Accommodating Spatio-Temporal Dependency in Airline Demand Modeling

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

Tanmoy Bhowmik
Assistant Professor
Department of Civil and Environmental Engineering
Portland State University
Tel: 1-503-725-7096
Email: t.bhowmik@pdx.edu
ORCiD number: 0000-0002-0258-1692

Naveen Eluru
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

Submitted to: Journal of Air Transport Management
Submission Date: February 16, 2024

Declaration of Interest: None
Funding Source: None
Data Statement: The research data will be available upon request.

*Corresponding author

ABSTRACT
The objective of the current study is to examine monthly air passenger departures at the airport level considering spatial interactions between airports. In this study, we develop a novel spatial grouped generalized ordered probit (SGGOP) model system of monthly air passenger departures at the airport level. Specifically, we estimate two variants of spatial models including spatial lag model and spatial error model. In the presence of repeated demand measures for the airports, we also consider temporal variations of spatial correlation effects among proximally located airports by employing space and time-based weight matrix. The proposed model is estimated using monthly air passenger departures for five years for 369 airports across the US. The proposed spatial model is implemented using composite marginal likelihood (CML) approach that offers a computationally feasible framework. From the estimation results, it is evident that air passenger departures at the airport level are influenced by different factors including MSA specific demographic characteristics, built environment characteristics, airport specific factors, spatial factors, and temporal factors. Moreover, spatial autocorrelation parameter is found to be significant validating our hypothesis of the presence of common unobserved factors associated with the spatial unit of analysis. In this study, we also perform a validation analysis to examine the predictive performance of the proposed spatial models. The results highlight the superiority of spatial error model compared to spatial lag model and the independent model that ignores the spatial interactions. Finally, we undertake an elasticity analysis to quantify the impact of the independent variables.

Keywords: Airline Demand; Spatial Interaction; Spatial Lag Model; Spatial Error Model; Weight Matrix

1 INTRODUCTION

Domestic airline industry plays an important role in the US economy. The net annual revenue of $488 billion from this industry contributes to 5.2% of US GDP (FAA, 2023). Domestic airline industry is interconnected with other sectors of the economy such as tourism, lodging, and related auxiliary business (Tirtha et al., 2023). To understand the health of this industry, air passenger demand serves as an important indicator. Domestic air passenger demand increased significantly by 31% between 2010 and 2019. However, due to the outbreak of novel Coronavirus, domestic airline industry experienced a sharp drop of 41% in 2020 compared to the previous year. As the country recovers from the pandemic, understanding the factors influencing airport demand is important for several reasons. First, analyzing air passenger demand is an integral part of long-term policy making such as airport runway and terminal design and expansion, intermodal transportation facilities. Second, understanding airline demand guides operational decisions for airport services such as aircraft and crew management. Finally, analyzing airline demand and identifying its contributing factors will allow us to build a template of possible demand recovery path in future months.

Given the importance of understanding airline demand, earlier studies examined airline demand at different spatial (airport level and regional level) and temporal (year, quarter, and month) resolutions. Traditionally, airports are mapped to spatial units such as metropolitan statistical area (MSA), county, or region in airport-level demand analysis. In such studies, characteristics of spatial unit of analysis including socio-demographics (population, education, age distribution), socio-economic factors (income, unemployment rate, GDP), built environment characteristics (number of trade centers, tourist attractions), level of service factors (average air fare and distance) and lag variables (historical demand) are considered to affect airline demand. In addition to these observed factors of airline demand, several unobserved factors associated with the spatial unit can possibly influence airport-level demand. For instance, consider multiple airports in proximally located MSAs. It is plausible that observed characteristics of these MSAs such as population and employment can impact demand across these airports. These impacts can be considered by generating these variables considering larger catchment areas for demand prediction (as opposed to using MSA attributes only). In addition, there might be some unobserved factors associated with closely linked spatial units that may cause demand correlation among the airports. For example, closer airports share passenger behavior trends that are less likely to be captured by attributes. For example, variations across how pandemic guidelines were considered and implemented is likely to be similar within proximal airports. Neglecting the presence of such unobserved spatial correlations in demand modelling may result in biased estimates.

Earlier research efforts on airline demand modeling have neglected to adequately consider for spatial interactions between air passenger demand at multiple airports. The main objective of this study is to analyze monthly air passenger departures at an aggregate level of airport while accommodating for spatial and temporal interactions (observed and unobserved). To achieve this goal, airline demand data for 5 years (2010, 2012, 2014, 2016 and 2018) sourced from the Bureau of Transportation Statistics (BTS) is employed to model monthly air passenger departures at the airport level. Air passenger demand data is augmented with several exogenous attributes including Metropolitan Statistical Area (MSA) specific demographic characteristics, built environment characteristics, airport specific factors, spatial factors, and temporal factors. The proposed research effort allows us to examine the impact of these aforementioned factors on airline demand while incorporating the spatial dependencies between spatially linked airports. Traditional approaches employing linear regression frameworks inherently impose a linear restriction on parameter impacts for independent variables. While these restrictions can be addressed to some extent by considering indicator variables and/or polynomial terms, the restrictions still exist. In this research, we discretize continuous log-transformed air passenger departures' variable into a categorical dependent variable with a series of ordinal levels (≤6, >6-7, >7-8, >8-9, >9-10, >10-11, >11-12, >12-13, >13-14, and >14). We recast the recently developed grouped generalized ordered probit (GGOP) framework to model the ordinal airline demand[1] variable. In previous research efforts, it has been shown that the proposed non-linear system subsumes the traditional linear regression model system (see Tirtha et al., 2023 for more details). The approach estimated on discrete bins still allows us to predict continuous airline demand similar to the traditional linear regression model (see Tirtha et al., 2023 for the details of prediction mechanism).

In the GGOP framework, we accommodate for spatial correlations among the airports. We consider two variants of spatial models, namely spatial lag model and spatial error model in our study. The spatial lag model incorporates the correlation using the dependent variables at multiple airports (excluding the current airport) in the form of spatially lagged dependent variables. The spatial error model captures the correlation using the error terms through the autocorrelated error term. Further, as we are considering spatial models in the discrete outcome paradigm, maximum likelihood approaches are infeasible (see Bhat et al., 2010 for a discussion). In the presence of complex spatial and temporal dependencies across observations, it is very difficult to estimate the model using full likelihood approach. Hence, we draw on recent advances in spatial econometrics employing composite maximum likelihood (CML) method to examine airline demand. In the CML approach, we maximize a surrogate log-likelihood function by computing pairwise joint probabilities of the observations. The GGOP model with CML is estimated using a host of independent variables including demographic characteristics, built environment characteristics, airport specific factors, spatial factors, and temporal factors. The model results offer intuitive and useful insights on airline demand. Finally, a validation exercise is conducted to present the value of the proposed models by comparing them with the traditional model that does not consider any spatial dependency.

The rest of the paper is organized as follows: Section 2 describes relevant earlier research and positions the current study. Section 3 presents the modeling approach employed in the research. Next, we describe the dataset employed in this study in Section 4. Section 5 presents the model estimation results. In Section 6, we undertake a validation exercise to compare predictive performance of alternative model frameworks. In Section 7, we perform an elasticity analysis to quantify the impact of independent variables. Finally, concluding remarks are included in Section 8.

2 LITERATURE REVIEW AND CURRENT STUDY

2.1 Earlier Studies

The literature review in the current study context can be categorized into two major streams: a) studies identifying key factors of airline demand, b) studies developing spatial panel models across transportation domains considering dependency between the spatial unit of analysis.

