

Mixed Generalized Ordered Logit Model: Application to Airline Demand Modelling

Sudipta Dey Tirtha

Post-doctoral Scholar

Department of Civil, Environmental & Construction Engineering

University of Central Florida

Tel: 407-543-7521

Email: sudiptadey.tirtha@ucf.edu

ORCID number: 0000-0002-6228-0904

Naveen Eluru¹

Professor

Department of Civil, Environmental & Construction Engineering

University of Central Florida

Tel: 407-823-4815, Fax: 407-823-3315

Email: naveen.eluru@ucf.edu

ORCID number: 0000-0003-1221-4113

ABSTRACT

Traditional ordered outcome/response models such as ordered logit/probit are widely employed by researchers to examine the effects of independent factors on ordered response variables. In ordered logit/probit models, a uni-dimensional latent propensity equation is partitioned using a set of fixed thresholds to compute the probability allocated to each of the ordered alternatives. The primary drawback of the traditional ordered model arises due to the implicit restriction that thresholds remain the same across the observations. Generalized ordered response (GOR) model addresses this limitation by allowing the thresholds to be flexible across observations as a function of independent variables. The GOR model can also be easily extended to accommodate unobserved or missing information in a framework analogous to the mixed ordered logit/probit model. In this chapter, we present a detailed formulation for the Mixed GOR model with an application to airport level air passenger demand data in the US. In the model application, we consider quarterly passenger demand records of 300 airports across the US from 2010 – 2018. The model estimation process is implemented in GAUSS to illustrate how the MGOL model relates the different independent variables to airline demand.

¹ Corresponding Author

1 Introduction

Discrete outcome variables can be broadly classified as ordered or unordered variables. Examples of ordered discrete variables include household vehicle ownership, crash injury severity, and three- or five-point Likert scale ratings. Example of unordered discrete variables include travel mode choice, activity type selection, and crash type outcomes. Unordered model frameworks including multinomial logit model (MNL) and other variants can be employed to analyze all discrete variables. However, for modeling ordered discrete variables a specific class of discrete outcome models referred to as ordered response models are widely employed. In this chapter, we provide a discussion on ordered response frameworks including mathematical details of ordered response model, generalized ordered response model and mixed generalized ordered response model with an example application for airline demand analysis (for discussion on unordered models please see Bhat et al. 2008).

The ordered response approaches accommodate the inherent ordering of the chosen/dependent variable while partitioning a uni-dimensional latent propensity equation using a set of fixed thresholds to compute the probability of the ordered alternatives. Further, the uni-dimensional propensity approach allows for identification of factors affecting the dependent variable in a parsimonious manner relative to the unordered model systems. In traditional ordered response systems, the number of parameters/coefficients estimated is tied to the number of independent variables (and is not impacted by the number of alternatives). While this ability to incorporate the impact of independent variables on the dependent variable parsimoniously through a single equation is useful, in some cases this can also result in restrictive specifications. The restrictive structure only allows for a single parameter to represent the impact of each independent variable. The structure does not allow for a variable to impact partial group of alternatives. For example, consider an analysis of household vehicle ownership variable defined as zero, one, two and three or more vehicles. Household income available in the data is defined as <30,000\$, between 30,000\$ and 60,000\$ and greater than 60,000\$. Let us consider the possibility that households with income greater than 60,000\$ are not likely to own zero cars while the exact number of vehicles owned is not affected by income. Incorporating this behavior in the traditional ordered model explicitly is not possible as reducing the probability of zero vehicles is also associated with a corresponding increase in the probability of three or more vehicles (see Eluru et al. 2008 for discussion in another empirical context). The restrictive assumption in traditional ordered models can be addressed by allowing for the impact of independent variables to vary across the alternative space in two forms. First, the change is accommodated by allowing the thresholds to vary across observations as a function of observed attributes. Second, the traditional approach is modified to allow for the impact of exogenous variables to vary across alternatives. The first approach gives rise to the model referred to as the Generalized Ordered Response Model (Terza 1985; Eluru et al. 2008; Srinivasan 2002; Yasmin and Eluru 2013) and the second approach results in a model referred to as the Partial Proportional Odds model (Peterson and Harrell 1990).

Eluru and Yasmin (2015) established that these two variants are mathematically identical. Further, they highlight the specific advantages of the generalized ordered response approach for accommodating for the presence of unobserved heterogeneity. Henceforth, in this chapter we refer to these models as generalized ordered response (GOR) models. In literature, GOR models have been widely employed for modelling ordered response variables in different domains such as crash severity analysis, transportation demand modelling, social science, tourism, medical science, agriculture, and education (see Section 3 for more details).

The objective of the current chapter is to offer the reader a detailed description of the generalized ordered response model from first principles. An overview of the chapter organization follows: An overview of the different variations of the ordered response model including traditional ordered response model, GOR model and mixed GOR model are presented in Section 2. Section 3 provides a summary of ordered response framework applications from various research domains. Section 4 provides a real-world application of the generalized ordered model system. Finally, Section 5 concludes the chapter.

2 Mathematical Formulation of Ordered Response Models

In this section, we present detailed mathematical formulations for traditional OR model and its advanced variants including GOR model and Mixed GOR model. In OR model, probabilities of ordered categories are estimated by partitioning a continuous latent propensity using a set of fixed thresholds. In GOR model, thresholds are parametrized to allow them to vary across observations. Finally, Mixed GOR model is formulated to accommodate for the potential influence of unobserved factors on the dependent variable.

