instrumental variables. From a practical point of view, this condition can generally be satisfied by taking several precautions (Wooldridge, 2002) : (i) the Mills ratio must be present as a covariate in the first steps of the instrumentation; (ii) all exogenous variables (including instruments) must be present in both the selection equation and instrumentation first steps; (iii) at least one exogenous variable, not present in the reduced form (22), must be included for each extra equation.[14]

The following tables present respectively the coefficient estimates for the selection equation and the reduced form of the set-aside rate. Table 3 presents two specifications, one with only structural variables (as in (19)) and the other with all exogenous variables to fit the Mills ratios and assure additional rank condition. Table 4 includes two specifications, one by OLS without taking account of endogeneity, and another with instrumentation (2 SLS) for acreage variations and specialization degree. Remember that these results depend on assumption (24) which is questionable, so interpretations must be made with caution.

[Table 3 about here.]

The above models demonstrate good explanatory power with about 80% of concordant previsions. According to Akaike's Information Criterion, the second specification dominates. In the first specification, all regressors are statistically different from zero at the 1% threshold. The coefficients have the expected sign: negative for fixed input which increase average variable profit

[14] With a selection equation and 2 endogenous regressors, this condition is reached with at least 3 extra exogenous variables. For the selection equation, $\ln(RY_{dep}S_i(p,c))$ is strongly suggested by the theoretical model to be an extra regressor. To fit $(\Delta TFA)/TFA$ and $GOAC/TFA$, instruments are selected on the basis of validity tests: age of principal farmer, land tenure proportions and presence of a successor are used for the former. For the latter, questions related to irrigation and presence of livestock are used on combination with the taxation scheme which depends indirectly on previous crop choices.

and decrease the probability of participation; positive for monetary incentives, with a strong effect of potential compensatory payments. The introduction of additional variables in probit (2) does not change the results substantially. The negative effect of family labour is more balanced with the waged labour. The effect of potential compensatory payments (indirectly farm size) is divided by more than 2, which probably means correlation between farm size and the additional variables. Not all the control variables are significant, the signs of the coefficients are globaly as expected with some exceptions. For example, an agricultural formation for the principal farmer means an increase in terms of participation probability (so one might assume the formation is productivity improving.) To estimate the reduced form (22), the sample is limited to farms that are defined in general arrangements.

[Table 4 about here.]

Between these last two models, the adjusted R squares are divided by two (80% for the OLS and 42% for the 2SLS.) This is principally due to the weak projection of endogenous regressors by the first stages of the 2SLS methodology.[15] Even if the Sargan test does not reject instrument validity (exogeneity), this well-known problem of weak instruments invalidates the statistical inference of the coefficients of the endogenous variables (for a survey of these questions, see Hahn and Hausman (2002) or Stock et al. (2002).) Since endogenous regressors are control variables and the interpretation of coefficients is not so important here, we do not dwell on these questions. Interestingly, the Mills ratio is not significant in either specification. This is an important result which means that the incentive scheme in the reforms performs well. The reforms create the idea of general arrangements to avoid compulsory set-aside by small farms which are regarded as economically more fragile. We find this separation based on payment claiming permits reaches the goal without any effects on set-aside rate for farms from general arrangement. This incentive effect would probably have some longstanding consequences for the arable sector because it strengthens the support duality between small and larger farms. If we envisage more long run consequences based on land adjustments, the area-payments differentiation, which is high to assure the participation of larger farms, may produce appropriation of land and a phenomenon of concentration. Moreover, this result means that analysis that consider only general arrangement (such as the Farm Accountancy Data Network, FADN) do not suffer from a sample selection bias, at least in terms of the arable sector's reaction to the MacSharry reforms.

