2022

An Econometric Analysis of Domestic Aviation in the US

Sudipta Dey Tirtha
University of Central Florida

AN ECONOMETRIC ANALYSIS OF DOMESTIC AVIATION IN THE US

by

SUDIPTA DEY TIRTHA
B.Sc. Bangladesh University of Engineering and Technology, 2017
M.Sc. University of Central Florida, 2020

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2022

Major Professor: Naveen Eluru

ABSTRACT

In this dissertation, we examine two dimensions of domestic aviation - demand and delay - that directly influence economic impact of the sector. We conduct a comprehensive analysis of airline demand employing airline data compiled by Bureau of Transportation Statistics. The demand analysis is conducted in three steps. First, we propose a novel modeling approach for modeling airline demand evolution over time. Specifically, we develop a joint panel group generalized ordered probit (GGOP) model system for modeling air passenger arrivals and departures in a discretized framework that subsumes the traditional linear regression approach. Further, we consider the influence of observed and unobserved effects on airline demand across multiple time periods. Second, we explore the impact of Coronavirus disease 2019 (COVID-19) on domestic airline demand in the US. The effect of COVID-19 on airline demand is identified by considering global and local COVID-19 transmission, temporal indicators of pandemic start and progress, and interactions of airline demand predictors with global and local COVID-19 indicators. Based on the results, we present a blueprint for airline demand recovery using three hypothetical scenarios of COVID-19 transmission rates – expected, pessimistic and optimistic. Finally, we build on the novel airline demand modeling framework by accommodating for observed and unobserved spatial and temporal effects. Specifically, we develop spatial lag model and spatial error model formulations of the GGOP model proposed in the first step. The second part of the dissertation is focused on flight level delay analysis. In this part, we identify the factors affecting flight level airline delay by jointly modeling departure and arrival delays. Towards this end, we develop a novel copula-based group generalized ordered logit model system that accommodates for the influence of common observed and unobserved effects on flight departure and arrival delays.

Keywords: Airline Demand, COVID-19, Demand Recovery, Spatial Dependency, Airline Delay

ACKNOWLEDGMENTS

I consider myself highly fortunate to be guided and supported by many nice people to complete my PhD dissertation. Without their consistent assistance and encouragements, completion of this dissertation might not have been possible.

The first person to whom I would like to convey my heartiest gratitude is my honorable supervisor, Dr. Naveen Eluru. I am highly grateful to Dr. Eluru for his excellent supervision, encouragement, and constant support in this dissertation. In every academic step during my graduate study, I have been privileged to be guided with his profound wisdom. I always consider myself very lucky to have the opportunity to work under Dr. Eluru's supervision.

The next person I want to mention is Dr. Tanmoy Bhowmik. Dr. Bhowmik is an excellent mentor and he provided continuous support throughout this research. In many difficult research problems, Dr. Bhowmik assisted me with his innovative ideas. I convey my gratitude to him for his instructions, and encouragement throughout this dissertation.

I would also like to gratefully acknowledge Dr. Mohamed Abdel-Aty and Dr. Adan Vela for their valuable comments and suggestions in this dissertation.

It would be very much unfair if I forgot to mention the contributions of my family members. I am highly grateful to my parents (Tapash Ranjan Dey and Dipti Rani Dey), my spouse (Spandita Roy), my sister (Sutapa Dey Tithi) and my little kid (Sattyik Dey) for their love, motivation, and immense support. Without their constant supports, it would be impossible for me to achieve this degree.

Finally, I would also like to gratefully acknowledge Bureau of Transportation Statistics (BTS) for providing access to airline data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Background

