

Sarrias, Mauricio

Article

Random parameters and spatial heterogeneity using Rchoice in R

REGION

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Sarrias, Mauricio (2020) : Random parameters and spatial heterogeneity using Rchoice in R, REGION, ISSN 2409-5370, European Regional Science Association (ERSA), Louvain-la-Neuve, Vol. 7, Iss. 1, pp. 1-19, <https://doi.org/10.18335/region.v7i1.279>

This Version is available at:

<https://hdl.handle.net/10419/235808>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc/4.0>

Random Parameters and Spatial Heterogeneity using *Rchoice* package in *R*

Mauricio Sarrias¹

¹ Universidad Católica del Norte, Antofagasta, Chile

Received: 17 September 2019/Accepted: 4 February 2020

Abstract. This document provides a brief introduction to models with spatial heterogeneity using a random parameter approach. Specifically, this paper shows how this modelling strategy can be used to capture and model spatial heterogeneity and locally varying coefficients for different latent structure. To show the main advantages of this modeling strategy, the *Rchoice* package (Sarrias 2016) in *R* is used. The examples will be focused on the ordered probit model with spatially varying coefficients using self-assessed health status as the dependent variable.

1 Introduction

Regional scientists, as well as many social researchers concerned on spatial relationships, analyze how the reciprocal geographical interaction of social agents generates spatial autocorrelation, affecting the bias and efficiency of standard econometric estimators. After Anselin (1988), a large number of papers dealt with the spatial autocorrelation using spatial versions of standard linear regression models, namely Spatial Autoregressive Regression (SAR) or Spatial Error Model (SEM) and even recent contributions extend the analysis toward Spatial Panel Data (Kelejian, Prucha 1998, 1999, Elhorst 2014). However, spatial interaction also is manifested through spatially varying coefficients referred to as: “structural instability over space, in the form of different response functions or systematically varying parameters” (Anselin 1988). In spite of the relevance of the concept, the evolution and development of econometric models that attempt to capture and model spatial heterogeneity has not been as euphoric as those focused on the spatial autocorrelation. The few attempts to capture this heterogeneity can be summarized by the spatial expansion method (SEM) (Casetti 1972), Geographically Weighted Regression (GWR) (Brunsdon et al. 1998) or assuming that the local relationship varies randomly over geographical space, a method also known as the Random Coefficient Model (RCM) (Swamy 1971). Each one of these methods enable estimation of model parameters locally, or they allow model parameter to vary as a function of location¹.

The three methods presented above share an important limitation: they require aggregating the variables at the location level. Therefore, we are prevented from using data at the individual level and capturing the spatial heterogeneity, simultaneously. This raises concerns about the misleading conclusions that can be derived at the individual level by using aggregate variables known as the ecological fallacy problem (Robinson 1950). A potential solution for this constraint is provided by multilevel modeling². This

¹For further review see for example Fotheringham, Brunsdon (1999).

²For other quantitative methods that avoid the ecological fallacy problem see Withers (2001).

approach separates the effect of personal and place characteristics to investigate the extent and nature of spatial variation in individual outcome measures (Goldstein 1987). The main drawback of multilevel modeling is that usually the random coefficients are assumed to be normally distributed. This makes the estimation process easier, but creates other problems. For example, this assumption implies that some locations might have positive or negative coefficients, whether or not this is true. In practice, this implies that occasionally researchers find sign reversals that are counterintuitive and difficult to explain. Furthermore, the domain of the normal distribution is $(-\infty, +\infty)$, which results in unreliable extreme coefficients and high coefficient variability. Those problems have also been found when applying the GWR approach (Jetz et al. 2005)³.

This study focusses on models with spatially varying coefficients using simulation as in Sarrias (2019) and Train (2009). This modeling strategy is intended to complement the existing approaches by using variables at the micro level – overcoming the problem associated with spatial aggregation – and by adding flexibility and realism to the potential domain of the coefficient on the geographical space. Spatial heterogeneity is modelled by allowing the parameters associated with each observed variable to vary “randomly” across space according to some distribution $g(\cdot)$. However, it is not known how the parameters vary across space. All that is known is that they vary locally with population probability density function (pdf) $g(\cdot)$, which is assumed to be well behaved and continuous.

To show the main advantages of this modeling strategy, the `Rchoice` package (Sarrias 2016) in R is used. The examples will be focused on the ordered probit model with spatially varying coefficients using self-assessed health status as the dependent variable.

The remainder of this paper is organized as follows. Section 2 discusses the modelling approach for incorporating continuous spatial heterogeneity using a random parameter approach. The main R packages needed for the examples are described in Section 3. The example using `Rchoice` package in R is presented in Section 4. Finally, Section 5 concludes.

2 Modelling approach

2.1 Continuous spatial heterogeneity

Consider the following structural model:

$$\begin{aligned} y_{ci}^* &= \mathbf{x}'_{ci}\boldsymbol{\beta}_c + \epsilon_{ci} & c = 1, \dots, C; \quad i = 1, \dots, n_c \\ \boldsymbol{\beta}_c &\sim g(\boldsymbol{\beta}_c) \end{aligned} \quad (1)$$

where y_{ci}^* is a latent (unobserved) process for individual i in geographical area c (e.g. region, city, country, census track) that we are trying to explain; \mathbf{x}_{ci} is a $K \times 1$ vector of individual and regional variables; and ϵ_i is the error term⁴. It is assumed that the vector $(y_{ci}, \mathbf{x}'_{ci}, \boldsymbol{\beta}'_c)'$ is independently and identically distributed. The conditional probability density function of the latent process, $f^*(y_{ic}|\mathbf{x}_{ci}, \boldsymbol{\epsilon}_c)$, is determined once the nature of the observed y_{ci} and the population pdf of ϵ_i is known. For example, if the observed y_{ci} is binary and ϵ_i is normal distributed, we obtain the traditional probit model. But if ϵ_i is distributed as logistic, then we obtain the binary logit model. Due to space restrictions, the applied example in this study will focus on the ordered model.

The key element in the structural model is $\boldsymbol{\beta}_c$. The notation implies that coefficients are associated with region c , representing those region-specific partial correlations on the latent dependent variable. Thus, all individuals located in the same region have the same coefficient, but there exists inter-spatial heterogeneity, i.e., the coefficients vary across regions but not within the region.

³There are some interesting extensions that have been recently developed. For example, Dong et al. (2015) extend the traditional multilevel models to incorporate spatial interaction effects at different level units. Dong et al. (2018) extend the GWR for ordinal categorical responses. Bayesian spatially varying coefficient models have been also suggested by Finley (2011) and Gelfand et al. (2003).

⁴Throughout this work I will use location unit, region, or geographical area interchangeably to refer to the subindex c .

However, we do not know how these parameters vary across regions. All we know is that they vary locally with population pdf $g(\beta_c)$. Once $g(\beta_c)$ is specified, we might have a fully parametric or a semi-parametric spatially random parameter model.

2.2 Choosing the distribution

Continuous spatial heterogeneity is introduced by assuming that the parameters vary “randomly” across regions according to some pre-specified “continuous” distribution. The pdf of the spatially random coefficients in the population is $g(\beta_c|\theta)$, where θ represents, for example, the mean and variance of β_c . The goal for the researcher is to estimate θ .

The distribution of the spatially random parameters can in principle take any shape. The researcher has to choose a priori the distribution according to his beliefs of the domain and boundedness of the coefficients.

Therefore, some prior theoretical knowledge of the spatial structure being modeled may lead to a more appropriate choice of the distribution. Below, some continuous distributions and their implications are discussed.

