Modeling and predicting agricultural land use in England based on spatially high-resolution data

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

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

Abstract

This paper addresses various statistical and empirical challenges associated with modelling farmers' decision-making processes concerning agricultural land-use. These include (i) use of spatially high-resolution data so that idiosyncratic effects of physical environment drivers, e.g. soil textures, can be explicitly modelled; (ii) modelling land-use shares as censored responses, which enables consistent estimation of the unknown parameters; (iii) incorporating spatial error dependence and heterogeneity, which leads to accurate formulation of the variances for the parameter estimates and more effective statistical inferences; and (iv) reducing the computational burden and improving estimation accuracy by introducing an alternative GMM/QML hybrid estimation procedure. We also provide extensive evidence, which suggests that our approach can construct more accurate land-use predictions than existing methods in the literature. We then apply our method to empirically investigate how the climatic, economic, policy and physical determinants influence the land-use patterns in England over time and spatial space. We are also interested in examining 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.

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: `wongsa-artp@cardiff.ac.uk` (Patrick Wongsa-art)

1. Introduction

By now, it is a common knowledge that changes in land-use patterns significantly affect both the environment (e.g. biodiversity, water pollution and soil erosion) and economic/social welfare (see e.g. Mattison and Norris (2005), Reidsma et al. (2006), and Chakir and Le Gallo (2013)). In the UK, there is also a strong belief that freeing up some types of agricultural land-use and converting it to alternative usages can help to achieve deep emission reductions and prepare for climate change (Committee on Climate Change (2018, 2020)). In 2018, the UK's Committee on Climate Change (CCC) proposed a number of changes to the way we use land. These changes aim at (a) reducing land-use for grasslands, i.e. permanent and temporary grasslands, and rough grazing by 26 to 36%, (b) introducing new woodlands to store carbon by 1.5 million hectares, and (c) increasing 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 35 to 80% overall reductions in Metric tons of carbon dioxide equivalent by 2050.

These factors give rise to the necessity to manage how land is used and the need for a better understanding of farmers' behavioral responses to drivers of land-use changes. In particular, we are interested in (i) climatic drivers (e.g. rain and temperature), (ii) economic drivers (e.g. input and output prices), and (iii) environmental policies and governmental schemes (e.g. greenbelts, set-asides, environmentally sensitive areas (ESAs), and other subsidies and grants). A better understanding of how these drivers influence the land-use patterns over time and spatial space should help policy makers not only to evaluate existing practices, but also to formulate new environmental policies.

In this paper, we embrace this necessity and aim to address various well-known methodological challenges associated with modelling farmers' decision-making processes about land-use allocation. Moving beyond that of existing studies (e.g. Fezzi and Bateman (2011), Chkir and Le Gallo (2013), Ay et al. (2017), and Marcos-Martinez et al. (2017)), our new framework (i) explores use of spatially disaggregated data, (ii) models land-use shares as censored responses, (iii) allows for potential spatial error dependence (SED), (iv) models unobserved heterogeneity in an error component structure, and (v) reduces computational burden via a hybrid estimation procedure. We shall discuss the statistical and empirical underpinnings of these explorations in detail in Section 2.1. Taking into account these features gives rise to estimation of 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 establishment 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.

Section 4 provides an insight into the predictive power under the newly proposed TL-RE-SED-Tobit specification. Subsecquently, in Section 5 we present an empirical investigation of how the climatic, economic and policy drivers influence the agricultural land-use patterns in England over time and spatial space. We also give special attention on the question whether existing environmental schemes and grants in England have helped to incentivise a switch in land-use. In this section, we show firstly that the newly introduced TL-RE-SED-Tobit specification leads to a significant improvement in the predictive accuracy of land-use. Since prediction of future land-use areas is essential, such a superior predictive power lays an important foundation for the use of our model. Finally, Section 6 draw some important conclusions, while mathematical proof is presented in the Appendix.

2. Tobit system with spatial error dependence and random effects

This section formulates the system of simultaneous Tobit equations of land-use shares. We begin with the key analytical considerations that lead to establishment of such a system, then discuss its formulation in detail.

2.1. Methodological explorations

We have identified a number of methodological shortfalls in the existing studies of agricultural land-use modelling and prediction (see, for example, Fezzi and Bateman (2011), Chkir and Le Gallo (2013), Ay et al. (2017), and Marcos-Martinez et al. (2017)). Below we discuss these shortfalls and suggest strategies to remedy.

2.1.1. Exploring the use of spatially disaggregated data

In an empirical analysis of land-use, different techniques are required for different data resolutions. At the two extreme ends of the spectrum, we have individual data, which correspond to the parcel level, and aggregated data on a larger geographical region (e.g. regional and national levels). Regarding the former, analysis is often conducted within a discrete choice modelling framework (see e.g. Li et al. (2013)). The analysis of the latter involves tools in panel-data regression models and seemingly unrelated regressions (see e.g. Baltagi and Pirotte (2011), Chakir and Gallo (2013), Marcos-Martinez et al. (2017), and Ay et al. (2017)). In this paper, we believe that there are benefits to be gained by exploring spatially disaggregated data, which represent a convenient middle ground. In this regard, individual choices are aggregated in order to construct land use shares. However, unlike the case for national level data, aggregation is done only on a scale that is (i) small enough to capture the spatial variation in the environmental and climatic drivers of farmers' behavior, and (ii) proportionate with the scale of the decision-making unit. The first characteristic suggests that an important benefit, which we are able to enjoy, is the ability to explicitly model the idiosyncratic effects of policies (e.g. ESA and subsidies) and other other physical environmental drivers (e.g. mean elevation, land slope and altitude). Moreover, to satisfy the second characteristic, we shall assume in Section 5 that each of the spatial units considered is a decision-making unit (see Remark 5.1 in particular).

2.1.2. Modeling land-use shares as censored responses

A well-known difficulty of modelling spatially disaggregated data resides in the problem referred to in the literature as censoring problem. In particular, we are likely to see a wide range of land-use share values between zero and one with pile-ups at the two endpoints of zero and one. In empirical models, the failure to account for these features likely lead 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 Sections 2.3 and 3 below.

2.1.3. Allowing for potential spatial error dependence:

Often the use of the spatially disaggregated data involves some degree of spatial dependence. This may be brought about by endogenous interaction effects (e.g. peer effects), which indicates a spatial lag specification, or by the so-called Durbin effects, which is simply the exogenous interaction counterparts. Nonetheless, in our view these effects seem to be secondary within the context of a land-use share model. In this regard, a more relevant type of dependence is the SED (see also Moscone et al. (2007), and Chakir and Le Gallo (2013)). 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. When SED is not properly addressed, the usual maximum likelihood and quasi maximum likelihood (QML) methods can be severely affected, which renders statistical inferences unreliable. In the current paper, we first construct a panel-data Tobit model with error components that allow both spatial and time-wise correlations. Then, this model forms the basis for the development of our system of simultaneous Tobit equations of land-use shares (see Sections 2.2 and 2.3 for details).

2.1.4. Modelling unobserved heterogeneity in an error component structure:

Some previous studies, which are based on the aggregated data e.g. Lacroix and Thomas (2011), control for unobserved heterogeneity among farmers by using a fixed-effects model. However, by using the spatially disaggregated data, we can explicitly model the idiosyncratic effects of soils and other physical environmental drivers, which render the fixed-effects specification unnecessary. Moreover, in the fixed-effect model, factors such as land quality and other soil characteristics are assumed to be time-invariant and are therefore swept away by the estimation. Nonetheless, since data limitations can hinder a complete assessment of the influence of inter-regional biophysical and socioeconomic differences on land-use dynamics, in this paper, we model such leftover individual-effects via a random-effects model. The random effects render our model closer to a traditional censored demand system discussed in e.g. Meyerhoefer et al. (2005).

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 the many spatially distinct variants of Cliff and Ord (1973, 1981) formulations. Practical estimation of the Cliff-Ord type specifications can be computationally burdensome. This is the case even for spatial panel-data models of uncensored responses (e.g. Kapoor et al. (2007), Yang (2013), and Liu and Yang (2015)). To lighten the computational burden, Liu and Yang (2015) suggested an alternative QML procedure that involves concentrating out a subset of the parameters and maximisation of 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, here, we discuss additional properties that are crucial to the statistical validity of our hybrid framework (see Section 3 for details).

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

The censoring problem, which was discussed in Section 2.1.2, suggests that we model the land-use shares using the the two-limit Tobit model of the form

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

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

where k specifies the land-use category, $x_{k,it} = [1, x_{k,1,it}, \ldots, x_{k,J,it}]$ with J denoting the number of land-use determinants included in the model, and i and t signify the i-th grid of land and t-th time period, respectively. In this regard, let $k = 1, \ldots, K$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

With regard to the latent specification in (2.1), the general form of model is obtained by replacing 0 and 1 in (2.1) and (2.2) with a and b, where $a, b \in \mathbb{R}$ and $a < b$. Furthermore, the model is well founded since it can be viewed as a reduced form model of the well-known structural profit-maximisation problem in Chambers and Just (1989), which is extended to the context of the agricultural land-use by Fezzi and Bateman (2011), and Bateman et al (2020)).

In addition, we incorporate the RE-SED component in the two-limit 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 disturbance terms), W_k is an $N \times N$ weighting matrix of known constants (which does not involve t), ρ_k is a scalar autoregressive parameter and $\varepsilon_k(t)$ is an $N \times 1$ vector of innovations in period t. We also assume that the 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. Equations (2.3) and (2.4) suggest that the disturbances are auto-correlated both spatially and timewise. This can be seen more clearly in the scalar notations

$$u_{k,it} = \rho_k \sum_{j=1}^{N} W_{k,ij} u_{k,jt} + \varepsilon_{k,it},\ \varepsilon_{k,it} = \mu_{k,i} + v_{k,it}.$$

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 $0 < \sigma_{k,v}^2 < b_v < \infty$, and finite fourth moment. In addition, $E(v_{k,it}|X_{k,it}) = 0$ 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 $0 < \sigma_{k,\mu}^2 < b_\mu < \infty$, and finite fourth moment. In addition, $E(\mu_{k,i}|X_{k,it}) = 0$ almost surely.

(d) The processes $\{v_{k,it}\}$ and $\{\mu_{k,i}\}$ are independent. □

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

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

In the other words, the innovations $\varepsilon_{k,it}$ are intertemporally correlated within a unit, but are not spatially correlated across units. Also, the concatenation of the innovation vector with respect to time periods $t = 1, \ldots, T$ leads to

$$\varepsilon_k = (e_T \otimes I_N)\mu_k + v_k,$$

where e_T is a $T \times 1$ vector of 1s and

$$v_k = (v_k^\top(1), v_k^\top(2), \ldots, v_k^\top(T))^\top = (v_{k,11}, v_{k,21}, \ldots, v_{k,N1}, v_{k,12}, \ldots, v_{k,NT})^\top.$$

Hence, $E(\varepsilon_k) = 0$ with the variance-covariance matrix of

$$\Omega_{k,\varepsilon} = E(\varepsilon_k \varepsilon_k^\top) = \sigma_{k,v}^2 I_{NT} + \sigma_{k,\mu}^2 (J_T \otimes I_N) = \sigma_{k,v}^2 Q_0 + \sigma_{k,1}^2 Q_1,$$

where $\sigma_{k,1}^2 = \sigma_{k,v}^2 + T\sigma_{k,\mu}^2$, $Q_0 = \left(I_T - \frac{J_T}{T}\right) \otimes I_N$, $Q_1 = \frac{J_T}{T} \otimes I_N$, and $J_T = e_T e_T'$ is a $T \times T$ matrix of unit elements. In this regard, Q_0 and Q_1 are transformation matrices often seen in the error component literature (see e.g. Baltagin (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, it is immediately the case that

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

Furthermore, the similar concatenation of (2.3) leads to

$$u_k = \rho_k (I_T \otimes W_k) u_k + \varepsilon_k = [I_T \otimes (I_N - \rho_k W_k)^{-1}] \varepsilon_k, \tag{2.5}$$

where $\otimes$ denotes the Kronecker product and the following assumptions are maintained 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, while 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 $E(u_k u_k^\top) = \Omega_{k,u}$, where

$$\Omega_{k,u} = \left[I_T \otimes (I_N - \rho_k W_k)^{-1}\right] \Omega_{k,\varepsilon} \left[I_T \otimes (I_N - \rho_k W_k^\top)^{-1}\right]. \tag{2.6}$$

It is immediately the case that

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

From (2.6), it is clear that the variance-covariance matrix of the disturbance vector $u_k(t)$ 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. The System of TL-RE-SED-Tobit Models

Having discussed the TL-RE-SED-Tobit panel data model of land-use shares for each $k = 1, \ldots, K$ category, we can now construct the system of Tobit equations as simply a collection of these models over K categories.

