Central European Journal of Economic Modelling and Econometrics

Accounting for Spatial Heterogeneity of Preferences in Discrete Choice Models

Wiktor Budziński* and Mikołaj Czajkowski†

Submitted: 9.08.2019, Accepted: 6.10.2020

Abstract

There are reasons researchers may be interested in accounting for spatial heterogeneity of preferences, including avoiding model misspecification and the resulting bias, and deriving spatial maps of willingness-to-pay (WTP), which are relevant for policy-making and environmental management. We employ a Monte Carlo simulation of three econometric approaches to account for spatial preference heterogeneity in discrete choice models. The first is based on the analysis of individual-specific estimates of the mixed logit model. The second extends this model to explicitly account for spatial autocorrelation of random parameters, instead of simply conditioning individual-specific estimates on population-level distributions and individuals' choices. The third is the geographically weighted multinomial logit model, which incorporates spatial dimensions using geographical weights to estimate location-specific choice models. We analyze the performance of these methods in recovering population-, region- and individual-level preference parameter estimates and implied WTP in the case of spatial preference heterogeneity. We find that, although ignoring spatial preference heterogeneity did not significantly bias population-level results of the simple mixed logit model, neither individual-specific estimates nor the geographically weighted multinomial logit model was able to reliably recover the true region- and individual-specific parameters. We show that the spatial mixed logit proposed in this study is promising and outline possibilities for future development.

Keywords: discrete choice experiment, discrete choice models, individual-, region- and population-level parameter estimates, spatial preference heterogeneity

JEL Classification: C31, C25, Q51

*University of Warsaw; e-mail: wbudzinski@wne.uw.edu.pl; ORCID: 0000-0001-9866-4267
†University of Warsaw; e-mail: mc@uw.edu.pl; ORCID: 0000-0001-5118-2308

1 Introduction

Preferences for environmental goods may follow spatial patterns. This becomes an important issue for discrete choice models, commonly used for modeling consumers' preferences and willingness-to-pay (WTP; Carson and Czajkowski, 2014; Hanley and Czajkowski, 2019). First, ignoring any important source of preference heterogeneity may lead to model misspecification and result in biased estimates. However, because modern developments in geographical information systems (GIS) allow researcher to easily combine them with individuals' locations (e.g., zip-codes) and provide detailed information about the spatial configuration of environmental goods, it is now easy to control for spatial patterns in stated and revealed preferences, such as spatial autocorrelation. Second, explicitly accounting for spatial dependencies is useful for policy-making and environmental management, for example by allowing derivation of detailed spatial maps of WTP.

These reasons spark increasing interest in using location-specific references or improvement levels of choice attributes, or including location-specific characteristics as explanatory variables of preferences in discrete choice models, mostly in stated preference setting (e.g., Campbell et al. 2008; Campbell et al. 2009; Hynes et al. 2010; Johnston et al. 2011). More importantly, however, parametric methods of accounting for spatial heterogeneity are being developed, which allow for uncovering spatial patterns that are otherwise difficult to attribute to any characteristic that is observable and available in the data (e.g., Johnston and Ramachandran, 2014).

The most common parametric way to account for spatial preference heterogeneity in discrete choice models is to include a two-step procedure, in which the Mixed Logit (MXL) model is estimated and individual-specific parameter estimates (conditional on respondents' choices) are derived; they are then used for spatial analysis, such as the spatial lag model, the spatial error model, or kriging (e.g., Abildtrup et al. 2013; Broch et al. 2013a; Yao et al. 2014; Czajkowski et al. 2017). This approach can be extended to explicitly allow random parameters to follow a spatial lag process. Such a spatial mixed logit (S-MXL) model is proposed in this study. Finally, the third approach considered in this study is the geographically weighted multinomial logit model (GW-MNL; Budziński et al. 2018). This is an extension of the geographically weighted regression, in which spatial dimension is incorporated using geographical weights to estimate location-specific models (Fotheringham et al. 1998) of discrete choice data.

In this study, we provide an overview of the three methods, reviewing their advantages and limitations, and employ a Monte Carlo simulation to investigate their performance in the case of spatial preference heterogeneity. We evaluate the models' performances in terms of the ability to correctly recover population-, region- and individual-level preferences and WTP.

Section 2 presents the three methods mentioned above with more detail and mathematical rigor. Technical details regarding the S-MXL model proposed here and its estimation are provided in Appendix A. Section 3 describes and justifies the data generating process used in the Monte Carlo simulation. Section 4 presents results in terms of model bias when recovering population-level preferences and WTP, individual-specific parameters, and region-specific estimates. The last section concludes.

2 Methods

We use a Monte Carlo simulation to compare three models that may be applied to discrete choice data with spatial dimension of preference heterogeneity. We start the description of these models with a standard mixed logit (MXL) model, and then follow it with a novel extension, a spatial mixed logit (S-MXL) model. In the last part, we describe a geographically weighted multinomial logit (GW-MNL) model. The software codes for estimating the models presented here were developed in Matlab and are available from `http://github.com/czaj/DCE` under Creative Commons BY 4.0 license. The simulation data and supplementary materials are available from `http://czaj.org/research/supplementary-materials`. We use maximum likelihood method to estimate MXL and GW-MNL models, whereas S-MXL is estimated with Bayesian inference. We use a Bayesian techniques to facilitate estimation, and avoid numerical issues related to spatially correlated random effects.

2.1 Mixed logit model

The theoretical foundation for the discrete choice model is random utility theory, which assumes that the utility a person derives depends on observed characteristics and unobserved idiosyncrasies, represented by a stochastic component (McFadden, 1974). As a result, individual i's utility resulting from choosing alternative j in choice set t can be expressed as:

$$U_{ijt} = \boldsymbol{\beta}_i \mathbf{X}_{ijt}^{\text{non-cost}} - \alpha_i X_{ijt}^{\text{cost}} + \varepsilon_{ijt}. \tag{1}$$

In what follows we use NP to denote number of people and NCT to denote number of choice tasks, namely $i \in \{1, 2, ..., NP\}$ and $t \in \{1, 2, ..., NCT\}$. In the simulation that follows we assume $NP = 1000$, $NCT = 6$ and number of alternatives equal to 3. In the mixed logit model, it is assumed that each individual i has a separate, independent set of parameters, $\boldsymbol{\beta}_i$ and α_i. Assuming extreme value distribution for the error component, ε_{ijt}, leads to a well-known formulation of conditional likelihood in a multinomial logit form. As individual-specific parameters are not directly observed, a distribution for them needs to be assumed and integrated. Such unconditional likelihood can then be used for estimation of parameters describing the distribution of random effects; for example, their means and variances, which we denote jointly by a vector of parameters, Ω.

Estimation of MXL is usually performed using the maximum simulated likelihood (MSL) method, and this is also the case in the present study. Let denote by $f(\alpha_i, \boldsymbol{\beta}_i \mid \Omega)$ a joint density function of the random parameters in (1). In MSL method, we take R quasi-random draws from this distribution to approximate a likelihood function. (For our simulation, we employ $R = 10,000$ draws from a scrambled Sobol sequence (Czajkowski and Budziński, 2015)). Denoting by y_{ijt} a binary variable, which equals 1 if individual i has chosen alternative j in choice situation t, probability of choosing alternative j, $p(y_{ijt} = 1 \mid \mathbf{X}_i, \Omega, \alpha_i, \boldsymbol{\beta}_i)$, is given by multinomial logit formula. As random parameters are unobserved, and each individual is assumed to make NCT choices, likelihood function and its approximation are given by

$$\begin{aligned} L_i &= \int \left(\prod_{t=1}^{NCT} \sum_{j=1}^{3} y_{ijt} p(y_{ijt} = 1 \mid \mathbf{X}_i, \Omega, \alpha_i, \boldsymbol{\beta}_i) \right) f(\alpha_i, \boldsymbol{\beta}_i \mid \Omega) \, d(\alpha_i, \boldsymbol{\beta}_i) \approx \\ &\approx \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{NCT} \sum_{j=1}^{3} y_{ijt} p(y_{ijt} = 1 \mid \mathbf{X}_i, \Omega, \alpha_i^r, \boldsymbol{\beta}_i^r), \end{aligned} \tag{2}$$

where α_i^r, $\boldsymbol{\beta}_i^r$ denote draws from the assumed distribution. Logarithm of likelihood function in (2) is then maximized with respect to parameters in Ω.

