WORKING PAPERS

A MONTE CARLO INVESTIGATION OF THE EFFECTS OF SPATIAL HETEROGENEITY OF PREFERENCES FOR DISCRETE CHOICE MODELS

WIKTOR BUDZIŃSKI
MIKOŁAJ CZAJKOWSKI

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES

WORKING PAPERS

A Monte Carlo investigation of the effects of spatial heterogeneity of preferences for discrete choice models

Wiktor Budziński*, Mikołaj Czajkowski

Faculty of Economics, University of Warsaw, Poland
* Corresponding author: wbudzinski@wne.uw.edu.pl

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 parametrically account for spatial auto-correlation 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 correlation, 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 autocorrelation. We find that, although ignoring spatial autocorrelation 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 spatially-autocorrelated 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; preference heterogeneity; spatial auto-correlation

JEL codes: Q51, C25, C31

1. Introduction

Preferences for environmental goods may follow spatial patterns (ref. intro paper of this special issue). This becomes an important issue for discrete choice models, commonly used for modeling consumers' preferences and willingness-to-pay (WTP). 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 (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 spatially autocorrelated mixed logit (SA-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. forthcoming). 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 unobserved spatial autocorrelation. 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 SA-MXL model proposed here and its estimation are provided in Appendix 1. 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 spatially autocorrelated mixed logit (SA-MXL) model. In the last part, we describe a geographically weighted multinomial logit (GW-MNL) model.[1]

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 the mixed logit model, it is assumed that each individual i has a separate, independent set of parameters. Assuming extreme value distribution for the error component, ε_{ijt}, leads to a well-known formulation of conditional likelihood in multinomial logit form. As individual-specific parameters are not directly observed, a distribution for them needs to be assumed and

---

[1] 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.

integrated. Such unconditional likelihood can then be used for estimation of parameters describing the distribution of random effects; for example, their means and variances. Estimation of MXL is usually performed using the maximum simulated likelihood (MSL) method, and this is also the case in the present study.[2]

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 will focus on predicting individual-specific marginal WTP with the following formula

$$E\left(\frac{\boldsymbol{\beta}_i}{\alpha_i} \mid \mathbf{y}_i, \mathbf{X}_i, \Omega\right) = \int \frac{\boldsymbol{\beta}_i}{\alpha_i} \frac{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \Omega, \alpha_i, \boldsymbol{\beta}_i\right) f\left(\alpha_i, \boldsymbol{\beta}_i \mid \Omega\right)}{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \Omega\right)} d\left(\alpha_i, \boldsymbol{\beta}_i\right), \qquad (2)$$

where $p\left(\mathbf{y}_i \mid \mathbf{X}_i, \Omega, \alpha_i, \boldsymbol{\beta}_i\right)$ is the likelihood of individual i making the observed choices conditional on the values of random parameters, $p\left(\mathbf{y}_i \mid \mathbf{X}_i, \Omega\right)$ is the same likelihood but unconditional (random parameters are integrated out, so it is the likelihood function for MXL) and $f\left(\alpha_i, \boldsymbol{\beta}_i \mid \Omega\right)$ is the assumed pdf function of random parameters. 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 (2) is due to conditioning on the vector of observed choices, $\mathbf{y}_i$.

2.2. Spatially autocorrelated mixed logit

In this study, we propose a novel extension of the MXL model, which directly accounts for spatial dependencies in preference heterogeneity. Although SA-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

---

[2] For our simulation, we employ 10,000 draws from a scrambled Sobol sequence (Czajkowski and Budziński, 2015).