Another point of interest in the above regressions is the possibility to obtain an estimated vector of residuals, orthogonal to the economic structure of farms and the payment scheme of the reforms. To develop a spatial analysis of residuals similar to that we presented in the previous section, the observational unit switches from farms to *communes*. Inside the 7,750

[15] In particular for acreage variations, R^2=0.15; because for specialization, R^2=0.5.

communes of the area, 3,494 have no observations, 2,324 have one observation, 1,092 have two, 500 have three and 340 have four or more. This farm geocodage permits us to conduct a spatial analysis of the residuals. By far, the most popular test for a kind of spatial dependence (autocorrelation) is that based on Moran's I test statistic. We compute it for the residuals from the previous estimations using Anselin and Kelejian (1997) methodology to take account for the presence of endogenous regressors. The non-standardized spatial weight matrix used is in the following form :

$$W_{ij} = \begin{cases} (dist_{jl})^{-2} & \text{if } dist_{jl} \leqslant 20km \text{ and } j \neq l \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

where $dist_{jl}$ is the centroid-to-centroid distance between *communes* ω_j and ω_l and 20km is the size of the ray, *a priori* fixed.[16] Both OLS and 2SLS residuals exhibit a spatial structure because we reject the null hypothesis of no spatial autocorrelation: the respective values of the statistics are 0.059 and 0.086 with p-values inferior to 1%. This inference result means that there is a non-explained part of the observed set-aside which is correlated with location. Our theoretical model identifies this component as the spatial heterogeneity of land. However, this relationship is not an exclusive explanation and there could be other causes of spatial dependences: neighbourhood effects, spillovers between land uses, spatial patterns of unobserved variable input use or any other spatially omitted variables which might reinforce the spatial dependence of the residuals. Moreover, spatial dependence among the errors could be due to measurement errors arising from a mismatch between *commune* boundaries and the boundaries of the spatial process. The application presented here identifies all the spatial components of the set-aside consequences (not explained by covariates) and analyses them as the effects of land quality. Interpolate values of residuals are computed with a kernel formula (25) to obtain the following spatial description of the patterns :

[Figure 2 about here.]

For kernel structure, we use an exponential form[17] which is square summable. With a bandwidth (b) of 20 km, the average number of considered observations by *commune* is 48, the minimum is 18 but only 3% of the sample have less than 30 neighbours. Another piece of information used to choose the 'optimal' bandwidth is the Cross Validation criteria (Bowman, 1984) which consists of minimizing the interpolation errors for all area. The Cross Validation statistic diminishes with decreased bandwidth, down to a minimal value which corresponds to the exigencies of the strong law of large numbers for the idiosyncratic component of residuals. In that sense, $b = 20km$ seems to be 'optimal.' The main hypothesis under such a usage of spatial weights is that the more that two *communes* are geographically nearby, the more they are alike, as suggested by Moran's I. The map presents interpolated lambdas computed on the basis of 2SLS-Heckit

[16] We tested several rays and the results were of the same order, we present the results for 20 km because this ray is described as 'optimal' in the kernel smoothing approach, as we will see later.

[17] $\kappa_{jl}(b) = exp(-b \times dist_{jl})$ if $dist_{jl} \leqslant b$ and $\kappa_{jl}(b) = 0$ otherwise.

residuals, by taking their logarithm transformation. This leads us to cluster the *communes* in accordance with the spatial component of the residuals. The land qualities of the *communes* in dark-red can be seen as being better than those in light-red, because the interpolate values of the residuals are inferior. The map also includes Paris, the borders of the *départements*, the two main rivers in the area and the 14 cities with more than 50,000 inhabitants. Discussion of location is not the focus of this paper, but we note that proximity to main cities (perhaps with the exception of Paris) or rivers does not seem to be important. Morover, the area in south west of Paris (*la Beauce*) is well known to be fertile. More investigations are needed to improve the interpretation of spatial distribution: with objectives and observables parameters such as agronomic land quality, climate, distance from cities or parcels' sizes for exemple.

6. Conclusion

In accordance with the hypothesis of farmers' rationality and consequently with the CAP MacSharry reforms, this paper presents an explanation for the fact that specialized arable farms do not claim compensatory payments on all eligible acreages. Under the usual hypothesis for the economic modelling of the reforms (lack of anticipation, mono-product farms, fixed land endowments, constant return to scale from land and preliminary long term equilibrium), we introduce the spatial heterogeneity of farmland. Based on differentiated land endowments, the model predicts a trade-off between compensatory payments for cultivated lands following the reforms and diversion of the poorest land qualities. The compulsory set-aside rate (15%) constrains participation in the general arrangement and the acreages claimed for compensatory payments. We show that such a model can constitute the basis for a structural estimation of general arrangement participation and set-aside rates using microeconomic data. The negative effects of per-land unit input intensity on general arrangement participation and set-aside rate is consistent with the model, and the absence of selection bias when we regress only observed set-aside rates is an important result. An illustration of the economic definition of land quality (omitted in first empirical estimations because it is assumed to not be observable) is provided by using an exploratory tool, the spatial kernel interpolation. The spatial patterns of land qualities are presented for an area around Paris on the basis of the residuals in previous models. This exploration of the spatial heterogeneity of farmland needs to be developped futher, in particular through the introduction of a multi-product framework and decomposition of the black box of land quality by its principal observable components.