Throughout the study we assume that individual-specific parameters for non-cost attributes, $\boldsymbol{\beta}_i$, follow normal distributions, whereas the individual-specific parameter for cost, α_i, follows a log-normal distribution. Although the individual specific parameters are not observed by the researcher, it is possible to estimate their values as implied by each respondents' choices conditional on the population-level estimates of parameter distributions (Bayesian posterior means) using the Bayes theorem. We note, that even though formula (3) employs a Bayes theorem, the model is estimated with MSL method, and therefore, we treat model parameters, Ω, as non-random, substituting the values obtained from the estimation procedure. We will focus on predicting individual-specific marginal WTP with the following formula

$$E\left(\frac{\boldsymbol{\beta}_i}{\alpha_i} \mid \mathbf{y}_i, \mathbf{X}_i, \Omega\right) = \int \frac{\boldsymbol{\beta}_i}{\alpha_i} \frac{p(\mathbf{y}_i \mid \mathbf{X}_i, \Omega, \alpha_i, \boldsymbol{\beta}_i) f(\alpha_i, \boldsymbol{\beta}_i \mid \Omega)}{p(\mathbf{y}_i \mid \mathbf{X}_i, \Omega)} d(\alpha_i, \boldsymbol{\beta}_i), \tag{3}$$

where $\mathbf{y}_i$ is a vector stacking all y_{ijt} variables, $p(\mathbf{y}_i \mid \mathbf{X}_i, \Omega, \alpha_i, \boldsymbol{\beta}_i)$ is the likelihood of individual i making the observed choices conditional on the values of random parameters, and $p(\mathbf{y}_i \mid \mathbf{X}_i, \Omega)$ is the same likelihood but unconditional (random parameters are integrated out, so it is equal to L_i in (2)). As model in (1) has linear form, marginal WTP is a ratio of parameter for a given attribute and a parameter for cost (marginal rate of substitution). For more details about this approach and examples of its applications, refer to Czajkowski et al. (2017), Abildtrup et al. (2013), Broch et al. (2013b) and Yao et al. (2014).

Note that the MXL model does not assume any spatial dependence in its specification. Any spatial effect we observe in individual-specific WTP obtained from formula (3) is due to conditioning on the vector of observed choices, $\mathbf{y}_i$.

2.2 Spatial mixed logit

In this study, we propose a novel extension of the MXL model, which directly accounts for spatial dependencies in preference heterogeneity. Although S-MXL follows the same utility specification as described in (1), the difference arises due to the specification of the distribution of random parameters. In standard MXL, it is assumed that parameters are independent among individuals; this is not the case here. In this specification, we follow Smith and LeSage (2004). Let us define random parameters in S-MXL as

$$\beta_{ik} = \mu_k + \theta_{ik}, \tag{4}$$

if the k-th attribute is not a cost ($k < K$), and as

$$\alpha_i = exp(\mu_K + \theta_{iK}), \tag{5}$$

for the cost attribute ($k = K$). In this model, μ_k are parameters to estimate (means of random parameters), whereas θ_{ik} are normally distributed auxiliary variables, used to incorporate preference heterogeneity with spatial dependence. Similarly, as in Train and Sonnier (2005) any distribution that is a transformation of a normal distribution can be specified here, but in this simulation, we limit ourselves to the most common specifications of normal and log-normal distributions. Auxiliary variables, θ_{ik}, are defined as:

$$\theta_{ik} = \rho_k \sum_{m=1}^{NP} w_{im}\theta_{mk} + u_{ik}, \tag{6}$$

where $u_{ik} \sim N(0, \sigma_2^k)$. We assume that u_{ik} are independent among individuals and attributes. The latter assumption is likely to be too restrictive, as it implies that there is no correlation between different random parameters, and therefore no scale heterogeneity or correlation of tastes (Train, 2009; Hess and Train, 2017). The model can be extended to incorporate such correlation, but this is beyond the scope of this study.

Equation (6) describes a spatial autoregressive process, which can be written in vectorized form as $\boldsymbol{\theta}_{\cdot k} = \mathbf{B}_{\rho_k}^{-1}\mathbf{u}_{\cdot k}$, where $B_{\rho_k} = I - \rho_k W$, and W is a spatial weight matrix, whose rows sum to 1. Throughout the study, we use the inverse squared distance as a weight matrix, which is a standard approach in spatial econometrics. From this it follows that $\boldsymbol{\theta}_{\cdot k}$ follows multivariate normal distribution, namely $\boldsymbol{\theta}_{\cdot k} \sim MVN\left(0, \sigma_k^2 \left(B'_{\rho_k} B_{\rho_k}\right)^{-1}\right)$. Note that in this model we assume that random parameters are spatially auto-correlated, rather than utilities or choices.

Note that in general this model cannot be estimated by MSL. This is because of the non-zero correlation between choices of all individuals. In MSL, the researcher would have to calculate the product of conditional probabilities of choices across all individuals and choice tasks. If the number of individuals is large, such a product would become effectively equal to 0 due to numerical precision. This issue can be, to some extent, resolved if the particular specification of W is assumed; for example, k

nearest neighbors. In this study, to test a more general case, we estimate the model using Bayesian inference, based on Train and Sonnier (2005) and Smith and LeSage (2004). The detailed specification of the model and its estimation details are provided in Appendix A.

Individual-specific estimates of WTP (analogous as in (3)) are easy to obtain from S-MXL, as we can simply use saved draws from posterior distributions to calculate them. Let us denote as μ_k^n, μ_K^n, θ_{ik}^n and θ_{iK}^n the n-th draws from the conditional posterior distribution of μ_k, μ_K, θ_{ik} and θ_{iK} , respectively. Then, having $R = 10,000$ draws generated with Markov Chain Monte Carlo method (consult Appendix A) for each of these variables, the mean WTP of individual i for the t-th attribute can be calculated as

$$E(WTP_{ik}) \approx \frac{1}{R}\sum_{n=1}^{R}\frac{\beta_{ik}^n}{\alpha_i^n} = \frac{1}{R}\sum_{n=1}^{R}\frac{\mu_k^n + \theta_{ik}^n}{exp(\mu_k^n + \theta_{iK}^n)}. \tag{7}$$

We calculate median WTP analogously.

To our knowledge, this specification of the model has never been used before for any application aimed at uncovering spatial heterogeneity of preferences.

2.3 Geographically weighted multinomial logit

The geographically weighted multinomial logit model (GW-MNL; Budziński et al. 2018) is an extension of the geographically weighted regression (Fotheringhamet al. 1998) of discrete choice data, in which the spatial dimension is incorporated using geographical weights to estimate location-specific models. The rationale of this approach is that if spatial clusters of preferences do exist, a locally-weighted maximum likelihood method can be used to account for spatial autocorrelation or other spatial patterns of preferences and welfare measures. Because this is a semi-parametric approach, no *a priori* assumptions about the spatial distribution of preferences are necessary.

The utility function in GW-MNL is defined analogously as in (1), with the only difference being that the parameters are now location-specific rather than individual-specific, and therefore, they are indexed by l. If the number of locations is the same as the number of individuals, then the two specifications are the same (which is the case in our simulation).

$$U_{ijt} = \boldsymbol{\beta}_l \mathbf{X}_{ijt}^{\text{non-cost}} - \alpha_l X_{ijt}^{\text{cost}} + \varepsilon_{ijt}. \tag{8}$$

The assumption that allows for the estimation of GW-MNL is that individuals located close to each other have more similar preference parameters than individuals located far away from each other. As a result, the parameters become spatially correlated.

The estimation is conducted by estimating L 'local' models, where L is a number of distinct locations. In the case of our simulation, it will be equal to the number of individuals, so there will be a separate local model estimated for each individual.