individuals; this is not the case here. In this specification, we follow LeSage and Smith (2004). Let us define θ_{ik}, which the random parameter for the k-th attribute (out of a total of K attributes) will depend on, as:

$$\theta_{ik} = \rho_k \sum_{m=1}^{NP} w_{im} \theta_{mk} + u_{ik}, \tag{3}$$

where $u_{ik} \sim N(0, \sigma_k)$. 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 (3) describes a spatial autoregressive process, which can be written in vectorized form as $\mathbf{\theta}_{\cdot k} = \mathbf{B}_{\rho_k}^{-1} \mathbf{u}_{\cdot k}$, where $B_{\rho_k} = I - \rho_k W$, and W is a spatial weight matrix, whose rows sum to 1.[3] From this it follows that $\mathbf{\theta}_{\cdot k} \sim MVN\left(0, \sigma_k^2 \left(B_{\rho_k}{}' B_{\rho_k}\right)^{-1}\right)$. If the k-th attribute is not a cost ($k < K$), we assume that the random parameter is specified as:

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

whereas if the k-th attribute is a cost ($k = K$), then we assume that:

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

where μ_k are parameters to estimate (means of random parameters). 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.

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

---

[3] Throughout the study, we use the inverse squared distance as a weight matrix, which is a standard approach in spatial econometrics.

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 LeSage and Smith (2004). The detailed specification of the model and its estimation details are provided in Appendix 1.

Individual-specific estimates of WTP are easy to obtain from SA-MXL, as we can simply use saved draws from posterior distributions to calculate them. For the n-th saved set of draws from a posterior distribution, we can define WTP for the k-th attribute as:

$$WTP_{nk} = \frac{\mu_k^n + \boldsymbol{\theta}_{\cdot k}^n}{\exp\left(\mu_K^n + \boldsymbol{\theta}_{\cdot K}^n\right)}, \tag{6}$$

and then we can calculate its mean, median, or any other characteristic over n.

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. forthcoming) is an extension of the geographically weighted regression (Fotheringham et 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).

$$U_{ijt} = \boldsymbol{\beta}_l \mathbf{X}_{ijt}^{\text{non-cost}} - \alpha_l X_{ijt}^{\text{cost}} + \varepsilon_{ijt} \tag{7}$$

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

$$L_i^l = \prod_{t=1}^{CT} \frac{\exp\left(\boldsymbol{\beta}_l \mathbf{X}_{iy_{it}t}^{\text{non-cost}} - \alpha_l X_{iy_{it}t}^{\text{cost}}\right)}{\sum_j \exp\left(\boldsymbol{\beta}_l \mathbf{X}_{ijt}^{\text{non-cost}} - \alpha_l X_{ijt}^{\text{cost}}\right)}. \tag{8}$$

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

$$WL^l = \sum_{i=1}^{NP} \lambda\left(Lat_i, Long_i, b, l\right) \log\left(L_i^l\right), \tag{9}$$

where $\lambda\left(Lat_i, Long_i, b, l\right)$ 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. There are a few functional forms of $\lambda(\cdot)$ proposed in the literature. In what follows, we use the Gaussian kernel[4] defined as:

$$\lambda\left(Lat_i, Long_i, b, l\right) = \exp\left(-0.5\frac{\left(Lat_i - Lat_l\right)^2 + \left(Long_i - Long_l\right)^2}{b^2}\right). \tag{10}$$

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.

---

[4] For other possible kernels, see Fotheringham, Brunsdon and Charlton, 2003.

In GW-MNL, individual-specific WTP can be calculated simply with marginal WTP for each local model, i.e. $WTP_l = \dfrac{\beta_l}{\alpha_l}$, 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. (forthcoming).

3. Simulation setting

Our simulation setting was aimed at recovering individual preferences in the case of their spatial autocorrelation. Even though we expect the SA-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 assumed 1,000 individuals with six choice tasks per individual, each choice consisting of three alternatives, defined by two attributes, denoted as *Quality* and *Cost*. 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 is given by:

$$U_{ijt} = \beta_i \text{Quality}_{ijt} - \alpha_i \text{Cost}_{ijt} + \varepsilon_{ijt}, \tag{11}$$

where, *Quality* and *Cost* attributes follow uniform distributions on a $[0, 2]$ interval, and the stochastic term, ε_{ijt}, follows a standard extreme value distribution. α_i follows a log-normal distribution. Spatial autocorrelation was assumed to be equal to 0.6 for both parameters. The values of parameters used in the data generating process are summarized in Table 1.