The first group of studies analyzing airline demand provides useful insights on the factors affecting airline demand. From the review, previous studies can be categorized into several streams by spatial resolution and method of analysis. In terms of spatial resolution, we can divide earlier studies into two major categories: disaggregate resolution and aggregate resolution. In

[1] The reader should note that "airline demand" represents monthly airport level domestic air passenger departures in the US.

disaggregate level studies, airline demand is examined mostly at the airport level. For example, Li and Wan (2019); Suryani et al. (2010) and Loo et al. (2005) developed airline demand models at the disaggregate resolution of airport. On the other hand, spatial resolution includes metropolitan statistical area (MSA), region or country in aggregate level analysis. Sample studies such as Chen et al. (2009); Chang (2014) and Abed et al. (2001) analyzed airline demand at the aggregate level. Airline demand analyses by earlier studies have employed a wide range of methodological approaches including regression analysis, artificial neural network, autoregressive moving average approach, gravity model and optimization models. For example, Tirtha et al. (2022); Valdes (2015) and Chi (2014) employed linear regression model and its variants to model airline demand. Mostafaeipour et al. (2018) employed artificial neural network to analyze airline demand. Xu et al. (2019) and Tsui et al. (2014) developed airline demand models using autoregressive moving average method. Zhou et al. (2018); Grosche et al. (2007) and Matsumoto (2004) identified pairwise OD demand using gravity model. Li and Wan (2019); Li et al. (2013) and Loo et al. (2005) employed optimization techniques to model airline demand. Finally, independent variables considered in the above-mentioned studies include socio-demographics (population, education, age distribution), socio-economic factors (income, unemployment rate, GDP), built environment characteristics (number of trade centers, tourist attractions), level of service factors (average air fare and distance) and lag variables (historical demand).

The second group of studies include research efforts identifying spatial dependencies between the spatial unit of analysis in the modelling approach. In terms of dependent variables, the studies considered cover a wide range of topics in transportation research domain including transportation demand modeling (Faghih-Imani and Eluru, 2016; Rahman et al., 2021), impact of transportation infrastructure on regional/agricultural growth (Tong et al., 2013; Yu et al., 2013; Chen and Haynes, 2015), land use modeling (Wang and Kockelman, 2006; Carrión-flores et al., 2009; Chakir and Parent, 2009; Ferdous and Bhat, 2013), crash injury severity modeling (Castro et al., 2013), recreational activity modeling (Bhat et al., 2010), and airfare analysis (Daraban and Fournier, 2008). Dependent variables in such studies can be categorized as either a continuous variable (Daraban and Fournier, 2008; Tong et al., 2013; Yu et al., 2013; Chen and Haynes, 2015; Faghih-Imani and Eluru, 2016; Rahman et al., 2021), or a categorical variable (Wang and Kockelman, 2006; Carrión-flores et al., 2009; Chakir and Parent, 2009; Bhat et al., 2010; Castro, Paleti and Bhat, 2013; Ferdous and Bhat, 2013). The aforementioned studies employ different variants of spatial models to capture spatial correlations including spatial lag or spatial autoregressive model (SAR) (Wang and Kockelman, 2006; Daraban and Fournier, 2008; Carrión-flores et al., 2009; Chakir and Parent, 2009; Lee and Yu, 2010; Castro et al., 2013; Ferdous and Bhat, 2013; Chen and Haynes, 2015; Faghih-Imani and Eluru, 2016; Rahman et al., 2021), spatial intermediate model (Castro et al., 2013), spatial error model (SEM) (Bhat et al., 2010; Castro et al., 2013; Chen and Haynes, 2015; Faghih-Imani and Eluru, 2016; Rahman et al., 2021), and Spatial Dublin model (SDM) (Tong et al., 2013; Yu et al., 2013; Chen and Haynes, 2015). A majority of these modeling approaches require a spatial weight matrix representing the spatial arrangements of the analysis units to incorporate spatial correlation among the units. Spatial weight matrices are generally formed based on pairwise distance between the spatial units. Various types of formulation of the weight matrix elements include neighborhood/within distance threshold indicator (Yu et al., 2013; Faghih-Imani and Eluru, 2016; Rahman et al., 2021), inverse of distance squared (Daraban and Fournier, 2008; Ferdous and Bhat, 2013), inverse of distance cubed (Castro et al., 2013), and inverse of exponential of distance (Ferdous and Bhat, 2013). In case of panel data, distance-based weight matrix may need some modifications to capture changes of spatial dependency effect over time. To consider for such temporal variability, Wang and Kockelman (2006) formulated spatial weight matrix as a function of distance and time difference. However, earlier research efforts analyzing spatially correlated discrete dependent variables indicated increased complexity in model estimation. In presence of complex correlation between the observations, full likelihood approach might be infeasible especially for discrete outcome variables. Therefore, earlier studies emphasized the application of methods estimating surrogate likelihood measures such as composite marginal likelihood method (CML) (see Bhat et al., 2010; Castro et al., 2013; Ferdous and Bhat, 2013 for more details) and Markov Chain Monte Carlo (MCMC) method (see Chakir and Parent, 2009 for more details).

2.2 The Current Study in the Context

While earlier studies in airline literature examined the impact of key factors on airline demand (as examined in Tirtha et al, 2023), spatial interaction between the airports has not been sufficiently considered in the demand analysis. The current study addresses this gap by developing a novel spatial grouped generalized ordered probit (SGGOP) model system of monthly air passenger departures at the airport level that explicitly accommodates the spatial interactions of the proximally located airports. The current study is the first attempt to accommodate spatial correlation in a grouped generalized ordered model framework. Further, we formulate weight matrix as a function of distance between the airports and temporal difference (measured in months) to capture temporal variation in spatial dependency.

In this study, we categorize log-transformed monthly air passenger departures into ten demand groups (≤6, >6-7, >7-8, >8-9, >9-10, >10-11, >11-12, >12-13, >13-14, and >14) and employ the recently developed GGOP model system to model the discretized dependent variable. The proposed grouped response model is a hybrid system that ties a continuous demand variable to a categorical demand variable. The proposed GGOP model system is analogous to the linear regression model system without the restrictions of linear regression (Tirtha et al., 2020; Tirtha et al., 2023). In addition, the proposed model system recognizes that there can be spatial correlations in the error terms of demand propensity of the spatially linked airports. In this study, we estimate two variants of spatial models including spatial lag model and spatial error model. In presence of repeated demand measures at the airport level, it is possible that spatial correlations between the observational units may vary over time. Therefore, we formulate weight matrix as a function of shortest geodesic distance between the airports and the absolute value of time difference (measured in months). The approach we followed in this study allows correlation between observations varying across both space and time (see Wang and Kockelman, 2006 for a similar approach). In the model development, we employ various functional forms of weight matrix (such as the inverse of square root of distance × time, the inverse of distance × time, and the inverse of distance × time squared) and select the best formulation based on data fit. In our analysis, we restrict spatial correlation to be present only within a distance and time threshold considering as the dependency is negligible between observations far apart in terms of space and time. The proposed spatial model is implemented using composite marginal likelihood (CML) approach that is easier compared to full likelihood approach due to the presence of complex spatial dependencies among the observations. Further, we perform spatial data enhancement by considering a large set of airports across the US to accommodate the effects of different spatial factors in the analysis. Finally, we compare the performance of spatial lag model and spatial error model with the traditional model without spatial effects to highlight the importance of accommodating spatial correlations while modeling airline demand at the airport level. It is important to recognize that model systems that ignore for the presence of spatial and temporal correlations when they exist are likely to result in inaccurate model estimates (Chamberlain, 1980; Bhat, 2001). The errors can spill-over into any policy analysis conducted using the independent models. For example, if the population or employment impact on airline demand are incorrect, a potential scenario analysis will under-predict or over-predict the influence of these variables thus affecting policy decisions.

In this study, airline demand data is sourced from T-100 Domestic Market (U.S. Carrier) dataset compiled by Bureau of Transportation Statistics (BTS). The demand dataset employed in this study includes monthly air passenger departure rate for 369 airports across the US for 5 annual time points (2010, 2012, 2014, 2016, and 2018). Airline demand data is further augmented with a comprehensive set of independent variables including a) demographic characteristics (population, median income, employment, and vehicle ownership level), b) built environment characteristics (number of airports in close proximity, and state level tourism ranking), c) airport specific factors (airport classification such as core airports, and Operational Evolution Partnership (OEP-35) airports), d) spatial factors (region of the airports), and e) temporal factors (month of analysis).

3 ECONOMETRIC METHODOLOGY

In this section, we first present the details of grouped generalized ordered probit (GGOP) model without considering any spatial dependencies between the airports. In the subsequent sub-sections, we present the formulations of spatial lag and spatial error GGOP models, respectively. Finally, we present model estimation procedure.