Each local model is estimated via the weighted maximum likelihood method. The likelihood of the choices of individual i, assuming the l-th local model, can then be written as

$$L_i^l = \prod_{t=1}^{NCT} \frac{\exp\left(\boldsymbol{\beta}_l \mathbf{X}_{iy_{it}t}^{\text{non-cost}} - \alpha_l X_{iy_{it}t}^{\text{cost}}\right)}{\sum_j \exp\left(\boldsymbol{\beta}_l \mathbf{X}_{ijt}^{\text{non-cost}} - \alpha_l X_{ijt}^{\text{cost}}\right)}. \tag{9}$$

The weighted log-likelihood for the l-th model is defined as follows:

$$WL^l = \sum_{i=1}^{NP} \lambda(Lat_i, Long_i, b, l) log\left(L_i^l\right), \tag{10}$$

where $\lambda(Lat_i, Long_i, b, l)$ is a geographical weight (kernel) that depends on the latitude and longitude of individual i's location, b which is called the 'bandwidth parameter' and the location l for which the local model is estimated. (Note that geographically weighted models normally use projected data, with the location given as metric coordinates X and Y (easting and northing), to avoid the complex and computationally time-consuming 3D calculation of geographic distance with the two angular coordinates (latitude and longitude). Nevertheless, this distinction is not relevant for current study, as in simulation we assume that each individual is placed within flat square). There are a few functional forms of $\lambda(\cdot)$ proposed in the literature. In what follows, we use the Gaussian kernel (for other possible kernels, see Fotheringham et al. 2003) defined as:

$$\lambda(Lat_i, Long_i, b, l) = exp\left(-0.5\frac{(Lat_i - Lat_l)^2 + (Long_i - Long_l)^2}{b^2}\right). \tag{11}$$

This is simply an exponential function of negative half of the squared Euclidean distance of individual i's location from location l divided by the square of the bandwidth parameter. If the individual lives exactly in location l, this weight is equal to 1. The use of this weight implies the clustering of similar values because observations near location l have a larger bearing on the local model's log-likelihood compared to observations that are further away. The bandwidth parameter therefore determines what "further away" means. If the bandwidth is low, then practically, only the observations in very close proximity of the given location influence the local model. Specifically, when $b \to 0$, each local model is estimated using observations only from the given location. Analogously, when bandwidth is high, all local models will have similar parameter estimates, with $b \to \infty$ leading to a simple MNL model for the whole sample.

In GW-MNL, individual-specific WTP can be calculated simply with marginal WTP for each local model, i.e. $WTP_l = \beta_l/\alpha_l$, and then assigning them to the individuals who live in those locations.

For more details about this approach and an example of its applications, see Budziński et al. (2018).

The main differences between GW-MNL and S-MXL models are that the former is a semi-parametric approach where the only assumption regarding preference heterogeneity is that individuals living closer have more similar preferences. Therefore, GW-MNL can theoretically recover a wide range of preference distributions, even if they are non-standard ones. On the other hand, preference heterogeneity in this model is mostly driven by the spatial dimension, and, therefore, this model is likely to ignore other sources of preference heterogeneity that may occur in the data. In contrast, the S-MXL model makes a relatively strong assumption about parametric distribution of the preferences (although assuming normal and log-normal distribution is a standard in discrete choice modelling literature), but can account for both spatial and non-spatial sources of heterogeneity. The spatial autocorrelation equation in (6) is used to obtain a distribution of preferences with spatial dependence $\left(\boldsymbol{\theta}_{\cdot k} \sim MVN\left(0, \sigma_k^2 \left(B'_{\rho_k} B_{\rho_k}\right)^{-1}\right)\right)$.

We do not postulate that there is some causality involved in preference formation, for example, individuals forming their tastes by observing behavior of other individuals residing nearby. We use equation (6) to formulate the probabilistic model in which individuals located closer to each other have higher probability of having similar preference parameters than individuals located far away from each other. There may be various reasons for this dependence such as residential sorting, availability of substitute goods or unobserved covariates which are spatially auto-correlated on their own. The same reasons are valid for the GW-MNL and the S-MXL model.

3 Simulation setting

Our simulation setting was aimed at recovering individual preferences in the case of the spatial dependencies in preference heterogeneity. Even though we expect the S-MXL model to perform best, as it is fully consistent with the data generating process, this model is relatively more complex and demanding in terms of estimation time (approximately 1 hour, compared to 1 minute for MXL and 40 minutes for all GW-MNL models). It is therefore useful to test if the alternative approach to uncovering spatial patterns (the GW-MNL model) or using individual-specific estimates from the MXL that ignores spatial dimension perform reasonably well.

The models will be compared using three different measures of performance. First, we compare the characteristics of the WTP distribution implied by the estimated models, such as WTP mean and standard deviation. Here, we are not directly interested in the spatial dimension of preferences, but we rather focus on obtaining a good description of the WTP distribution in the population. Second, we compare individual-specific WTP estimates with their true values and calculate mean absolute percentage error (MAPE). This analysis will allow us to conclude whether individual-specific estimates can be used for valid inference. Last, we will compare the models' "regional" predictions based on their individual-specific WTP estimates. Specifically,

we will divide the simulated area into nine squares of equal sizes, and predict mean WTP for each of them using our models. Then, we will compare these predictions with the true mean regional values and calculate absolute percentage errors. This measure will provide us with information about how useful these methods are when it comes to using national level samples for more "regional" analysis.
The data generating process assumes the sample of $1,000$ individuals, with each individual making six choices and each choice consisting of three alternatives. Simulated data has therefore a panel structure, as is usually the case in stated preference setting employing the discrete choice experiment method. Alternatives differ only in terms of two attributes (so $K = 2$), denoted here as *Quality* and *Cost*, which follow uniform distributions on $[0, 2]$ interval. For each alternative those attributes take different values, and it is assumed that individuals choose the best alternative for themselves from three such alternatives at the time. Such a simple setting leads to only one vector of WTP (one value for each individual), which we will use to evaluate the performance of the approaches presented in the previous section. Specifically, the utility function that drives individuals' choices is given by:

$$U_{ijt} = \beta_i \text{Quality}_{ijt} - \alpha_i \text{Cost}_{ijt} + \varepsilon_{ijt}. \tag{12}$$

Quality and *Cost* attributes are assumed to be observed by the researcher and therefore they are used in the estimation as the independent variables. On the other hand, the stochastic term, ε_{ijt}, follows a standard extreme value distribution, β_i follows a normal distribution, and α_i follows a log-normal distribution. The values of ε_{ijt}, β_i and α_i are unobserved by the researcher. In the simulation setting we assume to know distribution of these parameters. In the real-life case study, these assumptions would have to additionally be tested. Spatial autocorrelation coefficient was assumed to be equal to 0.6 for both parameters. We also conducted simulations in which ρ for the *Quality* attribute took values 0.2 or 0.9 to analyze different degrees of spatial autocorrelation. The obtained results were qualitatively the same, although we generally observed higher errors for all models when spatial dependence was stronger. These results are available from the authors upon request. The spatial weight matrix, W, is calculated using an inverse squared distance, with sum of rows normalized to 1. The values of parameters used in the data generating process are summarized in Table 1. Having utility function calculated for each individual, choice and alternative, we then generated the dependent variable, which would take value of 1 if given alternative had the highest utility in a given choice set, and 0 otherwise. We assumed the individuals' locations were distributed uniformly on a 10×10 square. An example of spatial distribution of WTP is presented in Figure 1. The spatial dependence is not straightforward to see, as there is still significant variation on the local level, but Moran's I statistic is actually equal to 0.67 in this case. Moran I statistic was calculated using the same spatial weights matrix that was assumed in the data generating process, namely inverse squared distance. This data generating process may correspond to a situation where spatial dependence is only one of

Table 1: True values of parameters used in the Data Generating Process of the Monte Carlo simulation

	μ_k	σ_k^2	ρ_k
Quality	3	1	0.6
Cost	−1	1	0.6

the factors driving the preference heterogeneity, which is likely to be the case in a real-world scenario. Nevertheless, significant spatial autocorrelation means that individuals located near each other are more likely to have similar preferences and values of WTP.