In the United States, commercial aviation sector is a significant contributor to the economy. About 7.3% of the US job sector is attributed to commercial aviation sector contributing about 5.2% of US Gross Domestic Product (FAA, 2022). Further, airline industry is closely intertwined with tourism, hospitality, and related auxiliary business (such as rental cars). An important metric to examine the health of the aviation sector is passenger demand – arrivals and departures - at airports. Thus, understanding the factors influencing air passenger demand is important for long term planning and operational decisions. While airline passenger demand and revenue have steadily increased at an annualized growth rate of 2.9% and 5.4% respectively between 2009 and 2019, airline industry has experienced a significant shock in passenger demand worldwide due to the recent outbreak of Coronavirus disease 2019 (COVID-19). COVID-19, as of January 29^{th}, 2022, with a reported 370.10 million cases and 5.67 million fatalities, has affected nearly every country in the world (Worldometer, 2022). In the United States, 73.51 million cases and 876.63 thousand fatalities have been reported (CDC, 2022). The pandemic has affected every facet of life in the world significantly burdening social, health and economic systems. Among these affected industries, airline industry ranks as one of the worst affected industries (S & P global, 2020). The estimated annual drop in global passenger demand and revenue amounts to 2.70 billion passenger trips and 372 billion dollars respectively (ICAO, 2022). The US airline domestic passenger demand reduced by 476 million in 2020 compared to the previous year (BTS, 2022a). Airline demand in the recent months has started to recover from April 2020 lows as precautions at airports, access to testing and mask mandates has encouraged some air travel. However, the magnitude of the challenge facing the airline industry is highlighted by the current state of operations. Airline demand in December 2020 still represents only 39.1% of the demand in December 2019. The emergency use authorization of vaccines offers promise in curbing the pandemic and supporting the recovery. As the recovery begins airlines and airports would need to address supply side shortages with growing demand. This is particularly critical as airline supply (flights) has reduced by about 70% relative to the previous year (BTS, 2020). Therefore, understanding the potential path to recovery will allow airlines, airport management agencies to design plans for increasing flight availability and hiring staff for airline and airport operations.

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. But there might be some observed and unobserved factors associated with closely linked spatial units that may cause demand correlation among the airports. Neglecting the presence of such observed and unobserved spatial correlations in demand modelling may result in biased estimates.

In addition to the airline demand challenges due to COVID-19, flight delays at airports have become recurrent events in recent years causing significant economic loss to commercial aviation industry. According to Bureau of Transportation Statistics (BTS), 20.79% of all flights operated in the US arrived late by 15 minutes or more in 2019 (the highest such percentage since

2014) (BTS, 2022b). Airline delays cause both direct and indirect costs to several components of the industry. The cost of airline delays attributed to passengers is estimated at $18.1 billion in 2019 (FAA, 2019). Costs attributed to airlines from additional expenses for crews, fuel and maintenance is estimated at $8.3 billion (FAA, 2019) not considering the impact of the worsening customer experience on airline attractiveness (Suzuki, 2000). Airline delays also cause indirect costs to different business sectors amounting to nearly $4.2 billion (FAA, 2019). Given such negative impacts of airline delays on aviation industry, it is important to identify the key factors of airline delays and quantify their impacts to plan policies for reducing or mitigating the delays.

1.2 Motivation for the Study

Given the importance of the airline industry to US economy, understanding the factors affecting airline demand at US airports is important for long-term planning (such as airport runway and terminal design and expansion, intermodal transportation facilities) and operational decisions (such as crew management for airport services). Also, analyzing how airline demand at airports evolved over time in presence of external/health shocks and identifying the factors contributing to this evolution will allow us to build a template of a possible recovery path in the future months. This will allow airlines, airport management agencies to design plans for increasing flight availability and hiring staff for airline and airport operations. Moreover, it is important to accommodate spatial and temporal dependencies between the spatial units (airport) in air passenger demand modeling. Neglecting such dependencies, when they actually exist, may result in biased estimates. Finally, given the substantial negative impacts of airline delays on the US economy, understanding the factors influencing airline on time performance will allow airlines to improve their on-time performance or mitigate the delays by increasing and reallocating their resources such as aircrafts, crews, and staff.

