Transportation Research Part B 95 (2017) 126–148

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

A spatial generalized ordered-response model with skew normal kernel error terms with an application to bicycling frequency

Chandra R. Bhat[a,b,*], Sebastian Astroza[a,1], Amin S. Hamdi[c,2]

[a] The University of Texas at Austin, Department of Civil, Architectural and Environmental Engineering, 301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA
[b] King Abdulaziz University, Jeddah 21589, Saudi Arabia
[c] King Abdulaziz University, Department of Civil Engineering, P.O. Box 80204, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Article history:
Received 1 July 2016
Revised 24 October 2016
Accepted 25 October 2016
Available online 10 November 2016

Keywords:
Generalized ordered response model
Skew normal distribution
Social interactions
Composite marginal likelihood
Spatial econometrics
Bicycling frequency

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.

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

* Corresponding author. Fax: 1-512-475-8744.
E-mail addresses: bhat@mail.utexas.edu (C.R. Bhat), sastroza@utexas.edu (S. Astroza), ahamdi@kau.edu.sa (A.S. Hamdi).
[1] Fax: 1 512 475 8744
[2] Fax: +966 2 695 2179

http://dx.doi.org/10.1016/j.trb.2016.10.014
0191-2615/

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., 2014), 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