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

We assumed the individuals were distributed uniformly on a 10x10 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. This data generating process may correspond to a situation where spatial dependence is only one of 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.

Figure 1. Example of spatial distribution of WTP in the data generating process
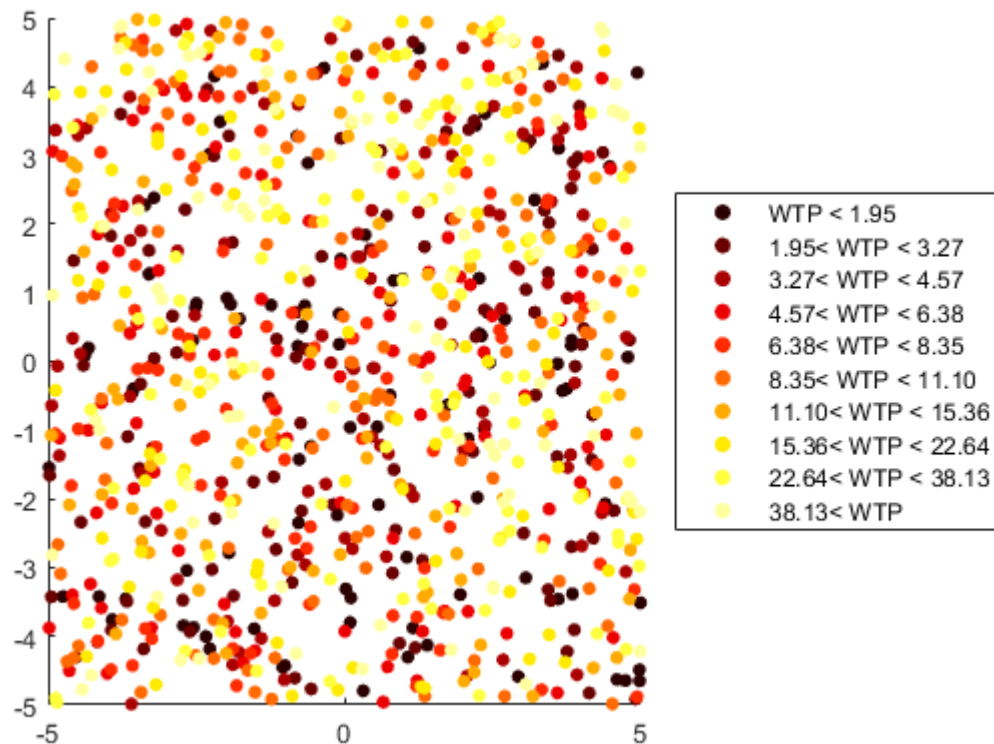
The simulation uses 100 artificial samples according to the described data generating process. We then estimate MXL, SA-MXL, and GW-MNL models for each sample, with 40 GW-MNL models for each sample (the GW-MNL models differed in the value of bandwidth parameters, ranging from 0.05 to 2 with 0.05 increases).

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 SA-MXL, they are additionally spatially autocorrelated) with unknown analytic formulas. For each of 100 generated datasets we calculated the implied characteristics of WTP for each model. Table 2 reports their means and standard deviations (in brackets).

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	15.4807	31.6163	1.617	3.4422	7.6421	16.7063	33.8252
SA-MXL	16.5795	39.5100	1.6188	3.4657	7.7832	17.3141	35.7787
	(4.8224)	(28.1746)	(0.2088)	(0.4335)	(1.2882)	(3.9655)	(10.5124)
MXL	16.3735	30.2507	1.5762	3.4088	7.8272	17.8172	37.2093
	(4.6594)	(14.7820)	(0.1993)	(0.4142)	(1.2514)	(3.9582)	(10.6940)
GW-MNL	-0.0191	302.5630	2.4358	3.2610	4.5097	6.7887	11.3977
(min. AIC)	(55.1850)	(1617.1076)	(0.4508)	(0.4026)	(0.4681)	(0.8828)	(2.3929)
GW-MNL	4.7185	9.8760	2.2433	3.3625	4.3330	5.5394	7.4295
(min. MAPE)	(1.4703)	(26.0672)	(3.4068)	(1.6781)	(1.0085)	(0.8548)	(1.5778)

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 SA-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 SA-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 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.[5]

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 5th and 95th percentiles of MAPE values. The cases analyzed here are similar to Figure 2, with the exception of SA-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.