While earlier research identifying the factors of airline travel demand has offered significant insights, there is still scope for enhancing our understanding of factors influencing airline demand. *First*, it is possible to enhance spatial and temporal data for airline demand analysis. While most of the earlier studies analyzed airline demand at aggregate levels such as country or region, airport level disaggregate analysis may better incorporate the local factors in modeling airline demand. In addition, earlier studies that conducted airport level prediction analysis have employed a small number of airports in the US. Spatially, our aim is to consider a large set of airports across the country. Temporally, our aim is to examine airline demand at a quarterly level for multiple years. Also, while previous studies focus on only one dimension of airline demand, we focus on two airport level variables - arrivals and departures. Given the obvious interaction between these two variables, we are motivated towards developing a bivariate multiple time period framework that recognizes the influence of common unobserved factors. *Second*, it is important to examine the appropriate hierarchy of unobserved factors that affect airline demand. The inclusion of observed factors within the model framework is reasonably straightforward. However, unobserved effects in the current context provide multiple levels of hierarchies including airport level, airport – year, airport – quarter, quarter only, departures and arrivals. *Finally*, earlier research has predominantly considered linear regression and its variants as a framework for such analysis. This is expected due to continuous nature of airline demand variables (such as natural logarithm of airline demand). However, linear regression models 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.

Further, it is far from straightforward to test for polynomial terms for all variables. To address this limitation, we are motivated towards recasting a recently developed model structure referred to as the grouped response framework for developing a non-linear regression framework that is analogous to the linear regression model system without the restrictions of linear regression (Tirtha et al., 2020; Bhowmik et al., 2019; Rahman et al., 2019).

Due to the outbreak of COVID-19 pandemic, airline industry has experienced a significant shock in passenger demand. The emergency use authorization of vaccines offers promise in curbing the pandemic and supporting the recovery. As the recovery begins, airlines and airports would need to address supply side shortages with growing demand. Therefore, it is important to examine the impact of local and global COVID-19 factors to build a template of the possible demand recovery path. However, the earlier research efforts on measuring the impact of shocks (external or health) on airline demand focused on a retrospective analysis as opposed to offering insights for the potential recovery of demand in response to the shock. While earlier research provides the building blocks of demand prediction systems and some insights on modeling demand in the presence of shocks, these frameworks have not been employed to study demand recovery patterns. Thus, there are research opportunities for enhancing our understanding of the impact of the pandemic and identifying potential recovery path. *First*, research on COVID-19 impact on airline industry is in the nascent stages and has predominantly focused on global or regional effects. Therefore, it is important to examine the influence of COVID-19 at the disaggregate resolutions to incorporate the interplay of local and global factors on airline demand. *Second*, the research study is motivated towards employing a robust modeling framework to analyze airline demand variable. An exhaustive specification exercise can be conducted to evaluate the impact of various COVID-19 factors while controlling for other attributes affecting airline demand. *Finally*, policy analysis based on different hypothetical scenarios can be effective to provide a blueprint to the path to recovery for airline demand.

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. While earlier research efforts on airline demand modeling have neglected to adequately consider for such spatial interactions, in this dissertation, we are motivated to address this gap by developing spatio-temporal models (spatial lag and spatial error) of monthly air passenger departures at the airport level that explicitly accommodates the spatial interactions between the proximally located airports.

In recent years, a significant proportion of all flights in the US have been delayed. Airline delay causes significant economic loss to different components of commercial aviation industry including passengers, airlines, business sectors, etc. Moreover, events like flight delays adversely affect carrier schedule reliability. Given the adverse effect of such delays on the US economy, understanding the factors affecting on time performance of the airlines at the flight level is important to plan policies (including but not limited to prioritizing certain regions/airports, resource allocation, expansion of facilities) to reduce the delays and subsequent economic loss. In this dissertation, we focus on formulating a mathematical model of airline delay (departure delay and arrival delay) at a disaggregate level of flight to identify the factors of such delays and quantify their impacts. In this dissertation, we are motivated to enhance flight delay data for delay analysis. The flight delay data we employed is sourced from 2019 marketing carrier on time performance dataset compiled by BTS. The flight delay data are augmented with a comprehensive set of independent variables sourced from secondary data sources including Automated Surface Observing System (ASOS) dataset (sourced from Iowa Environment Mesonet) and FAA's Aviation System Performance Metrics (ASPM). Using the ASOS dataset, we focus on building a detailed process that allows us to generate weather conditions for the entire duration of the flight. Subsequently, we employ ASPM data to determine air traffic conditions at the origin and destination airports in the hours preceding the flight's departure and arrival, respectively. The data for our analysis is also augmented with other independent variables including (a) trip specific factors (carrier and flight distance), (b) spatial factors (region of origin and destination airports) and (c) temporal factors (season, day of the week and time of the day). While earlier research efforts mostly analyzed arrival delay, we focus on analyzing both delay categories – departure delay and arrival delay in this research. Given the obvious interactions between two types of delay variables, we focus on developing a joint model framework that accommodates for the influence of common observed and unobserved effects on flight departure and arrival delays.