2.1 *Traditional Ordered Response Model*

Let i ($i = 1, 2, \dots, N$) and j ($j = 1, 2, \dots, J$) be indices to denote the observation and the different levels of an ordinal dependent variable respectively. The observation specific dependent variable y_i is postulated to be linked with an underlying continuous latent propensity variable y_i^* . This variable is parameterized as a linear function as follows:

$$y_i^* = X_i\beta + \varepsilon_i, \text{ for } i = 1, 2, \dots, N \quad (1)$$

Where, X_i is a vector of variables (excluding a constant), β is a vector of unknown parameters/coefficients to be estimated and ε_i is an idiosyncratic error term. The exact formulation of the model is determined based on the distribution of the error term. If ε_i is assumed to be identically and independently standard logistic (normal) distributed across observation i , then the model will result in ordered logit (probit) model. The continuous latent variable, y_i^* is mapped to actual ordered categories y_i using a set of fixed thresholds, τ ($-\infty < \tau_1 < \tau_2 < \dots < \tau_{J-1} < +\infty$). The variable y_i^* is linked to the observed variable y_i in the following manner:

$$y_i = j, \text{ if } \tau_{j-1} < y_i^* < \tau_j, \text{ for } j = 1, 2, \dots, J \quad (2)$$

Now, the probability expression for alternative j is computed as follows:

$$Pr(y_i = j) = \Lambda(\tau_j - X_i\beta) - \Lambda(\tau_{j-1} - X_i\beta) \quad (3)$$

Where, $\Lambda(\cdot)$ represents the standard logistic (normal) cumulative distribution function. The standard logistic (normal) assumption results in an ordered logit (probit) model. The unknown

parameters in equation (3) are τ and β which can be estimated by maximizing log-likelihood function presented below:

$$LL = \sum_{i=1}^N \sum_{j=1}^J \ln[\Pr(y_i = j|X_i)^{m_{ij}}] \quad (4)$$

In equation (4), m_{ij} is a binary variable where $m_{ij} = 1$ if j is the observed level and 0, otherwise. The reader would note that the ordering of thresholds must be maintained to ensure non-negativity of probability values.

2.2 Generalized Ordered Response Model

The generalized ordered response framework allows the thresholds to vary across the observations. This flexibility is achieved by the expression of thresholds as a linear function of independent variables (Terza 1985; Srinivasan 2002; Eluru and Yasmin 2015) using the following parametric form:

$$\tau_{i,j} = \alpha_j + \delta_j Z_i \quad (5)$$

where, α_j is a constant specific to category j , Z_i is a set of exogenous variables excluding a constant and δ_j is a vector of parameters to be estimated. Therefore, the probability expression in equation (3) can be modified as follows:

$$\Pr(y_i = j) = \Lambda(\tau_{i,j} - X_i\beta) - \Lambda(\tau_{i,j-1} - X_i\beta) \quad (6)$$

The unknown parameters, α , δ and β are estimated maximizing the same log-likelihood function presented in equation (4) with probability values generated employing equation (6).

2.3 Mixed Generalized Ordered Response Framework

The GOR framework described in Section 2.2 can be employed to examine the influence of observed variables on the decision process. However, in many contexts, it is also important for an analyst to recognize that variables that are missing or unobserved factors for a specific observation (or a set of observations) can impact the decision process. For example, individual characteristics and vehicle safety equipment specifications are often missing in police/hospital reported crash dataset. Therefore, in crash severity analysis using traditional OR and GOR model, it possible that the presence of unobserved factors can modify the effect of independent variables on outcome variables (Eluru and Yasmin 2015). In another context, consider that an analyst has access to a multi-year air passenger demand data at the airport level and is interested in identifying the key factors affecting the demand. In this context, airport specific factors such as number of terminals, service quality of terminals and intermodal transportation services affecting the demand at an

airport are unobserved or missing in the dataset. The unobserved factors can affect airline passenger demand at several resolutions such as airport, year, and airport-year (Tirtha et al. 2023). The Mixed GOR framework includes additional parameters to account for the presence of such unobserved effects. Several research efforts considered the impact of unobserved factors in GOR modelling approach (please see Eluru et al. 2008; Marcoux et al. 2018; Yasmin et al. 2015; Balusu et al. 2018; Balan and Paleti 2018; Rahman et al. 2019; Tirtha et al. 2023). In this section, we will present MGOR model structure that allows us to incorporate the potential effect of unobserved factors/variables on the dependent variable.

Let μ_i and γ_{ij} be two column vectors representing the vectors of unobserved factors modifying the effect of independent variables for observation i in equation (1) and (5), respectively. η_k represents the vector of parameters/coefficients representing the impact of these unobserved factors influencing outcome variable across repetition level k . Therefore, we can rewrite the propensity and threshold functions as follows:

$$y_i^* = (X_i + \mu_i)\beta + (\eta_k)X_i + \varepsilon_i, \text{ for } i = 1, 2, \dots, N \quad (7)$$

$$\tau_{i,j} = \alpha_j + (\delta_j + \gamma_{ij})Z_i \quad (8)$$

However, in ordered model framework, it is important to maintain non-negativity of the probabilities of ordered outcomes. To ensure non-negativity condition, it is necessary that $(\alpha_j + (\delta_j + \gamma_{ij})Z_i) > (\alpha_{j-1} + (\delta_{j-1} + \gamma_{i,j-1})Z_i)$. In GOR model (without any unobserved heterogeneity consideration), it is very unlikely that this condition will be violated. However, in MGOR model estimation ensuring that the thresholds ordering remains as required is not straightforward (see detailed discussion in Srinivasan 2002; Eluru et al. 2008). To avoid such potential negative probability scenarios, Eluru et al. (2008) proposed the following alternative formulation for threshold parameters:

$$\tau_{i,j} = \tau_{i,j-1} + \exp [\alpha_j + (\delta_j + \gamma_{ij})Z_i] \quad (9)$$

The reader would note that all thresholds are not identifiable. Hence, , we need to set either the constant in the propensity or one of the thresholds to 0. In our model structure, we include a constant in latent propensity, and we set $\tau_{i,1} = 0$. For probability computation, it is necessary to outline the structure for the different unobserved vectors μ_i, γ_{ij} and η_k . We consider that these three vectors are drawn independently from normal distributions as follows: $N(0, \sigma^2), N(0, \vartheta^2)$ and $N(0, \rho^2)$. Thus, conditional on σ, ϑ and ρ , the probability formula for observation i and alternative j in MGOL model is computed as shown below:

$$\begin{aligned} Pr(y_i = j | \mu_i, \gamma_{ij}, \eta_k) \\ = \Lambda(\tau_{i,j} - ((X_i + \mu_i)\beta + (\eta_k)X_i)) - \Lambda(\tau_{i,j-1} - ((X_i + \mu_i)\beta + (\eta_k)X_i)) \end{aligned} \quad (10)$$

The unconditional probability can be expressed as below (Eluru and Yasmin 2015):

$$Pr(y_i = j) = \int_{\mu_i, \gamma_{ij}, \eta_k} Pr(y_i = j | \mu_i, \gamma_{ij}, \eta_k) * dF(\mu_i, \gamma_{ij}, \eta_k) d(\mu_i, \gamma_{ij}, \eta_k) \quad (11)$$

To estimate the above MGOR model, we apply Quasi-Monte Carlo simulation techniques based on scrambled Halton sequences for the maximum simulated likelihood approach (see Bhat 2001; Eluru et al. 2008 for more details). The approach followed in this study employs Halton draws to calculate the multidimensional integrals presented in Equation 11. The model systems are programmed and estimated in GAUSS. As GAUSS is not readily accessible to most readers, we provide our model estimation procedures for OR and GOR models using statistical software SAS in the supplementary materials. Advanced readers with GAUSS access can contact the authors for GAUSS codes of advanced GOR variants.

