Summer 8-1-2012

Bayesian Bidimensional Regression and Its Extension

Haeri Lee
Washington University in St. Louis

WASHINGTON UNIVERSITY

Department of Mathematics

Bayesian Bidimensional Regression and Its Extension

by

Haeri Lee

A thesis presented to the
Graduate school of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the
degree of Masters of Arts

August 2012

Saint Louis, Missouri

Abstract

Bidimensional regression analysis is used for comparing the similarity between two plane figures (Tobler,1994). The basic bidimensional regression can be written as a linear regression model after re-parameterization and has been traditionally estimated by the ordinary least squares. In this dissertation, we propose a Bayesian approach to bidimensional regression and further consider its extension to a cognitive study that studies the relationship between a real map and memorized maps from many subjects. A hierarchical model is further proposed to incorporate random effects that describe the difference among subjects. Also, we develop a Gibbs sampler for estimating this hierarchical model. The proposed method is then applied to a real cognitive study.

Acknowledgement

This dissertation would not have been possible without the continual support from many people. I would like to express my deep gratitude to the following.

First and above all, I praise God, the Creator and the Guardian of my life, for giving me the capability to finish this thesis successfully. "The LORD is my shepherd, I lack nothing." (Psalm 23:1)

I must give my utmost respectful gratitude to my advisor, Dr. Nan Lin, for his support and advice throughout the project. I have learned a lot through challenging yet valuable experiences.

I would also like to thank the Department of Psychology, Washington University whose data made this project possible.

I also acknowledge the staff and faculty of the Mathematics Department for the help and encouragement during my degree.

My acknowledgment also goes to Okhwa Lee and Steve Han. It would have never been possible without your endless support, constant prayer and unflagging love.

Last but not least, I am deeply and forever indebted to my family, Jae Kwan Lee, Song Hui Yi and Sean Lee, for their constant prayer and confidence in me. It was under their watchful eye that I gained so much ability to tackle challenges head on. Their love provided my inspiration and was my driving force.

Table of Contents

Index of Tables and Figures

1 Introduction

Lots of times, there are situations where the degree of resemblance between two plane figures, e.g.maps, needs to be measured, such as computing the degree of similarity of two people's faces based on their pictures. This matter can be approached by using regression analysis. Tobler(1994) specifically used bidimensional regression as a tool for this type of tasks.

Bidimensional regression models the transformation between the coordinates of a set of objects(landmarks) on the two maps. For example, Kendra, David and Ashok (2009) described three different bidimensional regression models which are Euclidean, affine and projective models, in an order of increasing complexity. For psychological data, Llyod(1989) and Nakaya(1997) stated that only the Euclidean and affine models have provided practically useful descriptions. In this dissertation, we focus on the basic bidimensional regression model for Euclidean transformation, which assumes that the original coordinates are scaled, rotated and translated by the same values so the overall configuration remains in the same shape in the other map. The principle developed in this dissertation can be generalized to affine and projective transformations.

Unlike the common linear regressions, bidimensional regression assesses the relation between independent and dependent variables which are each two dimensional. Suppose that we have two maps with n objects marked on them. Let (u_i, v_i) denote the ith point of the target plane, the dependent variable, and (x_i, y_i) the matching point on the explanatory plane (the independent variable). Nakaya(1997) defined the basic bidimensional regression model for 'Euclidean transformation' as

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \phi \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} + \begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} , i = 1, \dots, n ,$$

where ϕ is a scaling parameter and θ is an angle of rotation. Also, ε_i and η_i are random errors assumed to be independent. Then, by re-parameterization, we get

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \beta_1 & -\beta_2 \\ \beta_2 & \beta_1 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} + \begin{pmatrix} \varepsilon_i \\ \eta_i \end{pmatrix} , i = 1, \dots, n , \tag{1}$$

where, $$\phi = \sqrt{\beta_1^2 + \beta_2^2}$$

and $$\theta = \tan^{-1}\left(\frac{\beta_2}{\beta_1}\right). \tag{2}$$

According to Friedman and Kohler (2003), the two parameters (α_1, α_2) indicate the magnitude of the horizontal and vertical translation, respectively and the remaining parameters (β_1, β_2) are to derive the scale and angle values. Hence, (α_1, α_2) represents a translation relative to the original location: up or down; and (β_1, β_2) indicates by how much the points have been rotated and expanded or contracted.

In the literature, the classical way of fitting a bidimensional regression model is in the frequentist way using the ordinary least squares (OLS). The goal of this dissertation is to derive statistical inference for bidimensional regression in the Bayesian way. The Bayesian approach is more appropriate for the situation where, the data are supported with additional prior information. This prior information is combined with the likelihood function of the data to yield the posterior belief about the coefficients and variance. Bayesian inference treats the unknown parameters as random variables when the observed data are treated as fixed and known. In the same context, the unobservable parameters are treated probabilistically, while the observed data are treated deterministically (Martin, 2005). The goal is to obtain the distribution of the parameters given the information in the data. To carry out Bayesian inference, prior information for unknown parameters has to be added. After the priors are chosen, multiplying the likelihood function and the prior results in a posterior distribution for the parameters.

Although the Bayesian principle seems simple, analyzing the posterior distribution is complicated because deriving posterior distributions often requires the integration of high-dimensional functions (Walsh 2004). Consequently, Monte Carlo (MC) methods are usually used as a tool of summarizing the posterior distribution. MC methods state that information about the target distribution can be learned by repeatedly drawing from it. Then, there has to be an algorithm that suits to produce draws from the target distribution. There are two popular algorithms, Gibbs sampling and the Metropolis-Hastings algorithm, used widely in Bayesian inference. In this thesis, we will use the Gibbs sampler.

In a Gibbs sampler, the sequences of draws are dependent and each draw depends on the previous draw thus they form a Markov chain. Before running a Gibbs sampling algorithm, full conditional distributions of each parameter has to be derived. Gibbs sampling draws from full conditional distributions instead of the joint distribution as simulating from the joint distribution is typically much more complicated.