1.3 Objective of the Dissertation

The <u>first objective</u> of this dissertation is to identify the factors of quarterly air passenger arrivals and departures at the airport level and quantify their impact. Towards achieving this goal, the current research develops a joint panel group generalized ordered probit model system with observed thresholds for modeling air passenger arrivals and departures while accommodating for the influence of observed and unobserved effects on airline demand across multiple time periods. The proposed model system is estimated using airline demand data from Bureau of Transportation Statistic (BTS) for 510 airports at quarterly level for 5 years (2010, 2012, 2014, 2016, and 2018). In preparation of dependent variables, we discretize log-transformed quarterly air passenger arrivals and departures into 14 demand categories (≤3; >3-4; >4-5, >5-6, >6-7, >7-8, >8-9, >9-10, >10-11, >11-12, >12-13, >13-14, >14-15 and >15). For the selected airports, we augmented the airline demand data with a host of independent variables including demographic characteristics and built environment characteristics at metropolitan statistical area (MSA), spatial and temporal factors. Demographic factors include population, household median income, employment, out of state employment and education level of the residents in the corresponding metropolitan statistical area (MSA). Built environmental characteristics include number of airports in 50 mile buffer area around the airport of interest and tourism ranking of the corresponding state. Spatial factors include location of the airport in terms of region including south, north-east, west, mid-west and pacific region. Temporal factors include year and quarter of the analysis. The current study also performs a validation exercise to compare the performance of the proposed model with traditional linear regression model. Finally, an elasticity analysis is undertaken to quantify impact of the factors of airline demand.

The second objective of this dissertation is to identify the impact of COVID-19 on domestic airline demand in the US and provide a blueprint of recovery path in the upcoming months. Towards achieving this broad objective, the current study develops a model for analyzing airport level passenger demand data characterized as monthly departures at the airport level. In this current study, we consider monthly airline demand for 380 airports in the US for 24 months from January 2019 through December 2020. The dependent variable is sourced from T-100 Domestic Market dataset provided by Bureau of Transportation Statistics (BTS). Flight passenger counts are aggregated over origin airports for each month to generate the dependent variable. In preparation of the dependent variable, we perform log-transformation of monthly air passenger departures at the airport level. We employ a linear mixed modeling method that examines the continuous monthly airport level passenger demand, and the model system is estimated with a host of independent variables including (a) global and local COVID-19 factors, (b) county level demographic characteristics, (c) built environment characteristics, (d) airport specific factors, (e) spatial factors, (f) temporal factors, and (g) adjoining county attributes. The model developed is employed to generate predictions for airline demand under various scenarios of future COVID-19 transmission in response to vaccinations and other guidelines. The research develops a potential band of airline demand recovery over time by considering expected, pessimistic and optimistic scenarios.

The third objective of this dissertation 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 in its discretized form 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. We recast the recently developed generalized group ordered probit (GGOP) framework to model the ordinal airline demand variable. In our first objective, we will present that the proposed non-linear system subsumes the traditional linear regression model system. In the proposed 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) methods 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 traditional model that does not consider any spatial dependency.

The fourth objective of this study is to formulate a copula based joint model system to identify important determinants of flight departure delay and arrival delay as well as quantifying their impact. In this study, we develop a novel copula-based group generalized ordered logit (GGOL) model of flight delay at the flight level. Departure delay and arrival delay (in minutes) are sourced from the BTS 2019 non-stop domestic marketing carrier on time performance dataset. In preparation of the dependent variables, departure delay and arrival delay are categorized (in minutes) into 6 groups (0-5, 5-10, 10-15, 15-30, 30-60, >60 minutes). The flight delay data is also augmented with a host of independent variables including (a) airport level traffic conditions, (b) trip level attributes, (c) weather factors, (d) spatial factors, and (e) temporal factors. The current research effort will allow us to examine the impact of these aforementioned factors on on-time flight performance. Further, the value of the proposed model system is illustrated by comparing predictive performance of the proposed model relative to independent models of flight departure and arrival on a holdout sample (records not used in estimation). Finally, we conduct an application analysis to present the policy implications of the current research. The illustration provides a mechanism for employing the proposed model as a tool for airline carrier level or airport level delay prediction analysis using weather forecasts.