3.1 Grouped Generalized Ordered Probit Model

Let *k* (*k*= 1, 2,…, *K=369*) be an index to represent airports, *t* (*t* = 1, 2, 3,…, *T*) represents the different years, *m* (*m*=1, 2, 3,…., *M*) represents different months of a year and *j* (*j* = 1, 2, 3,…, *J = 10*) be an index to represent the bins for the logarithm of monthly passenger departures. We consider ten categories for the air travel demand analysis and these categories are: Bin 1 = ≤6; Bin 2 = 6-7; Bin 3 = 7-8, Bin 4 = 8-9, Bin 5 = 9-10, Bin 6 = 10-11, Bin 7 = 11-12, Bin 8 = 12-13, Bin 9 = 13-14, and Bin 10 = >14. For ease of formulation, we express each observational unit as an unique combination of airport *k*, year *t*, and month m, using *q* (*q* = 1, 2, 3,…., Q). Then, the equation system for modeling demand may be written as follows:

$$y_q^* = \alpha' x_q + \varepsilon_q, y_q = j \ \ if \ \psi_{j-1} < y_q^* \le \psi_j \tag{1}$$

In Equation 1, y_q^* is the continuous latent propensity for total airline demand at airport *k*, for the year *t* and month *m*. This latent propensity y_q^* is mapped to the actual demand category *j* by the ψ thresholds, in the usual ordered-response modeling framework. In our case, we consider J = 10 and thus the 11 ψ values are as follows: -∞, 6, 7, 8, 9, 10, 11, 12, 13, 14, and +∞. x_q is a matrix of attributes that influence passenger departures (including the constant); α is the vector of coefficients corresponding to the attributes. Further, ε_q is an idiosyncratic random error term assumed independently normally distributed with variance λ^2.

The variance vector for passenger departures is parameterized as a function of independent variables as follows: $\lambda_q = \exp(\theta' x_q)$. The parameterization allows for the variance to be different across the airports accommodating for heteroscedasticity[2]. Finally, to allow for alternative specific

[2] Elements of error variance function do not include a constant as estimation result confirms strong correlation between the constant and spatial correlation parameter.

effects, we also introduce threshold specific deviations in the model as $\rho_j = \tau'_j x_q$. Here, τ_j is a vector of coefficients and x_q is a set of independent variables including constant. If ρ_j is positive, the threshold demarcating alternatives *j-1* and *j* shifts to the left and the probability of the lower-level category (higher level category) decreases (increases).

The probability for airport *k* to have departures in category *j* is given by:

$$P(y_q = j_q) = \Lambda\left[\frac{\psi_{q,j}-(\alpha' x_q+\rho'_{q,j})}{\lambda_q}\right] - \Lambda\left[\frac{\psi_{q,j-1}-(\alpha' x_q+\rho'_{q,j-1})}{\lambda_q}\right] \quad (2)$$

where Λ (.) is the cumulative standard normal distribution.

3.2 Spatial Lag GGOP Model

The spatial lag formulation includes spatial correlation in the latent propensity of airline demand presented in Equation 1 as follows (Castro et al., 2013):

$$y_q^* = \delta \sum_{q'=1}^{Q} w_{qq'} y_{q'}^* + \alpha' x_q + \varepsilon_q, y_q = j \ \ if \ \psi_{j-1} < y_q^* \leq \psi_j \quad (3)$$

Where, $w_{qq'}$ is an element of an exogenously defined distance-month difference based space and time weight matrix $\mathbf{W}$ calculated based on locations and month of analysis for airport k and k' (with $w'_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$), and δ $(0 < \delta < 1)$ is the spatial autoregressive parameter. For example, distance between two airports, A and B is 50 miles and months of analysis are January 2016 and June 2014. Therefore, month difference between the observations is 19 and we add 1 to the difference (=20) to ensure the denominator does not become 0 for spatial records in the same time period. In space-time weight matrix **W**, we employ different functional forms of $w_{qq'}$ including 1/$\sqrt{distance \times month}$, 1/(distance×$\sqrt{month}$), 1/(distance×ln(month+1)), 1/(distance×month), and 1/(distance×month)^2. Further, we restrict 3 groups of elements of W to be zero: a) diagonal elements to avoid self-inclusion of the observations ($w_{qq} = 0$), b) off-diagonal elements for same airport (as we have repeated records for the airports), and c) any element for future time points where month difference is negative. Then, we perform row normalization of W matrix to restrict $\sum_{q'} w_{qq'} = 1$.

Finally, to restrict δ between 0 and 1, we represent δ using a function: $\frac{e^{\delta'}}{1+e^{\delta'}}$ and estimate the parameter δ'. The latent demand propensity presented in Equation 3 can be re-written using vector notation as follows:

$$\boldsymbol{y}^* = \boldsymbol{\delta}\mathbf{W}\boldsymbol{y}^* + \mathbf{x}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \quad (4)$$

Now, the Equation 4 can be re-written as follows (Castro et al., 2013):

$$\boldsymbol{y}^* = \boldsymbol{S}(\mathbf{x}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}) \quad (5)$$

where $\mathbf{S} = [\mathbf{I}_Q - \delta\mathbf{W}]^{-1}$ is a ($Q\times Q$) matrix and $\mathbf{I}_Q$ is an identity matrix of size Q. The vector $\boldsymbol{y}^*$ is multivariate normally distributed as, $y^* \sim MVN_Q(\boldsymbol{B_{lag}}, \boldsymbol{\Sigma_{lag}})$. We represent $\boldsymbol{B_{lag}}$ and $\boldsymbol{\Sigma_{lag}}$ as follows:

$$\boldsymbol{B_{lag}} = \mathbf{S}\mathbf{x}\boldsymbol{\alpha} \text{ and } \boldsymbol{\Sigma_{lag}} = \mathbf{S}\mathbf{I_Q}\boldsymbol{S}' \quad (6)$$

3.3 Spatial Error GGOP Model

In spatial error model formulation, continuous latent propensity is expressed as follows (Castro et al., 2013):

$$y_q^* = \delta \sum_{q'=1}^{Q} w'_{qq'}\varepsilon_{q'} + \alpha' x_q + \varepsilon_q, y_q = j \;\; if\; \psi_{j-1} < y_q^* \leq \psi_j \quad (7)$$

Now, vector representation of Equation 7 is as follows:

$$\boldsymbol{y}^* = \boldsymbol{\delta}\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{x}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \quad (8)$$

We can re-write Equation 8 as follows:

$$\boldsymbol{y}^* = \mathbf{x}\boldsymbol{\alpha} + \boldsymbol{S\varepsilon} \quad (9)$$

The vector $\boldsymbol{y}^*$ is multivariate normally distributed as, $y^* \sim MVN_Q(\boldsymbol{B_{error}}, \boldsymbol{\Sigma_{error}})$. We represent $\boldsymbol{B_{error}}$ and $\boldsymbol{\Sigma_{error}}$ as follows:

$$\boldsymbol{B_{error}} = \mathbf{x}\boldsymbol{\alpha} \text{ and } \boldsymbol{\Sigma_{error}} = \boldsymbol{SS}' \quad (10)$$

3.4 Model Estimation

The vector of parameters to be estimated in both spatial lag and spatial error GGOP model is $\gamma = (\alpha', \rho_j', \lambda, \delta')$. While full likelihood approach is infeasible in presence of complex dependencies between the observations, composite marginal likelihood (CML) approach is simpler which is based on maximizing surrogate likelihood function. In this study, we follow pairwise CML method to compute log-composite likelihood as follows (see Castro et al., 2013 for similar formulation):

$$L_{CML}(\gamma) = \prod_{q=1}^{Q} \prod_{q'=1, q'\neq q}^{Q} \Pr(y_q = z_q, y_{q'} = z_{q'})$$
$$= \prod_{q=1}^{Q} \prod_{q'=1, q'\neq q}^{Q} [\Phi(\varphi_q, \varphi_{q'}, \nu_{qq'}) - \Phi(\varphi_q, \mu_{q'}, \nu_{qq'}) - \Phi(\mu_q, \varphi_{q'}, \nu_{qq'}) + \Phi(\mu_q, \mu_{q'}, \nu_{qq'})] \quad (11)$$

Where, $\varphi_q = \frac{\psi_{q,z} - ([\boldsymbol{B}]_q + \rho'_{q,z})}{\sqrt{\lambda_q * [\boldsymbol{\Sigma}]_{qq}}}$, $\mu_q = \frac{\psi_{q,z-1} - ([\boldsymbol{B}]_q + \rho'_{q,z-1})}{\sqrt{\lambda_q * [\boldsymbol{\Sigma}]_{qq}}}$, $\nu_{qq'} = \frac{[\boldsymbol{\Sigma}]_{qq'}}{\sqrt{[\boldsymbol{\Sigma}]_{qq} * [\boldsymbol{\Sigma}]_{q'q'}}}$ and $z_q(z_1, z_2, \ldots, z_Q)$ is the observed demand level.