A second purpose of this dissertation is to investigate the usage of Bayesian bidimensional regression for spatial cognition. It is based on an ongoing research on cognitive mapping conducted by Department of Psychology, Washington University. The spatial transformation will be examined when objects' positions are reconstructed from memory. Participants were shown a video with a set of objects in a scene and after watching the video, they were given a map for the same scene shown in the video and asked to indicate the locations of the objects. By comparing the actual locations of the objects and those the participants put on the map, it helps to understand how people's memory works in spatial mapping. A straightforward application of bidimensional regression is to assess the relationship between the real map and each participant's memorized map. However, such an individual-by-individual analysis does not model the commonality among all participants, which is the key in understanding the mechanism of human spatial cognition. In addition, it is statistically more efficient to combine data from all subjects. Therefore, we extend the basic bidimensional regression by allowing random effects on the transformation coefficients of each participant and establish a hierarchical model. We then further construct a Gibbs sampler for this hierarchical model.

In the next chapter, we will develop Bayesian inference for the basic bidimensional regression. Then in Chapter 3, we describe the hierarchical bidimensional regression model and the associated Gibbs sampler. A real data analysis is then presented in Chapter 4. And we conclude the thesis in Chapter 5.

2 Bayesian bidimensional regression

Suppose we have a study involving two maps with n objects. Let (x_i, y_i) be the actual location and (u_i, v_i) be the memorized location of the ith object. We can rewrite model (1) as a multiple linear regression, that is,

$$\begin{pmatrix} u_1 \\ . \\ . \\ . \\ u_n \\ v_1 \\ . \\ . \\ . \\ v_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & x_1 & -y_1 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & 0 & x_n & -y_n \\ 0 & 1 & y_1 & x_1 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 1 & y_n & x_n \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_n \\ \eta_1 \\ . \\ . \\ . \\ \eta_n \end{pmatrix} \tag{3}$$

Here, ε_i and η_i are independent and identically distributed (i.i.d.) random errors. These two terms can either be assumed having a common variance in basic bidimensional regression or with different variances in weighted bidimensional regression (Schmid et al 2011). In this study, we assume ε_i and η_i have different variances where $\varepsilon_i \sim N(0, \sigma_u^2)$ and $\eta_i \sim N(0, \sigma_v^2)$ respectively. Then we may write the above model in matrix notations as $y|\beta, \sigma_u^2, \sigma_v^2 \sim N(X\beta, W)$, where the response $y = (u_1, \dots, u_n, v_1, \dots, v_n)^T$ and $\beta = (\alpha_1, \alpha_{2,}\beta_1, \beta_2)^T$. The coefficients β measure the scaling and rotation transformations. We further denote $X = (X_u^T, X_v^T)^T$, where $X_u$and X_v are the upper and lower half of the design matrix. Under this model, the covariance matrix W of the random errors, is the following diagonal matrix,

$$W = \begin{pmatrix} \sigma_u^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & .. & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_u^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_v^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & .. & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_v^2 \end{pmatrix} \tag{4}$$

2.1 Gibbs Sampler

Next, we derive the full conditional distributions needed for the Gibbs sampler. The joint posterior distribution is obtained by multiplying the likelihood function and the prior and full conditional distributions can then be derived from the joint posterior distribution. Here, we assume prior independence among all parameters, which leads to the fact that the full conditional distribution of any parameter can be obtained by multiplying the likelihood function and its prior. The log-likelihood function for the basic bidimensional regression model is given by

$$l(\beta, W) = -\frac{1}{2}\{(y - X\beta)^T W^{-1}(y - X\beta) + \log|W|\} + const$$

To derive the full conditional distribution of $\beta|\sigma_u^2, \sigma_v^2, y$ the prior information has to be specified. Here, we set a flat prior, $\pi(\beta) \propto 1$. Then, the full conditional distribution of $\beta|\sigma_u^2, \sigma_v^2, y$ is given as

$$\propto f(y|\beta, W) \cdot \pi(\beta) = |W|^{-1/2} exp\,(-\tfrac{1}{2}(y - X\beta)^T W^{-1}(y - X\beta))$$

$$\propto exp\,(-\tfrac{1}{2}(\beta^T X^T W^{-1} X\beta - \beta^T X^T W^{-1} y - y^T W^{-1} X\beta))$$

$$\propto exp\,(-\tfrac{1}{2}(\beta - (X^T W^{-1} X)^{-1} X^T W^{-1} y)^T (X^T W^{-1} X)(\beta - (X^T W^{-1} X)^{-1} X^T W^{-1} y))$$

Hence, we have

$$\beta|\sigma_u^2, \sigma_v^2, y \sim N\big(\hat{\beta}, V_\beta\big) \tag{5}$$

where $\hat{\beta} = (X^T W^{-1} X)^{-1} X^T W^{-1} y$ and $V_\beta = (X^T W^{-1} X)^{-1}$.

Next, we derive the full conditional distributions of σ_u^2 and σ_v^2. Different from β, where a flat prior was assumed, for two variances, conjugate prior is chosen for both. Conjugate prior makes updating the parameters more straightforward as both prior and posterior have the same distributional form and the posterior's parameters have a simple functional form. For normal data, inverse-gamma distribution is usually used as conjugate prior of unknown variance parameters. Hence, we assume

$\pi(\sigma_u^2) \sim IG(a_0, b_0)$ and $\pi(\sigma_v^2) \sim IG(a_1, b_1)$, whose density functions are proportional to ${\sigma_u^2}^{-(a_0+1)} \cdot exp\,(-b_0(\sigma_u^2)^{-1})$ and ${\sigma_v^2}^{-(a_1+1)} \cdot exp\,(-b_1(\sigma_v^2)^{-1})$. As the prior follows the inverse-gamma distribution, we expect the full conditional distribution also follows the inverse-gamma distribution with different parameters. Note that vector $u = (u_1, \ldots, u_n)^T \sim N(X_{ui}\beta, \sigma_u^2)$ and $v = (v_{n+1}, \ldots, v_{2n})^T \sim N(X_{vi}\beta, \sigma_v^2)$, where X_{ui} and X_{vi} denote the ith row of X_u and X_v, respectively.