1.4 Outline of the Dissertation

The remainder of the dissertation is divided into six chapters. In chapter 2 through chapter 5, we focus on objective 1 through objective 4, respectively. In the last chapter, we provide important concluding remarks based on our findings in this dissertation and discuss contributions and limitations of this research, and future research scope.

Chapter two contributes to objective one by modeling air passenger travel demand using a joint panel group generalized ordered probit model system. First, this chapter describes relevant earlier research and positions the current study. Next, modeling approach and details of the dataset employed in the research are presented. The proposed model system is estimated using origin and destination survey data provided by BTS for 510 airports in the US. In preparation of dependent variables, we performed log transformation of arrivals and departures, and then considered 14 categories of the transformed variables. The following sections presents model selection procedure and estimation results. In the next section, we undertake a validation exercise to compare performance of the proposed model with traditional linear regression model. The results of elasticity analysis are presented in the following section. Finally, summary of this chapter is presented in the last section.

Chapter three contributes to objective two by identifying the impact of COVID-19 on domestic airline demand in the US. First, this chapter describes relevant earlier research and positions the current study. Then, modeling approach and details of the dataset employed in the research are presented. The proposed model system is estimated using T-100 domestic market dataset provided by BTS for 380 airports in the US. In preparation of dependent variables, we performed log transformation of airport level monthly departures. The following section presents the model estimation results. In the next section, we undertake a validation exercise to compare observed and predicted demand to evaluate the model performance. The results of policy analysis are presented in the next section. Finally, summary of this chapter is presented in the last section.

Chapter four contributes to objective three by developing novel spatial GGOP models (spatial lag and spatial error) of airport level monthly air passenger departures while capturing for spatial interactions of airports in close proximity. First, this chapter discusses earlier studies in airline demand modeling and presents research efforts for capturing spatial interactions. Also, contributions of the current research are highlighted. Next, econometric methodology and dataset description are provided in subsequent sections. The next section presents model selection steps and estimation results. In the following section, we undertake a validation exercise to evaluate the model performance of the alternative models. Finally, the last section summarizes the chapter.

Chapter five contributes to objective four by developing a novel copula based GGOL model of flight departure and arrival delays. First, this chapter discuss earlier studies in airline delay literature and positions the current study. The next section provides discussions on the methodology and estimation process of the proposed model. Next, data preparation procedures and description of the dataset employed for model estimation are presented. In the following section, we select the best copula model by comparing of the performance of independent GGOL models and joint models with different dependency structures. Then, the results of the best Copula model in terms of data fit are presented and discussed in detail. The next two sections conduct model validation and policy analysis, respectively. In the last section, we summarize the findings from this chapter.

Finally, Chapter six summarizes this dissertation with some concluding remarks. First, we present the objectives, methodologies, and key findings from chapter 2 through chapter 5. Next, we discuss the contributions of this dissertation. Finally, we present the limitations and scope for future research.

CHAPTER 2: UNDERSTANDING THE FACTORS AFFECTING AIRPORT LEVEL AIRLINE DEMAND

Commercial aviation industry significantly contributes to the US economy and air passenger demand (arrivals and departures) is a key indicator of the health of this industry. Understanding the factors of air passenger demand is important for long-term planning and operational decisions. This chapter presents a novel modeling approach for modeling air passenger arrivals and departures at the disaggregate resolution of airport. The proposed model is developed using a large set of airports across the US and analogous to linear regression model without the restrictions of linear regression. A validation exercise is undertaken to evaluate the performance of the proposed model. Finally, an elasticity analysis is performed to quantify the impact of the key factors.