In computing marginal likelihood function presented in Equation 11, we need to calculate $Q(Q-1)$numbers of joint probabilities. In Equation 11, $\nu_{qq'}$ represents correlation parameter in bivariate normal cumulative density function which is stronger for observations in close proximity in terms of time and space. $\nu_{qq'}$ is considerably small for observations with larger distance and month difference. In this study, we assume that spatial correlations are present within a distance band and a time threshold. Based on spatial distribution of the airports, we select 100 miles as the distance band and 36 months as the time threshold for our study. Therefore, we can re-write Equation 11 as follows:

$$L_{CML}(\gamma) = \prod_{q=1}^{Q} \prod_{q'=1, q'\neq q}^{Q} \Pr(y_q = z_q, y_{q'} = z_{q'})$$
$$= \prod_{q=1}^{Q} \prod_{q'=1, q'\neq q}^{Q} [\Phi(\varphi_q, \varphi_{q'}, R_{qq'}\nu_{qq'}) - \Phi(\varphi_q, \mu_{q'}, R_{qq'}\nu_{qq'}) - \Phi(\mu_q, \varphi_{q'}, R_{qq'}\nu_{qq'}) + \Phi(\mu_q, \mu_{q'}, R_{qq'}\nu_{qq'})] \quad (12)$$

Where, $R_{qq'} = 1\ if\ d_{qq'} \leq 100$miles and $0 < m_{qq'} \leq 36$months, 0 otherwise

In above Equation 12, $R_{qq'}$ is a dummy variable indicating the presence of spatial correlation between a pair of airports. $d_{qq'}$ and $m_{qq'}$ represent distance in miles and time difference in months between observations, q and q'. Finally, covariance matrix of the parameters is estimated by the inverse of Godambe's (Godambe, 1960) sandwich information matrix (see Bhat et al., 2010; Castro et al., 2013 for the details of covariance matrix).

4 DATASET DESCRIPTION

The airline demand data is sourced from T-100 Domestic Market (U.S. Carrier) dataset compiled by Bureau of Transportation Statistics. The domestic market dataset contains number of passengers carried by domestic carriers for each airport for each month. In this study, we analyze monthly air passenger departure rate at the airport level for five annual time points (2010, 2012, 2014, 2016, and 2018). Hence, we aggregate air passenger departures for each airport and each month in the analysis period. Initially, we selected 510 airports across the US from five major regions including South, West, Mid-West, North-East, and Pacific regions. Then, we remove all smaller airports having missing demand records. After removing the airports with missing records, we retain 369 airports resulting in a sample of 22,140 observations (369 airports * 60 months). In preparation of estimation sample, we randomly select 5 records from each airport resulting in 1845 records in total. The remaining 20,295 observations are employed for model validation as a holdout sample. From our initial analysis on the continuous monthly air passenger departures, we found that the distribution of the variable is right skewed. The descriptive statistics of the continuous demand variable is as follows: mean: 155273, median: 13061, standard deviation: 417368, minimum: 0, and maximum: 4103420. In analyzing the airline demand data, we perform natural logarithmic transformation of monthly departures and then categorize the log-transformed variables into 10 demand groups including ≤6, >6-7, >7-8, >8-9, >9-10, >10-11, >11-12, >12-13, >13-14 and >14.

The distribution of the categorical demand variable is presented in Figure 1. The figure shows that the dependent variable is approximately normally distributed.

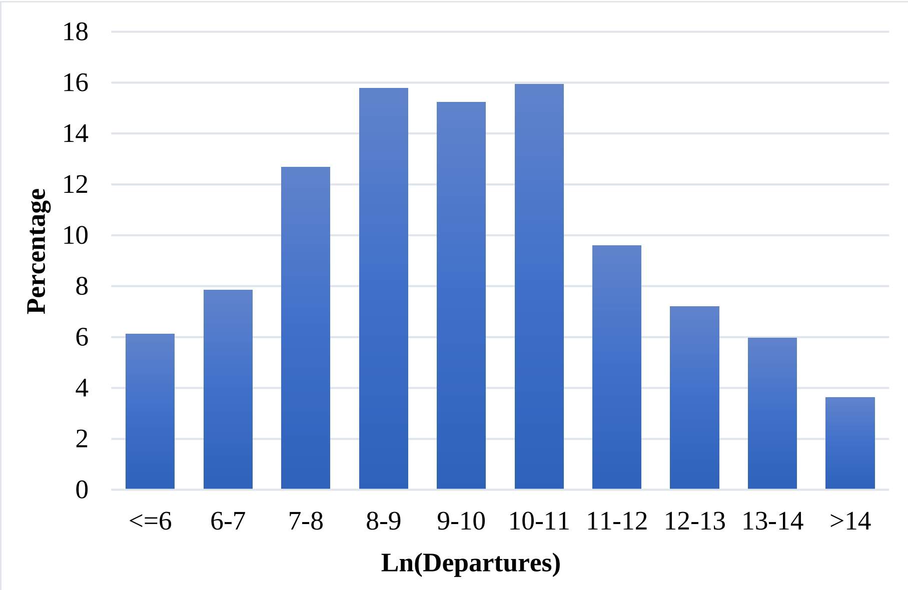


FIGURE 1 Distribution of dependent variable

The airline demand data is augmented with a comprehensive set of independent variables including a) demographic characteristics, b) built environment characteristics, c) airport specific factors, d) spatial factors, and e) temporal factors. Demographic characteristics includes Metropolitan Statistical Area (MSA) specific (Micropolitan Statistical Area where applicable) population, median income, employment, out of state employment rate, vehicle ownership level, etc. Demographic data is sourced from American Community Survey (ACS). Built environment characteristics include number of airports in close proximity of an airport, and tourism ranking of the corresponding state (Business Insider). Airport specific factors include airport classification such as core airports, and Operational Evolution Partnership (OEP-35) airports. Spatial factors include region of the airports including South, West, Mid-West, North-East, and Pacific regions. Temporal factors include month of the analysis ranging from January through December. The detailed description of the independent variables is presented in Table 1. Table 1 includes mean and standard deviation for continuous variables and frequency and percentage for categorical variables.

TABLE 1 Descriptive Statistics of the Independent Variables

Continuous Variables			
Variables	**Description**	**Mean**	**Std. Dev.**
Demographic Characteristics			
Population	Population in million in corresponding MSA	1.179	2.928
Median Income	Median income in 100K in corresponding MSA	0.544	0.119
Employment	Ln(number of workers in thousands residing in corresponding MSA)	0.464	0.046
Out of state employment	Fraction of job holders in corresponding MSA working out of state	0.029	0.046

Built Environment Characteristics			
No. of airports	Ln(Number of airports in 50 mile buffer area)	1.753	0.733
Categorical Variables			
Variables	**Description**	**Freq.**	**Percent**
Built Environment Characteristics			
Tourism Attraction			
Top10	The state is among top 10 tourist attraction states	105	28.455
Bottom10	The state is among bottom 10 tourist attraction states	38	10.298
Others	The state is other than top and bottom tourist attraction states	226	61.247
Airport Specific Effect			
Core airport in the US			
Yes		30	8.13
No		339	91.87
Spatial Factors			
Region			
South		114	30.894
West		88	23.848
Mid-West		85	23.035
North-East		46	12.466
Pacific		36	9.756
Temporal Factors			
Month			
January		158	8.564
February		145	7.859
March		133	7.209
April		155	8.401
May		156	8.455
June		147	7.967
July		165	8.943
August		155	8.401
September		149	8.076
October		157	8.509
November		156	8.455
December		169	9.160