Normal Distribution: The normal distribution is by far the most widely used distribution for the spatially random parameters. The density of the normal parameter has mean β and standard deviation σ_β , so that $\theta = (\beta, \sigma_\beta)'$. Thus, the coefficient for each region can be written as $\beta_c = \beta + \sigma_\beta \eta_c$, where $\eta_c \sim N(0, 1)$. An important feature of the normal density is its unboundedness. This implies that every real number has a positive probability of being drawn. Thus, specifying a given coefficient to follow a normal distribution is equivalent to making the a priori assumption that there is a proportion of regions with positive coefficients and another proportion with negative ones. As an illustration, consider a normally distributed coefficient with population parameters $\beta = 0.5$ and $\sigma_\beta = 1$. The proportion of regions with positive coefficients is approximately $\Phi(\beta/\sigma_\beta) \cdot 100 \approx 70\%$. This last fact makes this distribution quite suitable when the researcher assumes that the effect of x_k on y^* can have both signs depending in the local context of each region. For example, there exists an extensive literature that uses the city population as a proxy for urbanization economies (see for example [Duranton, Puga 2004](#)). However, in some regions, a large population may suggest agglomeration economies, while in others, it may suggest congestion effects ([Ali et al. 2007](#)). In other words, β_c for the population density can take positive or negative values across space. The normal distribution can be also used as an initial exploratory analysis to determine the domain of a coefficient. For example, if the estimated parameters are $\hat{\beta} = 2$ and $\hat{\sigma}_\beta = 1$, this implies that approximately $\Phi(\hat{\beta}/\hat{\sigma}_\beta) \cdot 100 \approx 98\%$ of the regions in the sample have a positive coefficient. Therefore, the researcher may be more inclined to choose a distribution with just a positive real domain. One major disadvantage of the normal distribution is that it has infinite tails, which might result in some regions having implausible extreme coefficient values.

Triangular Distribution: This is a continuous probability distribution with probability density function shaped like a triangle. The advantage of this distribution is that it has a definite upper and lower limit, so its tails are shorter than the normal distribution and we avoid extreme coefficients that may result for some regions. The density of a triangular distribution with mean β and spread s_β is zero beyond the range $(\beta - s_\beta, \beta + s_\beta)$, rises linearly from $\beta - s_\beta$ to β , and drops linearly to $\beta + s_\beta$. The parameters $\theta = (\beta, s_\beta)'$ are estimated.

Uniform Distribution: In this case the parameter for each location is equally likely to take on any value in some interval. Suppose that the spread of the uniform distribution is s_β , such that the parameter is uniformly distributed from $\beta - s_\beta$ to $\beta + s_\beta$. Then the parameter can be constructed as $\beta_c = \beta + s_\beta(2u_c - 1)$ where $u_c \sim U[0, 1]$ and the parameters $\theta = (\beta, s_\beta)$ are estimated. The new random draw $(2u_c - 1)$ is distributed as $U[-1, +1]$, therefore multiplying by s_β gives a uniformly distributed parameter $\pm s$ ([Train 2009](#), [Hensher, Greene 2003](#)). The standard deviation of the uniform distribution can be derived from the spread by dividing s_β

by $\sqrt{3}$. Note also that the uniform distribution with a $[0, 1]$ bound is very suitable when there exists spatial heterogeneity in a dummy variable. For this case the restriction is $\beta = s_\beta = 1/2$.

The normal, triangular and uniform distributions permit positive and negative coefficients. However, as I discussed above, the coefficient may present spatial heterogeneity but only in the positive or negative domain. For example, we may be confident that the coefficient for x_k is positive for all regions, but still there may exist spatial heterogeneity around the mean. Some widely used distributions with domain in the positive numbers are the log-normal, truncated normal, and Johnson S_b distribution⁵.

Log-normal Distribution: The support of the log-normal distribution is $(0, \infty)$. Formally, the coefficient for each region is specified as $\beta_c = \exp(\beta + \sigma_\beta \eta_c)$ where $\eta_c \sim N(0, 1)$. The parameters β and σ_β , which represent the mean and standard deviation of $\log(\beta_c)$, are estimated. The median, mean, and standard deviation of β_c are $\exp(\beta_c)$, $\exp(\beta_c + \sigma_\beta^2/2)$ and $\text{mean} \times \sqrt{\exp(\sigma_\beta^2) - 1}$, respectively (Revelt, Train 1998, Train 2009). The main drawback of the log-normal distribution is that it has a very long right-hand tail. This means that we might find regions with unreasonable extreme positive coefficients.

Truncated Normal Distribution: The domain of this distribution is $(0, \infty)$ if the normal distribution is truncated below at zero. The parameter for each region is created as $\beta_c = \max(0, \beta + \sigma_\beta \eta_c)$ where $\eta_c \sim N(0, 1)$ with the share below zero massed at zero equal to $\Phi(-\beta/\sigma_\beta)$. A normal distribution truncated at 0 can be useful when the researcher has a priori belief that for some regions the marginal latent effect of the variable is null. The parameters $\theta = (\beta, \sigma_\beta)$ are estimated.

Johnson S_b Distribution: The S_b distribution gives coefficients between 0 and 1, which is also very suitable for dummy variables. The parameter for region c is computed as $\beta_c = \frac{\exp(\beta + \sigma_\beta \eta_c)}{1 + \exp(\beta + \sigma_\beta \eta_c)}$ where $\eta_c \sim N(0, 1)$ and the parameters β and σ_β are estimated. The mean, variance, and shape are determined by the mean and variance of $\beta + \sigma_\beta \eta_c$ which is a normal distributed parameter. If the analyst needs the coefficient to be between 0 and k , then the variable can be multiplied by k . The logic behind this is the following. Since $\beta_c \times x_{ic}$ ranges between $[0, 1]$, then $\beta_c \times k \times x_{ic}$ is the same as $k[0, 1] = [0, k]$. The advantage of the Johnson S_b is that it can be shaped like log-normal distribution, but with thinner tails below the bound.

For any distribution, all the information about the unobserved spatial heterogeneity is captured by the spread or standard deviation parameter. For example, a significant standard deviation would reveal a spatially non-stationary relationship, and the higher the standard deviation the higher the unobserved spatial heterogeneity in the parameters. Finally, it is worth noting that if only the constant is assumed to be random, then the model is reduced to the random effect model also known as the spatially constant random parameter in the multilevel context (Jones 1991). If $n_c = 1$ for all C , then the model is reduced to the RCM.

2.3 Correlated spatially random parameters and observed variations around the mean

The random parameters can be generalized to include correlation across the parameters. For example, we may be interested in whether regions with greater (lower) β_1 have also greater (lower) values for β_2 . If it is true, we would say that both effects are positively correlated within regions. Furthermore, it is likely that the association between y_{ci}^* and x_{ci} is modified by unmeasured regional effects or region-specific unobserved factors. Therefore, by allowing the constant and the slope parameter to be correlated we might be able to identify whether those unobserved factors and the effect of x_{ci} are positively or negatively associated.

⁵If some coefficient is expected a priori to be negative for all the regions, one might create the negative of the variable and then include this new variable in the estimation. This “trick” allows the coefficient to be negative without imposing a sign change in the estimation procedure (Train 2009).

As an illustration of the usefulness of the correlated parameters, [Wheeler, Tiefelsdorf \(2005\)](#) raise the awareness of the potential dependencies (correlation) among the local regression coefficients associated with different exogenous variables in the GWR context. They use a GWR approach to explain the white male bladder cancer mortality rates in the 508 States Economic Areas of the United States. Using the population density and smoking as covariates, they find that those regions with high smoking parameter also have a low population density parameter. As they state, the important question is whether this negative correlation is real or an artifact of the statistical method. By allowing the parameters to be explicitly correlated, we are able to test whether the correlation among the parameters is in fact significant⁶.