3 Ordered Response Framework Applications

Several research studies have employed ordered response frameworks. In this section, we provide a review ordered response framework applications including traditional ordered models, generalized ordered models and mixed generalized ordered models. Table 1 provides a summary of a sample of earlier studies with information on study, study objective, data source, dependent variable (number of alternatives), alternative definitions, and the exact model structure considered. Given the dominance of ordered response model applications in the transportation field, studies in the table are organized by transportation and other fields. Table 1 provides a summary of 21 studies from the transportation domain and 11 studies from other domains. The reader would note that literature review was not intended to serve as an exhaustive review – it is developed to provide an overview of different fields of application as well as a representation of different ordered response variants employed across wide range of research domains.

Several observations can be made from our review. First, ordered response frameworks have been employed in transportation, social sciences, tourism, medical science, agriculture, and education. Second, among transportation studies, large number of studies focus on transportation safety examining injury severity. In fact, the reader would note that majority of seminal work on mixed generalized ordered response frameworks arose from injury severity modeling. Other fields of transportation where ordered response models are employed include public transport usage, ride-hailing demand, and airline demand. Third, among the different fields employing ordered response models the variables of interest include poverty levels, nurse expertise levels, bladder tumor stages and student grades. Fourth, several variants of ordered response frameworks are implemented in literature including ordered logit and probit, generalized ordered logit (including PPO) and probit, and mixed generalized ordered logit and probit model. Among alternative model structures considered, OR, GOR, PPO and MGOR formulations have been considered in 9, 1, 13, and 10 studies, respectively. While GOR formulation is commonly employed in transportation literature, studies in other domain mostly considered PPO variant. Fifth, the number of alternatives of ordered variables considered in the studies summarized in Table 1 ranges between 2 and 14. Finally, we observe that error term is assumed to follow logistic and normal distributions across these studies. Among the studies presented in Table 1, we observe that logistic (normal) distribution was employed in 26 (6) studies out of 32 indicating a preference for logit variants.

TABLE 1 Summary of Literature Review

Study	Study Objective	Data Source (Region; Year)	Dependent variable (No of Alternatives)	Alternatives	Model (logit/probit)
<i>Transportation Studies</i>					
Eluru et al. 2008	To study injury severity of non-motorists	General Estimates System (GES) database (US; 2004)	Injury severity (4)	No injury, non-incapacitating, incapacitating and fatal injury	MGOR (logit)
Yasmin et al. 2014	To estimate an econometric model of driver injury severity	Victorian Crash Database (Australia; 2006-2010)	Injury severity (4)	No injury, minor, serious and fatal injury	GOR (logit)
Kaplan and Prato 2012	To analyze the injury severity in bus accidents	GES database (US; 2005-2009)	Injury severity (5)	Property damage, light, severe non-incapacitating, incapacitating injury and fatality	PPO (logit)
Wang et al. 2018	To study e-bicyclists' injury severity	Police Department Database (China; 2008-2014)	Injury severity (4)	No injury, no treatment injury, treatment injury and fatality	PPO (logit)
Abegaz et al. 2014	To characterize the impact of speeding and falling asleep on crash injury severity	Crash detector data (Ethiopia; 2012-2013)	Injury severity (4)	Property damage, minor injury, serious injury and fatal	PPO (logit)
Marcoux et al. 2018	To quantify the temporal variation in exogenous factor/variable impact on crash severity	GES database (US; 1989-2014)	Injury severity (5)	No injury, possible injury, non-incapacitating, incapacitating and fatal injury	MGOR (logit)
AlKheder 2023	To identify how built environment affects traffic safety	Police reported crash data (UAE; 2008-2013)	Injury severity (4)	Minor, moderate, severe, and fatal injury	MGOR (logit)
Aidoo and Ackaah 2021	To study factors/variables influencing driver injury severity	National Road Traffic Accident Database (Ghana; 2014-2016)	Injury severity (3)	No injury, minor and fatal or seriously injured	PPO (logit)
Yasmin et al. 2012	To study the factors the severity of a crash involving emergency vehicles	Alberta Transportation and Infrastructure crash database (Canada; 1999-2008)	Injury severity (3)	Property damage, minor injury, and serious injury	PPO (logit)
Yasmin et al. 2015	To identify the various risk factors for driver fatalities	Fatality Analysis Reporting System dataset (US; 2010)	Fatality timeline (7)	6-30 days, 2-5 days and so on	MGOR (logit)
Balusu et al. 2018	To discuss potential limitations of the MGOR model	Simulated data	Injury severity (4)	No injury, non-incapacitating,	MGOR (logit)