References

Anselin, L. and Kelejian, H. H. (1997). Testing for spatial error autocorrelation in the presence of endogenous regressors. *International Regional Science Review* 20: 153–182.

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

Bell, K. P., Boyle, K. J. and Rubin, J. (2006). *Economics of rural land-use change*.

Bockstael, N. E. (1996). Modeling economics and ecology : The importance of a spatial perspective. *American Journal of Agricultural Economics* 78: 1168–1180.

Bourgeon, J.-M., Jayet, P.-A. and Picard, P. (1995). An incentive approach to land set-aside programs. *European Economic Review* 39: 1487–1509.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* 71: 353–360.

Caswell, M. F. and Zilberman, D. (1986). The effects of well depth and land quality on the choice of irrigation technology. *American Journal of Agricultural Economics* 68: 798–811.

Chow, Y. S. and Lai, T. L. (1973). Limiting behavior of weighted sums of independant random variables. *The Annals of Probability* 1: 810–824.

Cliff, A. D. and Ord, J. (1973). *Spatial autocorrelation*. London : Pion.

Colman, D. and Vavra, P. (2002). The optimal allocation of land between alternative support schemes : An exploration of claims for arable area payments. *Journal of Agricultural Economics* 53: 283–298.

Cressie, N. A. C. (1993). *Statitics for spatial data*. Wiley-Interscience.

Ervin, D. E. (1988). Cropland diversion (set-aside) in the us and uk. *Journal of Agricultural Economics* 39: 183–195.

FAO (1991). The digitized soil map of the world. Tech. Rep. World Soil Ressources Report 67/1 (Release 1.0), Food and Agriculture Organization of the United Nations.

Fraser, R. (1997). Land heterogeneity and the may 1992 reform of cap cereal price support. *Journal of Agricultural Economics* 48: 65–70.

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.

Hahn, J. and Hausman, J. (2002). A new specification test for the validity of instrumental variables. *Econometrica* 70: 163–189.

Hardie, I. W. and Parks, P. (1997). Land use with heterogeneous land quality : An application of an area base model. *American Journal of Agricultural Economics* 79: 299–310.

Heckman, J. J. (1979). Sample selection as a specification error. *Econometrica* 47: 153–161.

Hoag, D. L., Badcock, B. A. and Foster, W. E. (1993). Field-level measurement of land productivity and program slippage. *American Journal of Agricultural Economics* 75: 181–189.

Lankoski, J. and Ollikainen, M. (2003). Agri-environmental externalities : A framework for designing targeted policies. *European Review of Agricultural Economics* 30: 51–75.

Lee, L.-F., Maddala, G. S. and Trost, R. (1980). Asymptotic covariance matrices of two-stage probit and two-stage tobit methods for simultaneous equations models with selectivity. *Econometrica* 48: 491–503.

Lichtenberg, E. (1989). Land quality, irrigation development, and cropping patterns in the nothern high plains. *American Journal of Agricultural Economics* 71: 187–194.

Lichtenberg, E. (2002). *Agriculture and the environment*. Elsevier, *Hanbook of agricultural economics 2A-Agricultural and its external linkages*, chap. 23. 1249–1314.

Moro, D. and Sckokai, P. (1999). Modelling the cap arable crop regime in italy : Degree of decoupling and impact of agenda 2000. *Cahiers d'économie et sociologie rurales* 53: 49–73.

Nelson, A. C. (1992). Preserving prime farmland in the face of urbanization : Lessons from oregon. *Journal of the American Planning Association* 58: 467–488.

Palmquist, R. B. (1989). Land as a differentited factor of production : A hedonic model and its implications for welfare measurement. *Land Economics* 65: 23–28.