For simplicity of the notation, consider that the coefficients are distributed across space following a multivariate normal distribution, $\beta_c \sim \text{MVN}(\beta, \Sigma)$. In this case, the coefficient can be written as:

$$\beta_c = \beta + \mathbf{L}\eta_c,$$

where $\eta_c \sim N(\mathbf{0}, \mathbf{I})$, and \mathbf{L} is the lower-triangular Cholesky factor of Σ such that $\mathbf{L}\mathbf{L}' = \text{var}(\beta_c) = \Sigma$. When the off-diagonal elements of \mathbf{L} are zero, the parameters are independently normal distributed. If we assume that the model has only one covariate and the constant, then the extended form of the spatially random coefficient vector is

$$\begin{pmatrix} \alpha_c \\ \beta_c \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \sigma_{\alpha\alpha} & 0 \\ \sigma_{\beta\alpha} & \sigma_{\beta\beta} \end{pmatrix} \begin{pmatrix} \eta_{c\alpha} \\ \eta_{c\beta} \end{pmatrix}$$

$$\beta_c = \beta + \mathbf{L}\eta_c,$$

where:

$$\mathbf{L}\mathbf{L}' = \begin{pmatrix} \sigma_{\alpha\alpha} & 0 \\ \sigma_{\beta\alpha} & \sigma_{\beta\beta} \end{pmatrix} \begin{pmatrix} \sigma_{\alpha\alpha} & \sigma_{\beta\alpha} \\ 0 & \sigma_{\beta\beta} \end{pmatrix} = \begin{pmatrix} \sigma_{\alpha\alpha}^2 & \sigma_{\alpha\alpha}\sigma_{\beta\alpha} \\ \sigma_{\beta\alpha}\sigma_{\alpha\alpha} & \sigma_{\beta\alpha}^2 + \sigma_{\beta\beta}^2 \end{pmatrix} = \Sigma$$

If we need correlated parameters with positive domain, we might create a log-normal distributed parameter. For instance, if we need β_c to be log-normal distributed, then we can transform it in the following way:

$$\beta_c = \exp(\beta + \sigma_{\beta\alpha}\eta_{c\alpha} + \sigma_{\beta\beta}\eta_{c\beta})$$

Observed spatial heterogeneity – or deterministic spatial heterogeneity – can be also accommodated in the random parameters by including region-specific covariates. Specifically, the vector of random coefficient is:

$$\beta_c = \beta + \boldsymbol{\pi}z_c + \mathbf{L}\eta_c \quad (2)$$

where z_c is a set of M characteristics of region c that influences the mean of the spatial random coefficients, and $\boldsymbol{\pi}$ is a $K \times M$ matrix of additional parameters. The conditional mean becomes $E(\beta_c | z_c) = \beta + \boldsymbol{\pi}z_c$. The main drawback of this modeling strategy – and any type of spatial heterogeneity in the form of unobserved spatial heterogeneity – is that it assumes that the coefficients are drawn from some univariate or multivariate distribution and no attention is paid to the location of the regions ([Fotheringham, Brunsdon 1999](#)). However, the previous model can be very useful to consider regions' location explicitly in the random parameters if z_c includes any function of the geographical coordinates $(\mathbf{u}_c, \mathbf{v}_c)$. Thus, if $z_c = h(\mathbf{u}_c, \mathbf{v}_c)$, where $h(\cdot)$ is any function, and $\eta_c = \mathbf{0}$, then the model collapses into the Casetti's spatial expansion method.

⁶Those readers interested in modelling both spatial dependence and spatial heterogeneity are referred to [Dong et al. \(2016\)](#). They develop a spatial random slope multilevel modeling approach to account for the within-group dependence among individuals in the same area and the spatial dependence between areas simultaneously.

2.4 Estimation

Let $\mathbf{y}_c = \{y_{i1}, y_{i2}, \dots, y_{in}\}$ be the sequence of choices for all individuals in region c , where n_c is the total number of individuals in that region. Assuming that individuals are independent across regions, the joint probability density function, given β_c , can be written as

$$Pr(\mathbf{y}_c | \mathbf{X}_c, \beta_c) = \sum_{i=1}^{n_c} f^*(y_{ic} | \mathbf{x}_{ic}, \beta_c), \quad (3)$$

because, conditional on β_c , the observations are independent. Since β_c is common for individuals living in the region c , within each region individuals are not independent. Thus, the unconditional pdf of \mathbf{y}_c given \mathbf{X}_c will be the weighted average of the conditional probability evaluated over all possible values of β , which depends on the parameters of the distribution of β_c :

$$P_c(\boldsymbol{\theta}) = f(\mathbf{y}_c | \mathbf{X}_c, \boldsymbol{\theta}) = \int_{\beta_c} \left[\prod_{i=1}^{N_c} f^*(y_{ic} | \mathbf{x}_{ic}, \beta_c, \boldsymbol{\theta}) \right] g(\beta_c) d\beta_c, \quad (4)$$

The unconditional probability has no closed form solution, therefore the log-likelihood function is difficult to compute. However, we can simulate this probability and use the simulated maximum likelihood in order to estimate $\boldsymbol{\theta}$ (Gourieroux, Monfort 1997, Hajivassiliou, Ruud 1986, Stern 1997, Train 2009)⁷. In particular, $P_c(\boldsymbol{\theta})$ is approximated by a summation over randomly chosen values of β_c . For a given value of the parameters $\boldsymbol{\theta}$, a value of β_c is drawn from its distribution. Using this draw of β_c , $P_c(\boldsymbol{\theta})$ is calculated. This process is repeated for many draws, and the average over the draws is the simulated probability. Formally, the simulated probability for region c is

$$\tilde{P}_c(\boldsymbol{\theta}) = \frac{1}{R} \sum_{r=1}^R \prod_{i=1}^{N_c} \tilde{P}_{icr}(\boldsymbol{\theta}) \quad (5)$$

where \tilde{P}_{icr} is the probability for individual i in region c evaluated at the r^{th} draw of β_c , and R is the total number of draws. Then, the simulated log-likelihood function is:

$$\log L_s = \sum_{c=1}^C \log \left[\frac{1}{R} \sum_{r=1}^R \prod_{i=1}^{N_c} \tilde{P}_{icr}(\boldsymbol{\theta}) \right] \quad (6)$$

Lee (1992), Gourieroux, Monfort (1991) and Hajivassiliou, Ruud (1986) derive the asymptotic distribution of the simulated maximum likelihood (SML) estimator based on smooth probability simulators with the number of draws increasing with sample size. Under regularity conditions, the estimator is consistent and asymptotically normal. When the number of draws, R , rises faster than the square root of the number of observations, the estimator is asymptotically equivalent to the maximum likelihood estimator. It is worth noting that, even though the simulated probability is an unbiased estimate of the true probability, the log of the simulated probability with fixed number of repetitions is not an unbiased estimate of the log of the true probability. This bias in the SML decreases as the number of draws increases (see for example Gourieroux, Monfort 1997, Revelt, Train 1998).

One main limitation of these modeling strategies is that the performance of the maximum likelihood estimators may not be accurate or satisfactory when the number of individuals per region is large. The problem is that the log-likelihood function involves the integration or summation over a term involving the product of the probabilities for all the individuals in each location c . Borjas, Sueyoshi (1994) were the first in noticing this problem in the context of the probit model with random effects and using Gauss quadrature. Lee (2000) also gives more insights about this problem. For example,

⁷Other methods can be used in order to approximate the integrals. For example, Gauss-Hermite quadrature procedure is another numerical method widely used. However, it has been documented that for models with more than 3 random parameters SML performs better. Bayesian estimation is also suitable for continuous spatial heterogeneity. See for example Hashiguchi, Tanaka (2014).

assuming a sample of 500 individuals per group – or regions in our case – with a likelihood contribution of 0.5 per observation, [Borjas, Sueyoshi \(1994\)](#) show that the value of the integrand can be as small as $\exp(500 \times \ln(0.5)) \approx \exp(-346.6)$, which is below the existing absolute value for a computer. A consequence of this might be larger standard errors, explosive estimates and/or a singular Hessian. In the worst scenario, the computation will overflow, that is, it will exceed the computer's capacity to compute the value and the maximization procedure will stop. This issue should be borne in mind when applying these methods⁸.

3 Packages and dependencies

The main R packages used in the examples are the following:

Rchoice: This is the main package to estimate Binary, Poisson, and Ordered Models with Random parameters.

foreign: This package is used to read data in different formats (Stata, SPSS, etc).

car: This package will allow us to perform linear hypotheses.

lntest: This package has generic functions that allow to perform likelihood ratio tests for nested models.

All these packages can be installed using the `install.packages()` function.

4 Application using Rchoice in R: Self-assessed health status

4.1 Ordered Probit model with spatially homogeneous parameters

Suppose we are interested in the determinants of individuals' subjective evaluation of health. We assume that the health status of individual i in municipality c , h_{ic} , follows an underlying continuous but latent health process h_{ic}^* based on a linear combination of individual and municipal covariates given by:

$$h_{ic}^* = \mathbf{x}'_{ic}\boldsymbol{\beta} + \epsilon_{ic} \quad (7)$$

where \mathbf{x}_{ic} is a vector of individual and municipal characteristics; $\epsilon_{ic} \sim N(0, \sigma)$ is the error term, but since the scale of h_{ic}^* is not identified, we normalized $\sigma = 1$. Note that this model assumes that the partial correlation between the latent health status and the covariates follows a spatially stationary process.

