Spatial Economic Analysis

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rsea20

A spatial multinomial logit model for analysing urban expansion

Tamás Krisztin, Philipp Piribauer & Michael Wögerer

SPATIAL ECONOMIC ANALYSIS
https://doi.org/10.1080/17421772.2021.1933579

A spatial multinomial logit model for analysing urban expansion

Tamás Krisztin [a], Philipp Piribauer [b] and Michael Wögerer [c]

ABSTRACT
The paper proposes a Bayesian multinomial logit model to analyse spatial patterns of urban expansion. The specification assumes that the log-odds of each class follow a spatial autoregressive process. Using recent advances in Bayesian computing, our model allows for a computationally efficient treatment of the spatial multinomial logit model. This allows us to assess spillovers between regions and across land-use classes. In a series of Monte Carlo studies, we benchmark our model against other competing specifications. The paper also showcases the performance of the proposed specification using European regional data. Our results indicate that spatial dependence plays a key role in the land-sealing process of cropland and grassland. Moreover, we uncover land-sealing spillovers across multiple classes of arable land.

KEYWORDS
urban expansion, land-use change, spatial multinomial logit model, European regions, ,

JEL C11, C21, C25, O13, R14
HISTORY Received 30 July 2020; in revised form 3 May 2021

INTRODUCTION

Increased urbanization and expansion of cities as a direct result of economic and population growth, coupled with intensifying climate change, poses a key challenge for policymakers (IPBES, 2019). The location choice of new urban developments is of particular importance because land is a finite resource. Expanding artificial surfaces is both expensive and time-consuming to reverse, resulting in long-term impacts on land use and land cover. The conversion of natural habitats to artificial surfaces thus has a direct and potentially irreversible impact on biodiversity (Leclère et al., 2020). On the other hand, if arable land is built up, global food security is threatened and urban expansion might spillover to other types of land use.

Conversion of land to urban surfaces is a decision usually taken by the landowners, which are either regional governments or private land-holders. In an economic framework, this decision is understood as a trade-off between the relative profitabilities of land uses and respective conversion costs (Miller & Plantinga, 1999). Potential profits from land ownership are typically assessed

CONTACT
[a] **(Corresponding author)** krisztin@iiasa.ac.at
International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.
[b] philipp.piribauer@wifo.ac.at
Austrian Institute of Economic Research (WIFO), Vienna, Austria.
[c] wogerer@iiasa.ac.at
International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.
Supplemental data for this article can be accessed https://doi.org/10.1080/17421772.2021.1933579.

using various proxies for land rents (Chakir & Lungarska, 2017), while conversion costs rely on the quality of land and national regulations restricting land transformation. Land-use change models targeting aggregate administrative levels focus on capturing the outcomes of regional policies (Ay et al., 2017). This is of special importance within the European Union (EU), where regional policies (such as the Structural Funds) are aimed at this level (Alexiadis et al., 2013).

Within a regional econometric framework, the land-use expansion decision can be modelled as a random choice, with the multinomial logit model representing a popular option (Chakir, 2009; Lubowski et al., 2008). The particular advantage is that a joint modelling of land (or soil) sealing processes can take into account spillovers across land-use classes. When dealing with compositional (shares) data for land use, the multinomial logit random choice model can either be estimated directly from the multinomial logit form (Li et al., 2013) or from its log-linearized form (Chakir & Lungarska, 2017). While the log-linearized version of the model represents a popular choice due to its ease of transformation, it suffers from the usual problems of log-transformation, namely that frequently land-use shares are zero and accommodating these observations inherently biases the estimates.

Spatial dependence, from both unobserved spatially varying variables as well as contingent on the choice of neighbouring regions, is well documented in the land-use choice literature (Chakir & Le Gallo, 2013; Chakir & Parent, 2009; Li et al., 2013). In a regional econometrics context, a wide number of studies stress the inherent importance of spatial spillovers (LeSage & Pace, 2009). When estimating models for land-use change on a small-scale level, the problem of spatial dependence becomes even more central. Specifically, neglecting to account for spatial autocorrelation may result in severely biased and inefficient estimates and erroneous policy conclusions. However, spatial dependence in multinomial logit frameworks so far has been neglected by the spatial econometric literature, with the exception of generalized method of moments (GMM)-based approaches (Carrión-Flores et al., 2018; Klier & McMillen, 2008) or simulated log-likelihood estimators (Bhat & Guo, 2004; Bhat & Sener, 2009).

Within this paper our contribution to the existing literature is twofold. First and foremost, we present a novel Bayesian approach for capturing spatial dependence among land-use changes using a multinomial logit framework. By combining the spatial autoregressive (SAR) and multinomial logit frameworks, our specification can account for cross-regional and cross-land-use class spillovers. The estimation approach builds on recent advances in Bayesian modelling of logit-type specifications (Krisztin & Piribauer, 2020) and employs latent Pólya–Gamma-distributed variables. A particular virtue of this approach is the easy implementation in Bayesian Gibbs sampling algorithms.[1] We demonstrate the virtues of our approach in a series of Monte Carlo studies.

Our second contribution is a novel examination of land-use change processes on a regional pan-European level. For this we rely on an extensive dataset of land-use changes to assess the share of land sealing from cropland, grassland, forest and other fallow land. Additionally, we explore land sealing stemming from urban, artificial and settlement area expansion. Our framework allows us to shed a light on the small-scale spatial dynamics of land-sealing processes in European regions.

The remainder of the paper is structured as follows. The next section outlines the theoretical model of urban expansion, as well as its multinomial logit variant. The subsequent section focuses on the estimation framework. Afterwards we present the Monte Carlo benchmarks of the proposed econometric estimation approach. Finally, we discuss the results for urban expansion in Europe.

A SPATIAL AUTOREGRESSIVE MULTINOMIAL LOGIT MODEL

In this paper, we estimate an econometric model that aims at explaining the choice of land buyers (both public and private) for the purpose of converting it to urban, artificial and settlement surfaces in N regions. In a given region i (with $i = 1, \ldots, N$), land buyers may acquire land from J different land uses. In our case these are cropland, grassland, forest and other natural land. Within a region the buyers are assumed to be price taker and their choices are assumed to be homogeneous and risk neutral.

In an economic sense this constitutes a profit-maximization problem of land buyers (Lubowski et al., 2008; Miller & Plantinga, 1999), which is directly dependent on the associated profits and costs of the converted land. In addition, to account for the expected net present value of rents from urban land use and the respective conversion costs, land buyers also face the opportunity costs of alternatives usages.

Such frameworks have been adopted, among others, by Lubowski et al. (2008), Chakir and Parent (2009) and Li et al. (2013). For the estimation of parameters relating to observed buyers' choices, the profit-maximization problem can be formulated within a multinomial limited dependent variable framework. Let y_{ij} be the observed share of urban expansion from land use j relative to the total urban expansion in region i.[2] Econometric estimation thus concerns itself with modelling the probability of observing y_{ij}. Within the multinomial logit framework, this probability can be modelled as a function of choice specific log-odds μ_{ij}, weighted by the sum of log-odds over all choice alternatives $\mu_{ij'}$ ($j' = 1, \ldots, J$):

$$p(y_{ij}) = \frac{\exp\ \mu_{ij}}{\sum_{j'=1}^{J} \exp\ \mu_{ij'}}. \tag{1}$$

In the standard non-spatial multinomial framework μ_{ij} is specified as a function of k explanatory variables, with corresponding choice-specific slope coefficients, which are to be estimated. The explanatory variables correspond to the expected rents and conversion costs with respect to land use j.

Spatial dependence among log-odds μ_{ij} in equation (1) involves the assumption that the choices of urban land buyers do not solely depend on rent and conversion costs in their own region i, but also on other regions' characteristics as well. This assumption implies that the probability of observing a land-use choice in region i also depends on land-use choices of all other regions. This assumption is based on the spatial nature of land expansion: before construction, investors typically scope multiple investment opportunities, which might not be contiguous, but located across regions in spatial proximity to each other.

Following the spatial econometric literature, such dependencies can be incorporated by imposing an exogenous neighbourhood structure through a non-negative and row-stochastic spatial weight matrix. Let $\boldsymbol{W}$ be such an $N \times N$ spatial weight matrix. Two regions i and i' are assumed to be neighbours of $w_{ii'} > 0$, otherwise $w_{ii'} = 0$. No region is a neighbour to itself, thus $w_{ii} = 0$.