5 ANALYSIS AND RESULTS

In model development, we first estimate a simple grouped generalized ordered probit (GGOP) model system without considering any spatial dependencies between the observations. The estimated GGOP model serves as a benchmark for the spatial GGOP models. In this study, we estimate simple GGOP model using CML approach to compare the data fit with the spatial models. However, it is important to note that one can easily estimate simple GGOP model using maximum likelihood (ML) approach[3]. As we maximize log-composite likelihood in this study, traditional BIC measure might not be appropriate. Hence, we employ a modified version of Bayesian Information Criteria (BIC) measure to penalize the models for additional parameters. In this approach, we normalize log-composite likelihood by 2(Q-1). The modified BIC measure is presented in equation 13. The employed modified BIC for log-composite likelihood is equivalent to BIC measure for traditional maximum log-likelihood. Log-composite likelihood (LL) at convergence and modified BIC values of simple GGOP model (17 parameters) are -12,946,475.70 and 7148.71, respectively.

$$\text{Modified BIC} = -\frac{\text{LL}}{\text{Q}-1} + \text{No. of parameters} \times \ln(\text{Q}) \quad (13)$$

In the next step, we estimate a series of spatial lag and spatial error models considering various formulation of $w_{qq'}$ as discussed in methodology section. Based on the data fit and significance of spatial autoregressive/autocorrelation parameters, we select $1/(\sqrt{distance \times month})$ as the element of **W** matrix, and distance band and time threshold are set to be 100 miles and 36 months, respectively. The results shows that autoregressive parameter in spatial lag model is very close to zero and the model does not offer any data fit improvement compared to simple GGOP model. However, spatial autocorrelation parameter in the spatial error model is highly significant and the model offers considerable improvement in data fit. LL and modified BIC values of the proposed spatial error GGOP model (18 parameters) are -12,494,266.20 and 6911.00, respectively. Therefore, the proposed spatial error model is established to be superior to simple GGOP model in terms of LL and modified BIC measures. For the sake of brevity, only the spatial error GGOP model results are presented in this paper. The results of the independent model are presented in the Appendix. The reader should note that the spatial lag model collapsed to the independent model as discussed above.

5.1 Estimation Results

The proposed spatial error GGOP model is presented in Table 2. Positive (negative) value associated with a variable indicates that an increase of the variable increases (decreases) the propensity of higher demand. The effects of the variables on airline demand are discussed in detail as follows:

5.1.1 Demographic Characteristics

Estimation results indicate that airline demand is significantly influenced by MSA level demographics. From the results, it is evident that airport level passenger departure rate is positively associated with MSA level population. Thus, an increase in MSA population increases the

[3]As expected, estimates from CML and ML approaches are exactly same and LL value in CML approach is exactly 2(Q-1) times of LL value in ML approach.

propensity for higher monthly airline demand (see Grosche et al., 2007; Zhou et al., 2018; Tirtha et al., 2022 for similar results). The results show that airline demand is higher in MSAs with higher income level. Finally, we found that employment in the corresponding MSA significantly contributes to airport level airline demand. An increase in number of employees in the corresponding MSA significantly increases the propensity for higher demand. The results might indicate the fact that increased income and employment enhances business activities and also air travel affordability for residents in the MSA (see Chang, 2012 for similar findings regarding income and employment).

5.1.2 Built Environment Factors

Among built environment factors considered, number of airports in close proximity and state level tourism ranking affect airline demand. The effect of number of airports in a 50-mile buffer[4] is found to be positive indicating that as number of surrounding airports increases, departure rate at that airport will increase significantly. This may reflect the fact that number of airports in close proximity may be higher due to overall increased demand for air travel in an area. For example, residents' preference for air mode for long distance travel might be higher in certain metropolitan areas. Those MSAs may have increased number of airports to accommodate the higher air travel demand. It is important to recognize that these impacts go beyond the impact represented by population, income and employment variables. In addition to the number of airports, state level tourism ranking influences airport level air passenger demand. To identify the impact of tourism, we include top 10 and bottom 10 tourist attraction state indicators in the model. The results indicates that if an airport is present among the top 10 tourist attraction states in the US, the airport may experience higher demand in general. Inversely, if an airport is present among the bottom 10 tourist attraction states in the US, the airport, in general, may experience lower demand compared to other airports while controlling for remaining factors.

5.1.3 Airport Specific Factors

In this study, we include airport specific factors in the demand modelling. Airport specific factors include airport classifications such as core airports, and OEP-35 airports. From Table 2, it is evident that airport classification significantly affects airport level airline demand. The results show that core airports in the US experience increased demand compared to other airports if other factors remain the same. The result is intuitive as core airports are the largest airports in the US with the highest passenger share compared to the remaining airports.

5.1.4 Spatial Factors

Location of the airport in the US region is found to be significantly associated with total number of departures at an airport. From Table 2, it is evident that airports located in South region experience higher demand compared to airports in West and Mid-West regions controlling for other factors. On the other hand, airports in North-East and Pacific regions experience lower airline demand compared to airports in West and Mid-West regions. Significance of region indicators highlights the presence of spatial variations in airline demand across the airports in the US.

---

[4] In selection of the buffer area, we consider 50-mile and 70-mile radius of the buffers. Among these two variables, 50-mile buffer variable offers improved result in terms of variable significance and data fit.

5.1.5 Temporal Factors

From the analysis, we also found that there is temporal variability in airline demand. Compared to other months of the year, airline demand is lower in January controlling for other factors. In contrast, airline demand is higher in July compared to the other months. These temporal trends reflect the variation in airline demand that can be attributed to specific months while controlling for other factors.

5.1.6 Threshold Specific Deviations

The proposed model also allows for threshold specific deviations on various predefined thresholds. In our air passenger departure model, we consider various threshold specific deviations based on model fit and sample sizes across each category. The estimation result of these parameters is reported in the second-row panel of Table 2. The deviation parameter is similar to a constant in discrete choice models and does not have an interpretation after incorporating other variables. In our prosed model system, we define threshold parameters based on the observed bins. As thresholds are predefined, we estimate deviations from these values for selected thresholds based on data fit improvements. Based on the magnitude and direction of the deviation, a threshold may shift to the right or to the left. For example, threshold specific effect for threshold 2 is negative (constant=-0.389) indicating that value of threshold 2 is higher than its predefined value (=6) and it shifts to the right. Therefore, probabilities of alternative 1 (alternative 2) increase (decrease) across the observations. These factors might be considered as fixed effects specific to the alternatives affected by that threshold.

5.1.7 Variance Components

In the proposed model, we estimate and parameterize error variance. Variance components are presented in third-row panel of Table 2. From the results, it is evident that error variance is a function of region of the airports. Such parameterization of the variance component allows us to accommodate for heteroscedasticity in the data. The negative coefficient associated with south region indicates that error variance (scale parameter) is lower for airports in the south region. In contrast, scale parameter is higher for airports in the north-east region. These differences reflect the impact of spatial factors on airline demand. The consideration of these factors ensures that the parameters estimated in the model remain unbiased and offer improved accuracy.

5.1.8 Spatial Correlation

The main contribution of this paper arises from consideration of spatial dependency in the airline demand modelling. From the analysis results, we found spatial autocorrelation parameter as strong in magnitude (δ'= 2.378) and highly significant (t statistic is 93.538). The significance of the spatial dependency parameter indicates the presence of unobserved factors affecting airline demand at an airport also influence the demand at other proximally located airports. In the presence of time component in spatial weight matrix, we can conclude that such spatial correlation varies significantly over time. According to the formulation presented in Section 3, we can also conclude that airline demand at an airport is influenced by the unobserved factors at the airports which are closer spatially and temporally.