As typical in ordered models, we do not observe h_{ic}^* , but we instead observe the self-assessed health status (SAH) for each individual, h_{ic} , which ranges between 1 (very bad health) and 5 (very good health) in our sample. The link between h_{ic} and h_{ic}^* is the following:

$$h_{ic} = \begin{cases} 1 & \text{if } \kappa_0 < h_{ic}^* < \kappa_1 \\ 2 & \text{if } \kappa_1 < h_{ic}^* < \kappa_2 \\ \vdots & \\ 5 & \text{if } \kappa_4 < h_{ic}^* < \kappa_5 \end{cases}$$

where it is assumed that $\kappa_0 = -\infty$ and $\kappa_5 = \infty$ to cover the entire real line. Since having a constant is useful in our model to accommodate random effects, we set $x_{1ic} \equiv 1$ for all $i = 1, \dots, N$. Therefore, for identification we set $\kappa_1 \equiv 0$.

To estimate an ordered probit model with spatially homogeneous coefficients, we will use the **Rchoice** package which is loaded using the `library()` function:

```
[1]: > # Load package
      > library("Rchoice")
```

⁸For other estimation methods, such as Bayesian estimation of multi-level models, see for example [Bürkner \(2018\)](#).

Now, we load the dataset `sah.chile`. This dataset comes from the 2013 National Socioeconomic Characterization Survey (CASEN) from Chile. CASEN is a national, population-based survey which is representative at the municipal level and is carried out by the Ministry of Planning (MIDEPLAN) to describe the socioeconomic situation as well as the impact of social programs on the living conditions of the Chilean Population⁹.

In the following lines, the dataset in Stata format is downloaded. Then, the SAH variable (dependent variable) is recoded into 5 categories:

```
[2]: > # Load data set and recode SAH variable
> library("foreign") # package to load datasets
> library("car") # package with recode function
> data <- read.dta("https://msarrias.weebly.com/uploads/3/7/7/8/37783629/sah.chile.dta")
> data$sah2 <- recode(data$shealth, "1= 1; 2 = 2; 3 = 3; 5:6 = 4; 7 = 5")
```

The vector \mathbf{x}_{ic} includes the following controls at the individual level:

linch: log of household income.

agen: age in years / 10.

hsizen: household size / 10.

edun: years of schooling / 10.

male: =1 for men.

dcivil1: =1 if the individual is married.

dlstatus2: =1 if the individual is unemployed.

Some continuous variables are divided by 10 to improve convergence speed of the SML process and avoid singularities in the Hessian matrix.

In addition, a set of dummy variables indicating the self-perception of pollution and environmental problems is used. The dummy variables are obtained from the response to the question: *“What problems related to pollution and environmental degradation do you identify in your neighbourhood or location”*. Based on the answer, dummy variables were created for the following problems:

noise: noise pollution.

airpol: air pollution.

watpol: water pollution.

vispol: visual pollution.

waspol: garbage (rubbish) in the neighborhood.

The variables at the municipality level are:

lmdinc: log of median income (proxy for development).

lpop: log of population (size effect).

The following command lines show how to estimate the traditional ordered probit model. For other models such as the Binary (Logit and Probit) and Poisson model see [Sarrias \(2016\)](#) Sarrias (2016).

```
[3]: > # Ordered probit model
> oprobit <- Rchoice(sah2 ~ linch + agen + hsizen + edun + male +
+   dcivil1 + dlstatus2 +
+     noise + airpol + watpol + vispol + waspol +
+     lmdinc + lpop,
+     family = ordinal("probit"),
+     data = data)
> summary(oprobit)
```

⁹Chile has 346 municipalities of which 324 are representative in CASEN 2013.

```
[3]: ##
## Model: ordinal
## Model estimated on: Tue Jan 07 10:31:58 2020
##
## Call:
## Rchoice(formula = sah2 ~ lynch + agen + hsizen + edun + male +
## dcivil1 + dlstatus2 + noise + airpol + watpol + vispol +
## waspol + lmdinc + lpop, data = data, family = ordinal("probit"),
## method = "bfgs")
##
##
## Frequencies of categories:
## y
##      1      2      3      4      5
## 0.01087 0.01359 0.02897 0.64523 0.30133
## The estimation took: 0h:0m:5s
##
## Coefficients:
##              Estimate Std. Error z-value Pr(>|z|)
## kappa.1      0.341698   0.022434  15.231 < 2e-16 ***
## kappa.2      0.720961   0.027012  26.690 < 2e-16 ***
## kappa.3      3.015915   0.031564  95.548 < 2e-16 ***
## constant    -0.290407   0.481404  -0.603 0.546342
## lynch        0.132595   0.014377   9.223 < 2e-16 ***
## agen        -0.212654   0.008214 -25.889 < 2e-16 ***
## hsizen       0.242645   0.060359   4.020 5.82e-05 ***
## edun         0.211808   0.028537   7.422 1.15e-13 ***
## male         0.154095   0.019058   8.086 6.66e-16 ***
## dcivil1     -0.028142   0.021246  -1.325 0.185316
## dlstatus2   -0.093728   0.046169  -2.030 0.042347 *
## noise       -0.139891   0.026745  -5.231 1.69e-07 ***
## airpol      -0.084563   0.026219  -3.225 0.001259 **
## watpol      -0.120485   0.037756  -3.191 0.001417 **
## vispol      -0.069064   0.060747  -1.137 0.255573
## waspol      -0.041040   0.027421  -1.497 0.134482
## lmdinc       0.147852   0.043400   3.407 0.000658 ***
## lpop        -0.016493   0.009016  -1.829 0.067359 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Optimization of log-likelihood by BFGS maximization
## Log Likelihood: -13020
## Number of observations: 16188
## Number of iterations: 190
## Exit of MLE: successful convergence
```

The argument `family = ordinal("probit")` indicates that an ordered probit model will be estimated. If the user wants an ordered logit model the argument should be `family = ordinal("logit")`. For other models, see `help(Rchoice)`.

The results show that household income and education increase the probability of reporting better health status, whereas age decreases it. Men are more likely to report better health than women and being unemployed is detrimental for health. At the neighborhood level, noise, air, and water pollution reduce health perception and `vispol` and `waspol` apparently do not matter for health. The coefficient for the logarithm of population for each municipality, which is intended to capture agglomeration effects, is negative but weakly significant, whereas the level of development is positively correlated with individuals' health evaluation.

4.2 Ordered Probit models with spatial random coefficients

The standard ordered probit model does not allow for spatial heterogeneity in the coefficients. In this section, we estimate an Ordered Probit with Spatial Random Parameters (OPSRP) model. To reduce excessive computing time, we will only assume that the variables at the level of municipalities and neighborhood vary across space.

The first and more difficult task is to choose the distribution for each of them. As explained by [Hensher, Greene \(2003\)](#), distributions are essentially arbitrary approximations to the real behavioral profile. The researcher chooses a specific distribution because he has

a sense that the “empirical truth” is somewhere in their domain. The most widely used distribution in the empirical literature is the normal distribution due to its properties. If unobserved spatial heterogeneity is viewed as the sum of small random influences, then the central limit theorem can be invoked to justify the normality assumption [Greene, Hensher \(2010\)](#). Moreover, the normal distribution is unbounded, and therefore every real number has a positive probability of being drawn. Thus, specifying a given coefficient to follow a normal distribution is equivalent to making a priori assumption that both positive and negative coefficients exists across space ([Sarrias 2019](#)).

This last property is very appealing in our case, since theoretically we might observe municipalities with positive and negative sign for the population coefficient. For instance, municipalities with a positive coefficient might be characterized for having positive urban externalities that outweigh the negative ones. In those municipalities, inhabitants, on average, enjoy better health through local positive urban externalities. If the coefficient is negative, the opposite might be expected.