Plantinga, A. J. (1996). The effect of agricultural policies on land use and environmental quality. *American Journal of Agricultural Economics* 78: 1082–1091.

Roberts, D., Froud, J. and Fraser, R. (1996). Participation in set aside: What determines the opting in price? *Journal of Agricultural Economics* 47: 89–98.

Serra, T., Zilberman, D., Goodwin, B. K. and Hyvonen, K. (2005). Replacement of agricultural price supports by area payments in the european union and the effects on pesticide use. *American Journal of Agricultural Economics* 87: 870–884.

Stavins, R. N. (1999). The costs of carbon sequestration : A revealed-preference approach. *American Economic Review* 89: 994–1009.

Stavins, R. N. and Jaffe, A. B. (1990). Unintented impacts of public investments on private decisions : The depletion of forested wetlands. *American Economic Review* 80: 337–352.

Stock, J. H., Wright, J. H. and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20: 518–529.

USDA (1973). *Land Capability Classification*. Soil Conservation Service Agricultural Handbook No. 210.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. The MIT press.

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

A. Proofs for section 3

A.1. Proposition 1

1. Suppose a participating farm claims compensatory payments for only a fraction of its historical cropping acreages: $\theta^c < \overline{\theta}$. The set-aside rate equals 15% of the acreage claimed in order to fit exactly with the definition of general arrangements.[18] Set-aside acreage constitutes a constraint on compensation claims, inequality (10.1) is binding:

$$(1) \Rightarrow H_r(\theta^c) - H_r(\theta^r) = \frac{92}{RY \times \overline{\overline{T}}} \tag{28}$$

   By substituting the above equality in (10.2), we obtain :

$$\overline{T}(H_r(\theta^s) - H_r(\theta^r)) \times RY \geqslant 13,8 \tag{29}$$

2. Suppose a farm claims compensatory payments for all cropping acreages: $\theta^c = \overline{\theta}$ and $H_r(\theta^c) = 1$. Cropping acreage and claimed acreage are identical:

$$(1) \Rightarrow 1 - H_r(\theta^r) \geqslant \frac{92}{RY \times \overline{\overline{T}}} \tag{30}$$

$$(2) \Rightarrow H_r(\theta^s) - H_r(\theta^r) \geqslant 0,15 \times (1 - H_r(\theta^r)) \tag{31}$$

The system $\{(30) \cup (31)\}$ is always true when the inequality (29) is true. A farm that is in case 2 necessarily satisfies case 1. We can merge constraints and simplify θ^c.

A.2. Proposition 2

Keeping proposition 1's scheme, we have:

1. $CP_1(\theta^s) = H_r(\theta^s)(\tau^s + \frac{17}{3}\tau^c) - H_r(\theta^r)(\tau^s + \frac{17}{3}\tau^c)$
2. $CP_2(\theta^s) = H_r(\theta^s)(\tau^s - \tau^c) - H_r(\theta^r)\tau^s + \tau^c$

The main difference between the two functional forms is the set-aside marginal payment. Either it equals the set-aside compensatory rate plus extra rightfully claimed land units (case 1), or it equals the set-aside compensatory rate minus the claimed one (case 2.) We admit to simplification in order that the two cases are unified to the second:

$$\frac{\partial CP(\theta^s)}{\partial \theta^s} = h_r(\theta^s)(\tau^s - \tau^c) \equiv h_r(\theta^s)\psi \tag{32}$$

This simplification is oriented by empirical application. It enables a selection model instead of a switching regime model, which does not seem to be justified by the likeness between the two cases.

[18] A rational farm does not set-aside land units without claiming the rightful maximum compensatory payments.

Tables

N = 8,028 farms from arable sector					
Theoretical Production	Less than 92 t.w.e.		More than 92 t.w.e.		
Set aside rate	0	> 0	0	]0;15[	⩾ 15
	(a)	(b)	(c)	(d)	(e)
Arrangement	Simplified	Simplified	Simplified	General	General
Compensation Payments	simp.[1]	simp.[1] + voluntary set aside	simp.[1]	set-aside + partial claiming	set-aside + total claiming
Frequencies (%)	1,142 (15)	701 (9)	537 (6.5)	1,483 (18.5)	4,165 (51)
Weighted Frequencies[2] (%)	14,440 (20)	6,840 (9)	4,875 (6.5)	12,263 (16.5)	35,000 (48)

Sources: Structure des exploitations agricoles 1990 et 1993.

