A new model for agricultural land use modelling and prediction in England using spatially high-resolution data

Namhyun Kim[*,a], Patrick Wongsa-art[b], Ian J. Bateman[a]

[a] *University of Exeter Business School, University of Exeter, United Kingdom*
[b] *Cardiff Business School, Cardiff University, Cardiff, United Kingdom*

Abstract

In this paper, we contribute toward building a better understanding of farmers' responses to behavioural drivers of land-use decision by establishing an alternative analytical procedure, which can overcome various drawbacks suffered by methods currently used in existing studies. Firstly, our procedure makes use of spatially high-resolution data, so that idiosyncratic effects of physical environment drivers, e.g. soil textures, can be explicitly modelled. Secondly, we address the well-known censored data problem, which often hinders a successful analysis of land-use shares. Thirdly, we incorporate spatial error dependence and heterogeneity in order to obtain efficiency gain and a more accurate formulation of variances for the parameter estimates. Finally, we reduce the computational burden and improve estimation accuracy by introducing an alternative GMM-QML hybrid estimation procedure. We apply the newly proposed procedure to spatially high resolution data in England and found that, by taking these features into consideration, we are able to formulate conclusions about causal effects of climatic and physical environment, and environmental policy on land-use shares that differ significantly from those made based on methods that are currently used in the literature. Moreover, we show that our method enables derivation of a more effective predictor of the land-use shares, which is utterly useful from the policy making point of view.

Key words: Agro-environmental policy, land-use, multivariate Tobit, system of censored equation, spatial model, error component model.
JEL: C13, C21, C23, C34, Q15, Q53.

[*]Corresponding author
Email address: N.Kim@exeter.ac.uk (Namhyun Kim)

1. Introduction

Land is the most critical natural assets which provides us with various fundamentals of life, from clean water and food to the natural regulation of hazards, such as flooding. In the literature, the term “land-use” is used when describing socio-economic use of land, such as agricultural, recreational or residential use. In the current paper, we focus on agricultural land-use, e.g. arable land and pasture land (or grassland). In many countries, the biggest land-use category is agriculture. For example, agriculture is the biggest land-use category in England at 63% compared to transport/utilities and residential at 4% and 1%, respectively (Ministry of Housing, Communities and Local Government (2017)). Hence, it is a common knowledge that agricultural land-use decision (e.g. to grow wheat or barley instead of oilseed rape) significantly affects the environment (e.g. biodiversity) and socio-economic welfare (see also Mattison and Norris (2005), and Reidsma et al. (2006)). Moreover, there is a prevalent suggestion that introducing changes to agricultural land-use, i.e. how agricultural land is cultivated, can help to achieve deep emission reductions and prepare for climate change. In the UK, Committee on Climate Change (CCC) proposes various changes to the way we cultivate land to help to achieve net-zero emission target (CCC (2018)). These are (a) to reduce land-use for grasslands by 26 to 36%, (b) to introduce new woodlands by 1.5 million hectares, and (c) to increase land-use for bio-energy crops, e.g. oilseed rape, by up to 1.2 million hectares. It is believed that these changes should lead to between 35 and 80% overall reductions in Metric tons of carbon dioxide equivalent by 2050.

Although the above paragraph only briefly provides an insight into the importance of agricultural land-use to the environment and socio-economic welfare, from the policy making point of view, this is sufficient to highlight the necessity to manage how land is allocated between its alternative uses. To manage agricultural land requires a good understanding of farmers’ responses to behavioural drivers of land-use decision. Previous studies have suggested various behavioural drivers. These can be categorised into: (i) Climatic drivers, e.g. rain and temperature, (ii) Economic drivers, e.g. input/output prices, and (iii) Environmental policies and schemes, e.g. greenbelts and environmentally sensitive areas (ESAs) in the UK. A good understanding of how these drivers influence land-use decisions over time and spatial space should help the UK government to both evaluate existing practices and formulate new environmental policies, especially after Brexit.

In the current paper, we aim to contribute toward improving the ability to formulate a better understanding of farmers’ responses to the behavioural drivers of land-use decision. We achieve this objective by establishing an analytical procedure, which can handle complex data structures and is able to overcome various methodological drawbacks suffered by existing methods, and applying such a method to investigate how the climatic, economic and policy drivers influence agricultural land-use patterns in England. In this paper, these agricultural land-use patterns are depicted by "land-use share", which is defined hereafter as a proportion of a given plot of land used for cultivating a given crop. Moving beyond methods usually used in existing studies (Fezzi and Bateman (2011), Chkir and Le Gallo (2013), Ay et al. (2017), and Marcos-Martinez et al. (2017), for example), our analytical framework explores various directions. These are (i) to examine use of spatially disaggregated data, (ii) to model the land-use share as a censored response, (iii) to allow for potential spatial error dependence (SED), (iv) to model unobserved heterogeneity in an error component structure, and (v) to reduce computational burden by introducing a hybrid estimation procedure.

We discuss statistical and empirical underpinnings of these proposals in Section 2.1. Incorporating these features gives rise to a system of two-limit (TL) random-effect (RE) Tobit models with SED (TL-RE-SED-Tobit hereafter) for spatially high-resolution panel data. We thoroughly explain the construction of such a system in Sections 2.2 and 2.3, while introducing a new hybrid QML/generalised method of moments (GMM) estimation procedure in Section 3. We explain each of the necessary steps in detail in Sections 3.1 to 3.4. In Section 4 we apply our method to spatially high resolution data for England and formulate conclusions about causal effects of climatic and physical environment, and environmental policy on land-use shares. We note that these conclusions differ significantly from those made based on deficient methods that are currently used in the literature. Moreover, we show that our method enables derivation of a more effective predictor of the land-use shares, which is utterly useful from the policy making point of view. Section 5 draws some important conclusions. Finally, mathematical proof and other technical details are delegated to an online appendix.

2. System of TL-RE-SED Tobit Equations

We begin with a set of analytical considerations that lead to the need to formulate the system of TL-RE-SED-Tobit.

2.1. Methodological Explorations

These are intended to address drawbacks in the empirical methods used by previous studies for agricultural land-use modelling and prediction.

2.1.1. *Exploring the use of spatially disaggregated data*

In an analysis of agricultural land-use, different techniques are required for different data resolutions. At the extreme ends of the spectrum, we have individual data (e.g. parcel-level data) and aggregated data on a larger geographical region (e.g. national level). Regarding the former, an analysis is often conducted within a discrete-choice modelling framework (e.g. Li et al. (2013)). An analysis of the latter involves tools in panel-data regression models and seemingly unrelated regressions (e.g. Baltagi and Pirotte (2011), Chakir and Gallo (2013), Marcos-Martinez et al. (2017), and Ay et al. (2017)). We believe that there are benefits to be gained by exploring spatially-disaggregated data, i.e. a convenient middle ground. In this regard, individual choices are aggregated to construct the land-use shares. But, unlike the case for national level data, aggregation is done only on a scale that is small enough to capture the spatial variation in the environmental and climatic drivers of farmers' behaviour, and proportionate with the scale of the decision-making unit. The first characteristic suggests that an important benefit is the ability to explicitly model the idiosyncratic effects of policies and other other physical environmental drivers (e.g. mean elevation, land slope and altitude). Furthermore, to satisfy the second characteristic, we assume that each of the spatial units considered is a decision-making unit (see also the discussion in Remark 4.2 for details).

2.1.2. *Modeling land-use shares as censored responses*

A difficulty of modelling spatially disaggregated data resides in an issue often referred to as censoring problem, i.e. we are likely to see a wide range of land-use share values between 0 and 1 with pile-ups at the two endpoints. The failure to account for these features leads to numerous methodological shortfalls, especially the biasness and inconsistency of the parameter estimates (see e.g. Greene (2008), and Wooldridge (2010)). In this paper, we address the problem by modelling land-use share equations, which are based on farmers' profit maximisation, as a system of simultaneous Tobit equations. Hence, we have drawn upon a set of tools recently developed for estimating censored household demand systems (see e.g. Yen et al. (2003), and Dong et al. (2004)). These are explained in detail in Section 3 below.

2.1.3. *Allowing for potential spatial error dependence (SED)*

Often the use of the spatially disaggregated data involves some degree of spatial dependence. This may be brought about by endogenous interaction effects, which indicates a spatial lag specification, or by the Durbin effects, which is an exogenous interaction counterpart. Nonetheless, these effects seem to be secondary in the context of a land-use share model. A more relevant type of dependence is the SED. Measurement errors that spill across grid boundaries, for example, can lead to the SED. Otherwise, there may exist unobservable latent variables that might be unaccounted for in the model. For instance, some specific land characteristics, which cannot be accounted for in the model due to unavailability of the data, may lead to the SED if they are spatially correlated (see also Moscone et al. (2007), and Chakir and Le Gallo (2013)). When the problem is not properly addressed, usual maximum likelihood and quasi maximum likelihood (QML) methods can be severely affected. In the current paper, we first construct a panel-data Tobit model with error components, which allow both spatial and time-wise correlations. This model forms the basis for the development of our system of land-use shares (see Sections 2.2 and 2.3 for details).

2.1.4. Modelling unobserved heterogeneity in an error component structure

In an econometric point of view, heterogeneity can be handled via either a random-effect or a fixed effect model. The choice between these two alternatives is complex and depends on the model and data. In a spatial setting, using individual fixed effects might induce an incidental parameter problem as the asymptotics in the cross-sectional dimension is necessary. Some researchers, e.g. Lee and Yu (2010), suggested methods to overcome this problem. However, none of these papers deals with a system of equations with inter-equation correlation as in our case. Moreover, in a fixed-effect model, land quality, which is a time-invariant variable, is swept away by the within estimator and the associated coefficient is not identified. Unlike previous studies, the current paper explores the use of spatially disaggregated data. An important advantage for the use of the disaggregated data is the ability to explicitly model idiosyncratic effects of policies and other physical environmental drivers (e.g. land slope/altitude). In the other words, by using the spatially disaggregated data, we can almost completely capture the heterogeneity of spatial units, whose existence is due to the differences in geographical conditions of land. However, since data limitations can hinder a complete assessment of the influence of inter-regional biophysical and socio-economic differences on land-use dynamics, here we model such left-over individual-effects via a random-effect model. Taking into consideration the above discussion, an additional assumption that the unobserved variables are uncorrelated with the regressors seems to be less problematic than opting for a fixed effect model.

2.1.5. Reducing the computational burden via a hybrid estimation procedure:

In the literature, the SED is often modelled on the basis of one of many variants of the Cliff and Ord (1973, 1981) formulations. An estimation of the Cliff-Ord specifications can be computationally burdensome. This is so even for spatial panel data models of uncensored responses (e.g. Kapoor et al. (2007) and Yang (2013)). To lighten the computational burden, Liu and Yang (2015) suggested an alternative QML procedure that involves concentrating out a subset of the parameters and maximising a concentrated log likelihood function. Nonetheless, it is not straightforward to apply such a tool to our case of censored responses. Hence, in this paper, we formulate a hybrid method that is a combination of the QML and the GMM techniques for estimating our system of simultaneous Tobit equations of land-use shares. Even though Kelejian and Prucha (1999) and Kapoor et al. (2007) have presented some key asymptotic results for the GMM procedure, in Section 3, we discuss additional properties that are crucial to the statistical validity of our hybrid framework.