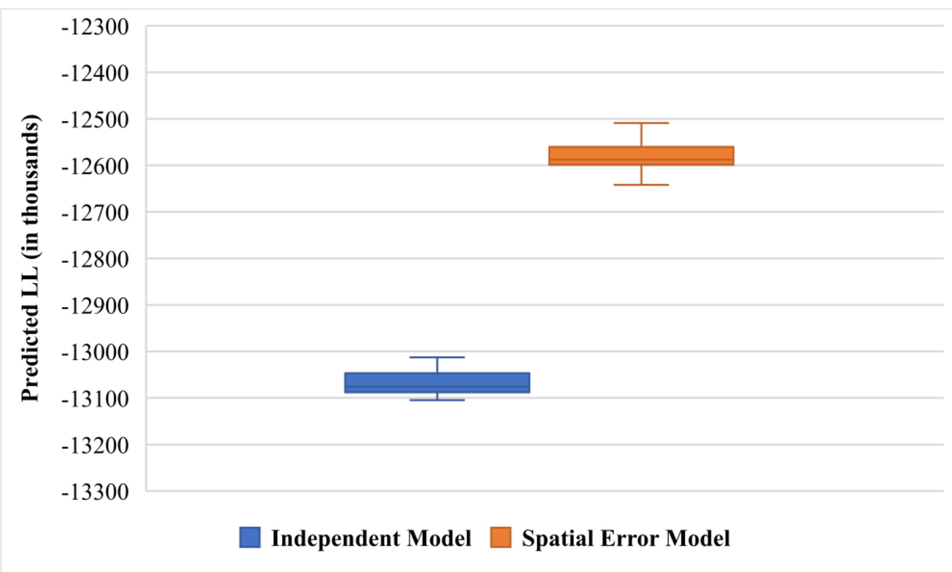
TABLE 2 Estimation Results for Spatial Error GGOP Model

Variables	**Estimates**	**t statistics**
Propensity Components		
Constant	3.977	13.384
Demographic Factors		
Population[5]	0.129	12.133
Median income	1.727	5.248
Employment	5.988	7.879
Built Environment Factors		
No. of airports	0.922	20.857
Tourism Ranking (Base: other states)		
Top 10	0.520	8.006
Bottom 10	-0.529	-5.690
Airport Specific Factors		
Core Airports (Base: No)		
Yes	2.879	31.969
Spatial Factors		
Region (Base: West and Mid-West)		
South	0.609	9.485
North-East	-0.885	-11.068
Pacific	-1.833	-14.196
Temporal Factors		
Month (Base: other months)		
January	-0.392	-4.135
July	0.379	4.522
Threshold Specific Effects		
Threshold 2	-0.389	-5.143
Threshold 3	-0.277	-5.266
Threshold 4	-0.167	-5.014
Variance Components		
Region (Base: other regions)		
South	-0.129	-5.004
North-East	0.152	4.802
Spatial Autocorrelation Parameter		
δ'	2.378	93.538

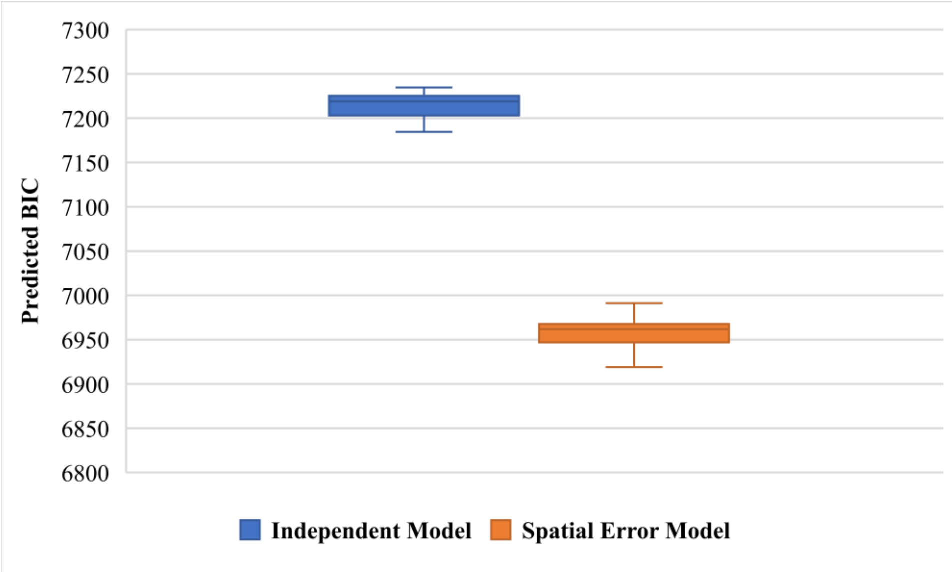
[5] The description of the variables (including exact variable transformation) is provided in the second column of Table 1.

6 MODEL VALIDATION

In this study, we undertake a validation exercise to compare the predictive performance of the alternative models developed. In this comparison, independent GGOP model (without spatial dependency parameter) of air passenger departures serves as the benchmark. To perform the validation test, we employ data from our hold out sample (observations not included in estimation set) consisting of 20,295 observations. From the hold out sample, we further create 20 data samples of 1845 observations by randomly choosing 5 monthly departure records for each airport. Next, we employ alternative models (independent GGOP and spatial Error GGOP) to generate prediction for each sample. Then, the predicted probabilities of the observed demand categories are used to estimate log-composite likelihood (LL) and modified Bayesian Information Criteria (as presented in equation 13) measures for the two model systems. The results from 20 samples are compiled to generate the average and range of the model performance measures across the two systems. The results from validation exercise are presented as a box plot in Figure 2. The result indicates that the average predicted LL and BIC values and the ranges (95% confidence interval) in parentheses for the model systems are as follows: (1) independent model: -13,070,043.02 [-13,090,956.47, -13,049,129.57] and 7215.72 [7204.38, 7227.06], (2) spatial error GGOP model: -12,585,319.26 [-12,603,396.76, -12,567,241.75] and 6960.37 [6950.57, 6970.18]. The results from the validation exercise confirm that spatial error model performs considerably better than the independent model that does not consider for spatial correlations between the observation units. The confirmation from our validation exercise highlights the importance of considering spatial and temporal dependency in airline demand models at the airport level.