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

	Mean Absolute Percentage Error (%) [90% confidence interval]
SA-MXL – posterior mean of WTP	398.0874 [160.29 - 744.49]
SA-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]

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 SA-MXL, with posterior means of WTP (398%). The best performance is provided by the GW-MNL, where the

---

[5] 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.

bandwidth is chosen to minimize the MAPE (149%) and SA-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 SA-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 SA-MXL, which works best in this setting, an average bias is 146%, which may be considered significant in real-world applications.

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 5th and 95th percentiles.

When comparing MAPE the conclusions from this Table are slightly different from those in Table 3 – here SA-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 SA-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 SA-MXL (with posterior median) and GW-MNL (with bandwidth that minimizes MAPE) provide more uniform distribution of errors across 9 "regions".

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

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 (SA-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 SA-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, while the GW-MNL is unable to recover the non-spatial unobserved preference heterogeneity implied by the data generating process in our experiment.[6]

It must be noted here that although in our simulation setting the proposed SA-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).

---

[6] 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.

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.

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.[7] This could be addressed by developing more advanced models, such as a geographically weighted latent class model (GW-LC; Campbell et al., this issue), 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).

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.

---

[7] 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.

Appendix 1. 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, $\mathbf{b}_i$ and α_i, are known, the likelihood of a vector of choices $\left(\mathbf{y}_i\right)$ is given by:

$$p\left(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{m}, \mathbf{q}_{i\cdot}\right) = \prod_{t=1}^{CT} \frac{\exp\left(\mathbf{b}_i \mathbf{X}_{iy_{it}t}^{\text{non-cost}} - \alpha_i X_{iy_{it}t}^{\text{cost}}\right)}{\sum_j \exp\left(\mathbf{b}_i \mathbf{X}_{ijt}^{\text{non-cost}} - \alpha_i X_{ijt}^{\text{cost}}\right)}. \tag{12}$$

To obtain the posterior distribution for all model coefficients, we need to first specify their priors. We have chosen standard diffuse priors, analogously as in LeSage and Smith (2004), which are presented in (13). For spatial autocorrelation coefficients, ρ_k, distribution is uniform on interval depending on minimal and maximal eigenvalues of the W matrix.

$$\begin{cases} \mu_k \sim N(0,1000) \\ 1/\sigma_k^2 \sim \Gamma(0,0) \\ \rho_k \sim U\left(Eig_{Min}^{-1}, Eig_{Max}^{-1}\right) \\ \mathbf{b}_k \sim MVN\left(\mu_k, \sigma_k^2\left(B_{\rho_k}'B_{\rho_k}\right)^{-1}\right) \\ \mathbf{a} \sim MVLN\left(\mu_K, \sigma_K^2\left(B_{\rho_K}'B_{\rho_K}\right)^{-1}\right) \end{cases} \tag{13}$$

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

$$p(\rho_k \mid \mathbf{q}_{\cdot k}) = \left|\mathbf{B}_{\rho_k}\right| \left[\frac{\mathbf{q}_{\cdot k}'\mathbf{B}_{\rho_k}'\mathbf{B}_{\rho_k}\mathbf{q}_{\cdot k}}{m}\right]^{-m/2} \frac{1}{Eig_{Max}^{-1} - Eig_{Min}^{-1}},$$ which does not describe any known

distribution. Other parameters have posterior distributions given as a product of (12) and their prior from (13). As some 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: $\mathbf{m}^0, \mathbf{s}^0, \mathbf{r}^0, \mathbf{q}^0$.