Derivation of the full conditional distribution of $\sigma_u^2|\beta, \sigma_v^2, y$ and $\sigma_v^2|\beta, \sigma_u^2, y$ is given below.

$$f(\sigma_u^2|\beta, \sigma_v^2, y) \propto f(\beta, \sigma_u^2, \sigma_v^2|y)$$

$$\propto {\sigma_u^2}^{-(a_0+1)} exp\,(-b_0(\sigma_u^2)^{-1}) \prod_{i=1}^{n} (\sigma_u^2)^{-1/2} exp\left\{-\frac{(u_i - X_{ui}\beta)^2}{2\sigma_u^2}\right\}$$

$$\times \prod_{i=n+1}^{2n} (\sigma_v^2)^{-1/2} exp\left\{-\frac{(v_i - X_{vi}\beta)^2}{2\sigma_v^2}\right\}$$

$$\propto {\sigma_u^2}^{-(a_0+1)} (\sigma_u^2)^{-n/2} exp\,(-b_0(\sigma_u^2)^{-1}) exp\left\{-\frac{\sum_{i=1}^{n}(u_i - X_{ui}\beta)^2}{2\sigma_u^2}\right\}$$

$$\propto {\sigma_u^2}^{-\left(a_0+\frac{n}{2}\right)-1} exp\left\{\frac{-b_0 - \frac{1}{2}\sum_{i=1}^{n}(u_i - X_{ui}\beta)^2}{\sigma_u^2}\right\}$$

and

$$f(\sigma_v^2|\beta, \sigma_u^2, y) \propto f(\beta, \sigma_u^2, \sigma_v^2|y)$$

$$\propto {\sigma_v^2}^{-(a_1+1)} exp\,(-b_1(\sigma_v^2)^{-1}) \prod_{i=n+1}^{2n} (\sigma_v^2)^{-1/2} exp\left\{-\frac{(v_i - X_{vi}\beta)^2}{2\sigma_v^2}\right\}$$

$$\times \prod_{i=1}^{n} (\sigma_u^2)^{-1/2} exp\left\{-\frac{(u_i - X_{ui}\beta)^2}{2\sigma_u^2}\right\}$$

$$\propto {\sigma_v^2}^{-(a_1+1)} (\sigma_v^2)^{-n/2} exp\,(-b_1(\sigma_v^2)^{-1}) exp\left\{-\frac{\sum_{i=n+1}^{2n}(v_i - X_{vi}\beta)^2}{2\sigma_v^2}\right\}$$

$$\propto {\sigma_v^2}^{-\left(a_1+\frac{n}{2}\right)-1} exp\left\{\frac{-b_1 - \frac{1}{2}\sum_{i=n+1}^{2n}(v_i - X_{vi}\beta)^2}{\sigma_v^2}\right\}$$

Note when deriving the full conditional distribution of $\sigma_u^2|\beta, \sigma_v^2, y$, terms in the log-likelihood function involving σ_v^2 can be considered as constant thus eliminated and vice versa. The full conditional distribution of each variance parameter is given as

$$\sigma_u^2|\beta, \sigma_u^2, y \sim Inv - Gamma(a_{0+}\frac{n}{2}\ , b_0 + \frac{1}{2}\sum_{i=1}^{n}(u_i - X_{ui}\beta)^2) \qquad (6)$$

and

$$\sigma_v^2|\beta, \sigma_u^2, y \sim Inv - Gamma(a_1 + \frac{n}{2}\ , b_1 + \frac{1}{2}\sum_{i=n+1}^{2n}(v_i - X_{vi}\beta)^2). \qquad (7)$$

3 Bayesian hierarchical bidimensional regression

Now we consider extending the basic bidimensional regression to a situation motivated by a spatial cognition study conducted by Department of Psychology, Washington University. In this study, each of 225 participants was shown a video with nine objects in a scene. And after watching the video, they were given a map for the same scene shown in the video and asked to indicate the locations of the nine objects. By comparing the actual locations of the objects and those the participants put on the map, it helps to understand how people's memory works in spatial mapping. The basic bidimensional regression is not suitable for this situation as it is only able to analyze the relationship between two maps whereas this study produces many memorized maps out of one real map.

The basic bidimensional regression model in the previous chapter can be used to estimate the transformation of a single participant (subject). Suppose we have k subjects and each provides memorized locations of n objects after viewing the video. We can then write the bidimensional regression model for the jth subject in the multiple linear regression form as

$$\begin{pmatrix} u_{ij} \\ \cdot \\ \cdot \\ \cdot \\ u_{nj} \\ v_{ij} \\ \cdot \\ \cdot \\ \cdot \\ v_{nj} \end{pmatrix} = \begin{pmatrix} 1 & 0 & x_i & -y_i \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & x_n & -y_n \\ 0 & 1 & y_i & x_i \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & y_n & x_n \end{pmatrix} \begin{pmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \beta_{1j} \\ \beta_{2j} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ij} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{nj} \\ \eta_{ij} \\ \cdot \\ \cdot \\ \cdot \\ \eta_{nj} \end{pmatrix} \tag{8}$$

where i denotes the ith object and j denotes the jth subject, and the errors are still assumed as $\varepsilon_{ij} \sim N(0, \sigma_u^2)$ and $\eta_{ij} \sim N(0, \sigma_v^2)$ i.i.d.. We allow the transformation coefficient for each individual ,β_j, to be different across j and consider them as random effects that vary around a common population parameter β. Coefficients for each individual, β_j's, $j = 1, \dots k$ are i.i.d. as

$$\beta_j \sim N(\beta, \Sigma_\beta). \tag{9}$$

Next, we need to assign hyper-priors on β and Σ_β. They are given as

$$\beta \sim N(\mu_\beta, C) \text{ and } {\Sigma_\beta}^{-1} \sim Wishart((\rho R)^{-1}, \rho). \tag{10}$$

The hyper-prior implies that $E\left({\Sigma_\beta}^{-1}\right) = R^{-1}$ and $\left(\Sigma_\beta\right) \propto \rho^{-1}$.