[1] *Simplified arrangements establish acreage-proportional compensatory payments independents from set-aside.*

[2] *Weighted frequencies consider the weights that come from stratified sampling.*

Table 1. Farms classification according to reform's reactions.

Variables	N=8,028 farms from arable sector						
Continues	Mean	Std.	1%	25%	50%	75%	99%
$(SAR_i)^s$	0.14	0.1	0	0.1	0.14	0.16	0.5
$\ln(1+k_i)^s$	1.97	0.64	0	1.65	2	2.3	3.65
$\ln(1+fl_i)^s$	3	0.82	1	2.5	3	3.5	5.4
$\ln(1+pl_i)^s$	0.48	1	0	0	0	0	4
$\ln(RY_{dep}S_i(p,c))^s$	5.36	1.23	1.87	4.63	5.6	6.28	7.4
$(\psi_{dep})^s$	11.68	1.07	9.91	10.84	12.02	12.64	13.047
$((\Delta TFA)/TFA)^{ec}$	0.17	0.97	-0.9	-0.02	0.04	0.22	2.8
$(GOAC/TFA)^{ec}$	0.7	0.26	0.08	0.53	0.76	0.95	1
AGE^i	47.8	11.6	25	39	48	57	76
$PRLOC^i$	0.004	0.04	0	0	0	0	0.14
$PELOC^i$	0.64	0.35	0	0.4	0.76	0.98	1
Dummies	N	%	Definition				
$PART^s$	5,648	70	General arrangements participation - 1993				
$IRRI^c$	799	10	Irrigation technology for crops - 1990				
$IRRP^i$	347	4.3	Planned irrigation in 1990				
$DRAI^c$	2,628	32.7	Presence of drained acreage - 1990				
WOM^c	1,237	15	The chief of the farm is a woman - 1990				
OOB^c	6,163	76.7	Legal statut : one's own business - 1990				
$AGRF^c$	4,575	57	Chief have an agricultural formation - 1990				
$PRLI^c$	2,007	25	Presence of livestock - 1990				
$PLLI^i$	733	9.1	Planned livestock in 1990				
CSW^c	959	12	Chief switching between 1990 and 1993				
$SUCC^i$	6,260	78	Presence of a successor - 1990				
TAX^i	2,960	36.9	Collective fixed rate for taxation scheme - 1990				

Sources: Structure des exploitations agricoles 1990 et 1993.

Variables : ^{s} structural, ^{c} control, ^{ec} endogenous control and ^{i} instruments.

GOAC : Grains and Other Arable Crops.

PRLOC and PELOC are respectively proportions of land in temporary tenancy and permanent tenancy; the third possibility of land tenure is the ownership.

Table 2. Summary statistics of the variables of the models.

Endogenous variable : *PART*				
	Probit (1)		Probit (2)	
Regressors	Coef.	Std.	Coef.	Std.
INT	-1.66**	(0.27)	-2.54**	(0.4)
$\ln(1+k_i)$	-0.2**	(0.03)	-0.17**	(0.03)
$\ln(1+lf_i)$	-0.65**	(0.03)	-0.25**	(0.03)
$\ln(1+lp_i)$	-0.06**	(0.02)	-0.11**	(0.03)
$\ln(RY_{dep}S_i(p,c))$	0.97**	(0.05)	0.42**	(0.02)
Ψ_{dep}	-		0.001**	(0.0002)
$IRRI$	-		-0.5**	(0.07)
$IRRP$	-		0.52	(0.14)
$DRAI$	-		0.09*	(0.04)
OOB	-		-0.03	(0.06)
$AGRF$	-		0.25**	(0.05)
WOM	-		-0.1*	(0.06)
CSW	-		0.16*	(0.07)
$PRLI$	-		-0.02	(0.04)
$PLLI$	-		-0.3**	(0.06)
AGE	-		0.042**	(0.01)
AGE^2	-		-0.000**	(0.000)
$PELOC$	-		0.3**	(0.06)
$POLOC$	-		0.23	(0.4)
$SUCC$	-		0.07	(0.05)
TAX	-		-0.58**	(0.05)
N	7,640		7,640	
% Conc.	80		89.5	
AIC	7,034		5,359	

**, *and †respectively significant at 1%, 5% and 10%.

Table 3. Results of selection equation with probit methodology.