The resulting SAR multinomial logit model can be expressed as:[3]

$$\begin{aligned} \boldsymbol{\mu}_j &= \rho_j \boldsymbol{W} \mu_j + \boldsymbol{X}\boldsymbol{\beta}_j \\ \boldsymbol{\mu}_j &= \boldsymbol{A}^{-1}\boldsymbol{X}\boldsymbol{\beta}_j, \end{aligned} \tag{2}$$

with $\boldsymbol{A}_j^{-1} = (\boldsymbol{I}_N - \rho_j \boldsymbol{W})^{-1}$ where $\boldsymbol{I}_N$ denotes an $N \times N$ identity matrix. The $N \times K$ matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K]$ collects the K vectors of explanatory variables and $\boldsymbol{\beta}_j$ denotes the respective $K \times 1$ vector of slope parameters related to choice j. The (scalar) parameter ρ_j measures the strength of spatial autocorrelation for land-use class j, with sufficient stability condition $\rho_j \in (-1, 1)$, where positive (negative) values of ρ indicate positive (negative) spatial autocorrelation. Note that the model allows for different ρ_j across land-use classes.[4] In the absence of spatial autocorrelation ($\rho_1 = \ldots = \rho_J = 0$), the model framework collapses to a classical multinomial logit set-up.

In such an SAR model specification, the $N \times 1$ vector of choice-specific log-odds $\boldsymbol{\mu}_j = [\mu_{1j}, \ldots, \mu_{Nj}]'$ thus also depends on the characteristics of other regions in the sample. Spatial dependence is introduced by the spatial multiplier $\boldsymbol{A}_j^1 = (\boldsymbol{I}_N - \rho_j \boldsymbol{W}^{-1}) = \sum_{r=0}^{\infty} \rho_j^r \boldsymbol{W}^r$.[5]

A core implication of the SAR modelling framework is that a change in the explanatory variables associated with region i would result in changes not only of the observed shares y_{ij} in the own region, but also in other regions. Through the nature of the multinomial logit model, where marginal impacts to one choice j also affect the shares of all other choices, this implies that in a spatial dependent setting marginal impacts of y_{ij} have spillover effects over regions and choices as well. While spatial dependence is only *explicitly* modelled among the log-odds of land-use class j, note that due to the nature of the multinomial logit model, where $p(y_{ij})$ depends on $\mu_{1j}, \ldots, \mu_{1J}$, spillovers across land-use classes are *implicitly* captured.

ESTIMATION STRATEGY

We propose a Bayesian estimation strategy for the SAR multinomial logit model which builds on the idea of introducing a latent variable in order to facilitate the estimation of the multinomial logit likelihood. This estimation strategy has been widely employed in recent Bayesian econometric literature for tackling models featuring non-Gaussian distributions (e.g., Frühwirth-Schnatter et al., 2009; Frühwirth-Schnatter & Frühwirth, 2012). To illustrate the core problem, consider the likelihood of the multinomial logit model in equation (1):

$$\prod_{i=1}^{N} \prod_{j=1}^{J} \frac{(\exp \mu_{ij})^{y_{ij}}}{\sum_{j'=1}^{J} \exp \mu_{ij'}}. \tag{3}$$

Note that the likelihood contribution of observation i relies not only on μ_{ij}, but also on the log-odds of making other choices. This well-known non-linearity in the likelihood greatly complicates the estimation of the unknown slope and SAR coefficients.

Within a Bayesian framework the focus of estimation frequently lies mainly on finding conditional posterior distributions for the parameters of interest. In fact, assuming suitable priors $p(\beta_j)$, the conditional posterior of β_j can be expressed conditional on all other slope coefficients β_{-j} and ρ (Holmes & Held, 2006):

$$\begin{gathered} p(\boldsymbol{\beta}_j | \boldsymbol{\beta}_{-j}, \rho) = p(\boldsymbol{\beta}_j) \prod_{i=1}^{N} \left(\frac{\exp \eta_{ij}}{1 + \exp \eta_{ij}} \right)^{y_{ij}} \left(\frac{1}{1 + \exp \eta_{ij}} \right)^{1 - y_{ij}} \\ \text{with } \eta_{ij} = \mu_{ij} - C_{ij} \text{ and } C_{ij} = \log \sum_{j' \neq j}^{J} \exp \mu_{ij'}. \end{gathered} \tag{4}$$

While this distribution cannot be easily sampled from, we follow the work of Polson et al. (2013), which has been adopted to the SAR variant of a bivariate logit distribution (Krisztin & Piribauer, 2020). A particularly useful result in Polson et al. (2013) is the fact that conditional on introducing a Pólya–Gamma-distributed latent random variable, exponential type distributions such as that in equation (4) can be recast as Gaussian, where posterior sampling can be easily achieved.

Particularly when conditioning on $\omega_{ij} \sim \mathcal{PG}(1, 0)$ – where $\mathcal{PG}(1, 0)$ denotes a Pólya–Gamma distribution with rate 1 and shape 0 – the conditional posterior of the slope parameters associated with choice j can be reformulated as:

$$\begin{aligned} p(\boldsymbol{\beta}_j|\boldsymbol{\beta}_{-j}, \rho, \omega_j) \quad & \alpha \quad p(\boldsymbol{\beta}_j)\prod_{i=1}^{N}\exp(k_{ij}\mu_{ij})\exp\left(\frac{\eta_{ij}^2\omega_{ij}}{2}\right)\mathcal{PG}(\omega_{ij}|1, 0) \\ & \alpha \quad p(\boldsymbol{\beta}_j)\exp\left\{-\frac{1}{2}([z_j - c_j] - \boldsymbol{A}_j^{-1}\boldsymbol{X})^{'}\Omega_j([z_j - c_j] - \boldsymbol{A}_j^{-1}\boldsymbol{X})\right\}, \end{aligned}$$

where $\omega_{1j} = [\omega_{1j}, \ldots, \omega_{Nj}]^{'}$ and $k_{ij} = y_{ij} - 1/2$. The conditional posterior has working responses $\boldsymbol{z}_j = [\kappa_{1j}/\omega_{1j}, \ldots, \kappa_{Nj}/\omega_{Nj}]'$ and $\boldsymbol{c}_j = [C_{1j}, \ldots, C_{Nj}]'$, with variance matrix $\Omega_j = diag(\omega_j)$. If we elicit a Gaussian prior distribution for the slope coefficients, with $p(\beta_j) = \mathcal{N}(\underline{\mu}_{\beta j}, \underline{\Sigma}_{\beta j})$, the conditional posteriors for the slope coefficients are also Gaussian:

$$\begin{aligned} p(\beta_j|\beta_{-j}, \rho, \omega_j) &= \mathcal{N}(\bar{\mu}_{\beta j}, \bar{\Sigma}_{\beta j}) \\ \bar{\mu}_{\beta j} &= \bar{\Sigma}_{\beta j}[(\boldsymbol{A}_j^{-1}\boldsymbol{X})'(\boldsymbol{\kappa}_j - \boldsymbol{\Omega}_j c_j) + \underline{\Sigma}_{\beta j}^{-1}\underline{\mu}_{\beta j}] \end{aligned} \tag{5}$$

$$\bar{\Sigma}_{\beta j} = [(\boldsymbol{A}_j^{-1}\boldsymbol{X})'\boldsymbol{\Omega}_j(\boldsymbol{A}_j^{-1}\boldsymbol{X}) + \underline{\Sigma}_{\beta j}^{-1}]^{-1}. \tag{6}$$

The Gaussian conditional posterior of the slope parameters reveals the particular appeal of using latent Pólya–Gamma-distributed variables. A wide variety of Bayesian model extension, such as variable selection, or uncertainty over the $\boldsymbol{W}$ can be easily introduced in the above framework.

Following Polson et al. (2013), the conditional distribution of ω_j is also a Pólya–Gamma distribution:

$$p(\boldsymbol{\omega}_j|\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J, \boldsymbol{\rho}_1, \ldots, \boldsymbol{\rho}_J, \boldsymbol{\omega}_{-j}) = \mathcal{PG}(1, \boldsymbol{\eta}_j),$$

where $\boldsymbol{\eta}_j = [\eta_1, \ldots, \eta_N]'$. Computationally efficient algorithms for sampling from the Pólya–Gamma distribution are readily available in the R package *BayesLogit*.

The conditional posterior of ρ relates directly to the multinomial logit:

$$p(\boldsymbol{\rho}_j|\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_J, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)\alpha\, p(\boldsymbol{\rho}_j)\prod_{i=1}^{N}\prod_{j=1}^{J}\frac{(\exp\mu_{ij})^{y_{ij}}}{\sum_{j'=1}^{J}\exp\mu_{ij'}} \tag{8}$$

$$\boldsymbol{\mu}_j = \boldsymbol{A}^{-1}\boldsymbol{X}\boldsymbol{\beta}_j \tag{9}$$

where $p(\rho_j)$ denotes the prior distribution of ρ_j. The conditional posterior in equation (8) is not from a well-known form and thus cannot be sampled from easily. This is usual in the spatial econometric literature, and the standard solution is to use a Metropolis–Hastings step, as in LeSage and Pace (2009).