3.1 Gibbs Sampler

Now we need to derive the full conditional distribution of each parameter. First, consider the full conditional distribution of $\beta_j | \beta_{-j}, y_j, \mu_\beta, W^{-1}$ where β_j is the transformation coefficient for each subject. Given the prior and hyper-prior information, it follows that

$$f(\beta_j | \beta_{-j}, y_j, \mu_\beta, W^{-1})$$

$$\propto exp\,(-0.5(y_j - X_j\beta_j)^T W^{-1}(y_j - X_j\beta_j))exp\,(-0.5(\beta_j - \beta)^T \Sigma_\beta^{-1}(\beta_j - \beta))$$

$$\propto exp\,(-0.5(\beta_j^T \Sigma_\beta^{-1}\beta_j - 2\beta_j^T \Sigma_\beta^{-1}\beta + \beta_j^T X_j^T W^{-1} X_j \beta_j - 2\beta_j^T X_j^T y_j + \beta^T \Sigma_\beta^{-1}\,\beta + y_j^T y_j)$$

$$\propto exp\,(-0.5(\beta_j^T (\Sigma_\beta^{-1} + X_j^T W^{-1} X_j)\beta_j - 2\beta_j^T (\Sigma_\beta^{-1}\beta + X_j^T y_j) + \beta^T \Sigma_\beta^{-1}\,\beta + y_j^T y_j)).$$

Therefore, we have

$$\beta_j | \beta_{-j}, y_j, \mu_\beta, W^{-1} \sim N(\widehat{\beta_J}, \widehat{\Sigma_\beta}), \tag{11}$$

where $\widehat{\beta_J} = \widehat{\Sigma_\beta}(\Sigma_\beta^{-1}\beta + X_j^T y_j)$ and $\widehat{\Sigma_\beta} = (\Sigma_\beta^{-1} + X_j^T W^{-1} X_j)^{-1}$.

Note that X_j is the design matrix of jth subject and $y_j = (u_1, \dots, u_n, v_1, \dots, v_n)_j^T$ for each subject.

Next, we derive the full conditional distribution of the population mean parameter, $\beta | \{\beta_j\}, y_j, W^{-1}$.

$$f(\beta | \{\beta_j\}, y_j, W^{-1})$$

$$\propto \prod_{j=1}^{k} exp\,(-0.5\,(\beta_j-\beta)^T\Sigma_\beta^{-1}(\beta_j-\beta))\; exp\,(-0.5(\beta-\mu_\beta)^T C^{-1}(\beta-\mu_\beta))$$

$$\propto exp\,(-0.5(\sum_{j=1}^{k}(\beta_j-\beta)^T\Sigma_\beta^{-1}(\beta_j-\beta)+(\beta-\mu_\beta)^T C^{-1}(\beta-\mu_\beta)))$$

$$\propto exp\,(-0.5(\sum_{j=1}^{k}({\beta_j}^T\Sigma_\beta^{-1}\beta_j-2{\beta_j}^T\Sigma_\beta^{-1}\beta+\beta^T\Sigma_\beta^{-1}\beta)+\beta^T C^{-1}\beta-2\beta^T C^{-1}\mu_\beta+ {\mu_\beta}^T C^{-1}\mu_\beta))$$

$$\propto exp\,(-0.5(\beta^T(k\Sigma_\beta^{-1}+C^{-1})\beta-2\beta^T\left(C^{-1}\mu_\beta+k\Sigma_\beta^{-1}\sum_{j=1}^{k}\beta_j\right)+\sum_{j=1}^{k}\left({\beta_j}^T\Sigma_\beta^{-1}\beta_j\right) + {\mu_\beta}^T C^{-1}\mu_\beta)).$$

Then we have

$$\beta|\{\beta_j\},y_j,W^{-1}\sim N(\widehat{\mu_\beta},\hat{C}), \quad (12)$$

where $\widehat{\mu_\beta}=\hat{C}(C^{-1}\mu_\beta+k\Sigma_\beta^{-1}\sum_{j=1}^{k}\beta_j)$ and $\hat{C}=(k\Sigma_\beta^{-1}+C^{-1})^{-1}$.

Then, the full conditional distribution of each variance parameter, σ_u^2 and σ_v^2, is derived. We are still using conjugate priors for each variance parameter, $\pi(\sigma_u^2)\sim IG(a_0,b_0)$ and $\pi(\sigma_v^2)\sim IG(a_1,b_1)$. Hence, we expect the full conditional distribution of each variance parameter is also an inverse-gamma distribution. Full conditional distributions of $\sigma_u^2|\sigma_v^2,y_j,\{\beta_j\},\mu_\beta\, W^{-1}$ and $\sigma_v^2|\sigma_u^2,y_{j,}\{\beta_j\},\mu_\beta,W^{-1}$ are given as

$$f(\sigma_{u\,j}^2|\sigma_{v\,j}^2,y_j,\{\beta_j\},\mu_\beta)\propto f(\sigma_{u\,j}^2,\sigma_{v\,j}^2,\{\beta_j\},\mu_\beta|y_j)$$

$$\propto {\sigma_{u\,j}^2}^{-(a_0+1)}\exp\,(-b_0(\sigma_{u\,j}^2)^{-1})\prod_{j=1}^{k}\left(\sigma_{u\,j}^2\right)^{-n/2} exp\left\{-\frac{\sum_{i=1}^{n}\left(u_{ij}-X_{uij}\beta_j\right)^2}{2\sigma_{u_j}^2}\right\}$$

$$\times\prod_{j=1}^{k}(\sigma_{v\,j}^2)^{-n/2}\,exp\left\{-\frac{\sum_{i=n+1}^{2n}(v_{ij}-X_{vij}\beta_j)^2}{2\sigma_{v\,j}^2}\right\}$$

$$\propto {\sigma_{u\,j}^2}^{-(a_0+1)}(\sigma_{u\,j}^2)^{-kn/2}exp\,\left\{-b_0(\sigma_{u\,j}^2)^{-1}-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n}\left(\frac{(u_{ij}-X_{uij}\beta_j)^2}{\sigma_{u_j}^2}\right)\right\}$$

$$\propto {\sigma_{u\,j}^2}^{-\left(a_0+\frac{kn}{2}\right)-1}\exp\,\left\{\frac{-b_0-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n}(u_{ij}-X_{uij}\beta_j)^2}{\sigma_{u_j}^2}\right\}$$

and

$$f(\sigma_{v\,j}^2|\sigma_{u\,j}^2, y_j, \{\beta_j\}, \mu_\beta\,) \propto\; f(\sigma_{v\,j}^2, \sigma_{u\,j}^2, \{\beta_j\}, \mu_\beta, |y_j)$$