Study	Study Objective	Data Source (Region; Year)	Dependent variable (No of Alternatives)	Alternatives	Model (logit/probit)
				incapacitating and fatal injury	
Chang et al. 2022	To analyze e-bike safety data	Incident Case Data Report (China; 2014-2016)	Injury severity (3)	Minor, major, and fatal injury	MGOR (probit)
Chen and Shen 2016	To determine built environment effects on bicyclist injury severity	Seattle Department of Transportation (US; 2004-2013)	Injury severity (4)	Property damage, possible, evident, severe injury or fatality	PPO (logit)
Balan and Paleti 2018	To develop a modified MGOR model of crash severity	GES database (US; 2014)	Injury severity (4)	No injury, possible injury, non-incapacitating and incapacitating or fatal injury	MGOR (probit)
Castro et al. 2013	To identify factors/variables affecting crash severity while accommodating spatial dependency	TxDOT Crash Records Information System (Texas, US; 2009)	Injury severity (4)	No injury, possible injury, non-incapacitating, incapacitating or fatal injury	MGOR (logit)
Fountas et al. 2018	To accommodate for the influence of unobserved heterogeneity on crash injury severity	Single-vehicle crash data (Washington, US; 2011-2013)	Injury severity (4)	No evident injury, non-incapacitating, incapacitating and fatality	OR (probit)
Michalaki et al. 2015	To identify the factors/variables affecting crash severity	Police reported crash data (Great Britain; 2005-2011)	Injury severity (3)	Slight injury, serious injury and fatality	PPO (logit)
Rahman et al. 2019	To study how new public transportation infrastructure affects transit infrastructure usage	Lynx bus ridership (Orlando, US; 2013-2015)	Bus ridership: average weekday boarding and alighting (13)	Discretized log-transformed ridership (0-5, 5-10, so on)	MGOR (logit)
Tirtha et al. 2023	To identify and quantify the impact of independent variables on airline demand	Bureau of Transportation Statistics (US; 2010-2018)	Air passenger departures and arrivals (14)	Discretized log-transformed demand (1-2, 2-4 and so on)	MGOR (probit)
Bellizzi et al. 2018	To identify the factors influencing air passenger satisfaction	Survey data (Italy; 2012-2014)	Passenger satisfaction (5)	Very poor to Excellent	OR (logit)
Alemi et al. 2019	To identify the factors/variables impacting the frequency of ride-hailing service usage	Survey data (California, US)	Ride-hailing service usage (3)	Non-user, Infrequent user, and frequent user	OR (probit)
<i>Studies in other disciplines</i>					

Study	Study Objective	Data Source (Region; Year)	Dependent variable (No of Alternatives)	Alternatives	Model (logit/probit)
Guagnano et al. 2016	To identify the association between social capital and subjective poverty	EU-SILC survey and Eurostat statistics database (Europe; 2009)	Subjective poverty (6)	Very easily, easily, fairly easily, with some difficulty, with difficulty and with great difficulty	PPO (logit)
Millán et al. 2013	To determine the factors influencing job satisfaction	European Community Household Panel (Europe; 1994-2001)	Job satisfaction (3)	Dissatisfied, neither dissatisfied nor satisfied and satisfied	PPO (logit)
Abrudan et al. 2020	To study how hotel facilities impact customer rating	Online data (Romania; 2019)	Customer rating (4)	Acceptable, medium, good, and excellent	PPO (logit)
Brida et al. 2012	To study resident's preferences in cruise tourism investment	Survey data (Italy; 2011)	Level of investment (5)	Very low, low, medium, high and very high	OR (logit)
Khedhiri et al. 2010	To examine the risk to develop bladder cancer due to smoking and metabolizing enzyme genotypes	Patient data (Tunisia)	Stage of bladder tumors (3)	3 stages	PPO (logit)
McHugh and Lake 2010	To study the factors affecting clinical nursing expertise	Statewide survey of registered nurses (Pennsylvania, US; 1999)	Level of Expertise (4)	Advanced beginner, competent, proficient and expert	PPO (logit)
Cyrenne and Chan 2022	To determine the factors determining student success in university	Administrative data (Canada; 1997)	Ordered Outcomes: Success (3)	Withdrawn, continuing and graduated	OR and PPO (logit)
Lin and Zhang 2020	To represent how COVID-19 affected agricultural exports	Survey data (China; 2020)	Change of exports (5)	Increased or no change, decreased by less than 10%, decreased between 10 and 30% and so on	OR (logit)
Islam et al. 2017	To identify the determinants of crop production loss severity	Survey data (Bangladesh)	Loss severity (3)	Low, medium, and high	OR (probit)
Zaurin et al. 2020a	To investigate how incorporating an Adaptive Learning Module (ALM) in Statics course affected student learning	Class performance data (US; 2019)	Binary outcome (2)	Withdraw/fail and pass	OR (logit)
Zaurin et al. 2020b	To compare between different instructional delivery methods	Class performance data (US; 2019)	Students' grade (6)	A through F	OR (logit)

4 Model Application

In this section, we present an application of MGOR model system using domestic airline demand data for US airports. Domestic aviation sector contributes to more than 5% of US Gross Domestic Product (FAA 2023). Analyzing the factors affecting airport level passenger demand is useful for informing facility level (e.g., runways, terminals, and intermodal transportation system) improvement and operational decisions (e.g., crew management).

Previous studies analyzed air passenger demand/flows at regional/country level (Chang 2014; Chi 2014; Kalić et al. 2014; Tsui et al. 2014; Valdes 2015; Mostafaeipour et al. 2018; Albayrak et al. 2020) and airport level (Suryani et al. 2010; Zhou et al. 2018; Li and Wan 2019; Tirtha et al. 2022). Among different mathematical approaches, linear regression model and its variants (Chang 2014; Chi 2014; Valdes 2015; Kağan Albayrak et al. 2020; Tirtha et al. 2022), autoregressive integrated moving average (Tsui et al. 2014), artificial neural networks (Mostafaeipour et al. 2018), fuzzy models (Kalić et al. 2014), gravity models (Zhou et al. 2018) and optimization techniques (Li and Wan 2019) are commonly employed to model air passenger demand. Tirtha et al. (2023) developed a joint panel generalized ordered probit model of domestic airline demand at US airports and offered a formulation to predict continuous demand using the developed model. Independent variables considered in airline demand analysis include socio-demographic factors (population, education, age distribution, median income, employment, and GDP), built environment factors (number of business centers and tourism), airport classification, level of service factors (average airfare and distance), temporal lag variables, and spillover effects (see Tirtha et al. 2023 for more details). In this chapter, we apply MGOR framework to model domestic airline demand to illustrate how the framework can be applied in practice.