Assuming that the cross-equation correlations satisfy

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

for all $k, l = 1, \ldots, K$, leads to the covariance matrix of the innovations of the form

$$\Omega_{kl,\varepsilon} = E(\varepsilon_k \varepsilon_l^\top) = \sigma^2_{kl,\mu}(J_T \otimes I_N) + \sigma^2_{kl,v} I_{NT}, \tag{2.7}$$

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

$$\Omega_{kl,\varepsilon} = \sigma^2_{kl,v} Q_0 + \sigma^2_{kl,1} Q_1, \tag{2.8}$$

which is obtained by defining $\sigma^2_{kl,1} = \sigma^2_{kl,v} + T\sigma^2_{kl,\mu}$. The covariance matrix of $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_K)^\top$ is then

$$\Omega_\varepsilon = E(\varepsilon \varepsilon^\top) = \Sigma_\varepsilon \otimes I_N,$$

where $\Sigma_\varepsilon = \Omega_\mu \otimes J_T + \Omega_v \otimes I_T$ is $KT \times KT$ in which $\Omega_\mu = [\sigma^2_{kl,\mu}]$ and $\Omega_v = [\sigma^2_{kl,v}]$ both with dimension $K \times K$. Alternatively,

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

where $\Omega_1 = [\sigma^2_{kl,1}]$.

Regarding the disturbances, we note first that

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

where $H_k = B_k^{-1}$ with $B_k = (I_N - \rho_k W_k)$, so that

$$\Omega_{kl,u} = \left\{\sigma^2_{kl,1} \bar{J}_T + \sigma^2_{kl,v}(I_T - \bar{J}_T)\right\} \otimes H_k H_l^\top \tag{2.9}$$

by (2.7) and (2.8), where $\bar{J}_T = J_T/T$. The variance-covariance matrix of the system disturbances is then

$$\Omega_u = E(uu^\top) = A\Omega_\varepsilon A^\top,$$

where $u = (u_1, u_2, \ldots, u_K)^\top$ and $A = \mathrm{diag}(A_{11}, A_{22}, \ldots, A_{KK})$ with $A_{kk} = I_T \otimes H_k$.

Finally, we conclude this section by introducing spatial transformations that will be useful for discussion in the next section. Let e_{NT} denote an $NT \times 1$ vector of ones, $Y = [Y_1, \ldots, Y_K]^\top$, where $Y_k = [Y_k(1), \ldots, Y_k(T)]^\top$ with $Y_k(t) = [y_{k,1t}, \ldots, y_{k,Nt}]^\top$, and $X = \text{diag}[x_1, x_2, \ldots, x_K]$, in which $x_k = [x_k(1), \ldots, x_k(T)]^\top$ and $x_k(t) = [x_{k,1}(t), \ldots, x_{k,J}(t)]$. In this regad,

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

are the Cochrance-Orcutt-type spatial transformations, which will be essential components of the estimation procedure in the next section.

3. Hybrid QML/GMM estimation procedure

This section proposes a hybrid QML/GMM estimation procedure for estimating the system of TL-RE-SED-Tobit equations for the K land-use categories. Overall, this hybrid procedure consists of four main steps, namely (1) estimating the TL-RE-SED-Tobit panel data model for the k-th category of land-use, (2) performing GMM estimation of the spatial parameter ρ_k, (3) constructing the spatial Cochrane-Orcutt-type transformation, and (4) estimating the system of TL-RE-SED Tobit equations for the K land-use categories. We will now discuss these steps in more detail.

Step 1: Estimating the TL-RE-SED-Tobit panel data model

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

(a) Regarding the QML estimation, it is well known that the presence of heteroskedasticity is likely to lead to inconsistent estimates. However, consistent estimation can made be possible by specifying a model for heteroskedasticity, particularly

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

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

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

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

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

and $1[\,\cdot\,]$ signifies an indicator function. Lemma 3.1 below confirms that consistent estimates of the disturbances can be obtained from this QML estimation. Mathematical proof of this lemma requires imposing further assumptions as follows.

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. $x_{k,l}$,*

$$\lim_{N\to\infty} x_{k,l}^{\top}x_{k,l} \to \infty, \;\; \lim_{N\to\infty} x_{k,l,it}^{2}/x_{k,l}^{\top}x_{k,l} \to 0 \text{ and } E(x_{k,l,it}^{4}) < \infty$$

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 $k = 1, \ldots, K$. □

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

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 (e.g. Amemiya (1973)). Furthermore, the proof of Lemma 3.1 can be found in Amemiya (1973). The only exception resides in an additional random variable, which reflects that our model is a two-limit case, as follows

$$\begin{aligned} v_{k,1,it} &= \begin{cases} 1 & \text{with probability } F_{k,1,it} \\ 0 & \text{with probability } 1 - F_{k,1,it} \end{cases} \\ v_{k,2,it} &= \begin{cases} 1 & \text{with probability } F_{k,2,it} \\ 0 & \text{with probability } 1 - F_{k,2,it} \end{cases} \\ v_{k,3,it} &= \begin{cases} 1 & \text{with probability } 1 - F_{k,1,it} - F_{k,2,it} \\ 0 & \text{with probability } F_{k,1,it} + F_{k,2,it} \end{cases}, \end{aligned}$$

where $F_{k,1,it} = \text{Prob}(y_{k,it}^{*} \leq 0)$ and $F_{k,2,it} = \text{Prob}(y_{k,it}^{*} \geq 1)$.

Step 2: Performing the GMM estimation of the spatial parameter ρ_k

Upon the completion of Step 1, consistent residuals are obtained for the k-th land-use category. Then, for uncensored observations, the required devolatilised residuals are constructed as

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

where $\widetilde{\sigma}_{k,u,it}$ and $\widetilde{\beta}_k$ represent the QML estimates for the standard errors defined in (3.1) and parameter estimates obtained in Step 1, respectively. Otherwise, generalised residuals for censored observations are computed via the inverse Mills ratio

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

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal density and cumulative distribution functions, respectively. For example, for observations left-censored at 0, the generalised residuals can be computed as $\widetilde{u}_{k,it} = -\widetilde{\lambda}_{k,it}$. Let $\widetilde{u}_k = (\widetilde{u}_{k,11}, \widetilde{u}_{k,21}, \ldots, \widetilde{u}_{k,N1}, \widetilde{u}_{k,12}, \ldots, \widetilde{u}_{k,NT})^\top$.

The current step involves estimating the SAR parameter by using the GMM procedure introduced in Kapoor et al. (2007), in which $\widetilde{u}_k$ is used in place of the true disturbances.To be accustomed to such practice, one only has to note that generalised residuals are commonly used for performing diagnostic tests in the standard Tobit model literature (e.g. Cameron and Trivedi (2005)). Although the QML estimation is used in Step 1, unlike Kapoor et al. (2007) who employed the ordinary least squares, Lemma 3.1 suggests that consistency of our GMM estimators for ρ_k, $\sigma^2_{k,v}$ and $\sigma^2_{k,1}$ can be shown in a similar fashion. In particular:

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

Step 3: Constructing the spatial Cochrane-Orcutt-type transformations

Upon the completion of Steps 1 and 2, for all $k = 1, \ldots, K$, the GMM estimates $\widehat{\rho}_1, \ldots, \widehat{\rho}_K$ are readily available. These are then used in the estimation of the matrix A_{kk}, namely

$$\widehat{A}_{kk} = I_T \otimes \widehat{H}_k,$$

where $\widehat{H}_k = (I_N - \widehat{\rho}_k W_k)^{-1}$. The current step involves computation of the Cochrane-Orcutt-type spatial transformations of Y and X. Particularly,

$$\dot{\mathcal{X}} = \widehat{A}^{-1}\widetilde{X} \quad \text{and} \quad \dot{\mathcal{Y}} = \widehat{A}^{-1}\widetilde{Y},$$

where $\widetilde{Y} = [\widetilde{Y}_1, \widetilde{Y}_2, \ldots, \widetilde{Y}_K]$ and $\widetilde{X} = [\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_K]$ are the devolatilised versions of Y and X based on $\widetilde{\sigma}_{k,u,it}$, respectively.

Step 4: Estimating the system of TL-RE-SED Tobit equations

We first note an important drawback of the traditional Amemiya-Tobin mechanism, which resides in the fact that the adding-up restriction holds only for the latent equations but not for the observed land-use shares. We address this issue by treating the K-th land-use share as a residual category with no specific land-use demand of its own. As the results, the current step focuses on estimating the system of TL-RE-SED Tobit models of $(K-1)$ land-use shares based on the Cochrane-Orcutt-type spatial transformations and the QML estimation.

This step involves the QML estimation of the system of TL-RE-SED Tobit models of land-use shares. We follow the suggestion made by Yen et. al. (2003) and specify the quasi-likelihood function based on a sequence of bivariate Tobit likelihoods. In this regard, since a pooled estimation is considered, it is notationally simpler to write

$$\dot{\mathcal{Y}}_k = (\dot{y}_{k,11}, \dot{y}_{k,21}, \ldots, \dot{y}_{k,N1}, \dot{y}_{k,12}, \ldots, \dot{y}_{k,NT})^\top \equiv (\dot{y}_{k,1}, \dot{y}_{k,2}, \ldots, \dot{y}_{k,NT})^\top$$

and $\iota = 1, 2, \ldots, NT$. For the k-th and j-th TL-RE-SED equations, let $\dot{u}_{k,\iota} = [\dot{y}_{k,\iota} - \dot{x}_{k,\iota}\bar{\beta}_k]/\bar{\sigma}_k$ and $\dot{u}_{j,\iota} = [\dot{y}_{j,\iota} - \dot{x}_{j,\iota}\bar{\beta}_j]/\bar{\sigma}_j$, respectively, where $\dot{x}_{k,\iota} = [\dot{x}_{k,1,\iota}, \ldots, \dot{x}_{k,J,\iota}]$ (the ι-th row of $\dot{x}_k$). In addition, $1[\dot{y}_{k,\iota} = 0, \dot{y}_{j,\iota} = 0]$ is a dichotomous indicator which equals 1 when $\dot{y}_{k,\iota} = 0$ and $\dot{y}_{j,\iota} = 0$. The bivariate Tobit likelihood for the ι-th observation is then

