Technical Report 117

The Formulation and Estimation of a Spatial Skew-Normal Generalized Ordered-Response Model

Chandra Bhat
Center for Transportation Research

June 2016

Data-Supported Transportation Operations & Planning Center (D-STOP)

A Tier 1 USDOT University Transportation Center at The University of Texas at Austin

D-STOP is a collaborative initiative by researchers at the Center for Transportation Research and the Wireless Networking and Communications Group at The University of Texas at Austin.

1. Report No. D-STOP/2016/117	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle The Formulation and Estimation of a Spatial Skew-Normal Generalized Ordered-Response Model		5. Report Date June 2016
		6. Performing Organization Code
7. Author(s) Chandra R. Bhat and Sebastian Astroza of CTR Amin S. Hamdi of King Abdulaziz University		8. Performing Organization Report No. Report 117
9. Performing Organization Name and Address Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 1616 Guadalupe Street, Suite 4.202 Austin, Texas 78701		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DTRT13-G-UTC58
12. Sponsoring Agency Name and Address Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 1616 Guadalupe Street, Suite 4.202 Austin, Texas 78701		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes Supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program.		

16. Abstract

This paper proposes a new spatial generalized ordered response model with skew-normal kernel error terms and an associated estimation method. It contributes to the spatial analysis field by allowing a flexible and parametric skew-normal distribution for the kernel error term in traditional specifications of the spatial model. The resulting model is estimated using Bhat's (2011) maximum approximate composite marginal likelihood (MACML) inference approach. The model is applied to an analysis of bicycling frequency, using data from the 2014 Puget Sound household travel survey undertaken in the Puget Sound region in the State of Washington in the United States. Our results underscore the important effects of demographic variables, as well as the miles of bicycle lanes in an individual's immediate residential neighborhood, on bicycling propensity. An interesting finding is that women and young individuals (18-34 years of age) in particular "warm up" to bicycling as more investment is made in bicycling infrastructure, thus leading not only to a larger pool of bicyclists due to bicycling infrastructure enhancements, but also a more diverse and inclusive one. The results highlight the importance of introducing social dependence effects and non-normal kernel error terms from a policy standpoint. Specifically, our results suggest that ignoring these effects, as has been done by all earlier bicycling studies, can underestimate the impacts of bicycling infrastructure improvements and public campaigns on bicycle use frequency, potentially leading to under-investments in bicycling infrastructure projects.

17. Key Words Generalized ordered response model, skew normal distribution, social interactions, composite marginal likelihood, spatial econometrics, bicycling frequency	18. Distribution Statement No restrictions. This document is available to the public through NTIS (http://www.ntis.gov): National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161		
19. Security Classif.(of this report) Unclassified	20. Security Classif.(of this page) Unclassified	21. No. of Pages 50	22. Price

Form DOT F 1700.7 (8-72)   **Reproduction of completed page authorized**

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. *Mention* of trade names or commercial products does not constitute endorsement or recommendation for use.

Acknowledgements

.

Table of Contents

Chapter 1. Introduction

Ordered-response (OR) choice models are now widely used in many different disciplines, including sociology, biology, political science, marketing, and transportation. OR models may be used when analyzing ordinal discrete outcome data that may be considered as manifestations of an underlying scale that is endowed with a natural ordering. Examples include ratings data (for instance, of consumer products and movies), or likert-scale type attitudinal/opinion data (for example, of traffic congestion levels and teacher evaluations), or intensity data (such as of land use development levels and pain levels). In all of these situations, the observed outcome data may be considered as censored (or coarse) measurements of an underlying latent continuous random variable. The censoring mechanism is usually characterized as a partitioning or thresholding of the latent continuous variable into mutually exclusive (non-overlapping) intervals. The reader is referred to McKelvey and Zavoina (1971) and Winship and Mare (1984) for some early expositions of the ordered-response model formulation, and Liu and Agresti (2005) and Greene and Hensher (2010) for a survey of more recent developments.

The standard ordered-response model of McKelvey and Zavoina (1971) has been generalized in many different directions. One important direction is the extension to allow the thresholds (that map the latent underlying continuous variable to the observed ordinal outcomes) to vary across individuals due to observed individual characteristics, while also ensuring (through functional form specifications) that the resulting thresholds satisfy, for each individual in the sample, the ordering needed to ensure positive probabilities of each ordinal outcome (see Eluru *et al.*, 2008 and Greene and Hensher, 2010). As indicated by Greene and Hensher (2010) in Chapter 7 of their book, the resulting generalized ordered-response (GOR) model has been recently applied to many different application contexts. Castro *et al.* (2013) have also shown how a specific functional form parameterization of the thresholds leads to a generalized count model.

In this paper, we use the GOR structure as the starting point, and extend the formulation in two different directions. The first direction relates to the distribution of the kernel error distribution, and the second relates to spatial dependence. Each of these is discussed in turn in the next two sections.

1.1 The Kernel Error Term Structure

The estimation of ordered-response models is based on potentially noisy observations of ordinal outcomes, and thus there is little *a priori* information to specify the probability distribution form for the data generation process conditional on the observed explanatory variables. But it is typical in the literature to impose an *a priori* and convenient, but potentially very restrictive, kernel error distributional assumption for the underlying data generation process. Two of the most dominant error distribution assumptions are the logistic and normal distributions, leading to the familiar logit-based GOR and probit-based GOR models, respectively. But the actual functional form of the latent variable (conditioned on observed covariate) that underlies the observed discrete choice is seldom known in practice. It also, however, is widely recognized that mis-specification of the

kernel error distribution will, in general, lead to inconsistent estimates of the choice probabilities as well as the effects of exogenous variables (Geweke and Keane, 1999, Caffo *et al.*, 2007). This has led to the use of non-parametric as well as semi-parametric (or flexibly parametric) methods to characterize the error distribution (many studies using such methods are focused on binary choice models, though the same methods are applicable to ordered-response models). The non-parametric methods (see Berry and Haile, 2010 and Greene and Hensher, 2010, Chapter 12 for reviews) allow consistent estimates of the observed variable effects under broad model contexts by making regularity (for instance, differentiability) assumptions on an otherwise distribution-free density form. But the flexibility of these methods comes at a high inferential cost since consistency is achieved only in very large samples, parameter estimates have high variance, and the computational complexity/effort can be substantial (Mittlehammer and Judge, 2011). On the other hand, the semi-parametric methods, while not guaranteeing consistency in as broad a sense as the non-parametric methods, are somewhat easier to implement. They also allow asymmetric and flexible kernel error distribution forms. While the class of semi-parametric (or flexibly parametric) methods subsumes many different approaches, the ones that are used quite widely fall under the finite discrete mixture of normals (FDMN) approach (see Geweke and Keane, 1999, Caffo *et al.*, 2007, Fruhwirth-Schnatter, 2011a,b, Ferdous *et al.*, 2011, and Malsiner-Walli *et al.*, 2016) or the best fit parametric distribution selection approach through generalized link functions (see, for example, Stewart, 2005, Czado and Raftery, 2006 and Canary *et al.*, 2016).

1.2 Spatial Dependence

There is increasing interest and attention in discrete choice modeling on recognizing and explicitly accommodating spatial dependence among decision-makers, based on spatial lag and spatial error-type specifications (and their variants) that have been developed for continuous dependent variables. Further, the importance of spatial modeling, while originating initially in urban and regional modeling, is now permeating into economics and mainstream social sciences, including agricultural and natural resource economics, public economics, geography, sociology, political science, epidemiology, and transportation. Some examples in these fields include assessing the harvest level of agricultural products (Ward *et al.*, 2010), determining the siting location for an industry (Alamá-Sabater *et al.*, 2011, Bocci and Rocco, 2016), analyzing voter turnout in an election (Facchini and François, 2010), and investigating crashes and accident injury severity (Rhee *et al.*, 2016, Castro *et al.*, 2013). The reader is referred to a special issue of the *Journal of Regional Science* edited by Partridge *et al.* (2012) for collections of recent papers on spatial dependence. Other sources for good overviews include LeSage and Pace (2009), Anselin (2010), Arbia (2014), Franzese *et al.* (2016) and Elhorst *et al.* (2016).

Of course, the same mis-specification-in-distribution form considerations that lead to inconsistent maximum likelihood estimation in aspatial ordered-response models also lead to inconsistent estimation in spatial ordered-response models when an incorrect

distributional form is assumed for the kernel error term and the model coefficients.[1] If at all, mis-specifications lead to even more severe problems in spatial models because the spillover effects result in dependence and heteroscedasticity across the unobserved components of decision agents or units. In this context, and unlike the case of aspatial ordered-response models, there has been no prior research that we are aware of that explicitly accommodates non-normal error terms. This has been explicitly discussed as an issue of serious concern in spatial analysis papers within the past decade. For example, McMillen (2010, 2012) suggests that distributional form mis-specification of errors can themselves lead to spurious spatial correlation in residuals, and Pinkse and Slade (2010) identifies the normality assumption as being "implausible". Further, extant semi-parametric and flexible parametric methods developed for the aspatial case (and discussed above) are all but infeasible for the spatial case. For instance, the most commonly used and easily implemented (in the aspatial case) finite scale mixture of normals method will lead to L^Q different mixtures for each error term in the reduced form spatial ordered-response model, where L is the number of mixtures assumed for each individual error term in the structural spatial model and Q is the number of decision agents in the spatial setting (this is because the reduced form error term for each decision agent is an affine transformation of the original structural error terms). Similarly, the generalized link functions approach for aspatial binary or ordered-response cases is not suitable for spatial (and, therefore, multivariate) binary or ordered-response cases where the spatial dependence between dependent variables is generated in a specific form in terms of the latent underlying variables. Even without this spatial dependence form issue, the use of multivariate link functions with flexible marginal error distributions to handle multivariate binary or ordered-response variables is difficult and cumbersome to work with.[2]

1.3 The Current Paper

In the current paper, the key innovation is that we propose the use of a flexible parametric approach to incorporate asymmetry and skewness in the structural (kernel) error terms within a spatial GOR model (for ease, we will also refer to this model henceforth as a spatial skew-normal GOR or SSN-GOR model). That is, we accommodate both a flexible parametric non-normal kernel error term as well as potential spatial dependence effects in

---

[1]A terminology issue here. We use the term "aspatial model" strictly to refer to the case where there is no direct effect whatsoever on the dependent variable at one point in space through observed covariates (that is, explanatory variables) or unobserved factors at another point in space. A "spatial model", on the other hand, is one in which the dependent variable at one point is space is explicitly influenced by observed covariates and/or unobserved factors at another point is space.

[2]Some researchers have used a Generalized method of moments (GMM)-based estimator for the estimation of spatial qualitative dependent variable models, which is relatively robust to mis-specifications of the error distribution form (see, for example, McMillen, 1995, Pinske *et al.*, 2006, Klier and McMillen, 2008, and Flores-Lagunes and Schnier, 2012). However, these methods do not explicitly accommodate flexible error structures in the specification per se. Further, these estimators are typically based on a two-step instrumental variable technique after linearizing around zero interdependence, and work well only for cases with large sample sizes and low spatial interdependence (see Franzese *et al.*, 2016). There is also no guarantee in these methods that the spatial autoregressive term will be confined to its legitimate parameter space (Elhorst, 2010).

a GOR model. We achieve this through the use of a skew-normal distribution for the kernel error terms.[3] The skew-normal is a flexible density function that allows a "seamless" and "continuous" variation from normality to non-normality, and can replicate a variety of smooth density shapes with tails to the left or right as well as with a high modal value (sharp peaking) or low modal value (flat plateau). It is also tractable for practical applications and parsimonious in general in the number of parameters that regulate skewness. Further, the multivariate normal distribution is obtained as a specific restricted case of the multivariate skew-normal distribution (see Azzalini and Dalla Valle, 1996, Azzalini and Capitanio (1999), Arrellano-Valle and Azzalini, 2006, Lee and McLachlan, 2013, 2014). Bhat and Sidharthan (2012) used the skew-normal distribution for the aspatial case in a discrete choice context, but did not consider spatial dependence. Indeed, we are not aware of any linear regression or discrete choice model in the literature that considers a skew-normal distribution for the error terms within a spatial econometric context, though there have been a few applications of the skew-normal distribution in the context of aspatial linear regression type models with continuous observations (see, for example, Meintanis and Hlávka, 2010, Molenaar *et al.*, 2010, Smith *et al.*, 2012, and Lin *et al.*, 2016). In this paper, we show how the skew-normal is particularly well suited for spatial analysis because it leads to just one additional parameter to be estimated relative to traditional spatial models. This is because the structural (*i.e.*, kernel) error terms are marginally skew-normal and distributed with the same amount of skew across observations. We exploit this characteristic and impose a specific restrictive form on the multivariate skew-normal distribution that has not appeared and been used in the literature.

The second key contribution is that we show how spatial GOR models with error terms of the skew-normal variety can be estimated with relative ease using Bhat's (2011) maximum approximate marginal composite likelihood (MACML) estimation approach. Traditional frequentist and Bayesian methods, on the other hand, are still relatively cumbersome and involve rather long estimation times. While important strides have been made in reducing computational times in the context of error terms with normally distributed errors by recognizing the sparse covariance matrix structure of error terms (see Pace and LeSage, 2011 and Liesenfeld *et al.*, 2013; Elhorst *et al.*, 2016 provides a good review), the effectiveness of these methods for skew-normally distributed error terms is still in question.[4] Finally, we demonstrate an application of the proposed model.

Overall, we contribute to the spatial analysis field by allowing a general and robust formulation for the error terms using a skew-normal distribution. In the current paper, we adopt a spatial lag specification because we believe it is grounded in a structural basis that is certainly plausible in many empirical settings of spatial interaction, diffusion, and

---

[3]The normal distribution is symmetric and bell-shaped, while the skew-normal distribution introduces asymmetry (through a skewness parameter) to the normal distribution. From this standpoint, the skew-normal is one type of a non-normal distribution. Note also that unlike, for example, a truncated normal distribution that retains the overall density function shape of the normal distribution over the real line but truncates it at a specific point, the skew-normal distribution is fundamentally different in shape over the real line from the normal distribution.

[4]Besides, there is substantial scope for combining Bhat's MACML approach with the sparse matrix approaches of Pace and LeSage (2011) and Liesensfeld *et al.* (2013) in future studies for even more substantial computational benefits.

spillover.[5] However, our methodology itself is immediately applicable to the spatial error specification and more general spatial specifications too.

The rest of this paper is structured as follows. The next section provides an overview of the multivariate skew normal distribution as a prelude to its introduction to develop a spatial model with each observation's kernel error term specified as being univariate skew-normal. The third section presents the model framework and estimation procedure for the proposed spatial skew-normal GOR model. Section 4 demonstrates an application of the model, and the final section concludes the paper.

---

[5]See Beck *et al.* (2006), McMillen (2010), and Sidharthan and Bhat (2012) for a discussion of why the spatial lag specification is a more sound basis for specifying spatial dependence than is the spatial error specification. In this regard, in our opinion, the negative assessments of the spatial lag specification found in McMillen (2012) and Gibbons and Overman (2012) go a little too far. McMillen views the spatial lag model specification itself as being very restrictive and fully parametric, which need not be the case. For instance, McMillen says "What is ironic is that the common estimation methods are derived from parametric specifications and log-likelihood functions that rely heavily on a known model structure." As we show in this paper, there can be semi-parametric versions of the spatial lag model specification. And, in terms of modeling causality, we do believe that Gibbon and Overman's criticism of the spatial lag and related traditional structural specifications overreach. Keane (2010) provides an excellent and strong rebuttal to those who espouse the experimentalist paradigm and denounce the structural paradigm, while also underscoring the issue of understanding and isolating the actual pathways of effects in econometric models.

Chapter 2. The Skew-Normal Distribution

In this section, we provide an overview of the multivariate skew-normal distribution, and briefly present the properties of the distribution that are most relevant in the context of application for spatial GOR models.