2.2. Constructing the TL-RE-SED-Tobit Model for Panel Data

The censoring problem explained in Section 2.1.2 suggests that we model land-use shares based on the two-limit Tobit model of the form

$$y^*_{k,it} = x_{k,it}\beta_k + u_{k,it} \tag{2.1}$$

$$y_{k,it} = \begin{cases} 0 & \text{if } y^*_{k,it} \le 0 \\ y^*_{k,it} & \text{if } 0 < y^*_{k,it} < 1, \\ 1 & \text{if } y^*_{k,it} \ge 1 \end{cases} \tag{2.2}$$

where k signifies the kth alternative crop grown on the land, e.g. arable, $x_{k,it} = [1, x_{k,2,it}, \ldots, x_{k,J,it}]$, J denotes the number of land-use determinants included in the model, i and t signify the i-th grid of land and t-th time period, respectively. Let $k = 1, \ldots, K$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The models in (2.1) is well founded since it can be viewed as a reduced form of a well-known structural profit-maximisation problem discussed in Chambers and Just (1989), and extended to the context of agricultural land-use by Fezzi and Bateman (2011). A general form of the model is obtained by replacing 0 and 1 in (2.2) with a and b, where $a, b \in \mathbb{R}$ and $a < b$.

We now incorporate the RE-SED component into the TL-Tobit model by specifying the disturbance process in each time period as following the first order spatial autoregressive (SAR) process

$$u_k(t) = \rho_k W_k u_k(t) + \varepsilon_k(t), \tag{2.3}$$

where $u_k(t) = (u_{k,1t}, u_{k,2t}, \ldots, u_{k,Nt})^\top$ (i.e. an $N \times 1$ vector of disturbances), ρ_k is a scalar autoregressive parameter, $\varepsilon_k(t)$ is an $N \times 1$ vector of innovations in period t, W_k is an $N \times N$ weighting matrix of known constants. We also assume that innovation vector $\varepsilon_k(t)$ follows the error component structure

$$\varepsilon_k(t) = \mu_k + v_k(t), \tag{2.4}$$

where μ_k denotes a vector of the unit specific error component, which suggests that the disturbances are auto-correlated both spatially and time-wise.

With regard to the above model and specifications, we maintain the following assumptions throughout this paper.

Assumption 2.1. *(a) Let T be a fixed positive integer. (b) For all $1 \leq t \leq T$ and $1 \leq i \leq N$, where $N \geq 1$, $v_{k,it}$ are identically and independently distributed (iid) with zero mean, variance of $0 < \sigma^2_{k,v} < b_v < \infty$, and finite fourth moment. Also, $E(v_{k,it}|x_{k,it}) = 0$ almost surely. (c) For all $1 \leq i \leq N$, where $N \geq 1$, the unit-specific error components $\mu_{k,i}$ are iid with zero mean, the variance of $0 < \sigma^2_{k,\mu} < b_\mu < \infty$, and finite fourth moment. Also, $E(\mu_{k,i}|x_{k,it}) = 0$ almost surely. (d) The processes $\{v_{k,it}\}$ and $\{\mu_{k,i}\}$ are independent.* $\square$

Assumption 2.1(a) suggests that our analysis concerns the case where T is fixed and $N \to \infty$. Assumptions 2.1(b) and (c) imply $E(\varepsilon_{k,it}) = 0$ and

$$E(\varepsilon_{k,it}\varepsilon_{k,js}) = \begin{cases} \sigma^2_{k,\mu} + \sigma^2_{k,v} & \text{if } i = j;\ t = s \\ \sigma^2_{k,\mu} & \text{if } i = j;\ t \neq s. \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

In the other words, the innovations $\varepsilon_{k,it}$ are temporally correlated within a unit, but are not spatially correlated across units.

Moreover, concatenation of the innovation vector with respect to time $t = 1, \ldots, T$ leads to $\varepsilon_k = (e_T \otimes I_N)\mu_k + v_k$, where $\otimes$ denotes the Kronecker product, e_T is a $T \times 1$ vector of 1s, I_N is an identity matrix of size N, and $v_k = (v_k^\top(1), v_k^\top(2), \ldots, v_k^\top(T))^\top = (v_{k,11}, v_{k,21}, \ldots, v_{k,N1}, v_{k,12}, \ldots, v_{k,NT})^\top$. This suggests that $E(\varepsilon_k) = 0$ and covariance matrix $E(\varepsilon_k\varepsilon_k^\top)$ of the form

$$\Omega_{k,\varepsilon} = \sigma^2_{k,v} I_{NT} + \sigma^2_{k,\mu}(J_T \otimes I_N) = \sigma^2_{k,v} Q_0 + \sigma^2_{k,1} Q_1, \tag{2.6}$$

where I_{NT} denotes an identity matrix of size NT, $\sigma^2_{k,1} = \sigma^2_{k,v} + T\sigma^2_{k,\mu}$, $Q_0 = \left(I_T - \frac{J_T}{T}\right) \otimes I_N$ and $Q_1 = \frac{J_T}{T} \otimes I_N$ in which $J_T = e_T e_T'$ is a $T \times T$ matrix of unit elements. In (2.6), Q_0 and Q_1 are transformation matrices often used in the error component literature (see e.g. Baltagi (2008)). These matrices are symmetric, idempotent, orthogonal to each other and satisfy the following properties: (i) $Q_0 + Q_1 = I_{NT}$, (ii) $TR(Q_0) = N(T-1)$ and $TR(Q_1) = N$, and (iii) $Q_0 Q_1 = 0$. In the light of these properties,

$$\Omega_{k,\varepsilon}^{-1} = \sigma_{k,v}^{-2} Q_0 + \sigma_{k,1}^{-2} Q_1 \quad \text{and} \quad \Omega_{k,\varepsilon}^{-1/2} = \sigma_{k,v}^{-1} Q_0 + \sigma_{k,1}^{-1} Q_1. \tag{2.7}$$

A similar concatenation to (2.3) also leads to

$$u_k = \rho_k (I_T \otimes W_k) u_k + \varepsilon_k = [I_T \otimes (I_N - \rho_k W_k)^{-1}] \varepsilon_k, \tag{2.8}$$

which we maintain the following assumptions throughout this paper.

Assumption 2.2. *(a) The matrix $I_N - \rho_k W_k$ is nonsingular. (b) $|\rho_k| < 1$. (c) All diagonal elements of W_k are zero.* □

Assumption 2.2(a) ensures that the model is closed, in the sense that it can be uniquely solved for the disturbance u_k in terms of the innovation ε_k, whereas Assumption 2.2(c) is a normalisation, which implies that no unit is related in a meaningful way or being a neighbour to itself. Although the elements of W_k are assumed to be nonvarying over t, they are allowed to depend on the cross-sectional dimension N (i.e. they are allowed to form a triangular array). This corresponds to models in which the weighting matrix is row-normalised and the number of neighbors for a given unit depends on the sample size. In this respect, we also assume:

Assumption 2.3. *Row and column sums of W_k and $H_k = (I_N - \rho_k W_k)^{-1}$ are bounded in absolute values by $c_W < \infty$ and $c_H < \infty$, respectively.* □

Accordingly, $E(u_k) = 0$ and covariance matrix $E(u_k u_k^\top)$ is of the form

$$\Omega_{k,u} = \left[I_T \otimes (I_N - \rho_k W_k)^{-1}\right] \Omega_{k,\varepsilon} \left[I_T \otimes (I_N - \rho_k W_k^\top)^{-1}\right]. \tag{2.9}$$

It is useful to also note $\Omega_{k,u}^{-1} = \left[I_T \otimes (I_N - \rho_k W_k^\top)\right] \Omega_{k,\varepsilon}^{-1} \left[I_T \otimes (I_N - \rho_k W_k)\right].$ From (2.9), it is clear that the variance-covariance matrix of the disturbance vector u_k is proportional to H_k. Since this property is preserved under matrix multiplication, Assumption 2.3 implies that the row/column sums of this matrix are bounded uniformly in absolute values, which restricts the degree of cross-sectional correlation between the model disturbances.

2.3. System Variance-Covariance Structure

We now specify the system variance-covariance structure. Consider first the disturbance $u = (u_1^\top, \dots, u_K^\top)^\top$. In accordance with the definition in (2.8), covariance matrix of the system disturbance $E(uu^\top)$ is

$$\Omega_u = A \Omega_\varepsilon A^\top, \tag{2.10}$$

where $A = \text{diag}(A_{11}, \ldots, A_{KK})$ and $A_{kk} = I_T \otimes H_k$. In the other words, covariance matrix $E[u_k u_l^\top]$ can be expressed as

$$\Omega_{kl,u} = E[\varepsilon_k \varepsilon_l^\top]\{I_T \otimes H_k H_l^\top\}, \tag{2.11}$$

where $H_k = B_k^{-1}$ and $B_k = (I_N - \rho_k W_k)$. Now we discuss the covariance of the innovations $\varepsilon = (\varepsilon_1^\top, \ldots, \varepsilon_K^\top)^\top$. We assume that the cross-equation correlation of the innovations is driven by

$$E\begin{pmatrix} \mu_k \\ v_k \end{pmatrix}\begin{pmatrix} \mu_l^\top & v_l^\top \end{pmatrix} = \begin{pmatrix} \sigma^2_{kl,\mu}(J_T \otimes I_N) & 0 \\ 0 & \sigma^2_{kl,v} I_{NT} \end{pmatrix}, \tag{2.12}$$

where $k, l = 1, \ldots, K$. Accordingly, covariance matrix of the innovations, i.e. $E(\varepsilon\varepsilon^\top)$, is

$$\Omega_\varepsilon = \Omega_v \otimes Q_0 + \Omega_1 \otimes Q_1 = [\Omega_{kl,\varepsilon}], \tag{2.13}$$

where $\Omega_\mu = [\sigma^2_{kl,\mu}]$ and $\Omega_v = [\sigma^2_{kl,v}]$ both with dimension $K \times K$, such that $\Omega_{kl,\varepsilon}$ is $E(\varepsilon_k \varepsilon_l^\top)$ defined as

$$\Omega_{kl,\varepsilon} = \sigma^2_{kl,\mu}(J_T \otimes I_N) + \sigma^2_{kl,v} I_{NT}, \tag{2.14}$$

which is in line with (2.6), where $\sigma^2_{kl,v} = E(v_k v_l^\top)$ and $\sigma^2_{kl,\mu} = E(\mu_k \mu_l^\top)$. Alternatively,

$$\Omega_{kl,\varepsilon} = \sigma^2_{kl,v} Q_0 + \sigma^2_{kl,1} Q_1 \tag{2.15}$$

obtained by defining $\sigma^2_{kl,1} = \sigma^2_{kl,v} + T\sigma^2_{kl,\mu}$. In relation to (2.7), we also write

$$\Omega_\varepsilon^{-1/2} = \Omega_v^{-1/2} \otimes Q_0 + \Omega_1^{-1/2} \otimes Q_1. \tag{2.16}$$

Finally, the use of (2.14) in (2.11) leads to

$$\Omega_{kl,u} = \left\{\sigma^2_{kl,1} \bar{J}_T + \sigma^2_{kl,v}(I_T - \bar{J}_T)\right\} \otimes H_k H_l^\top, \tag{2.17}$$

where $\bar{J}_T = J_T/T$.