MARKOV CHAIN MONTE CARLO SAMPLING PROCEDURE

Given the conditional posterior distributions stated above, Markov chain Monte Carlo (MCMC) algorithms can be employed by sequentially sampling from the conditional posteriors. We follow the usual identification assumption of the multinomial logit model in that we set $\boldsymbol{\beta}_J = 0$, $\rho_J = 0$ and $\omega_J = 0$. With suitable starting values for $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{J-1}$ and $\rho_1, \ldots, \rho_{J-1}$, our sampler involves the following steps:

I. For $j = 1, \ldots, J-1$, update ω_j by drawing from $p(\boldsymbol{\omega}_j|\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J, \boldsymbol{\rho}, \boldsymbol{\omega}_{-j})$ using equation (7).
II. For $j = 1, \ldots, J-1$, update β_j by drawing from $p(\boldsymbol{\beta}_j|\boldsymbol{\beta}_{-j}, \boldsymbol{\rho}, \boldsymbol{\omega}_j)$ using equation (5).

III. Update ρ_j using a Metropolis–Hastings step from $p(\boldsymbol{\rho}_j|\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}, \boldsymbol{\rho}_{-j}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)$ based on equation (8).

The MCMC algorithm cycles through steps I–III B times by excluding the first B_0 draws as burn-ins. Inference on the parameters is conducted using the $B - B_0$ remaining draws.[6]

SIMULATION STUDY

For the simulation study we use a SAR multinomial logit model as a benchmark data generating process with three choice classes ($J = 3$) and two randomly generated explanatory variables ($k = 2$). The data-generating process can be written as follows, where variables with a tilde denotes generated quantities:

$$\tilde{y}_{ij} = \frac{\exp \tilde{\mu}_{ij}}{\sum_{j'=1}^{J} \tilde{\mu}_{ij'}} \tag{10}$$

where:

$$\tilde{\boldsymbol{\mu}}_j = (\boldsymbol{I} - \rho_j \tilde{\boldsymbol{W}})^{-1} \boldsymbol{X} \boldsymbol{\beta}_j,$$

with $\tilde{\boldsymbol{\beta}}_{-J} = [\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2]$ and $\tilde{\boldsymbol{\beta}}_j = 0$. The slope coefficients and the explanatory variables are generated anew in each Monte Carlo iteration, where $\tilde{\boldsymbol{X}}$ stems from a standard normal distribution. The true slope coefficients are generated in each Monte Carlo iteration from a Gaussian distribution, where the means for $\tilde{\boldsymbol{\beta}}_1$ are $[1, 0.5]'$ and the means for $\tilde{\boldsymbol{\beta}}_2$ are $[0.5, 1]'$. The variance corresponds in both cases to 0.1. The row-stochastic spatial weight matrix $\tilde{\boldsymbol{W}}$ is based on a random spatial pattern generated from a Gaussian distribution for latitude and longitude, and constructed using seven nearest neighbours.[7] Note that our dependent variable $\tilde{y}_{ij}$ is a share variable, as is often used in land-use share models (e.g., Chakir & Parent, 2009).

In a Monte Carlo study we benchmark the SAR multinomial logit model in order to assess the predictive performance of our proposed modelling framework against two competing specifications: (1) a non-spatial version of the SAR multinomial logit, where the SAR coefficient $\rho_j = 0$ for all j; and (2) $J - 1$ bivariate SAR logit models where each logit model captures the log-odds of not choosing option J. The latter bivariate SAR logit models are in fact the same as the model in equation (1), albeit with the restriction $J = 2$. The estimation exactly corresponds to that laid out in the previous section.

To assess the strength of the specifications along multiple scenarios, we vary the strength of spatial dependence $\rho_j \in \{0, 0.5, 0.8\}$. To evaluate the accuracy of the sampler with respect to the chosen sample size, we consider $N \in \{400, 1000, 1400\}$. Particularly, the sample size $N = 1,400$ was chosen as it corresponds most closely to the number of observations in the empirical application of this paper. Across all models, our prior set-up is as follows: we use a rather uninformative Gaussian prior for $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{J-1}$ with zero mean and variance 10^8 and for $\rho_1, \ldots, \rho_{J-1}$ we use a the standard beta prior specification as proposed in LeSage and Pace (2009).

The results of the Monte Carlo study are summarized in Table 1. Each element of the table corresponds to the average over 1000 runs for a particular model specification and Monte Carlo scenario. The first and second columns contain information on the sample size N and the model specifications. Corresponding to the choice of spatial dependence, Table 1 reports the average root mean squared error (RMSE) point estimates for slope coefficients, average direct and indirect impacts, as well as average estimates for ρ_j for all j.

In the case of no spatial autocorrelation ($\rho_j = 0$), the non-spatial multinomial logit exhibits the highest estimation accuracy for all sample sizes under scrutiny. It is worth noting that this

Table 1. Root mean squared error measures for the Monte Carlo runs.

		RMSE											
		$\rho_j = 0.0$				$\rho_j = 0.5$				$\rho_j = 0.8$			
N	**Model**	$\boldsymbol{\beta}$	**Direct**	**Indirect**	ρ	$\boldsymbol{\beta}$	**Direct**	**Indirect**	ρ	$\boldsymbol{\beta}$	**Direct**	**Indirect**	ρ
400	SAR multinomial logit	0.019	0.003	0.005	0.086	**0.022**	**0.003**	**0.021**	**0.086**	**0.032**	**0.003**	**0.031**	**0.023**
	Multinomial logit	**0.018**	**0.003**	**0.000**	**0.000**	0.033	0.004	0.110	0.500	0.069	0.008	0.393	0.800
	Bivariate logit	0.533	0.156	0.077	0.262	0.534	0.147	0.179	0.179	0.523	0.118	0.445	0.349
1000	SAR multinomial logit	**0.011**	0.002	0.003	0.039	**0.014**	**0.002**	**0.010**	**0.034**	**0.016**	**0.002**	**0.013**	**0.009**
	Multinomial logit	0.012	**0.002**	**0.000**	**0.000**	0.024	0.003	0.109	0.500	0.049	0.005	0.391	0.800
	Bivariate logit	0.533	0.155	0.076	0.265	0.533	0.145	0.178	0.175	0.519	0.114	0.442	0.341
1400	SAR multinomial logit	**0.010**	**0.002**	0.003	0.036	**0.011**	**0.002**	**0.007**	**0.022**	**0.016**	**0.002**	**0.011**	**0.008**
	Multinomial logit	0.010	0.002	**0.000**	**0.000**	0.024	0.002	0.109	0.500	0.047	0.004	0.391	0.800
	Bivariate logit	0.534	0.154	0.076	0.266	0.532	0.145	0.178	0.174	0.519	0.115	0.442	0.346

Note: Results are based on 1000 Monte Carlo runs. For each Monte Carlo run, the corresponding sampling algorithms are run using 1000 draws, where the initial 700 draws were discarded as burn-in. The columns *direct* and *indirect* correspond to summary marginal effects (for details, see the Appendix in the supplemental data online). The values given for $\boldsymbol{\beta}$, *direct*, *indirect* and ρ correspond to the average **RMSE**$(\cdot)$ over all Monte Carlo iterations. Bold values denote the lowest average **RMSE** scores.

result is hardly surprising, as in the absence of spatial autocorrelation this model resembles the true data-generating process most closely. However, the SAR multinomial logit closely tracks the estimates of its non-spatial counterpart. In the case of larger sample sizes (N = 1000 and N = 1400), our proposed model specification even slightly outperforms all considered competitors in terms of average direct effects and slope coefficients.

For a moderate degree of spatial autocorrelation ($\rho_j = 0.5$), the SAR multinomial logit model outperforms all other specifications under scrutiny for all three considered sample sizes. In terms of direct average impacts, the non-spatial multinomial logit model exhibits roughly similar performance. However, in the case of the smallest sample size ($N = 400$), the bias in terms of point predictions clearly increases. Note that the competing bivariate SAR logit specification shows considerable bias in estimating the spatial autocorrelation parameters, albeit the bias is less than that of the non-spatial multinomial logit.

Turning attention to a high degree of spatial autocorrelation ($\rho_j = 0.8$), we observe that the SAR multinomial logit model significantly outperforms its alternatives. Furthermore, when high spatial autocorrelation is present, the bivariate SAR logit exhibits lower bias in terms of point prediction of the SAR parameters ρ, as the non-spatial multinomial logit model.

Overall, we can conclude that the SAR multinomial logit model outperforms both a non-spatial multinomial logit as well as the application of bivariate SAR logit models. This result applies both in moderate and large sample sizes. Even when no spatial autocorrelation is present, the SAR multinomial logit model produces rather promising results in terms of predictive performance because it closely tracks the results of its non-spatial counterpart.