$$\begin{aligned}
L_{k,j,\iota} \;=\; & \left\{\Psi(\dot{u}_{k,\iota}\dot{u}_{j,\iota}; \bar{r}_{kj})\right\}^{1[\dot{y}_{k,\iota}=0, \dot{y}_{j,\iota}=0]} \\
\times & \left\{\bar{\sigma}_k^{-1}\bar{\sigma}_j^{-1}(1-\bar{r}_{kj}^2)^{-1/2}\psi(\dot{u}_{k,\iota}, \dot{u}_{j,\iota}; \bar{r}_{kj})\right\}^{1[0<\dot{y}_{k,\iota}<1, 0<\dot{y}_{j,\iota}<1]} \\
\times & \left\{\Psi(-\dot{u}_{k,\iota}, \dot{u}_{j,\iota}; -\bar{r}_{kj})\right\}^{1[\dot{y}_{k,\iota}=1, \dot{y}_{j,\iota}=0]} \\
\times & \left\{\Psi(\dot{u}_{k,\iota}, -\dot{u}_{j,\iota}; \bar{r}_{kj})\right\}^{1[\dot{y}_{k,\iota}=0, \dot{y}_{j,\iota}=1]} \\
\times & \left\{\bar{\sigma}_k^{-1}\phi(\dot{u}_{k,\iota})\Phi\left[(\dot{u}_{j,\iota} - \bar{r}_{kj}\dot{u}_{k,\iota})/(1-\bar{r}_{kj}^2)^{1/2}\right]\right\}^{1[0<\dot{y}_{k,\iota}<1, \dot{y}_{j,\iota}=0]} \\
\times & \left\{\bar{\sigma}_j^{-1}\phi(\dot{u}_{j,\iota})\Phi\left[(\dot{u}_{k,\iota} - \bar{r}_{kj}\dot{u}_{j,\iota})/(1-\bar{r}_{kj}^2)^{1/2}\right]\right\}^{1[\dot{y}_{k,\iota}=0, 0<\dot{y}_{j,\iota}<1]},
\end{aligned} \tag{3.2}$$

where $\psi(\cdot,\cdot,\cdot)$ and $\Psi(\cdot,\cdot,\cdot)$ are the bivariate standard normal probability density function and corresponding cumulative distribution, respectively. Hence, the QML estimators of the vector of true parameters

$$\theta = \begin{pmatrix} \beta_1^\top & \ldots & \beta_{K-1}^\top & \sigma_{1,\varepsilon}^2 & \ldots & \sigma_{K-1,\varepsilon}^2 & r_{12} & \ldots & r_{K-2,K-1} \end{pmatrix}^\top$$

can be obtained by maximising the quasi-likelihood of

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

Let us conclude the current section by discussing the consistency of the proposed QML estimation. We first define

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

in which

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

where $\dot{u}^0_{k,\iota} = [\dot{y}^0_{k,\iota} - \dot{x}^0_{k,\iota}\bar{\beta}_k]/\bar{\sigma}_k$, $\dot{u}^0_{j,\iota} = [\dot{y}^0_{j,\iota} - \dot{x}^0_{j,\iota}\bar{\beta}_j]/\bar{\sigma}_j$, $\dot{y}^0_{k,\iota}$ and $\dot{x}^0_{k,\iota}$ denote elements of the transformed $\widetilde{Y}$ and $\widetilde{X}$, respectively, with A^{-1} instead of $\widehat{A}^{-1}$. Then, establishing the consistency of the proposed QML estimation requires showing that

$$\mathcal{L}(\bar{\theta}) = \mathcal{L}^0(\bar{\theta}) + O_P\left((NT)^{-1/2}\right) \tag{3.3}$$

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, whereas its proof is relegated to the appendix.

Theorem 3.1. *Let* $\inf\limits_{u_{k,\iota}u_{j,\iota}\in\mathbb{R}^2} \psi(u_{k,\iota},u_{j,\iota}) = \delta_1$ *and* $\inf\limits_{u_{k,\iota}\in\mathbb{R}} \phi(u_{k,\iota}) = \delta_2$, *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 $N \to \infty$. □

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

For a given $\Omega_{kk,u}$, studies in spatial econometrics (e.g. Baltagi and Li (2006), Baltagi et al. (2012), Chakir and Gallo (2013), and Ay et al. (2017)) suggest that the best linear unbiased predictor for the i-th individual at a future period $T+\tau$ is of the form

$$\widehat{y}^{*}_{k,i,T+\tau} = X_{k,i,T+\tau}\widehat{\beta}_k + \omega'_{k,i}\Omega^{-1}_{kk,u}\widehat{u}_k, \tag{4.1}$$

where $\widehat{\beta}_k$ denotes the parameter estimate obtained in Step 4 of Section 3 and $\omega_{k,i} = E[u_{k,i,T+\tau}u_k]$ is the co-variance between future and current disturbances. Below, we construct the predictor $\widehat{y}^{*}_{k,i,T+\tau}$ in the context of the above-discussed system.

To derive $\omega_{k,i}$, we first note that $u_k(t) = B_k^{-1}(\mu_k + v_k(t))$ and $u_k = (e_T \otimes H_k)\mu_k + (I_T \otimes H_k)v_k$, then write

$$\begin{aligned} E[u_k(T+\tau)u'_k] &= E[B_k^{-1}(\mu_k + v_k(T+\tau))((e_T \times H_k)\mu_k + (I_T \otimes H_k)v_k)'] \\ &= \sigma^2_{\mu,kk}H_k(e'_T \times H'_k), \end{aligned}$$

which is $N \times TN$. These lead to

$$E[u_{k,i,T+\tau}u'_k] = \sigma^2_{\mu,kk}h_{k,i}(e'_T \times H'_k), \tag{4.2}$$

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

Equations (2.9) and (4.2) suggest collectively that

$$\omega'_{k,i}\Omega^{-1}_{kk,u} = \frac{\sigma^2_{\mu,kk}}{\sigma^2_{1,kk}}h_{k,i}(e'_T \otimes B_k), \tag{4.3}$$

which is obtained due to $e'_T = e'_T\bar{J}_T$ and $\frac{\sigma^2_{\mu,kk}}{\sigma^2_{v,kk}} - \frac{\sigma^2_{\mu,kk}}{\sigma^2_{v,kk}} \cdot \frac{T\sigma^2_{\mu,kk}}{\sigma^2_{1,kk}} = \frac{\sigma^2_{\mu,kk}}{\sigma^2_{1,kk}}$. Since $h_{k,i}$ is the i-th row of $H_k = B_k^{-1}$ and $B_k^{-1}B_k = I_N$, it is clear that

$h_{k,i}B_k = l'_{k,i}$, where $l'_{k,i}$ is the i-th row of I_N. Accordingly, $h_{k,i}(e'_T \otimes B_k) = (1 \otimes h_{k,i})(e'_T \otimes B_k) = (e'_T \otimes l'_{k,i})$, which is $(1 \times TN)$, and hence

$$\omega'_{k,i}\Omega_{kk}^{-1} = \frac{\sigma^2_{\mu,kk}}{\sigma^2_{1,kk}}(e'_T \otimes l'_{k,i}). \tag{4.4}$$

Substitution of these results into (4.1) leads to two equivalent unbiased predictors for the i-th individual at a future period $T+\tau$, namely

$$\widehat{y}^*_{k,i,T+\tau} = X_{k,i,T+\tau}\widehat{\beta}_k + \frac{\sigma^2_{\mu,kk}}{\sigma^2_{1,kk}}h_{k,i}(e'_T \otimes B_k)\widehat{u}_k, \tag{4.5}$$

and

$$\widehat{y}^*_{k,i,T+\tau} = X_{k,i,T+\tau}\widehat{\beta}_k + \frac{\sigma^2_{\mu,kk}}{\sigma^2_{1,kk}}(e'_T \otimes l'_{k,i})\widehat{u}_k. \tag{4.6}$$

The following remarks provide further clarifications of these predictors. Firstly, $\widehat{y}^*_{k,i,T+\tau}$ 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 5.3. Secondly, although a single equation model should lead to the same formulas, here via practical computation of such predictors relies on the residual vectors obtained via a system that takes inter-equation correlations into account. If the innovations are known, Baltagi (1980) (also Baltagi and Pirotte (2011)) suggests that the required variances could be calculated via

$$O_v = \epsilon^\top Q_0 \epsilon / N(T-1) \quad \text{and} \quad O_1 = \epsilon^\top Q_1 \epsilon / N,$$

where $\epsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_K]$ is the $NT \times K$ matrix of innovations. However, in an empirical analysis, we must rely on residuals formulated from the system estimation in Step 4 of Section 3.

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

Our objective is twofold. Firstly, it is to examine the predictive accuracy of the TL-RE-SED specification. Since prediction of future land-use areas is essential, superior predictive power should lay an important foundation for the use of such a specification in an empirical analysis. Secondly it is to investigate how the above-mentioned determinants influence the land-use patterns in England over time and spatial space. It is also our goal to examine 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.

To achieve these goals, the analysis in this section focuses on five types of land-use, namely (1) arable, i.e. land planted to cereals (e.g. wheat, barley, and oats) and root crops (e.g. potatoes and sugar beet excluding oilseed rape), (2) temporary grassland, i.e. grassland that is typically part of an arable crop rotation, (3) permanent grassland, i.e. grassland maintained perpetually without reseeding, (4) rough grazing, i.e. uncultivated land used for grazing livestock, and (5) oilseed rape. The first four categories are the main land-use types for the English agricultural sector, while 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. These land-use shares are computed by constructing the total farm land as a summation of these five land-uses. Moreover, land-use determinants can be classified into three categories, namely (i) economics, (ii) climatic and physical environment, and (iii) environmental policy. Moreover, Table 2 presents a full list of the exogenous variables used in our analysis.

In the sections that follow we first present detailed descriptions of the data and their sources, then specify a set of empirical specifications, which form the basis of our analysis.

5.1. Data descriptions and sources

We use a unique database that consists of data compiled from various sources at the Land, Environment, Economics and Policy (LEEP) Institute, University of Exeter. Below, we give some details about these data.

Agricultural land-use data: Data on agricultural land-use are derived from the June Agricultural Census on a 2-km^2 (400 ha) grid, which are available online from the Edinburgh University Data Library. These data cover England and Wales for seventeen 17 spaced years between 1969 and 2006 and yield roughly 38,000 grid-square records each year.

Physical environment and climatic data: These data concern climate, environmental and topographic variables as follows. (i) Climate related variables are for a growing season (i.e. April to September), namely average temperature and accumulated rainfall. These data were initially obtained on a 5-km^2 grid from the Met Office, which are calculated as the average climate between years 1981 and 2010, then interpolated to the 2-km^2 grid. (ii) Environmental and topographic variables that may influence farmers' decisions. The former include soil characteristics (namely the portion of peat, stones, gravel or fragipan soil) and three dummy variables for various soil textures (specifically the proportion of fine, medium, and coarse soils), all obtained from the Harmonised World Soil Database. The latter include the mean altitude and slope, which are both derived from the 50-m resolution Integrated Hydrological Digital Terrain Model (licensed from the Centre for Ecology and Hydrology).

Policy determinants of land-use decisions: These include share of each grid square designated as National Parks, ESA, set-asides and greenbelts. ESAs were introduced in 1987 (and extended in subsequent years) to conserve and enhance areas of particular landscape and wildlife significance. In addition, the spatial data for English greenbelts were licensed by Defra from the Ordnance Survey.

Transportation cost: Transportation cost is proxied by the the distance to the closest major market (defined as an urban centre with more than 300,000 inhabitants according to the 2011 Census data).

Remark 5.1. *(a) 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 analysis these are 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). (b) We suggested in Section 2.1.1 that here the aggregation of the individual data is done only on a scale that is (i) small enough to capture the spatial variation in the environmental and climatic drivers of farmers' behavior, and (ii) 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. the 2-km^2 grid, represents a decision-making unit. The average farm size in the North East of England, for example, was just below 1.5-km^2 (150 hectares) in 2018 (Department for Environment, Food and Rural Affairs (2020)).*

The formulation in Section 2.2 suggests that the TL-RE-SED-Tobit Model only accepted 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.

Table 1 presents descriptive statistics for the areas of land used in hectares. The total agricultural land is computed as the sum of land used for each of the categories. In addition, the table 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 steadily (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. Moreover, Table 2 presents a full list of the exogenous variables used in our analysis and their definitions. These will form competing empirical specifications to be used in conjunction with the TL-RE-SED Tobit model, in which an interaction term is included to allow for the effect of temperature to depend on rainfall and vice versa.