There are several multivariate versions of the skew-normal distribution in the literature (see Arellano-Valle and Azzalini, 2006 for a discussion of these many variants, and a unified treatment of these; Lee and McLachlan, 2013, 2014 also present the many multivariate variants). All of these share several properties similar to the multivariate normal distribution. In this paper, we select the restricted multivariate skew normal (MSN) distribution version originally proposed by Azzalini and Dalla Valle (1996) and labeled as the rMSN distribution by Lee and McLachlan (there are several parameterization variants within the rMSN specification, all equivalent to one another; here, we will use the unified skew-normal (or SUN) representation of Arellano-Valle and Azzalini, 2006). This representation of the rMSN distribution is based on a conditioning mechanism as will be discussed later. The rMSN version is particularly well suited for spatial analysis, especially because of the nature of the kernel error terms with the same level of skew across observations. It is also closed under any affine transformation of the skew-normally distributed vector as well as is closed under marginalization (both of which are the key to the MACML estimation of the spatial skew-normal GOR model). Of particular importance is that the cumulative distribution function of a D-variate skew normally distributed variable of the rMSN distribution requires only the evaluation of a $(D+1)$-dimensional multivariate cumulative normal distribution function. In the context of the spatial GOR model, this implies that one can use a composite marginal likelihood approach (see Paleti and Bhat, 2013 for a recent review of this approach) for estimation that entails only the evaluation of a three-dimensional multivariate cumulative normal distribution (MVNCD). When supplemented with an analytic approximation to compute this three-dimensional integral, as proposed by Bhat (2011) in his MACML approach, the net result is the need to evaluate only univariate and bivariate cumulative normal distribution functions. This enables the practical estimation of the spatial skew-normal GOR (or SSN-GOR) model.

The SUN representation of the rMSN distribution may be obtained as follows. Consider a $(D+1)$-variate normally distributed vector $\tilde{\mathbf{M}} = (\tilde{M}_1, \tilde{\mathbf{M}}_2')'$, where $\tilde{M}_1$ is a latent (1×1)-vector and $\tilde{\mathbf{M}}_2$ is a $(D\times1)$-vector:

$$\begin{pmatrix} \tilde{M}_1 \\ \tilde{\mathbf{M}}_2 \end{pmatrix} \sim N_{d+1}\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \mathbf{\Omega}_+^* \right), \text{ where } \mathbf{\Omega}_+^* = \begin{pmatrix} 1 & \mathbf{\rho}' \\ \mathbf{\rho} & \mathbf{\Omega}^* \end{pmatrix}. \tag{1}$$

$\mathbf{0}$ represents a column vector of zeros of the appropriate dimension (of dimension D in the above case), and $\mathbf{\rho}$ is a $(D\times1)$-vector, each of whose elements may lie between -1 and $+1$. The matrix $\mathbf{\Omega}^*$ is a positive-definite $D\times D$ correlation matrix (this is later parameterized in a specific form in the context of a spatial model, as discussed in the next

section). Then, $\mathbf{Z} = \tilde{\mathbf{M}}_2 \mid (\tilde{M}_1 > 0)$ has the standard multivariate skew-normal (SMVSN) density function shown below:

$$\tilde{\phi}_D(\mathbf{z}; \mathbf{\Omega}_+^*) = 2\phi_D(\mathbf{z}; \mathbf{\Omega}^*)\Phi(\mathbf{\alpha}'\mathbf{z}), \text{ where } \mathbf{\alpha} = \frac{(\mathbf{\Omega}^*)^{-1}\mathbf{\rho}}{\left(1 - \mathbf{\rho}'(\mathbf{\Omega}^*)^{-1}\mathbf{\rho}\right)^{1/2}}. \tag{2}$$

where $\phi_D(.)$ and $\Phi(.)$ represent the standard multivariate normal density function of D dimensions and the standard univariate cumulative distribution function, respectively.[6] The cumulative distribution function for $\mathbf{Z}$ may be obtained as:

$$P(\mathbf{Z} < \mathbf{z}) = \tilde{\Phi}_D(\mathbf{z}; \mathbf{\Omega}_+^*) = 2\Phi_{D+1}([0, \mathbf{z}], \mathbf{\Omega}_-^*); \ \mathbf{\Omega}_-^* = \begin{pmatrix} 1 & -\mathbf{\rho}' \\ -\mathbf{\rho} & \mathbf{\Omega}^* \end{pmatrix}, \quad \text{and} \quad \text{where} \quad \Phi_{D+1}(.,.)$$

represents the represents the MVNCD function of dimension $D+1$. In notation form, we write $\mathbf{Z} \sim \text{MVSN}(\mathbf{0}, \mathbf{IDEN}, \mathbf{\Omega}_+^*)$, where $\mathbf{IDEN}$ is an appropriately sized identity matrix (of dimension D in the current case) representing the standard deviation of the elements of the $\tilde{\mathbf{M}}_2$ vector. The first three moments of the distribution may subsequently be obtained from the function above in a straightforward fashion with $b = \sqrt{2/\pi}$:

$$E(\mathbf{Z}) = \mathbf{\mu}_Z = b\mathbf{\rho}; \ \text{Var}(\mathbf{Z}) = \mathbf{\Omega}^* - \mathbf{\mu}_Z\mathbf{\mu}_Z', \text{ and} \tag{3}$$

$$\text{Skew}(\mathbf{Z}) = \mathbf{\gamma}_Z = \left(\frac{4-\pi}{2}\right)^2 \left(\frac{\mathbf{\mu}_Z'(\mathbf{\Omega}^*)^{-1}\mathbf{\mu}_Z}{1 - \mathbf{\mu}_Z'(\mathbf{\Omega}^*)^{-1}\mathbf{\mu}_Z}\right)^3, \tag{4}$$

Defining $\mathbf{\omega}$ as a $(D \times D)$-diagonal matrix formed by the standard deviations of a $(D \times D)$ covariance matrix $\mathbf{\Omega}$ (i.e., $\mathbf{\Omega}^* = \mathbf{\omega}^{-1}\mathbf{\Omega}\mathbf{\omega}^{-1}$ and $\mathbf{\Omega} = \mathbf{\omega}\mathbf{\Omega}^*\mathbf{\omega}$), the probability density function of the D-dimensional random variable $\mathbf{G} = \mathbf{\xi} + \mathbf{\omega}\mathbf{Z}$ $[\mathbf{G} \sim \text{MVSN}(\mathbf{\xi}, \mathbf{\omega}, \mathbf{\Omega}_+^*)]$ may be written in terms of the standard MVSN density function in Equation (2) as:

$$f_G(\mathbf{g}; \mathbf{\xi}, \mathbf{\omega}, \mathbf{\Omega}_+^*) = \left(\prod_{j=1}^{D} \omega_j\right)^{-1} \tilde{\phi}_D(\mathbf{\omega}^{-1}(\mathbf{g} - \mathbf{\xi}); \mathbf{\Omega}_+^*), \tag{5}$$

where ω_j is the jth diagonal element of the matrix $\mathbf{\omega}$. The moments for the variable $\mathbf{G} = \mathbf{\xi} + \mathbf{\omega}\mathbf{Z}$, which is non-standard skew-normally distributed, may be obtained as $\mathbf{\mu}_G = \mathbf{\xi} + \mathbf{\omega}\mathbf{\mu}_Z$, $\text{Var}(\mathbf{G}) = \mathbf{\omega}\text{Var}(\mathbf{Z})\mathbf{\omega}$, and $\mathbf{\gamma}_G = \mathbf{\gamma}_Z$. The corresponding cumulative distribution function for $\mathbf{G}$ is:

$$P(\mathbf{G} < \mathbf{g}) = \tilde{\Phi}_D\left(\mathbf{\omega}^{-1}(\mathbf{g} - \mathbf{\xi}); \mathbf{\Omega}_+^*\right) = 2\Phi_{D+1}\left(0, \mathbf{\omega}^{-1}(\mathbf{g} - \mathbf{\xi}), \mathbf{\Omega}_-^*\right). \tag{6}$$

---

[6]The presence of the factor "2" in Equation (2) may be explained in many ways, including based on a conditioning stochastic representation of the skew-normal distribution (see Equations (2) and (3), and Equations (5) and (6), of Sidharthan and Bhat (2012)). A more intuitive way to explain it is that it is a normalizing factor needed to ensure that the cumulative distribution function as $\mathbf{z} \to \infty$ reaches the value of one (see Appendix A.1 of Sidharthan and Bhat, 2012).

A couple of properties of the MVSN distribution that will be useful in the development of the SSN-GOR model are now provided below without proof (the proofs may be found in many different sources, including Bhat and Sidharthan, 2012 and Arellano-Valle and Azzalini, 2006):

2.1 Property 1

The affine transformation of the MVSN distributed vector $\mathbf{G}$ (of dimension $d \times 1$) $[\mathbf{G} \sim \text{MVSN}(\boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\Omega}_{+}^{*})]$ as $\mathbf{a} + \mathbf{BG}$, where $\mathbf{B}$ is a $(h \times d)$ matrix is also a MVSN distributed vector of dimension $h \times 1$:

$$\mathbf{a} + \mathbf{BG} \sim \text{MVSN}(\mathbf{a} + \mathbf{B}\boldsymbol{\xi}, \tilde{\boldsymbol{\omega}}, \tilde{\boldsymbol{\Omega}}_{+}^{*})], \text{ where } \tilde{\boldsymbol{\Omega}}_{+}^{*} = \begin{pmatrix} 1 & \tilde{\boldsymbol{\rho}}' \\ \tilde{\boldsymbol{\rho}} & \tilde{\boldsymbol{\Omega}}^{*} \end{pmatrix}, \tilde{\boldsymbol{\Omega}}^{*} = (\tilde{\boldsymbol{\omega}})^{-1}\tilde{\boldsymbol{\Omega}}(\tilde{\boldsymbol{\omega}})^{-1}, \tilde{\boldsymbol{\Omega}} = \mathbf{B}\boldsymbol{\Omega}\mathbf{B}',$$

$\tilde{\boldsymbol{\rho}} = (\tilde{\boldsymbol{\omega}})^{-1}\mathbf{B}\boldsymbol{\omega}\,\boldsymbol{\rho}$, and $\tilde{\boldsymbol{\omega}}$ is the diagonal matrix of standard deviations of $\tilde{\boldsymbol{\Omega}}$.

2.2 Property 2

If $[\mathbf{G} \sim \text{MVSN}(\boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\Omega}_{+}^{*})]$ is partitioned into two-subvectors $\mathbf{G}_1$ and $\mathbf{G}_2$, with corresponding partitions of the other vectors as follows:

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \quad \boldsymbol{\omega} = \begin{pmatrix} \boldsymbol{\omega}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\omega}_2 \end{pmatrix}, \text{ and } \boldsymbol{\Omega}_{+}^{*} = \begin{pmatrix} 1 & \boldsymbol{\rho}_1' & \boldsymbol{\rho}_2' \\ \boldsymbol{\rho}_1 & \boldsymbol{\Omega}_1^{*} & \boldsymbol{\Omega}_{12}^{*} \\ \boldsymbol{\rho}_2 & \boldsymbol{\Omega}_{12}^{*} & \boldsymbol{\Omega}_2^{*} \end{pmatrix}, \text{ then}$$

$$\mathbf{G}_1 \sim \text{MVSN}(\boldsymbol{\xi}_1, \boldsymbol{\omega}_1, \boldsymbol{\Omega}_{1+}^{*})], \boldsymbol{\Omega}_{1+}^{*} = \begin{pmatrix} 1 & \boldsymbol{\rho}_1' \\ \boldsymbol{\rho}_1 & \boldsymbol{\Omega}_1^{*} \end{pmatrix}, \text{ and } \mathbf{G}_2 \sim \text{MVSN}(\boldsymbol{\xi}_2, \boldsymbol{\omega}_2, \boldsymbol{\Omega}_{2+}^{*})], \boldsymbol{\Omega}_{2+}^{*} = \begin{pmatrix} 1 & \boldsymbol{\rho}_2' \\ \boldsymbol{\rho}_2 & \boldsymbol{\Omega}_2^{*} \end{pmatrix}.$$

That is, the rMSN distribution is closed under marginalization, which can be shown by straightforward integration (Azzalini, 2005).

Chapter 3. Modeling Framework

3.1 The Basic Structure

Let q ($q = 1,2,...,Q$) be an index to represent observations of an ordered response variable taking on the values $k \in \{1,2,...,K\}$. The latent variable underlying the ordered response observation y_q is denoted by y_q^*. The equation system for the spatial skew-normal GOR (or SSN-GOR) model then takes the following form:

$$y_q^* = \delta \sum_{q'=1}^{Q} w_{qq'} y_{q'}^* + \boldsymbol{b}' \boldsymbol{x}_q + \varepsilon_q, \ y_q = k \text{ if } \psi_{q,k-1} < y_q^* < \psi_{q,k}, \tag{7}$$

where $\boldsymbol{x}_q$ is an ($L \times 1$)-column vector of exogenous attributes (excluding a constant), and $\boldsymbol{b}$ is a corresponding ($L \times 1$)-column vector. The $w_{qq'}$ terms are the elements of an exogenously defined distance-based spatial weight matrix $\mathbf{W}$ corresponding to observations q and q' (with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$), and δ ($-1 < \delta < 1$) is the spatial autoregressive parameter. To ensure the constraints on the spatial lag term δ, we parameterize it as $\delta = \pm e^{\tilde{\delta}}/(1+e^{\tilde{\delta}})$. Once estimated, the $\tilde{\delta}$ estimate can be translated back to an estimate of δ. In our empirical analysis, we expect δ to be positive because attitudes/preferences are likely to be reinforcing through social interactions.

The formulation above generates spatial dependence through the spatial lag term, the nature of which is related to the specification of the weight terms $w_{qq'}$. The weight matrix can take the form of a contiguity specification ($w_{qq'} = 1$ if the parcels q and q' are adjacent and 0 otherwise), or a specification based on a distance threshold ($w_{qq'} = c_{qq'} / \sum_{q'} c_{qq'}$, where $c_{qq'}$ is a dummy variable taking the value 1 if the parcel q' is within the distance threshold and 0 otherwise), or based on the inverse of distance $d_{qq'}$ and its power functions $\left(w_{qq'} = (1/d_{qq'}^n) \left[\sum_{q'} 1/d_{qq'}^n \right]^{-1} \right)$ ($n > 0$), or the shared border length $\tilde{d}_{qq'}$ between parcels $w_{qq'} = \tilde{c}_{qq'} \tilde{d}_{qq'} / \left(\sum_{q'} \tilde{c}_{qq'} \tilde{d}_{qq'} \right)$ (where $\tilde{c}_{qq'}$ is a dummy variable taking the value 1 if the parcels q and q' are adjoining, and 0 otherwise). All of these functional forms and others for the weight matrix may be tested empirically. Finally, in Equation (7), ε_q is an error term capturing the effects of unobserved factors on the latent propensity.

In the usual ordered-response fashion, the latent propensity y_q^* is mapped to the observed level y_q through the thresholds $\psi_{q,k}$ ($\psi_{q,0} = -\infty$ and

$\psi_{q,K} = \infty$; $\psi_{q,1} < \psi_{q,2} < \ldots < \psi_{q,K-1}$ $\forall q$). To allow heterogeneity (across observations) in the thresholds, they are parameterized as a function of a relevant exogenous variable vector z_{qk} (excluding a constant), as in Eluru *et al.* (2008):

$$\psi_{q,k} = \psi_{q,k-1} + \exp(\lambda_k + \boldsymbol{\varphi}_k' z_{qk}).\tag{8}$$

where λ_k is a scalar, and $\boldsymbol{\varphi}_k$ is a vector of coefficients associated with level $k = 1, 2, \ldots, K-1$. For identification reasons, impose the normalization $\psi_{q,1} = \exp(\lambda_1)$ for all q (that is, all elements of the vector $\boldsymbol{\varphi}_1$ take the value of zero). For future use, define the following, let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_K)'$, and $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_2', \ldots, \boldsymbol{\varphi}_K')'$. If all elements of the vector $\boldsymbol{\varphi}$ are uniformly zero, the GORL formulation collapses to the standard ordered-response formulation.