A OPSRP model with normally distributed parameters is estimated as follows:

```
[4]: > # Spatial random parameter model
> ran_1 <- Rchoice(sah2 ~ lynch + agen + hsize + edun + male + dcivil1 + dlstatus2 +
+ noise + airpol + watpol + vispol + waspol +
+ lmdinc + lpop,
+ family = ordinal('probit'),
+ data = data,
+ rang = c(noise = "n", airpol = "n", watpol = "n", vispol = "n",
+ waspol = "n", lmdinc = "n", lpop = "n"),
+ panel = TRUE,
+ index = "idc",
+ R=30,
+ method = "bfgs")
> summary(ran_1)
```

```
[4]: ##
## Model: ordinal
## Model estimated on: Tue Dec 31 09:11:29 2019
##
## Call:
## Rchoice(formula = sah2 ~ lynch + agen + hsize + edun + male +
## dcivil1 + dlstatus2 + noise + airpol + watpol + vispol +
## waspol + lmdinc + lpop, data = data, family = ordinal("probit"),
## rang = c(noise = "n", airpol = "n", watpol = "n", vispol = "n",
## waspol = "n", lmdinc = "n", lpop = "n"), R = 30, panel = TRUE,
## index = "idc", method = "bfgs", iterlim = 2000)
##
##
## Frequencies of categories:
## y
##      1      2      3      4      5
## 0.01087 0.01359 0.02897 0.64523 0.30133
## The estimation took: 0h:9m:48s
##
## Coefficients:
##              Estimate Std. Error z-value Pr(>|z|)
## kappa.1      0.3417117  0.0162495  21.029 < 2e-16 ***
## kappa.2      0.7209893  0.0209411  34.429 < 2e-16 ***
## kappa.3      3.0170292  0.0268859 112.216 < 2e-16 ***
## constant    -0.2905344  0.4958543  -0.586 0.557925
## lynch        0.1310569  0.0143865   9.110 < 2e-16 ***
## agen        -0.2132221  0.0082107 -25.969 < 2e-16 ***
## hsize        0.2425766  0.0604009   4.016 5.92e-05 ***
## edun         0.2116421  0.0285525   7.412 1.24e-13 ***
## male         0.1540426  0.0190620   8.081 6.66e-16 ***
## dcivil1     -0.0281834  0.0212486  -1.326 0.184717
## dlstatus2   -0.0937187  0.0462089  -2.028 0.042545 *
## mean.noise  -0.1399619  0.0269286  -5.198 2.02e-07 ***
## mean.airpol -0.0845976  0.0264687  -3.196 0.001393 **
## mean.watpol -0.1205199  0.0379779  -3.173 0.001507 **
## mean.vispol -0.0690556  0.0611251  -1.130 0.258585
## mean.waspol -0.0410117  0.0276460  -1.483 0.137953
## mean.lmdinc  0.1463353  0.0445929   3.282 0.001032 **
## mean.lpop   -0.0180293  0.0091612  -1.968 0.049067 *
```

```
## sd.noise      0.0986997  0.0261224  3.778 0.000158 ***
## sd.airpol    0.0986787  0.0254373  3.879 0.000105 ***
## sd.watpol    0.0994208  0.0398379  2.496 0.012573 *
## sd.vispol    0.0998332  0.0688382  1.450 0.146987
## sd.waspol    0.0988309  0.0273009  3.620 0.000295 ***
## sd.lmdinc    0.0050337  0.0008714  5.776 7.64e-09 ***
## sd.lpop      0.0012177  0.0011066  1.100 0.271156
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Optimization of log-likelihood by BFGS maximization
## Log Likelihood: -11320
## Number of observations: 16188
## Number of iterations: 391
## Exit of MLE: successful convergence
## Simulation based on 30 Halton draws
```

The argument `ranp` indicates which variables are random in the formula and their distributions. In this example, all the random variables are assumed to be normally distributed using "n". The remaining distribution discussed in Section 2.2 can be used using the following shorthands:

- Triangular = "t",
- Uniform = "u",
- Truncated normal = "cn",
- Log-normal = "ln",
- Johnson's Sb = "sb".

The number of draws for the simulation of the probability is set using the argument `R`. To keep the estimation time manageable, we use 30 draws for each individual. However, consistency of the SML requires a higher number of draws (see for example Train 2009).

The argument `index` is a string indicating the id for the municipalities in the data, whereas `panel=TRUE` allows for the spatial structure of the sample.

The previous model assumes that the coefficients has the following form:

$$\beta_k = \bar{\beta}_k + \sigma_k v_{ir}$$

where $v_{ir} \sim N(0,1)$. Thus, the coefficients with the `mean.` and `sd.` prefix represent the estimated mean, $\hat{\beta}$, and standard deviation, $\hat{\sigma}$, for variable k , respectively. If $\sigma_k = 0$, then there is no evidence of systematical variation for regression coefficient over space. The output shows that there is evidence of spatial heterogeneity for most of the variables, except for `vispol` and `lpop`.

To test the joint hypothesis of coefficient homogeneity across space we can perform a Likelihood Ratio test using `lrtest` function from `lmtest` package.

```
[5]: > # Testing spatial heterogeneity
> library("lmtest")
> lrtest(oprobit, ran_1)
```

```
[5]: ## Likelihood ratio test
##
## Model 1: sah2 ~ lynch + agen + hsizen + edun + male + dcivil1 + dlstatus2 +
## noise + airpol + watpol + vispol + waspol + lmdinc + lpop
## Model 2: sah2 ~ lynch + agen + hsizen + edun + male + dcivil1 + dlstatus2 +
## noise + airpol + watpol + vispol + waspol + lmdinc + lpop
## #Df LogLik Df Chisq Pr(>Chisq)
## 1 18 -13018
## 2 25 -11318 7 3400.6 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The test rejects the null hypothesis providing empirical evidence of spatial heterogeneity for those variables.

Since the parameters are allowed to vary across space following a normal distribution, we can also compute the proportion of municipalities with positive coefficients using $\Phi(\hat{\beta}/\hat{\sigma})$. For example, for `noise` and `lmdinc` the results are:

```
[6]: > # Computing proportions
> pnorm(coef(ran_1)["mean.noise"] / coef(ran_1)["sd.noise"])
```

```
[6]: ## mean.noise
## 0.07808686
```

```
[7]: > pnorm(coef(ran_1)["mean.lmdinc"] / coef(ran_1)["sd.lmdinc"])
```

```
[7]: ## mean.lmdinc
## 1
```

Thus, we can say that for 100% of the municipalities development is positively correlated with individuals' health, whereas for around 8% of the municipalities, higher noise pollution increases health. This last result can be true or an artifact of the normality assumption.

4.3 Correlated parameters

The previous model specifies the coefficients to be independently distributed, while one would expect correlation. To show this, the model `ran_1` we will be estimated but assuming that the spatially random coefficients are correlated adding the argument `correlation = TRUE`:

```
[8]: > # Spatially random parameters with correlated coefficients
> ran_2 <- Rchoice(sah2 ~ linc + agen + hsize + edun + male + dcivil1 +
+ dlstatus2 + noise + airpol + watpol + vispol + waspol +
+ lmdinc + lpop,
+ family = ordinal("probit"),
+ data = data,
+ ranp = c(noise = "n", airpol = "n", watpol = "n", vispol = "n",
+ waspol = "n", lmdinc = "n", lpop = "n"),
+ panel = TRUE,
+ index = "idc",
+ R=30,
+ method = "bfgs",
+ correlation = TRUE)
> summary(ran_2)
```

```
[8]: ##
## Model: ordinal
## Model estimated on: Tue Dec 31 09:21:06 2019
##
## Call:
## Rchoice(formula = sah2 ~ linc + agen + hsize + edun + male +
## dcivil1 + dlstatus2 + noise + airpol + watpol + vispol +
## waspol + lmdinc + lpop, data = data, family = ordinal("probit"),
## ranp = c(noise = "n", airpol = "n", watpol = "n", vispol = "n",
## waspol = "n", lmdinc = "n", lpop = "n"), R = 30, correlation = TRUE,
## panel = TRUE, index = "idc", method = "bfgs", iterlim = 2000)
##
##
## Frequencies of categories:
## y
## 1 2 3 4 5
## 0.01087 0.01359 0.02897 0.64523 0.30133
## The estimation took: 0h:9m:36s
##
## Coefficients:
## Estimate Std. Error z-value Pr(>|z|)
## kappa.1 0.3412626 0.0180952 18.859 < 2e-16 ***
## kappa.2 0.7195420 0.0226625 31.750 < 2e-16 ***
## kappa.3 3.4970489 0.0294076 118.917 < 2e-16 ***
## constant -0.2901421 0.5322664 -0.545 0.585680
## linc 0.1457106 0.0149581 9.741 < 2e-16 ***
## agen -0.2696671 0.0086621 -31.132 < 2e-16 ***
## hsize 0.2440816 0.0629962 3.875 0.000107 ***
## edun 0.2202293 0.0300012 7.341 2.12e-13 ***
```