2.4. Other Useful Results and Transformations

We finish this section by presenting a set of results that will be useful for the discussion that follows. Firstly, let $\omega_{k,i}$ signify covariance between future and the current disturbances, $E[u_{k,i,T+\tau} u_k^\top]$. Deriving $\omega_{k,i}$ in the context of the TL-RE-SED model requires first noting that $u_k(t) = B_k^{-1}(\mu_k + v_k(t))$ and $u_k = (e_T \otimes H_k)\mu_k + (I_T \otimes H_k)v_k$. In this regard,

$$\begin{aligned} E[u_k(T+\tau)u_k^\top] &= E[B_k^{-1}(\mu_k + v_k(T+\tau))((e_T \otimes H_k)\mu_k + (I_T \otimes H_k)v_k)^\top] \\ &= \sigma^2_{kk,\mu} H_k (e_T^\top \otimes H_k^\top), \end{aligned}$$

where $\sigma^2_{kk,\mu} = E[\mu_k \mu_k^\top]$, which is $N \times TN$. As the results,

$$E[u_{k,i,T+\tau} u_k^\top] = \sigma^2_{kk,\mu} h_{k,i} (e_T^\top \otimes H_k^\top), \tag{2.18}$$

where $h_{k,i}$ is the i-th row of $H_k = B_k^{-1}$, for an individual i at time $T + \tau$. Moreover, recall

$$\begin{aligned} \Omega_{kk,u} &= E[\varepsilon_k \varepsilon_k^\top]\{I_T \otimes H_k H_k^\top\} \\ &= \left\{\sigma^2_{kk,1} \bar{J}_T + \sigma^2_{kk,v}(I_T - \bar{J}_T)\right\} \otimes H_k H_k^\top, \end{aligned} \tag{2.19}$$

which were presented previously in (2.11) and (2.17). In this regard, (2.19) and (2.18) suggest collectively that

$$\begin{aligned} \omega_{k,i}^\top \Omega_{kk,u}^{-1} &= \frac{\sigma^2_{kk,\mu}}{\sigma^2_{kk,v}} h_{k,i}(e_T^\top \otimes H_k^\top) \left[I_T \otimes B_k B_k^\top - \frac{T\sigma^2_{kk,\mu}}{\sigma^2_{kk,1}} \bar{J}_T \otimes B_k B_k^\top \right] \\ &= \frac{\sigma^2_{kk,\mu}}{\sigma^2_{kk,1}} h_{k,i}(e_T^\top \otimes B_k) \end{aligned} \tag{2.20}$$

since $e_T^\top = e_T^\top \bar{J}_T$ and $\frac{\sigma^2_{kk,\mu}}{\sigma^2_{kk,v}} - \left(\frac{\sigma^2_{kk,\mu}}{\sigma^2_{kk,v}} \times \frac{T\sigma^2_{kk,\mu}}{\sigma^2_{kk,1}} \right) = \frac{\sigma^2_{kk,\mu}}{\sigma^2_{kk,1}}$. Since $h_{k,i}$ is the i-th row of $H_k = B_k^{-1}$ and $B_k^{-1} B_k = I_N$, $h_{k,i} B_k = l_{k,i}^\top$, where $l_{k,i}^\top$ is the i-th row of I_N,$h_{k,i}(e_T^\top \otimes B_k) = (1 \otimes h_{k,i})(e_T^\top \otimes B_k) = (e_T^\top \otimes l_{k,i}^\top)$, which is $(1 \times TN)$. Hence,

$$\omega_{k,i}^\top \Omega_{kk,u}^{-1} = \frac{\sigma^2_{kk,\mu}}{\sigma^2_{kk,1}} (e_T^\top \otimes l_{k,i}^\top). \tag{2.21}$$

We shall revisit this result in Section 3.4 when we discuss prediction under the TL-RE-SED Tobit specification.

In addition, results in the previous section allows derivation of a set of transformations that are essential for the discussion in the next section. To this end, let $Y = [Y_1^\top, \ldots, Y_K^\top]^\top$, where $Y_k = [Y_k^\top(1), \ldots, Y_k^\top(T)]^\top$ and $Y_k(t) = [y_{k,1t}, \ldots, y_{k,Nt}]^\top$, and let $X = \text{diag}[x_1, x_2, \ldots, x_K]$, where $x_k = [x_k^\top(1), \ldots, x_k^\top(T)]^\top$ and $x_k(t) = [x_{k,1t}^\top, \ldots, x_{k,Nt}^\top]^\top$. Firstly, it is the Cochrane-Orcutt-type transformation

$$\dot{X} = A^{-1} X \quad \text{and} \quad \dot{Y} = A^{-1} Y. \tag{2.22}$$

Guided by the classical error component literature, we can also include the RE-GLS-type transformation to obtain the *"Cochrane-Orcutt plus RE-GLS transformations"* of the form

$$\ddot{X} = \Omega_\varepsilon^{-1/2} \dot{X} \ \text{ and } \ \ddot{Y} = \Omega_\varepsilon^{-1/2} \dot{Y}. \tag{2.23}$$

In relation to (2.22) and (2.23), let us also define

$$\dot{u} = A^{-1}u \quad \text{and} \quad \ddot{u} = \Omega_{\varepsilon}^{-1/2}\dot{u}. \tag{2.24}$$

While the former is equivalent by definition to ε, we assume that the latter has a contemporaneous error correlation matrix $[r_{kl}]$ for $k, l = 1, \ldots, K$.

3. Hybrid QML/GMM Estimation Procedure

In the current section, we propose a hybrid QML-GMM procedure for estimating the above-discussed system of simultaneous Tobit equations for land-use shares. Overall, our procedure consists of four key steps, namely: (3.1) Estimating the TL-Tobit panel data model of land-use shares for all the K categories. Our objective is to obtain consistent estimates of the disturbances $u_{k,it}$. (3.2) Performing a GMM estimation to obtain consistent estimates of the spatial parameters ρ_k for all $k = 1, \ldots, K$ and constructing the Cochrane-Orcutt transformations $\dot{Y}$ and $\dot{X}$. (3.3) Estimating system of TL-RE-SED Tobit equations of the land-use shares under consideration with QML in order to obtain estimates of the parameters $\beta = [\beta_1^\top, \ldots, \beta_K^\top]^\top$, and the elements of $\Omega_v = [\sigma^2_{kl,v}]$ and $\Omega_1 = [\sigma^2_{kl,1}]$. (3.4) Constructing land-use prediction. Performing Steps 3.1 to 3.3 should provide sufficient information for analysing causal effects of the various drivers on agricultural land-use shares. Therefore, the final step can be viewed as a supplemental step, which is useful for policy making. Below we will discuss these steps in turn.

3.1. Estimating the TL-Tobit Panel Data Models

The current step involves estimating the TL-Tobit panel data model for the k-th category of land-use using QML estimation. Our objective is to obtain a consistent estimate of the disturbance, $u_{k,it}$. A number of issues must be taken into consideration to this end.

3.1.1. Heteroscedasticity

With regard to the QML estimation, it is well known that presence of heteroscedasticity is likely to lead to inconsistent estimates. However, consistent estimation is possible by specifying a model for heteroscedasticity. Particularly, let

$$\sigma_{k,u,it} = \exp(z_{k,it}\alpha_k), \tag{3.1}$$

where $z_{k,it} = [1, z_{k,1,it}, \ldots, z_{k,J,it}]$ and "J" is used with a slight abuse of notation since it may not be the same as the number of determinants in (2.1). Here, we assume a multiplicative error specification as is often done in the auto-regressive heteroskedasticity literature (see e.g. Tsay (2005)).

3.1.2. Pooled QML estimation

Following the popular pooled method, pooled QML estimators maximise the quasi-log-likelihood function

$$\mathcal{L}_{k,N} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \ell_{k,it}(\bar{\beta}_k, \bar{\alpha}_k), \tag{3.2}$$

where $\ell_{k,it}(\bar{\beta}_k, \bar{\alpha}_k)$ is log-likelihood function for the it-th observation, i.e.

$$\begin{aligned} \ell_{k,it}(\bar{\beta}_k, \bar{\alpha}_k) \;=\; & 1[y_{k,it} = 0] \;\log\left[\Phi((-x_{k,it}\bar{\beta}_k)/\sigma_{k,u,it}(\bar{\alpha}_k))\right] \\ + \; & 1[0 < y_{k,it} < 1] \;\log\left[(1/\sigma_{k,u,it}(\bar{\alpha}_k)\phi((y_{k,it} - x_{k,it}\bar{\beta}_k)/\sigma_{k,u,it}(\bar{\alpha}_k))\right] \\ + \; & 1[y_{k,it} = 1] \;\log\left[\Phi(-(1 - x_{k,it}\bar{\beta}_k)/\sigma_{k,u,it}(\bar{\alpha}_k))\right], \end{aligned}$$

and $1[\,\cdot\,]$ signifies an indicator function. Lemma 3.1 below confirms that consistent estimates of the disturbances can be obtained using the above QML estimation. The proof of this lemma requires some additional assumptions.

Assumption 3.1. *(a) x_k has a full column rank, i.e. $rank(x_k) = J$, where $J < \infty$. (b) For a column of x_k, i.e. $x_{k,l}$, $\lim_{N\to\infty} x_{k,l}^\top x_{k,l} \to \infty$, $\lim_{N\to\infty} x_{k,l,it}^2/x_{k,l}^\top x_{k,l} \to 0$ and $E(x_{k,l,it}^4) < \infty$ for all $l = 1, \ldots, J$ and $it = 11, 21, \ldots, N1, 12, \ldots, NT$. (c) The empirical distribution function $G_{k,N}$ (defined by $G_{k,N}(x_k) = j/NT$ where j is the number of points $x_{k,it} \leq x_k$) converges to a distribution function G_k for all $it = 11, 21, \ldots, N1, 12, \ldots, NT$ and $k = 1, \ldots, K$.* □

With the exception of the finite fourth moment condition on $x_{k,l,it}$, which is necessary for the proof of Theorem 1 below, Assumption 3.1 is standard in the Tobit model literature (see e.g. Amemiya (1973)).