$$\propto \sigma_{v\,j}^2{}^{-(a_1+1)} exp\,\left(-b_1\left(\sigma_{v\,j}^2\right)^{-1}\right)\prod_{j=1}^{k}\left(\sigma_{v\,j}^2\right)^{-n/2} exp\left\{-\frac{\sum_{i=n+1}^{2n}\left(v_{ij}-X_{vij}\beta_j\right)^2}{2\sigma_{v\,j}^2}\right\}$$

$$\times \prod_{j=1}^{k}(\sigma_{u\,j}^2)^{-n/2} exp\,\left\{-\frac{\sum_{i=1}^{n}\left(u_{ij}-X_{uij}\beta_j\right)^2}{2\sigma_{u\,j}^2}\right\}$$

$$\propto \sigma_{v\,j}^2{}^{-(a_1+1)}(\sigma_{v\,j}^2)^{-kn/2}\exp\left\{\,-b_1(\sigma_{v\,j}^2)^{-1}-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=n+1}^{2n}\left(\frac{(v_{ij}-X_{vij}\beta_j)^2}{\sigma_{v\,j}^2}\right)\right\}$$

$$\propto \sigma_{v\,j}^2{}^{-\left(a_1+\frac{kn}{2}\right)-1} exp\left\{\frac{-b_1-\frac{1}{2}\sum_{j=1}^{k}\sum_{i=n+1}^{2n}(v_{ij}-X_{vij}\beta_j)^2}{\sigma_{v\,j}^2}\right\},$$

where $X_{ui\,j}$ and $X_{vi\,j}$ denote the ith row of X_u and X_v for the jth subject, respectively. Therefore,

$$\sigma_{u\,j}^2|\sigma_{v\,j}^2, y_j, \{\beta_j\}, \mu_\beta\; \sim$$

$$Inv-Gamma(a_0+\frac{kn}{2}, b_0+\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n}\left(u_{ij}-X_{uij}\beta_j\right)^2) \tag{13}$$

and

$$\sigma_{v\,j}^2|\sigma_{u\,j}^2, y_j, \{\beta_j\}, \mu_\beta\; \sim$$

$$Inv-Gamma\,(a_1+\frac{kn}{2}, b_1+\frac{1}{2}\sum_{j=1}^{k}\sum_{i=n+1}^{2n}\left(v_{ij}-X_{vij}\beta_j\right)^2)\,. \tag{14}$$

4 Real data analysis

The data set is provided by Department of Psychology, Washington University. There are 231 subjects enrolled by December 12, 2011. Each subject provided memorized x and y coordinates of nine objects, whose actual coordinates are also recorded from the real map. Six individuals who have missing values in any of their coordinates have been eliminated from the sample, so the actual number of subjects is 225.

4.1 Basic bidimensional regression

After full conditional distributions of all parameters have been derived, we can run the Gibbs sampler. The algorithm was run for each individual separately with 3,000 iterations for the iteration to converge. The first 20 were discarded as burn-in. The initial values for β and hyper-parameters of each variance need to be pre-specified. Initial value for β is set as $\hat{\beta}$ = (-18.244, 53.087, 0.902,-0.084), which is the regression coefficient estimate from non-Bayesian bidimensional regression, i.e.OLS. The hyper-parameters in the inverse-gamma distribution, a_0, a_1, b_0 and b_1, are chosen to make the prior non-informative. The typical non-informative prior for a variance parameter is $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ which corresponds to the case both hyper-parameters are close to zero. Hence, hyper-parameter values are defined as $a_0 = a_1 = b_0 = b_1$ equal to 0.001.

The Gibbs sampler then iteratively samples from the full conditional of β, σ_u^2 and σ_v^2. We implement the Gibbs sampler in R and use the R package `coda` to summarize the MCMC output and monitor the convergence.

Table 1 summarizes bidimensional regression analysis of one individual, Subject 80002, including posterior mean, standard error and quantiles.

Table 1: Summary of individual-by-individual bidimensional regression analysis for Subject 80002

	Mean	SD	Naive SE	Time-series SE
alpha1	-1.183e+01	4.0395188	3.343e-01	2.711e-01
alpha2	3.503e+01	1.0460379	8.657e-02	8.410e-02
beta1	9.449e-01	0.0103824	8.593e-04	6.951e-04
beta2	-3.390e-02	0.0134462	1.113e-03	9.192e-04
sigmau	8.851e-04	0.0004979	4.121e-05	3.710e-05
sigmav	1.146e-03	0.0005436	4.499e-05	3.939e-05

	2.5%	25%	50%	75%	97.5%
alpha1	-2.039e+01	-1.482e+01	-1.121e+01	-8.693141	-5.459371
alpha2	3.400e+01	3.420e+01	3.476e+01	35.494349	37.510156
beta1	9.228e-01	9.377e-01	9.466e-01	0.954051	0.958092
beta2	-6.659e-02	-4.249e-02	-2.998e-02	-0.023335	-0.017007
sigmau	2.990e-04	5.649e-04	7.586e-04	0.001030	0.002253
sigmav	4.219e-04	7.381e-04	1.066e-03	0.001366	0.002445

Using Equation (2), these regression coefficients can be converted into scaling and rotation transformation parameters. In this case, ϕ=$\sqrt{0.95^2 + 0.034^2}$ and $\theta = \tan^{-1}\left(\frac{0.034}{0.95}\right)$. Hence, one can state that the map for this individual is scaled by ϕ=0.951 and rotated by 2.05 degree of angle.

Also, to help understanding the result, we also present more detailed output of a randomly selected subject (id=80002). Two plots were given which are traceplot (Figure 1) and autocorrelation function (ACF) plot (Figure 2). The traceplot is useful in assessing convergence because the trace helps to see if the chain has not yet converged to its stationary distribution or whether the chain has mixed well. For a well-mixed Markov chain, the chain traverses the posterior space rapidly in the traceplot. ACF plot gives correlations between the series and lagged values of the series for lags. In the ACF plot, a low autocorrelation is expected for chains with good mixing.

Figure 1: Traceplot for each parameter of Subject 80002.