```

## male          0.1579018  0.0199386  7.919  2.44e-15 ***
## dcivil1      -0.0361313  0.0222373 -1.625  0.104204
## dlstatus2    -0.0930049  0.0482324 -1.928  0.053822 .
## mean.noise   -0.1446350  0.0281478 -5.138  2.77e-07 ***
## mean.airpol  -0.0855360  0.0275906 -3.100  0.001934 **
## mean.watpol  -0.1204861  0.0399076 -3.019  0.002535 **
## mean.vispol  -0.0689847  0.0636902 -1.083  0.278751
## mean.waspol  -0.0371335  0.0288215 -1.288  0.197609
## mean.lmdinc  0.1607965  0.0478379  3.361  0.000776 ***
## mean.lpop    0.0069806  0.0101792  0.686  0.492860
## sd.noise.noise 0.0050271  0.0333140  0.151  0.880055
## sd.noise.airpol 0.0058448  0.0324393  0.180  0.857013
## sd.noise.watpol 0.0539982  0.0478928  1.127  0.259539
## sd.noise.vispol 0.0739318  0.0776348  0.952  0.340943
## sd.noise.waspol 0.0182864  0.0331744  0.551  0.581483
## sd.noise.lmdinc 0.0618425  0.0086254  7.170  7.51e-13 ***
## sd.noise.lpop -0.0813180  0.0105174 -7.732  1.07e-14 ***
## sd.airpol.airpol 0.0160025  0.0319875  0.500  0.616881
## sd.airpol.watpol 0.0559519  0.0459907  1.217  0.223760
## sd.airpol.vispol 0.0816862  0.0775592  1.053  0.292245
## sd.airpol.waspol 0.0208368  0.0330971  0.630  0.528979
## sd.airpol.lmdinc 0.0434195  0.0084910  5.114  3.16e-07 ***
## sd.airpol.lpop -0.0588680  0.0102640 -5.735  9.73e-09 ***
## sd.watpol.watpol 0.0628180  0.0456531  1.376  0.168826
## sd.watpol.vispol 0.0818528  0.0779919  1.050  0.293946
## sd.watpol.waspol 0.0233655  0.0330245  0.708  0.479244
## sd.watpol.lmdinc 0.0408594  0.0085428  4.783  1.73e-06 ***
## sd.watpol.lpop -0.0505877  0.0103099 -4.907  9.26e-07 ***
## sd.vispol.vispol 0.0850337  0.0779709  1.091  0.275457
## sd.vispol.waspol 0.0187284  0.0330837  0.566  0.571331
## sd.vispol.lmdinc 0.0471032  0.0083197  5.662  1.50e-08 ***
## sd.vispol.lpop -0.0644627  0.0100842 -6.392  1.63e-10 ***
## sd.waspol.waspol 0.0224287  0.0329135  0.681  0.495591
## sd.waspol.lmdinc 0.0491523  0.0082301  5.972  2.34e-09 ***
## sd.waspol.lpop -0.0642418  0.0099655 -6.446  1.15e-10 ***
## sd.lmdinc.lmdinc 0.0353881  0.0082862  4.271  1.95e-05 ***
## sd.lmdinc.lpop -0.0450126  0.0099427 -4.527  5.98e-06 ***
## sd.lpop.lpop  0.0002568  0.0011373  0.226  0.821375
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Optimization of log-likelihood by BFGS maximization
## Log Likelihood: -11210
## Number of observations: 16188
## Number of iterations: 315
## Exit of MLE: successful convergence
## Simulation based on 30 Halton draws

```

It is important to note that the output prints the elements of the lower-triangular Cholesky factor \mathbf{L} . The variance-covariance matrix, $\mathbf{\Sigma}$, can be extracted using the `vcov` function in the following way:

```
[9]: > # Obtain Sigma
> vcov(ran_2, what = "ranp", type = "cov", se = TRUE)
```

```

##
## Elements of the variance-covariance matrix
##
##              Estimate Std. Error z-value Pr(>|z|)
## v.noise.noise  2.5271e-05  3.3494e-04  0.0754  0.93986
## v.noise.airpol  2.9382e-05  2.3810e-04  0.1234  0.90179
## v.noise.watpol  2.7145e-04  1.7934e-03  0.1514  0.87969
## v.noise.vispol  3.7166e-04  2.4696e-03  0.1505  0.88037
## v.noise.waspol  9.1927e-05  6.3478e-04  0.1448  0.88486
## v.noise.lmdinc  3.1089e-04  2.0673e-03  0.1504  0.88046
## v.noise.lpop   -4.0879e-04  2.7200e-03 -0.1503  0.88053
## v.airpol.airpol  2.9024e-04  1.0858e-03  0.2673  0.78923
## v.airpol.watpol  1.2110e-03  2.4984e-03  0.4847  0.62788
## v.airpol.vispol  1.7393e-03  3.7015e-03  0.4699  0.63843
## v.airpol.waspol  4.4032e-04  1.0616e-03  0.4148  0.67829
## v.airpol.lmdinc  1.0563e-03  2.4234e-03  0.4359  0.66293
## v.airpol.lpop   -1.4173e-03  3.2248e-03 -0.4395  0.66029

```

```
## v.watpol.watpol 9.9925e-03 9.2513e-03 1.0801 0.28009
## v.watpol.vispol 1.3705e-02 9.7826e-03 1.4009 0.16124
## v.watpol.waspol 3.6211e-03 3.6712e-03 0.9863 0.32397
## v.watpol.lmdinc 8.3355e-03 4.0026e-03 2.0825 0.03730 *
## v.watpol.lpop -1.0863e-02 5.2619e-03 -2.0644 0.03898 *
## v.vispol.vispol 2.6069e-02 2.4122e-02 1.0807 0.27982
## v.vispol.waspol 6.5591e-03 5.8208e-03 1.1268 0.25981
## v.vispol.lmdinc 1.5469e-02 7.5381e-03 2.0521 0.04016 *
## v.vispol.lpop -2.0443e-02 9.9442e-03 -2.0558 0.03981 *
## v.waspol.waspol 2.1683e-03 2.9746e-03 0.7289 0.46604
## v.waspol.lmdinc 4.9749e-03 3.5556e-03 1.3992 0.16176
## v.waspol.lpop -6.5438e-03 4.7027e-03 -1.3915 0.16407
## v.lmdinc.lmdinc 1.3266e-02 1.8869e-03 7.0309 2.052e-12 ***
## v.lmdinc.lpop -1.7439e-02 2.3778e-03 -7.3341 2.232e-13 ***
## v.lpop.lpop 2.2946e-02 3.0048e-03 7.6365 2.243e-14 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The estimated coefficients represent the variance and covariance of the randomly distributed parameters. Their standard errors are estimated using the Delta Method. To obtain the standard deviations for the random parameters, one might use the following code:

```
[10]: > # Obtain standard deviations
> vcov(ran_2, what = "ranp", type = "sd", se = TRUE)
```

```
[10]: ##
## Standard deviations of the random parameters
##
## Estimate Std. Error z-value Pr(>|z|)
## noise 0.0050271 0.0333140 0.1509 0.88006
## airpol 0.0170365 0.0318660 0.5346 0.59291
## watpol 0.0999627 0.0462737 2.1602 0.03075 *
## vispol 0.1614594 0.0746999 2.1614 0.03066 *
## waspol 0.0465651 0.0319403 1.4579 0.14487
## lmdinc 0.1151790 0.0081909 14.0617 < 2e-16 ***
## lpop 0.1514789 0.0099181 15.2730 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finally, the correlation matrix for the estimated coefficients is:

```
[11]: > # Obtain correlation matrix of estimated coefficients
> vcov(ran_2, what = "ranp", type = "cor")
```

```
[11]: ##          noise      airpol      watpol      vispol      waspol      lmdinc
## noise  1.0000000  0.3430771  0.5401842  0.4578971  0.3927053  0.5369247
## airpol  0.3430771  1.0000000  0.7110814  0.6323119  0.5550469  0.5383007
## watpol  0.5401842  0.7110814  1.0000000  0.8491072  0.7779253  0.7239690
## vispol  0.4578971  0.6323119  0.8491072  1.0000000  0.8724090  0.8317965
## waspol  0.3927053  0.5550469  0.7779253  0.8724090  1.0000000  0.9275751
## lmdinc  0.5369247  0.5383007  0.7239690  0.8317965  0.9275751  1.0000000
## lpop   -0.5368276 -0.5492089 -0.7173734 -0.8358492 -0.9277189 -0.9995225