EUROPEAN LAND-USE CHANGE

The recent literature focused attention on land sealing resulting from urban sprawl, and associated spillovers with other land-use classes. Results from van Vliet (2019) suggest that in the last decade in Europe 8.4 Mha of land has been converted to urban, out of which 6.3 Mha was converted from cropland. However, this land sealing led to 13.1 Mha displacement of other land-use classes, as cropland was expanded elsewhere, to compensate for the lack of production resources, out of which the majority (13 Mha) was expanded in other regions. These spillover effects are well documented in the literature (e.g., Coisnon et al., 2014; Guastella et al., 2017; Zoppi & Lai, 2014), and serve as a motivation for an empirical application of the spatial multinomial logit model. Both global (Ay et al., 2017) and local (Deng et al., 2008) spillovers are considered of importance.

In the spirit of Chakir and Parent (2009), Zoppi and Lai (2014) and Lai and Zoppi (2017), we model the areal share of urban sprawl stemming from non-urban land in a given region within a spatial Durbin multinomial logit model, where the log odds take the following form:

$$\boldsymbol{\mu}_j = \rho_j \boldsymbol{W} \boldsymbol{\mu}_j + \alpha + \boldsymbol{X} \boldsymbol{\beta}_j + \boldsymbol{W} \boldsymbol{X} \theta_j. \tag{11}$$

The scalar α is an intercept and the term $\boldsymbol{W}\boldsymbol{X}$ is a spatial lag of the matrix of covariates with associated vector of parameters $\boldsymbol{\theta}_j$. This lag explicitly controls for the regions' characteristics of their neighbours.

REGIONS, DATA AND SPATIAL WEIGHTS

Our sample covers a cross-section of 1316 European regions across 27 countries. The regions are classified under the NUTS 2013 classification at the NUTS-3 level. The NUTS-3 regions used, though varying in size, are generally considered to be appropriate spatial units for modelling and analysis purposes. But the delineation of the NUTS-3 regions is formal rather than functional in nature, and they do not necessarily represent the boundaries of the regional processes under scrutiny. However, it is worth noting that spatial aggregation appears necessary for a meaningful spatial analysis on a pan-European regional level.[8]

The regions included in the sample are located in Austria (35 regions), Belgium (44 regions), Bulgaria (28 regions), Cyprus (one region), Czech Republic (14 regions), Denmark (11 regions), Estonia (five regions), Finland (19 regions), France (96 regions), Germany (402 regions), Greece (52 regions), Hungary (20 regions), Italy (110 regions), Latvia (six regions), Lithuania (10 regions), Luxembourg (one region), Malta (two regions), Netherlands (40 regions), Poland (72 regions), Portugal (25 regions), Republic of Ireland (eight regions), Romania (42 regions), Slovakia (eight regions), Slovenia (12 regions), Spain (59 regions), Sweden (21 regions) and the UK (173 regions).

The dependent variable of our analysis describes the share of land sealing emanating from any non-urban type of land within the period from 2000 to 2018. More formally, it is defined as the land area of a certain type of land use that is being transformed to urban land use between 2000 and 2018 divided by the whole area of land-sealing expansion that took place in the respective period. As a result, we obtain a compositional data vector that – by definition – sums to unity.

The types of land use we consider follow the empirical literature on land-use changes and urban expansion (Chakir & Le Gallo, 2013; Chakir & Parent, 2009; Lai & Lombardini, 2016; Lai & Zoppi, 2017; Zoppi & Lai, 2014). We distinguish between the five classes cropland, grassland, forest, other and urban. It is worth noting that we classify *urban* land use as both *settlement* areas (that is, man-made buildings) as well as *artificial* surfaces (such as roads, mines or construction sites). The main focus of our analysis is on land sealing by *urban* expansion. However, in order to make our results more robust, we also model land sealing from expanding settlement and artificial areas in particular. The raw data stem from the CORINE Land Cover (CLC) maps provided by Copernicus Land Monitoring Service (CLMS). Their maps are based on satellite data with minimum mapping units (MMU) of 25 ha for areal phenomena and a minimum width of 100 m for linear phenomena. The data consist of an inventory of land cover in 44 classes, which we summarize to the classes stated above.[9] We use CLC change-layers also provided by CLMS, designed to capture the land cover changes at a higher resolution between two neighbour surveys. Regional aggregates at the NUTS-3 level are obtained by simple summation of all changes of the corresponding raster elements. Likewise, changes for the whole investigated period are obtained by addition of the three sub-periods for which CLC change-layers are provided. Further data sources are (1) the Urban Data Platform Plus provided as a joint initiative of the Joint Research Centre (JRC) and the Directorate General for Regional and Urban Policy (DG REGIO) of the European Commission, (2) Eurostat (the statistical office of the European Union) and (3) the European Observation Network for Territorial Development and Cohesion (ESPON).

Our set of covariates consists of $K' = 19$ variables that are commonly employed in the land-use change literature (for an overview, see Shaw et al., 2020). Further, to capture the complex spatial structure we include not only the spatially lagged dependent vector but also the spatially lagged forms of the explanatory variables (except for the dummy variables). We also include a vector of ones as intercept. After including the spatially lagged covariates, the resulting design matrix is of column dimension 38. Table 2 provides a short technical description for the variables included in our estimation.

Since the rent of a certain land-use class is assumed to affect the decision of land-owners – yet it is usually not observed – many recent studies consider various proxies to control for the variation in returns from different land uses (e.g., Livanis et al., 2006; Lubowski et al., 2008). Chakir and Parent (2009) conclude that agricultural gross value added divided by the respective land-use area serves as a reasonably good proxy. Higher rents are therefore assumed to reduce the amount of land that is converted to artificial area.

Table 2. Variables used in the empirical analysis.

Variable	Description	Source
Cropland to …	Sum of 2000–06, 2006–12 and 2012–18 CLC land-cover changes from cropland to urban/settled/artificial land, divided by the total change of area in the same period	CLC
Forest to …	Sum of 2000–06, 2006–12 and 2012–18 CLC land-cover changes from forests to urban/settled/artificial land, divided by the total change of area in the same period	CLC
Grassland to …	Sum of 2000–06, 2006–12 and 2012–18 CLC land-cover changes from pastures and grassland to urban/settled/artificial land, divided by the total change of area in the same period	CLC
Other to …	Sum of 2000–06, 2006–12 and 2012–18 CLC land-cover changes from area of other use to urban/settled/artificial land, divided by the total change of area in the same period	CLC
Crop rent	Share of agricultural gross value added, divided by km^2 of area used to grow crops, 2000	JRC, CLC
Forest rent	Share of agricultural gross value added, divided by km^2 of forest area, 2000	JRC, CLC
Grass rent	Share of agricultural gross value added, divided by km^2 of pasture and grassland, 2000	JRC, CLC
Initial sealed area	Area of urban/settled/artificial land cover, 2000	CLC
Sealed area growth	Growth of urban/settled/artificial areas between 2000 and 2018 (%)	CLC
Employment primary	Share of employment in the primary sector (NACE A), in total employment, 2000	JRC
Employment tertiary	Share of employment in the tertiary sector (NACE F to Q) in total employment, 2000	JRC
GDP per capita	Gross domestic product divided by population, 2000	JRC
Population density	Population/km^2, 2000	JRC
Elevation	Average elevation (m)	Copernicus
Slope	Average slope (°)	Copernicus
Soil moisture	Content of liquid water in a surface soil layer of 2–5 cm depth expressed as qubic m water per qubic m of soil, 2000	Copernicus
N2000 cropland	Share of protected area used to grow crops over total area used to grow crops, 2000	Natura 2000
N2000 forest	Share of protected area of forests over total area of forests, 2000	Natura 2000
N2000 grassland	Share of protected area of pastures and grassland over total area of pastures and grassland, 2000	Natura 2000

(*Continued*)

Table 2. Continued.

Variable	Description	Source
N2000 other	Share of protected area of other use over total area of other use, 2000	Natura 2000
Objective 2 region	Dummy variable, 1 denotes region eligible under Objective 2, 2000–06; 0 otherwise	ESPON
Farm density	Number of farms divided km^2, measured at a NUTS-2 level, 2000	Eurostat
Farm size	Total farm area divided by the number of farms, measured at a NUTS-2 level, 2000	Eurostat

Note: CLC, CORINE Land Cover; ESPON, European Observation Network for Territorial Development and Cohesion; JRC, Joint Research Centre.

The initial level of land-sealing areas and especially land-sealing expansion rates are discussed in the literature in the context of the level of available agricultural amenities (Coisnon et al., 2014; Wu, 2006; Wu & Plantinga, 2003). Based on this strain of the literature, lower initial sealed area expansion would lead to higher land-sealing rates, as regions surrounding population centres with low share of built-up land are in higher demand.

On the other hand, quantities on employment, population and income are typical variables to represent the degree of economic development. Employment enters the model in the form of sectoral shares, with manufacturing (secondary) as baseline. Region-specific population, a particularly important driver of land take (e.g., Guastella et al., 2017; Paulsen, 2012; Terama et al., 2019), is divided by the respective area and therefore captured as density. Income is measured as gross domestic product (GDP) per inhabitant. High shares of tertiary employment, paired with high income and population density, is usually observed around the city centres and, therefore, associated with the expansion of housing supply, which again should translate into urban expansion.