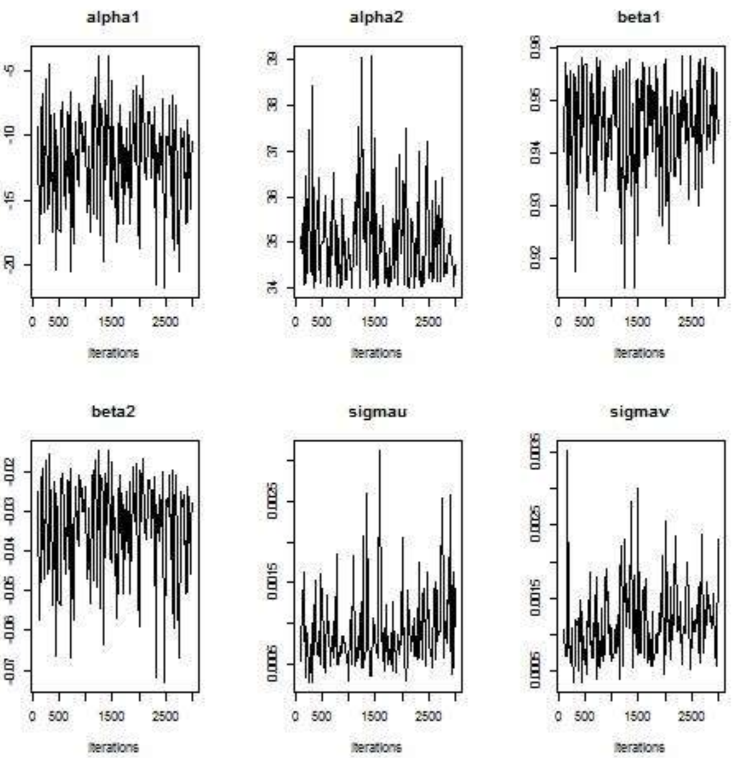
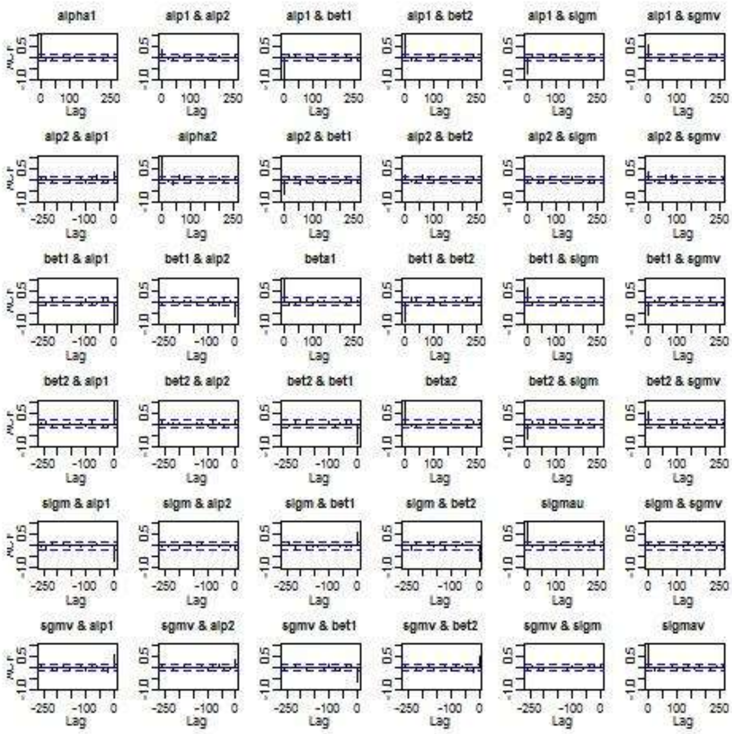


Figure 2: Autocorrelation function plot for Subject 80002.



Lastly, to check the convergence numerically, Geweke's convergence diagnostic is performed. Geweke's diagnostic is based on the tests for equality of the means of the first and last part of a Markov chain. If the samples are drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic has an asymptotically standard normal distribution (Bernado et al,1992). Table 2 is the output from Geweke's convergence diagnostic test for the same individual.

Table 2: Geweke's convergence diagnostic test for Subject 80002

```
Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5
alpha1    alpha2    beta1     beta2     sigmau    sigmav
0.02251   1.33663   -0.33745  -0.15876  -1.13749  -0.73409
```

Geweke's convergence diagnostic test provides a z-score for each parameter. As a Markov chain progresses to infinity, the sampling distribution of Geweke's z-score goes to N(0,1) if the chain has converged (Best and Cowles 1996). Hence, values of z –score which fall in the extreme tails of a standard normal distribution suggest that the chain is not fully converged. In Table 2, z-scores for most parameters are insignificant except for alpha2 and σ_u^2.

In this individual-by-individual analysis, we ran the Gibbs sampler for each individual with 3,000 iterations. And first 100 were discarded as burn-in. Figures 3 and 4 are the histograms and boxplots for each regression coefficient, $(\alpha_1, \alpha_2, \beta_1, \beta_2)$, across all individuals

Figure 3: Histograms of regression coefficient, $(\alpha_1, \alpha_2, \beta_1, \beta_2)$