The simulation uses 100 artificial samples according to the described data generating process. We then estimate MXL, S-MXL, and GW-MNL models for each sample, with 40 GW-MNL models for each sample (the GW-MNL models differed in the value of bandwidth parameters, ranging from 0.05 to 2 with 0.05 increases). For GW-MNL, Lat_i and $Long_i$ were defined as coordinates of locations, as presented in Figure 1. We note as a reminder that the models are estimated with different procedures; MXL uses maximum simulated likelihood method with quasi Monte Carlo 'draws' (Czajkowski and Budziński, 2019), S-MXL uses the Bayesian estimation procedure, and GW-MNL employs a locally-weighted maximum likelihood method.

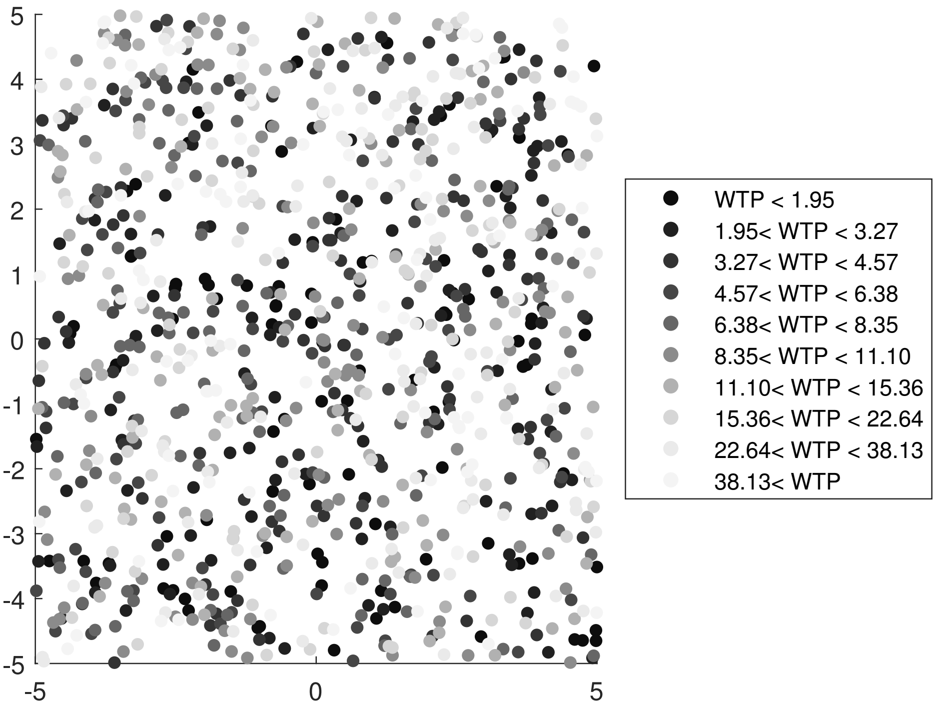
Table 2: Comparison of selected moments and quantiles of the recovered willingness-to-pay distribution with their true values implied by the data generating process

	Mean	Std. Dev	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
True	15.4547	31.4383	1.6151	3.4386	7.6339	16.693	33.8052
S-MXL	16.5629	39.1574	1.6188	3.4655	7.7824	17.3120	35.7671
	(4.8087)	(25.4946)	(0.2094)	(0.4327)	(1.2865)	(3.9552)	(10.4885)
MXL	16.3096	30.0706	1.5780	3.4112	7.8236	17.7638	37.1703
	(4.6023)	(14.7791)	(0.1964)	(0.4141)	(1.2556)	(3.9752)	(10.7712)
GW-MNL (min. AIC)	−0.0191	302.5630	2.4358	3.2610	4.5097	6.7887	11.3977
	(55.1850)	(1617.1076)	(0.4508)	(0.4026)	(0.4681)	(0.8828)	(2.3929)
GW-MNL (min. MAPE)	4.7185	9.8760	2.2433	3.3625	4.3330	5.5394	7.4295
	(1.4703)	(26.0672)	(3.4068)	(1.6781)	(1.0085)	(0.8548)	(1.5778)

4 Results

Table 2 provides the characteristics of WTP, as implied by the data generating process and the estimated models. Each characteristic is simulated, as the WTP distribution is non-standard (normal random variable divided by log-normal random variable; in the case of true values and S-MXL, they are additionally spatially autocorrelated) with unknown analytic formulas. For each of 100 generated datasets we calculated

Figure 1: Example of spatial distribution of WTP in the data generating process



the implied characteristics of WTP for each model. Table 2 reports their means and standard errors (in brackets).

In the case of GW-MNL, it was not clear which value of bandwidth parameter should be selected. There are several methods available in the literature, such as the corrected Akaike information criterion (AIC, Dekker et al. 2014), taking the lowest bandwidth at which all local models converge (Dekker et al. 2014), a leave-one-individual-out cross-validation criterion (Fotheringham et al. 2003), or simply "eye-balling" as suggested by Koster and Koster (2015). We compare the bandwidth chosen based on the AIC method, with the bandwidth that minimizes MAPE for individual-specific WTP. The latter should result in the best possible performance of the GW-MNL, as it is optimized by considering the true values of the parameters. In a real-life situation, however, it would not be possible to use this method, as true individual-specific WTP would be unknown.

The results presented in Table 2 show that S-MXL and MXL provide estimates that are close to each other, and at the same time, close to the true characteristics. The only difference is in the standard deviation, which S-MXL seems to overestimate. As percentiles of the distribution are well recovered, and standard deviation is not, this would imply that there are some outliers within the 100 estimated models. We attempted to re-estimate the models that implied values of the standard deviations of WTP that were too large using different starting values and a greater number of iterations in Gibbs sampler, but this did not qualitatively change the results. It may be the case that a more sophisticated Bayesian algorithm needs to be used, such as the Hamiltonian Monte Carlo (Gelman et al. 2014), or that some other candidate function could be considered in Metropolis-Hastings algorithm instead of normal distribution. For example, LeSage (1999) proposes to use t-student distribution. It is also possible, that the complexity of the model makes it difficult to properly recover the variance of preferences, as σ_k and ρ_k parameters can both influence it.

In the case of GW-MNL, the results depend on the choice of bandwidth parameter. If the bandwidth is chosen to minimize AIC (on average $b = 0.47$), means and standard deviations are highly biased because some local models did not converge, resulting in unreasonably high or low values of local parameters. On the other hand, estimates of percentiles are too high for low percentiles, and too low for higher ones. This implies that GW-MNL cannot provide a proper estimate of the distribution of WTP. This is likely because this model only considers spatial dimension of data, ignoring any other possible sources of heterogeneity. When the bandwidth is chosen to minimize the MAPE of individual-specific WTP (on average $b = 1.01$), the means and standard deviations are less biased, but the percentile estimates are similar to the ones from GW-MNL with the AIC-based bandwidth. In the supplementary materials available online, we provide the results of a simulation in which preference heterogeneity depends deterministically on some spatial process (e.g., distance decay). In such a case GW-MNL works much better and different methods of choosing bandwidth provide similar results.

In Table 3, we present a summary of the results of the MAPE, calculated by comparing true individual-specific WTP with individual-specific WTP predicted by the models. As we have a separate MAPE calculated for each of 100 artificial datasets, we present a summary, with the mean MAPE value, and the 5^{th} and 95^{th} percentiles of MAPE values. The cases analyzed here are similar to Table 2, with the exception of S-MXL, which was divided into two cases. We calculate the posterior mean of individual-specific WTP, as well as the posterior median of individual-specific WTP, as the latter seems to have significantly lower error. This is an interesting finding, as the mean is usually used for inference, while the median is known to be less sensitive to outliers. Nevertheless, for MXL we calculate only the posterior mean, as the posterior median is not easily obtainable.

As for the performance of the models, MXL has the highest MAPE on average (535%). Next are GW-MNL with bandwidth chosen based on AIC (417%) and S-MXL, with posterior means of WTP (398%). The best performance is provided by the GW-MNL, where the bandwidth is chosen to minimize the MAPE (149%) and S-MXL with posterior medians of individual-specific WTP (146%). It should be noted again that the former could not be chosen in a real-life scenario, as the values of MAPE could not be calculated. As a result, although GW-MNL can follow S-MXL when it comes to the prediction of individual-specific WTP, it remains limited by the lack of a robust method for choosing a bandwidth parameter. Last, it should be noted that even for S-MXL, which works best in this setting, an average bias is 146%, which may be considered significant in real-world applications.