The latent propensity representation of Equation (7) can be written equivalently in vector notation as:

$$\boldsymbol{y}^* = \delta \mathbf{W} \boldsymbol{y}^* + \mathbf{x} \boldsymbol{b} + \boldsymbol{\varepsilon},\tag{9}$$

where $\boldsymbol{y}^* = (y_1^*, y_2^*, \ldots, y_Q^*)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_Q)'$ are $(Q \times 1)$ vectors, and $\mathbf{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_Q)'$ is a $(Q \times L)$ matrix of exogenous variables for all Q units. Defining $\mathbf{S} = \left[\mathbf{IDEN}_Q - \delta \mathbf{W} \right]^{-1}$ [$(Q \times Q)$ matrix], where $\mathbf{IDEN}_Q$ is an identity matrix of dimension Q. Equation (9) may be re-written as:

$$\boldsymbol{y}^* = \mathbf{S}\mathbf{x}\boldsymbol{b} + \mathbf{S}\boldsymbol{\varepsilon}.\tag{10}$$

The distribution of the vector $\boldsymbol{y}^*$ is determined by the distribution of the vector $\boldsymbol{\varepsilon}$. In this paper, we assume a specific rMSN distributional form for $\boldsymbol{\varepsilon}$ that satisfies the following properties: (1) The marginal distribution of each element of $\boldsymbol{\varepsilon}$ is identically univariate skew-normal distributed, and (2) there is no dependence in the elements of $\boldsymbol{\varepsilon}$ other than what is automatically generated because of the skew-normality of each ε_q term. The first property is maintained in a simple way by specifying each ε_q term as a standard univariate skew-normal random term with identical skew determined by the parameter ρ. That is, $\varepsilon_q \sim \text{MVSN}\left(0, 1, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \forall q$. It is the second property where we propose a specific form of the multivariate skew-normal (rMSN) distribution introduced in the previous section as the distributional form for $\boldsymbol{\varepsilon}$: $\boldsymbol{\varepsilon} \sim \text{MVSN}(\mathbf{0}, \boldsymbol{\omega}, \boldsymbol{\Omega}_+^*)$ with $\boldsymbol{\omega}$ being the identity matrix of size Q (that is, $\boldsymbol{\omega} = \mathbf{IDEN}_Q$) (this is needed for identification, similar to the use of a standard normal or standard logistic distribution in traditional ordered-response models) and $\boldsymbol{\Omega}_+^*$ as follows:

$$\mathbf{\Omega}_+^* = \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho^2 & \cdots & \rho^2 \\ \rho & \rho^2 & 1 & \cdots & \rho^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho & \rho^2 & \rho^2 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{\rho}' \\ \mathbf{\rho} & \mathbf{\Omega}^* \end{pmatrix}, \ \mathbf{\rho} = \begin{pmatrix} \rho \\ \rho \\ \vdots \\ \rho \end{pmatrix} \text{ and } \mathbf{\Omega}^* = \begin{pmatrix} 1 & \rho^2 & \cdots & \rho^2 \\ \rho^2 & 1 & \cdots & \rho^2 \\ \vdots & \rho^2 & \vdots & \vdots \\ \rho^2 & \rho^2 & \cdots & 1 \end{pmatrix} \quad (11)$$

The specification above is very parsimonious (dependent on only one term ρ), and satisfies both the desired properties mentioned earlier. It is straightforward to note that the marginal distribution of each ε_q is standard univariate skew-normal, as defined earlier. Further, it is straightforward to note that, when $\rho=0$, the multivariate vector becomes multivariate normally distributed with independence across the marginals. That is, the only source of dependence among the elements of $\mathbf{\varepsilon}$ in our specification is the element that generates the skew in each of the elements. After accommodating the skew, there is no additional dependence in the elements of $\mathbf{\varepsilon}$.[7] Another way to notice this is that the Cholesky decomposition matrix of our specification of $\mathbf{\Omega}_+^*$ in the equation above is given by:

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \rho & \sqrt{1-\rho^2} & 0 & 0 & 0 \\ \rho & 0 & \sqrt{1-\rho^2} & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & 0 & 0 & 0 & \sqrt{1-\rho^2} \end{pmatrix}.$$

Except for the first column that has entries of ρ, the Cholesky matrix is otherwise diagonal. Another benefit of our specification is that $\mathbf{\Omega}_+^*$ is immediately positive definite as long as $|\rho|<1$. Overall, we have not seen this special form of the rMSN proposed here used anywhere before.

Finally, from Equation (10) and property 1, we obtain the multivariate skew-normal distribution of the vector $\mathbf{y}^*$ as follows:

$$\mathbf{y}^* \sim \mathrm{MVSN}(\tilde{\mathbf{B}}, \tilde{\mathbf{\omega}}, \tilde{\mathbf{\Omega}}_+^*)],$$

$$\tilde{\mathbf{B}} = \mathbf{S}\mathbf{x}b, \ \tilde{\mathbf{\Omega}}_+^* = \begin{pmatrix} 1 & \tilde{\mathbf{\rho}}' \\ \tilde{\mathbf{\rho}} & \tilde{\mathbf{\Omega}}^* \end{pmatrix}, \ \tilde{\mathbf{\Omega}}^* = (\tilde{\mathbf{\omega}})^{-1}\tilde{\mathbf{\Omega}}(\tilde{\mathbf{\omega}})^{-1}, \tilde{\mathbf{\Omega}} = \mathbf{S}\mathbf{\Omega}\mathbf{S}', \ \tilde{\mathbf{\rho}} = (\tilde{\mathbf{\omega}})^{-1}\mathbf{S}\mathbf{\rho}, \quad (12)$$

where $\tilde{\mathbf{\omega}}$ is the diagonal matrix of standard deviations of $\tilde{\mathbf{\Omega}}$.

---

[7] By construction, the rMSN distribution generates a dependence across the skew-normally distributed variables. That is, the rMSN distribution does not allow independence among the marginals, except in the case when all elements of $\mathbf{\rho}$ in Equation (11) are zero (which is the case of multivariate normality). When all elements of $\mathbf{\rho}$ are equal to zero, independence among marginals is achieved in the special case when $\mathbf{\Omega}^*$ is a diagonal matrix.

An important point here before moving on to estimation. We are not encouraging the use of a skew-normal kernel error term within a spatial context as an alternative to (or as an escape from) a good model specification in the first place based on theory and past studies. A variety of model mis-specifications by way of omitted variables, a potentially incorrect weight matrix for accommodating spatial/social dependence, or even a different process other than the spatial lag process may all manifest themselves incorrectly in the form of a skew-normal error distribution and/or the presence of a statistically significant spatial auto-regressive parameter. However, it is also true of any econometric model that one has to start from some a priori information and functional form. After all, an econometric model is not intended to be reality, but as close of an abstraction of that reality as possible. In other words, any econometric model can potentially be mis-specified on multiple grounds, and this is particularly the case when one introduces spatial dependence. However, this should not also be taken as reason to reject more flexible model forms, as long as the analyst continues to place due emphasis on developing as cogent a specification as possible. Our introduction of the skew-normal distribution for the kernel error form should be viewed in this light of supplementing a rich specification of other elements of the model form.

3.2 Model Estimation

The parameter vector to be estimated in the model is denoted $\boldsymbol{\theta} = (\boldsymbol{b}', \delta, \rho, \boldsymbol{\varphi}', \boldsymbol{\lambda}')'$. Several restrictive models are obtained from the spatial model formulation developed here. If $\delta = 0$, but $\rho \neq 0$, the result is the aspatial skew-normal GOR (ASSN-GOR) model. If $\delta \neq 0$, but $\rho = 0$, the result is the spatial probit GOR (SP-GOR) model that has been used in the spatial econometric literature. If both $\delta = 0$ and $\rho = 0$, the result is the aspatial probit GOR (ASP-GOR) model. Of course, in each of these instances, if all the elements of $\boldsymbol{\varphi}$ are also zero, the result is the corresponding standard ordered-response (SOR) model.

The likelihood function $L(\boldsymbol{\theta})$ for the SRC-GORP model may then be obtained from the fact that $\boldsymbol{y}^*$ is multivariate skew-normally distributed:

$$L(\boldsymbol{\theta}) = P(\boldsymbol{y} = \boldsymbol{m}) = \int_{D_{\boldsymbol{y}^*}} \left(\prod_{q=1}^{Q} \tilde{\omega}_j \right)^{-1} \left[2\phi_Q(\tilde{\boldsymbol{\omega}}^{-1}(\boldsymbol{y}^* - \tilde{\mathbf{B}}); \tilde{\boldsymbol{\Omega}}^*) \right] \left[\Phi(\boldsymbol{\alpha}'\boldsymbol{y}^*) \right] d\boldsymbol{y}^*, \tag{13}$$

where $\boldsymbol{y} = (y_1, y_2, ..., y_Q)'$, $\boldsymbol{m} = (m_1, m_2, ..., m_Q)'$ is the corresponding ($Q \times 1$) vector of the actual observed ordered-response outcomes, $D_{\boldsymbol{y}^*}$ is the integration domain defined as $D_{\boldsymbol{y}^*} = \{\boldsymbol{y}^* : \psi_{q, m_q - 1} < y_q^* < \psi_{q, m_q}, \forall q = 1, 2, ..., Q\}$, $\phi_Q(.)$ is the multivariate normal density function of dimension Q, and $\boldsymbol{\alpha} = \dfrac{(\tilde{\boldsymbol{\Omega}}^*)^{-1} \tilde{\boldsymbol{\rho}}}{\left(1 - \tilde{\boldsymbol{\rho}}'(\tilde{\boldsymbol{\Omega}}^*)^{-1} \tilde{\boldsymbol{\rho}}\right)^{1/2}}$.

The rectangular integral in the likelihood function is of dimension Q, which becomes difficult to evaluate and increasingly less accurate using traditional numerical simulation techniques (Bhat *et al.*, 2010; Müller and Czado, 2005; Bhat *et al.*, 2016). The alternative

is to use the composite marginal likelihood (CML) approach, exploiting the closure under marginalization property of the multivariate skew-normal distribution (Property 2).

The CML inference approach is based on maximizing a surrogate likelihood function that compounds much easier-to-compute, lower-dimensional, marginal likelihoods (see Varin *et al.*, 2011 for an extensive review of CML methods; Lindsay *et al.*, 2011, Bhat, 2011, and Yi *et al.*, 2011 are also useful references). The CML approach, which belongs to the more general class of composite likelihood function approaches (see Lindsay, 1988), may be explained in a simple manner as follows. In the SSN-GOR model, instead of developing the likelihood function for the entire set of Q observations, as in Equation (13), one may compound (multiply) pairwise probabilities of observation q being in ordinal state m and observation q' being in ordinal state m', of observation q being in ordinal state m and observation q'' being in ordinal state m'', and so on. The CML estimator (in this instance, the pairwise CML estimator) is then the one that maximizes the compounded probability of all pairwise events. The properties of the CML estimator may be derived using the theory of estimating equations (see Cox and Reid, 2004, Yi *et al.*, 2011). Specifically, under usual regularity assumptions (Molenberghs and Verbeke, 2005, page 191, Xu and Reid, 2011), the CML estimator is consistent and asymptotically normal distributed (this is because of the unbiasedness of the CML score function, which is a linear combination of proper score functions associated with the marginal event probabilities forming the composite likelihood; for a formal proof, see Yi *et al.*, 2011 and Bhat, 2014).

The pairwise CML function for the SSN-GOR model may be written as follows:

$$L_{CML}(\boldsymbol{\theta}) = \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} L_{qq'}, \text{ where } L_{qq'} = P\left([\boldsymbol{y}]_q = [\boldsymbol{m}]_q, [\boldsymbol{y}]_{q'} = [\boldsymbol{m}]_{q'}\right)$$

$$= 2 \times \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} \left[\begin{array}{l} \Phi_3\left(0, \tau_q, \tau_{q'}; \breve{\boldsymbol{\Omega}}^*_{-qq'}\right) - \Phi_3\left(0, \tau_q, \pi_{q'}; \breve{\boldsymbol{\Omega}}^*_{-qq'}\right) \\ -\Phi_3\left(0, \pi_q, \tau_{q'}; \breve{\boldsymbol{\Omega}}^*_{-qq'}\right) + \Phi_3\left(0, \pi_q, \pi_{q'}; \breve{\boldsymbol{\Omega}}^*_{-qq'}\right) \end{array} \right], \tag{14}$$

where $\tau_q = \dfrac{\psi_{q,m_q} - [\tilde{\mathbf{B}}]_q}{\omega_q}$, $\pi_q = \dfrac{\psi_{q,m_q-1} - [\tilde{\mathbf{B}}]_q}{\omega_q}$, $[\tilde{\mathbf{B}}]_q$ represents the qth element of the

column vector $\tilde{\mathbf{B}}$, and $\breve{\boldsymbol{\Omega}}^*_{-qq'} = \Delta_{qq'} \tilde{\boldsymbol{\Omega}}^*_{-} \Delta'_{qq'}$, with $\Delta_{qq'}$ being a $(3 \times Q)$ selection matrix constructed as follows: (1) First fill the entire matrix with values of zero, (2) Place a value of '1' in the first row and first column, and (3) Place a value of '1' in the qth column of the second row, and a value of '1' in the q'^{th} column of the third row.

The CML function in the equation above can be computed using the maximum simulated likelihood function to evaluate the three-dimensional multivariate normal cumulative distribution (MVNCD) function. While this can be done relatively easily using quasi-Monte Carlo methods (see Bhat, 2001 and Bhat, 2003), we use an analytic approximation embedded in the MACML approach of Bhat (2011) to evaluate the MVNCD function. This is because the analytic approximation requires only the evaluation of univariate and bivariate MVNCD functions, and is computationally much faster than the MSL approach. Further, the analytic approximation embedded in the MACML inference approach

generally recovers parameters much better than the simulated approach of MSL, because of a much smoother analytic approximated surface to be maximized than a simulation approximated surface to be maximized. This latter result is illustrated in Bhat *et al*., 2010 and Bhat and Sidharthan, 2012.

The pairwise marginal likelihood function of Equation (14) comprises four MVNCD function computations per pair of observation units, and $Q(Q-1)/2$ pairs of observation units. This can itself become quite time consuming. However, previous studies (Varin and Vidoni, 2008, Bhat *et al*., 2010, Varin and Czado, 2010) have shown that spatial dependency drops quickly with inter-observation distance. Therefore, there is no need to retain all observation pairs because the pairs formed from the closest observations provide much more information than pairs far from one another. The "optimal" distance for including pairings can be based on minimizing the trace of the asymptotic covariance matrix. Thus, the analyst can start with a low value of the distance threshold (leading to a low number of pairwise terms in the CML function) and then continually increase the distance threshold up to a point where the gains from increasing the distance threshold is very small or even drops. To be specific, for a given threshold, construct a $Q \times Q$ matrix $\tilde{\mathbf{R}}$ with its qth column filled with a $Q \times 1$ vector of zeros and ones as follows: if the observational unit q' is not within the specified threshold distance of unit q, the q'^{th} row has a value of zero; otherwise, the q'^{th} row has a value of one. By construction, the qth row of the qth column has a value of one. Let $[\tilde{\mathbf{R}}]_{qq'}$ be the qq^{th} element of the matrix $\tilde{\mathbf{R}}$, and let $\tilde{W} = \sum_{q=1}^{Q-1} \sum_{q'=q+1}^{Q} [\tilde{\mathbf{R}}]_{qq'}$. Define a set $\tilde{\mathbf{C}}_q$ of all individuals (observation units) that have a value of '1' in the vector $[\tilde{\mathbf{R}}]_q$, where $[\tilde{\mathbf{R}}]_q$ is the qth column of the vector $\tilde{\mathbf{R}}$. Then, the CML function is as follows:

$$L_{CML}(\boldsymbol{\theta}) = \prod_{\substack{q=1}}^{Q-1} \prod_{\substack{q'=q+1 \\ q' \in \tilde{C}_q}}^{Q} L_{qq'} . \tag{15}$$