Endogenous variable : $\ln(SAR_i)$				
	OLS		2SLS	
Regressors	Coef.	Std.	Coef.	Std.
INT	0.156^{**}	(0.005)	0.187^{**}	(0.013)
$MILLS$	0.003	(0.018)	-0.002	(0.04)
$\ln(1+k_i)$	-0.002^{*}	(0.0007)	0.0001	(0.0007)
$\ln(1+lf_i)$	0.004^{**}	(0.0007)	0.0008	(0.001)
$\ln(1+lp_i)$	-0.001^{**}	(0.0004)	-0.002^{**}	(0.0004)
Ψ_{dep}	-0.000	(0.000)	-0.0006	(0.0005)
$(\Delta TFA)/TFA$	0.12^{**}	(0.001)	0.132^{**}	(0.006)
$GOAC/TFA$	-0.03^{**}	(0.002)	-0.054^{**}	(0.01)
$IRRI$	0.000	(0.001)	0.0002	(0.001)
$DRAI$	-0.0017^{*}	(0.0007)	0.0016	(0.0007)
OOB	-0.002^{*}	(0.000)	-0.0002	(0.0009)
$AGRF$	0.002^{**}	(0.0007)	0.0002	(0.001)
WOM	0.0001	(0.001)	0.0016	(0.0012)
CSW	0.003^{**}	(0.001)	0.0035^{**}	(0.001)
$PRLI$	0.008^{**}	(0.001)	0.0012	(0.0025)
N	5,524		5,524	
Adjusted R^2	0.79		0.42	
Sargan	-		0.718	

**, *and †respectively significant at 1%, 5% and 10%.

Table 4. Set aside rate regressions accounting for sample selection.

Figures

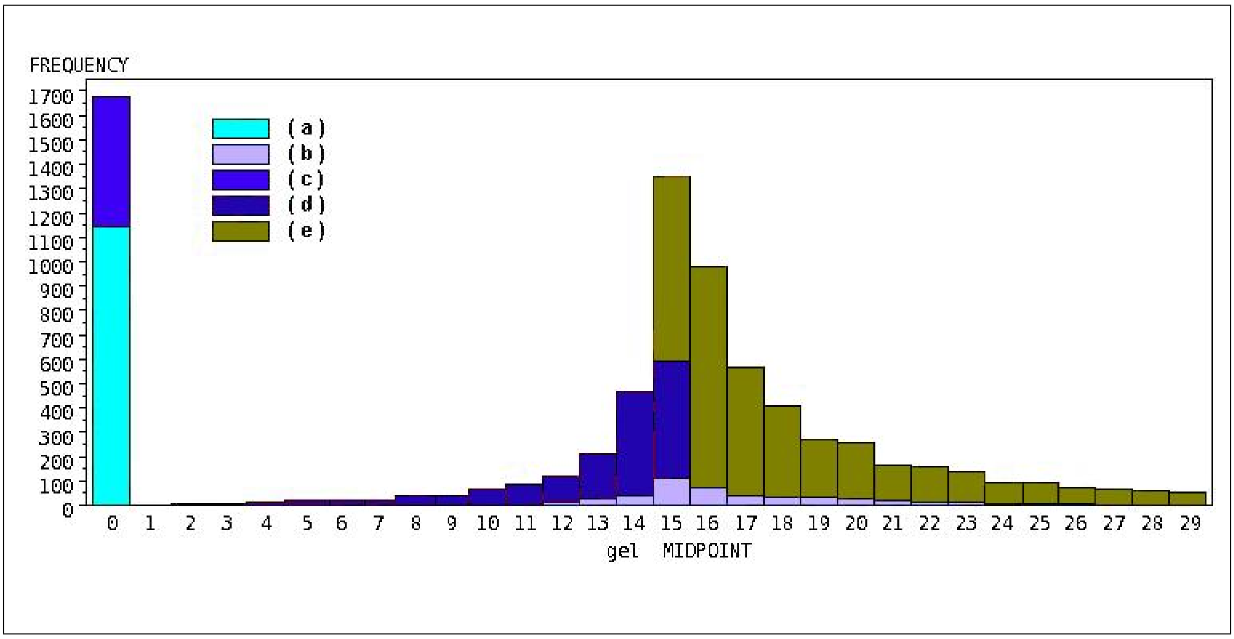

Figure 1. Set aside rate empirical distribution with table 1's classification.