2. For the each individual, we draw the candidate vector of individual-specific parameters, $\bar{\mathbf{q}}_{i\cdot} = \mathbf{q}^0_{i\cdot} + \mathbf{t}_1\mathbf{h}$, where $\mathbf{h}$ is a vector of random variables from a multivariate normal distribution centered around 0, with standard deviations equal to vector $\mathbf{s}^0$, 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 $\mathbf{q}^l_{\cdot k} = \left(\theta^1_{1k}, \theta^1_{2k}, \ldots, \bar{\theta}_{lk}, \theta^0_{l+1k}, \ldots, \theta^0_{NPk}\right)'$, where, for example, θ^1_{1k} denotes the saved draw for individual 1. Analogously we define $\mathbf{q}^{-l}_{\cdot k} = \left(\theta^1_{1k}, \theta^1_{2k}, \ldots, \theta^0_{lk}, \theta^0_{l+1k}, \ldots, \theta^0_{NPk}\right)'$. Then $R = \dfrac{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{m}^0, \mathbf{q}^l_{i\cdot}\right) \prod_k f\left(\mathbf{q}^l_{\cdot k}\right)}{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{m}^0, \mathbf{q}^{l-1}_{i\cdot}\right) \prod_k f\left(\mathbf{q}^{l-1}_{\cdot k}\right)}$ is compared with a random draw from the uniform distribution, ω. Specifically, if $\omega < R$ we set $\theta^1_{lk} = \bar{\theta}_{lk}$ for each k, and $\theta^1_{lk} = \theta^0_{lk}$ otherwise. In the formula for R $f\left(\mathbf{q}^l_{\cdot k}\right)$ 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 $\left(\mathbf{s}^1\right)^2$ is drawn from the inverse gamma distribution conditional on $\mathbf{q}^1$ matrix and $\mathbf{r}^0$ vector.

4. We draw a candidate vector for means of random parameters, $\bar{\mathbf{m}} = \mathbf{m}^0 + \mathbf{t}_2\mathbf{h}$, where $\mathbf{h}$ 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, $R = \dfrac{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \bar{\mathbf{\mu}}, \mathbf{\theta}_{i\cdot}\right) \varphi\left(\dfrac{\bar{\mathbf{\mu}}}{\sqrt{1000}}\right)}{p\left(\mathbf{y}_i \mid \mathbf{X}_i, \mathbf{\mu}^0, \mathbf{\theta}_{i\cdot}\right) \varphi\left(\dfrac{\mathbf{\mu}^0}{\sqrt{1000}}\right)}$ is compared with a random draw from the uniform distribution, ω. Specifically, if $\omega < R$, we set $\mathbf{m}^1 = \bar{\mathbf{m}}$, and $\mathbf{m}^1 = \mathbf{m}^0$ otherwise.

5. We draw a candidate vector $\bar{\mathbf{r}} = \mathbf{r}^0 + \mathbf{t}_3\mathbf{h}$, where $\mathbf{h}$ 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 $\left[Eig_{Min}^{-1}, Eig_{Max}^{-1} \right]$, we set $\mathbf{r}_1 = \mathbf{r}_0$. Otherwise, we compare

$R = \dfrac{\prod_k p(\bar{\rho}_k \mid \mathbf{q}_{\cdot k}^1)}{\prod_k p(\rho_k^0 \mid \mathbf{q}_{\cdot k}^1)}$ with a random draw from the uniform distribution, ω. Specifically,

if $\omega < R$ we set $\mathbf{r}^1 = \bar{\mathbf{r}}$, and $\mathbf{r}^1 = \mathbf{r}^0$ otherwise.

After all steps, the procedure is repeated but the initial values $\mathbf{m}^0, \mathbf{s}^0, \mathbf{r}^0, \mathbf{q}^0$ are replaced by $\mathbf{m}^1, \mathbf{s}^1, \mathbf{r}^1, \mathbf{q}^1$. 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.

Appendix 2. 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 A1 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.