2.1 Earlier Studies

To be sure, several studies have examined airline passenger demand. Table 2.1 provides a summary of earlier research efforts related to air passenger travel demand modeling with information on the study, study region, demand resolution, study objectives, methodology and independent variables considered[1]. From Table 2.1, we can make several important observations. First, earlier research on air travel demand can be categorized into two groups based on the spatial unit of demand data analyzed: (a) airport level and (b) regional level. In the former category,

---

[1] The reader would note that we focused on earlier research examining airline demand. For studies exploring itinerary shares or individual level airline survey data analysis see Li & Wan, 2019; Chi, 2014; Carson et al., 2011; Wei & Hansen, 2006 and Coldren et al., 2003.

studies analyze passenger demand data for individual airports while in the latter category, the analysis is conducted by aggregating demand at a regional level. From the review, a majority of earlier research focused on analyzing aggregate demand (we found only three studies that explored data at the airport level). Second, the factors identified to affect airline demand have been consistent including socio-demographic factors (population, education, age distribution), socio-economic factors (income, unemployment rate, GDP), built environment (number of trade centers, tourist attractions), level of service factors (average air fare and distance) and lag variables (historical demand). Third, in terms of mathematical frameworks employed for analyzing data, we found two predominant approaches: (a) prediction methods using data and (b) distribution or assignment methods. The majority of prediction methods focused on one dimension – trip departures from the spatial unit of interest. Thus, these studies resorted to employing univariate models of passenger demand such as regression models and their variants such as repeated measures models and regression trees, Artificial neural networks and Fuzzy models. The second set of studies employ approaches to match the pairwise origin destination demand using approaches such as gravity models, bi-level optimization and continuous equilibrium approach. Finally, studies of air travel demand have primarily employed cross-sectional data for estimating demand. In fact, we only found 3 studies (Li & Wan, 2019; Suryani et al., 2010; Loo et al., 2005) that considered air travel demand at the airport level employing data from multiple time points.

2.2 Contributions of the Current Study

While earlier research has offered significant insights on airline travel demand, there is scope for enhancing our understanding of factors influencing airline demand. The ***first contribution*** of our study to the literature arises from spatial and temporal data enhancement of airline demand data from Bureau of Transportation Statistic (BTS). Spatially, the proposed research is conducted at the disaggregate resolution of airport to better incorporate the local factors in modeling airline demand. Earlier studies that conducted airport level prediction analysis have employed a small number of airports in the US (with the highest number of airports considered being 176[2]). In our study, we conduct our analysis considering 510 airports across the country. For these airports, we augmented the airline demand data with a host of independent variables including demographic characteristics and built environment characteristics at metropolitan statistical area (MSA), spatial and temporal factors. Temporally, the current study examines airline demand at a quarterly level for five annual time points (2010, 2012, 2014, 2016 and 2018). Thus, for every airport, we have 20 observations (5 years * 4 quarters per year). Also, in our study we consider two airport level variables - arrivals and departures. Given the obvious interaction between these two variables, we develop a bivariate multiple time period framework that recognizes the influence of common unobserved factors.

The presence of multiple dependent variables and repeated observations requires the analysis methodology to accommodate for the influence of observed and unobserved factors affecting airline demand. The inclusion of observed factors within the model framework is reasonably straightforward. However, unobserved effects in the current context provide multiple levels of hierarchies including airport level, airport – year, airport – quarter, quarter only, departures and arrivals. The reader would note that in some cases there is an apparent nesting across the hierarchies while in other cases there is some overlap. The ***second contribution*** of the

[2] Li & Wan, 2019 considered 449 airports in their analysis. However, their approach involved a bi-level optimization model that is different from the proposed data driven exercise.