Table 3: Mean absolute percentage error calculated for individual-specific willingness to pay estimates

	Mean Absolute Percentage Error (%) [90% confidence interval]
S-MXL – posterior mean of WTP	398.0874 [160.29 – 744.49]
S-MXL – posterior median of WTP	145.8917 [88.49 – 284.06]
MXL – posterior mean of WTP	535.2666 [166.38 – 1226.17]
GW-MNL (min. AIC)	417.3288 [113.95 – 1607.45]
GW-MNL (min. MAPE)	148.8819 [98.32 - 243.85]

Table 4: Minimum, Mean and Maximum absolute percentage errors calculated for region-specific willingness to pay estimates

	Minimum Absolute Percentage Error (%) [90% confidence interval]	Mean Absolute Percentage Error (%) [90% confidence interval]	Maximum Absolute Percentage Error (%) [90% confidence interval]
S-MXL – posterior	9.3109	44.0709	110.4981
mean of WTP	[0.39 – 37.93]	[13.82 – 127.33]	[31.65 – 327.31]
S-MXL – posterior	24.1706	42.3833	59.8446
median of WTP	[5.48 – 43.58]	[26.79 – 55.07]	[41.46 – 72.44]
MXL – posterior	31.8365	71.7537	122.3254
mean of WTP	[0.41 – 186.43]	[13.85 – 266.29]	[30.47 – 372.05]
GW-MNL	29.9781	130.9899	658.9347
(min. AIC)	[5.77 – 55.33]	[52.05 – 170.34]	[71.77 – 982.27]
GW-MNL	49.0685	68.6270	88.0015
(min. MAPE)	[18.66 – 62.75]	[60.08 – 96.86]	[69.44 – 161.82]

In Table 4, we present a summary of absolute percentage errors when comparing mean WTP across nine “regions” (which we defined simply as nine squares of equal size). We calculated minimal, mean, and maximal absolute percentage error (across nine “regions”) for each model and each artificial dataset. To summarize the results, we once again present mean results, as well as 5^{th} and 95^{th} percentiles.

When comparing MAPE the conclusions from this Table are slightly different from those in Table 3 – here S-MXL performs best irrespectively whether it uses mean (44%) or median (42%) posterior WTP. GW-MNL (69%, with bandwidth that minimizes MAPE from Table 3) and MXL (72%) follow, although we note that MXL’s MAPE has relatively large interquantile range here. Lastly, GW-MNL with bandwidth that minimizes AIC (131%) performs worst. When comparing MAPE the order is the same as for Table 3. It seems that although S-MXL (with posterior mean of WTP) has low error on average, it may produce much larger error (up to 110%) for some “regions”. The results for S-MXL (with posterior median) and GW-MNL (with bandwidth that minimizes MAPE) provide more uniform distribution of errors across 9 “regions”.

5 Summary and conclusions

In this study, we compared the performance of three models that can be used to analyze discrete choice data in which preference heterogeneity depends on some spatial factors, which affect spatially autocorrelated preferences. The models compared include 1) the traditional MXL model, which does not explicitly assume any spatial dependence but recovers spatial effects by conditioning individual-specific WTP on the vector of observed choices, 2) the novel specification for the MXL model, in which spatial dependence of preference heterogeneity is explicitly accounted for (S-MXL), and 3) the geographically-weighted (locally estimated) MNL model (GW-MNL). The comparison was based on the models’ ability to recover true parameters of the data generating process, as assumed in the simulation.

Our results show that the S-MXL model generally performs best. This should not come as a surprise, as this is the only currently available model which is specified consistently with our data generating process, involving both unobserved and spatially autocorrelated preference heterogeneity. The individual-specific estimates resulting from the traditional MXL model are conditional on population-level distributions and observed choices without explicitly allowing for spatial auto-correlation in random parameters, while the GW-MNL is unable to recover the non-spatial unobserved preference heterogeneity implied by the data generating process in our experiment. The model performs well in an idealized situation of no such source of preference heterogeneity; this is equivalent to the MNL model failure in the case of, for example, normally distributed preference parameters, as implied by the common MXL model specification.

It must be noted here that although in our simulation setting the proposed S-MXL model shows promising results, further work is needed to confirm its performance in real-life studies. In particular, we acknowledge two limitations of this model, which future research could focus on: 1) establishing more efficient algorithms for faster estimation and better convergence (e.g., other Bayesian estimation techniques, use of probit instead of logit kernel) and 2) allowing for correlation between random parameters, which in our opinion could be a significant limiting factor for empirical applications.

The good news of our experiment is that if the goal of the analysis is to estimate the overall distribution of WTP in the population, in our simulation setting the standard MXL model was sufficient. Not accounting for spatial dependencies did not significantly bias estimates of the mean or some other characteristics of the WTP distribution (Table 2).

On the contrary, we find that if one is interested in deriving individual-specific or region-specific estimates of WTP, using a model that accounts for spatial dependencies may be necessary and the results of commonly used approaches may be misleading. In our simulation, the MXL and GW-MNL models resulted in substantial bias of individual-specific (Table 3) and region-specific (Table 4) estimates. Therefore, researchers should be careful when using a popular two-step method with individual-specific estimates from MXL as, in this setting, even the correct model (S-MXL) led to the bias of nearly 150%, on average.

The likely reason the GW-MNL model fails in these settings is that although the model accounts for spatial dimension of heterogeneity, it does not allow for other sources of (unobserved) preference heterogeneity. (In a real-life situation, one would typically have several individuals per location (e.g., per ZIP-code area), which would likely render GW-MNL's performance even worse). This could be addressed by developing more advanced models, such as a geographically weighted latent class model, or geographically weighted mixed logit (GW-MXL) models. The development and proliferation of these approaches to applied studies would remain limited, however, until reliable methods for selection of the bandwidth parameter, responsible for weighting in other nearby locations in geographically weighted models, are found (Budziński et al., forthcoming).

In Appendix B, we present results from simulation with different data generating process, where preference heterogeneity is driven solely by a deterministic factor (for example, a distance decay function). We find that in such a setting the proposed S-MXL performed rather poorly, and could not provide reliable results. On the other hand, GW-MNL performed quite well. This indicates that the proposed S-MXL model is appropriate for settings which lie in between the two extreme cases of no spatial dependence in preference heterogeneity (as in MXL) and preference heterogeneity being driven purely by the spatial pattern (as in GW-MNL). As in real-life there can be multiple factors driving preference heterogeneity (both spatial and non-spatial), we find it to be more realistic description of preference heterogeneity.

Overall, our results are encouraging, showing that it is generally possible to mitigate bias resulting from spatial autocorrelation of individual preferences for environmental goods. At the same time, we demonstrate that by employing correct parametric methods to explicitly account for spatial dependencies, it is possible to derive unbiased individual- and region-specific preference and WTP estimates. We call for the further development of these methods, however, so that they better fit the real-life situations that are encountered in many applied studies.

Acknowledgements

Wiktor Budziński gratefully acknowledges the support of the Foundation for Polish Science (FNP) and the National Science Centre of Poland (Sonata, 2015/19/D/HS4/01972). Mikołaj Czajkowski gratefully acknowledges the support of the National Science Centre of Poland (Sonata Bis, 2018/30/E/HS4/00388).

References

[1] Abildtrup J., Garcia S., Olsen S. B., Stenger A., (2013), Spatial preference heterogeneity in forest recreation, *Ecological Economics* 92(1), 67–77.

[2] Broch S. W., Strange N., Jacobsen J. B., Wilson K. A., (2013a) Farmers' willingness to provide ecosystem services and effects of their spatial distribution, *Ecological Economics* 92, 78–86.

[3] Broch S. W., Strange N., Jacobsen J. B., Wilson K. A., (2013b), Farmers' willingness to provide ecosystem services and effects of their spatial distribution, *Ecological Economics* 92(0), 78–86.

[4] Budziński W., Campbell D., Czajkowski M., Demsar U., Hanley N., (2018), Using geographically weighted choice models to account for spatial heterogeneity of preferences, *Journal of Agricultural Economics* 69(3), 606–626.