Under usual regularity assumptions (Molenberghs and Verbeke, 2005, Xu and Reid, 2011, Bhat, 2014), the CML estimator of $\boldsymbol{\theta}$ is consistent and asymptotically normal distributed with asymptotic mean $\boldsymbol{\theta}$ and covariance matrix given by $\dfrac{\left[\hat{\boldsymbol{H}}^{-1}\right]\left[\hat{\boldsymbol{J}}\right]\left[\hat{\boldsymbol{H}}^{-1}\right]}{\tilde{W}}$, where

$$\hat{\boldsymbol{H}} = -\frac{1}{\tilde{W}} \left[\sum_{\substack{q=1}}^{Q-1} \sum_{\substack{q'=q+1 \\ q'' \in \tilde{C}_q}}^{Q} \frac{\partial^2 \log L_{qq'}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \right]_{\hat{\boldsymbol{\theta}}_{CML}}, \text{ or alternatively,} \tag{16}$$

$$\hat{\boldsymbol{H}} = \frac{1}{\tilde{W}} \sum_{\substack{q=1}}^{Q-1} \sum_{\substack{q'=q+1 \\ q' \in \tilde{C}_q}}^{Q} \left(\left[\frac{\partial \log L_{qq'}}{\partial\boldsymbol{\theta}} \right] \left[\frac{\partial \log L_{qq'}}{\partial\boldsymbol{\theta}'} \right] \right)_{\hat{\boldsymbol{\theta}}_{CML}} \tag{17}$$

14

However, the estimation of the "vegetable" matrix $\boldsymbol{J}$ is more difficult in this case. One cannot empirically estimate $\boldsymbol{J}$ as the sampling variance of the individual contributions to the composite score function because of the underlying spatial dependence in observation units. But a windows resampling procedure (see Heagerty and Lumley, 2000) may be used to estimate $\boldsymbol{J}$. While there are several ways to implement this, Bhat (2011) suggests overlaying the spatial region under consideration with a square grid providing a total of $\tilde{Q}$ internal and external nodes.[8] Then, select the observational unit closest to each of the $\tilde{Q}$ gird nodes to obtain $\tilde{Q}$ observational units from the original Q observational units ($\tilde{q} = 1, 2, 3, \ldots, \tilde{Q}$). Let $\tilde{\boldsymbol{R}}_{\tilde{q}}$ be the $Q \times 1$ matrix representing the $\tilde{q}^{th}$ column vector of the matrix $\tilde{\boldsymbol{R}}$, let $\tilde{\boldsymbol{C}}_{\tilde{q}}$ be the set of all individuals (observation units) that have a value of '1' in the vector $\tilde{\boldsymbol{R}}_{\tilde{q}}$, and let $\boldsymbol{y}_{\tilde{q}}$ be the sub-vector of $\boldsymbol{y}$ with values of '1' in the rows of $\tilde{\boldsymbol{R}}_{\tilde{q}}$. Let $N_{\tilde{q}}$ be the sum (across rows) of the vector $\tilde{\boldsymbol{R}}_{\tilde{q}}$ (that is $N_{\tilde{q}}$ the cardinality of $\tilde{\boldsymbol{C}}_{\tilde{q}}$), so that the dimension of $\boldsymbol{y}_{\tilde{q}}$ is $N_{\tilde{q}} \times 1$. Let $l_{\tilde{q}}$ be the index of all elements in the vector $\boldsymbol{y}_{\tilde{q}}$, so that $l_{\tilde{q}} = 1, 2, \ldots, N_{\tilde{q}}$. Next, define $\ddot{C}_{\tilde{q}} = \left[N_{\tilde{q}} (N_{\tilde{q}} - 1) \right] / 2$. Then, the $\boldsymbol{J}$ matrix maybe empirically estimated as:

$$\hat{\boldsymbol{J}} = \frac{1}{\tilde{Q}} \left[\sum_{\tilde{q}=1}^{\tilde{Q}} \left[\frac{1}{\ddot{C}_{\tilde{q}}} \left(\left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}-1} \sum_{l'=l_{\tilde{q}}+1}^{N_{\tilde{q}}} \frac{\partial \log L_{l_{\tilde{q}} l'}}{\partial \boldsymbol{\theta}} \right] \left[\sum_{l_{\tilde{q}}=1}^{N_{\tilde{q}}-1} \sum_{l'=l_{\tilde{q}}+1}^{N_{\tilde{q}}} \frac{\partial \log L_{l_{\tilde{q}} l'}}{\partial \boldsymbol{\theta}} \right]' \right) \right] \right]_{\hat{\boldsymbol{\theta}}_{CML}}. \tag{18}$$

The CML estimator loses some asymptotic efficiency from a theoretical perspective relative to a full likelihood estimator (Lindsay, 1988, Zhao and Joe, 2005, Bhat, 2011), though several studies have found this efficiency loss to be negligible to small from an empirical standpoint (see Zhao and Joe, 2005, Lele, 2006, Joe and Lee, 2009). Further, when simulation methods have to be used to evaluate the likelihood function, there is also a loss in asymptotic efficiency in the maximum simulated likelihood (MSL) estimator relative to a full likelihood estimator (see McFadden and Train, 2000). In earlier empirical comparisons of the CML and MSL estimators in cases where the MSL estimation is feasible, little difference has been found in the efficiency of the two estimators, and the CML estimator has the benefit of a very substantial reduction in computation time and much superior convergence properties (see Bhat *et al.*, 2010, and Paleti and Bhat, 2013).

---

[8] Note also that in the case of spatial models, we need a fast enough rate at which the Godambe information converges to infinity (for good asymptotic properties). This holds when the spatial correlation fades quickly over distance, as is the case when using spatial lag and related models (see Varin *et al.*, 2011 for a good review).

Chapter 4. Application to Bicycling Frequency

4.1 Background

In 2007, 130 million bicycles were produced around the world, more than twice the number of cars produced (Worldwatch Institute, 2008). In 2012, more bicycles than cars were sold in all of the member countries of the European Union, except Belgium and Luxembourg (COLIBI/COLIPED, 2013; ACEA, 2013). Although bicycles have been perceived as a transportation mode exclusively for health-conscious and environmentally conscious people (see Horton, 2006, Pretty *et al.*, 2007, Steinbach *et al.*, 2011, and Jensen, 2013 for a detailed study of the relation between bicycle use and healthy/green lifestyles), the number of bicycle riders has been increasing since cities started promoting bicycling as a practical way to reduce traffic congestion and smog. Government leaders have been attempting to bring cycling to prominence in the urban transport mix and use this healthy, affordable, and compact mode of transportation to fight not only congestion, but also climate change and the emerging obesity epidemic.[9] Cities worldwide, particularly in Europe, have promoted pro-bicycling transportation and land use policies and heavily funded bicycle infrastructure and public education (Roney, 2008, Pucher *et al.*, 2010). Although the modal share of bicycle trips is very low in the United States (about 1% of all trips, according to Pucher *et al.*, 2010), bicycling advocacy has been growing in recent years in the United States too. The League of American Bicyclists now honors 371 U.S. towns and cities as Bicycle Friendly Communities, compared with 52 in 2005 (League of American Bicyclists, 2015). The United States federal government provided $900 million a year to promote bicycling and walking between 2005 and 2009. The 50 largest U.S. cities planned to double their pedestrian/bicycle infrastructure during this period, including the installation of additional parking (racks), bicycle-friendly roads, and designated lanes (Roney, 2008).

The promotion of bicycling as an alternative mode appears to have been facilitated by the presence of a social interaction effect (through a peer effect or a peer pressure effect) inside urban communities (Salvy *et al.*, 2009, Ferdous *et al.*, 2011). As with other sustainable technologies, such as electric cars or solar panels, decisions about using a bicycle has been shown to be strongly related to social group influences. For example, Dill and Voros (2007) found that if an individual's co-workers bicycle to work, the individual is more likely to bicycle to work too. The notion of norms in one's social group impacting bicycling behavior is also consistent with the theory of planned behavior and the theory of interpersonal behavior (see Heinen *et al.*, 2010 for a good discussion). The increasing and reinforcing popularity of bicycle sharing systems provides another indirect evidence of social interaction effects. These sharing systems were available in 500 cities in 40 countries around the world by 2012 (Midgley, 2009, Shaheen *et al.*, 2010,

---

[9]The popular belief that bicycling is a dangerous activity has been questioned and refuted. Many studies have juxtaposed the obesity reduction and cardiovascular health improvement of bicycling relative to the potential health risks from traffic injuries (see, for example, Hamer and Chida, 2008, Huy *et al.*, 2008, Gordon-Larsen *et al.*, 2009, Elvik, 2009, and Pucher *et al.*, 2010), and concluded that the health benefits exceed the health risks.

and Meddin and DeMaio, 2012)[10], providing the riders with more opportunities to use multi-modal alternatives such as bicycling and the use of public transportation. Social interaction effects are further intensified by the use of smartphones. Currently several bicycling-related applications ("apps") are available that share with friends and family— or any social network—favorite routes, personal records, and challenges (Reddy *et al.*, 2010, Ertiö, 2015). Finally, formal and informal bicycling associations have arisen in various cities in the last few years, organizing massive rides across communities, initiating ciclovia (events that close streets on weekends to promote family-friendly bicycling-related activities), demanding better bicycling infrastructure and support from their respective governmental authorities, sharing and promoting the bicycling experience in different neighborhoods, facilitating public access to tools and training for bicycle maintenance and repair, and—most importantly—providing more opportunities for social interaction effects associated with bicycling propensity.

Despite the increasing recognition of social interaction effects in the bicycling literature, few empirical studies of bicycling frequency consider such effects. Indeed, as indicated by Heinen *et al.* (2010) in their review of the bicycling literature, research into psychological constructs, including social interaction and dependence effects, has been more theoretical. Though this situation has been changing in the past few years, with studies such as Maldonado-Hinarejos *et al.* (2014) and Ma *et al.* (2014) focusing on attitudes and perceptions, most of these studies still do not consider social interaction effects in bicycling frequency. Even when very occasionally considered, such as in Dill and Voros (2007), these effects are considered in simple univariate terms without controlling for other variables or other effects that may manifest as social interaction effects (for example, an individual may be more likely to bicycle to work if more co-workers bicycle not because of social interactions, but because bicycling facilities, such as a wash room and bicycle racks, at the common work place are good). In contrast, in this paper, we use a well-established conceptual and structural model to capture social dependence effects in the underlying bicycling propensity that gets manifested in the observed bicycling frequency. As we also demonstrate, introducing social dependence effects is not simply an esoteric econometric nuance; ignoring social interaction effects can underestimate the impacts of bicycling infrastructure improvements and public campaigns on bicycle use frequency, potentially (and incorrectly) under-projecting the benefits of bicycling-oriented projects. Further, incorrectly considering a normally distributed kernel error term in the bicycling propensity when a skew-normal distribution is more appropriate can lead to additional mis-projections of bicycling infrastructure investment improvements.

4.2 Data and Sample Formation

The data used in the analysis is based on the 2014 Puget Sound household travel survey. The Puget Sound Regional Council (PSRC) collected complete travel information of household members (over the age of 5 years) for a weekday (Tuesday, Wednesday, or Thursday) during the spring (April–June) of 2014 in the Puget Sound region (King,

---

[10]Bicycle-sharing systems are expected to expand even more in the near future (Faghih-Imani *et al.*, 2014).

Kitsap, Pierce, and Snohomish counties in the state of Washington, USA). Socio-demographic information was collected as provided by one household member (age 18 or older) deemed the *householder*. That information included household demographics (income, number of household members, and number and type of motorized vehicles), person-level demographics (including age, gender, education level, and employment status), and current home location characteristics (postal address, census tract, and census block group). The survey also collected individual responses on how many times in the past 30 days they have traveled using a bicycle for more than 15 minutes, which constituted the dependent variable in the current analysis.

From the original data set of 4,092 householders who were also workers, we removed those records that contained incomplete information on socioeconomic, household location and tenure/type, and travel characteristics relevant to the current analysis (332 records were deleted). The final data sample used in the estimation included 3,760 individuals, with about an equal percentage of men (49%) and women (51%). As expected in this pool of workers, a high percentage had attained a bachelor's (38.5%) or graduate (31.3%) education, with relatively low percentages of individuals with some college degree (21.5%) and high school or less education (8.7%). Most individuals (75.3%) were full-time workers, with 12.9% being part-time and 11.8% being self-employed. A majority of the individuals (44.4%) were between 35-54 years, with 29.3% in the 18–34-year age group, 20.1% in the 55–64-year age group, and a much lower percentage of 6.2% in the ">65 years" age category. 30% of the individuals reported one or more children in the household, and a majority of the individuals (58.9%) reported an annual household income of more than US $75,000.

4.3 Variable Specification

The dependent variable in our analysis is the bicycling frequency reported for each individual. The survey recorded bicycling frequency in ordinal categories, and for the purpose of this analysis, we identified five categories with a reasonable share of individuals in each response category: (1) 'I never do this' (56.5% of the sample), (2) 'I do this, but not in the past 30 days' (24.0%), (3) '1-3 times in the past 30 days' (7.2%), (4) '1-4 days per week' (9.0%), (5) '5 days or more per week' (3.3%). For ease in presentation, we will refer to these categories as "not a bicyclist", "occasional bicyclist", "infrequent bicyclist", "frequent bicyclist", and "everyday bicyclist".[11]

The smallest space unit available in the survey for household location identification is the census block group. The Puget Sound region has 2,647 census block groups, with an average size of 2.36 square miles (considering only the area of land, *i.e.*, excluding water). To construct our distance matrix for characterizing spatial proximity-based dependence, we located the centroid of each census block group and computed the Euclidean distance between centroids (in miles). For those individuals belonging to the same census block group, we assumed that the distance between them is 0.5 miles (our

---

[11]The trips considered for the individuals when they reported bicycling frequency are not necessarily confined to only commute bicycling. That is, the model developed in this paper examines bicycling frequency in general, commuting included.

results were not very sensitive to variations of this value from 0.1 miles to 1 mile). The largest inter-individual distance between the residences of two individuals in our sample is about 99 miles.

Earlier studies have identified several groups of variables that may affect bicycling frequency. Heinen *et al*. (2010) and Pucher *et al*. (2010) provide a good literature review of bicycling behavior determinants, and acknowledge that, in good part, the reason for different variables being used in different studies is purely based on the practical consideration of data availability. Further, in many cases, statistical problems of multicollinearity are also encountered in including multiple variables capturing similar effects, such as income and car ownership in the case of demographic variables, or residential density, bicycle facility mileage in the residential neighborhood, and street connectivity in the case of built-environment measures. In addition to the (individual and household) demographic variables and residential neighborhood attributes identified above, which are two sets of variables almost always included in empirical studies of bicycling, socio-ecological and socio-cognitive theoretical models (see Sallis *et al*., 2002 and Ma *et al*., 2014) suggest that bicycling behavior should be impacted by a third set of such subjective factors as individual attitudes and perceptions. These theoretical models also suggest that the impact of objective built environment measures should be moderated by subjective factors as well as demographic factors. The underlying notion is that individual factors such as gender, place attachment, local culture, societal norms, safety attitudes, and physical capability color the lens through which individuals sense and perceive the objective built environment, leading to potentially different mental maps across individuals experiencing the same objective environment (see, for example, Gebel *et al*., 2009, Lackey and Kaczynski, 2009, Van Acker *et al*., 2013, and Maldonado-Hinarejos *et al*., 2014). A fourth set of variables correspond to the natural environment, including seasonality/climate, terrain grades, and daylight length times (see, for example, Stinson and Bhat, 2004, Parkin *et al*., 2008, and Gatersleben and Appleton, 2007). A fifth set of variables used in some earlier commute mode choice studies including bicycling correspond to cycling infrastructure considerations (such as travel time to work and presence of showering facilities at work). Another set of variables that are occasionally used refer to bicycling orientation and experience, with some studies developing cyclist typologies (such as dedicated or diehard bicyclist and experienced bicyclist) as a precursor to modeling frequency within each cyclist typology (see, for example, Hunt and Abraham, 2007 and Damant-Sirois *et al*., 2014). A problem though with doing so is that there is an endogeneity problem given the typologies themselves are a function of bicycling frequency.