5.2. 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 environmental, climatic and policy variables is described by

$$E\left[y^*_{k,it}|x_{k,it}\right] = x_{k,it}\beta_k, \tag{5.1}$$

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 2.

Since the specification in (5.1) 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 $\zeta_k(temp_{it}) = \beta_{k,t9}temp_{9,it} + \cdots + \beta_{k,t14}temp_{14,it}$. Such a specification leads to the inclusion of 41 exogenous variables in the model.

Spatial Weighting Matrices Let us recall that W_k denotes an $N \times N$ weighting matrix of known constants that satisfies Assumption 2.3. A number of previous studies suggest that predictive accuracy and empirical results in general may be sensitive to the choice of W_k (i.e. Anselin and Bera (1998) and Bhattacharjee, and Jensen-Butler (2006)). To investigate such sensitivity, we consider weighting matrices based on both the inter-point-distance and the graph-based neighbours. In particular, 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}. Finally, all the spatial weighting matrices are row-normalized (i.e. each elements is divided by the sum of its rows).

The 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 land-use 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 land use 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.

5.3. Predictive Evaluation

First observe that $\widehat{y}_{k,i,T+\tau}$ in (4.6) is the best linear unbiased predictor of the latent variable $y^*_{k,i,T+\tau}$. As a result, the 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. In addition, the training data are from 1976, 1979, 1981, 1988, 2000 and 2004, which suggests that $T = 6$, whereas the loss function of interest is the root mean squared errors (RMSE), by which the observed land-use shares in 2010 are treated as the validation dataset.

Moreover, our examination focuses on comparing RMSEs from a number of alternative predictors. These are formulated 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:

$$\widehat{y}^*_{k,i,T+\tau} = x_{k,i,T+\tau}\widehat{\beta}_k + \widehat{\vartheta}_k(rain_{it}) + \widehat{\zeta}_k(temp_{it}) + \frac{\widehat{\sigma}^2_{\mu,kl}}{\widehat{\sigma}^2_{1,kl}}(e'_T \otimes l'_{k,i})\widehat{u}_k$$

We are interested in two set of comparisons. Firstly, they are comparisons between the predictors listed under categories A and B. These are important because they can help to confirm the asymptotic compatibility between $\tilde{\beta}_k$ and $\widehat{\beta}_k$. Secondly, they are comparisons between the predictors listed under categories B and C. These are significant since they can affirm the need to incorporate the random effects and spatial error dependence into the model in order to improve the predictive accuracy.

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

$$LM_{BP} = \frac{NT}{2(T-1)} \left[\left(\frac{\sum_{i=1}^{N} \left[\sum_{t=1}^{T} \varepsilon_{it} \right]^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2} \right) - 1 \right]^2 . \tag{5.2}$$

The limiting distribution of LM_{BP} is chi-squared with one degree of freedom under the null hypothesis $H_0 : \sigma_{k,\mu}^2 = 0$. 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 (2.5) or $\varepsilon_k = u_k - \rho_k(I_T \otimes W_k)u_k$.

Tables 3 to 6 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 $[a]$ in each table do not vary significantly from one another. These suggest that $\tilde{\beta}_k$ and $\widehat{\beta}_k$ do not vary significantly from one another. Such findings are expected and theoretically deducible from the estimation consistency. (ii) The RMSEs reported in rows $[b]$ in each table (even under the 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}. In addition, the corresponding estimates of $\widehat{\rho}_k$ are also relatively close between the two weighting matrices (see rows [c] in particular). The most likely reason underpinning such closeness is the similarity in the degree of spareness implied by these weighting matrices. Furthermore, we find that a higher degree of spareness (in the weighting matrix) is usually associated with higher estimates of $\widehat{\rho}_k$ but lower estimates of $\widehat{\sigma}^2_{k,\mu}$. Nonetheless, the evidence is not conclusive on which specifications of the weighting matrices are able to bring about better forecasting.

In Tables 3 to 6, rows [d] present the corresponding LM_{BP} test statistic and the associated p-value under different weighting matrices. In all cases, the LM_{BP} test statistics far exceed the 95% critical value for the chi-squared distribution with one degree of freedom, which is 3.84. These lead to rejection of the null hypothesis and therefore suggest that the superiority in predictive accuracy reported above 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.

5.4. Estimation results and important findings

Having taken into consideration the empirical and statistical issues as outlined in Section 2.1, our TL-RE-SED framework is able to formulate consistent parameter estimates and more accurate variance and covariance structures. In addition, we have shown that the TL-RE-SED specification possesses a greater predictive power of future land-use. Hence, it serves an excellent platform for investigating (i) how the climatic, economic and physical-environment determinants influence land-use patterns in England over time and spatial space, and (ii) 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, especially oilseed rape, to help to achieve deep emission reductions and prepare for climate change.

Tables 8 to 11 present the estimation results in five columns. The second, third and fourth columns present the parameter estimates associated with W_k^{5NN}, W_k^{2NN}, and W_k^{SOI}, respectively. For the sake of comparison, we also present in the fifth column the parameter estimates obtained without taking into consideration the random effects and spatial error dependence. Below, we discuss a number of important findings.

We begin with some general observations. Although they are based on different weighting matrices, the parameter estimates presented in the second to fourth columns are fairly similar with respect to both their signs and magnitude. In addition, except for those of the time dummies, these estimates show similar congruence to that in the fifth column. However, the same cannot be said for the associated p-values. Clearly, in the fifth column, a higher proportion of the computed p-values are less than 0.01, 0.05 and 0.1 than in other columns. In this regard, since the spatial error dependence is entirely ignored, the variance estimates might be significantly under-estimated, which could lead to a superfluous number of rejections. It should be also noted that among the three spatial weighting matrices, W_k^{2NN} contains the largest proportion of zeros. In this regard, the resulting model offers the most congruent results in terms of the number of rejections with the model without RE-SED.

We now focus on a list of policy-related findings. Among the schemes being considered, set-asides and national parks are the only two schemes whose results agree across all model specifications. Our results suggest that set-asides are the most effective scheme, which achieves the intended negative effect statistically significantly across all types of agricultural land-use shares. The opposite is true for national parks, however. For the remaining schemes, there appears to be conflicting findings about their effectiveness among models with and without the RE-SED. With regard to greenbelts, the RE-SED-based models suggest that there are effective in reducing the share of permanent-grassland, but the effect becomes positive and statistically insignificant for a model without the RE-SED. In addition, the scheme seems to have an unwanted positive effect on the shares of rough grazing and temporary grassland. For the former, this effect is statistically significant in all models, for the latter, it is significant only in the model without the RE-SED. Moreover, only the models without the RE-SED suggest that the ESAs has statistically significant effects. These effects are positive, however.

Next, we proceed to the results that involve the time dummies, where 2004 is used as the reference year. Let us recall that these time dummies are instrumental in capturing the effects of economic and financial incentives, such as prices and other environmental grants, the spatial observations of which cannot be accurately formulated in practice. With only a few exceptions, the parameter estimates are generally positive and statistically significant. Here, the positivity suggests that on increase in time (e.g. from 2000 to 2004) is associated with negative effects on the corresponding land-uses. Clearly, these effects can be brought about by financial incentives to the farmers via price changes and other government grants. Estimating the model without the RE-SED results in statistically significant effects for all cases. The spatial models suggest, nonetheless, that the effects of government grants might only become statistically significant in recent years (i.e.

from 1988 onward). This finding is congruent with the fact that the governmental grants have been more widely and systematically introduced since the early 1990s.

We now shift our attention to the physical environment and climatic drivers. Our results suggest that physical environment (e.g. soil texture and characteristics) are more influential on farmers' land-use decisions than temperature and rainfall. All the models (both with and without RE-SED) suggest that rainfall does not statistically and significantly affect the land-use shares in question. We think the underpinning reason for this might be the fact that farmers usually have more complete information about these physical determinants. On the other hand, only the RE-SED models suggest that temperature affects the shares of temporary and permanent grasslands. The effect is negative for the former, while it is positive for latter. Slope is shown by all the models to have positive effect on the rough grazing share and a negative effect on arable. Similar effects are also reported for permanent and temporary grasslands, but are not statistically significant (with an exception of a case under the model without RE-SED).

A general observation can be made regarding the estimation results for soil characteristics and textures. Across all types of land-use, models with W_k^{2NN} and without RE-SED are more likely to have larger p-values than their the W_k^{5NN}-based and W_k^{SOI}-based counterparts. Our estimation results also suggest that differences in soil characteristics have significant effect on the share of temporary grassland. A higher share of peat seems to show some negative effects, whereas the opposite is true for the shares of gravel and stone. The same can also be said about the share of permanent grassland, but only the effect of gravel share appears to be statistically significant across all models. Regarding rough grazing, the only statistically significant effect is the negative effect of the share of stone. By holding "medium" as the reference category, our estimation results suggest that an increase in the proportion of fine texture with a proportionate decrease in medium texture, leads to a statistically significant positive effect on the share of temporary grassland. In addition, an increase in the proportion of coarse texture, with a proportionate decrease in medium texture, leads to a statistically significant negative effect on the share of permanent grassland. Such an increase has a positive effect on the share of rough grazing, however. Lastly, an increase in the collective proportion of fine and fragipan soil with a proportionate decrease in medium texture leads to a statistically significant negative effect on the share of arable land-use. Finally, the shares of temporary grassland, rough grazing and arable tend to be higher for lands located in the south of England than those in the central part. The opposite is true, however, for the share of permanent grassland. Land being located in the north of England does not seem to cause the land-use shares to differ from those in the Midlands.

6. Conclusions

This paper addressed a number of statistical and empirical challenges associated with modelling farmers' decision-making processes concerning the agricultural land-use. These included (i) the use of spatially disaggregated data so that the idiosyncratic effects of soils and other physical environment drivers could be explicitly modelled; (ii) modelling land-use shares as censored responses which enabled consistent estimation of the unknown parameters; (iii) incorporating SED and heterogeneity within an error component structure, which led to accurate formulation of the variances for the parameter estimates and, consequently, to better effective statistical inferences; and (iv) reducing the computational burden (and therefore improving estimation accuracy) under an alternative GMM/QML hybrid estimation procedure. In addition to these statistical advantages, we also provided extensive empirical evidence, which illustrated that our approach was able to construct more accurate land-use predictions than existing models in the literature. We then applied our method to empirically investigate how the climatic, economic, policy and physical determinants influence the land-use patterns in England over time and spatial space. We were also interested in examining 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 found that the effects of government grants might only become statistically significant in recent years (i.e. from 1988 onward), which was congruent with the fact that the governmental grants have been more widely and systematically introduced since the early 1990s. We also found that set-asides were the most effective scheme, which achieved the intended statistically significant negative effects across all types of agricultural land-uses. However, the opposite was true for national parks.

7. Appendix: Proof of Theorem 3.1:

The proof of Theorem 3.1 consists of two steps. The first step is to prove (3.3). The identification condition is established by discussing a counter-argument. Let us show the convergence of $\mathcal{L} \to_P \mathcal{L}^0$ uniformly over the parameter space, Θ. Firstly, let us note that $\ln L - \ln L^0 = \ln L - \ln L^* + \ln L^* - \ln L^0$, where

$$L^* = \prod_{\iota=1}^{NT} L^*_{1,K-1,\iota} \prod_{k=2}^{K-1} \prod_{j=1}^{k-1} L^*_{k,j,\iota}$$

with

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

Clearly, $\mathcal{L}^* - \mathcal{L}^0 \to_P 0$ uniformly over Θ, where $\mathcal{L}^* = \frac{1}{NT}\sum_{\iota=1}^{NT} \ln L^*$. Since the only difference between $\mathcal{L}^*$ and $\mathcal{L}^0$ is the number of corresponding dichotomous indicators for each component of the quasi-likelihood functions, establishing $\mathcal{L} - \mathcal{L}^0 \to_P 0$ is equivalent to showing that $\mathcal{L} - \mathcal{L}^* \to_P 0$ uniformly over Θ.