[5] Budziński W., Campbell D., Czajkowski M., Demsar U., Hanley N., Using geographically weighted choice models to account for spatial heterogeneity of preferences, *Journal of Agricultural Economics*, forthcoming.

[6] Campbell D., Hutchinson W. G., Scarpa R., (2009), Using Choice Experiments to Explore the Spatial Distribution of Willingness to Pay for Rural Landscape Improvements, *Environment and Planning A* 41(1), 97–111.

[7] Campbell D., Scarpa R., Hutchinson W. G., (2008), Assessing the spatial dependence of welfare estimates obtained from discrete choice experiments, *Letters in Spatial and Resource Sciences* 1(2-3), 117–126.

[8] Carson R. T., Czajkowski M., (2014), The Discrete Choice Experiment Approach to Environmental Contingent Valuation, [in:] *Handbook of choice modelling*, Hess S., Daly A., [eds.] Elgar E., Northampton, MA.

[9] Czajkowski M., Budziński W., (2015), An insight into the numerical simulation bias – a comparison of efficiency and performance of different types of quasi Monte Carlo simulation methods under a wide range of experimental conditions, *Environmental Choice Modelling Conference*, Copenhagen.

[10] Czajkowski M., Budziński W., (2019), Simulation error in maximum likelihood estimation of discrete choice models, *Journal of Choice Modelling* 31, 73–85.

[11] Czajkowski M., Budziński W., Campbell D., Giergiczny M., Hanley N., (2017), Spatial Heterogeneity of Willingness to Pay for Forest Management, *Environmental and Resource Economics* 68(3), 705–727.

[12] Dekker T., Koster P., Brouwer R., (2014), Changing with the Tide: Semiparametric Estimation of Preference Dynamics, *Land Economics* 90(4), 717–745.

[13] Fotheringham A. S., Brunsdon C., Charlton M., (2003), *Geographically weighted regression: the analysis of spatially varying relationships*, John Wiley & Sons.

[14] Fotheringham S., Charlton M., Brunsdon C., (1998), Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis, *Environment and Planning A* 30(11), 1905–1927.

[15] Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., (2014), *Bayesian data analysis*, CRC press Boca Raton, FL.

[16] Hanley N., Czajkowski M., (2019), The Role of Stated Preference Valuation Methods in Understanding Choices and Informing Policy, *Review of Environmental Economics and Policy* 13(2), 248–266.

[17] Hess S., Train K., (2017), Correlation and scale in mixed logit models, *Journal of Choice Modelling* 23, 1–8.

[18] Hynes S., Hanley N., O'Donoghue C., (2010), A Combinatorial Optimization Approach to Nonmarket Environmental Benefit Aggregation via Simulated Populations, *Land Economics* 86(2), 345–362.

[19] Johnston R. J., Ramachandran M., (2014), Modeling Spatial Patchiness and Hot Spots in Stated Preference Willingness to Pay, *Environmental and Resource Economics* 59(3), 363–387.

[20] Johnston R. J., Ramachandran M., Schultz E. T., Segerson K., Besedin E. Y., (2011), Characterizing spatial pattern in ecosystem service values when distance decay doesn't apply: choice experiments and local indicators of spatial association. Paper number 103374 provided by Agricultural and Applied Economics Association in its series 2011 Annual Meeting, July 24-26, 2011, Pittsburgh, Pennsylvania.

[21] Koster P. R., Koster H. R. A., (2015), Commuters' preferences for fast and reliable travel: A semi-parametric estimation approach, *Transportation Research Part B: Methodological* 81, Part 1, 289–301.

[22] LeSage J. P., (1999), The Theory and Practice of Spatial Econometrics, unpublished manuscript available at: `http://www.spatial-econometrics.com`.

[23] McFadden D., (1974), Conditional Logit Analysis of Qualititative Choice Behaviour, [in]: *Frontiers in Econometrics*, [ed.:] Zarembka P., Academic Press, New York, NY, 105–142.

[24] Smith T. E., LeSage J. P., (2004), A bayesian probit model with spatial dependencies, *Spatial and Spatiotemporal Econometrics* 18(18), 127–160.

[25] Train K., Sonnier G., (2005), Mixed Logit with Bounded Distributions of Correlated Partworths, [in:] *Applications of Simulation Methods in Environmental and Resource Economics*, [eds.:] Scarpa R., Alberini A., Springer Netherlands, 117–134.

[26] Train K. E., (2009), *Discrete Choice Methods with Simulation*, 2 Ed., Cambridge University Press, New York.

[27] Yao R. T., Scarpa R., Turner J. A., Barnard T. D., Rose J. M., Palma J. H. N., Harrison D. R., (2014), Valuing biodiversity enhancement in New Zealand's planted forests: Socioeconomic and spatial determinants of willingness-to-pay, *Ecological Economics* 98(0), 90–101.

Appendix

A Econometric specification and estimation of the spatially autocorrelated mixed logit model

To specify the posterior distribution of this hierarchical Bayes model, we start with the conditional probability of making an observed choice. If random parameters, β_i and α_i, are known, the likelihood of a vector of choices ($\mathbf{y_i}$) is given by:

$$p\left(\mathbf{y_i} \mid \mathbf{X_i}, \boldsymbol{\mu}, \boldsymbol{\theta}_{i\cdot}\right) = \prod_{t=1}^{NCT} \frac{\exp\left(\boldsymbol{\beta}_i \mathbf{X}_{iy_{it}t}^{\text{non-cost}} - \alpha_i X_{iy_{it}t}^{\text{cost}}\right)}{\sum_j \exp\left(\boldsymbol{\beta}_i \mathbf{X}_{ijt}^{\text{non-cost}} - \alpha_i X_{ijt}^{\text{cost}}\right)}, \tag{13}$$

where variable y_{it} denotes which alternative individual i has chosen in a choice situation t. The joint distribution of observed choices, and model parameters is therefore proportional to:

$$p\left(\mathbf{y_i}, \boldsymbol{\mu}, \boldsymbol{\theta}_{i\cdot}, \boldsymbol{\sigma}^2, \boldsymbol{\rho} \mid \mathbf{X_i}\right) \propto p\left(\mathbf{y_i} \mid \mathbf{X_i}, \boldsymbol{\mu}, \boldsymbol{\theta}_{i\cdot}\right) p\left(\boldsymbol{\theta}_{i\cdot} \mid \boldsymbol{\sigma}^2, \boldsymbol{\rho}\right) p\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\rho}\right). \tag{14}$$

We note that conditional distribution of θ was obtained in Section 2.2, namely $\boldsymbol{\theta}_{\cdot k} \sim MVN\left(0, \sigma_k^2\left(B'_{\rho_k} B_{\rho_k}\right)^{-1}\right)$. $p\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\rho}\right)$ denotes priors of these parameters. We have chosen standard diffuse priors, analogously as in Smith and LeSage (2004), which are presented in (15). We assume that a priori each parameter is independent. For spatial autocorrelation coefficients, ρ_k, distribution is uniform on interval depending on minimal and maximal eigenvalues of the W matrix.

$$\begin{cases} p(\mu_k) \propto exp\left(-\frac{1}{2}\left(\frac{\mu_k}{1000}^2\right)\right) & \text{for } k \in \{1, 2, ..., K\}, \\ p\left(\frac{1}{\sigma_k^2}\right) \propto \frac{1}{\sigma_k^2} & \text{for } k \in \{1, 2, ..., K\}, \\ p(\rho_k) \propto \mathbf{1}_{\{Eig_{Min}^{-1} < \rho_k < Eig_{Max}^{-1}\}} & \text{for } k \in \{1, 2, ..., K\}. \end{cases} \tag{15}$$