Clearly, there are many sets of variables that have been used in the literature, though, due to data limitations, it is literally impossible for any single study to consider all the variables at the same time. For instance, in the current analysis, there is no information on subjective attitudes and preferences related to bicycling behavior, and the narrow time window of survey data collection precludes the inclusion of natural environment variables. Further, as with many earlier studies (such as Sallis *et al*., 2013 and Ma *et al*., 2014), we focus on bicycling frequency in general for all purposes and frequency over a longer period of time than a single day rather than daily commuting mode choice, and so do not include cycling infrastructure variables. Thus, we include the "staple" variables corresponding to residential neighborhood and demographic variables, along with

interactions of the two sets of variables to accommodate the potential different mental maps across individuals of the same objective environment based on observed demographics. In the category of residential neighborhood, we construct two variables based on GIS data available at the PSRC web site and the location of the household of each individual. The first variable characterizes residential density, computed as the density or number of households per square mile in the Census block group of the household's residence, as obtained from the 2010 decennial Census data. The second represents a measure of the intensity of bicycling facilities available, computed as the total length of bicycle lanes (in miles) within a one-mile radius of the centroid of the census block group of the individual's household (for presentation ease, we will henceforth refer to this variable simply as "intensity of bicycle lane infrastructure" or IBLI; note that a bicycle lane, as considered in this study, can be a striped bicycle lane, a protected bicycle lane, a marked shared lane, or a separated bicycle lane).

Overall, while our empirical model is by no means comprehensive in the consideration of bicycling frequency determinants, it exploits the information available in the data used here. Most importantly, we explicitly consider social dependence effects using a spatial lag formulation, which, to our knowledge, is a first in the bicycling literature.[12]

4.4 Estimation Process

The selection of variables in the final specification was based on previous research (as discussed earlier), intuitiveness, and statistical testing. Note that, while many earlier studies (such as Sallis *et al*., 2013 and Ma *et al.,* 2014), like our study, consider the frequency variable in ordinal discrete categories (such as never bicycle to bicycle everyday), they have assumed the variable as a continuous dependent variable, which is fundamentally inappropriate from an econometric perspective. Further, while some other studies (such as Sener *et al*., 2009a, and Noland *et al*., 2011) have used a standard ordered response (SOR) formulation, we improve upon this restrictive formulation by generalizing it and allowing the thresholds themselves to be functions of exogenous variables through the use of a generalized ordered response (GOR) formulation. This allows the exogenous variables to have a less restrictive monotonic and even quite

---

[12]As discussed at the end of Section 3.1, an issue is whether what is being captured through the spatial lag formulation is a social dependence effect or a manifestation of other model mis-specifications. Of course, one can never be sure that things have been appropriately disentangled in any empirical application with purely choice outcome data. In the empirical analysis of this paper, we pre-specify the elements of $\mathbf{W}$ to be a fixed decreasing function of a single exogenous variable (distance between the residences of individuals), and assume that these represent social dependence effects due to peer interactions. This is standard in much of the transportation literature to acknowledge that the home-end generally tends to be the hub of socialization and interaction. However, our framework is extendable to include more general forms of spatial and social dependence. For example, $\mathbf{W}$ itself can be parameterized as a finite mixture of several weight matrices, each weight matrix being related to a specific proximity measure k: $\mathbf{W} = \sum_{k=1}^{K} \varphi_k \mathbf{W}_k$, where φ_k is the weight on the kth proximity variable in determining dependency between individuals ($\sum_{k=1}^{K} \varphi_k = 1$), and $\mathbf{W}_k$ is a matrix with its elements representing a measure of distance between individuals on the kth covariate (see Yang and Allenby, 2003).

different direction of effects on each ordinal category of bicycling frequency than the SOR formulation (see Eluru *et al*., 2008). As indicated in Section 3.2, if no exogenous variables affect the thresholds, the GOR collapses to the SOR formulation. To our knowledge, this is the first use of the GOR formulation in the bicycling frequency literature, in addition to the inclusion of social interdependence while also including a skew-normal error term for the kernel error term.

Our progression in terms of estimation began with developing the best specification in the context of a simple aspatial GOR formulation. Then, we retained this specification to explore alternative spatial weight specifications. Specifically, we considered several functional forms for the spatial weight matrix (**W**) based on the inverse of inter-individual distance to accommodate fading social dependence based on decreasing spatial proximity of individuals' residential locations, including inverse distance, inverse distance squared, inverse distance cubed, and inverse of exponential distance. The best weight configuration is chosen based on a composite likelihood information criterion (CLIC) statistic. The weight configuration that provides the highest value of the CLIC statistic is the preferred one (see Bhat, 2011, 2014). In our analysis, this came out to be the inverse of the cube of distance, which was retained in all subsequent specifications. The spatial specification above considered all pairs of individuals. We next explored alternative distance bands (2 miles, 5 miles, 7.5 miles, 10 miles, 20 miles, 50 miles, and 100 miles, the last of which corresponds to considering all pairs of individuals) to select the distance band to include for the pairing of individuals in the composite marginal likelihood (CML) estimation. This is because, as discussed in detail in Bhat, 2011 and Bhat, 2014, a higher efficiency of the CML estimator can be achieved by lowering the number of pairings used in the CML estimation. The best estimator efficiency, based on minimizing the trace of the asymptotic covariance matrix (see Bhat, 2014), was obtained with a distance band of 7.5 miles, which was then retained in subsequent estimations. Using the spatial specifications just discussed, we once again explored several alternative variable specifications, but the original variable specification without the spatial dependence continued to be the best specification for the spatial GOR (or S-GOR) model. Finally, we added skew-normality in the kernel error term. To help the estimation process, we fixed the skew parameter at successively increasing values in increments of 0.1 (that is, $\rho = 0.1, 0.2, 0.3, ..., 0.9$) to determine that the best value of the skew parameter was obtained at 0.2. Next, the estimation was continued at increments of 0.01 between 0.1 and 0.3, and the best fit was at the value of 0.19. The reason for using this approach to obtain the skew parameter is that convergence of the model with a skew-normal kernel is much faster by fixing the skew parameter. Finally, we estimated the entire spatial skew-normal GOR (or SSN-GOR) model all at once (including the skew parameter and all other parameters) to obtain the final results for the SSN-GOR model. Doing so provides the t-statistics for the skew parameter too.

4.5 Estimation Results

Table 1 present the estimation results. As discussed in Section 3.1, for reasons of identification, there is no constant in the latent bicycling propensity (second column in Table 1). Also, there are no exogenous variables identified for the threshold between the first and second ordinal categories of "not a bicyclist" and "occasional bicyclist". That is,

there is only a constant identified in the first threshold (*i.e.*, $\psi_{q1} = \exp(\alpha_1)$), and this constant is listed in the first row of Table 1.

There are four main columns in Table 1, in addition to the first "variables" column. The first numeric column corresponds to the estimates of the $\boldsymbol{b}$ vector elements that characterize the latent bicycling propensity (with no constant). The second column corresponds to ψ_{q2}, and the estimates presented are the λ_2 constant, and the $\boldsymbol{\varphi}_2$ parameters corresponding to the second threshold demarcating the "occasional" and "infrequent" bicycling frequency categories. The final two columns correspond to λ_3 and $\boldsymbol{\varphi}_3$ parameters for the ψ_{q3} threshold (threshold between "infrequent" and "frequent" bicycling frequency categories), and λ_4 and $\boldsymbol{\varphi}_4$ parameters for the ψ_{q4} threshold (threshold between "frequent" and "everyday" bicycling frequency categories).

The effect of each category of variables on the latent injury risk propensity and the three thresholds are discussed in the next section, followed by the spatial dependency parameter and the skew parameter. We should note that, for continuous variables such as age, income, and the IBLI variable, we attempted a variety of functional forms, including the continuous value, piecewise linear functions, and a dummy variables specification (with many different combinations of the cut points for the piecewise linear and dummy variable specifications). At the end, the dummy variables specification came out to be the best for the age and income variables, with the boundary cut-off points as in Table 1. For the IBLI variable, the continuous value came out to be the best specification in general, though a cut-off representation of this variable was superior for interactions of this variable with demographic variables.

4.5.1 Variable effects

The individual characteristics that impact bicycling propensity are gender, educational attainment, age, and work status. According to our results, men typically have a higher propensity to bicycle than women, as several previous studies have found (Parkin *et al.*, 2008, Sener *et al.*, 2009a, Ferdous *et al.*, 2011, Sallis *et al.*, 2013, Ma *et al.*, 2014). Heinen *et al.* (2010) indicate that in countries where the bicycling share is very low (such as the United States), studies have almost universally found that men bicycle more often than women. However, studies in other countries where there is a high share of bicycling (such as Netherlands and Belgium) have typically not found any statistically different effects between men and women. This is likely to be the result of bicycling being viewed as a "risky" proposition in terms of safety in the countries where the bicycling share is very low. In this context, the study by Bhat *et al.* (2015) suggests that the difference between men and women in the United States in bicycling orientation may be related to safety-consciousness attitudes. Based on their results that women are more safety-conscious than men in the context of bicycling safety from traffic crashes, an issue that has consistently been identified as a barrier to bicycling in U.S. cities, Bhat *et al.* (2015) reach into the psychological literature to offer three explanations for the gender difference in risk-taking (which is on the reverse scale of safety-consciousness). The first is based on the notion of "risk as feelings" in which our instinctive and intuitive emotions dominate reasoned approaches when faced with risk. In this regard, given that women

experience feelings of nervousness and fear more than men in anticipation of negative outcomes, the net result is a heightened risk-averseness (or higher safety consciousness) among women. The second is based on the notion of confidence; many psychological studies (see, for example, Niederle and Vesterlund, 2007) indicate that men tend to be more overconfident in uncertain situations, translating to more risk-taking (and less safety-consciousness) in men than women. The third explanation is tied to the notion of believed appropriate response, with men tending to view a risky situation as a challenge that warrants participation, while women tend to view risky situations as threats that must be avoided.

The Table 1 results for the effect of education level on bicycling indicate that highly educated individuals have a higher propensity to bicycle (see the positive coefficient of 0.190 in the column labeled "latent bicycling propensity" corresponding the "bachelor's degree or higher education level" variable; we attempted a more extensive multiple dummy variable specification for introducing the effect of education, but found no differences between the effects of "high school or less" and "some college" education categories, as well as between the "bachelor's degree" and "graduate degree" education categories). The result is not surprising because a high education level is typically associated with a high level of environmental awareness (see McCright, 2010, Lo and Jim, 2012, and Paleti *et al*., 2013) and, consequently, a higher tendency to use environmental-friendly modes of transportation such as bicycling. However, this may also be because individuals with higher levels of education tend to be more aware of traffic safety rules and regulations, and have a more objective and less negative perspective of safety from traffic crashes. In addition to the effect on the bicycling propensity, the education variable also impacts the thresholds in the framework of the GORL model. As should be clear from Equation (8), a negative coefficient in the vector φ_k for a variable has the effect of moving the corresponding threshold to the left. The net effect for the probability of each ordinal category will depend on the entire sequence of threshold effects. In the case of the education variable, the pattern of threshold effects in the flexible GOR model indicates that higher educated individuals, compared to lower educated individuals, will be substantially more likely to be frequent or everyday bicyclists than what would be predicted by the standard ordered-response model. In Section 4.5.3, we compute elasticity effects to better represent the composite effects of variables on each ordinal bicycling category.

Individuals in the intermediate age range (between 35 and 54 years old) have a higher propensity to bicycle than those in other age groups. The effects on the thresholds indicate that, relative to a standard ordered-response model, the GOR model predicts a lower probability of such individuals to be occasional bicyclists and a higher probability of such individuals to be everyday bicyclists. Overall, individuals in the middle age group are more bicycle prone (in terms of propensity of use) than individuals in the younger or older age groups. However, as we will discuss later, the difference between individuals in the younger age group of 18-34 years and middle-aged individuals becomes less pronounced as the lane mileage around the household's residence increases). The lower propensity of older individuals (>54 years of age) to bicycle, which has been found in many earlier studies too (see, for example, Kemperman and Timmerman, 2009, and Ma

et al., 2014), is likely explained by a worry about slower reflexes and recovering from bicycle-related crash injuries as individuals reach their 50s and beyond.

Finally, full-time workers are less inclined to bicycle frequently than other non-full time workers, another common finding in the literature that is attributed to longer commute distances and less free time to engage in recreational bicycling for full-time workers relative to those not employed full-time (see Heinen *et al*., 2010, Sener *et al*., 2009a, and Ferdous *et al*., 2011).

The household demographics that play an important role in bicycling frequency decisions are number of automobiles, presence of children, and annual household income. Individuals living in households with more automobiles have a lower bicycling propensity. The effect on the threshold shows that, compared to the standard ordered-response model, the GOR model predicts an even higher probability of individuals with a high number of autos never bicycling or bicycling only occasionally. Of course, it is possible that this car ownership effect is not a causal one, but really an association effect because of, for example, individuals with a green lifestyle orientation owning fewer cars and bicycling often (see, for example, Bhat *et al*., 2015). The presence of children in the household also reduces bicycling propensity, presumably because of time poverty effects (see Bernardo *et al*., 2015) that precludes participation in relaxing activities such as recreational bicycling. Besides, earlier studies in psychology (see, for example, Turner and McClure, 2003 and Dohmen *et al*., 2011) suggest that humans tend to become less adventurous and more risk-averse when a child is present in the household, which perhaps gets manifested in the form of less bicycling to reduce exposure to perceived unsafe travel conditions. Finally, in terms of household attributes, individuals from progressively lower income households have a lower propensity of bicycle use relative to individuals from progressively higher income households. Although this result has been confirmed by many other studies (Stinson and Bhat, 2004, Dill and Voros, 2007, Parkin *et al*., 2008), there is no clear consensus regarding the role of income in the frequency of bicycling. Bicycling is not a cost-prohibitive mode of transportation as the car is, so it does not represent an economic barrier for potential riders. However, low income households tend to locate in neighborhoods that are far from the work place, making commute bicycling more difficult and perhaps leading to less time for recreational bicycling too. Another explanation is that low household income is a proxy for a variety of household residential neighborhood crime-related considerations, and/or blatant discrimination in bicycling facility investments in minority and/or low income areas (see, for example, Rietveld and Daniel, 2004).

Among the two neighborhood variables considered; residential density and the IBLI variable; we discerned clear multicollinearity. In particular, when we introduced each variable separately, each had a statistically significant impact, with the IBLI variable specification providing a far superior data fit. Further, when we introduced both variables, the residential density variable turned out to be statistically insignificant. Thus, we retained only the IBLI variable and its interactions with demographic variables. As Table 1 indicates, the more the intensity of bicycle lanes, the higher in general is the

propensity of bicycle use.[13] However, there are interactions of this variable (in a dummy variable form based on distance thresholds) with the "male" and "individuals aged 18-34 years" (*i.e.*, young individuals) (Technically, the discussions of the male and age variables discussed earlier should have been done in combination with these interaction terms, but the effects of these demographic variables as discussed earlier holds regardless of the intensity of bicycle lanes, as we will now discuss; thus, to streamline the presentation, we adopt this layered presentation scheme). The interaction with the male dummy variable suggests that while men generically have a higher propensity to bicycle, the difference in propensity between men and women reduces from 0.311 (the coefficient on the male dummy variable in Table 1) in residential locations where the mileage of bicycle lanes is less than four miles to 0.085 (0.311-0.226; the difference between the coefficient of the male dummy variable and the male dummy variable interaction with the intensity of bicycle lanes) in residential locations where the mileage of bicycle lanes is more than four miles. As importantly, in locations where the mileage of bicycle lanes is more than four miles, the difference between men and women becomes statistically insignificant (the standard error on the coefficient of 0.085 is 0.149, with a t-statistic of 0.57 for testing the null hypothesis that the true difference between men and women in bicycling propensity in residential locations with more than four miles of bicycling lanes is zero). This result is consistent with the notion of bicycling being viewed as a "risky" proposition in terms of safety when bicycling facilities are limited (and bicycling share is low) and women being more sensitive to risk (see earlier discussion). But, when IBLI increases, bicycling is viewed less as a risky proposition, thus reducing the differences between men and women in their bicycling propensity (as has been found in many European cities). Further, the interaction of the IBLI variable with the young age group in Table 1 suggests that, in residential locations with more than three miles of bicycle lanes, young individuals (18-34 years of age) have a statistically significantly higher propensity of bicycle use than the oldest category of individuals (55 years or older) (while there is no statistically significant difference between the youngest and oldest age groups when the intensity of bicycle lanes is less than three miles). However, even in locations with more than three miles of bicycle lanes, middle-aged individuals (aged 35-54 years) have a marginally significant higher bicycling propensity than the youngest age group (the difference in propensity is 0.136 (=0.320-0.184), with a standard error of 0.125 and a t-statistic of zero difference being 1.1). Overall, the reluctance of younger individuals to bicycle relative to middle-aged individuals in locations with a low intensity of bicycle lanes, and the less pronounced difference between these two age groups in locations with a high intensity of bicycle lanes (>3 miles), needs further exploration in future studies. However, one important implication from the interaction effects just discussed is that a

---

[13]A caveat about the effects of the mileage of bicycle lanes is in order here. In this analysis, we use the miles of bicycle lanes in the household residence neighborhood of the individual as an independent variable. In fact, this may be an endogenous variable in the sense that people who are bicycle use-prone may locate themselves in areas with good bicycling facilities. But, as in all earlier bicycling frequency studies we are aware of, we do not consider this residential self-selection issue here. The very few studies that consider residential self-selection in the effects of built environment measures on bicycling-related travel have done so in the context of commuting mode choice to work or non-motorized (bicycling and walking together) tour making (see, for example, Pinjari *et al*., 2011 and Bhat, 2014). We leave the consideration of residential self-selection in examining the impact of built environment measures on bicycling frequency to future studies, and focus on the spatial dependence issue here.

more inclusive and diverse set of individuals (men and women, and young and middle-aged individuals) are likely to take up bicycling as more and more resources are invested in bicycling facilities.