Quantities usually associated with the quality of soil include measures of slope, elevation and moisture (usually in the form of precipitation or humidity). Following Chang-Martínez et al. (2015) we include these physical drivers of land-use conversion, as they implicitly influence the cost of land conversion. We consider slope and elevation in average meters and degrees, respectively. Soil moisture is captured as volumetric measure of liquid water in a surface soil layer of 2–5 cm depth. Variables capturing the quality of the land are assumed to have a negative impact on conversion of productive land, as they are to be interpreted as costs of conversion (Huang et al., 2006; Shaw et al., 2020).

Additionally, national regulations, as the amount of nature conservation areas, restrict the potential conversion. We include the share of area being protected under the Natura 2000 network of nature protection. The Natura 2000 network's main objective is to preserve natural habitats and secure biodiversity in the EU, hence forest and grassland areas are of main concern (Lai & Zoppi, 2017).

In the discussion of steering soil sealing, subsidies and taxes play a key role (Artmann, 2014; Shaw et al., 2020). As a proxy for European-level subsidies, we use observations on whether a region received Objective 2 level regional funding within the period, because this type of funding is also used to enhance infrastructure in the region.[10] An additional major source of subsidy for land-use management are agricultural subsidies of countries, as well as the EU. These are not divided on the regional level, but by farm size and productivity. Therefore, to control for the heterogeneous structures of agricultural actors across Europe, variables that account for farm-specific characteristics are incorporated (for a discussion, see Delbecq et al., 2014).

For the spatial weights matrix $\boldsymbol{W}$, we use a seven nearest-neighbour specification, where every region is constrained to be a neighbour of its seven closest regions. Our results, however, prove robust to variations in the assumed spatial dependence structure.

EMPIRICAL RESULTS

This subsection presents the MCMC results obtained from 10,000 posterior draws for our spatial multinomial logit specification, where the first 5000 were discarded as burn-in.[11] Straightforward interpretation of coefficient estimates in spatial models could lead to deceptive or misleading conclusions (e.g., Anselin, 1988; LeSage & Fischer, 2009). One possibility is to provide summary metrics in form of direct, indirect (spillover) and total effects. Following LeSage and Pace (2009) we present marginal effects in Table 3. Direct effects (in the top panel) are then to be interpreted similarly to regular slope coefficients. In turn, indirect effects (bottom panel) account for the impacts due to changes in other regions and are therefore to be interpreted as spillover effects. We find a significant class-specific spatial parameter ρ_j for cropland (across all land-sealing process) as well as grassland (only for urban and settlement sources), highlighting the necessity of incorporating the spatial dependence structure in the model. This result confirms the findings of Guastella et al. (2017) and especially of van Vliet (2019), in that land sealing on productive land leads to further spillover land conversions in surrounding regions.

In addition, Table 3 reports the McFadden pseudo-R^2, which serves as a measure of the goodness of fit in limited dependent variable models. McFadden (1974) highlights that values between 0.2 and 0.4 already indicate a rather good fit, which is true for both the urban and settlement models, while the multinomial logit model with artificial area expansion as a dependent variable is, with a pseudo-R^2 of 0.193 relatively close to this rule of thumb.

The rest of the reported results are to be interpreted as follows: each set of columns corresponds to a specific type of land use (cropland, grassland, forest and other land). Within a specific land use, the three columns correspond to the sources of land sealing under scrutiny: land sealing from urban, artificial or settlement surfaces (which are in turn individual multinomial logit models). The posterior mean estimates correspond to the changes of the probabilities to convert the respective class in that region to urban, artificial or settlement area, respectively. The top panel corresponds to estimates effecting the own region, while the bottom panel corresponds to estimates with respect to a change in neighbouring regions.

The direct effects of the three types of land rent proxies (crop, forest and grass) confirm the results from Chakir and Lungarska (2017) and Chakir and Parent (2009), in that for each land-use class higher rents imply a significantly lower chance of land sealing. Additionally, the joint modelling in a multinomial model indicates that significant spillover effects to other classes are present. Most notably, an increase in cropland rents in a region would also increase the chance of grassland conversion to sealed surfaces. We find analogous relationships for forest rent and cropland, as well as grass rent and cropland. The results seem robust, even if only land sealing from artificial or settlement is taken into account. Note that in the case of urban expansion, higher cropland rents also significantly increase the chance of forest being converted to urban land (but not its component classes artificial and settlement). In terms of indirect effects, an increased rent from crops and forest land would also lead to significantly lower chance of land sealing of the respective land-use class in neighbouring region. Additionally, higher rents from cropland would also increase the chance of neighbouring regions sealing grassland in preference of other land-use classes, though the estimate is only significant in terms of urban or settlement areas, which seems to support the conclusions of Shaw et al. (2020).

A higher initial level of artificial areas indicates that land sealing of other natural vegetation in the own region has a significantly higher probability as compared with land sealing of the other land covers under scrutiny. If settled areas are present (for the urban and settlement dependent

Table 3. Summary impact measures for sealed area expansion from cropland, grassland, forest and other land.

	Cropland to …			Grassland to …			Forest to …			Other to …		
	Urban	**Artificial**	**Settlement**	**Urban**	**Artificial**	**Settlement**	**Urban**	**Artificial**	**Settlement**	**Urban**	**Artificial**	**Settlement**
Direct effects												
Crop rent	**−0.053**	**−0.047**	**−0.034**	**0.033**	**0.029**	**0.020**	**0.018**	0.018	0.008	0.000	−0.002	0.006
Grass rent	**0.044**	**0.043**	**0.033**	**−0.029**	**−0.028**	**−0.035**	−0.012	−0.013	0.003	−0.003	−0.003	−0.001
Forest rent	**0.078**	**0.090**	**0.053**	−0.012	−0.006	−0.015	**−0.056**	**−0.067**	**−0.030**	−0.010	**−0.017**	**−0.010**
Initial sealed area	**−0.102**	−0.033	**−0.076**	0.031	0.007	0.008	0.020	−0.008	**0.037**	**0.050**	**0.037**	**0.029**
Sealed area growth	−0.006	−0.025	**0.026**	0.010	0.007	0.007	−0.010	0.004	**−0.023**	0.005	**0.015**	−0.010
Employment primary	0.005	−0.011	0.023	0.010	0.025	−0.002	−0.026	−0.020	−0.027	0.006	0.008	0.010
Employment tertiary	−0.033	−0.042	**−0.043**	0.010	0.014	0.015	−0.007	−0.006	0.009	**0.029**	**0.031**	0.022
GDP per capita	0.012	0.006	**0.034**	**−0.050**	**−0.053**	−0.029	0.021	0.025	−0.003	0.018	0.019	−0.003
Population density	**0.095**	0.044	0.054	0.017	**0.050**	0.007	**−0.052**	**−0.040**	**−0.038**	**−0.060**	**−0.054**	−0.020
Elevation	0.010	0.011	0.001	0.026	0.012	0.031	−0.014	−0.005	−0.027	−0.018	−0.016	−0.003
Slope	**−0.055**	**−0.053**	−0.026	−0.001	0.007	−0.011	0.024	0.016	**0.024**	**0.032**	**0.031**	0.011
Soil moisture	−0.011	−0.009	−0.014	0.012	0.012	**0.026**	0.005	0.007	−0.003	−0.008	−0.011	−0.007
N2000 cropland	−0.016	−0.013	−0.016	0.005	0.004	0.017	0.010	0.010	−0.002	0.000	0.001	0.002
N2000 forest	**0.047**	**0.046**	**0.045**	−0.014	−0.015	**−0.032**	**−0.025**	**−0.025**	−0.010	−0.007	−0.006	−0.004

(*Continued*)

Table 3. Continued.