In the proposed model, only the conditional posterior distribution for σ_k^2 is a known distribution, namely $1/\sigma_k^2 \mid \boldsymbol{q}_{\cdot k}, \rho_k, \mathbf{y}_i \sim \Gamma\left(\frac{m}{2}, \boldsymbol{q}'_{\cdot k} \mathbf{B}'_{\rho_k} \mathbf{B}_{\rho_k} \boldsymbol{q}_{\cdot k}\right)$, where m is a number of distinct spatial locations (in our case, equal to the number of respondents). Spatial autocorrelation has a conditional posterior distribution proportional to

$$p(\rho_k \mid \boldsymbol{\theta}_{\cdot k}, \mathbf{y}_i) = \mathbf{1}_{\{Eig_{Min}^{-1} < \rho_k < Eig_{Max}^{-1}\}} \mid \mathbf{B}_{\rho_k} \mid \left[\frac{\boldsymbol{\theta}'_{\cdot k} \mathbf{B}'_{\rho_k} \mathbf{B}_{\rho_k} \boldsymbol{\theta}_{\cdot k}}{m}\right]^{-\frac{m}{2}} \frac{1}{Eig_{Max}^{-1} - Eig_{Min}^{-1}}$$

which does not describe any known distribution. Other parameters have conditional posterior distributions given as a product of (13) and their prior from (15). As some conditional posteriors are unknown, the proposed estimation algorithm is a Gibbs sampler, with four steps in total, where three of them employ the Metropolis-Hastings algorithm. The estimation process proceeds as follows:

1. Some initial values of all parameters are assumed: μ^0, σ^0, ρ^0, θ^0.

2. For the each individual, we draw the candidate vector of individual-specific parameters, $\overline{\theta}_{i\cdot} = \theta^0_{i\cdot} + \tau_1\eta$, where η is a vector of random variables from a multivariate normal distribution centered around 0, with standard deviations equal to vector σ^0, and τ_1 is the tuning parameter, used to assure an acceptance rate of approximately 0.3. This step has numerous sub-steps equal to the number of individuals in the sample. For individual l, there is the following procedure:

   a. We define $\theta^l_{\cdot k} = (\theta^1_{1k}, \theta^1_{2k}, ..., \overline{\theta}_{lk}, \theta^0_{l+1k}, ..., \theta^0_{NPk})'$, where, for example, θ^1_{1k} denotes the saved draw for individual 1. Analogously we define $\theta^{-l}_{\cdot k} = (\theta^1_{1k}, \theta^1_{2k}, ..., \theta^0_{lk}, \theta^0_{l+1k}, ..., \theta^0_{NPk})'$. Then

   $$R = \frac{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \mu^0, \theta^l_{i\cdot}\right) \prod_k f(\theta^l_{\cdot k})}{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \mu^0, \theta^{l-1}_{i\cdot}\right) \prod_k f(\theta^{l-1}_{\cdot k})}$$

   is compared with a random draw from the uniform distribution, ω. Specifically, if $\omega < R$ we set $\theta^1_{lk} = \overline{\theta}_{lk}$ for each k, and $\theta^1_{lk} = \theta^0_{lk}$ otherwise. In the formula for R $f(\theta^l_{\cdot k})$ is a pdf function of a multivariate normal distribution with mean equal to 0, and variance matrix equal to $\sigma^2_k \left(B'_{\rho_k} B_{\rho_k}\right)^{-1}$.

3. The vector of variances $\left(\rho^1\right)^2$ is drawn from the inverse gamma distribution conditional on θ^1 matrix and ρ^0 vector.

4. We draw a candidate vector for means of random parameters, $\overline{\mu} = \mu^0 + \tau_2\eta$, where η is a vector of random variables from a multivariate normal distribution centered around 0, and τ_2 is the tuning parameter, used to assure an acceptance rate of approximately 0.3. Then,

   $$R = \frac{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \overline{\boldsymbol{\mu}}, \boldsymbol{\theta}_{i\cdot}\right) \varphi\left(\frac{\overline{\boldsymbol{\mu}}}{\sqrt{1000}}\right)}{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \boldsymbol{\mu}^0, \boldsymbol{\theta}_{i\cdot}\right) \varphi\left(\frac{\boldsymbol{\mu}^0}{\sqrt{1000}}\right)}$$

   is compared with a random draw from the uniform distribution, ω, where $\varphi(\cdot)$ is a density function of a normal distribution. Specifically, if $\omega < R$, we set $\mu^1 = \overline{\mu}$, and $\mu^1 = \mu^0$ otherwise.

5. We draw a candidate vector $\overline{\rho} = \rho^0 + \tau_3\eta$, where η is a vector of random variables from a multivariate normal distribution centered around 0, and τ_3 is the tuning parameter, used to assure an acceptance rate of approximately 0.3. If the candidate draw is not from the interval $[Eig^{-1}_{Min}, Eig^{-1}_{Max}]$, we set $\rho_1 = \rho_0$. Otherwise, we compare

   $$R = \frac{\prod_k p(\overline{\rho}_k \mid \boldsymbol{q}^1_{\cdot k}, \mathbf{y}_i)}{\prod_k p(\rho^0_k \mid \boldsymbol{q}^1_{\cdot k}, \mathbf{y}_i)}$$

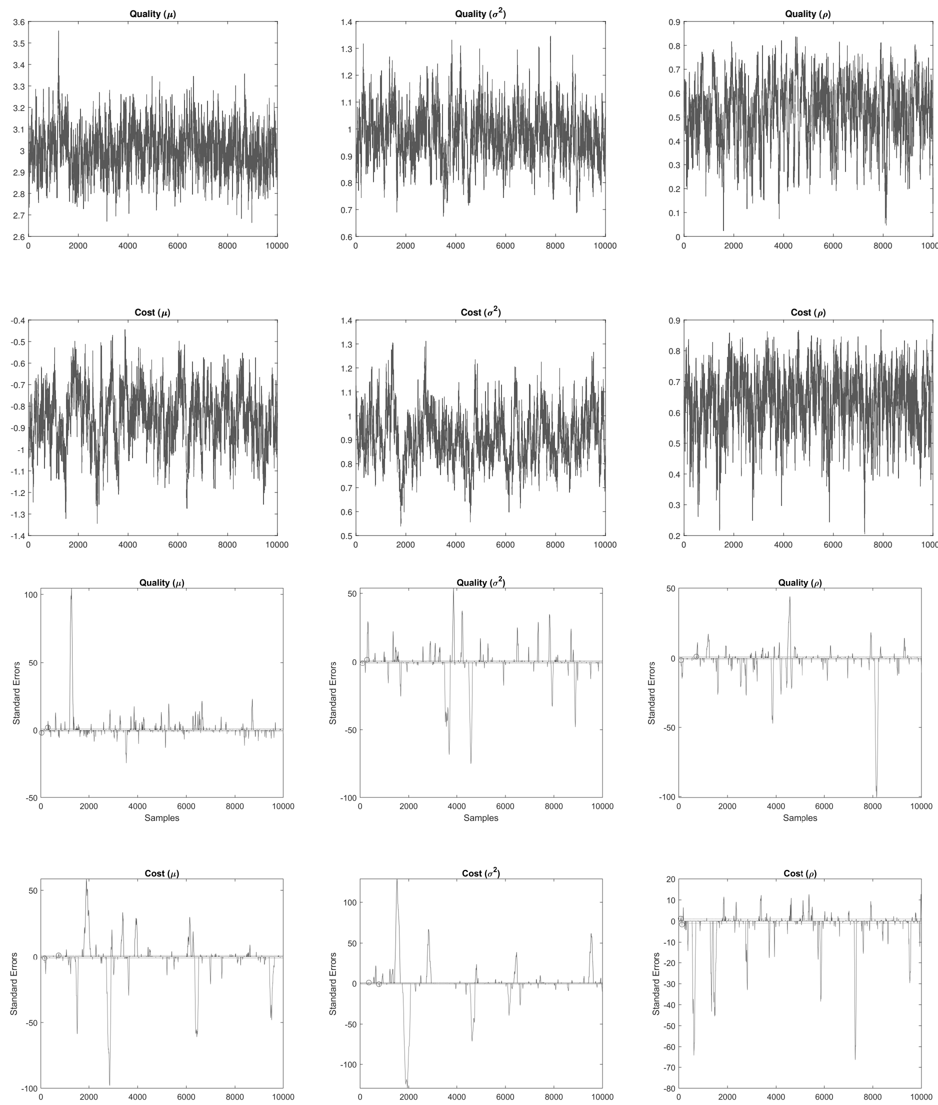
with a random draw from the uniform distribution, ω. Specifically, if $\omega < R$ we set $\rho^1 = \overline{\rho}$, and $\rho^1 = \rho^0$ otherwise.