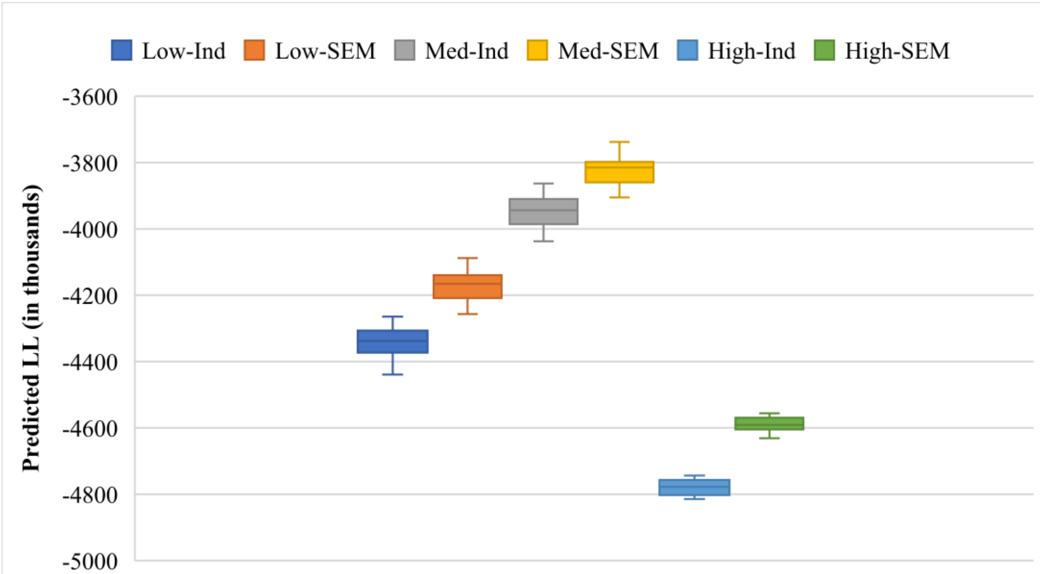


(a) Comparison of predicted LL values

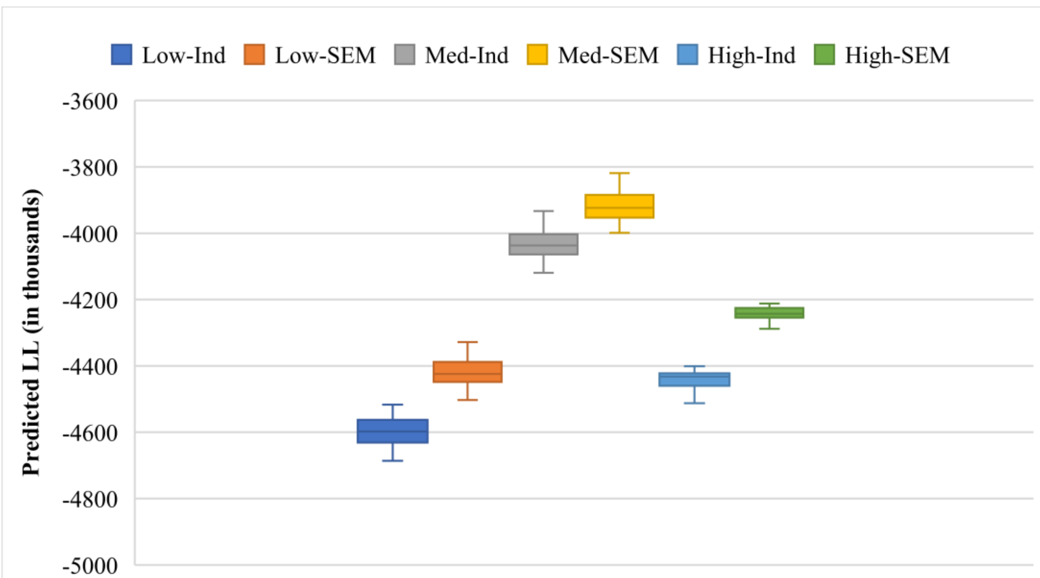


(b) Comparison of predicted BIC values
FIGURE 2 Comparison between two model systems

In this validation exercise, we also investigate how the proposed model performs for various MSA groups. To be specific, the prediction exercise is performed for different airport subgroups based on MSA characteristics including population and employment. For both variables, MSAs are categorized into three groups: low, medium, and high. In categorization of the MSAs, we consider population <=150,000 as low, 150,000<population<=400,000 as medium and population>400,000 as high population MSAs. For employment, we consider employment<=70,000 as low, 70,000<employment<=200,000 as medium and employment>200,000 as high employment MSAs. Similar to the above analysis, we estimate predicted LL values for both independent model and spatial error model using 20 random samples. The results of the analysis are presented in Figure 3. Figure 3a compares model performance of alternative models using population sub-samples and Figure 3b compares model performance using employment sub-samples. From the figures, it is clearly evident that the proposed spatial error model performs better than the independent model for all groups of MSAs. For both population and employment groups, the average LL improvement is the highest for the "high" category (high population: 4.0% and high employment: 4.4% improvements) and lowest for the "medium" category (medium population: 3.2% and medium employment: 2.8% improvements). Overall, the sub-sample analysis highlights the improved performance of the proposed model across all samples.



(a) MSA population sub-samples



(b) MSA employment sub-samples

FIGURE 3 LL comparison between models by MSA characteristics

7 ELASTICITY ANALYSIS

The results presented in Table 2 offers important insights on the relationship between airline demand and the independent variables. However, it is not possible to quantify the impact of the variables based on parameter estimates in a non-linear model structure. To quantify the variable impact, we undertake an elasticity analysis in this study. In this analysis, we determine variable impacts on the continuous airline demand. In predicting continuous airline demand under various scenarios, we follow the prediction procedure developed by Tirtha et al., 2023. For continuous demand variable, we identify the percentage change of aggregated demand due to the change in the independent variable. To be specific, we consider a 20% increase for the continuous independent variables. For example, we increase MSA population by 20% across the dataset to quantify its impact on airline demand. For indicator variables, we convert 20% of zero records to ones. For example, we convert 20% non-core airports to core airports to quantify its impact on airline demand. The result of the analysis is presented in Figure 4. In Figure 4, we plot the aggregated continuous demand predictions. From the results, it is evident that among the selected variables MSA level employment, airport classification (classified as core airport in the US), and number of airports in close proximity have considerably higher impact compared to other variables. The approach allows us to develop a non-linear model that can provide elasticity at the continuous resolution. The findings from the elasticity analysis can be useful for airport agencies, airlines and metropolitan agencies to plan policies for attracting air travelers and accommodating increased demand through resource allocation and existing facility improvements.

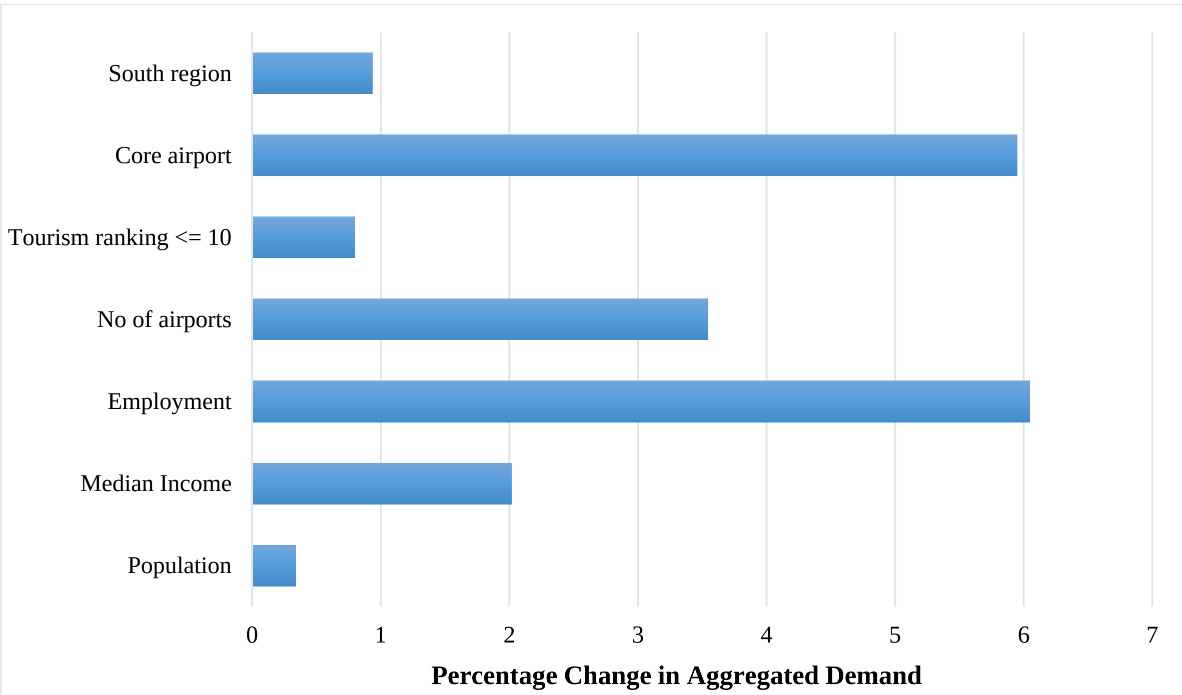


FIGURE 4 Results of the elasticity analysis

8 CONCLUSIONS

The current study aims to analyze monthly air passenger departures at the airport level accommodating for spatial interactions between the airports in close proximity. Towards this end, we develop a novel spatial grouped generalized ordered probit (SGGOP) model system of monthly air passenger departures at the airport level. Specifically, we estimate two variants of spatial models including spatial lag model and spatial error model. In presence of repeated demand measures for the airports, we also consider temporal variations of spatial correlation effects among proximally located airports by employing space and time-based weight matrix. The proposed model is estimated using monthly air passenger departures for five years for 369 airports across the US. The proposed spatial model is implemented using composite marginal likelihood (CML) approach that offers a computationally feasible framework compared to sheer dimensionality challenge associated with the full likelihood approach for discrete outcome spatial models.

In model development, we employed various functional forms for the weight matrix and model selection was based on data fit. Among the three model systems we estimated, spatial error GGOP model was found to be the best in terms of BIC measure. While spatial error model significantly improves the data fit, spatial lag model does not offer any improvement compared to independent GGOP model. From the estimation results, it is evident that air passenger departures at the airport level are influenced by different factors including MSA specific demographic characteristics, built environment characteristics, airport specific factors, spatial factors, and temporal factors. Moreover, spatial autocorrelation parameter is found to be significant supporting our hypothesis of the presence of common unobserved factors associated with the spatial unit of analysis. In this study, we also perform a validation analysis to examine the predictive performance of the proposed spatial error GGOP model compared to independent GGOP model. The result of validation exercise indicates the superiority of spatial error model relative to the independent model.