Lemma 3.1. *Let $\mathbb{B}_1$ denote the vector of true parameters $(\beta_k^\top, \alpha_k^\top)^\top$ and $\widehat{\mathbb{B}}_1$ be the QML estimator of $\mathbb{B}_1$. Under Assumptions 2.1 to 2.3 and 3.1, $\mathbb{B}_1$ is uniquely identifiable and $\widehat{\mathbb{B}}_1 = \mathbb{B}_1 + O_p((NT)^{-1/2})$ as $N \to \infty$.* □

3.1.3. Standardised residuals

Upon completion of the above estimation, the required standardised residuals for uncensored observations are constructed as

$$\widetilde{u}_{k,it}/\widetilde{\sigma}_{k,u,it} = (y_{k,it} - x_{k,it}\widetilde{\beta}_k)/\widetilde{\sigma}_{k,u,it}, \tag{3.3}$$

where $\widetilde{\sigma}_{k,u,it} = \exp(z_{k,it}\widetilde{\alpha}_k)$ as in equation (3.1), and $\widetilde{\beta}_k$ and $\widetilde{\alpha}_k$ are the QML parameter estimates from Step 3.1.2. Otherwise, generalised residuals for censored observations are computed via the inverse Mills ratio

$$\lambda_{k,it} = \phi(x_{k,it}\beta_k/\sigma_{k,u,it})/\{\Phi(x_{k,it}\beta_k/\sigma_{k,u,it})\}, \tag{3.4}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote normal density function and the cumulative distribution function, respectively. For instance, the generalised residuals can be computed as $\widetilde{u}_{k,it} = -\widetilde{\lambda}_{k,it}$ for observations left-censored at 0, where $\widetilde{\lambda}_{k,it}$ is obtained using (3.4) and by plugging in the parameter estimates from Section 3.1.2.

3.2. Estimating ρ_k & Constructing Cochrane-Orcutt Transformations

Now we use $\widetilde{u}_k = (\widetilde{u}_{k,11}, \widetilde{u}_{k,21}, \ldots, \widetilde{u}_{k,N1}, \widetilde{u}_{k,12}, \ldots, \widetilde{u}_{k,NT})^\top$ in place of the true disturbances in order to obtain estimates for ρ_k by using the GMM procedure introduced in Kapoor et al. (2007). To be accustomed to such a practice, one only has to note that generalised residuals are commonly used for performing diagnostic tests in a standard Tobit model literature (see e.g. Cameron and Trivedi (2005)). Although the QML estimation is used in Step 3.1 unlike Kapoor et al. (2007), who employed the ordinary least squares, consistency of our GMM estimators for ρ_k, $\sigma^2_{k,v}$ and $\sigma^2_{k,1}$ can be shown in a similar fashion.

Lemma 3.2. *Let $\mathbb{B}_2$ denote the vector of true parameters $(\rho_k, \sigma^2_{k,v}, \sigma^2_{k,1})^\top$ and $\widehat{\mathbb{B}}_2$ is the GMM estimator of $\mathbb{B}_2$. Under Assumptions 2.1 to 2.3 and 3.1, $\mathbb{B}_2$ is uniquely identifiable and $\widehat{\mathbb{B}}_2 = \mathbb{B}_2 + O_p((NT)^{-1/2})$ as $N \to \infty$.* □

The underlying moment conditions and weight matrices for the above GMM estimators, and the proof of Lemma 3.2 are discussed in detail in the online appendix. Once the above estimation is completed for all K categories of land-use, GMM estimates $\widehat{\rho}_1, \ldots, \widehat{\rho}_K$ of the autoregressive parameters are readily available. The remainder of the current step focuses on estimating the matrix A_{kk} via expression (2.10). Particularly, $\widehat{A}_{kk} = I_T \otimes \widehat{H}_k$, where $\widehat{H}_k = (I_N - \widehat{\rho}_k W_k)^{-1}$. These are then used for computing the Cochrane-Orcutt transformations of Y and X, $\dot{\mathcal{X}} = \widehat{A}^{-1}\widetilde{X}$ and $\dot{\mathcal{Y}} = \widehat{A}^{-1}\widetilde{Y}$ where $\widetilde{Y} = [\widetilde{Y}_1^\top, \widetilde{Y}_2^\top, \ldots, \widetilde{Y}_K^\top]^\top$ and $\widetilde{X} = \mathrm{diag}[\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_K]$ are the standardised versions of Y and X with respect to $\widetilde{\sigma}_{k,u,it}$, respectively.

In order to discuss the estimation in the next step, it is useful to clarify the following notational issues. Firstly, it is a rewriting of the space-time subscripts it to $\iota = 1, 2, \ldots, NT$. Particularly let,

$$\dot{\mathcal{Y}}_k = (\dot{y}_{k,11}, \ldots, \dot{y}_{k,N1}, \dot{y}_{k,12}, \ldots, \dot{y}_{k,NT})^\top \equiv (\dot{y}_{k,1}, \dot{y}_{k,2}, \ldots, \dot{y}_{k,NT})^\top$$

and $\dot{x}_{k,\iota} = [\dot{x}_{k,1,\iota}, \ldots, \dot{x}_{k,J,\iota}]$, which is the ι-th row of $\dot{\mathcal{X}}_k$. Secondly, it is the notational distinction between these transformations, and

$$\dot{Y}_k = (\dot{y}_{k,1}, \dot{y}_{k,2}, \ldots, \dot{y}_{k,NT})^\top \text{ and } \dot{x}_{k,\iota} = [\dot{x}_{k,1,\iota}, \ldots, \dot{x}_{k,J,\iota}], \tag{3.5}$$

which are the k-th element of $\dot{Y}$ and ι-th row of $\dot{X}_k$, respectively. We recall that $\dot{Y}$ and $\dot{X}_k$ denote the conceptual Cochrane-Orcutt transformations in which the autoregressive parameters are known.

3.3. Estimating system of TL-RE-SED Tobit equations

We first note an important drawback of the traditional Amemiya-Tobin mechanism. This resides in the fact that the adding-up restriction, which was discussed in Section 2, holds only for the latent equations (i.e. equation (2.1)), but not for the observed land-use shares. Here, we address such an issue by treating the Kth use of land as a residual category with no specific land-use demand of its own. Therefore, the current step focuses on estimating the system of TL-RE-SED Tobit models of land-use shares for the total of $\mathcal{K} = K - 1$ categorises (see e.g. Fezzi and Bateman (2011), and Chakir and Le Gallo (2013) who have also followed this approach).

Moreover, an estimation of a Tobit system requires evaluating multiple Gaussian integrals, which is computationally expensive when there are more than three equations. Recent studies on the estimation of consumer demand system suggested a few approaches to alleviate this problem. In this paper, we follow a suggestion made by Yen et. al. (2003) and specify the likelihood function based on a sequence of bivariate Tobit likelihoods. To elaborate, let $\theta = [\beta^\top, S_v^\top, S_1^\top, S_{\ddot{u}}^\top]^\top$ be vector of true parameters in a system of $\mathcal{K}$ TL-RE-SED Tobit equations, where $\beta = [\beta_1^\top, \ldots, \beta_{\mathcal{K}}^\top]^\top$, $S_m = [\sigma^2_{11,m}, \ldots, \sigma^2_{\mathcal{K}\mathcal{K},m}, \sigma^2_{12,m}, \ldots, \sigma^2_{\mathcal{K}-1,\mathcal{K},m}]^\top$, $S_{\ddot{u}} = [\sigma_1, \ldots, \sigma_{\mathcal{K}}, r_{11}, \ldots, r_{\mathcal{K},\mathcal{K}-1}]^\top$, and $\mathcal{K} = K - 1$. QML estimators of the vector of true parameters θ can be obtained by maximising the quasi-likelihood

$$L = \prod_{\iota=1}^{NT} \left(L_{1,\mathcal{K},\iota} \prod_{k=2}^{\mathcal{K}} \prod_{j=1}^{k-1} L_{k,j,\iota} \right) \tag{3.6}$$

in which

$$\begin{aligned} L_{k,j,\iota} &= \{\Psi(\hbar_{k,\iota}\hbar_{j,\iota}; \bar{r}_{kj})\}^{1[y_{k,\iota}=0, y_{j,\iota}=0]} \\ &\times \left\{\bar{\sigma}_k^{-1}\bar{\sigma}_j^{-1}(1-\bar{r}_{kj}^2)^{-1/2}\psi(\hbar_{k,\iota}, \hbar_{j,\iota}; \bar{r}_{kj})\right\}^{1[0<y_{k,\iota}<1, 0<y_{j,\iota}<1]} \\ &\times \{\Psi(-\hbar_{k,\iota}, \hbar_{j,\iota}; -\bar{r}_{kj})\}^{1[y_{k,\iota}=1, y_{j,\iota}=0]} \\ &\times \{\Psi(\hbar_{k,\iota}, -\hbar_{j,\iota}; \bar{r}_{kj})\}^{1[y_{k,\iota}=0, y_{j,\iota}=1]} \\ &\times \left\{\bar{\sigma}_k^{-1}\phi(\hbar_{k,\iota})\Phi\left[(\hbar_{j,\iota} - \bar{r}_{kj}\hbar_{k,\iota})/(1-\bar{r}_{kj}^2)^{1/2}\right]\right\}^{1[0<y_{k,\iota}<1, y_{j,\iota}=0]} \\ &\times \left\{\bar{\sigma}_j^{-1}\phi(\hbar_{j,\iota})\Phi\left[(\hbar_{k,\iota} - \bar{r}_{kj}\hbar_{j,\iota})/(1-\bar{r}_{kj}^2)^{1/2}\right]\right\}^{1[y_{k,\iota}=0, 0<y_{j,\iota}<1]}, \end{aligned} \tag{3.7}$$

where $\hbar_{k,\iota} = [\ddot{y}_{k,\iota} - \ddot{x}_{k,\iota}\bar{\beta}_k]/\bar{\sigma}_k$, $\hbar_{j,\iota} = [\ddot{y}_{j,\iota} - \ddot{x}_{j,\iota}\bar{\beta}_j]/\bar{\sigma}_j$, $1[y_{k,\iota} = 0, y_{j,\iota} = 0]$ is a dichotomous indicator which equals 1 when $y_{k,\iota} = 0$ and $y_{j,\iota} = 0$, and $\psi(\cdot, \cdot, \cdot)$

and $\Psi(\cdot,\cdot,\cdot)$ are the bivariate standard normal probability density function and corresponding cumulative distribution, respectively. Furthermore, $\ddot{\mathcal{Y}}_{k,\iota}$ and $\ddot{\mathcal{X}}_{k,\iota}$ are elements of $\ddot{\mathcal{Y}} = \bar{\Omega}_\varepsilon^{-1/2}\dot{\mathcal{Y}}$ and $\ddot{\mathcal{X}} = \bar{\Omega}_\varepsilon^{-1/2}\dot{\mathcal{X}}$ in which

$$\bar{\Omega}_\varepsilon^{-1/2} = \bar{\Omega}_v^{-1/2} \otimes Q_0 + \bar{\Omega}_1^{-1/2} \otimes Q_1,$$

where $\bar{\Omega}_v^{-1/2} = [\bar{\sigma}^2_{kl,v}]$ and $\bar{\Omega}_1^{-1/2} = [\bar{\sigma}^2_{kl,1}]$.