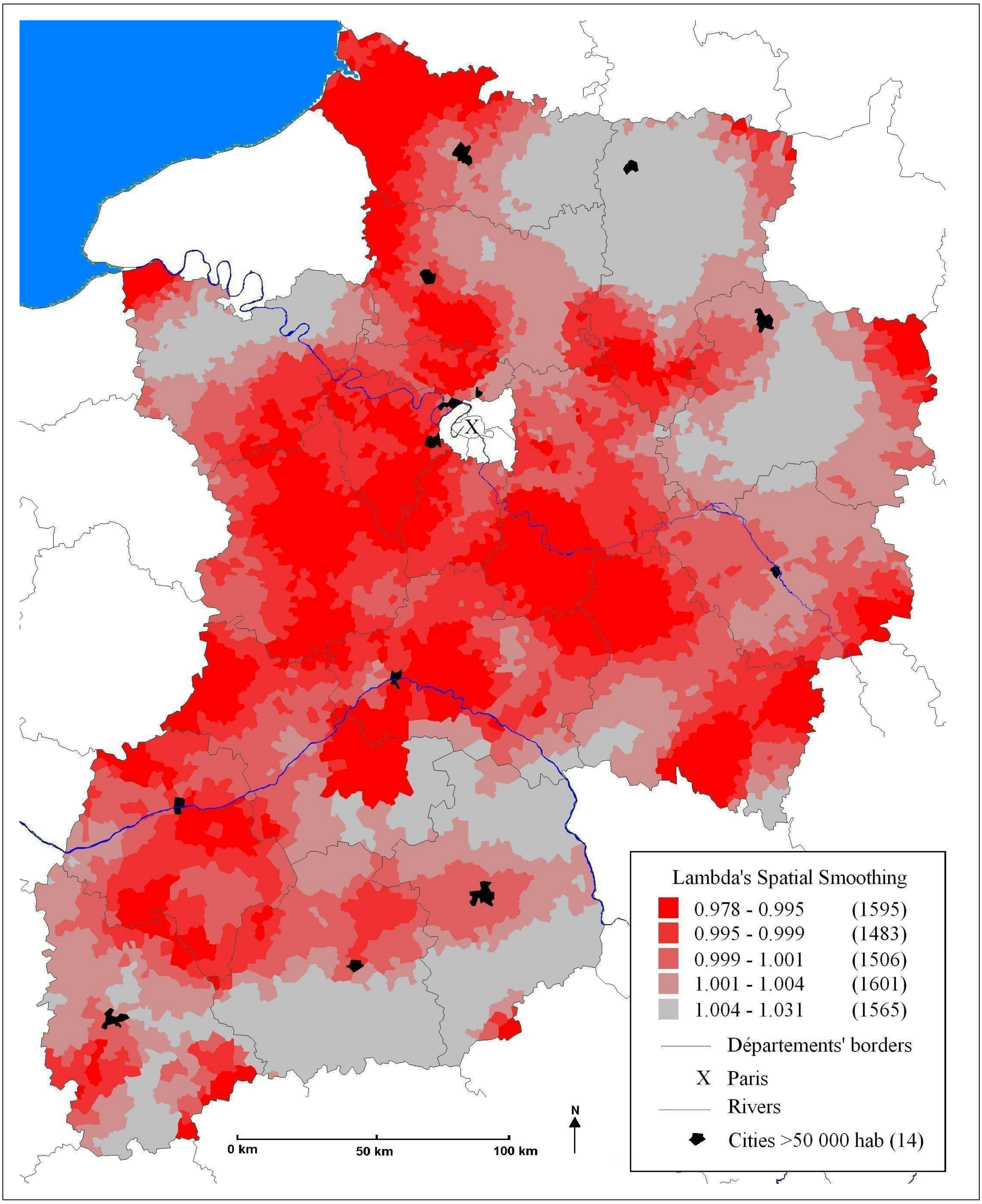

Figure 2. Spatial distribution of land quality index, with b=20 km.