	Cropland to …			Grassland to …			Forest to …			Other to …		
	Urban	**Artificial**	**Settlement**	**Urban**	**Artificial**	**Settlement**	**Urban**	**Artificial**	**Settlement**	**Urban**	**Artificial**	**Settlement**
N2000 grassland	0.009	0.005	0.015	−0.028	−0.023	−0.032	0.014	0.014	0.014	0.004	0.006	0.002
N2000 other	−0.006	−0.003	0.000	0.017	0.015	**0.023**	0.003	0.002	−0.012	−0.013	−0.015	−0.011
Objective 2 region	**−0.078**	−0.072	**−0.080**	−0.036	−0.034	0.002	0.043	0.027	0.033	**0.080**	**0.077**	**0.045**
Farm density	0.000	0.000	0.019	−0.029	−0.020	−0.018	0.024	0.017	0.011	0.005	0.006	−0.007
Farm size	−0.011	−0.013	−0.014	0.015	0.019	0.008	−0.004	−0.008	0.001	0.001	−0.001	0.007
Indirect effects												
Crop rent	**−0.083**	**−0.084**	**−0.064**	**0.044**	0.050	**0.035**	0.018	0.016	0.018	0.018	0.011	0.011
Grass rent	−0.008	−0.007	0.002	0.007	0.014	−0.012	0.010	0.002	0.017	−0.006	−0.007	−0.008
Forest rent	0.018	0.028	0.021	**−0.042**	**−0.048**	**−0.039**	0.003	−0.007	0.030	0.026	0.028	−0.011
Initial sealed area	−0.030	0.029	−0.031	−0.018	−0.013	0.001	0.010	−0.032	0.008	0.027	0.012	0.019
Sealed area growth	−0.006	−0.050	0.024	−0.001	0.015	−0.008	0.020	0.037	−0.010	−0.009	0.000	−0.006
Employment primary	−0.091	**−0.127**	−0.086	0.009	0.011	0.005	0.076	**0.098**	**0.088**	−0.005	0.015	−0.005
Employment tertiary	−0.057	−0.067	−0.078	−0.002	−0.007	0.023	0.079	**0.075**	0.069	−0.015	−0.005	−0.015
GDP per capita	0.071	0.061	0.090	−0.015	−0.013	−0.029	−0.036	−0.032	−0.066	−0.021	−0.017	0.000
Population density	−0.022	−0.042	−0.042	0.012	−0.002	−0.009	0.024	0.043	0.062	−0.006	0.002	−0.011
Elevation	0.016	0.030	0.025	−0.020	−0.032	−0.013	0.005	0.011	−0.002	−0.004	−0.004	−0.014

Slope	0.011	0.001	−0.004	0.005	0.015	0.008	−0.052	−0.056	−0.009	0.028	0.032	0.006
Soil moisture	−0.039	−0.049	−0.028	−0.002	−0.004	0.006	**0.060**	**0.061**	0.046	−0.019	−0.009	−0.022
N2000 cropland	−0.023	−0.026	−0.006	0.013	0.019	0.002	−0.010	−0.011	−0.011	0.018	0.019	0.019
N2000 forest	0.025	0.022	0.035	−0.014	−0.012	−0.007	−0.029	−0.022	−0.031	0.017	0.014	−0.006
N2000 grassland	0.033	0.043	0.051	0.005	0.008	−0.003	−0.018	−0.017	−0.048	−0.020	−0.031	−0.006
N2000 other	0.017	0.022	−0.033	−0.013	−0.015	0.005	0.034	0.029	0.021	−0.034	**−0.032**	0.010
Objective 2 region	−0.004	−0.004	−0.003	0.000	0.000	0.000	−0.002	−0.001	0.000	**0.000**	**0.000**	**0.000**
Farm density	0.049	0.050	0.052	−0.027	−0.018	−0.009	0.005	−0.001	−0.035	−0.031	−0.029	−0.015
Farm size	−0.014	−0.010	−0.007	−0.006	−0.002	−0.014	0.011	0.008	0.034	0.013	0.010	−0.006
ρ_j	**0.057**	**0.067**	**0.042**	**0.062**	0.047	**0.112**	0.017	0.022	−0.002	0.000	0.000	0.000
McFadden's R^2	**0.210**	**0.193**	**0.244**									

Note: Land-sealing classes defined as: urban (CLC111–142), artificial (CLC121–142) and settlement (CLC111–112). Summary metrics are based on 10,000 Markov chain Monte Carlo iterations, where the first 5000 were discarded as burn-in. Bold written estimates indicate statistical significance under a 90% credible interval.

variables), a higher initial area in the own region would imply a significantly lower chance of productive cropland being sealed under artificial surfaces. Burnett (2012) provides similar findings, where urbanization is a process which enforces itself. Moreover, as the crop, grass, and forest land surrounding cities is frequently the most productive (Shaw et al., 2020), it seems intuitive that urban expansion would take from the comparatively less productive other natural vegetation. This result for the own region is robust (albeit somewhat more muted) for artificial and settlement expansion as well. The growth of sealed area in the observed period – whether from urban, artificial or settlement – appears to have no indirect impacts on the allocation to the land-use classes. In the own region, however, having an increased growth of settlement would lead to a higher probability of being sealed for cropland and to a lower probability for forest. Additionally, a higher increase in artificial area growth implies a significantly increased probability of sealing off other natural vegetation, as opposed to more agriculturally productive land-use classes.

Regarding the sectoral mix of employment, our results indicate that a higher share of tertiary employment in the own region implies a significantly higher probability of other natural vegetation being sealed. This finding is robust across different types of land sealing. This reflects the findings of Salvati (2016) and Salvati and Carlucci (2016), where higher tertiary employment is found to mainly reflect the presence of urban fabric. The positive spillover effects of primary and tertiary employment to neighbouring regions' grassland can be contextualized as the effect of industrial belts on forestry. Interestingly, this result seems to be only significantly present in land sealing from artificial and settlement sources, not their aggregate.

Our results with regards to GDP per capita suggest that it is not a significant driver of land sealing from urban expansion in a European context. Merely in terms of settlements (i.e., exclusively residential land use) does a higher average income imply more cropland being sealed. This is opposed to findings of, for example, Deng et al. (2008) for developing countries, where GDP per capita is found to be one of the main drivers of urbanization. Moreover, we find that a higher GDP per capita in fact significantly lowers the probability of sealing grassland with artificial surfaces in the own region. When observed jointly with the direct effects of population density, this supports findings by McGrath (2005) and Guiling et al. (2009), who find that population is a more significant driver of urbanization, as opposed to personal income. This seems to significantly apply to cropland being sealed by urban areas, as well as grassland being sealed by artificial areas. However, note that land takes from forest and other natural land are negative and significant. That is a higher population density in fact results in a lower chance of land conversion. This result can be interpreted on the one hand with the fact that regions with a higher endowment of population density are more urban in nature and contain a much lower percentage of cropland or other natural vegetation. On the other hand, work by Delbecq et al. (2014) and Wu (2006) provide evidence that private homeowners exhibit strong preferences for surrounding grassland amenities.

Turning our attention to the estimated impacts of the biophysical drivers elevation, slope and soil moisture, we can largely confirm the overall conclusions of Shaw et al. (2020) and Chang-Martínez et al. (2015) in that the biophysical processes play a secondary role to socio-economic ones in explaining land-sealing processes. For the own-region, only slope has a small, albeit significant, impact on the probability of sealing other natural vegetation, which is offset by a negative impact on sealing cropland. Additionally, a higher percentage of soil moisture indicates a significantly higher chance of sealing forest land in neighbouring regions.

Our results indicate that the Natura 2000 protection programme has intended effects, as higher shares of protected forested land would – according to our results – lead to a significantly lower chance of the respective land cover being sealed. Simultaneously, we see that with a higher percentage of forest land under natural protection, the probability of cropland being sealed under urban surfaces increases significantly in the own region. Additionally, more protected forests seem to also significantly decrease the likelihood of grassland being a source of settlement expansion. Note that if a region has a higher share of other natural vegetation under Natura 2000 protection, this would lower the chances of neighbouring regions converting this land cover to artificial land. This largely confirms the findings of Lai and Lombardini (2016), Zoppi and Lai (2014), and Lai and Zoppi (2017). Our joint multinomial logit framework, however, allows us to uncover additional interdependencies among the natural protection of land covers.

The estimated results with regard to our subsidy proxies seem to show that regional funding plays a comparatively larger role as farm-specific subsidies. The own-regional effect of regional level Objective 2 subsidies is significant and negative for cropland, and positive for other natural vegetation, across all types of land sealing. This finding supports the hypothesis that subsidies increase land conversion (Shaw et al., 2020). Particularly noteworthy is the result that the land take comes more significantly from natural vegetation (which is highest in biodiversity) as opposed to more productive cropland. Additionally, neighbours of regions under Objective 2 funding also have a very small, albeit statistically significant, decreased chance of sealing other land.

Overall, our findings appear robust across multiple definitions of land sealing, particularly as pertaining to the significance of rents, the role of initial sealed surface endowments, as well as protected areas. An additional source of uncertainty for our results could be that cropland and grassland share many similar characteristics and often are strongly interrelated in terms of

Table 4. Summary impact measures for land sealing from urban area expansion from agricultural, forest, and other land.