To establish consistency of the proposed estimation requires first defining the following counterpart of (3.6)

$$L^0 = \prod_{\iota=1}^{NT}\left(L^0_{1,\mathcal{K},\iota}\prod_{k=2}^{\mathcal{K}}\prod_{j=1}^{k-1}L^0_{k,j,\iota}\right), \tag{3.8}$$

which is constructed under an assumption that the spatial parameter ρ_k is known. In this regard,

$$\begin{aligned}
L^0_{k,j,\iota} \;&=\; \{\Psi(h_{k,\iota}h_{j,\iota};\bar{r}_{kj})\}^{1[y_{k,\iota}=0,y_{j,\iota}=0]} \\
&\times\; \left\{\bar{\sigma}_k^{-1}\bar{\sigma}_j^{-1}(1-\bar{r}_{kj}^2)^{-1/2}\psi(h_{k,\iota},h_{j,\iota};\bar{r}_{kj})\right\}^{1[0<y_{k,\iota}<1,0<y_{j,\iota}<1]} \\
&\times\; \{\Psi(-h_{k,\iota},h_{j,\iota};-\bar{r}_{kj})\}^{1[y_{k,\iota}=1,y_{j,\iota}=0]} \\
&\times\; \{\Psi(h_{k,\iota},-h_{j,\iota};\bar{r}_{kj})\}^{1[y_{k,\iota}=0,y_{j,\iota}=1]} \\
&\times\; \left\{\bar{\sigma}_k^{-1}\phi(h_{k,\iota})\Phi\left[(h_{j,\iota}-\bar{r}_{kj}h_{k,\iota})/(1-\bar{r}_{kj}^2)^{1/2}\right]\right\}^{1[0<y_{k,\iota}<1,y_{j,\iota}=0]} \\
&\times\; \left\{\bar{\sigma}_j^{-1}\phi(h_{j,\iota})\Phi\left[(h_{k,\iota}-\bar{r}_{kj}h_{j,\iota})/(1-\bar{r}_{kj}^2)^{1/2}\right]\right\}^{1[y_{k,\iota}=0,0<y_{j,\iota}<1]},
\end{aligned} \tag{3.9}$$

where $h_{k,\iota} = [\ddot{y}_{k,\iota} - \ddot{x}_{k,\iota}\bar{\beta}_k]/\bar{\sigma}_k$ and $h_{j,\iota} = [\ddot{y}_{j,\iota} - \ddot{x}_{j,\iota}\bar{\beta}_j]/\bar{\sigma}_j$. Furthermore, $\ddot{y}_{k,\iota}$ and $\ddot{x}_{k,\iota}$ are elements of $\ddot{Y} = \bar{\Omega}_\varepsilon^{-1/2}\dot{Y}$ and $\ddot{X} = \bar{\Omega}_\varepsilon^{-1/2}\dot{X}$, where $\dot{Y}$ and $\dot{X}$ are defined in (2.22), i.e. under the assumption that the spatial parameter ρ_k is known (see also (3.5)). In this regard, establishing the consistency of the proposed QML estimation requires showing that

$$\mathcal{L}(\bar{\theta}) = \mathcal{L}^0(\bar{\theta}) + O_P\left((NT)^{-1/2}\right) \tag{3.10}$$

uniformly over a compact parameter space Θ, where $\bar{\theta} \in \Theta$, and $\mathcal{L}$ and $\mathcal{L}^0$ represent $\frac{1}{NT}\ln L$ and $\frac{1}{NT}\ln L^0$, respectively. Theorem 3.1 below presents the consistency of the proposed QML estimation.

Theorem 3.1. *Let* $\inf_{u_{k,\iota}u_{j,\iota}\in\mathbb{R}^2}\psi(u_{k,\iota},u_{j,\iota})=\delta_1$ *and* $\inf_{u_{k,\iota}\in\mathbb{R}}\phi(u_{k,\iota})=\delta_2$*, where* $\delta_l>0$ *is an arbitrary small value for* $l=1$ *or* 2*. In addition, let*

$$\limsup_{N\to\infty}\left\{\max_{\bar{\theta}\in\bar{D}_\delta(\theta)\cap\Theta}E\mathcal{L}^0(\bar{\theta})\right\}\neq\limsup_{N\to\infty}E\mathcal{L}^0(\theta)$$

for any $\bar{\theta}$*, where* $\bar{D}_\delta(\theta)$ *is the complement of the* δ*-neighborhood of* θ*. Then, under the conditions of Lemma 3.2,* θ *is uniquely identified and* $\hat{\theta}=\theta+O_P\left((NT)^{-1/2}\right)$ *as* $N\to\infty$*.* $\square$

By taking into consideration the consistency of the GMM estimation, i.e. Lemma 3.2, and that presented in Theorem 3.1, the asymptotic normality and variance formula of our estimators are in line with those of a standard QML for system of Tobit equations. This is in conformity with standard results in the literature, e.g. Theorem 4 of Kapoor et al. (2007) who based their claim of asymptotic normality of their feasible GLS estimators on a similar set of consistency. To discuss asymptotic normality of a standard QML for system of Tobit equations, let us begin with Amemiya (1973) who showed such a result under mild regularity conditions for a univariate Tobit model with normal disturbances. With respect to our model, since there are no irregularities for our multivariate generalisations, Amemiya's analysis can be generalised to establish asymptotic normality of the QML estimators for the multivariate Tobit models. Such a generalisation was previously discussed in e.g. Lee (1993), Wooldridge (2010), and Deng and Xue (2014). An important point to note, however, is the fact that, by specifying the likelihood function based on a sequence of bivariate Tobit likelihoods (e.g. (3.8)), the QML estimators provide the most efficient parameter estimates if and only if the quasi-likelihood function is the true likelihood function of the data. However, it is not possible to theoretically derive such loss of efficiency without imposing further assumptions on the data generating process.

Remark 3.1. *In practice, we may perform the estimation discussed in Step 3.3 by using a similar iterative steps to that in Wang and Kockelman (2007), and Baltagi and Pirotte (2011). That is to first estimate* S_v *and* S_1 *conditional upon the estimate of* β *from Step 3.1.2. Secondly, it is to estimate* β *conditional upon the above estimates of* S_v *and* S_1*. These two steps are iterated until the optimal estimates of* β*,* S_v *and* S_1 *are found.*

Remark 3.2. *An alternative way to estimate the causal parameters* β *and to obtain efficiency gain is to make use of the the knowledge about* S_v *and* S_1*. This involves computing the Cochrane-Orcutt plus RE-GLS transformations*

$$\ddot{\mathcal{X}}=\widehat{\Omega}_\varepsilon^{-1/2}\dot{\mathcal{X}}\quad\textit{and}\quad\ddot{\mathcal{Y}}=\widehat{\Omega}_\varepsilon^{-1/2}\dot{\mathcal{Y}},\tag{3.11}$$

where $\widehat{\Omega}_{\varepsilon}^{-1/2} = \widehat{\Omega}_{v}^{-1/2} \otimes Q_0 + \widehat{\Omega}_{1}^{-1/2} \otimes Q_1$, and performing the final equation-by-equation TL-Tobit estimation, which was explained in Section 3.1.2, by using the resulting transformations. Since we will make use of this estimation strategy in our empirical analysis in Section 4, we provide a discussion of the validity of this procedure in detail. However, such a discussion is delegated to the online appendix due to space limitation.

3.4. Prediction under the TL-RE-SED Tobit Model

Previous studies in econometrics (e.g. Baltagi and Li (2006), Chakir and Gallo (2013), Baltagi et al. (2012) and Ay et al. (2017)) show that

$$\widehat{y}^{*}_{k,i,T+\tau} = X_{k,i,T+\tau}\widehat{\beta}_k + \widehat{\omega}_{k,i}^{\top}\widehat{\Omega}_{kk,u}^{-1}\widehat{u}_k \tag{3.12}$$

is the best linear unbiased predictor of the i-th individual at the future period $T+\tau$. The derivation in Section 2.4 suggests that we compute

$$\widehat{\omega}_{k,i}^{\top}\widehat{\Omega}_{kk,u}^{-1} = \frac{\widehat{\sigma}^2_{kk,\mu}}{\widehat{\sigma}^2_{kk,1}}(e_T^{\top} \otimes l_{k,i}^{\top}) \quad \text{and} \quad \widehat{\sigma}^2_{kk,\mu} = (\widehat{\sigma}^2_{kk,1} - \widehat{\sigma}^2_{kk,v})/T. \tag{3.13}$$

Moreover, $\widehat{\beta}_k$, $\widehat{\sigma}^2_{kk,1}$, and $\widehat{\sigma}^2_{kk,v}$ are parameter estimates obtained in Step 3.3. Using $\widetilde{u}_k = (\widetilde{u}_{k,11}, \widetilde{u}_{k,21}, \ldots, \widetilde{u}_{k,N1}, \widetilde{u}_{k,12}, \ldots, \widetilde{u}_{k,NT})^{\top}$ from Step 3.1.3 to represent $\widehat{u}_k$ suggests that $\widehat{y}^{*}_{k,i,T+\tau}$ is the predictor of the latent variable $y^{*}_{k,i,T+\tau}$, which brings about inclusion of the "$*$" superscript. As a result, the best linear unbiased predictor of the land-use share for the i-th plot of land at the future period $T+\tau$ is computed as

$$\widehat{y}_{k,i,T+\tau} = \begin{cases} 0 & \widehat{y}^{*}_{k,i,T+\tau} \le 0 \\ \widehat{y}^{*}_{k,i,T+\tau} & \text{if } 0 < \widehat{y}^{*}_{k,i,T+\tau} < 1, \\ 1 & \text{if } \widehat{y}^{*}_{k,i,T+\tau} \ge 1 \end{cases} .$$

Finally, it should be noted that $\widehat{y}^{*}_{k,i,T+\tau}$ modifies the usual predictor simply by adding a fraction of the corresponding residuals to the i-th unit of land. A similar result was obtained in Baltagi and Li (2006), Baltagi et al. (2012), and Ay et al. (2017). Here, the addition is equivalent to that of a random-effects model without the spatial autocorrelation, which deviates from the results formulated in Baltagi and Li (2004, 2006). This is because our SAR random effects model differs from that of Anselin et al. (1988) in that the disturbance term itself follows a SAR process whereas the remainder term follows an error component structure. This point will be useful when performing the hypothesis test for comparing our model's predictive accuracy in Section 4.4.

Table 1: Land-use Determinants†

Abbreviations	Definitions
Group 1:	
$alt0$	$d_{eb200} \times elev$, where $d_{eb200} = 1$ if $elev < 200$ and 0 otherwise
$alt200$	$d_{ea200} \times elev$, where $d_{ea200} = 1$ if $elev > 200$ and 0 otherwise
$alt200d$	$alt200d = 1$ if $elev > 200$ and 0 otherwise
$slope6$	Share of each grid square with a slope higher than 6^o
$rain$	Accumulated rainfall for the growing season
$temp$	Average temperature for the growing season
$ratemp$	$rain \times temp$ (i.e., an interaction term)
$dist300$	Distance to the closet major market
$speat$	Proportion of soil characteristic "Peat"
$sgravel$	Proportion of soil characteristic "Gravel"
$sstoney$	Proportion of soil characteristic "Stone"
$sfragipan$	Proportion of soil characteristic "Fragipan Soil"
$scoarse$	Proportion of soil texture "Coarse"
$sfine$	Proportion of soil texture "Fine"
$smedium$	Proportion of soil texture "Medium"
sud	$sud = 1$, if the grid square is located in the Southern England
nor	$nor = 1$, if the grid square is located in the Northern England
mid	$mid = 1$, if the grid square is located in the Midlands
y_ℓ	Yearly dummies, where $\ell = 1976, 1979, 1981, 1988, 2000, 2004$
$npark$	Share of each grid square designated as a National Park
esa	Share of each grid square designated as an Environmentally Sensitive Area
$greenbelt$	Share of each grid square designated as a Greenbelt
$setaside$	Share of each grid square designated as a Set-aside
Group 2:	
$rain_\ell$	$rain_\ell = (rain - \ell)d_{r\ell}$ for $\ell = 300, 350, 400, 450, 500, 600$
$temp_\ell$	$temp_\ell = (temp - \ell)d_{t\ell}$ for $\ell = 9, 10, 11, 12, 13, 14$

† Since *sud*, *nor* and *mid* are summed to one, *mid* is omitted in the estimation because of multicollinearity. *smedium* is also omitted for a similar reason.