The airline demand data is sourced from origin and destination survey dataset for 5 years (2010, 2012, 2014, 2016 and 2018) hosted by Bureau of Transportation Statistics (BTS) for 300 airports. The dataset provides detailed information on 10% tickets of domestic flights operated in the US. Airport level quarterly passenger departures are processed from the dataset by aggregating the departing passengers for year, quarter, and airport ID². The resulting demand dataset consists of 20 demand measures for each airport (5 years*4 quarters). Therefore, the estimation sample consists of 6,000 demand records. To generate the dependent variable, we categorize airline demand into 6 equal bins based on their quarterly ranking as follows: Bin 1: rank 251-300, Bin 2: 201-250, Bin 3: 151-200, Bin 4: 101-150, Bin 5: 51-100 and Bin 6: 1-50. Therefore, grouping of the airports has an inherent ordering: 1 – the lowest demand group and 6 – the highest demand group. This ordered dependent variable is supplemented with a number of independent variables including metropolitan statistical area (MSA) specific socio-demographic characteristics, built environment factors, spatial and temporal factors. Socio-demographic characteristics include population, household median income, employment, and educational attainment status in the corresponding MSA. Built environment factors include the number of airports in close proximity and tourism ranking of the corresponding state. Spatial factors include airport location by region (including south, north-east, west, mid-west and pacific). Temporal factors include the quarter of passenger demand data considered for the analysis. A summary of the variables considered in the specification of models is provided in Table 2. Using the processed dataset, we develop airline demand models using logistic error distribution resulting in OL, GOL, and mixed GOL

² As the dataset represents 10% tickets of domestic flights operated in the US, we appropriately scale the aggregated numbers.

formulations. In OL and GOL models, we do not consider the effect of unobserved factors on the outcome variable. In MGOL model, we consider the potential effect of unobserved factors at different repetition levels and also consider the factors modifying the impact of independent variables (random effect).

TABLE 2 Descriptive Statistics of Independent Variables (Number of airports = 300)

Variables	Description	Mean	Standard Deviation
Continuous Variables			
<i>Demographic Characteristics</i>			
Population	MSA population in million	1.430	3.241
Median Income	MSA level median income in 100K	0.544	0.112
Employment	Ln(MSA employment in thousand+1)	5.223	1.503
<i>Built Environment Factors</i>			
No. of airports	Ln(No. of airports in 50-mile buffer area+1)	1.803	0.737
Categorical Variables			
Variables	Description	Freq.	Percentage
<i>Demographic Characteristics</i>			
Educational Status			
Low	% of adults without a high school degree > 12%	2748	45.80
High	% of adults without a high school degree ≤ 12%	3252	54.20
<i>Built Environment Factors</i>			
Tourism ranking			
Top 10	Airport is located in the top 10 tourist attraction states	1900	31.67
Other	Airport is located in other states	4100	68.33
<i>Spatial Factors</i>			
Region			
South	Airport lies in the south region in the US	2160	36.00
North-East	Airport lies in the north-east region in the US	660	11.00
West	Airport lies in the west region in the US	1500	25.00
Mid-West	Airport lies in the mid-west region in the US	1340	22.33
Pacific	Airport lies in the pacific region in the US	340	5.67
<i>Temporal Factors</i>			
Quarter 1	From January to March	1500	25.00
Quarter 2	From April to June	1500	25.00
Quarter 3	From July to September	1500	25.00
Quarter 4	From October to December	1500	25.00

4.1 *Model Development*

The parameter estimates and t-statistics (in parenthesis) from OL, GOL and MGOL model are presented in Table 3, Table 4, and Table 5, respectively. To be consistent across the alternative models, we employ the formulation presented in Section 2.3. We consider a constant in the demand propensity and set the first threshold $\tau_{i,1} = 0$. The model specification begins with the estimation of an OL model. Subsequently, a GOL model is estimated allowing for additional parameters in

the thresholds. Finally, a MGOL model is estimated where the parameters are more flexible as they are allowed to vary across observations through random parameters or unobserved factors impact accommodated at the airport level (random parameter for the constant). In our estimation process, the OL model with 11 parameters offered the best fit with a LL value of -9347. The GOL model with 13 parameters offered the best fit with a LL value of -8405. The improvement in GOL model can be attributed to the two additional threshold variables for the population variable. The additional degrees of freedom allowed within the GOL framework allows for improving the model flexibility. Finally, the MGOL model with 13 parameters offered the best with a LL value of -7408. The improvement in the MGOL model can be attributed to the random effect estimated for the population variable. For sake of completeness, we also computed the Bayesian Information Criterion (BIC) values. The BIC measures for the three models are: OL 18789, GOL 16924 and MGOL 14929. Clearly, the BIC values confirm that the MGOL model offers a superior fit for our data. We will only discuss the results for MGOL model in this chapter.

4.2 *MGOL Model Results*

The results from the MGOL model are shown in Table 5. Positive (negative) value of a propensity parameter (β) indicates that an increase of the variable increases (decreases) the propensity of higher outcome levels. The threshold parameters (δ) indicate shifting of thresholds depending on sign of the parameters. Positive (negative) value of a threshold parameter indicates that an increase of the variable shifts the threshold to the right (left). For threshold formulation presented in equation (9), an increase (decrease) of a threshold increases (decreases) the probability of the alternative immediately to the left of the threshold and decreases (increases) the probability of alternatives to the right of the threshold. In the following sections, we discuss the results from MGOL model by variable group.

4.2.1 *Socio-Demographic Characteristics*

Among the socio-demographic characteristics considered, MSA specific population, median income and education status are found to have significant impact on the airport level demand variable. The results suggest that population in the corresponding MSA significantly influences air passenger demand at US airports. As expected, it is found that an increase in population is associated with an increase in the propensity of airline demand (see Zhou et al. 2018; Tirtha et al. 2022; Tirtha et al. 2023 for similar findings). However, the effect of population in MGOL framework is modified by unobserved effects as illustrated by the statistically significant standard deviation parameter. The estimated parameter distribution indicates that more than 99% of the distribution is higher than zero. We also found that MSA population influences threshold parameters. Positive coefficient of the parameter demarcating Bin 3 and Bin 4 indicates that an increase in MSA population increases the likelihood of demand category 3 and decreases the likelihood of higher order demand groups (Bin 4 and higher). Further, the results show that the air demand has significant income association. An increase in MSA median income contributes to higher demand. This result may suggest that higher affordability of the residents associated with increased median income may result in higher demand. From the results, we also observe that airline demand is significantly low if overall education status of a MSA is lower.