##          lpop
## noise  -0.5368276
## airpol  -0.5492089
## watpol  -0.7173734
## vispol  -0.8358492
## waspol  -0.9277189
## lmdinc  -0.9995225
## lpop    1.0000000
```

The results show, for example, that noise pollution is positively correlated with other forms of pollution. Therefore, in those municipalities where noise pollution is detrimental to health, so are the other forms of pollution. It is also important to note that the municipalities where noise has a negative effect are also municipalities where lower development and higher population impact negatively the self-perception of health status.

4.4 Region-specific coefficients

In the applied literature it is very common to map the region-specific estimates to display the spatial heterogeneity for certain coefficients. This cannot be done using just the distribution of the parameters across regions, $g(\beta_c|\theta)$. The population distributions give us just the average affect, β , and the spatial variation around this mean, σ_β , when in fact we would like to know where each region's β_c lies in $g(\beta_c|\theta)$. We might be able to find the likely location of a given region on the heterogeneity distribution by moving from the conditional to the unconditional distribution (Revelt, Train 2000, Brunson et al. 1999, Sarrias 2019). Using Bayes' theorem, we obtain:

$$f(\beta_c|\mathbf{y}_c, \mathbf{X}_c, \theta) = \frac{f(\mathbf{y}_c|\mathbf{X}_c, \beta_c)g(\beta_c|\theta)}{f(\mathbf{y}_c|\mathbf{X}_c, \theta)} = \frac{f(\mathbf{y}_c|\mathbf{X}_c, \beta_c)g(\beta_c|\theta)}{\int_{\beta_c} f(\mathbf{y}_c|\mathbf{X}_c, \beta_c)g(\beta_c|\theta)d\beta_c} \quad (8)$$

where $f(\beta_c|\mathbf{y}_c, \mathbf{X}_c, \theta)$ is the distribution of the regional parameters β_c conditional on the sequence of choices of all the individuals in region c , whereas $g(\beta_c|\theta)$ is the unconditional distribution. The conditional expectation of β_c is given by

$$\bar{\beta}_c = E[\beta_c|\mathbf{y}_c, \mathbf{X}_c, \theta] = \frac{\int_{\beta_c} \beta_c f(\mathbf{y}_c|\mathbf{X}_c, \beta_c)g(\beta_c|\theta)d\beta_c}{\int_{\beta_c} f(\mathbf{y}_c|\mathbf{X}_c, \beta_c)g(\beta_c|\theta)d\beta_c} \quad (9)$$

This expectation gives us the conditional mean of the distribution of the spatially random parameter. The simulator for this expectation is:

$$\hat{\beta}_c = \hat{E}[\beta_c|\mathbf{y}_c, \mathbf{X}_c, \hat{\theta}] = \frac{\frac{1}{R} \sum_{r=1}^R \hat{\beta}_{cr} \prod_{i=1}^{n_c} f^*(y_{ci}|\mathbf{x}_{ci}, \hat{\beta}_{cr})}{\frac{1}{R} \sum_{r=1}^R \prod_{i=1}^{n_c} f^*(y_{ci}|\mathbf{x}_{ci}, \hat{\beta}_{cr})} \quad (10)$$

This estimator is the region-specific estimate, and can be computed in Rchoice using `effect.Rchoice` function and plotted using the function `plot`. In the following lines the municipality's coefficient for all the random parameters is plotted using a kernel approximation:

```
[12]: > # Plot municipality-specific coefficient
> par(mfrow = c(3, 3))
> plot(ran_2, par = "noise", type = "density", main = "Noise Pol.")
> plot(ran_2, par = "airpol", type = "density", main = "Air Pol.")
> plot(ran_2, par = "watpol", type = "density", main = "Water Pol.")
> plot(ran_2, par = "vispol", type = "density", main = "Visual Pol.")
> plot(ran_2, par = "waspol", type = "density", main = "Garbage Pol.")
> plot(ran_2, par = "lmdinc", type = "density", main = "Development")
> plot(ran_2, par = "lpop", type = "density", main = "Population")
```

[12]: Output reproduced in Figure 1

The red area under the kernel distribution illustrates the proportion of municipalities with a positive conditional mean. The most relevant result is that size (`lpop`) seems to be a positive externality for almost 50% of the municipalities, evidencing substantial spatial heterogeneity. This important result is obscured by the traditional ordered probit model.

We might also plot the 95% confidence interval for the conditional means of `airpol` and `noise` for the first 50 municipalities by typing:

```
[13]: > # Plot region-specific confidence intervals.
> par(mfrow = c(1, 2))
> plot(ran_2, par = "airpol", ind = TRUE, id = 1:50, ylab = "Municipalities")
> plot(ran_2, par = "noise", ind = TRUE, id = 1:50, ylab = "Municipalities")
```

[13]: Output reproduced in Figure 2

In terms of consistency of the regional-specific estimates, it is expected that $\hat{\beta}_c \xrightarrow{p} \beta_c$ as $n_c \rightarrow \infty$. That is, if we have more information about the choices made by the individuals in each region, then we are in better position to identify where each region coefficient lies on $g(\beta_c)$ (see for example Train 2009, Revelt, Train 2000, Sarrias, Daziano 2018).

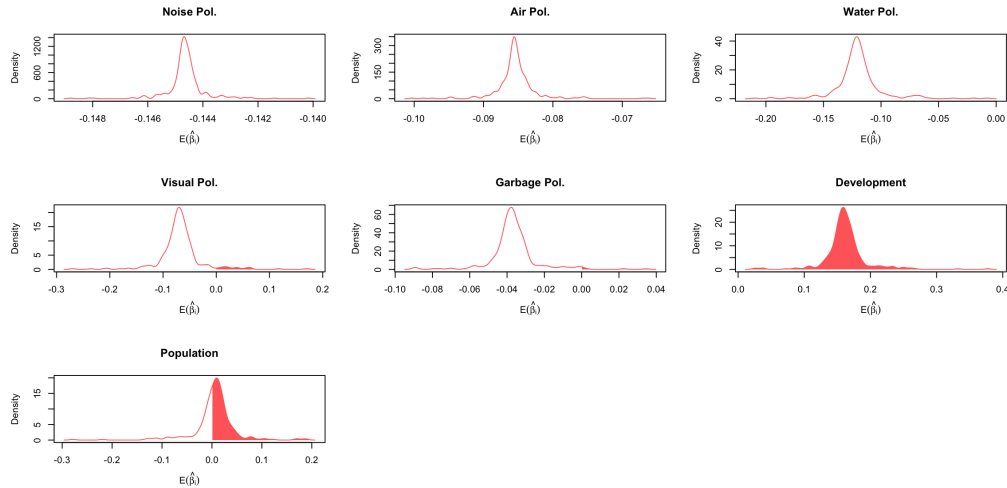


Figure 1: Plot of municipality-specific coefficients

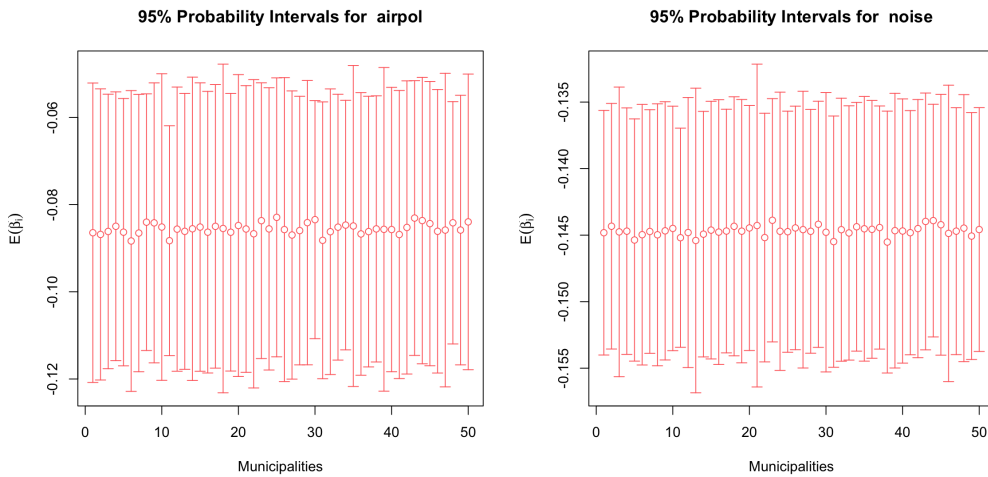


Figure 2: Plot of region-specific confidence intervals

5 Conclusion

This paper contributes to the literature of spatial econometric models that deal with spatially non-stationary process by assuming unobserved heterogeneity. This modelling approach has been widely used in discrete choice modeling, but it can also be implemented to capture and model observed and unobserved spatial heterogeneity. One of the main advantages of this modelling approach is that allows the analyst to include variables at the individual level, which mitigate the ecological fallacy problem, and to add more flexibility regarding the shape and boundedness of the coefficients.

Spatial heterogeneity is represented by some distribution $g(\beta_c)$, which can take any continuous shape, and the analyst must choose the distribution a priori. The choice of the distribution may be guided by theoretical reasons regarding the domain and bound of the coefficients. It also discussed some extensions that can be useful in order to take into consideration the geographical location of the regions, as well as the spatial correlation of the parameters. Although the unobserved spatial heterogeneity using continuous distributions has very appealing features, the probability for each region does not have a closed form solution. Therefore, we need to simulate this probability and estimate the parameters using SML, which can be very costly in terms of computational time.

This study also shows how the `Rchoice` package can be used to estimate this type of models. To do so, we provide a simple example for ordered probit models, focusing on how the determinants of individuals' self-assessed health status might vary across space.

This work can be extended in different ways. First, one of the main concerns and limitations of the model is that the estimation requires computing the product of the probabilities for all individuals in a given region. Thus, if the number of individuals is too high, the estimation method may run into numerical difficulties. To overcome this problem some of the two methods proposed by Lee (2000) can be studied under the spatial context. These methods alleviate the numerical problems by interchanging the inner product with the outer summation. Another possible extension is to study the behavior of the parameters with small and large samples using Bayesian and EM algorithms. Finally, more empirical applications are needed in order to understand the strengths and weaknesses for estimating models with locally varying coefficients using unobserved heterogeneity.

References

- Ali K, Partridge MD, Olfert MR (2007) Can geographically weighted regressions improve regional analysis and policy making? *International Regional Science Review* 30[3]: 300–329. [CrossRef](#).
- Anselin L (1988) *Spatial Econometrics: Methods and Models*, Volume 4. Springer. [CrossRef](#).
- Borjas GJ, Sueyoshi GT (1994) A two-stage estimator for probit models with structural group effects. *Journal of Econometrics* 64[1]: 165–182. [CrossRef](#).
- Bürkner P (2018) Advanced bayesian multilevel modeling with the r package brms. *R Journal* 10[1]. [CrossRef](#).
- Brunsdon C, Aitkin M, Fotheringham S, Charlton M (1999) A comparison of random coefficient modelling and geographically weighted regression for spatially non-stationary regression problems. *Geographical and Environmental Modelling* 3: 47–62
- Brunsdon C, Fotheringham S, Charlton M (1998) Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (the Statistician)* 47[3]: 431–443. [CrossRef](#).
- Casetti E (1972) Generating models by the expansion method: Applications to geographical research. *Geographical Analysis* 4[1]: 81–91. [CrossRef](#).
- Dong G, Harris R, Jones K, Yu J (2015) Multilevel modelling with spatial interaction effects with application to an emerging land market in beijing, china. *PloS One* 10[6]: e0130761. [CrossRef](#).
- Dong G, Ma J, Harris R, Pryce G (2016) Spatial random slope multilevel modeling using multivariate conditional autoregressive models: A case study of subjective travel satisfaction in beijing. *Annals of the American Association of Geographers* 106[1]: 19–35. [CrossRef](#).
- Dong G, Nakaya T, Brunsdon C (2018) Geographically weighted regression models for ordinal categorical response variables: An application to geo-referenced life satisfaction data. *Computers, Environment and Urban Systems* 70: 35–42. [CrossRef](#).
- Duranton G, Puga D (2004) Micro-foundations of urban agglomeration economies. In: Henderson V, Thisse J (eds), *Handbook of Regional and Urban Economics*, Volume 4. Elsevier, 2063–2117. [CrossRef](#).
- Elhorst JP (2014) *Spatial Panel Data Models*. Springer. [CrossRef](#).

- Finley AO (2011) Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution* 2[2]: 143–154. [CrossRef](#).
- Fotheringham AS, Brunsdon C (1999) Local forms of spatial analysis. *Geographical Analysis* 31[4]: 340–358. [CrossRef](#).
- Gelfand AE, Kim HJ, Sirmans CF, Banerjee S (2003) Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* 98[462]: 387–396. [CrossRef](#).
- Goldstein H (1987) *Multilevel Models in Education and Social Research*. Oxford University Press
- Gourieroux C, Monfort A (1991) Simulation based inference in models with heterogeneity. *Annales d'Economie et de Statistique* 20-21: 69–107. [CrossRef](#).
- Gourieroux C, Monfort A (1997) *Simulation-Based Econometric Methods*. Oxford University Press. [CrossRef](#).
- Greene WH, Hensher DA (2010) *Modeling Ordered Choices: A Primer*. Cambridge University Press. [CrossRef](#).
- Hajivassiliou VA, Ruud PA (1986) Classical estimation methods for ldv models using simulation. In: Engle R, McFadden D (eds), *Handbook of Econometrics*, Volume 4. Elsevier. [CrossRef](#).
- Hashiguchi Y, Tanaka K (2014) Agglomeration and firm-level productivity: A bayesian spatial approach. *Papers in Regional Science* 94[S1]: S95–S114. [CrossRef](#).
- Hensher D, Greene WH (2003) The mixed logit model: The state of practice. *Transportation* 30[2]: 133–176. [CrossRef](#).
- Jetz W, Rahbek C, Lichstein JW (2005) Local and global approaches to spatial data analysis in ecology. *Global Ecology and Biogeography* 14[1]: 97–98. [CrossRef](#).
- Jones K (1991) Specifying and estimating multi-level models for geographical research. *Transactions of the Institute of British Geographers* 16: 148–159. [CrossRef](#).
- Kelejian HH, Prucha IR (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics* 17[1]: 99–121
- Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40[2]: 509–533. [CrossRef](#).
- Lee L (2000) A numerically stable quadrature procedure for the one-factor random-component discrete choice model. *Journal of Econometrics* 95[1]: 117–129. [CrossRef](#).
- Lee LF (1992) On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. *Econometric Theory* 8[4]: 518–552. [CrossRef](#).
- Revelt D, Train K (1998) Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics* 80[4]: 647–657. [CrossRef](#).
- Revelt D, Train K (2000) Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier. Working Paper. Department of Economics, UCB
- Robinson WS (1950) Ecological correlations and the behavior of individuals. *American Sociological Review* 15[3]: 351–357. [CrossRef](#).
- Sarrias M (2016) Discrete choice models with random parameters in R: The Rchoice package. *Journal of Statistical Software* 74[10]: 1–31. [CrossRef](#).

- Sarrias M (2019) Do monetary subjective well-being evaluations vary across space? comparing continuous and discrete spatial heterogeneity. *Spatial Economic Analysis* 14[1]: 53–87. [CrossRef](#).
- Sarrias M, Daziano RA (2018) Individual-specific point and interval conditional estimates of latent class logit parameters. *Journal of Choice Modelling* 27: 50–61. [CrossRef](#).
- Stern S (1997) Simulation-based estimation. *Journal of Economic Literature* 35[4]: 2006–2039
- Swamy PAVB (1971) *Statistical Inference in Random Coefficient Regression Models*. Springer Berlin. [CrossRef](#).
- Train K (2009) *Discrete Choice Methods with Simulation*. Cambridge University Press. [CrossRef](#).
- Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* 7[2]: 161–187. [CrossRef](#).
- Withers SD (2001) Quantitative methods: Advancement in ecological inference. *Progress in Human Geography* 25[1]: 87–96. [CrossRef](#).