research is on empirically examining the appropriate hierarchy of unobserved factors that affect airline demand. *Finally*, earlier research has predominantly considered linear regression and its variants as a framework for such analysis. This is expected due to continuous nature of airline demand variables (such as natural logarithm of airline demand). However, linear regression models 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. Further, it is far from straightforward to test for polynomial terms for all variables. To address this limitation, we recast a recently developed model structure referred to as the grouped response framework for developing a non-linear regression framework that is analogous to the linear regression model system without the restrictions of linear regression (Tirtha et al., 2020; Bhowmik et al., 2019; Rahman et al., 2019). The proposed non-linear system is a recasting of the group generalized ordered probit (GGOP) model. In the traditional GGOP model, the ordered alternatives are modeled by estimating the threshold parameters that demarcate the different alternatives. For identification reasons, the variance of the GGOP error term is normalized to 1. However, in our current context, the data is a continuous value, and the demarcations can be predefined. To elaborate, we are translating the scale of the latent propensity to actual observed data. Thus, in the proposed approach, with observed thresholds, we can estimate the variance of the error term. The only data processing required is categorizing the data appropriately. If the data are finely categorized the model will represent a non-linear version of the traditional linear regression. In fact, we can establish that the proposed non-linear system subsumes the linear regression model system. Further, the proposed framework can be employed to generate a prediction output that is analogous to the linear regression model (details presented in Section 2.3).

Table 2.1 Summary of Literature Review

Study (Study region)	Demand resolution (dependent variable definition)	Objectives	Methodology	Independent Variables Considered				
				Socio-Demo.	Socio-Econ.	Built Env.	Service Factors	Lag Variable
Li & Wan, 2019 (US; 2017)	Airport (Departures)	Model originating air travel demand and its geographical distribution	Bi-level optimization model	Yes	Yes	No	No	No
Mostafaeipour et al., 2018 (Iran; 2011-2015)	Regional (Pairwise; total passenger)	Predict air travel demand	Artificial neural network	Yes	Yes	No	No	No
Zhou et al., 2018 (22 airports, Western Australia; 2016-2017)	Airport (Pairwise; total available seats)	Model air travel demand and find the effects of catchment area on the factors	Gravity model	Yes	Yes	Yes	Yes	No
Valdes, 2015 (32 middle income countries; 2002-2008)	Regional (Total passenger)	Find air travel demand determinants	Linear regression	No	Yes	No	No	No
Chang, 2014 (Countries in APEC region; 2006-2007)	Regional (Pairwise; total passenger)	Identify determinants of air passenger flows	Non-parametric multivariate adaptive regression spline	No	Yes	No	Yes	No
Chi, 2014 (US and 11 other countries; 2012)	Regional (Arrivals and departures)	Identify socio-economic factors on air travel demand	Autoregressive lag modeling approach	No	Yes	No	No	Yes
Kalić et al., 2014 (Serbia, 2001-2011)	Regional (Pairwise; Total passengers)	Model trip generation and trip distribution	Fuzzy models	Yes	Yes	No	No	No
Li et al., 2013 (US; 1995)	Airport (Pairwise; total passengers)	Estimate historical air travel demand	Route-based optimization model	No	No	No	Yes	No
Ba-Fail et al., 2000 (Soudi Arabia; 1971-1994)	Regional (Total passengers)	Estimate domestic air travel demand	Regression analysis	Yes	Yes	No	No	No
Hwang & Shiao, 2011 (Taiwan; 2007)	Airport (Pairwise; air cargo)	Determine the factors of international air cargo flows	Gravity model	Yes	Yes	No	Yes	No

Study (Study region)	Demand resolution (dependent variable definition)	Objectives	Methodology	Independent Variables Considered				
				Socio-Demo.	**Socio-Econ.**	**Built Env.**	**Service Factors**	**Lag Variable**
Carson et al., 2011) (US; 1990-2004)	Regional and airport (logarithm of departures/population)	Forecast originating air travel demand	Quasi-AIM approach	No	Yes	No	No	Yes
Suryani et al., 2010 (Taiwan; 1996-2007)	Airport (Total passengers)	Forecast air passenger demand	System dynamics model	Yes	Yes	No	Yes	No
Endo, 2007 (US and Japan; 2000-1992)	Regional (Pairwise; import and export)	Identify effect of bi-lateral aviation framework on air service imports	Regression analysis	No	Yes	No	Yes	No
Grosche et al., 2007 (Germany and 28 European countries; 2004)	Regional (Pairwise, total passengers)	Model air passenger volume between cities	Gravity model	Yes	Yes	No	Yes	No
Loo et al., 2005 (Hong Kong–Pearl River Delta region; 2000)	Airport (passengers/year)	Model geography of air passenger flows	Continuous equilibrium approach	No	Yes	No	No	No
Wei & Hansen, 2006 (Hub Airports, US; 2000)	Airport and airlines (Pairwise; logarithm of departures)	Model aggregate air passenger traffic	Log-linear demand model	Yes	Yes	No	Yes	No
Matsumoto, 2004 (Asia and outside Asia; 1998)	Regional (Pairwise; Total passengers and cargo)	Identify the pattern of international air passenger and cargo flows	Gravity model	Yes	Yes	No	Yes	No
Coldren et al., 2003) (US; 2000)	Air carrier (Pairwise; Total passengers)	Model market share of air carriers	Aggregate multinomial logit	No	No	No	Yes	No
Abed et al., 2001 (Saudi Arabia; 1971-1992)	Regional (Total passengers)	Model the demand for international air travel	Stepwise regression analysis	No	Yes	No	No	No
Rengaraju & Arasan, 1992 (40 city pairs, India; 1986)	Regional (Pairwise; total passenger)	Model demand of air travel	Stepwise multiple linear regression	Yes	Yes	No	No	No