This can be achieved by the following two steps. We first show the pointwise consistency of $\mathcal{L}$ to $\mathcal{L}^*$, then establish the stochastic equi-continuity by showing the uniform Lipschitz continuity over Θ. For the sake of convenience, let us present $\mathcal{L} - \mathcal{L}^*$ as follows

$$\begin{aligned} \frac{1}{NT}\ln L - \frac{1}{NT}\ln L^* &= \frac{1}{NT}\sum_{\iota=1}^{NT} \Bigg(\ln L_{1,K-1,\iota} - \ln L^*_{1,K-1,\iota} \\ &\quad + \sum_{k=2}^{K-1}\sum_{j=1}^{k-1}\left\{\ln L_{k,j,\iota} - \ln L^*_{k,j,\iota}\right\}\Bigg), \end{aligned} \tag{5.3}$$

where

$$
\begin{aligned}
\sum_{\iota=1}^{NT} \ln L_{k,j,\iota} \quad = \quad & \sum_{S_1} \ln \Psi(\dot{u}_{k,\iota}, \dot{u}_{j,\iota}; \bar{r}_{kj}) + \sum_{S_2} \ln \left(\bar{\sigma}_k^{-1} \bar{\sigma}_j^{-1} (1-\bar{r}_{kj}^2)^{-1/2} \psi(\dot{u}_{k,\iota}, \dot{u}_{j,\iota}; \bar{r}_{kj}) \right) \\
& + \sum_{S_3} \ln \Psi(-\dot{u}_{k,\iota}, \dot{u}_{j,\iota}; -\bar{r}_{kj}) + \sum_{S_4} \ln \Psi(\dot{u}_{k,\iota}, -\dot{u}_{j,\iota}; \bar{r}_{kj}) \\
& + \sum_{S_5} \ln \left(\bar{\sigma}_k^{-1} \phi(\dot{u}_{k,\iota}) \Phi \left[(\dot{u}_{j,\iota} - \bar{r}_{kj} \dot{u}_{k,\iota})/(1-\bar{r}_{kj}^2)^{1/2} \right] \right) \\
& + \sum_{NT - \sum_{i=1}^5 S_i} \ln \left(\bar{\sigma}_j^{-1} \phi(\dot{u}_{j,\iota}) \Phi \left[(\dot{u}_{k,\iota} - \bar{r}_{kj} \dot{u}_{j,\iota})/(1-\bar{r}_{kj}^2)^{1/2} \right] \right)
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{\iota=1}^{NT} \ln L^*_{k,j,\iota} \quad = \quad & \sum_{S_1} \ln \Psi(\dot{u}^0_{k,\iota}, \dot{u}^0_{j,\iota}; \bar{r}_{kj}) + \sum_{S_2} \ln \left(\bar{\sigma}_k^{-1} \bar{\sigma}_j^{-1} (1-\bar{r}_{kj}^2)^{-1/2} \psi(\dot{u}^0_{k,\iota}, \dot{u}^0_{j,\iota}; \bar{r}_{kj}) \right) \\
& + \sum_{S_3} \ln \Psi(-\dot{u}^0_{k,\iota}, \dot{u}^0_{j,\iota}; -\bar{r}_{kj}) + \sum_{S_4} \ln \Psi(\dot{u}^0_{k,\iota}, -\dot{u}^0_{j,\iota}; \bar{r}_{kj}) \\
& + \sum_{S_5} \ln \left(\bar{\sigma}_k^{-1} \phi(\dot{u}^0_{k,\iota}) \Phi \left[(\dot{u}^0_{j,\iota} - \bar{r}_{kj} \dot{u}^0_{k,\iota})/(1-\bar{r}_{kj}^2)^{1/2} \right] \right) \\
& + \sum_{NT - \sum_{i=1}^5 S_i} \ln \left(\bar{\sigma}_j^{-1} \phi(\dot{u}^0_{j,\iota}) \Phi \left[(\dot{u}^0_{k,\iota} - \bar{r}_{kj} \dot{u}^0_{j,\iota})/(1-\bar{r}_{kj}^2)^{1/2} \right] \right)
\end{aligned}
$$

where S_i represents the number of corresponding dichotomous indicators in (3.2) for $i = 1, \ldots, 5$ (e.g. S_1 is the number of dichotomous indicators of $1[\ddot{y}_{k,\iota} = 0, \ddot{y}_{j,\iota} = 0]$). In this proof, we mainly focus on the three main terms of the differences in a single bivariate quasi log-likelihood function in (5.3). Because other two terms can be shown in a similar manner to those three terms, the rest of the bivariate quasi log-likelihood functions can be dealt in the same manner. In addition, the parameter space, Θ, is a countable union of the compact parameter spaces of $\Theta_k s$, where Θ_k is the compact subset of Θ such that $\bar{\theta}_k \in \Theta_k$ with $\bar{\theta}_k = \begin{pmatrix} \bar{\beta}_k^\top & \bar{\sigma}_k^2 & \bar{r}_{kj} \end{pmatrix}^\top$. The three main terms to be considered are:

$$
\ln \Psi(\dot{u}_{k,\iota}, \dot{u}_{j,k}; \bar{r}_{kj}) - \ln \Psi(\dot{u}^0_{k,\iota}, \dot{u}^0_{j,\iota}; \bar{r}_{kj}),
$$

$$
\ln \left(\psi(\dot{u}_{k,\iota}, \dot{u}_{j,\iota}; \bar{r}_{kj}) \right) - \ln \left(\psi(\dot{u}^0_{k,\iota}, \dot{u}^0_{j,\iota}; \bar{r}_{kj}) \right)
$$

$$
\left(\ln \phi(\dot{u}_{k,\iota}) + \ln \Phi \left[\frac{(\dot{u}_{j,\iota} - \bar{r}_{kj} \dot{u}_{k,\iota})}{(1-\bar{r}_{kj}^2)^{1/2}} \right] \right) - \left(\ln \phi(\dot{u}^0_{k,\iota}) + \ln \Phi \left[\frac{(\dot{u}^0_{j,\iota} - \bar{r}_{kj} \dot{u}^0_{k,\iota})}{\sqrt{1-\bar{r}_{kj}^2}} \right] \right).
$$

By using Taylor expansion and triangular inequality, the first term is

$$\left|\ln\Psi(\dot{u}_{k,\iota},\dot{u}_{j,k};\bar{r}_{kj})-\ln\Psi(\dot{u}^0_{k,\iota},\dot{u}^0_{j,\iota};\bar{r}_{kj})\right|\leq\left|\left|\frac{\Psi'_\rho(\dot{u}^0_{k,\iota},\dot{u}^0_{j,\iota};\bar{r}_{kj})}{\Psi(\dot{u}^0_{k,\iota},\dot{u}^0_{j,\iota};\bar{r}_{kj})}\right|\right|\cdot||\widehat{\rho}-\rho||, \quad (5.4)$$

where $\rho=\begin{pmatrix}\rho_k & \rho_j\end{pmatrix}^\top$, $\widehat{\rho}=\begin{pmatrix}\widehat{\rho}_k & \widehat{\rho}_j\end{pmatrix}^\top$, and

$$\begin{aligned}\Psi'_\rho(\dot{u}^0_{k,\iota},\dot{u}^0_{j,\iota};\bar{r}_{kj}) &= \phi(\dot{u}^0_{k,\iota})\Phi\left(\frac{\dot{u}^0_{j,\iota}-\bar{r}_{kj}\dot{u}^0_{k,\iota}}{\sqrt{1-\bar{r}^2_{kj}}}\right)R_{111N} \\ &\quad -\bar{r}_{kj}\phi(\dot{u}^0_{j,\iota})\Phi\left(\frac{\dot{u}^0_{k,\iota}-\bar{r}_{kj}\dot{u}^0_{j,\iota}}{\sqrt{1-\bar{r}^2_{kj}}}\right)R_{112N}\end{aligned} \quad (5.5)$$

with $R_{111N}=\sum_{l=1}^N\mathcal{A}_{k,nl}u_{k,lt}$ and $R_{112N}=\sum_{l=1}^N\mathcal{A}_{j,nl}u_{j,lt}$ where $\mathcal{A}_m=\frac{\partial A^{-1}_{mm}}{\partial\rho_m}$ for $m=k$ or j, $u_{m,lt}=(y_{m,lt}-x_{m,lt}\bar{\beta}_m)/\bar{\sigma}_m$ and $l\in N$ and $t\in T$. For the sake of notational simplicity, let $\frac{\Psi'_\rho(\cdot)}{\Psi(\cdot)}=A_{1N}$. The convergence of (5.4) $\to_P 0$ can be achieved by showing that $A_{1N}=O_p(1)$ because $||\widehat{\rho}-\rho||=O_P((NT)^{-1/2})$ as a result of Lemma 2

$$E(A^2_{1N}) \quad\leq\quad \frac{E\left(\Psi'_\rho(\cdot)\right)^2}{\delta}=O(1),$$

under Assumptions 2.1 to 2.3 and the additional condition of $\inf\limits_{u_{k,\iota},u_{j,\iota}\in\mathbb{R}^2}\psi_\iota=\delta_1$, where δ represents an arbitrary small value. The uniform Lipschitz continuity of (5.4) over Θ_k, can be established as follows. By letting $\Psi(\cdot)=\Psi(\dot{u}_{k,\iota},\dot{u}_{j,\iota};\bar{r}_{kj})$ and $\Psi^0(\cdot)=\Psi(\dot{u}^0_{k,\iota},\dot{u}^0_{j,\iota};\bar{r}_{kj})$, we have

$$\begin{aligned}&\sup_{||\bar{\theta}_k-\breve{\theta}_k||\leq\delta}|\ln\Psi(\bar{\theta}_k)-\ln\Psi^0(\bar{\theta}_k)-\{\ln\Psi(\breve{\theta}_k)-\ln\Psi^0(\breve{\theta}_k)\}| \\ &\leq\sup_{||\bar{\theta}_k-\breve{\theta}_k||\leq\delta}\left|\left|\{\ln\Psi(\tilde{\theta}_k)\}'-\{\ln\Psi^0(\tilde{\theta}_k)\}'\right|\right|\cdot||\bar{\theta}_k-\breve{\theta}_k|| \\ &=o_P(1),\end{aligned} \quad (5.6)$$

where $\breve{\theta}_k\in\Theta_k$ lies on an δ-neighbourhood of $\bar{\theta}_k$ such that $||\bar{\theta}_k-\breve{\theta}_k||\to 0$ as $\delta\to 0$, $\tilde{\theta}_k$ lies on the line segment of $\{\lambda\bar{\theta}_k+(1-\lambda)\breve{\theta}_k:\lambda\in(0,1)\}$, and $\{\ln\Psi(\cdot)\}'$ and $\{\ln\Psi^0(\cdot)\}'$ denote the gradients of $\ln\Psi(\cdot)$ and $\ln\Psi^0(\cdot)$ with respect to $\tilde{\theta}_k$, respectively. (5.6) can be established by showing

$$\left|\left|\{\ln\Psi(\tilde{\theta}_k)\}'-\{\ln\Psi^0(\tilde{\theta}_k)\}'\right|\right| \quad=\quad O_P(1). \quad (5.7)$$