After all steps, the procedure is repeated but the initial values μ^0, σ^0, ρ^0, θ^0 are replaced by μ^1, σ^1, ρ^1, θ^1. We run this process 40,000 times, considering the first 10,000 draws as the "burn-ins" sample, and then taking every third draw of the rest of the generated draws. In total we obtain 10,000 draws from the posterior distribution for inferences.
In Figure 2 we illustrate a convergence of the algorithm, for the first sample in the simulation. First, we plotted draws from the conditional posterior distribution for each of the 6 coefficients against the number of the algorithm's iterations. For all parameters, we observe that draws oscillate around true values of the coefficients assumed in the data generating process. Next, we employed CUSUM plots. It reveals that in some cases the algorithm suffers from the autocorrelation of Markov chain. For example, in the case of μ parameter for cost, it takes long time for the algorithm, to reverse back to the mean value. Nevertheless, we do not observe consistent drifting away from the mean on the CUSUM plot (it always reverses back), which indicates convergence.

B Results of a simulation with a deterministic spatial heterogeneity of preferences

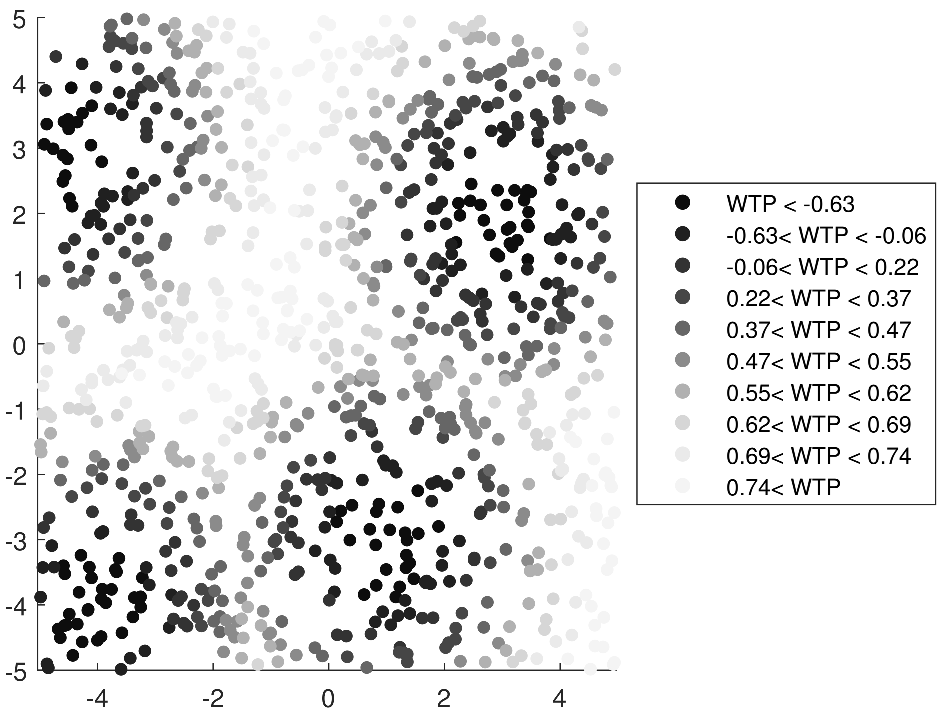
In this Appendix we provide the results of an analogous simulation as in the main text, with the difference that now preference heterogeneity is solely driven by a deterministic spatial process. Specifically, we assume that researcher is trying to value some public good, for which there are available 4 substitutes in the area of interest. The distance to those substitutes will be a driver of preference heterogeneity. We assume that individual-specific parameters are given by $\beta_i = -1 + 3\log(Dist_i)$ and $\alpha_i = 2\sqrt{Dist_i}$, where $Dist_i$ is a distance to the nearest substitute for individual i. In Figure 3 we present a distribution of WTP implied by these parameters, for substitutes located at points $(-4, -4)$, $(3, 2)$, $(-4.5, 3)$ and $(1, -3)$. This distribution follows an intuitive dependence, that individuals living further away from the substitutes are willing to pay more for a public good. The difference with respect to Figure 1 is that now, the dependence is deterministic. The question we investigate with such data generating process is whether proposed models can recover such spatial process, if, e.g., substitutes are unknown or we cannot control for them for some reason.
Below we present 3 tables analogous to tables 2, 3 and 4 from the main text. In this comparison we do not include S-MXL model as we obtained very strange results with this data generating process, with implied WTP distribution characteristics sometimes more than 100 times larger than the true values. We believe that this is because for such data generating process spatial autocorrelation parameter, ρ_k, is equal to 1, which is its upper boundary. We think that this may cause some

Figure 2: Example of convergence of the Gibbs sampler, for the first simulated dataset



identification issues, which could possibly be resolved with development of spatial autoregressive process of second order. We decided to not pursue this path, as to the best of our knowledge, such processes are not really analyzed in spatial econometrics, as they do not arise very often in practice. Nevertheless, we note that if researcher suspects similar process of preference heterogeneity in his study, S-MXL may not be

Figure 3: Example of spatial distribution of WTP in the deterministic data generating process



an appropriate approach.

We report three interesting findings based on Tables 5, 6 and 7. First, in all Tables GW-MNL significantly outperforms MXL. This is different from the main text, where in some cases MXL outperformed GW-MNL (especially with wrongly chosen bandwidth parameter). Second, in here the method of choosing bandwidth matters less. Indeed, whether we minimize AIC or MAPE, we obtain quite similar results, as on average the optimal bandwidth was 0.45 for the former, and 0.59 for the latter, so the difference is much smaller than in the main text. Lastly, as can be seen in Table A1, although GW-MNL recovers percentiles of WTP distribution well, the mean and standard deviation are still off, therefore we recommend using characteristics such as median, when using GW-MNL for inference.

The results reported in this Appendix indicate that GW-MNL is much better suited for analysis of data generating processes as presented in Figure 3, rather than the one as in Figure 1 in the main text.

Table 5: Comparison of selected moments and quantiles of the recovered willingness-to-pay distribution with their true values implied by the data generating process

	Mean	Std. Dev	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
True	0.1221	1.1546	−0.6342	0.1046	0.4717	0.6526	0.7407
MXL	0.3515	0.6889	−0.4614	−0.0804	0.3218	0.7539	1.2055
	(0.0198)	(0.0327)	(0.0317)	(0.0215)	(0.0182)	(0.0250)	(0.0460)
GW-MNL (min. AIC)	0.2125	0.6631	−0.6646	0.0711	0.4342	0.6287	0.7418
	(0.0212)	(0.0455)	(0.0528)	(0.0302)	(0.0208)	(0.0210)	(0.0279)
GW-MNL (min. MAPE)	0.2308	0.5657	−0.5729	0.0587	0.4154	0.6126	0.7190
	(0.0240)	(0.0747)	(0.0694)	(0.0272)	(0.0237)	(0.0223)	(0.0280)

Table 6: Mean absolute percentage error calculated for individual-specific willingness to pay estimates

	Mean Absolute Percentage Error (%) [90% confidence interval]
MXL – posterior mean of WTP	185.2022 [156.21 – 223.35]
GW-MNL (min. AIC)	49.0520 [39.99 – 58.51]
GW-MNL (min. MAPE)	46.1726 [38.86 – 55.00]

Table 7: Minimum, Mean and Maximum absolute percentage errors calculated for region-specific willingness to pay estimates

	Minimum Absolute Percentage Error (%) [90% confidence interval]	Mean Absolute Percentage Error (%) [90% confidence interval]	Maximum Absolute Percentage Error (%) [90% confidence interval]
MXL – posterior	9.6101	63.3339	145.9910
mean of WTP	[0.57 – 20.28]	[53.71 – 71.94]	[119.23 – 173.70]
GW-MNL	1.7086	34.3477	94.9462
(min. AIC)	[0.16 – 4.65]	[26.84 – 40.62]	[68.29 – 129.28]
GW-MNL	2.1241	42.1818	114.5984
(min. MAPE)	[0.13 – 5.72]	[30.17 – 51.70]	[84.91 – 150.12]