The proposed model allows us to identify key factors affecting airline demand at the airport resolution. Therefore, the proposed model will be useful for airport agencies, airlines and metropolitan agencies to plan policies for attracting air travelers and accommodating increased demand through resource allocation and existing facility improvements (see Tirtha et al., 2023 for a detailed discussion). For example, regions with growing population and tourism investments might examine how airline demand is likely to be altered based on our proposed model. The metropolitan regions (and airports) located in the South region of US with rapidly growing population will be able to predict airline demands for future years using our model. To be sure, the independent variables considered in the model should be carefully constructed for the future year of interest to generate these estimates. Finally, the proposed model allows us to consider for the impact of demands at proximal airports.

To be sure, the current study is not without limitations. It would be useful to accommodate for other socio-economic factors in the proposed model such as MSA specific GDP and business-related indicators. Moreover, the dataset considered in this study allows us to estimate monthly airline demand measure. If available, consideration of more disaggregated data at the daily/weekly level may enhance the proposed model. We employ state level tourism ranking to capture the effect of tourism on airline demand. MSA specific tourism measures (For example: number of hotel beds), if available, may further enhance the demand model.

ACKNOWLEDGMENT

The authors would like to acknowledge Bureau of Transportation Statistics (BTS) for providing access to their datasets.

REFERENCES

Abed, S. Y., Ba-Fail, A. O., Jasimuddin, S. M., 2001. An econometric analysis of international air travel demand in Saudi Arabia. Journal of Air Transport Management. 7(3), 143–148.

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: Methodological. 44(8–9), 903–921.

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

Business Insider. Detailed Look At How Americans Travel Within The US. Available at: https://www.businessinsider.com/the-most-popular-us-states-for-tourism-2014-10 (Accessed: 18 July 2020).

Carrión-flores, C. E., Flores-lagunes, A., Guci, L., 2009. Land use change: a spatial multinomial choice analysis. AAEA & ACCI Joint Annual Meeting. Available at: http://ageconsearch.umn.edu/bitstream/49403/2/CFG_AAEA_612952.pdf.

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.

Chakir, R., Parent, O., 2009. Determinants of land use changes: A spatial multinomial probit approach. Papers in Regional Science. 88(2), 327–344.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. The review of economic studies. 47(1), 225-238.

Chang, L. Y., 2012. International air passenger flows between pairs of APEC countries: A non-parametric regression tree approach. Journal of Air Transport Management. 20, 4–6.

Chang, L. Y., 2014. Analysis of bilateral air passenger flows: A non-parametric multivariate adaptive regression spline approach. Journal of Air Transport Management. 34, 123–130.

Chen, C. F., Chang, Y. H., Chang, Y. W., 2009. Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan. Transportmetrica. 5(2), 125–140.

Chen, Z., Haynes, K. E., 2015. Public surface transportation and regional output: A spatial panel approach. Papers in Regional Science. 94(4), 727–751.

Chi, J., 2014. A cointegration analysis of bilateral air travel flows: The case of international travel to and from the United States. Journal of Air Transport Management. 39, 41–47.

Daraban, B., Fournier, G. M., 2008. Incumbent responses to low-cost airline entry and exit: A spatial autoregressive panel data analysis. Research in Transportation Economics. 24(1), 15–24.

FAA, 2023. Air Traffic By The Numbers. Available at: https://www.faa.gov/air_traffic/by_the_numbers/ (Accessed: 17 May 2023).

Faghih-Imani, A., Eluru, N., 2016. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. Journal of Transport Geography. 54, 218–227.

Ferdous, N., Bhat, C. R., 2013. A spatial panel ordered-response model with application to the analysis of urban land-use development intensity patterns. Journal of Geographical Systems. 15(1), 1–29.

Godambe, V. P., 1960. An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics. 31(4), 1208–1211.

Grosche, T., Rothlauf, F., Heinzl, A., 2007. Gravity models for airline passenger volume estimation. Journal of Air Transport Management. 13(4), 175–183.

Lee, L. fei, Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics. 154(2), 165–185.

Li, T., Baik, H., Trani, A. A., 2013. A method to estimate the historical US air travel demand. Journal of Advanced Transportation. 47(3), 249–265.

Li, T., Wan, Y., 2019. Estimating the geographic distribution of originating air travel demand using a bi-level optimization model. Transportation Research Part E: Logistics and Transportation Review. 131(October), 267–291.

Loo, B. P. Y., Ho, H. W., Wong, S. C., 2005. An application of the continuous equilibrium modelling approach in understanding the geography of air passenger flows in a multi-airport region. Applied Geography. 25(2), 169–199.

Matsumoto, H., 2004. International urban systems and air passenger and cargo flows: Some calculations. Journal of Air Transport Management. 10(4), 239–247.

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. Journal of Supercomputing. 74(10), 5461–5484.

Rahman, M., Yasmin, S., Faghih-Imani, A., Eluru, N., 2021. Examining the Bus Ridership Demand: Application of Spatio-Temporal Panel Models. Journal of Advanced Transportation. 2021.

Suryani, E., Chou, S. Y., Chen, C. H., 2010. Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework. Expert Systems with Applications. 37(3), 2324–2339.

Tirtha, S. D., Bhowmik, T. and Eluru, N., 2023. Understanding the Factors Affecting Airport Level Demand (Arrivals and Departures) Using a Novel Modeling Approach. Journal of Air Transport Management. 106, 102320.

Tirtha, S.D., Bhowmik, T., Eluru, N., 2022. An airport level framework for examining the impact of COVID-19 on airline demand. Transportation Research Part A: Policy and Practice. 159(March), 169–181.

Tirtha, S.D., Yasmin, S., Eluru, N., 2020. Modeling of Incident Type and Incident Duration Using Data from Multiple Years. Analytic Methods in Accident Research. 28, 100132.

Tong, T., Yu, T. H. E, Cho, S. H., Jensen, K., De La Torre Ugarte, D, 2013. Evaluating the spatial spillover effects of transportation infrastructure on agricultural output across the united states. Journal of Transport Geography. 30, 47–55.

Tsui, W. H. K., Ozer Balli, H., Gilbey, A., Gow., H., 2014. Forecasting of Hong Kong airport's passenger throughput. Tourism Management. 42(2014), 62–76.

Valdes, V., 2015. Determinants of air travel demand in Middle Income Countries. Journal of Air Transport Management. 42, 75–84.

Wang, X., Kockelman, K. M., 2006. Tracking land cover change in mixed logit model: Recognizing temporal and spatial effects. Transportation Research Record (1977), 112–120.

Xu, S., Chan, H. K., Zhang, T., 2019. Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. Transportation Research Part E: Logistics and Transportation Review. 122(October 2018), 169–180.

Yu, N., Jong, M., Storm, S., Mi, J., 2013. Spatial spillover effects of transport infrastructure: Evidence from Chinese regions. Journal of Transport Geography. 28, 56–66.

Zhou, H., Xia, J., 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. Journal of Air Transport Management. 70(May), 91–103.

APPENDIX

The estimation results for alternative models estimated in this paper have been included in this Appendix section. In this study, we estimated an independent model and a spatial lag model along with the spatial error model presented in Table 2 in the manuscript. The spatial autoregressive parameter in spatial lag model is close to zero and it does not offer any data fit improvement compared to the independent model. Therefore, spatial lag model is no different than the independent model. The results from the estimated independent model are presented below:

TABLE A1 Estimation Results of the Independent Model

Variables	**Estimates**	**t statistics**
Propensity Components		
Constant	5.180	19.778
Demographic Factors		
Population	0.139	13.335
Median income	1.848	6.194
Employment	4.120	6.177
Built Environment Factors		
No. of airports	0.787	19.905
Tourism Ranking (Base: other states)		
Top 10	0.573	8.163
Bottom 10	-0.365	-4.365
Airport Specific Factors		
Core Airports (Base: No)		
Yes	2.945	35.847
Spatial Factors		
Region (Base: West and Mid-West)		
South	0.431	7.146
North-East	-0.833	-11.457
Pacific	-1.563	-14.101
Temporal Factors		
Month (Base: other months)		
January	-0.324	-3.469
July	0.374	4.523
Threshold Specific Effects		
Threshold 2	-1.312	-27.480
Threshold 3	-0.890	-27.693
Threshold 4	-0.467	-22.582
Variance Components		
Region (Base: other regions)		
South	0.236	13.250
North-East	0.496	16.161