Studying the impact of agricultural subsidies across Europe using a Bayesian spatio-temporal clustering model

Alexander Mozdzen[a], Gregor Kastner[a], Tamas Krisztin[b], Felicity Addo[a]
[a]Alpen-Adria University of Klagenfurt, Klagenfurt, Austria
[b]International Institute for Applied Systems Analysis, Laxenburg, Austria

Abstract

The global climate crisis has conceived the need for impactful policies reducing greenhouse gas emissions across all sources, including emissions stemming from agricultural expansion. In order to study the effectiveness of mitigation policies, statistical methods need to take into account complex biophysical and socio-economic processes. We propose a Bayesian spatio-temporal model for exploring the impact of agricultural subsidies on land usage while simultaneously controlling for other relevant drivers. We combine recent developments in the literature on land use models with a Bayesian nonparametric prior in order to cluster areas which exhibit similar results of the policy in question. We control for individual impacts of essential spatial processes and explicitly model spillovers between regions. Additionally, we develop a suitable Markov chain Monte Carlo (MCMC) algorithm and test the model in an extensive simulation study.

Using European regional data, we investigate the effectiveness of mitigation policies concerning agricultural expansion across Europe and reveal the diversity of the problem. Our model offers a novel approach in understanding the geographical variability in policy impacts, taking into account local environmental, economic, and social conditions. By identifying regions with similar policy outcomes, our study not only aids in assessing current policy effectiveness but also provides valuable insights for future policy formulation. This is particularly crucial in tailoring location-specific strategies that can more effectively address the unique challenges and opportunities in different areas.

The outcomes of our research have wide-reaching implications for policymakers, environmentalists, and agricultural stakeholders. By offering a nuanced understanding of how agricultural subsidies influence land use across varied European regions, our study contributes to the global effort in combating climate change through more informed and effective agricultural policies. Our findings highlight the importance of considering spatial heterogeneity in policy impacts, advocating for more regionalized approaches in environmental planning and management.

Keywords: Bayesian inference, Markov chain Monte Carlo (MCMC), land-use change, European regions

1 Introduction

The European Union's Green Deal represents a pivotal shift in environmental and agricultural policy, with a primary goal of substantially reducing greenhouse gas emissions by 2030. Central to this initiative is the Common Agricultural Policy (CAP), which plays a crucial role in regulating land use and management throughout the EU. Given the diverse impacts of such policies across varied European regions, our study is designed to methodically examine the effectiveness of these initiatives. We aim to identify regions where the CAP has had significant impacts, both positive and negative. This analysis will not only illustrate the varied effects of these policies but also underscore the necessity for region-specific policy evaluations and modifications to enhance their effectiveness.

Advancements in spatial statistics have opened new avenues for analyzing land usage data. Lungarska and Chakir (2018) utilized a spatial Durbin error model to explore the influence of climate change adaptation and mitigation policies on land use changes in France. Their findings underscore the importance of including spatial dynamics in land use analysis. Similarly, Chakir and Le Gallo (2013) developed an aggregated land use share model that emphasizes the necessity of accounting for unobserved individual heterogeneity and spatial autocorrelation to predict future land use accurately. Their research highlights the critical role of spatial information in modeling land-use allocation.

Building on these foundations, the spatial multinomial logit model proposed by Krisztin, Piribauer, and Wögerer (2022) integrates spatial information through a spatial autoregressive (SAR) component within a multinomial framework. Utilizing a Bayesian approach, they focus on modeling land-sealing activities across European

NUTS-3 level regions. Our study aims to expand upon this model to investigate the uniformity and effectiveness of mitigation policies across various European regions using this advanced methodological approach.

In modeling binary or categorical data, the selection of an appropriate link function, such as a probit or logit, is crucial. The logistic regression model, employing the logit function, has been successfully applied in various contexts, as evidenced by studies like (Chakir and Le Gallo, 2013) and (Krisztin et al., 2022). In a Bayesian framework, researchers such as Ferguson (1983) and Holmes and Held (2006) have innovatively applied data augmentation strategies. These strategies effectively convert the probability of observing one of J categories into a utility-based representation, enhancing the model's applicability and interpretability.

The multinomial logit model enriches the classical regression model by introducing the dimension of categories to be analyzed. This addition opens up opportunities for more versatile modeling options, such as the conditional logit model. In this model, as described by (Cameron and Trivedi, 2005), some regressors vary across alternatives, offering a more nuanced understanding of the effects of different variables on the outcomes. Our study seeks to harness these advanced statistical techniques to provide a comprehensive and insightful analysis of EU agricultural policies and their regional impacts.

2 Methodology

Our primary objective is to enhance the spatial multinomial logit model by incorporating a Bayesian nonparametric prior, specifically the Dirichlet process (DP), into the regression parameters. This approach follows the footsteps of Mozdzen, Cremaschi, Cadonna, Guglielmi, and Kastner (2022), who utilized a similar method to examine unemployment rates in Italy. Their technique introduces model-based clustering, which offers several key advantages. Primarily, clustering the parameters leads to a more parsimonious model, simplifying both estimation and interpretation processes. Furthermore, the flexibility of the DP allows us to tailor the model to our specific research question. By employing the DP as a prior for the coefficients of a particular explanatory variable, we can discern the shared impact of this variable on different areal units in relation to our dependent variable, namely, the land use share of cropland, grassland, forest, or other natural land. The main explanatory variable under investigation is agricultural subsidies, sourced from the farm accountancy data network (FADN).

To assess the impact of various covariates on the different categories of land use, we implement a multinomial logit model. Traditional Bayesian methods for estimating this model, such as those used by Holmes and Held (2006), involve auxiliary variables and utility representations. A significant advancement in this area is the seminal work of Polson, Scott, and Windle (2013), which introduced auxiliary Polya-gamma distributed random variables. This innovative augmentation converts the likelihood function in terms of β to be proportional to a multivariate normal distribution, thus enabling conjugate posterior sampling.