By using Taylor expansion and triangular inequality, we have

$$\left|\left|\{\ln \Psi(\tilde{\theta}_k)\}' - \{\ln \Psi^0(\tilde{\theta}_k)\}'\right|\right| \leq \left|\left|\frac{\partial^2 \ln \Psi^0(\tilde{\theta}_k)}{\partial\tilde{\theta}_k \partial\rho}\right|\right| \cdot ||\widehat{\rho} - \rho||,$$

where

$$\frac{\partial^2 \ln \Psi^0(\tilde{\theta}_k)}{\partial\tilde{\theta}_k \partial\rho} = \frac{\Psi^{0\prime\prime}_{\rho,\tilde{\theta}_k}(\cdot)\Psi^0(\cdot) - \Psi^{0\prime}_{\rho}(\cdot)\Psi^{0\prime}_{\tilde{\theta}_k}(\cdot)}{\{\Psi^0(\cdot)\}^2}$$

with

$$\begin{aligned}\Psi^{0\prime}_{\tilde{\theta}_k} &= \frac{\partial \Psi^0_{kj}(\tilde{\theta}_k)}{\partial\tilde{\theta}_k} \\ &= -\frac{1}{\tilde{\sigma}_k}\dot{x}^0_{k,\iota}\phi^0_{k,\iota}\Phi^0_{jk,\iota} - \frac{\dot{u}^0_{k,\iota} - \tilde{r}_{kj}\dot{u}^0_{j,\iota}}{2\tilde{\sigma}^2_k\tilde{\sigma}^2_j(1-\tilde{r}^2_{kj})}\{\tilde{\sigma}^2_j(1-\tilde{r}^2_{kj}) - 2\tilde{\sigma}^2_k\tilde{\sigma}^2_j\tilde{r}_{kj}\}\psi^0_{kj,\iota}\end{aligned}$$

by letting $\phi^0_{k,\iota} = \phi(\dot{u}^0_{k,\iota})$, $\Phi^0_{jk,\iota} = \Phi\left(\frac{\dot{u}^0_{j,\iota} - \tilde{r}_{kj}\dot{u}^0_{k,\iota}}{\sqrt{1-\tilde{r}^2_{kj}}}\right)$ and where $\psi^0_{kj,\iota}$ denotes the corresponding probability density function (pdf) of $\Psi(\dot{u}^0_{k,\iota}, \dot{u}^0_{j,\iota}; \tilde{r}_{kj})$,

$$\begin{aligned}&\Psi^{0\prime\prime}_{\rho,\tilde{\theta}_k}(\cdot) = \frac{1}{\tilde{\sigma}_k}\dot{x}^0_{k,\iota}\dot{u}^0_{k,\iota}\phi^0_{k,\iota}\Phi^0_{jk,\iota}R_{111N} + \tilde{r}_{kj}\dot{x}^0_{k,\iota}\phi^0_{k,\iota}\phi^0_{jk,\iota}R_{111N} \\ &-\frac{1}{\tilde{\sigma}_k}\phi^0_{k,\iota}\Phi^0_{jk,\iota}\sum_{l=1}^{N}\mathcal{A}_{k,nl}x_{k,lt} + \tilde{r}_{kj}\dot{x}^0_{k,\iota}\phi^0_{j,\iota}\phi^0_{kj,\iota}R_{112N} + \frac{(\dot{u}^0_{k,\iota})^2 - 1}{2\tilde{\sigma}^2_k}\phi^0_{k,\iota}\Phi^0_{jk,\iota}R_{111N} \\ &+\frac{1}{2\tilde{\sigma}_k}\tilde{r}_{kj}\dot{u}^0_{k,\iota}\phi^0_{k,\iota}\phi^0_{jk,\iota}R_{111N} - \frac{1}{2\tilde{\sigma}^2_k}\phi^0_{k,\iota}\Phi^0_{jk,\iota}\sum_{l=1}^{N}\mathcal{A}_{k,nl}u_{k,nt} + \frac{1}{2\tilde{\sigma}_k}\tilde{r}_{kj}\dot{u}^0_{k,\iota}\phi^0_{j,\iota}\phi^0_{kj,\iota}R_{112N} \\ &+\frac{\tilde{r}_{kj}(\dot{u}^0_{j,\iota} - \tilde{r}_{kj}\dot{u}^0_{k,\iota})}{(1-\tilde{r}^2_{kj})}\phi^0_{k,\iota}\phi^0_{jk,\iota}R_{111N} - \dot{u}^0_{k,\iota}\phi^0_{k,\iota}\phi^0_{jk}R_{111N} \\ &-\frac{\tilde{r}^2_{kj}(\dot{u}^0_{k,\iota} - \tilde{r}_{kj}\dot{u}^0_{j,\iota})}{(1-\tilde{r}^2_{kj})}\phi^0_{j,\iota}\phi^0_{kj,\iota}R_{112N} + \tilde{r}_{kj}\dot{u}^0_{j,\iota}\phi^0_{j,\iota}\phi^0_{kj,\iota}R_{112N} - \phi^0_{k,\iota}\Phi^0_{kj,\iota}R_{112N},\end{aligned}$$

there $\Psi^{0\prime}_{\rho}(\cdot)$ is as defined in (5.5). (5.7) holds because $E\left|\left|\frac{\partial^2 \ln \Psi^0(\tilde{\theta}_k)}{\partial\tilde{\theta}_k\partial\rho}\right|\right| = O(1)$ under Assumptions 2.1 to 3.1 and the condition of $\inf_{u_{k,\iota},u_{j,\iota}\in\mathbb{R}^2}\psi_\iota = \delta_1$.

The second term can be shown to be $o_P(1)$ uniformly over Θ_k by using similar arguments to those above. Let us represent the second term as follows

$$\begin{aligned}|\ln\psi_{kj,\iota} - \ln\psi^0_{kj,\iota}| &\leq \left|\left|\frac{\partial \ln\psi^{0\prime}_{kj,\iota,\rho}}{\partial\rho}\right|\right| \cdot ||\widehat{\rho} - \rho|| \\ &= o_P(1), \end{aligned} \tag{5.8}$$

where $\psi(\dot{u}_{k,\iota}, \dot{u}_{j,\iota}; \bar{r}_{kj}) = \psi_{kj,\iota}$ and $\psi(\dot{u}^0_{k,\iota}, \dot{u}^0_{j,\iota}; \bar{r}_{kj}) = \psi^0_{kj,\iota}$. By Taylor expansion and triangular inequality, where

$$\frac{\partial \ln \psi^{0\prime}_{kj,\iota,\rho}}{\partial \rho} = -\frac{\dot{u}^0_{k,\iota} - \bar{r}_{kj}\dot{u}^0_{j,\iota}}{(1-\bar{r}^2_{kj})}\{R_{111N} - \bar{r}_{kj}R_{112N}\}.$$

Here, (5.8) holds because $E\left|\left|\frac{\partial \psi^{0\prime}_{kj,\rho}}{\partial \rho}\right|\right|^2 = O(1)$ under Assumptions 2.1 to 2.3. The uniform Lipschitz continuity over Θ_k can be shown by

$$\left|\left|\{\ln \psi_{kj}(\tilde{\theta}_k)\}' - \{\ln \psi^0_{kj}(\tilde{\theta}_k)\}'\right|\right| = O_P(1),$$

where$\{\ln \psi_{kj}\}'$ and $\{\ln \psi^0_{kj}\}'$ denote the gradients of $\ln \psi$ and $\ln \psi^0$ with respect to $\tilde{\theta}_k$, respectively. By using the same arguments to the above, we have

$$\begin{aligned}\left|\left|\{\ln \psi_{kj}(\tilde{\theta}_k)\}' - \{\ln \psi^0_{kj}(\tilde{\theta}_k)\}'\right|\right| &\leq \left|\left|\frac{\partial^2 \ln \psi^0_{kj}(\tilde{\theta}_k)}{\partial \tilde{\theta}_k \partial \rho}\right|\right| \cdot ||\hat{\rho} - \rho|| \\ &= o_P(1), \end{aligned} \tag{5.9}$$

where

$$\begin{aligned}\frac{\partial^2 \ln \psi^0_{kj}(\tilde{\theta}_k)}{\partial \tilde{\theta}_k \partial \rho} &= \frac{\dot{x}^0_{k,\iota}/\tilde{\sigma}_k\{R_{111N} - \tilde{r}_{kj}R_{112N}\} + (\dot{u}^0_{k,\iota} - \tilde{r}_{kj}\dot{u}^0_{j,\iota})\sum_{l=1}^N \mathcal{A}_{k,nl}x_{k,lt}/\tilde{\sigma}_k}{(1-\tilde{r}^2_{kj})} \\ &+ \frac{\dot{u}^0_{k,\iota}\{R_{111N} - \bar{r}_{kj}R_{112N}\} + (\dot{u}^0_{k,\iota} - \tilde{r}_{kj}\dot{u}^0_{j,\iota})\sum_{l=1}^N \mathcal{A}_{k,nl}u_{k,lt}}{2(1-\tilde{r}^2_{kj})\tilde{\sigma}^2_k} \\ &+ \frac{\dot{u}^0_{j,\iota}(1-\tilde{r}^2_{kj}) - 2\tilde{r}_{kj}(\dot{u}^0_{k,\iota} - \tilde{r}_{kj}\dot{u}^0_{j,\iota})}{(1-\tilde{r}^2_{kj})^2}R_{111N} \\ &+ \frac{(\dot{u}^0_{k,\iota} - 2\tilde{r}_{kj}\dot{u}^0_{j,\iota})(1-\tilde{r}^2_{kj}) + 2\tilde{r}^2_{kj}(\dot{u}^0_{k,\iota} - \tilde{r}_{kj}\dot{u}^0_{j,\iota})}{(1-\tilde{r}^2_{kj})^2}R_{112N}.\end{aligned}$$

(5.9) holds because $E\left|\left|\frac{\partial^2 \ln \psi^0_{kj}(\tilde{\theta}_k)}{\partial \tilde{\theta}_k \partial \rho}\right|\right| = O(1)$ under Assumptions 2.1 to 3.1.

Finally, the last term is

$$\ln \phi(\dot{u}_{k,\iota}) - \ln \phi(\dot{u}^0_{k,\iota}) + \{\ln \Phi_{jk} - \ln \Phi^0_{jk}\} \tag{5.10}$$

by letting $\Phi_{jk} = \Phi\left(\frac{\dot{u}_{j,\iota} - \bar{r}_{kj}\dot{u}_{k,\iota}}{\sqrt{1-\bar{r}^2_{kj}}}\right)$. The convergence of (5.10) $\rightarrow_P 0$ can be established by using similar arguments to those above. By using Taylor expansion and triangular inequality, the first term of (5.10) is

$$|\ln \phi(\dot{u}_{k,\iota}) - \ln \phi(\dot{u}^0_{k,\iota})| \leq \left|\frac{\phi'_\rho(\dot{u}^0_{k,\iota})}{\phi(\dot{u}^0_{k,\iota})}\right| \cdot |\widehat{\rho}_k - \rho_k|,$$

where $\phi'_\rho(\dot{u}^0_{k,\iota}) = -\dot{u}^0_{k,\iota}\phi(\dot{u}^0_{k,\iota})R_{111N}$. The second term of (5.10) is

$$|\ln\Phi_{jk} - \ln\Phi^0_{jk}| \leq \left|\left|\frac{\Phi^{0\prime}_{jk,\rho}}{\Phi^0_{jk}}\right|\right| \cdot ||\widehat{\rho} - \rho||$$

according to the same arguments, where

$$\Phi^{0\prime}_{jk,\rho} \;\; = \;\; \phi^0_{jk,\iota}\left(R_{112N} - \bar{r}_{kj}R_{111N}\right). \tag{5.11}$$