4. Empirical Analysis of Selected Land-Use Shares in England

The objective of this section is to illustrate the applicability of our framework by using it to investigate farmers' responses to the behavioural drivers of land-use decision in England. We are particularly interested in investigating whether environmental schemes and grants have assisted in freeing up land used for arable, rough grazing, temporary and permanent grasslands and converting it to bio-energy crops to help to achieve deep emission reductions and prepare for climate change. We are also interested in finding out whether our predictor, which includes the above-derived fraction of the residuals to the i-th unit of land, can improve accuracy for the prediction of future land-use share.

4.1. Data descriptions and sources

To achieve these goals, our analysis focuses on land-use shares of (i) arable, i.e. cereals (wheat, barley, and oats) and root crops (excluding oilseed rape), (ii) temporary grassland (grassland typically part of an arable crop rotation), (iii) permanent grassland (grassland maintained perpetually without reseeding), (iv) rough grazing (uncultivated land used for grazing livestock), and (v) oilseed rape. While the first four categories are the main land-use types for the English agricultural sector, the fifth is a representation of bio-energy crops, which should be financially incentivised in order to help to reduce emissions and prepare for climate change (CCC (2018)). Moreover, we consider a set of land-use determinants and drivers, which can be classified into three categories, namely (i) economics, (ii) climatic and physical environment, and (iii) environmental policy.

Regarding the data used, they are from a unique database that consists of data compiled from various sources at the Land, Environment, Economics and Policy (LEEP) Institute. Data on agricultural land-use are derived from the June Agricultural Census on a 2-km^2 (400 ha) grid available online from Edinburgh University Data Library. These data cover England and Wales for 17 irregular spaced years between 1969 and 2006 and yield roughly 38,000 grid-square records each year. Table 1 presents a full list of exogenous variables considered in our model. Due to space limitation, more details of the data are presented in the online appendix.

Remark 4.1. *A lack of information on the spatial variation of market input and output prices hinders an explicit modelling of their effects on land-use shares. Hence, in our empirical analysis these will be accounted for by a set of yearly and regional dummy variables (see also Sterling et al. (2013) and Fezzi et al. (2015) who used a similar approach).*

Remark 4.2. *We suggested in Section 2.1.1 that aggregation of the individual data is done only on a scale that is small enough to capture the spatial variation in the environmental and climatic drivers of farmers' behavior, and proportionate with the scale of the decision-making unit. In order to satisfy the second characteristic, we shall assume that the spatial unit considered, i.e. 2-km^2 grid, represents a decision-making unit. The average farm size in the North East of England, for example, was about 1.5-km^2 (150 hectares) in 2018 (Department for Environment, Food and Rural Affairs (2020)).*

The formulation in Section 2.2 suggests that TL-RE-SED-Tobit Model only accepts balanced panel data. To satisfy such a condition, a subset of the data in the space dimension is selected by randomly extracting one grid square and then sampling every fourth grid cell along both the latitude and longitude axes. In the time dimension, since the original data cover unevenly spaced years, only observations from 1976, 1979, 1981, 1988, 2000 and 2004 are selected to stay as close to a regular time series as possible. In the other words, $T = 6$ years. For England, this leads to $NT = 10,034$ or $N = 1,729$ observations. This spatial sampling method has been used extensively in the literature (see e.g. Nelson and Hellerstein (1997), Carrión-Flores and Irwin (2004), and Fezzi and Bateman (2011)) and should help to improve estimation performance since undesirable noises are also removed.

Finally, to compute the land-use share, we first calculate total amount of land within a 2-km^2 grid used for cultivating arable, temporary grassland, permanent grassland, rough grazing, oilseed rape, then compute land-use share of a given crop as a percentage of such a total. Table 2 presents descriptive statistics for the areas of land used in hectares. The table also indicates cases in which p-values for Welch's unequal variances t-test for mean-comparison are less than 0.01, 0.05 and 0.1, respectively. These results suggest that only the area used for temporary grassland has statistically significantly declined between 1976 and 2004. The level of land used for permanent grassland (arable) remained unchanged between 1976 and 1988, then fluctuated slightly (decreased steadily) between 1988 and 2004.

4.2. Empirical Specifications

This section discusses a number of empirical specifications, which are important to the analysis that follows.

Conditional mean: Regarding the empirical specifications of the conditional mean, the most basic specification is to impose linear effects on all the determinants of the land-use shares. In the other words, how the expected value of the unobserved and censored land-use share $y^*_{k,it}$ varies with the

Table 2: Descriptive Statistics for Land-Uses (in ha)

	1976	1979	1981	1988	2000	2004
Temp. grassland	35.929	30.133 ^{c}	29.645	25.363 ^{c}	23.428 ^{b}	20.753 ^{c}
Perm. grassland	96.705	96.623	94.938	91.338	81.641 ^{c}	88.755 ^{c}
Rough grazing	25.772	25.274	24.995	24.341	23.622	24.948
Arable	113.255	117.333	121.042	119.114	107.460 ^{c}	99.035 ^{c}
Total arg. land†	272.910	271.415	274.330	269.853^{a}	245.715^{c}	248.451

† Total agricultural land is computed as the summation of temporary, permanent, rough grassland and oilseed rape. ^{c}, ^{b} and ^{a} signify cases where p-values for Welch's unequal variances t-test (e.g. $H_0 : \mu_{k,1979} - \mu_{k,1976} = 0$ or $H_0 : \mu_{k,1988} - \mu_{k,1981} = 0$) are less than 0.01, 0.05 and 0.1, respectively.

environmental, climatic and policy variables is described by

$$E\left[y^*_{k,it}|x_{k,it}\right] = x_{k,it}\beta_k, \tag{4.1}$$

where $x_{k,it}$ is an 1×29 row-vector whose elements are a constant one and the variables listed under Group 1 in Table 1. Since the specification in (4.1) can be overly restrictive, we also consider an alternative which (i) allows for some nonlinear flexibility within the parametric specification, and (ii) does so without imposing too much computational pressure. This is to capture the potential nonlinear effects of climatic factors by modelling the measures of rainfall and temperature as piecewise linear functions. In particular, how the expected value of the unobserved and censored land-use share $y^*_{k,it}$ varies with the environmental, climatic and policy variables is described by

$$E\left[y^*_{k,it}|x_{k,it}\right] = x_{k,it}\beta_k + \vartheta_k(rain_{it}) + \zeta_k(temp_{it}),$$

where $\vartheta_k(rain_{it}) = \beta_{k,r300}rain_{300,it} + \cdots + \beta_{k,r600}rain_{600,it}$ and $\zeta_k(temp_{it}) = \beta_{k,t9}temp_{9,it} + \cdots + \beta_{k,t14}temp_{14,it}$.

Spatial Weighting Matrices: Various studies have reported that predictive accuracy and empirical results in general are sensitive to the choice of spatial weighting matrix W_k (e.g. Anselin and Bera (1998) and Bhattacharjee, and Jensen-Butler (2006)). To investigate such sensitivity, we consider weighting matrices based on two types of schemes, namely "inter-point-distance" and "graph-based-neighbours". Particularly, we construct the κ-Nearest-Neighbours weighting matrices, $W_k^{\kappa NN}$, where either $\kappa = 2$ or $\kappa = 5$, and the Sphere-of-Influence-Neighbours weighting matrix, W_k^{SOI}. All these spatial weighting matrices are row-normalized.

Reference Land Use Category: Note that the adding-up restriction on the land-use shares only holds for the latent shares in (2.1), but it is unsatisfiable for the observed shares. In the demand study literature, such a problem is avoided by treating one of the categories as a reference and omitting it from the system (i.e. Yen et al. (2003), Chakir and Le Gallo (2013), and Rarcos-Martinez et al. (2017)). In the study that follows, we drop the category "oilseed rape" and jointly estimate a system of four TL-RE-SED-Tobit models for arable, temporary grassland, permanent grassland and rough grazing for England.

4.3. Estimation results and important findings

We have prepared detailed results for each of the steps discussed in Section 3. They are presented in the online appendix due to space limitation. In the current section, we first provide a brief description of these results, then thoroughly explain important findings from each of the estimation steps.

In the online appendix, Tables 3 to 10 present estimation results for the four land-use shares under consideration. In these tables, we present these results for four modelling strategies: (i) Without RE-SED, (ii) RE-SED under W_k^{2NN}, (iii) RE-SED under W_k^{5NN}, and (iv) RE-SED under W_k^{SOI}. In addition, for each strategy, we present the results in three columns, namely parameter estimates ($\widehat{\beta}_k$), associated standard errors (SEs), and p-values (p-vals). At the bottom of the tables, we also present numbers of coefficient estimates that are statistically significant at 0.01, 0.05 and 0.1 significance levels. Furthermore, Table 11 presents results of the GMM estimation for the autoregressive parameters ρ_k, where $k = 1, \ldots, \mathcal{K} = 4$. For each of the four land-use shares under consideration, we compute six estimates, i.e. three based on the linear specification under W_k^{5NN}, W_k^{2NN} and W_k^{SOI}, and the remainders based on the partial linear specification. In addition, Tables 12 to 14 presents the resulting estimates for Ω_v and Ω_μ denoted by $\widehat{\Omega}_v$ and $\widehat{\Omega}_\mu = (1/T)(\widehat{\Omega}_1 - \widehat{\Omega}_v)$. Based on these, we also compute correlations matrices $\widehat{\rho}_v$ and $\widehat{\rho}_\mu$, which gauge the cross-equation correlations in the error terms v_k and the random effects μ_k, respectively. These estimates are for the RE-SED models under W_k^{2NN}, W_k^{5NN} and W_k^{SOI} matrices, respectively, and both the linear and partial linear specifications.

4.3.1. Pooled QML estimation of the TL-Tobit panel data models

The results, which concern the pooled QML estimation of the TL-Tobit panel data models, are presented under W/O RE-SED (Without RE-SED) in Tables 3 to 10. We shall discuss these numbers more thoroughly below.