Our analysis centers on a multinomial response variable, observed across $i = 1, .., N$ areal units, $t = 1, ..., T$ time points, and $j = 1, ..., J$ categories. The data for each time point t for unit i across all categories is represented by the vector $y_{i,t} \in (0, 100)^J$. The likelihood of a single observation in the multinomial logistic model is expressed as:

$$p(y_{i,t}|\boldsymbol{x}_{i,t}, \boldsymbol{\beta} i, j) = \prod j = 1^J \left(\frac{\exp(\boldsymbol{x}_{i,t}^T \boldsymbol{\beta} i, j)}{\sum j = 1^J \exp(\boldsymbol{x}_{i,t}^T \boldsymbol{\beta} i, j)} \right)^{yi,t,j} \tag{1}$$

Here, $\boldsymbol{x}_{i,t} \in \mathbb{R}^{K+1}$ includes an intercept and K explanatory variables for unit i at time t, while $\boldsymbol{\beta}_{i,j} \in \mathbb{R}^{K+1}$ denotes the coefficient vector for unit i and category j.

Furthermore, the likelihood over all time points and categories is formulated as:

$$p(\boldsymbol{y}_{i,\cdot}|\boldsymbol{x}_{i,t}, \boldsymbol{\beta} i, j) = \prod t = 1^T \prod_{j=1}^{J} \left(\frac{\exp(\boldsymbol{x}_{i,t}^T \boldsymbol{\beta} i, j)}{\sum j = 1^J \exp(\boldsymbol{x}_{i,t}^T \boldsymbol{\beta} i, j)} \right)^{yi,t,j} \tag{2}$$

Given the i.i.d. assumption on $y_{i,t,j}$, the order of products in the likelihood expression can be interchanged as appropriate.

Assigning a unique set of β coefficients to each areal unit could lead to an excessive number of parameters, complicating estimation and interpretation, and increasing the risk of overfitting. Conversely, estimating a single β coefficient set for all units might oversimplify the model's representation of covariate effects on land use categories. Our goal is to find a middle ground by clustering areal units based on the similarities in their β coefficients. We utilize a mixture model with covariates in the cluster-specific kernels for this purpose. Similar approaches have been successfully implemented in studies by Greenwade (1993), Ferguson (1983) using MCMC in a Bayesian setting. The R-package `multinomialLogitMix`, introduced by Papastamoulis (2023), facilitates both EM and MCMC estimation for finite mixtures of multinomial distributions, with or without covariates.

In this study, we choose the Dirichlet Process mixtures (DPM), a popular nonparametric alternative to finite mixtures, first described in Lo (1984) and Ferguson (1983). Despite their theoretical complexity, DPMs benefit from analogous representations of the DP, such as the Chinese restaurant process and the Stick-Breaking construction (Sethuraman, 1994a), which allow for sophisticated sampling algorithms.

Various computationally efficient representations of the DP, like the Chinese restaurant process and the Stick-Breaking construction (Sethuraman, 1994b), have led to the development of effective sampling mechanisms. Key examples include the MCMC algorithms by Neal (2000), MacEachern and Müller (1998), Escobar and West (1995), and the variational inference algorithm by Blei and Jordan (2006). Moreover, Miller and Harrison (2018) demonstrated that these advantageous properties are applicable within the mixtures of finite mixtures framework, thus enabling the use of these effective nonparametric algorithms in finite contexts.

References

Blei, D. M. and M. I. Jordan (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis 1*(1), 121 – 143.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: methods and applications.* Cambridge university press.

Chakir, R. and J. Le Gallo (2013). Predicting land use allocation in France: A spatial panel data analysis. *Ecological Economics 92*, 114–125. Land Use.

Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association 90*(430), 577–588.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In M. H. Rizvi, J. S. Rustagi, and D. Siegmund (Eds.), *Recent Advances in Statistics*, pp. 287–302. Academic Press.

Greenwade, G. D. (1993). The Comprehensive Tex Archive Network (CTAN). *TUGBoat 14*(3), 342–351.

Holmes, C. C. and L. Held (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis 1*(1), 145–168.

Krisztin, T., P. Piribauer, and M. Wögerer (2022). A spatial multinomial logit model for analysing urban expansion. *Spatial Economic Analysis 17*(2), 223–244.

Lo, A. Y. (1984). On a class of bayesian nonparametric estimates: I. density estimates. *The Annals of Statistics 12*(1), 351–357.

Lungarska, A. and R. Chakir (2018). Climate-induced land use change in France: Impacts of agricultural adaptation and climate change mitigation. *Ecological Economics 147*, 134–154.

MacEachern, S. N. and P. Müller (1998). Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics 7*(2), 223–238.

Miller, J. W. and M. T. Harrison (2018). Mixture models with a prior on the number of components. *Journal of the American Statistical Association 113*(521), 340–356.

Mozdzen, A., A. Cremaschi, A. Cadonna, A. Guglielmi, and G. Kastner (2022). Bayesian modeling and clustering for spatio-temporal areal data: An application to Italian unemployment. *Spatial Statistics 52*, 100715.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics 9*(2), 249–265.

Papastamoulis, P. (2023, Jul). Model based clustering of multinomial count data. *Advances in Data Analysis and Classification*.

Polson, N. G., J. G. Scott, and J. Windle (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association 108*(504), 1339–1349.

Sethuraman, J. (1994a). A constructive definition of Dirichlet priors. *Statistica Sinica 4*(2), 639–650.

Sethuraman, J. (1994b). A constructive definition of Dirichlet priors. *Statist. Sinica 4*(2), 639–650.