4.5.2 Spatial dependency and kernel error skewness

The spatial dependency parameter is moderate in magnitude (0.152) but highly statistically significant, supporting the hypothesis that bicycling propensity of individuals in households located in close proximity of each other are indeed positively correlated. That is, our results reinforce the theoretical notion in the bicycling use literature of the presence of social group influences in bicycling use propensity.

The skew parameter came out to be 0.190, and is statistically significant. Thus, the hypothesis of the kernel error term being normal (as has been the norm in all earlier bicycling frequency studies) is soundly rejected. The implied shape of the marginal skew-normal density function for the error term is provided in Figure 1. We have included, as reference, the shape of the marginal standard normal density function (red line in Figure 1). Since the skew parameter is positive (even if small in magnitude), there is a slight right skew in the density function of the kernel error term. That is, relative to the normal distribution, the effects of unobserved characteristics make a larger fraction of individuals more inclined to bicycle often. The difference in shape between the implied skew-normal distribution and the normal distribution may appear rather minimal, but, as we will note in Section 4.7, the difference is enough to make substantial differences in the impacts of specific policy actions. In particular, using the normal distribution when the skew-normal is the appropriate distribution leads to an underestimation in the benefits of bicycling facility improvements. This is particularly so when the skew-normal distribution is combined with spatial dependence effects, because consideration of spatial dependence leads to a "spillover" or "multiplier" effect at the overall community level that accentuates the benefits of bicycling facility improvements at the individual level. Thus, ignoring spatial dependence as well as ignoring the positive skew leads to a substantial underestimation of the benefits of bicycling facility improvements.

4.5.3 Elasticity effects

As mentioned earlier, the coefficients in Table 1 do not directly provide a sense of the magnitude and direction of effects of each variable on each bicycling frequency category. But one can compute aggregate-level effects to characterize the overall impacts of each variable. To do so, we compute the aggregate-level "pseudo-elasticity effects" of exogenous variables. In the current analysis, we have three types of exogenous variables: discrete variables, a count variable, and a continuous variable. The discrete variables include binary variables (the male dummy variable, the "bachelor's degree or higher education level" dummy variable, the full-time work status dummy variable, and the "presence of children" dummy variable) and multinomial variables (age categorized in three groups: 18-34 years, 35-54 years, and >54 years; and annual household income in four categories: <25,000, 25,000-34,999, 35,000-50,000, and >50,000). But because the binary variables are simply one type of a multinomial variable, we will discuss the methodology we use to compute elasticity effects only for one of the multinomial variables (the age variable) within the class of discrete variables. For these variables, we

first predict the probabilities of each bicycling frequency level for each individual, assigning the base value of "0" for all dummy variables characterizing the multinomial exogenous discrete variable (that is, assigning zero values for all the three age categories). All other exogenous variables are at their values in the original data. The procedure to compute the individual-level probabilities in our spatial model (for one realization of the parameters from their sampling distributions) is similar to the one discussed in detail by Castro *et al*. (2013), with the important difference that the kernel error terms are generated using a skew-normal distribution in the current paper as opposed to a simple normal distribution in Castro *et al*. (2013).[14] Then, the individual-level probabilities are added to obtain the expected value of the number of individuals at each bicycling frequency level in the base case (label the resulting vector of five values, one each for each of the five frequency levels, in this base case as BASE). Subsequently, the same procedure as above is undertaken but after changing the value of the "35-54 years" dummy variable for each individual from the value of zero to the value of one, and obtaining the expected value of the number of individuals at each frequency level in this new case (label the resulting vector of five values as AGE3554). Next, the same procedure as above is implemented, but now starting with the base data and changing the value of the ">54 years" dummy variable for each individual from the value of zero to the value of one (label the resulting vector of five expected values as AGEGT54). Subsequently, to obtain an aggregate-level elasticity of the "Age – 35-54 years" dummy variable, we compute the change between the AGE3554 and BASE vectors as a percentage of the BASE vector, yielding five elasticity values (one for each frequency level). Similarly, to obtain an aggregate-level elasticity of the "Age > 54 years" dummy variable, we compute the change between the AGEGT54 and BASE vectors as a percentage of the BASE vector, once again yielding five elasticity values. Finally, we compute the mean and standard errors of the aggregate-level elasticity effects as computed above across 200 bootstrap draws (realizations) taken from the sampling distributions of the parameters. These are available from the authors, but are suppressed from Table 2 to avoid clutter.

For the one count variable (number of vehicles), the procedure is simpler. We simply change the count variable (number of vehicles) for each individual by the value of one, and compute the percentage change in the expected number of individuals at each bicycling frequency level. For the one continuous variable (the IBLI variable), we increase the value of the variable by 10% for each observation.[15]

---

[14]Essentially, the procedure entails, for one realization (say realization A) of the parameters from their sampling distributions, the generation of one realization of the latent vector y^* in Equation (12). This is then translated into an ordinal category assignment for each individual based on the threshold values for each individual (these thresholds themselves are computed from Equation (8) for the one realization A of the parameters). The above procedure is repeated N times to obtain N realizations of the vector y^* and N corresponding ordinal category assignments for the one realization A of the parameters. The ordinal category assignments (converted to 0/1 dummy variables of choosing each ordinal category) across the N repetitions are averaged to obtain probabilities for each ordinal category for each individual for the realization A of the parameters.

[15]We can also compute pseudo-elasticity effects for interaction variables. For example, to compute the elasticity effect of the male variable, we can develop separate estimates for the population of individuals in residential areas with less than or equal to four miles of bicycling lanes and the population of individuals in residential areas with

Table 2 provides the pseudo-elasticity effects. The numbers in the table may be interpreted as the percentage change in the probability of each bicycling frequency level due to a change in the exogenous variable. For example, the first entry in the table indicates that the probability of a man not being a bicyclist is 13.6% lower than the probability of a woman not being a bicyclist, everything else being equal (that is, there will be about 14 less non-bicyclist individuals in a random sample of 100 men compared to in a random sample of 100 women). For the count variable of "number of vehicles", the entry in the first column indicates that an additional vehicle in the household increases the probability of an individual in the household not being a bicyclist by 30.1% (that is, there will be about 30 less non-bicyclists in a random sample of 100 households with one less car than another random sample of 100 households). For the continuous IBLI variable, the probability of an individual not being a bicyclist reduces by 14.9% if the mileage of bicycle lanes in the individual's residential neighborhood is increased by 10%. The directions of the elasticity effects of the model are consistent with the discussions in the previous section. The table suggests that the most important variables affecting bicycling frequency are (1) education level (if the individual has a bachelor's degree or higher), (2) work status (if the individual is a full-time worker), (3) the number of vehicles in the individual's household, and (4) household income (if individual's household earns less than $25K annually). While the continuous variable effect of IBLI cannot be strictly compared with the other effects, Table 2 indicates that increasing the mileage of bicycle lanes has a substantial positive impact on the percentage of everyday bicyclists. The implication is that investing in bicycling facilities is an effective way to increase the share of those who use bicycling on a regular basis as well draw an inclusive (and diverse) set of demographic groups (in terms of gender and age) to bicycling.

4.6 Measures of Data Fit

In this section, we examine the data fit of the proposed spatial skew-normal GOR (SSN-GOR) models with its more restricted versions: (1) the traditional probit SOR model typically used in bicycling frequency analysis, (2) the aspatial probit GOR (ASP-GOR) model, (3) the spatial probit (SP-GOR) model, and (4) the aspatial skew-normal GOR (ASSN-GOR) model. As discussed in Section 3.2, all of these four models are restrictive versions of the SSN-GOR model, and may be tested against the SSN-GOR model using the adjusted composite likelihood ratio test (ADCLRT) statistic, which is asymptotically chi-squared distributed similar to the likelihood ratio test statistic for the maximum likelihood approach. The reader is referred to Bhat (2011) for details regarding the ADCLRT test statistic. In addition to testing the models using the ADCLRT statistic, we also compute the average probability of correct prediction of each model using the procedure discussed in Castro *et al.,* 2013. Finally, we also predict the shares for each bicycling frequency category with each model and compare the predicted shares with the actual sample shares using the Mean Absolute Percentage Error (MAPE) measure.

---

more than four miles of bicycling lanes. But, to keep the presentation simple, we focus on only the overall effects in the entire population of individuals for each individual variable.

The results of the data fit comparisons are presented in Table 3. For completeness, we note that the composite log-likelihood (CLL) value of the naïve model with only a constant in the underlying bicycling propensity and only constants in the thresholds (equivalent to a sample shares model) is -4822776.3 (the second numeric row). The number of parameters and the CLL values for the five models (from the simplest SOR model to the proposed general SSN-GOR model) are provided in the third and fourth numeric rows, respectively, of Table 3. All the models in the table reject the naïve model based on ADCLRT tests. The fifth row compares the SSN-GOR model with its restrictive versions using ADCLRT tests, and indicates the clear superior performance of the SSN-GOR model relative to other models. While not shown in the table, an ADCLRT of the SOR and ASP-GOR models (the first and second models in Table 3) clearly rejects the SOR model in favor of the ASP-GOR model, supporting the view that the traditional SOR formulations in the bicycling frequency literature are restrictive and likely to be mis-specified. Also, both the SP-GOR and the ASSN-GOR models reject the ASP-GOR model, supporting the notion that spatial interactions and a non-normal kernel distribution underlie bicycling propensity. This should also be obvious from the statistically significant spatial autoregressive parameter (in the SP-GOR model; see the sixth numeric row) and the statistically significant skew parameter (in the ASSN-GOR model; see the seventh numeric row). The result is further reinforced by the average probability of correct prediction statistics listed in the eighth numeric row of the table. Between the non-nested SP-GOR and ASSN-GOR models, which may be compared using the composite likelihood information criterion (CLIC) (Varin and Vidoni, 2005, Bhat, 2014), the SP-GOR model is superior to the ASSN-GOR model. However, from an aggregate prediction (MAPE) perspective (see the last row panel of Table 3 that provides the predicted percentages of each bicycling frequency level as well as the actual sample percentages, and the MAPE statistic in the last row), the ASSN-GOR model does better than the SP-GOR model. This suggests that the best model should accommodate both spatial dependence and a non-normal kernel error term, as does the SSN-GOR model proposed in this paper. Indeed, the SSN-GOR model rejects all other models based on not only the disaggregate ADCLRT test and the probability of correct prediction statistic (as discussed earlier), but also based on the aggregate MAPE statistic.

Another interesting observation from Table 3 is that the aggregate bicycling frequency predictions for the high frequency categories ("frequent bicyclist" and "everyday bicyclist") are under-projected when spatial dependence is ignored (see the predictions from the SP-GOR and ASP-GOR models) as well as when the positive skewness of the kernel error term is ignored (see the predictions from the ASSN-GOR and ASP-GOR models). The first is because of ignoring the multiplier effect, and the second is because unobserved factors tend to make individuals have a higher bicycling propensity than that specified by a normal distribution. Further, the consequence of incorrectly imposing a normal distribution is particularly severe in the presence of spatial interaction. This is evidenced in the MAPE percentage increase of 56% from 5.22% to 8.15% between the ASSN-GOR to ASP-GOR models, but an even larger percentage increase of 104% from 3.54% to 7.21%, between the SSN-GOR and the SP-GOR models, again because of the multiplier effect when spatial dependence is considered.

4.7 Policy Implications

As discussed in the previous sub-section, the SSN-GOR model provides a better data fit than other models based on both disaggregate composite likelihood ratio tests as well as aggregate sample fit predictions. To further demonstrate the potential policy pitfalls of using the inaccurate models, we examine the predicted changes in the aggregate shares of bicycling frequency levels due to a bicycling infrastructure investment program that increases the "intensity of bicycle lanes" by 10% for each individual. To streamline and simplify the presentation, we combine the two lowest frequency categories (not a bicyclist and occasional bicyclist) in a single category labeled as "little to no bicycling" (LNB), and the two highest frequency categories (frequent and everyday bicyclist) in another single category labeled as "regular bicycling" (RB). Further, because the middle category of "infrequent bicycling" is a sandwich category that is essentially determined by the two extreme categories, we will focus only on the two extreme categories of LNB and RB. The pseudo-elasticities for these two bicycling frequency categories are presented in Table 4. As expected, all the five models present the same trend: the LNB category share reduces and the RB category share increases. However, the restricted models clearly underestimate the reduction in the LNB category share as well as the increase in the RB category share. The P-values for the comparison of the elasticities with the proposed SSN-GOR model show that the under-estimations from the SOR and ASP-GOR models are statistically significant at the 4% significance level or lower for the LNB category, though not so for the SP-GOR and ASSN-GOR models for the LNB category. But the under-estimations from all the restrictive models are statistically significant at any reasonable level for the RB category. Finally, among the many models, the pattern of elasticities mimics the pattern of the predicted shares discussed in the previous section, with clear evidence of the multiplier effect combined with the positive skew effect as represented in the SSN-GOR model. Of particular note is that even the two closest models (SP-GOR and ASSN-GOR) to the proposed SSN-GOR model underestimate the increase in the RB category by as much as about one third of the total increase.

Overall, ignoring spatial interactions and/or non-normal kernel error terms will generally lead to both inconsistent estimates of the choice probabilities as well as the effects of exogenous variables. As a consequence, there is a very real possibility of mis-informed policy-making if these important model specification issues are ignored. In our specific empirical context, the results indicate that improvements in bicycling facilities would be quite substantially (and incorrectly) under-estimated, potentially leading to a rejection of bicycling infrastructure improvements when actually warranted, due to an inappropriate low-balling of benefits relative to costs.

Chapter 5. Conclusions

This paper has proposed a new spatial skew normal generalized ordered response (or SSN-GOR) model and an associated estimation method. It contributes to the spatial analysis field by allowing a non-normal (though parametric) kernel error term in traditional specifications of the spatial model. As discussed in the paper, while the use of an incorrect kernel distribution in aspatial models will, in general, lead to inconsistent estimates of the choice probabilities as well as the effects of exogenous variables, the situation gets exacerbated in spatial models because of the multiplier effect. To our knowledge, this is the first spatial non-normal GOR model proposed in the economic literature. The skew normal distribution that we use for introducing non-normality is tractable, parsimonious in parameters that regulate the distribution and its skewness, includes the normal distribution as a special interior point case, and is flexible because it allows a continuity of shapes from normality to non-normality, including skews to the left or right. We have implemented Bhat's (2011) maximum approximate composite marginal likelihood (MACML) inference approach for estimation of the SSN-GOR model.