4.3.2. GMM estimation of the spatial parameters

In Table 11, it is clear that at $\kappa = 5$, the corresponding estimates of $\widehat{\rho}_k$ for $W_k^{\kappa NN}$ are close to those of W_k^{SOI} for all cases. The most likely reason underpinning such a phenomenon is the similarity in the degree of spareness of these weighting matrices. Furthermore, a higher degree of spareness in the weighting matrix is usually associated with higher estimates of ρ_k. The estimates for shares of arable and permanent grassland are statistically significant at 0.05 significance level for all cases. For the share of temporary grassland, the estimates are significant at 0.1 significance level for W_k^{2NN} and at 0.05 for both W_k^{5NN} and W_k^{SOI}. For the share of rough grazing, the estimates are significant at 0.1 level, except that of W_k^{2NN}. Moreover, these estimates enable computation of $\widehat{A}_{kk} = I_T \otimes \widehat{H}_k$, $\widehat{H}_k = (I_N - \widehat{\rho}_k W_k)^{-1}$, and the Cochrane-Orcutt transformations $\dot{\mathcal{X}} = \widehat{A}^{-1}\widetilde{X}$ and $\dot{\mathcal{Y}} = \widehat{A}^{-1}\widetilde{Y}$ as explained in Step 3.2.

4.3.3. Iterative QML method

This step performs the iterative QML method discussed in Remark 3.1. Our main focus is on estimating $S_v = [\sigma^2_{11,v}, \dots, \sigma^2_{\mathcal{K}\mathcal{K},v}, \sigma^2_{12,v}, \dots, \sigma^2_{\mathcal{K}-1,\mathcal{K},v}]^\top$ and $S_1 = [\sigma^2_{11,1}, \dots, \sigma^2_{\mathcal{K}\mathcal{K},1}, \sigma^2_{12,1}, \dots, \sigma^2_{\mathcal{K}-1,\mathcal{K},1}]^\top$, where $\mathcal{K} = 4$. We use these estimates to construct the Cochrane-Orcutt plus RE-GLS transformations defined in (3.11), then perform the final Tobit QML estimation of the causal parameters (as discussed in Remark 3.2) in order to obtain the associated standard errors. Estimation results for the four land-use shares in question are also presented in Tables 3 to 10 under the headings RE-SED under W_k^{2NN}, RE-SED under W_k^{5NN} and RE-SED under W_k^{SOI}. Below, let us summarise a number of key findings.

Form these tables, it is clear that taking into consideration cross-equation correlations and RE-SED reduces the number of coefficient estimates that are considered statistically significant in all cases. Let us take the share of arable as an example. The number of estimates that are significant at 0.01, 0.05 and 0.1 significance levels are 17, 20 and 21 (15, 16 and 17) under the linear (partial linear) specification and without RE-SED. These reduces to 13, 15 and 16 (11, 13 and 13) when the RE-SED is modelled under W_k^{2NN}. The increase in the degree of spareness in the weighting matrices leads to further reduction of the figures to 10,11 and 12 (8, 10 and 10) for modelling under W_k^{5NN}. Similar results are also obtained for modelling the RE-SED under W_k^{SOI}. The most likely reason underpinning such a phenomenon is the similarity in the degree of spareness of these weighting matrices.

Inevitably, these lead to differences in the conclusions drawn about the causal effects of climatic and physical environment, and environmental policy on the land-use shares. We now discuss the implications of the above findings on the individual land-use shares.

Share of Arable: The coefficient estimates associated with different measures of altitude are statistically significant under the model without RE-SED, but are insignificant under RE-SED irrespective of the weighting matrices used. The coefficient estimates associated with *slope* (negative), *rain* (positive) and *temp* (positive) are statistically significant across the modelling strategies considered. However, only that associated with $sfragipan$ (negative) remains statistically significant after taking into consideration RE-SED. Being in the north of England has a negative effect on the share of arable compared to Midlands. Year dummies are all statistically significant irrespective of the modelling strategies. Finally, the coefficient estimates associated with environmental policies becomes insignificant when RE-SED is modelled irrespective of the weighting matrices used.

Share of Permanent Grassland: Unlike those of arable, the coefficients estimates associated with measures of altitude are statistically significant under all the models considered for share of permanent grassland. Also unlike those for arable, the coefficient estimates associated with *slope*, *rain*, and *temp*, which are statistically significant under the model without RE-SED, become insignificant when RE-SED is taken into consideration. Although, soil characteristics remains matter for the share of permanent grassland, locations of the land (i.e. *nor* or *sud*) become insignificant after taking into consideration RE-SED. A similar conclusion can also be drawn for all the year dummies (except $y4$ and $y5$ which remains statistically significant). Finally, the coefficient estimates associated with *esa* becomes statistically significant after including RE-SED.

Share of Temporary Grassland: We now shift our attention to the share of temporary grassland. The coefficient estimates associated with measures of altitude and *slope* are not statistically significant in all models. On the contrary, those associated with *rain* and *temp* are significant irrespective of the modelling strategies used. Soil characteristics seem to matter when modelling without RE-SED, but the coefficient estimates become insignificant under weighting matrices with higher degree of spareness. A similar conclusion can also be drawn for *nor* and *sud*, and $y1$ and $y2$. Finally, all the coefficient estimates associated with the environmental policies (except that of *setaside* (positive)) are not statistically significant.

Share of Rough Grazing: Regarding share of rough grazing, the coefficient estimates associated with measures of altitude are statistically insignificant, while that associated with *slope* are statistically significant irrespective of the modelling strategies considered. A similar conclusion can also be drawn for *rain* and *temp*. Nonetheless, the coefficient estimates associated with soil characteristics become insignificant after taking into consideration RE-SED. The location of the land, i.e. *nor* and *sou*, contribute positively to the share of rough grazing compared to the Midlands. Furthermore, all year dummies are statistically significant across all the model used. Finally, all coefficient estimates associated with *npark*, *esa*, and *greenbelt* are significant when modelled without RE-SED, but only that of *greenbelt* remains significant after taking RE-SED into consideration.

We complete this section by discussing the resulting estimates for Ω_v and Ω_μ denoted by $\widehat{\Omega}_v$ and $\widehat{\Omega}_\mu = (1/T)(\widehat{\Omega}_1 - \widehat{\Omega}_v)$. Firstly, we find that the cross-equation correlations μ_k are much stronger than those of v_k. In the other words, the cross-equation correlations in the TL-Tobit system of the land-use shares are dominated by those of the random effects. Secondly, switching between the weighting matrices does not alter the signs of the estimated cross-equation correlations. However, their magnitudes change more significantly when switching from W_k^{2NN} to W_k^{5NN} than from W_k^{5NN} to W_k^{SOI}. These changes are largely dominated by those in the cross-equation correlations of μ_k. Finally, we find that the above findings hold for both the linear and partial linear specifications.

4.4. Improvement in Prediction Accuracy

In the current section, we investigate whether inclusion the above-derived fraction of the residuals to the i-th unit of land is able to improve accuracy for the prediction of future land-use share. To this end, note that predictive evaluation must be performed by treating

$$\widehat{y}_{k,i,T+\tau} = \begin{cases} 0 & \widehat{y}^*_{k,i,T+\tau} \leq 0 \\ \widehat{y}^*_{k,i,T+\tau} & \text{if } 0 < \widehat{y}^*_{k,i,T+\tau} < 1, \\ 1 & \text{if } \widehat{y}^*_{k,i,T+\tau} \geq 1 \end{cases}$$

as the test dataset since $\widehat{y}_{k,i,T+\tau}$ is the best linear unbiased predictor of the latent variable $y^*_{k,i,T+\tau}$. Moreover, our examination focuses on comparing root mean squared errors (RMSEs) from a number of alternative predictors. These are computed based on: (A.1) Linear two-limit Tobit model without the random effects and spatial error dependence $\tilde{y}^*_{k,i,T+\tau} = x_{k,i,T+\tau}\tilde{\beta}_k$. (A.2) Partially linear two-limit Tobit model without the random effects and spatial error dependence $\widetilde{y}^*_{k,i,T+\tau} = x_{k,i,T+\tau}\widetilde{\beta}_k + \widetilde{\vartheta}_k(rain_{it}) + \widetilde{\zeta}_k(temp_{it})$. (B.1) Linear two-limit Tobit model with the random effects and spatial error dependence, but without the fraction of the residuals corresponding to the i-th unit of land $\widehat{y}^*_{k,i,T+\tau} = x_{k,i,T+\tau}\widehat{\beta}_k$. (B.2) Partially linear two-limit Tobit model with the random effects and spatial error dependence, but without the fraction of the residuals corresponding to the i-th unit of land

$$\widehat{y}^{\,*}_{k,i,T+\tau} = x_{k,i,T+\tau}\widehat{\beta}_k + \widehat{\vartheta}_k(rain_{it}) + \widehat{\zeta}_k(temp_{it}).$$

(C.1) Linear two-limit Tobit model with the random effects and spatial error dependence, and with the fraction of the residuals corresponding to the i-th unit of land $\widehat{y}^{\,*}_{k,i,T+\tau} = x_{k,i,T+\tau}\widehat{\beta}_k + \frac{\widehat{\sigma}^2_{\mu,kl}}{\widehat{\sigma}^2_{1,kl}}(e'_T \otimes l'_{k,i})\widehat{u}_k$. (C.2) Partially linear two-limit Tobit model with the random effects and spatial error dependence, and with the fraction of the residuals corresponding to the i-th unit of land $\widehat{y}^{\,*}_{k,i,T+\tau} = x_{k,i,T+\tau}\widehat{\beta}_k + \widehat{\vartheta}_k(rain_{it}) + \widehat{\zeta}_k(temp_{it}) + \frac{\widehat{\sigma}^2_{\mu,kl}}{\widehat{\sigma}^2_{1,kl}}(e'_T \otimes l'_{k,i})\widehat{u}_k$.

In this regard, we are interested in two set of comparisons: (a) It is the comparison between the predictors listed under categories A and B. These are important because they can help to confirm the asymptotic compatibility between $\widetilde{\beta}_k$ and $\widehat{\beta}_k$. (b) Comparison between the predictors listed in categories B and C. These are significant since they can affirm that improvement in predictive accuracy can be achieved by including the above-derived fraction of the corresponding residuals to the i-th unit of land without adjusting the causal specification. In this regard, it should also be noted that such an inclusion is not possible without incorporating the random effects and spatial error dependence into the model.

Furthermore, we reinforce these results by conducting hypothesis testing for the equivalence of predictors listed under categories B and C. To this end, it is useful to recall the difference between the predictors in these categories, namely the added fraction of the corresponding residuals to the i-th land. Unlike other error component models (e.g. those formulated in Baltagi and Li (2004, 2006)), the addition here is equivalent to that of a random-effect model without the spatial autocorrelation. This suggests that the absence of random effect should lead to simplification of the predictors in category C to those in B, and therefore that a testing procedure such as that of Breusch and Pagan (1980), which tests for the random-effects model, could be used for checking the equivalence of these predictors. Breusch and Pagan (1980) devised a Lagrange multiplier test for the random-effects model, in which the test statistic is

$$LM_{BP} = \frac{NT}{2(T-1)}\left[\left(\frac{\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\varepsilon_{it}\right]^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2}\right) - 1\right]^2. \qquad (4.2)$$

The limiting distribution of LM_{BP} is chi-squared with one degree of freedom under the null hypothesis $H_0 : \sigma^2_{k,\mu} = 0$. The practical implementation of the test relies on computation of $\widehat{\varepsilon}_{k,it}$ using $\widehat{\rho}_k$ and $\widehat{u}_k$, and based either on (2.8) or $\varepsilon_k = u_k - \rho_k(I_T \otimes W_k)u_k$.