By letting $A_{3N} = \frac{\phi'_\rho(\dot{u}^0_{k,\iota})}{\phi(\dot{u}^0_{k,\iota})}$ and $A_{4N} = \frac{\Phi^{0\prime}_{jk,\rho}}{\Phi^0_{jk}}$, we need to show that $A_{3N} = O_P(1)$ and $A_{4N} = O_P(1)$. These results are easily obtained because $E(A^2_{3N}) = O(1)$ under Assumptions 2.1 to 2.3 and $E(A^2_{4N}) = O(1)$ with the additional condition, $\inf_{u_{k,\iota}\in\mathbb{R}}\phi(u_{k,\iota}) = \delta_2$. The uniform Lipschitz continuity of (5.10) over Θ_k can then be shown as follows. By denoting the gradients of $\ln\phi$ and $\ln\phi^0$ with respect to $\tilde{\theta}_k$ as $\{\ln\phi\}'$ and $\{\ln\phi^0\}'$, respectively, and by using Taylor expansion and triangular inequality, we have

$$\begin{aligned} \left|\left|\{\ln\phi(\tilde{\theta}_k)\}' - \{\ln\phi^0(\tilde{\theta}_k)\}'\right|\right| &\leq \left|\left|\frac{\partial^2\ln\phi^0(\tilde{\theta}_k)}{\partial\tilde{\theta}_k\partial\rho}\right|\right| \cdot |\widehat{\rho}_k - \rho_k| \\ &= o_P(1), \end{aligned}$$

where

$$\begin{aligned} \frac{\partial^2\ln\phi^0(\tilde{\theta}_k)}{\partial\tilde{\theta}_k\partial\rho} = & \frac{1}{\tilde{\sigma}_k}\left(\dot{x}^0_{k,\iota}R_{111N} + \dot{u}^0_{k,\iota}\sum_{l=1}^{N}\mathcal{A}_{k,nl}x_{k,lt}\right) \\ & + \frac{1}{2\tilde{\sigma}^2_k}\left(\dot{u}^0_{k,\iota}R_{111N} + \dot{u}^0_{k,\iota}\sum_{l=1}^{N}\mathcal{A}_{k,nl}u_{k,lt}\right) \end{aligned}$$

and

$$E\left|\left|\frac{\partial^2\ln\phi^0(\tilde{\theta}_k)}{\partial\tilde{\theta}_k\partial\rho}\right|\right|^2 = O(1)$$

under Assumptions 2.1 to 3.1. Hence, $\left|\left|\{\ln\phi(\tilde{\theta}_k)\}' - \{\ln\phi^0(\tilde{\theta}_k)\}'\right|\right| = O_P(1)$. Now let us consider the stochastic Lipschitz continuity of the second term in (5.10). By denoting $\{\ln\Phi_{jk}(\tilde{\theta}_k\}'$ and $\{\ln\Phi^0_{jk}(\tilde{\theta}_k)\}'$ as the gradients of $\ln\Phi_{jk}$ and $\ln\Phi^0_{jk}$ with respect to $\tilde{\theta}_k$, respectively, and using the same arguments as the above, we have

$$\begin{aligned} \left|\left|\{\ln\Phi_{jk}(\tilde{\theta}_k)\}' - \{\ln\Phi^0_{jk}(\tilde{\theta}_k)\}'\right|\right| &\leq \left|\left|\frac{\partial^2\ln\Phi^0_{jk}(\tilde{\theta}_k)}{\partial\tilde{\theta}_k\partial\rho}\right|\right| \cdot ||\widehat{\rho} - \rho|| \\ &= o_P(1), \end{aligned}$$

where

$$\frac{\partial^2 \ln \Phi^0_{jk}(\tilde{\theta}_k)}{\partial \tilde{\theta}_k \partial \rho} = \frac{\Phi^{0\prime\prime}_{jk,\rho,\tilde{\theta}_k}(\tilde{\theta}_k)\Phi^0_{jk}(\tilde{\theta}_k) - \Phi^{0\prime}_{jk,\rho}(\tilde{\theta}_k)\Phi^{0\prime}_{jk,\tilde{\theta}_k}(\tilde{\theta}_k)}{\{\Phi^0_{jk}(\tilde{\theta}_k)\}^2},$$

$$\begin{aligned}
\Phi^{0\prime\prime}_{jk,\rho,\tilde{\theta}_k}(\tilde{\theta}_k) &= \frac{\partial^2 \Phi^0_{jk,\iota}(\tilde{\theta}_k)}{\partial \rho \partial \tilde{\theta}_k} \\
&= \left\{ -\frac{1}{(1-\tilde{r}^2_{kj})}(\dot{u}^0_{j,\iota} - \tilde{r}_{kj}\dot{u}^0_{k,\iota})\tilde{r}_{kj}\frac{\dot{x}^0_{k,\iota}}{\tilde{\sigma}_k}\phi^0_{jk,\iota} \right\}(R_{112N} - \tilde{r}_{kj}R_{111N}) + \tilde{r}_{kj}\sum_{l=1}^{N}\mathcal{A}_{k,nl}\frac{x_{k,lt}}{\tilde{\sigma}_k}\phi^0_{jk,\iota} \\
&\quad + \left\{ \frac{1}{(1-\tilde{r}^2_{kj})}(\dot{u}^0_{j,\iota} - \tilde{r}_{kj}\dot{u}^0_{k,\iota})\tilde{r}_{kj}\frac{\dot{u}^0_{k,\iota}}{2\tilde{\sigma}^2_k}\phi^0_{jk,\iota} \right\}(R_{112N} - \tilde{r}_{kj}R_{111N}) + \tilde{r}_{kj}\sum_{l=1}^{N}\mathcal{A}_{k,nl}\frac{u_{k,lt}}{2\tilde{\sigma}^2_k}\phi^0_{jk,\iota} \\
&\quad \left\{ \left(\frac{\tilde{r}_{kj}\{(\dot{u}^0_{j,\iota} - \tilde{r}_{kj}\dot{u}^0_{k,\iota})^2 - (1-\tilde{r}_{kj})\} - 2\dot{u}^0_{k,\iota}(\dot{u}^0_{j,\iota} - \tilde{r}_{kj}\dot{u}^0_{k,\iota})(1-\tilde{r}^2_{kj})}{(1-\tilde{r}^2_{kj})^2} \right) \right. \\
&\quad \left. \times \phi^0_{jk,\iota}(R_{112N} - \tilde{r}_{kj}R_{111N}) \right\} - \phi^0_{jk,\iota}R_{111N}
\end{aligned}$$

and

$$\begin{aligned}
\Phi^{0\prime}_{jk,\tilde{\theta}_k}(\tilde{\theta}_k) &= \frac{\partial \Phi^0_{jk}(\tilde{\theta}_k)}{\partial \tilde{\theta}_k} \\
&= \phi^0_{jk,\iota}\left(\tilde{r}_{kj}\frac{\dot{x}^0_{k,\iota}}{\tilde{\sigma}_k} + \tilde{r}_{kj}\frac{\dot{u}^0_{k,\iota}}{2\tilde{\sigma}^2_k} - \dot{u}^0_{k,\iota} + \frac{\tilde{r}_{kj}(\dot{u}^0_{j,\iota} - \tilde{r}_{kj}\dot{u}^0_{k,\iota})}{(1-\tilde{r}^2_{kj})} \right).
\end{aligned}$$

Here $\Phi^{0\prime}_{\rho}(\tilde{\theta}_k)$ is as defined in (5.11). Hence,

$$\left|\left|\{\ln \Phi_{jk}(\tilde{\theta}_k)\}' - \{\ln \Phi^0_{jk}(\tilde{\theta}_k)\}'\right|\right| = O_P(1)$$

under Assumptions 2.1 to 3.1 and the condition $\inf_{u_{k,it}\in\mathbb{R}} \phi(u_{k,it}) = \delta_2$.

The final step of this proof is to establish the unique identification condition for θ_k. Let us establish the identification condition in Theorem 1 by presenting a counter argument. By Jensen's inequality, we have

$$E\mathcal{L}^0_{k,j}(\bar{\theta}_k) - E\mathcal{L}^0_{k,j}(\theta_k) \leq 0.$$

This equality holds when $\bar{\theta}_k = \theta_k$. Hence θ_k is not uniquely identified when there is a sequence such that $\bar{\theta}_{k_N} \in D_\delta(\theta^*_k)$, where $D_\delta(\theta^*_k)$ is the δ-neighbourhood of θ^*_k, converges to $\theta^*_k \in \bar{D}_\delta(\theta_k)\cap\Theta_k$, where $\bar{D}_\delta(\theta_k)$ is the complement of the δ-neighbourhood of θ_k, and $\lim_{N\to\infty}\mathcal{L}^0_{k,j}(\theta^*_k) \to \lim_{N\to\infty}\mathcal{L}^0_{k,j}(\theta_k)$. Hence the identification condition requires

$$\limsup_{N\to\infty}\left(\max_{\bar{\theta}_k\in\bar{D}_\delta(\theta_k)\cap\Theta_k} E\mathcal{L}^0_{k,j}(\bar{\theta}_k) \right) \neq E\mathcal{L}^0_{k,j}(\theta_k)$$

for any $\bar{\theta}_k$ for all $k = 1, \cdots, K-1$. □

8. 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, Heidelberg.

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, Finance, and the Firm Discussion Paper, 519.

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. 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.

Chambers, R. G., & Just, R. E. (1989). Estimating multioutput technologies. American Journal of Agricultural Economics, 71(4), 980-995.

Chakir, R., & Gallo, J.L. (2013). Predicting land use allocation in France: A spatial panel data analysis. Ecological Economics, 92(0), 114-125.

Cliff, A. D. (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 https://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 https://www.theccc.org.uk/publication/land-use-policies-for-a-net-zero-uk/.

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 modeling for spatial agro-environmental policy analysis. American Journal of Agricultural Economics, 93(4), 1168-1188.

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.

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.

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.

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.

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.

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

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.

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.

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.

Table 1: Descriptive Statistics for Land-Uses (in ha)

	1976	1979	1981	1988	2000	2004
Temp. grassland	35.929	30.133 ***	29.645	25.363 ***	23.428 **	20.753 ***
Perm. grassland	96.705	96.623	94.938	91.338	81.641 ***	88.755 ***
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 ***	99.035 ***
Total arg. land†	272.910	271.415	274.330	269.853*	245.715***	248.451

† Total agricultural land is computed as the summation of temporary, permanent, rough grassland and oilseed rape. ***, ** and * 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.

Table 2: 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
$setaside$	
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. Similarly, $scoarse$, $fine$ and $smedium$ are summed to one and therefore $smedium$ is omitted.