The paper demonstrates the application of the proposed model through an analysis of bicycling frequency among workers of the Puget Sound Region in the U.S. state of Washington. The sample is drawn from the 2014 Puget Sound household travel survey and includes a variety of socio-demographic characteristics of the individuals and household demographics, and a couple of indicators serving as proxies for bicycle facilities at the residential location. Our results underscore the important effects of demographic variables on bicycling propensity, especially education level, full-time work status, the number of vehicles in the individual's household, and household income. Unlike many earlier bicycling studies, we draw from the social-psychological concepts and theories to provide possible explanations for these demographic effects. The results also indicate the positive effects of increasing the miles of bicycle lanes in residential neighborhoods. As importantly, the results suggest that women and young individuals (18-34 years of age) are much more likely to "warm up" to bicycling as more investment is made in bicycling infrastructure, thus leading not only to a larger pool of bicyclists but also a more diverse and inclusive one. This is an interesting result that suggests that information campaigns on bicycling improvements may benefit from targeting women and young individuals. We have also demonstrated that introducing social dependence effects and non-normal kernel error terms have real policy implications; ignoring these effects can underestimate the impacts of bicycling infrastructure improvements and public campaigns on bicycle use frequency, potentially leading to under-investments in bicycling infrastructure projects. Also, the fact that our study found positive multiplier effects implies that information and promotion campaigns to foster frequent bicycle use (either as part of congestion mitigation efforts or sustainable transportation policies or energy consumption reduction strategies or public health improvement considerations) would do well to bring individuals from disparate neighborhoods together (rather than "preaching" within neighborhoods). Doing so would exploit the neighborhood-based virtuous snowballing cycle through local social interactions when each individual goes back to her or his residential community.

An important direction for future methodological research would be to accommodate potential residential self-selection in built environment effects (such as bicycling infrastructure investments within neighborhoods). This would help disentangle associative effects from "true" built environment effects. Another methodological extension would be to introduce the proposed formulation within an integrated choice latent variable system, so that subjective attitudes and perceptions can be included along with the social interactions and general kernel error distribution of the proposed formulation (this would entail combining Bhat *et al*.'s (2016) formulation with the current one). A third methodological extension would be to develop a longitudinal component within the current framework to accommodate dynamics in bicycling behavior and help tease out better the causal impacts of exogenous variables. Another extension would be to introduce a spatial dependence process across individuals in the thresholds of the GOR model. In terms of empirical extensions, the current paper uses a general travel survey data set that has good spatial resolution of residential locations to develop relatively accurate variables of residential density and bicycle lane mileage, as well as generate inter-individual distances for proximity-based social interactions. Future studies can introduce a more comprehensive set of exogenous variables, which may be captured through bicycling-oriented travel surveys (see, for example, Sener *et al*., 2009a,b), to accommodate the effects of additional neighborhood built environment attributes, subjective factors (individual attitudes and perceptions), and natural environment characteristics.

References

ACEA (European Automobile Manufacturers Association), 2013. Automobile industry pocket guide 2013. Available at: http://www.acea.be/images/uploads/files/POCKET_GUIDE_13.pdf

Alamá-Sabater, L., Artal-Tur, A., Navarro-Azorín, J.M., 2011. Industrial location, spatial discrete choice models and the need to account for neighbourhood effects. *The Annals of Regional Science* 47(2), 393-418.

Anselin, L., 2010. Thirty years of spatial econometrics. *Papers in Regional Science* 89(1), 3-25.

Arbia, G., 2014. *A Primer for Spatial Econometrics*. Palgrave MacMillan, Basingstoke.

Arellano-Valle, R.B., Azzalini, A., 2006. On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* 33(3), 561-574.

Azzalini, A., 2005. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* 32(2), 159-188.

Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B* 61(3), 579-602.

Azzalini, A., Dalla Valle, A., 1996. The multivariate skew-normal distribution. *Biometrika* 83(4), 715-726.

Beck, N., Gleditsch, K.S., Beardsley, K., 2006. Space is more than geography: Using spatial econometrics in the study of political economy. *International Studies Quarterly* 50(1), 27-44.

Bernardo, C., Paleti, R., Hoklas, M., Bhat, C.R., 2015. An empirical investigation into the time-use and activity patterns of dual-earner couples with and without young children. *Transportation Research Part A* 76, 71-91.

Berry, S.T., Haile, P.A., 2010. Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Cowles Foundation Discussion Paper 1718, Cowles Foundation, Yale University. Available at: http://cowles.econ.yale.edu/P/cd/d17a/d1718.pdf.

Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B* 35(7), 677-693.

Bhat, C.R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled halton sequences. *Transportation Research Part B* 37(9), 837-855.

Bhat, C.R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B* 45(7), 923-939.

Bhat, C.R., 2014. The composite marginal likelihood (CML) inference approach with applications to discrete and mixed dependent variable models. *Foundations and Trends in Econometrics* 7(1), 1-117.

Bhat, C.R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B* 46(7), 817-833.

Bhat, C.R., Sener, I.N., Eluru, N., 2010. A flexible spatially dependent discrete choice model: Formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B* 44(8-9), 903-921.

Bhat, C.R., Dubey, S.K., Nagel, K., 2015. Introducing non-normality of latent psychological constructs in choice modeling with an application to bicyclist route choice. *Transportation Research Part B* 78, 341-363.

Bhat, C.R., Pinjari, A.R., Dubey, S.K., Hamdi, A., 2016. On accommodating spatial interactions in a generalized heterogeneous data model (GHDM) of mixed types of dependent variables. *Transportation Research Part B*, forthcoming.

Bocci, C., Rocco, E., 2016. Modelling the location decisions of manufacturing firms with a spatial point process approach. *Journal of Applied Statistics* 43(7), 1226-1239.

Caffo, B., An, M.W., and Rohde, C., 2007. Flexible random intercept models for binary outcomes using mixtures of normals. *Computational Statistics & Data Analysis* 51(11), 5220-5235.

Canary, J.D., Blizzard, L., Barry, R.P., Hosmer, D.W., Quinn, S.J., 2016. Summary goodness-of-fit statistics for binary generalized linear models with noncanonical link functions. *Biometrical Journal* 58(3), 674-690

Castro, M., Paleti, R., Bhat, C.R., 2013. A spatial generalized ordered response model to examine highway crash injury severity. *Accident Analysis and Prevention* 52, 188-203.

COLIBI (Association of the European Bicycle Industry)/COLIPED (Association of the European Two-Wheeler Parts' & Accessories' Industry), 2013. European bicycle market 2013 edition. Available at: http://edoc.difu.de/edoc.php?id=A27P3B9D

Cox, D.R., Reid, N., 2004. A note on pseudolikelihood constructed from marginal densities. *Biometrika* 91(3), 729-737.

Czado, C., Raftery, A., 2006. Choosing the link function and accounting for link uncertainty in generalized linear models using Bayes factors. *Statistical Papers* 47(3), 419-442.

Damant-Sirois, G., Grimsrud, M., El-Geneidy, A.M., 2014. What's your type: A multidimensional cyclist typology. *Transportation* 41(6), 1153-1169.

Dill, J., Voros, K., 2007. Factors affecting bicycling demand. *Transportation Research Record: Journal of the Transportation Research Board* 2031, 9-17.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association* 9(3), 522-550.

Elhorst, J.P., 2010. Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis* 5(1), 9-28.

Elhorst, J.P., Heijnen, P., Samarina, A., Jacobs, J., 2016. Transitions at different moments in time: A spatial probit approach. *Journal of Applied Econometrics*, forthcoming.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis & Prevention* 40(3), 1033-1054.

Elvik, R., 2009. The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis & Prevention* 41(4), 849-855.

Ertiö, T.P., 2015. Participatory apps for urban planning—space for improvement. *Planning Practice & Research* 30(3), 303-321.

Facchini, F., François, A., 2010. A border effect in political mobilization? Territorial dependence and electoral turnout in national election. Working paper, Université de Lille 1, Sciences et Technologie, Villeneuve d'Ascq, France.

Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., Haq, U., 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography* 41, 306-314.

Ferdous, N., Pendyala, R.M., Bhat, C.R., Konduri, K.C., 2011. Modeling the influence of family, social context, and spatial proximity on use of non-motorized transport mode. *Transportation Research Record: Journal of the Transportation Research Board* 2230, 111-120.

Flores-Lagunes, A., Schnier, K.E., 2012. Estimation of sample selection models with spatial dependence. *Journal of Applied Econometrics* 27(2), 173-204.

Franzese, R.J., Hays, J.C., Schaeffer, L.M., 2016. Spatial-and spatiotemporal-autoregressive probit models of interdependent binary outcomes. *Political Science Research and Methods*, 4(1), 151-173.

Fruhwirth-Schnatter, S., 2011a. Logit, probit, or robit – Efficient Bayesian inference using scale-mixture of probit models. Department of Applied Statistics and Econometrics, Johannes Kepler Universitat Linz, Austria.

Fruhwirth-Schnatter, S., 2011b. Dealing with label switching under model uncertainty. In K. Mengersen, C.P. Robert, and D. Titterington (Eds.), *Mixture Estimation and Applications*, Chapter 10, 193-218, Chichester, Wiley.

Gatersleben, B., Appleton, K.M., 2007. Contemplating cycling to work: Attitudes and perceptions in different stages of change. *Transportation Research Part A* 41(4), 302-312.

Gebel, K., Bauman, A., Owen, N., 2009. Correlates of non-concordance between perceived and objective measures of walkability. *Annals of Behavioral Medicine* 37(2), 228-238.

Geweke, J., Keane, M., 1999. Mixture of normals probit models. In C. Hsiao, M.H. Pesaran, K.L. Lahiri, and L.F. Lee (eds.) *Analysis of Panel and Limited Dependent Variables*, 49-78, Cambridge University Press, Cambridge.

Gibbons, S., Overman, H.G., 2012. Mostly pointless spatial econometrics. *Journal of Regional Science* 52(2), 172-191.

Gordon-Larsen, P., Boone-Heinonen, J., Sidney, S., Sternfeld, B., Jacobs, D.R., Lewis, C.E., 2009. Active commuting and cardiovascular disease risk: The CARDIA study. *Archives of Internal Medicine* 169(13), 1216-1223.

Greene, W.H., Hensher, D.A., 2010. *Modeling Ordered Choices: A Primer*. Cambridge University Press, Cambridge.

Hamer, M., Chida, Y., 2008. Active commuting and cardiovascular risk: A meta-analytic review. *Preventive Medicine* 46(1), 9-13.

Heagerty, P.J., Lumley, T., 2000. Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association* 95(449), 197-211.

Heinen, E., Van Wee, B., Maat, K., 2010. Commuting by bicycle: An overview of the literature. *Transport Reviews* 30(1), 59-96.

Horton, D., 2006. Environmentalism and the bicycle. *Environmental Politics* 15(1), 41-58.

Hunt, J.D., Abraham, J.E., 2007. Influences on bicycle use. *Transportation* 34(4), 453-470.

Huy, C., Becker, S., Gomolinsky, U., Klein, T., Thiel, A., 2008. Health, medical risk factors, and bicycle use in everyday life in the over-50 population. *Journal of Aging and Physical Activity* 16(4), 454-464.

Jensen, A., 2013. The power of urban mobility: Shaping experiences, emotions, and selves on a bike. In S. Witzgall, G. Vogl, and S. Kesselring (eds.), *New Mobilities Regimes in Art and Social Sciences*, 273-286, Routledge, London.

Joe, H., Lee, Y., 2009. On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis* 100(4), 670-685.

Keane, M.P., 2010. Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics* 156(1), 3-20.

Kemperman, A., Timmerman, H., 2009. Influences of built environment on walking and cycling by latent segments of aging population. *Transportation Research Record: Journal of the Transportation Research Board* 2134, 1-9.

Klier, T., McMillen, D.P., 2008. Clustering of auto supplier plants in the United States: Generalized method of moments spatial logit for large samples. *Journal of Business and Economic Statistics* 26(4), 460-471.

Lackey, K.J., Kaczynski, A.T., 2009. Correspondence of perceived vs. objective proximity to parks and their relationship to park-based physical activity. *International Journal of Behavioral Nutrition and Physical Activity* 6(1), 53.

League of American Bicyclists, 2015. Fall 2015 round bicycle friendly community awards and honorable mentions. Available at: http://bikeleague.org/sites/default/files/BFC_Awards_and_HMs_Fall_2015.pdf.

Lee, S.X., McLachlan, G.J., 2013. On mixtures of skew normal and skew t-distributions. *Advances in Data Analysis and Classification* 7(3), 241-266.

Lee, S., McLachlan, G.J., 2014. Finite mixtures of multivariate skew t-distributions: Some recent and new results. *Statistics and Computing* 24(2), 181-202.

Lele, S.R., 2006. Sampling variability and estimates of density dependence: A composite-likelihood approach. *Ecology* 87(1), 189-202.

LeSage, J.P., Pace, R., 2009. *Introduction to Spatial Econometrics*. CRC Press, Taylor & Francis Group, Boca Raton, FL.

Liesenfeld, R., Richard, J-F., Vogler, J., 2013. Analysis of discrete dependent variable models with spatial correlation, available at SSRN: http://ssrn.com/abstract=2196041.

Lin, T.I., McLachlan, G.J., Lee, S.X., 2016. Extending mixtures of factor models using the restricted multivariate skew-normal distribution. *Journal of Multivariate Analysis* 143, 398-413.

Lindsay, B.G., 1988. Composite likelihood methods. *Contemporary Mathematics* 80, 221-239.

Lindsay, B.G., Yi, G.Y., Sun, J., 2011. Issues and strategies in the selection of composite likelihoods. *Statistica Sinica* 21(1), 71-105.

Liu, I., Agresti, A., 2005. The analysis of ordered categorical data: An overview and a survey of recent developments. *Test* 14(1), 1-73.

Lo, A.Y., Jim, C.Y., 2012. Citizen attitude and expectation towards greenspace provision in compact urban milieu. *Land Use Policy* 29(3), 577-586.

Ma, L., Dill, J., Mohr, C., 2014. The objective versus the perceived environment: What matters for bicycling? *Transportation* 41(6), 1135-1152.

Maldonado-Hinarejos, R., Sivakumar, A., Polak, J.W., 2014. Exploring the role of individual attitudes and perceptions in predicting the demand for cycling: A hybrid choice modelling approach. *Transportation* 41(6), 1287-1304.

Malsiner-Walli, G., Fruhwirth-Schnatter, S., Grun, B., 2016. Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing* 26(1-2), 303-324.

McCright, A.M., 2010. The effects of gender on climate change knowledge and concern in the American public. *Population and Environment* 32(1), 66-87.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15(5), 447-470.

McKelvey, R., Zavoina, W., 1971. An IBM Fortran IV program to perform n-chotomous multivariate probit analysis. *Behavioral Science* 16(2), 186-187.

McMillen, D.P., 1995. Selection bias in spatial econometric models. *Journal of Regional Science* 35(3), 417-436.

McMillen, D.P., 2010. Issues in spatial data analysis. *Journal of Regional Science* 50(1), 119-141.

McMillen, D.P., 2012. Perspectives on spatial econometrics: Linear smoothing with structured models. *Journal of Regional Science* 52(2), 192-209.

Meddin, R., DeMaio, P., 2012. The bike-sharing world map. Available at: http://www.metrobike.net.

Meintanis, S.G., Hlávka, Z., 2010. Goodness-of-fit tests for bivariate and multivariate skew-normal distributions. *Scandinavian Journal of Statistics* 37(4), 701-714.

Midgley, P., 2009. The role of smart bike-sharing systems in urban mobility. *Journeys* 2, 23-31.

Mittlehammer, R.C., Judge, G., 2011. A family of empirical likelihood functions and estimators for the binary response model. *Journal of Econometrics* 164(2), 207–217.

Molenaar, D., Dolan, C.V., Verhelst, N.D., 2010. Testing and modeling non-normality with the one-factor model. *British Journal of Mathematical and Statistical Psychology* 63(2), 293-317.

Molenberghs, G., Verbeke, G., 2005. *Models for Discrete Longitudinal Data*. Springer.

Müller, G., Czado, C., 2005. An autoregressive ordered probit model with application to high-frequency financial data. *Journal of Computational and Graphical Statistics* 14(2), 320-338.

Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 1067-1101.

Noland, R.B., Deka, D., Walia, R., 2011. A statewide analysis of bicycling in New Jersey. *International Journal of Sustainable Transportation* 5(5), 251-269.