In the empirical analysis, the observed land-use shares in 2010 are treated as the validation dataset. In this regard, let $\widehat{y}^*_{k,i,2010}$ denote the best linear unbiased predictor of crop k's land-use share on the i-th plot of land in 2010. Hence, $\widehat{y}^*_{k,i,2010}$ quantifies the expected level of crop k's land-use share on the i-th plot of land in 2010 for a given scenario of climatic, economic and policy drivers based on our estimated models and specifications. Clearly, if changes in environmental policies (e.g. an increase in farming in London's greenbelt) or climate (e.g. a higher level of rainfall and/or temperature in some regions) are expected in 2010, then $\widehat{y}^*_{k,i,2010}$ is adjusted accordingly. This suggests an important benefit of land-use prediction in practice that is the ability to accurately forecast the effects of policy and/or climate changes on agricultural production and land-use in the UK.

In the online appendix, Tables 15 to 18 present the RMSEs for the out-of-sample predictions of the land-use shares in 2010. Some important findings are as follows: (i) It is clear that the RMSEs reported in rows [a] within each of the tables do not vary significantly from one another. These suggest that $\widetilde{\beta}_k$ and $\widehat{\beta}_k$ are closely similar. Such findings are as anticipated and theoretically deducible from the estimation consistency. (ii) The RMSEs reported in rows [b] in each table (even under different weighting matrices) are always smaller than those in rows [a]. These differences are particularly significant for permanent grassland and arable. Such findings stress the need to incorporate the random effects and spatial error dependence into the model in order to improve the predictive accuracy. (iii) It seems that at $\kappa = 5$, the RMSEs reported for $W_k^{\kappa NN}$ are relatively close to those of W_k^{SOI}. Nonetheless, the evidence is not conclusive on which specifications of the weighting matrix is able to bring about better forecasts. Moreover, in Tables 15 to 17, rows [c] present the corresponding LM_{BP} test statistics and p-values under the different weighting matrices. In all cases, the LM_{BP} test statistics far exceed 3.84, which is the 95% critical value for the chi-squared distribution with one degree of freedom. These lead to rejection of the null hypothesis and a suggestion that superiority in the predictive accuracy reported in the previous paragraph was not caused by measurement error. The predictors under category C are statistically different from those listed under category B and are able to provide more accurate prediction.

5. Conclusions

We contributed toward building a better understanding of farmers' responses to behavioural drivers of land-use decision by establishing a new analytical procedure that can handle complex data structures and overcome various drawbacks suffered by existing methods. Firstly, our procedure made use of spatially high-resolution data so that idiosyncratic effects of physical environment drivers could be explicitly modelled. Secondly, we addressed the famous censored data problem to ensure theoretical consistency of the parameter estimates. Thirdly, we incorporated spatial error dependence and heterogeneity in order to gain efficiency, more accurate formulation of the variances for the parameter estimates and hence more effective statistical inferences. Finally, we reduced the computational burden and improved estimation accuracy by introducing a GMM/QML hybrid estimation procedure. We applied our method to spatially high resolution data in England and found that the number of coefficient estimates that are statistically significant reduces significantly when the SED and heterogeneity are taken into consideration. Inevitably, this leads to conclusions about causal effects of climatic and physical environment, and environmental policy on land-use shares that differed significantly from those made based on methods that are currently used in the literature. Moreover, we showed that our method enables derivation of a more effective predictor of the land-use shares, which is utterly useful from the policy making point of view.

6. References

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. Econometrica: Journal of the Econometric Society, 997-1016.

Anselin, L. (1988). Spatial econometrics: Methods and models. Kluwer Academic: Boston, MA.

Anselin, L. & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics: Regression models. In Handbook of applied economic statistics (pp. 257-259). CRC Press: Boca Raton, FL.

Ay, J. S., Chakir, R., & Le Gallo, J. (2017). Aggregated versus individual land-use models: Modeling spatial autocorrelation to increase predictive accuracy. Environmental Modeling & Assessment, 22(2), 129-145.

Baltagi, B. H. (1980). On seemingly unrelated regressions with error components. Econometrica: Journal of the Econometric Society, 1547-1551.

Baltagi, B. H. (2008). Econometric analysis of panel data. John Wiley & Sons.

Baltagi, B. H., Bresson, G., & Pirotte, A. (2012). Forecasting with spatial panel data. Computational Statistics & Data Analysis. 56(11):3381-97.

Baltagi, B. H., & Li, D. (2004). Prediction in the panel data model with spatial correlation. In Advances in spatial econometrics (pp. 283-295). Springer: Berlin.

Baltagi, B. H., & Li, D. (2006). Prediction in the panel data model with spatial correlation: the case of liquor. Spatial Economic Analysis, 1(2), 175-185.

Baltagi, B. H., & Pirotte, A. (2011). Seemingly unrelated regressions with spatial error components. Empirical Economics, 40(1), 5-49.

Bhattacharjee, A., & Jensen-Butler, C. (2006). Estimation of spatial weights matrix, with an application to diffusion in housing demand. Centre for Research into Industry, Enterprise and the Firm Discussion Paper, 519.
Breusch, T. S., & Pagan, A. R. (1980). Lagrange multiplier test and its applications to model specification. The Review of Economic Studies, 47(1), 239-253.
Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: methods and applications. Cambridge University Press.
Carrion-Flores, C., & Irwin, E. G. (2004). Determinants of residential land-use conversion and sprawl at the rural-urban fringe. American Journal of Agricultural Economics, 86(4), 889-904.
Chakir, R., & Gallo, J.L. (2013). Predicting land use allocation in France: A spatial panel data analysis. Ecological Economics, 92(0), 114-125.
Chambers, R. G., & Just, R. E. (1989). Estimating multioutput technologies. American Journal of Agricultural Economics, 71(4), 980-995.
Cliff, A. D., & Ord, J. K. (1973). Spatial autocorrelation (No. 04; QA278. 2, C5.).
Cliff, A. D., & Ord, J. K. (1981). Spatial processes: models & applications. Taylor & Francis.
Committee on Climate Change (2018). Land use: Reducing emissions and preparing for climate change. Retrieved from www.theccc.org.uk/publication/land-use-reducing-emissions-and-preparing-for-climate-change.
Committee on Climate Change (2020). Land use: Policies for a Net Zero UK. Retrieved from www.theccc.org.uk/publication/land-use-policies-for-a-net-zero-uk/.
Deng, Q., & Xue, J. (2014). Multivariate Tobit system estimation of education expenditure in urban China. The Singapore Economic Review, 59(01), p.1450005.
Department for Environment, Food and Rural Affairs (2020). Defra Statistics: Agricultural Facts England Regional Profiles February 2020. Retrieved from www.gov.uk/government/statistics/agricultural-facts-england-regional-profiles.
Dong, D., Gould, B. W., & Kaiser, H. M. (2004). Food demand in Mexico: an application of the Amemiya-Tobin approach to the estimation of a censored food system. American Journal of Agricultural Economics, 86(4), 1094-1107.
Fezzi, C., & Bateman, I. J. (2011). Structural agricultural land use modelling for spatial agro-environmental policy analysis. American Journal of Agricultural Economics, 93(4), 1168-1188.
Fezzi, C., Harwood, A., Lovett, A., & Bateman, I. J. (2015). The environmental impact of climate change adaptation on land use and water quality. Nature Climate Change, 5, 255–260.
Greene, W. H. (2008). The econometric approach to efficiency analysis. The measurement of productive efficiency and productivity growth, 1(1), 92-250.
Kapoor, M., Kelejian, H. H., & Prucha, I. R. (2007). Panel data models with spatially correlated error components. Journal of Econometrics, 140(1), 97-130.
Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review, 40(2), 509-533.
Lacroix, A., & Thomas, A. (2011). Estimating the environmental impact of land and production decisions with multivariate selection rules. American Journal of Agricultural Economics, 93(3), 784-802.
Lee, L. F. (1993). Multivariate Tobit models in econometrics. In G.S. Maddala, C.R. Rao, and H.D.Vinod, eds. Handbook of Statistics, vol. 11., chap. 6. Amsterdam: North-Holland, 1993, pp. 145–73.

Lee, L. F. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. Journal of econometrics, 154(2), pp.165-185.
Li, M., JunJie, W., & Deng, X. (2013). Identifying drivers of land use change in China: A spatial multinomial logit model analysis. Land Economics, 89(4), 632-654.
Liu, S. F., & Yang, Z. (2015). Asymptotic distribution and finite sample bias correction of QML estimators for spatial error dependence model. Econometrics, 3(2), 376-411.
Marcos-Martinez, R., Bryan, B. A., Connor, J. D., & King, D. (2017). Agricultural land-use dynamics: Assessing the relative importance of socioeconomic and biophysical drivers for more targeted policy. Land Use Policy, 63, 53-66.
Mattison, E. H., & Norris, K. (2005). Bridging the gaps between agricultural policy, land-use and biodiversity. Trends in Ecology & Evolution. 20(11):610-6.
Meyerhoefer, C. D., Ranney, C. K. & Sahn, D. E. (2005). Consistent estimation of censored demand systems using panel data. American Journal of Agricultural Economics. 87(3), pp.660-672.
Ministry of Housing, Communities & Local Government (2017). Land use in England 2017. Retrieved from
www.gov.uk/government/collections/land-use-in-england-experimental-statistics
Moscone, F., Knapp, M., & Tosetti E. (2007). Mental health expenditure in England: a spatial panel approach. Journal of Health Economics. 26(4):842-64.
Nelson, G. C., & Hellerstein, D. (1997). Do roads cause deforestation? Using satellite images in econometric analysis of land use. American Journal of Agricultural Economics, 79(1), 80-88.
Reidsma, P., Tekelenburg, T., Van den Berg M., & Alkemade R. (2006) Impacts of land-use change on biodiversity: An assessment of agricultural biodiversity in the European Union. Agriculture, ecosystems & environment. 114(1):86-102.
Sterling, S. M., Ducharne, A., & Polcher, J. (2013). The impact of global land-cover change on the terrestrial water cycle. Nature Climate Change, 3(4), 385-390.
Tsay, R. S. (2005). Analysis of financial time series (Vol. 543). John Wiley & Sons.
Wang, X., & Kockelman, K. M. (2007). Specification and estimation of a spatially and temporally autocorrelated seemingly unrelated regression model: application to crash rates in China. Transportation, 34(3), pp.281-300.
Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. Journal of Econometrics, 68(1), 115-132.
Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.
Yang, Z. (2013). Quasi-Maximum Likelihood Estimation for Spatial Panel Data Regressions. Research Collection School Of Economics. Available at:
https://ink.library.smu.edu.sg.
Yen, S. T., Lin, B. H., & 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(2), 458-478.