Table 3: Out-of-Sample Root Mean Squared Errors (RMSE) Land-Use Share Prediction: Shares of Temporary Grassland for 2010

	Partial Linear	Linear
TL without RE-SED:		
[a] RMSE	5.016	5.469
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 5$)		
[a] RMSE without added residual	4.140	4.344
[b] RMSE with added residual	3.791	3.795
[c] SAR parameter, ρ_k	0.452	0.480
[d] LM_{BP} statistic	4.4660e+03 (0.000)	4.4666e+03 (0.000)
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 2$)		
[a] RMSE without added residual	4.225	4.470
[b] RMSE with added residual	3.910	3.948
[c] SAR parameter, ρ_k	0.215	0.224
[d] LM_{BP} statistic	5.1330e+03 (0.000)	5.4210e+03 (0.000)
TL with RE-SED & W_k^{SOI}		
[a] RMSE without added residual	4.158	4.470
[b] RMSE with added residual	3.799	3.948
[c] SAR parameter, ρ_k	0.450	0.478
[d] LM_{BP} statistic	4.4434e+03 (0.000)	4.4466e+03 (0.000)

Table 4: Out-of-Sample RMSE Land-Use Share Prediction: Shares of Permanent Grassland for 2010

	Partial Linear	Linear
TL without RE-SED:		
[a] RMSE	12.817	14.874
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 5$)		
[a] RMSE with added residual	7.141	7.278
[b] RMSE without added residual	11.452	12.994
[c] SAR parameter, ρ_k	0.498	0.534
[d] LM_{BP} statistic	1.0844e+03 (0.000)	1.0649e+03 (0.000)
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 2$)		
[a] RMSE with added residual	7.034	7.097
[b] RMSE without added residual	10.948	12.595
[c] SAR parameter, ρ_k	0.274	0.292
[d] LM_{BP} statistic	1.1956e+03 (0.000)	1.2041e+03 (0.000)
TL with RE-SED & W_k^{SOI}		
[a] RMSE with added residual	7.123	7.248
[b] RMSE without added residual	11.496	10.296
[c] SAR parameter, ρ_k	0.506	0.539
[d] LM_{BP} statistic	1.0858e+03 (0.000)	1.0658e+03 (0.000)

Table 5: Out-of-Sample RMSE Land-Use Share Prediction: Shares of Rough Grazing for 2010

	Partial Linear	Linear
TL without RE-SED:		
[a] RMSE	11.874	12.009
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 5$)		
[a] RMSE with added residual	11.853	11.818
[b] RMSE without added residual	12.013	12.121
[c] SAR parameter, ρ_k	0.114	0.157
[d] LM_{BP} statistic	4.4097e+03 (0.000)	4.7439e+03 (0.000)
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 2$)		
[a] RMSE with added residual	11.854	11.814
[b] RMSE without added residual	11.995	12.074
[c] SAR parameter, ρ_k	0.061	0.074
[d] LM_{BP} statistic	4.4502e+03 (0.000)	4.4868e+03 (0.000)
TL with RE-SED & W_k^{SOI}		
[a] RMSE with added residual	11.844	11.841
[b] RMSE without added residual	12.043	12.074
[c] SAR parameter, ρ_k	0.113	0.158
[d] LM_{BP} statistic	4.4159e+03 (0.000)	4.7639e+03 (0.000)

Table 6: Out-of-Sample RMSE Land-Use Share Prediction: Shares of Arable for 2010

	Partial Linear	Linear
TL without RE-SED:		
[a] RMSE	12.995	15.693
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 5$)		
[a] RMSE with added residual	7.936	8.025
[b] RMSE without added residual	13.596	16.156
[c] SAR parameter, ρ_k	0.555	0.602
[d] LM_{BP} statistic	1.2320e+03 (0.000)	1.1872e+03 (0.000)
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 2$)		
[a] RMSE with added residual	8.132	8.147
[b] RMSE without added residual	13.072	15.378
[c] SAR parameter, ρ_k	0.323	0.360
[d] LM_{BP} statistic	1.3265e+03 (0.000)	1.3071e+03 (0.000)
TL with RE-SED & W_k^{SOI}		
[a] RMSE with added residual	7.955	8.052
[b]RMSE without added residual	13.651	16.218
[c] SAR parameter, ρ_k	0.563	0.606
[d] LM_{BP} statistic	1.2227e+03 (0.000)	1.3071e+03 (0.000)

Table 7: Out-of-Sample RMSE Land-Use Share Prediction: Shares of Permanent Grassland for 2010

	Partial Linear	Linear
TL without RE-SED:		
[a] RMSE	12.817	14.874
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 5$)		
[a] RMSE with added residual	7.141	7.278
[b] RMSE without added residual	11.452	12.994
[c] SAR parameter, ρ_k	0.498	0.534
[d] LM_{BP} statistic	1.0844e+03 (0.000)	1.0649e+03 (0.000)
TL with RE-SED & $W_k^{\kappa NN}$ ($\kappa = 2$)		
[a] RMSE with added residual	7.034	7.097
[b] RMSE without added residual	10.948	12.595
[c] SAR parameter, ρ_k	0.274	0.292
[d] LM_{BP} statistic	1.1956e+03 (0.000)	1.2041e+03 (0.000)
TL with RE-SED & W_k^{SOI}		
[a] RMSE with added residual	7.123	7.248
[b] RMSE without added residual	11.496	10.296
[c] SAR parameter, ρ_k	0.506	0.539
[d] LM_{BP} statistic	1.0858e+03 (0.000)	1.0658e+03 (0.000)

Table 8: Parameter Estimates for the Share of Temporary Grassland

	W_k^{5NN}	W_k^{2NN}	W_k^{SOI}	Without RE-SED
alt0	0.003	0.006	0.001	0.010 **
alt200	0.000	-0.006	-0.006	-0.004
alt200d	-0.488	1.178	0.442	1.168
slopef6	-1.796 *	-0.681	-1.541	-0.577
rain	-0.081	-0.064	-0.076	-0.097
temp	-14.370 **	-16.239 ***	-17.942 *	-23.742 **
ratemp	0.014 **	0.013 **	0.014 **	0.014 **
dist300	0.004	0.004	0.005	0.006 **
speat	-2.064 **	-3.252 ***	-2.002 **	-4.100 ***
sgravel	1.605 *	3.234 ***	1.858 **	4.657 ***
sstoney	1.322 **	1.669 ***	1.227 *	1.777 ***
sfragipan	1.494	2.978 **	1.672	3.808 ***
scoarse	-0.449	-0.763 **	-0.459	-1.203 ***
sfine	0.767 **	0.732 **	0.797 **	0.534 *
smedium	·	·	·	·
sud	0.641 **	0.636 **	0.643 **	1.080 ***
north	0.449	0.509	0.468	0.958 *
mid	·	·	·	·
y_{1976}	4.013 ***	3.909 ***	4.004 ***	2.662 ***
y_{1974}	1.779 ***	1.661 ***	1.737 ***	0.370 ***
y_{1981}	1.584 ***	1.462 ***	1.557 ***	0.171 ***
y_{1988}	-0.024	-0.115	-0.043	-1.415 ***
y_{2000}	0.404 **	0.312 **	0.401 **	0.480 ***
y_{2004}	·	·	·	·
npark	-0.005	-0.003	-0.005	-0.008
esa	-0.003	-0.004	-0.006	-0.005
greenbelt	0.004	0.006	0.005	0.008 *
setaside	-0.084 ***	-0.089 ***	-0.086 ***	-0.167 ***

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1, · signifies the reference variable

Table 9: Parameter Estimates for the Share of Permanent Grassland

	W_k^{5NN}	W_k^{2NN}	W_k^{SOI}	Without RE-SED
alt0	-0.025 ***	-0.013 **	-0.023 ***	0.015
alt200	-0.146	-0.105 **	-0.155	-0.051
alt200d	26.142	20.136	28.443	16.042 *
slopef6	5.077	8.493	3.908	10.943 ***
rain	-0.008	0.077	-0.004	0.269
temp	26.754 ***	23.841 ***	-11.095 ***	81.078
ratemp	0.030 ***	0.021 ***	0.029 ***	0.012
dist300	-0.041	-0.046	-0.036	-0.023 **
speat	-0.960	-1.724	-1.152	1.339
sgravel	1.808 ***	6.354 ***	1.690 ***	10.069 ***
sstoney	1.210	0.447 ***	0.601	-1.553
sfragipan	6.737	8.370 **	5.946	13.665 ***
scoarse	-2.777 ***	-4.092 ***	-2.745 ***	-5.256 ***
sfine	5.762	6.277	5.894	7.463 ***
smedium	·	·	·	·
sud	-0.822	-0.756	-1.027	-2.357 **
north	-0.947	-1.042	-0.705	-2.190
mid	·	·	·	·
y_{1976}	0.901	2.351 ***	1.002	-5.554 ***
y_{1979}	0.681	2.107 ***	0.857	-5.864 ***
y_{1981}	-0.142	1.228 **	-0.073	-6.679 ***
y_{1988}	-1.668 **	-0.335	-1.597 **	-8.282 ***
y_{2000}	-2.300 ***	-2.250 ***	-2.223 ***	-2.932 ***
y_{2004}	·	·	·	·
npark	-0.025	-0.004	-0.029	-0.016
esa	0.068	0.059	0.068	0.015
greenbelt	-0.008 ***	-0.005 ***	-0.007 ***	0.008
setaside	-0.141 ***	-0.061 ***	-0.136 ***	-0.463 ***

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1, · signifies the reference variable

Table 10: Parameter Estimates for the Share of Rough Grazing

	W_k^{5NN}	W_k^{2NN}	W_k^{SOI}	Without RE-SED
alt0	-0.001	-0.002	-0.001	-0.006 **
alt200	-0.008	-0.010	-0.005	-0.020
alt200d	2.237	2.523	1.765	4.695
slopef6	4.820 ***	4.973 ***	4.811 ***	4.328 ***
rain	0.098	0.107	0.092	0.164
temp	-16.808	-9.154	117.054	-1.396
ratemp	-0.010	-0.011	-0.010	-0.009
dist300	0.002	0.002	0.002	0.003
speat	0.303	0.244	0.241	-0.484
sgravel	0.010	-0.066	0.005	-0.233
sstoney	-1.618 **	-1.748 ***	-1.636 **	-1.753 ***
sfragipan	1.224	1.196	1.144	0.809
scoarse	0.557 **	0.583 **	0.551 **	0.574 **
sfine	-0.172	-0.165	-0.166	-0.120
smedium	·	·	·	·
sud	0.396 ***	0.396 ***	0.384 ***	0.647 ***
north	0.179	0.193	0.182	-0.039
mid	·	·	·	·
y_{1976}	-0.455 **	-0.459 **	-0.455 **	-1.901 ***
y_{1979}	-0.432 **	-0.437 **	-0.434 **	-1.713 ***
y_{1981}	-0.461 **	-0.465 **	-0.463 **	-1.816 ***
y_{1988}	-0.470 **	-0.473 ***	-0.472 **	-1.687 ***
y_{2000}	0.151 ***	0.149 ***	0.149 ***	0.092 **
y_{2004}	·	·	·	·
npark	0.021	0.017	0.020	0.028
esa	0.002	0.003	0.003	0.016 **
greenbelt	0.006 **	0.005 **	0.005 **	0.005 **
setaside	-0.016 ***	-0.017 ***	-0.016 ***	-0.037 ***

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1, · signifies the reference variable

Table 11: Parameter Estimates for the Share of Arable

	W_k^{5NN}	W_k^{2NN}	W_k^{SOI}	Without RE-SED
alt0	-0.018	-0.018	-0.015	-0.038 ***
alt200	0.006	0.002	0.001	-0.005
alt200d	-4.846	-3.372	-3.249	-7.541 **
slopef6	-6.476 ***	-9.074 ***	-6.850 ***	-8.837 ***
rain	-0.047	-0.061	-0.083	-0.253 *
temp	5.311	3.434	-0.372	17.177
ratemp	-0.024 **	-0.028 ***	-0.020 **	-0.021 ***
dist300	0.002	0.013	-0.004	0.010
speat	3.214	6.791 **	3.556	6.116
sgravel	-0.227	-6.629 ***	-0.856	-13.005 ***
sstoney	0.769	1.190	0.974	1.410
sfragipan	-8.500 ***	-11.263 ***	-6.957 ***	-16.179 ***
scoarse	1.383	3.399 ***	1.403	5.217 ***
sfine	-7.132 ***	-7.966 ***	-7.458 ***	-8.731 ***
smedium	·	·	·	·
sud	2.277 **	1.670 **	2.404 **	3.045 ***
north	-1.461	0.003	-1.851 *	0.117
mid	·	·	·	·
y_{1976}	0.634	-0.375	0.456	4.808 ***
y_{1979}	2.219 ***	1.126 ***	2.060 ***	6.482 ***
y_{1981}	2.405 ***	1.308 ***	2.271 ***	6.754 ***
y_{1988}	3.028 ***	1.941 ***	2.912 ***	7.203 ***
y_{2000}	3.806 ***	4.005 ***	3.862 ***	3.932 ***
y_{2004}	·	·	·	·
npark	0.010	0.010	0.009	0.003
esa	-0.013	-0.004	-0.008	0.041 ***
greenbelt	0.020	0.021	0.013	-0.012
setaside	-0.059 ***	-0.146 ***	-0.069 ***	0.240 ***

*** p-value < 0.01, ** p-value < 0.05, * p-value < 0.1, · signifies the reference variable