Figure A1. Example of spatial distribution of WTP in the deterministic data generating process
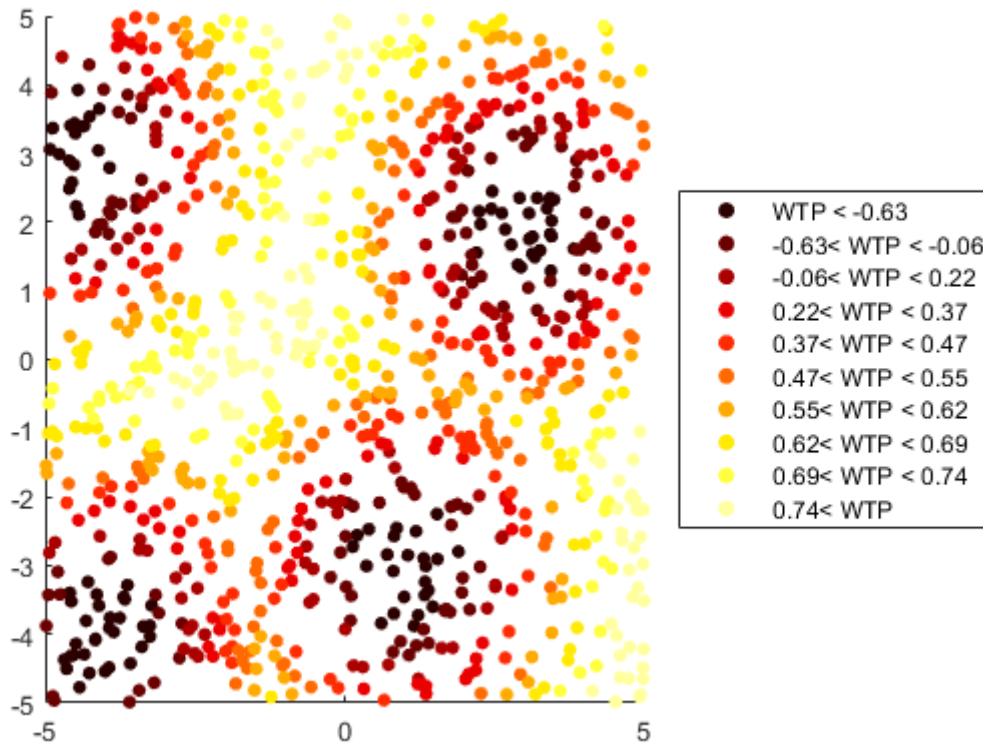
Below we present 3 tables analogous to tables 2, 3 and 4 from the main text. In this comparison we do not include SA-MXL model as we obtained very strange results with this data generating process, with implied WTP distribution characteristics sometimes more than 100 times larger than the true values. 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 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, SA-MXL may not be an appropriate approach.

We report three interesting findings based on Tables A1, A2 and A3. 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 A1, rather than the one as in Figure 1 in the main text.

Table A1. 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 A2. 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 A3. 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 mean of WTP	9.6101 [0.57 - 20.28]	63.3339 [53.71 - 71.94]	145.9910 [119.23 - 173.70]
GW-MNL (min. AIC)	1.7086 [0.16 - 4.65]	34.3477 [26.84 - 40.62]	94.9462 [68.29 - 129.28]
GW-MNL (min. MAPE)	2.1241 [0.13 - 5.72]	42.1818 [30.17 - 51.70]	114.5984 [84.91 - 150.12]

References

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

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.

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.

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

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.

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.

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.

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.

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

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

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.

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.

Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. Journal of Choice Modelling, 23, 1-8.

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.

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.

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.

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.

LeSage, J. P., Smith, T. E., 2004. A bayesian probit model with spatial dependencies. Spatial and Spatiotemporal Econometrics 127-160.

McFadden, D., 1974. Conditional Logit Analysis of Qualititative Choice Behaviour. Frontiers in Econometrics, P. Zarembka, ed., Academic Press, New York, NY, 105-142.

Train, K., and Sonnier, G., 2005. Mixed Logit with Bounded Distributions of Correlated Partworths. Applications of Simulation Methods in Environmental and Resource Economics, R. Scarpa and A. Alberini, eds., Springer-Verlag, 117-134.

Train, K. E., 2009. Discrete Choice Methods with Simulation. 2 Ed., Cambridge University Press, New York.

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.