4.2.2 *Built Environment Factors*

From the table, we also observe that built environment factors also determine air passenger demand at US airports. If number of airports near an airport of interest increases, the probabilities of higher demand categories increase. The result might suggest that higher number of airports in an area indicates overall higher airline demand in that area. We also consider the impact of tourism on

demand. From the result, we found that if an airport is located in the top 10 tourist attraction states, the propensity of airline demand increases.

4.2.3 Spatial Factors

We consider spatial variation in airline demand using US region indicators and found that airports in the south experience higher airline demand compared to other airports.

TABLE 3 Traditional Ordered Logit Model

Variables	Latent propensity	Threshold 2-3	Threshold 3-4	Threshold 4-5	Threshold 5-6
Constant	-1.483 (-8.329)	0.084 (2.847)	-0.120 (-4.114)	-0.039 (-1.278)	0.345 (10.081)
<i>Demographic Characteristics</i>					
Population	0.417 (6.524)	--	--	--	--
Median Income	2.977 (9.105)	--	--	--	--
Education status (Base: High)					
Low	-0.547 (-10.627)	--	--	--	--
<i>Built Environment Factors</i>					
No. of airports	0.773 (15.410)	--	--	--	--
Tourism ranking (Base: Other)					
Top 10	0.220 (3.838)	--	--	--	--
<i>Spatial Factors</i>					
Region (Base: Other regions)					
South	0.910 (15.179)	--	--	--	--

TABLE 4 Generalized Ordered Logit Model

Variables	Latent propensity	Threshold 2-3	Threshold 3-4	Threshold 4-5	Threshold 5-6
Constant	-1.143 (-4.974)	1.066 (27.335)	0.342 (2.333)	0.135 (4.418)	0.516 (14.783)
<i>Demographic Characteristics</i>					
Population	0.119 (5.658)	-2.762 (-25.335)	-0.614 (-2.580)	--	--
Median Income	3.581 (7.751)	--	--	--	--
Education status (Base: High)					
Low	-0.682 (-12.121)	--	--	--	--
<i>Built Environment Factors</i>					
No. of airports	0.653 (13.626)	--	--	--	--
Tourism ranking (Base: Other)					
Top 10	0.274 (4.087)	--	--	--	--

<i>Spatial Factors</i>					
Region (Base: Other regions)					
South	0.705 (12.224)	--	--	--	--

TABLE 5 Mixed Generalized Ordered Logit Model

Variables	Latent propensity	Threshold 2-3	Threshold 3-4	Threshold 4-5	Threshold 5-6
Constant	-2.106 (-9.125)	0.396 (12.667)	0.351 (9.796)	0.865 (20.509)	1.996 (33.209)
<i>Demographic Characteristics</i>					
Population	8.425 (24.679)		0.236 (128.211)	--	--
<i>SD population</i>	3.637 (14.048)	--	--	--	--
Median Income	3.002 (7.103)	--	--	--	--
Education status (Base: High)					
Low	-1.057 (-14.036)	--	--	--	--
<i>Built Environment Factors</i>					
No. of airports	0.366 (7.022)	--	--	--	--
Tourism ranking (Base: Other)					
Top 10	0.743 (9.139)	--	--	--	--
<i>Spatial Factors</i>					
Region (Base: Other regions)					
South	0.183 (2.262)	--	--	--	--

4.3 *Policy Exercise*

As is evident from the results section, the effect of independent variables in MGOL model are complex and can only be determined by examining the variable impacts in propensity and thresholds collectively. Hence, to allow for an easier understanding of variable impacts across the various alternatives, we conducted a policy exercise. In this policy exercise, we examine the changes in alternative level aggregate probability due to changes in independent variables. For our exercise, we quantify the impact of a subset independent variables including continuous and categorical variables. Among continuous variables, we consider a 25% increase in population, median income, and number of airports. Among categorical variables, we select education status – low indicator variable and modify 25% of high-education status (indicator value=0) records and set them to 1. The changes to alternative probabilities are computed in two steps. In step one, for each record, we compute the predicted probability for each alternative for the original independent variables and aggregate the probabilities across all records. In step two, we repeat the same procedure for modified variables (as described above) and compute the new aggregate values. A percentage change measure is computed using the alternative specific aggregate values from the two computations. The result of the analysis is presented in Figure 1. From the figure, we can see

that 25% increase of population, median income and number of airports increases the probabilities of higher order categories. Therefore, as expected, the overall impact of population, median income, and number of airports on airline demand are positive. Conversely, the impact of low education status is negative as aggregate probabilities of higher order alternatives decrease due to increase in proportion of education status – low. The analysis illustrates how the results from the MGOL model can be employed for policy exercises³.