Pace, R.K., LeSage, J.P., 2011. Fast simulated maximum likelihood estimation of the spatial probit model capable of handling large samples, Available at SSRN: http://ssrn.com/abstract=1966039, accessed September 21, 2016.

Paleti, R., Bhat, C.R., 2013. The composite marginal likelihood (CML) estimation of panel ordered-response models. *Journal of Choice Modelling* 7, 24-43.

Paleti, R., Bhat, C.R., Pendyala, R., 2013. Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. *Transportation Research Record: Journal of the Transportation Research Board* 2382, 162-172.

Partridge, M.D., Boarnet, M., Brakman, S., Ottaviano, G., 2012. Introduction: Whither spatial econometrics? *Journal of Regional Science* 52(2), 167-171.

Parkin, J., Wardman, M., Page, M., 2008. Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation* 35(1), 93-109.

Pinjari, A.R., Pendyala, R.M., Bhat, C.R., Waddell, P.A., 2011. Modeling the choice continuum: An integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation* 38(6), 933-958.

Pinkse, J., Slade, M., Shen, L., 2006. Dynamic spatial discrete choice using one-step GMM: An application to mine operating decisions. *Spatial Economic Analysis* 1(1), 53-99.

Pinkse, J., Slade, M., 2010. The future of spatial econometrics. *Journal of Regional Science* 50(1), 102–118.

Pretty, J., Peacock, J., Hine, R., Sellens, M., South, N., Griffin, M., 2007. Green exercise in the UK countryside: Effects on health and psychological well-being, and implications for policy and planning. *Journal of Environmental Planning and Management* 50(2), 211-231.

Pucher, J., Dill, J., Handy, S., 2010. Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive Medicine* 50, S106-S125.

Reddy, S., Shilton, K., Denisov, G., Cenizal, C., Estrin, D., Srivastava, M., 2010. Biketastic: Sensing and mapping for better biking. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1817-1820.

Rietveld, P., Daniel, V., 2004. Determinants of bicycle use: Do municipal policies matter? *Transportation Research Part A* 38(7), 531-550.

Rhee, K.A., Kim, J.K., Lee, Y.I., Ulfarsson, G.F., 2016. Spatial regression analysis of traffic crashes in Seoul. *Accident Analysis and Prevention* 91, 190-199.

Roney, J.M., 2008. Bicycles pedaling into the spotlight. *Eco-Economy Indicator*, 239-42.

Sallis, J.F., Owen, N., Fisher, E.B., 2002. Ecological models of health behavior. In: Glanz, K., Rimer, B.K., Lewis, F.M. (eds.), *Health Behavior and Health Education: Theory, Research, and Practice*, Wiley, San Francisco.

Sallis, J.F., Conway, T.L., Dillon, L.I., Frank, L.D., Adams, M.A., Cain, K.L., Saelens, B.E., 2013. Environmental and demographic correlates of bicycling. *Preventive Medicin*e 57(5), 456-460.

Salvy, S.J., Roemmich, J.N., Bowker, J.C., Romero, N.D., Stadler, P.J., Epstein, L.H., 2009. Effect of peers and friends on youth physical activity and motivation to be physically active. *Journal of Pediatric Psychology* 34(2), 217-225.

Sener, I.N., Eluru, N., Bhat, C.R., 2009a. Who are bicyclists? Why and how much are they bicycling? *Transportation Research Record: Journal of the Transportation Research Board* 2134, 63-72.

Sener, I.N., Eluru, N., Bhat, C.R., 2009b. An analysis of bicycle route choice preferences in Texas, U.S. *Transportation* 36(5), 511-539.

Shaheen, S., Guzman, S., Zhang, H., 2010. Bikesharing in Europe, the Americas, and Asia: Past, present, and future. *Transportation Research Record: Journal of the Transportation Research Board* 2143, 159-167.

Sidharthan, R., Bhat, C.R., 2012. Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis* 44(4), 321-349.

Smith, M.S., Gan, Q., Kohn, R.J., 2012. Modelling dependence using skew t copulas: Bayesian inference and applications. *Journal of Applied Econometrics* 27(3), 500-522.

Steinbach, R., Green, J., Datta, J., Edwards, P., 2011. Cycling and the city: A case study of how gendered, ethnic and class identities can shape healthy transport choices. *Social Science & Medicine* 72(7), 1123-1130.

Stewart, M., 2005. A comparison of semiparametric estimators for the ordered response model. *Computational Statistics and Data Analysis* 49(2), 555-573.

Stinson, M., Bhat, C.R., 2004. Frequency of bicycle commuting: Internet-based survey analysis. *Transportation Research Record: Journal of the Transportation Research Board* 1878, 122-130.

Turner, C., McClure, R., 2003. Age and gender differences in risk-taking behaviour as an explanation for high incidence of motor vehicle crashes as a driver in young males. *Injury Control and Safety Promotion* 10(3), 123-130.

Van Acker, V., Derudder, B., Witlox, F., 2013. Why people use their cars while the built environment imposes cycling. *Journal of Transport and Land Use* 6(1), 53-62.

Varin, C., Czado, C., 2010. A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics* 11(1), 127-138.

Varin, C., Vidoni, P., 2005. A note on composite likelihood inference and model selection. *Biometrika* 92(3), 519-528.

Varin, C., Vidoni, P., 2008. Pairwise likelihood inference for general state space models. *Econometric Reviews* 28(1-3), 170-185.

Varin, C., Reid, N., Firth, D., 2011. An overview of composite marginal likelihoods. *Statistica Sinica* 21(1), 5-42.

Ward, P.S., Florax, R.J., Flores-Lagunes, A., 2010. Agricultural productivity and anticipated climate change in Sub-Saharan Africa: A spatial sample selection model. Selected Paper prepared for presentation at the Agricultural & Applied Economics Association.

Winship, C., Mare, R.D., 1984. Regression models with ordinal variables. *American Sociological Review* 49(4), 512-525.

Worldwatch Institute, 2008. Bicycle production reaches 130 million units. Available at: http://www.worldwatch.org/node/5462.

Xu, X., Reid, N., 2011. On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference* 141(9), 3047-3054.

Yang, S., Allenby, G.M., 2003. Modeling interdependent consumer preferences. *Journal of Market Research* 40(3), 282-294.

Yi, G.Y., Zeng, L., Cook, R.J., 2011. A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. *Canadian Journal of Statistics* 39(1), 34-51.

Zhao, Y., Joe, H., 2005. Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* 33(3), 335-356.
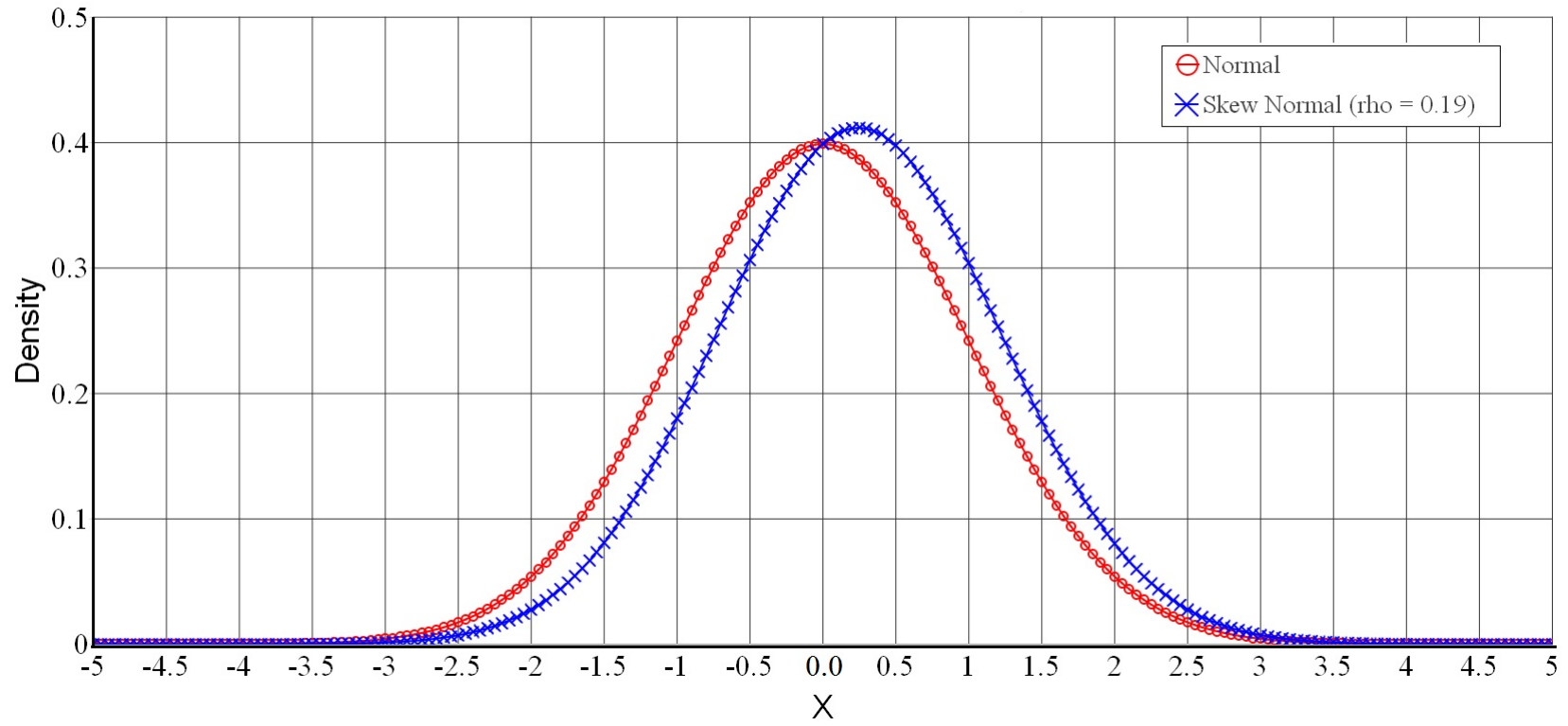
Figure 1: Marginal Probability Density Plot

Table 1. Estimation results of the SSN-GOR model

Variables	Latent bicycling propensity		Threshold between occasional bicycling and infrequent bicycling		Threshold between infrequent bicycling and frequent bicycling		Threshold between frequent bicycling and everyday bicycling	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Threshold constant between not a bicyclist and occasional bicyclist α_1	-10.144 (t-stat: -8.11)							
Threshold constants α_k			0.053	1.79	1.202	4.22	1.361	3.600
Individual Characteristics								
Male	0.311	2.25						
Bachelor's degree or higher educational level	0.190	2.06	-0.234	-3.05	-0.192	-3.00		
Age variables: between 35 and 54 years old	0.320	2.75	-0.170	-2.53			-0.101	-2.03
Work status: full-time worker	-0.108	-2.10	0.287	2.91				
Household demographics								
Number of automobiles	-1.034	-3.65	0.675	3.65				
Presence of children	-0.431	-2.69						
Household income: (base is more than $50,000)								
Less than $25,000	-0.321	-2.50	0.279	3.21				
$25,000-$34,999	-0.088	-2.22						
$35,000-$49,999	-0.057	-2.10						
Environmental characteristics								
Intensity of Bicycle Lane Infrastructure (IBLI)	1.298	7.21	-0.370	-1.99			-0.102	-2.49
(IBLI > 4 miles) interacted with Male	-0.226	4.02						
(IBLI > 3 miles) interacted with '18 to 34 years old' age category	0.184	3.97						
Spatial parameter δ	0.152 (t-stat: 3.50)							
Skew parameter ρ	0.190 (t-stat: 2.16)							
Log-composite likelihood at convergence	-2,823,142.9							
Number of parameters estimated	27							
Number of observations	3,760							

Table 2. Pseudo-elasticity effects of variables

Variables	Not a bicyclist	Occasional bicyclist	Infrequent bicyclist	Frequent bicyclist	Everyday bicyclist
Individual Characteristics					
Male	-13.6	-9.4	8.1	6.7	5.4
Bachelor's degree or higher educational level	-28.7	-20.3	4.7	20.7	7.3
Age variable: between 18 and 34 years old	-7.0	-8.5	6.7	7.8	7.2
Age variable: between 35 and 54 years old	-10.2	-9.6	10.2	9.3	11.5
Work status: full-time worker	27.5	25.3	-3.6	-8.1	-11.9
Household demographics					
Number of automobiles	30.1	27.4	-11.2	-15.3	-17.6
Presence of children	20.8	19.3	-7.2	-12.4	-13.9
Household income: (base is more than $50,000)					
Less than $25,000	28.6	20.4	-6.2	-15.2	-18.4
$25,000-$34,999	20.9	12.7	-5.3	-10.7	-13.2
$35,000-$49,999	14.3	9.5	-4.8	-7.6	-6.0
Environmental characteristics					
Intensity of Bicycle Lane Infrastructure (IBLI)	-14.9	-19.5	8.7	11.6	16.2

Table 3. Measures of fit

Summary Statistic		SOR	ASP-GOR	SP-GOR	ASSN-GOR	SSN-GOR
Number of Observations		3,760				
Composite log-likelihood (CLL) at convergence of the naïve model		-4,832,776.3				
Number of parameters		16	25	26	26	27
Composite log-likelihood (CLL) at convergence		-2,823,251.3	-2,823,224.1	-2,823,208.8	-2,823,175.7	-2,823,142.9
Adjusted composite likelihood ratio test (ADCLRT) between SSN-GOR and the corresponding model		Test statistic [-2*(CLL$_{SOR}$-CLL$_{SSN-GOR}$)]=216.8 > Chi-Squared statistics with 2 degrees of freedom at any reasonable level of significance	Test statistic [-2*(CLL$_{ASP-GOR}$-CLL$_{SSN-GOR}$)]=162.4 > Chi-Squared statistics with 2 degrees of freedom at any reasonable level of significance	Test statistic [-2*(CLL$_{SP-GOR}$-CLL$_{SSN-GOR}$)]=133.8 > Chi-Squared statistics with 1 degree of freedom at any reasonable level of significance	Test statistic [-2*(CLL$_{ASSN-GOR}$-CLL$_{SSN-GOR}$)]=66.0 > Chi-Squared statistics with 1 degree of freedom at any reasonable level of significance	NA
Spatial correlation		0.0 (fixed)	0.0 (fixed)	0.150 (t-stat: 3.60)	0.0 (fixed)	0.152 (t-stat: 3.50)
Skew parameter		0.0 (fixed)	0.0 (fixed)	0.0 (fixed)	0.178 (t-stat: 2.11)	0.190 (t-stat: 2.16)
Average probability of correct prediction		0.517	0.528	0.539	0.555	0.579
Bicycling frequency categories	Actual percentage	Predicted percentage				
Not a bicyclist	56.5	60.5	60.0	59.2	58.8	58.3
Infrequent bicyclist	24.0	25.1	25.0	25.0	24.6	24.4
Occasional bicyclist	7.2	5.8	6.0	6.3	6.5	6.7
Frequent bicyclist	9.0	6.3	6.6	7.1	7.4	7.6
Everyday bicyclist	3.3	2.3	2.4	2.4	2.7	3.0
Mean Absolute Percentage Error (MAPE)		9.42%	8.15%	7.21%	5.22%	3.54%

NA: Not applicable

Table 4. Percentage change in bicycling frequency category shares due to a 10% increment of miles of bicycling lanes within a 1-mile distance (standard error in parenthesis)

Bicycling frequency category	SOR		ASP-GOR		SP-GOR		ASSN-GOR		SSN-GOR
	Elasticity	P*	Elasticity	P*	Elasticity	P*	Elasticity	P*	Elasticity
Little to No Bicycling (LNB)	-7.6% (0.68)	0.00	-8.5% (0.73)	0.04	-9.5% (0.49)	0.14	-10.3% (1.97)	--	-11.2% (1.03)
Regular Bicycling (RB)	11.8% (1.02)	0.00	13.2% (0.95)	0.00	14.6% (1.17)	0.00	15.8% (0.96)	0.00	23.7% (0.91)

*P value of the difference between the model elasticity and the SSN-GOR model elasticity.

A "—" implies that the difference is not statistically significant even at the 0.2 level of significance.