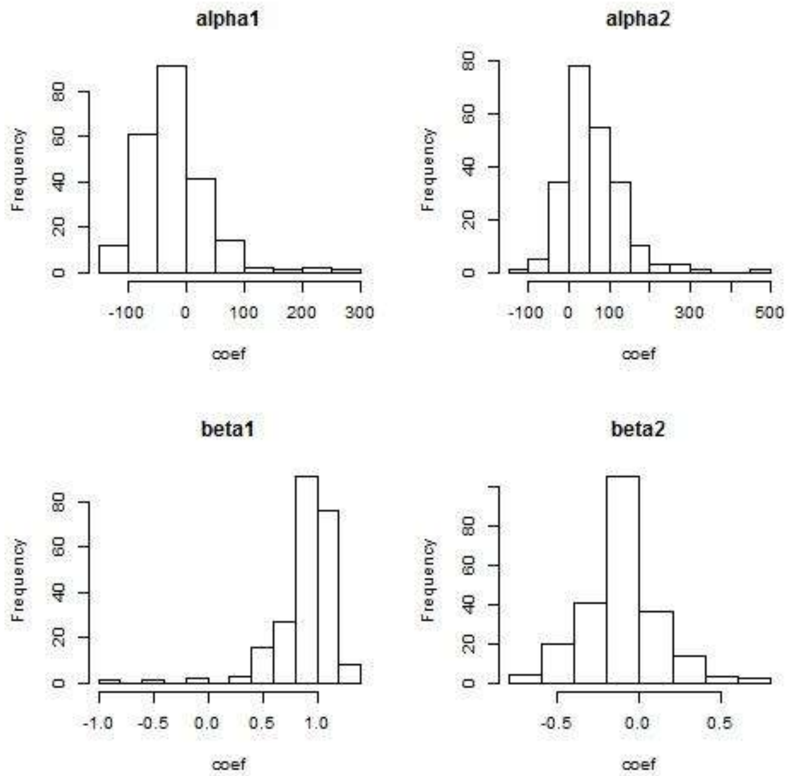
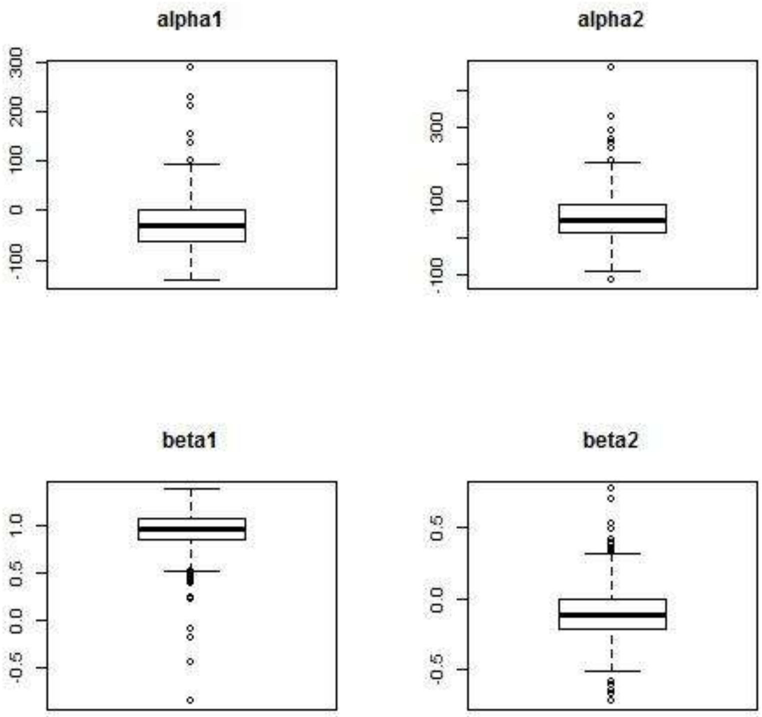


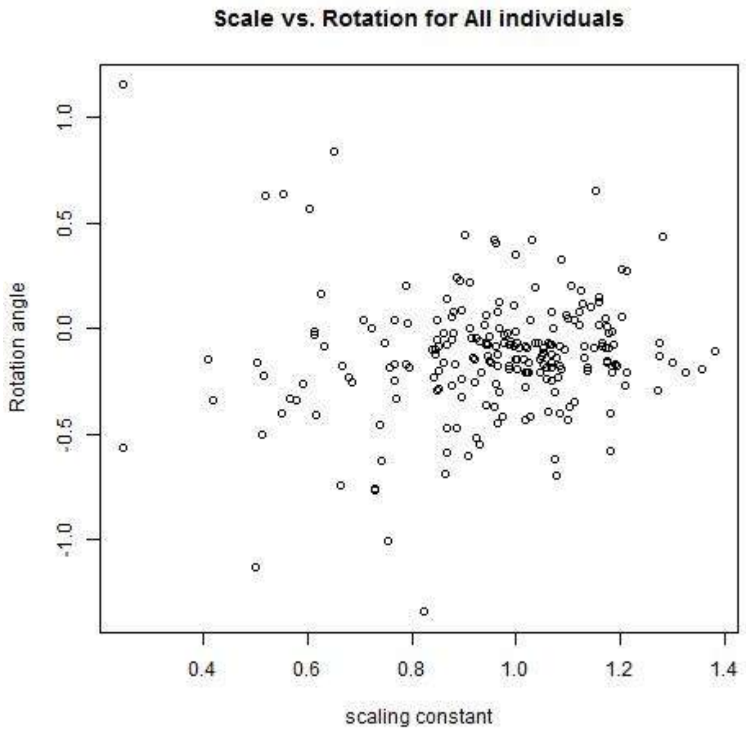
Figure 4: Boxplots of regression coefficient, $(\alpha_1, \alpha_2, \beta_1, \beta_2)$



From Figures 3 and 4, alpha1 and alpha2 have little heavier tail than normal distribution and beta2 has approximately normal distribution. Meanwhile, we also notice that the boxplot for all four coefficients show a few outliers which suggests that some subjects' spatial transformation works differently from the majority. And they may deserve further scientific investigation.

Then, the posterior median of the regression coefficients are transformed to more interpretable rotation and scaling parameters using Equation (2). Rotation angle and scaling constant are collected from all individuals and they were plotted in a two-dimensional scatterplot (Figure 5).

Figure 5: Scatter plot of Rotation angle vs. Scaling constant



Most number of points is located between 0.8 and 1.2 for scaling constant and around 0 for rotation angle values. This implies majority of people located objects almost correctly rotation-wise but they enlarged the objects by the rate of 0.8 to 1.2.

Marginal distributions of the rotation angle(θ) and scaling constants (ϕ) are shown in Figures 6 and 7.