	Direct			Indirect		
	Agricultural	Forest	Other	Agricultural	Forest	Other
Agricultural rent	−0.015	0.020	−0.003	0.037	−0.049	0.013
Forest rent	**0.073**	**−0.062**	−0.010	−0.026	−0.009	0.036
Sealed initial area	**−0.078**	0.018	**0.062**	0.003	0.005	−0.009
Sealed area growth	0.012	−0.010	0.000	−0.022	0.012	0.009
Employment primary	0.021	−0.021	0.002	−0.020	0.004	0.015
Employment tertiary	−0.025	−0.007	**0.033**	0.064	−0.038	−0.028
GDP per capita	−0.028	0.015	0.014	0.003	0.006	−0.012
Population density	**0.122**	**−0.056**	**−0.063**	0.017	−0.024	0.010
Elevation	0.025	−0.010	−0.012	−0.031	0.009	0.019
Slope	**−0.047**	0.019	**0.026**	0.007	0.006	−0.014
Soil moisture	−0.002	0.009	−0.008	0.000	−0.014	0.017
N2000 cropgrass	−0.005	0.005	0.000	0.043	−0.015	−0.027
N2000 forest	0.025	−0.019	−0.004	−0.003	0.000	**0.000**
N2000 other	0.014	0.003	−0.016	**0.057**	−0.037	−0.021
Objective 2 region	**−0.113**	0.030	**0.089**	−0.001	−0.009	0.012
Farm density	−0.023	0.020	0.003	**−0.059**	0.036	0.022
Farm size	−0.001	−0.002	0.002	−0.008	0.009	−0.001
ρ_j	0.032	0.022	0.000			
McFadden R^2	**0.273**					

Note: Agricultural land is defined as the sum of cropland and grassland (CLC211–242). Summary metrics are based on 10,000 Markov chain Monte Carlo iterations, where the first 5000 were discarded as burn-in. Bold written estimates indicate statistical significance under a 90% credible interval.

agricultural significance. To make our results more robust to this source of uncertainty, we present in Table 4 the estimate from a multinomial logit model, where the dependent variables relate to urban expansion; however, the share calculated is based on only three land-use classes: agricultural, forest and other land. Agricultural land in this sense refers to the sum of cropland and grassland from our original model. The first three columns correspond to direct effects, while the final three correspond to indirect effects.

The estimates from the aggregate land-use model in Table 4 largely correspond to our baseline model in Table 3, albeit with fewer results that are statistically significant. The SAR coefficients are, in fact, insignificant. Moreover, the aggregated rent of cropland and grassland does not seem to play a significant role (although the effect sign remains as in the baseline model). Posterior estimates of direct impacts of Natura 2000 regions also appear insignificant, as opposed to the baseline results. However, having a higher percentage of protected other natural vegetations in neighbouring areas significantly increases the chance of total agricultural land being sealed. Additionally, an increase in farm density – proxying the role of the Common Agricultural Policy – would imply that neighbouring NUTS-3 regions have a significantly decreased chance of sealing total agricultural land.

CONCLUSIONS

In this paper we put forth a Bayesian estimation approach for a multinomial logit specification for the modelling of land-use conversion, which has a SAR structure in the log odds, with a differing strength of spatial autocorrelation for each choice alternative. The virtue of our specification is that it combines an SAR framework (allowing for cross-regional spillovers), and a joint multinomial framework (allowing for cross-land-use class dependencies). The proposed approach is based on recent spatial econometric advances dealing with Bayesian estimation of the logit model (Krisztin & Piribauer, 2020). The core step of the estimation procedure relies on introducing a latent Pólya–Gamma variable (Polson et al., 2013). Through the latent variable, the conditional posterior distribution of the slope parameters in the SAR logit specification is rendered in a Gaussian form, which allows us to tackle the MCMC estimation in a particularly efficient way. We demonstrate in a simulation study the advantages and behaviour of our proposed model specification, benchmarking it against simpler alternatives.

The virtues of the spatial multinomial logit model are illustrated by modelling the land-sealing activities in European NUTS-3 level regions. We consider the areal share of urban, artificial and settlement sprawl emanating from cropland, grassland, forest and other natural vegetation from 2000 to 2018. The data on observed land use stem from the CLC maps. Our results suggest that spatial dependence indeed plays a small, but significant, role, particularly for the land-use classes cropland and grassland. For all land covers proxied, land rents are of central importance. Additionally, our findings corroborate evidence from the recent literature that socio-economic drivers play a much more central role, as opposed to biophysical ones (for an overview, see Shaw et al., 2020). The key role of population density (Guastella et al., 2017; Deng et al., 2008; Lai & Lombardini, 2016) in urban land take is confirmed by our results. Moreover, we confirm on a larger level that environmental protection not only has effects in the own but also in neighbouring regions (Lai & Lombardini, 2016; Lai & Zoppi, 2017; Zoppi & Lai, 2014). Through the virtue of our multinomial analysis, we also find evidence for protected land exerting spillover effects to neighbouring regions and other land covers.

NOTES

[1] Particularly, as demonstrated by Polson et al. (2013), the approach relies on comparatively fewer number of Gibbs sampling steps and latent parameters as compared with alternative

Bayesian approaches, such as the one by Holmes and Held (2006). This allows one to readily extend the standard SAR logit type of models to more flexible specifications.

2 When land-use specific observations y_{ij} are shares, a popular choice for estimating the multinomial model is to apply a log-linear transformation, where the dependent variables correspond to $\log(y_{ij}/y_{iJ})$ and perform standard regression analysis (e.g., Chakir & Lungarska, 2017; Chakir, 2009). The main drawbacks of this approach are twofold. First, Jensen's inequality states that the expectation of a logarithm is not equal to the logarithm of the expectation. Therefore, log-linearization inherently introduces a bias into the estimated slope coefficients. Second, in empirical applications frequently a large number of observed choices y_{ij} are equal to 0, thus necessitating either a censoring of observations or adding a constant to all observations, both of which have been demonstrated to lead to substantial bias.

3 The specification of the log-odds $\boldsymbol{\mu}_j$ may be easily extended by an additional error term in order to capture the spatial dependence structure in a more flexible way (e.g., Krisztin & Piribauer, 2020).

4 With the identifying restriction that the spatial autocorrelation coefficient associated with the J-th land-use class $\rho_J = 0$.

5 The standard SAR model can be extended to more flexible spatial econometric model specifications in a straightforward way. Specifically, one may additionally include spatially lagged explanatory variables, resulting in a so-called spatial Durbin model (SDM) specification (e.g., LeSage & Pace, 2009). A similar extension is presented in the empirical exercise.

6 Convergence of the MCMC algorithm was checked using the convergence diagnostics proposed by Geweke (1992) and Raftery and Lewis (1992). Convergence diagnostics have been calculated using the R package *coda*.

7 As a robustness check, we have also considered an alternative number of nearest neighbour specifications. However, these alternative neighbourhood structures produced highly similar results.

8 Spatial aggregation might lead to the well-known change of support problem (COSP) in general and the modifiable areal unit problem (MAUP) in particular. For a thorough discussion, see Gotway and Young (2002).

9 The Appendix in the supplemental data online contains a detailed aggregation of CLC land cover classes used in the analysis.

10 A total of 361 out of the 1316 regions from the sample, which is roughly 27%, received subsidies under the Objective 2 regional funding programme.

11 Convergence of the sampler was checked using the diagnostics by (Geweke, 1992).

ACKNOWLEDGEMENTS

The authors thank the editor, Jean Paul Elhorst, an anonymous managing editor as well as three anonymous referees for constructive comments that significantly improved the article.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

FUNDING

The research carried out in this paper was supported by funds of the Oesterreichische Nationalbank (Jubilaeumsfond project number 18043) and Austrian Science Fund (FWF) (grant number ZK 35).

ORCID

Tamás Krisztin http://orcid.org/0000-0002-9241-8628
Philipp Piribauer http://orcid.org/0000-0002-3414-3101
Michael Wögerer http://orcid.org/0000-0002-1990-841X

REFERENCES

Alexiadis, S., Hasanagas, N., & Christos, L. (2013). Examining agriculture from a regional perspective: Implications for the common agricultural policy. *Land Use Policy*, *30*(1), 665–669. https://doi.org/10.1016/j.landusepol.2012.05.013

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

Artmann, M. (2014). Assessment of soil sealing management responses, strategies, and targets toward ecologically sustainable urban land use management. *Ambio*, *43*(4), 530–541. https://doi.org/10.1007/s13280-014-0511-1

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. https://doi.org/10.1007/s10666-016-9523-5

Bhat, C. R., & Guo, J. (2004). A mixed spatially correlated logit model: Formulation and application to residential choice modeling. *Transportation Research Part B: Methodological*, *38*(2), 147–168. https://doi.org/10.1016/S0191-2615(03)00005-5

Bhat, C. R., & Sener, I. N. (2009). A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *Journal of Geographical Systems*, *11*(3), 243–272. https://doi.org/10.1007/s10109-009-0077-9

Burnett, P. (2012). Urban industrial composition and the spatial expansion of cities. *Land Economics*, *88*(4), 764–781. https://doi.org/10.3368/le.88.4.764

Carrión-Flores, C. E., Flores-Lagunes, A., & Guci, L. (2018). An estimator for discrete-choice models with spatial lag dependence using large samples, with an application to land-use conversions. *Regional Science and Urban Economics*, *69*, 77–93. https://doi.org/10.1016/j.regsciurbeco.2017.12.005

