Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 23 (2017) 460–479

Transportation Research Procedia

www.elsevier.com/locate/procedia

22nd International Symposium on Transportation and Traffic Theory

Models of Count with Endogenous Choices

Roger B. Chen*

Rochester Institute of Technology, One Lomb Memorial Drive, Rochester, NY 14623-5608, USA

Abstract

In transportation and traffic analysis count data arises frequently, collectively emerging from individual traveler choices from a choice set of alternatives. Examples include network origin-destination (OD) flow rates and visitor counts at transit stations. From a modeling perspective, these data are aggregate counts at the top level, but comprised of individual discrete choices at the lower level. Models of count data are widely applied in the transportation and traffic fields. However, only a moderate level of applications jointly model count observations at the top level with discrete choice models at the bottom level under a random utility maximization (RUM) framework. This paper considers modeling count data with an underlying choice process as a joint model that merges an observed event count process with a discrete choice process, where the count level is Poisson distributed. This model captures both processes within a single random utility framework that preserves a direct mapping between the count intensity and the utility of the chosen alternative, including unobserved variables and latent factors. The decision-making context presented examines discretionary activity type choice for activities completed within a one-day period. The estimation results for this model are compared against (i) a mixed-logit model and (ii) a mixed-Poisson model, each with normally distributed parameters. The results indicate that a model of count with endogenous choices can account for the randomness associated with the utility of choice alternatives from lower level discrete choices, consequently leading to significantly different utility parameter estimates for the Poisson rate parameter in the upper level. Furthermore, while the linkage between the maximizing utility and rate parameter is preserved in this joint model, identifying the contribution of attributes between the two levels requires further parameterization.

Peer review under responsibility of the scientific committee of the 22nd International Symposium on Transportation and Traffic Theory.

Keywords: Poisson Models, Choice Models, Cumulants, Computational Statistics, Hierarchical Bayes, Markov-Chain-Monte-Carlo

* Corresponding author. Tel.: +1-585-475-5802; fax: +0-000-000-0000 .
E-mail address: rbcgis@rit.edu

Peer review under responsibility of the scientific committee of the 22nd International Symposium on Transportation and Traffic Theory.
10.1016/j.trpro.2017.05.026

1. Introduction and Background

Observed count data arises frequently in transportation and traffic analysis, leading analysts to produce several methods for estimating and forecasting count rates, largely within an econometric framework. Example contexts include network origin-destination (OD) flow rates (Parry and Hazelton 2012) and visitor counts at transit stations and other facilities (Ceder 2007). In these examples, the observed aggregate event counts arise collectively from an underlying individual choice from a choice set of alternatives. For example, visitor counts at transit stations are count observations at the top level, but comprised of individual choices of stations at a lower level.

Within the applied econometrics literature, models that characterize the distribution of aggregate level count or rate data are extensive. Models of count data, specifically the Poisson or negative binomial models, have seen wide application in the transportation and traffic field. These applications include daily activity frequency (Ma and Goulias 1999), accident rates (Lord and Mannering 2010; Karlaftis and Tarko 1998) and transit station arrival rates (Ceder 2007; Salek and Machemehl 1999). Outside of transportation, other areas related to travel behavior have also modeled the intensity of behavioral processes at an aggregate level. Exponential random graph models (ERGM) have been used to model the degree of social connectivity within social networks randomly (Wasserman 1994; Jackson 2010). In the area of network science, human mobility pattern are described using power law distributions fitted to cell phone location data (Gonazles et al. 2008).While distributions like the Poisson, ERGMs and power law are used extensively in transportation systems analysis to describe aggregate level observations, the mapping and linkage with discrete choice modeling at the individual traveler level has received less attention. Even for models of count data (i.e. Poisson, Negative Binomial, etc.), few applications jointly model count observations at the upper level with discrete choice models for individual travelers at the lower level, within a single random utility maximization (RUM) framework. Given the wide application and extensive development of disaggregate discrete choice models in transportation analysis (Train 2009; Ben-Akiva and Lerman 1985), integrating them with other models that account for distributions at a more aggregate level, such as those mentioned, may be beneficial. By preserving a direct mapping between the distribution parameters at both levels, including unobserved variables and latent factors, this integration could permit econometrically estimating the marginal impact of individual choices on collective effects. Two areas of related literature are reviewed briefly before presenting the model formulation and estimation results.

1.1. Collective Effects in Transportation

Collective effects that emerge from the collection of individual traveler choices are prevalent in the transportation and traffic analysis, ranging from network link flows that arise from route choices to OD flows that arise from destination choices. Within an econometric framework, travelers' route and destination choices have been modeled and parameters estimated with empirical data (Prato 2009; Bekhor and Prashker 2008). The dominant approach for capturing the interdependence between individual travel choices and network performance has been to solve for an assumed equilibrium under various assumptions on user behavior and estimated parameters from choice modeling. Simulation and laboratory experiments has also been used that simulate either the network conditions in response to decisions from real, actual commuters (Mahmassani et al. 1986, Mahmassani 1990, Helbing et al 2002) or both the network and individual user decisions (Chen and Mahmassani 2004; Mahmassani and Chang 1986), such as agent-based simulation. These approaches provide the ability to investigate dynamic system evolution, in particular convergence and stability, and the mechanisms underlying the day-to-day choice behavior of system users.

In the field of network or system science, there is also interest in collective effects where the aggregate distribution of a collection of individual decisions is modeled. Studies in this field have found that human mobility patterns, characterized by displacement over time, follow power law or Levy flight distributions (Gonzalez et al. 2009; Brockmann 2011). A second example includes patterns in social networks, where individual decisions to connect to others collectively give rise to these networks (Jackson 2004; Wasserman and Faust 1994). Exponential random graph models are used to explain the distribution of these connections. While these studies provide an improved understanding of the mechanics underlying the systems, the linkage with discrete choice models that capture