Figure 6: Histogram and Boxplot of Rotation angle parameter of all individuals

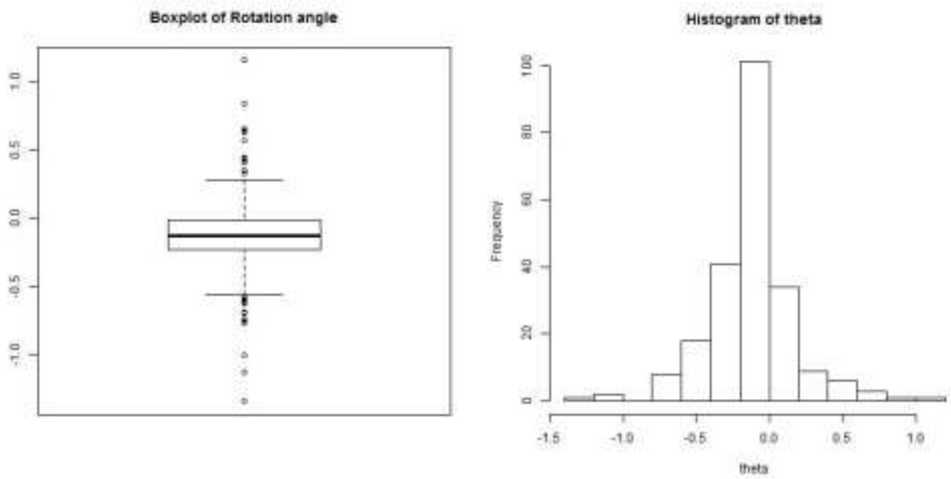
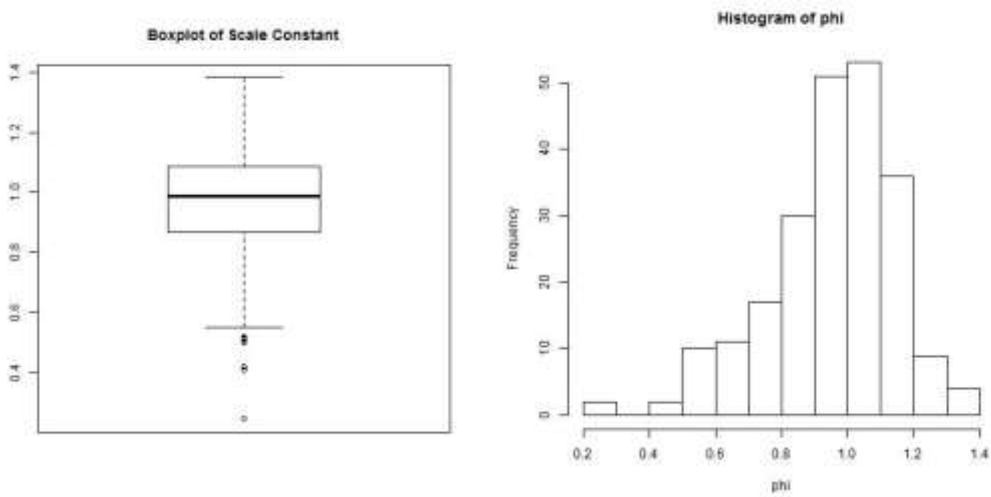


Figure 7: Histogram and Boxplot of Scaling constant parameter of all individuals



From Figure 6 and 7, we see that the degree of rotation of most individuals is not that large and the frequency around 0 values is the highest; on the other hand, the heavy tail of the distribution of the scaling constant indicates a much larger variation in how people rescale the locations of the nine objects.

Also, the histograms show that the rotation and scaling parameters are approximately normal. This suggests that it might be more reasonable to assume normal random effects on these parameters. However, this makes the computation much more complicated since there is no easy conjugate form is available hence it is worth future efforts to explore which way specifying the random effects is more reasonable.

5 Conclusion

In this thesis, we developed Bayesian inference for the basic bidimensional regression with Euclidean transformation. Furthermore, we also constructed a hierarchical bidimensional regression model motivated by a spatial cognition study and presented a Gibbs sampling procedure for estimating the model. Unlike classic bidimensional regression, we approach this in the Bayesian way that allows incorporating prior information. Since the joint posterior distribution is too complicated to directly sample from, we use Gibbs sampling to instead draw from the full conditional distributions of each parameter. We first developed a Gibbs sampler for the basic bidimensional regression, which can be used to analyze the spatial mechanism from one individual's data. In order to better utilize the information from many individuals in the spatial cognition study, we construct a hierarchical bidimensional regression model that allows borrow information cross different subject. We implemented the proposed methodology in the statistical software, R. We then apply the basic Bayesian bidimensional regression to a real data set. Results indicate that the normal assumption needs to be placed on the random effects. And in the future, we will explore the most reasonable way to specify the random effects and apply the hierarchical model the same data set and see whether it provides an improvement to the model fit.

Bibliography

Bernardo, J. M., Berger, J.O., Dawid, A. P. and Smith, A. F. M., Eds. (1992). *Bayesian Statistics 4.* Oxford University Press.

Best, N. G., Cowles, M. K. and Vines, S. K. (1995). CODA: Convergence diagnostics and output analysis software for Gibbs sampling. Technical report, Cambridge MRC Biostatistics Unit.

Chib, S., and Greenberg, E. (1995). “Understanding the Metropolis-Hastings Algorithm.” *The American Statistician* 49:327-336.

Cowles, M. K., and Carlin, B. P. (1995). “Markov Chain Monte Carlo Diagnostics: A Comparative Review.” *Journal of the American Statistical Association* 91:883–904.

Friedman, A. and Kohler, B. (2003). Bidimensional regression: Assessing the configural similarity and accuracy of cognitive maps and other two-dimensional data sets, *Psychological Methods*, Vol. 8, 468-491.

Gammerman, D. (1997). *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Chapman and Hall, London.

Leisch, Friedrich. (2004b). “FlexMix: A general framework for finite mixture models and latent class regression in R.” *Journal of Statistical Software*, **11**(8), 1-18. http://www.jstatsoft.org/v11/i08/

Martin, A. D. (2008). "Bayesian Analysis." In *Oxford Handbook of Political Methodology* (Box-Steffensmeier, J. M., Brady, H. E., and Collier, D., Eds.) Oxford University Press.

Nakaya, T. (1997). Statistical inferences in bidimensional regression models, *Geographical Analysis*, Vol. 29, 169-186.

Schmid, K. K., Marx, D. B. and Samal, A. (2011) Weighted bidimensional regression, *Geographical Analysis*, Vol. 43, 1-13.

Tobler, W. R. (1994), Bidimensional regression, *Geographical Analysis*, Vol. 26, No. 3, 187-212.