Chakir, R. (2009). Spatial downscaling of agricultural land-use data: An econometric approach using cross entropy. *Land Economics*, *85*(2), 238–251. https://doi.org/10.3368/le.85.2.238

Chakir, R., & Le Gallo, J. (2013). Predicting land use allocation in France: A spatial panel data analysis. *Ecological Economics*, *92*(3), 114–125. https://doi.org/10.1016/j.ecolecon.2012.04.009

Chakir, R., & Lungarska, A. (2017). Agricultural rent in land-use models: Comparison of frequently used proxies. *Spatial Economic Analysis*, *12*(2–3), 279–303. https://doi.org/10.1080/17421772.2017.1273542

Chakir, R., & Parent, O. (2009). Determinants of land use changes: A spatial multinomial probit approach. *Papers in Regional Science*, *88*(2), 327–344. https://doi.org/10.1111/j.1435-5957.2009.00239.x

Chang-Martínez, L. A., Mas, J. F., Valle, N. T., Torres, P. S. U., & Folan, W. J. (2015). Modeling historical land cover and land use: A review from contemporary modeling. *ISPRS International Journal of Geo-Information*, *4*(4), 1791–1812. https://doi.org/10.3390/ijgi4041791

Coisnon, T., Oueslati, W., & Salanié, J. (2014). Urban sprawl occurrence under spatially varying agricultural amenities. *Regional Science and Urban Economics*, *44*(1), 38–49. https://doi.org/10.1016/j.regsciurbeco.2013.11.001

Delbecq, B. A., Kuethe, T. H., & Borchers, A. M. (2014). Identifying the extent of the urban fringe and its impact on agricultural land values. *Land Economics*, *90*(4), 587–600. https://doi.org/10.3368/le.90.4.587

Deng, X., Huang, J., Rozelle, S., & Uchida, E. (2008). Growth, population and industrialization, and urban land expansion of China. *Journal of Urban Economics*, *63*(1), 96–115. https://doi.org/10.1016/j.jue.2006.12.006

Frühwirth-Schnatter, S., & Frühwirth, R. (2012). Bayesian inference in the multinomial logit model. *Austrian Journal of Statistic*, *41*(1), 27–43. https://doi.org/10.17713/ajs.v41i1.186

Frühwirth-Schnatter, S., Frühwirth, R., Held, L., & Rue, H. (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing*, *19*(4), 479–492. https://doi.org/10.1007/s11222-008-9109-4

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics* (4th ed., pp., 167–193). Oxford University Press.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, *97*(458), 632–648. https://doi.org/10.1198/016214502760047140

Guastella, G., Pareglio, S., & Sckokai, P. (2017). A spatial econometric analysis of land use efficiency in large and small municipalities. *Land Use Policy*, *63*, 288–297. https://doi.org/10.1016/j.landusepol.2017.01.023

Guiling, P., Brorsen, B. W., & Doye, D. (2009). Effect of urban proximity on agricultural land values. *Land Economics*, *85*(2), 252–264. https://doi.org/10.3368/le.85.2.252

Holmes, C. C., & Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, *1*(1), 145–168. https://doi.org/10.1214/06-BA105

Huang, H., Miller, G. Y., Sherrick, B. J., & Gómez, M. I. (2006). Factors influencing Illinois farmland values. *American Journal of Agricultural Economics*, *88*(2), 458–470. https://doi.org/10.1111/j.1467-8276.2006.00871.x

IPBES. (2019). *Global assessment report on biodiversity and ecosystem services*.

Klier, T., & McMillen, D. P. (2008). Clustering of auto supplier plants in the United States: Generalized method of moments spatial logit for large samples. *Journal of Business and Economic Statistics*, *26*(4), 460–471. https://doi.org/10.1198/073500107000000188

Krisztin, T., & Piribauer, P. (2020). A Bayesian spatial autoregressive logit model with an empirical application to European regional FDI flows. *Empirical Economics*. https://doi.org/10.1007/s00181-020-01856-w

Lai, S., & Lombardini, G. (2016). Regional drivers of land take: A comparative analysis in two Italian regions. *Land Use Policy*, *56*, 262–273. https://doi.org/10.1016/j.landusepol.2016.05.003

Lai, S., & Zoppi, C. (2017). The influence of natura 2000 sites on land-taking processes at the regional level: An empirical analysis concerning Sardinia (Italy). *Sustainability (Switzerland)*, *9*(2), 259. https://doi.org/10.3390/su9020259

Leclère, D., Obersteiner, M., Barrett, M., Butchart, S. H., Chaudhary, A., De Palma, A., DeClerck, F. A., Di Marco, M., Doelman, J. C., Dürauer, M., Freeman, R., Harfoot, M., Hasegawa, T., Hellweg, S., Hilbers, J. P., Hill, S. L., Humpenöder, F., Jennings, N., Krisztin, T., … Young, L. (2020). Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature*, *585*(7826), 551–556. https://doi.org/10.1038/s41586-020-2705-y

LeSage, J. P., & Fischer, M. M. (2009). Spatial growth regressions: Model specification, estimation and interpretation. *Spatial Economic Analysis*, *3*(3), 275–304. https://doi.org/10.1080/17421770802353758

LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. CRC Press.

Li, M., Wu, J. J., & Deng, X. (2013). Identifying drivers of land use change in China: A spatial multinomial logit model analysis. *Land Economics*, *89*(4), 632–654. https://doi.org/10.3368/le.89.4.632

Livanis, G., Moss, C. B., Breneman, V. E., & Nehring, R. F. (2006). Urban sprawl and farmland prices. *American Journal of Agricultural Economics*, *88*(4), 915–929. https://doi.org/10.1111/j.1467-8276.2006.00906.x

Lubowski, R. N., Plantinga, A. J., Stavins, R. N., Ruben, N., Plantinga, J., & Robert, N. (2008). What drives land-use change in the United States? A national analysis of landowner decisions. *Land Economics*, *84*(4), 529–550. https://doi.org/10.3368/le.84.4.529

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of econometrics* (pp. 105–142). Academic Press.

McGrath, D. T. (2005). More evidence on the spatial scale of cities. *Journal of Urban Economics*, *58*(1), 1–10. https://doi.org/10.1016/j.jue.2005.01.003

Miller, D. J., & Plantinga, A. J. (1999). Modeling land use decisions with aggregate data. *American Journal of Agricultural Economics*, *81*(1), 180–194. https://doi.org/10.2307/1244459

Paulsen, K. (2012). Yet even more evidence on the spatial size of cities: Urban spatial expansion in the US, 1980–2000. *Regional Science and Urban Economics*, *42*(4), 561–568. https://doi.org/10.1016/j.regsciurbeco.2012.02.002

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Polya–Gamma latent variables. *Journal of the American Statistical Association*, *108*(504), 1339–1349. https://doi.org/10.1080/01621459.2013.829001

Raftery, A. E., & Lewis, S. (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics* (4th ed., pp. 763–773). Oxford University Press.

Salvati, L. (2016). Soil sealing, population structure and the socioeconomic context: A local-scale assessment. *GeoJournal*, *81*(1), 77–88. https://doi.org/10.1007/s10708-014-9608-0

Salvati, L., & Carlucci, M. (2016). The way towards land consumption: Soil sealing and polycentric development in Barcelona. *Urban Studies*, *53*(2), 418–440. https://doi.org/10.1177/0042098014563486

Shaw, B. J., van Vliet, J., & Verburg, P. H. (2020). The peri-urbanization of Europe: A systematic review of a multifaceted process. *Landscape and Urban Planning*, *196*(December 2019), 103733. https://doi.org/10.1016/j.landurbplan.2019.103733

Terama, E., Clarke, E., Rounsevell, M. D., Fronzek, S., & Carter, T. R. (2019). Modelling population structure in the context of urban land use change in Europe. *Regional Environmental Change*, *19*(3), 667–677. https://doi.org/10.1007/s10113-017-1194-5

van Vliet, J. (2019). Direct and indirect loss of natural area from urban expansion. *Nature Sustainability*, *2*(8), 755–763. https://doi.org/10.1038/s41893-019-0340-0

Wu, J. J. (2006). Environmental amenities, urban sprawl, and community characteristics. *Journal of Environmental Economics and Management*, *52*(2), 527–547. https://doi.org/10.1016/j.jeem.2006.03.003

Wu, J. J., & Plantinga, A. J. (2003). The influence of public open space on urban spatial structure. *Journal of Environmental Economics and Management*, *46*(2), 288–309. https://doi.org/10.1016/S0095-0696(03)00023-8

Zoppi, C., & Lai, S. (2014). Land-taking processes: An interpretive study concerning an Italian region. *Land Use Policy*, *36*(2014), 369–380. https://doi.org/10.1016/j.landusepol.2013.09.011