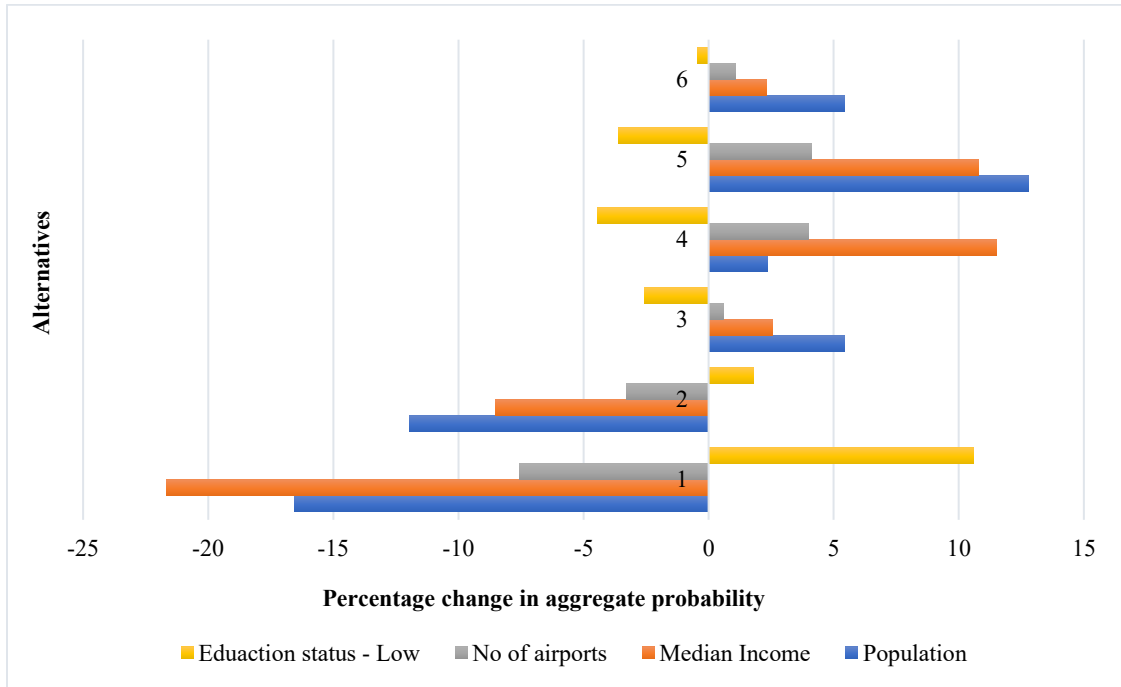


FIGURE 1 Results of Policy Exercise

5 Concluding Remarks

Ordered response models, a specific class of discrete outcome models, are widely employed for modeling ordered discrete variables. In ordered response frameworks, a uni-dimensional latent propensity equation is partitioned using a set of fixed thresholds to compute the probability of the ordered alternatives. The restrictive assumption in traditional ordered models can be relaxed by allowing for the impact of exogenous/independent variables vary across the alternative space using two formulations: Generalized Ordered Response model and Partial Proportional Odds model. In this chapter, we provide a discussion on ordered response frameworks including mathematical details of traditional ordered response model, generalized ordered response model and mixed generalized ordered response model. We also provide an application of MGOL model using air passenger demand data at US airports for multiple years. In model development, we consider

³ The dependent variable was characterized as a ranking among airports – the 300 airports were ranked into 6 equal bins of 50 airports each. The policy exercise is conducted examining how each airport probability of belonging to a particular group changes with changes to different variables. In this process, we recognize that we are examining changes assuming a static cut-off demand based on initial ranking for the original set of independent variables. In our policy exercise, we are not ranking them anymore. We are examining if the airport demand can cross the corresponding ranking threshold employed during dependent variable creation. So, after employing the model with new policy variables it is possible that all 6 categories of dependent variables will not have equal numbers (as was the case in the model estimation).

different variants of order response model including traditional ordered logit model, generalized ordered logit model and mixed generalized ordered logit model. Among the alternative models, MGOL model offers the best data fit in terms of LL and BIC measures. The results from MGOL model highlight that MSA specific socio-demographic characteristics, built environment factors and spatial factors are strongly associated with passenger demand at the US airports. Moreover, we found population standard deviation parameter significant in MGOL model highlighting how unobserved factors influence the effect of population on airline demand. The book chapter provides an illustration of how various ordered response models can be applied in practice.

References

- Abegaz T, Berhane Y, Worku A, Assrat A, Assefa A (2014) Effects of excessive speeding and falling asleep while driving on crash injury severity in Ethiopia: A generalized ordered logit model analysis. *Accid Anal Prev* 71:15–21.
- Abrudan IN, Pop CM, Lazăr PS (2020) Using a general ordered logit model to explain the influence of hotel facilities, general and sustainability-related, on customer ratings. *Sustain* 12(21):1–21.
- Aidoo EN, Ackaah W (2021) A generalized ordered logit analysis of risk factors associated with driver injury severity. *J Public Heal* 29(2):471–477.
- Alemi F, Circella G, Mokhtarian P, Handy S (2019) What drives the use of ridehailing in California? Ordered probit models of the usage frequency of Uber and Lyft. *Transp Res Part C Emerg Technol* 102(May 2018):233–248.
- AlKheder S (2023) Studying the effect of built environment on traffic safety with random parameter and generalized ordered logit models. *Urban Clim* 47(December 2022):101388.
- Balan L, Paleti R (2018) Modified mixed generalized ordered response model to handle misclassification in injury severity data. *Transp Res Rec* 2672(30):53–63.
- Balusu SK, Pinjari AR, Mannering FL, Eluru N (2018) Non-decreasing threshold variances in mixed generalized ordered response models: A negative correlations approach to variance reduction. *Anal Methods Accid Res* 20(October):46–67.
- Bellizzi MG, Eboli L, Forciniti C, Mazzulla G (2018) Air Transport Passengers' Satisfaction: An Ordered Logit Model. *Transp Res Procedia* 33:147–154.
- Bhat CR (2001) Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transp Res Part B Methodol* 35(7):677–693.
- Bhat CR, Eluru N, Copperman RB (2008) Flexible Model Structures for Discrete Choice Analysis. In: D.A. Hensler, Button KJ (eds) *Handbook of Transport Modelling*, 2nd edn. Elsevier Science, pp 75–104.
- Brida JG, Del Chiappa G, Meleddu M, Pulina M (2012) Cruise tourism externalities and residents' support: A mixed approach. *Economics* 6. <https://doi.org/10.5018/economics-ejournal.ja.2012-40>
- Castro M, Paleti R, Bhat CR (2013) A spatial generalized ordered response model to examine highway crash injury severity. *Accid Anal Prev* 52:188–203.
- Chang F, Haque MM, Yasmin S, Huang H (2022) Crash injury severity analysis of E-Bike Riders: A random parameters generalized ordered probit model with heterogeneity in means. *Saf Sci* 146(April 2021):105545.
- Chang LY (2014) Analysis of bilateral air passenger flows: A non-parametric multivariate adaptive regression spline approach. *J Air Transp Manag* 34:123–130.

- Chen P, Shen Q (2016) Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. *Accid Anal Prev* 86:239–246.
- Chi J (2014) A cointegration analysis of bilateral air travel flows: The case of international travel to and from the United States. *J Air Transp Manag* 39:41–47.
- Cyrenne P, Chan A (2022) The Determinants of Student Success in University: A Generalized Ordered Logit Approach. *Can J High Educ* 52(1):84–107.
- Eluru N, Bhat CR, Hensher DA (2008) A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accid Anal Prev* 40(3):1033–1054.
- Eluru N, Yasmin S (2015) A note on generalized ordered outcome models. *Anal Methods Accid Res* 8:1–6.
- FAA (2023) Air Traffic By The Numbers. https://www.faa.gov/air_traffic/by_the_numbers/. Accessed 02 Feb 2023
- Fountas G, Anastasopoulos PC, Mannering FL (2018) Analysis of vehicle accident-injury severities: A comparison of segment- versus accident-based latent class ordered probit models with class-probability functions. *Anal Methods Accid Res* 18:15–32.
- Guagnano G, Santarelli E, Santini I (2016) Can Social Capital Affect Subjective Poverty in Europe? An Empirical Analysis Based on a Generalized Ordered Logit Model. *Soc Indic Res* 128(2):881–907.
- Islam Z, Alauddin M, Sarker MAR (2017) Determinants and implications of crop production loss: An empirical exploration using ordered probit analysis. *Land use policy* 67(May):527–536.
- Kağan Albayrak MB, Özcan İÇ, Can R, Dobruszkes F (2020) The determinants of air passenger traffic at Turkish airports. *J Air Transp Manag* 86(February). <https://doi.org/10.1016/j.jairtraman.2020.101818>
- Kalić M, Kuljanin J, Dožić S (2014) Air travel demand fuzzy modelling: Trip generation and trip distribution. *Adv Intell Syst Comput* 223:279–289.
- Kaplan S, Prato CG (2012) Risk factors associated with bus accident severity in the United States: A generalized ordered logit model. *J Safety Res* 43(3):171–180.
- Khedhiri S, Stambouli N, Ouerhani S, Rouissi K, Marrakchi R, Gaaied AB, Slama MB (2010) The impact of smoking and polymorphic enzymes of xenobiotic metabolism on the stage of bladder tumors: A generalized ordered logistic regression analysis. *J Cancer Res Clin Oncol* 136(7):1111–1116.
- Li T, Wan Y (2019) Estimating the geographic distribution of originating air travel demand using a bi-level optimization model. *Transp Res Part E Logist Transp Rev* 131(October):267–291.
- LIN B xi, ZHANG YY (2020) Impact of the COVID-19 pandemic on agricultural exports. *J Integr Agric* 19(12):2937–2945.
- Marcoux R, Yasmin S, Eluru N, Rahman M (2018) Evaluating temporal variability of exogenous variable impacts over 25 years: An application of scaled generalized ordered logit model for driver injury severity. *Anal Methods Accid Res* 20(September):15–29.
- McHugh MD, Lake ET (2010) Understanding clinical expertise: nurse education, experience, and the hospital context. *Res Nurs Health* 33(4):276–287.
- Michalaki P, Quddus MA, Pitfield D, Huetson A (2015) Exploring the factors affecting motorway accident severity in England using the generalised ordered logistic regression model. *J Safety Res* 55:89–97.
- Millán JM, Hessels J, Thurik R, Aguado R (2013) Determinants of job satisfaction: A European comparison of self-employed and paid employees. *Small Bus Econ* 40(3):651–670.

- Mostafaeipour A, Goli A, Qolipour M (2018) Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: a case study. *J Supercomput* 74(10):5461–5484.
- Peterson B, Harrell FEJ (1990) Partial Proportional Odds Models for Ordinal Response Variables Author (s): Bercedis Peterson , Frank E . Harrell and Jr Published by : Wiley for the Royal Statistical Society Stable URL : <https://www.jstor.org/stable/2347760> Partial Proportional Odds M. Wiley *R Stat Soc* 39(2):205–217
- Rahman M, Yasmin S, Eluru N (2019) Evaluating the impact of a newly added commuter rail system on bus ridership: a grouped ordered logit model approach. *Transp A Transp Sci* 15(2):1081–1101.
- Srinivasan KK (2002) Injury severity analysis with variable and correlated thresholds: Ordered mixed logit formulation. *Transp Res Rec* (1784):132–142.
- Suryani E, Chou SY, Chen CH (2010) Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework. *Expert Syst Appl* 37(3):2324–2339.
- Terza JV (1985) Ordinal probit: a generalization. *Communications in Statistics-Theory and Methods* 14(1): 1–11.
- Tirtha SD, Bhowmik T, Eluru N (2022) An airport level framework for examining the impact of COVID-19 on airline demand. *Transp Res Part A Policy Pract* 159(March):169–181.
- Tirtha SD, Bhowmik T, Eluru N (2023) Understanding the factors affecting airport level demand (arrivals and departures) using a novel modeling approach. *J Air Transp Manag* 106(September 2022):102320.
- Tsui WHK, Ozer Balli H, Gilbey A, Gow H (2014) Forecasting of Hong Kong airport’s passenger throughput. *Tour Manag* 42(2014):62–76.
- Valdes V (2015) Determinants of air travel demand in Middle Income Countries. *J Air Transp Manag* 42:75–84.
- Wang T, Chen J, Wang C, Ye X (2018) Understand e-bicyclist safety in China: Crash severity modeling using a generalized ordered logit model. *Adv Mech Eng* 10(6):1–11.
- Yasmin S, Anowar S, Tay R (2012) Effects of drivers’ actions on severity of emergency vehicle collisions. *Transp Res Rec* 1(2318):90–97.
- Yasmin S, Eluru N (2013) Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accid Anal Prev* 59:506–521.
- Yasmin S, Eluru N, Bhat CR, Tay R (2014) A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. *Anal Methods Accid Res* 1:23–38.
- Yasmin S, Eluru N, Pinjari AR (2015) Analyzing the continuum of fatal crashes: A generalized ordered approach. *Anal Methods Accid Res* 7:1–15.
- Zaurin R, Tirtha SD, Eluru N (2020a) A comparative analysis of the students’ performance in two statics courses due to the inclusion of an Adaptive Learning Module (ALM) to review the mathematics pre-requisite knowledge. *ASEE Annu Conf Expo Conf Proc* 2020-June. <https://doi.org/10.18260/1-2--33982>
- Zaurin R, Tirtha SD, Eluru N (2020b) A comparison between mixed-mode and face-to-face instructional delivery approaches for engineering analysis: Statics. *ASEE Annu Conf Expo Conf Proc* 2020-June. <https://doi.org/10.18260/1-2--33985>
- Zhou H, Xia J (Cecilia), Luo Q, Nikolova G, Sun J, Hughes B, Kelobonye K, Wang H, Falkmer T (2018) Investigating the impact of catchment areas of airports on estimating air travel demand: A case study of regional Western Australia. *J Air Transp Manag* 70(May):91–103.