2.3 Econometric Methodology

2.3.1 Model Formulation

Let q (q = 1, 2,…, Q) be an index to represent airports, r represent the demand dimension (r =1 represents arrivals and r =2 represents departures), t (t = 1, 2, 3,…, $T = 5$) represent the different years, l (l=1, 2, 3,…., $L = 4$) represent different quarters and j (j = 1, 2, 3,…, $J = 14$) be an index to represent the logarithm of quarterly passenger arrivals or departures data. We consider fourteen categories for the air travel demand analysis and these categories are: Bin 1 = ≤3; Bin 2 = 3-4; Bin 3 = 4-5, Bin 4 = 5-6, Bin 5 = 6-7, Bin 6 = 7-8, Bin 7 = 8-9, Bin 8 = 9-10, Bin 9 = 10-11, Bin 10 = 11-12, Bin 11 = 12-13, Bin 12 = 13-14, Bin 13= 14-15 and Bin 14 = >15. Then, the equation system for modeling demand may be written as follows:

$$D_{qrtl}^{*} = \left(\alpha_{r}' + \gamma_{qr}'\right)x_{qrtl} + (\eta_k)x_{qrtl} + \varepsilon_{qrt}, D_{qrtl} = j \ \ if \ \psi_{j-1} < D_{qrtl}^{*} \leq \psi_j \qquad (2.1)$$

In equation 2.1, D_{qrtl}^{*} is the latent (continuous) propensity for total airline demand dimension r at airport q, for the year t and quarter l. This latent propensity D_{qrtl}^{*} 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 = 14 and thus the 15 ψ values are as follows: -∞, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 and +∞. x_{qrtl} is a matrix of attributes that influence passenger arrivals and departures (including the constant); α is the vector of coefficients corresponding to the attributes and γ_q is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element of x_{qtl}. Further, ε_{qrt} is an idiosyncratic random error term assumed independently normally distributed with variance λ_{Dr}^{2}.

The variance vectors for arrivals and departures are parameterized as a function of independent variables as follows: $\lambda_{Dr} = \exp\left(\theta'_r x_{qrtl}\right)$. The parameterization allows for the variance to be different across the airports accommodating for heteroscedasticity. Finally, to allow for alternative specific effects, we also introduce threshold specific deviations in the model as $\rho_{jr} = \tau'_{jr} x_{qrtl}$.

η_k represents the vector of coefficients representing the impact of common unobserved factors that jointly influence quarterly passenger arrivals and departures across repetition level k. As discussed earlier, in the current study context, we estimate η_k for different levels (k) of repetition measures including airport specific, year specific, quarter specific, airport-year specific, airport-quarter specific and year-quarter specific. The flexibility offered by testing for unobserved heterogeneity enhances the model development exercise. In accommodating unobserved effects at different repetition levels, random numbers are assigned to the appropriate observations of the repetition measures. For example, at airport level, we have 510 airports. Thus, in evaluating unobserved effect at the airport level, 510 sets of different random numbers are generated specific to 510 airports and assigned to the data records based on their airport ID. The random numbers are assigned for other repetition levels following the same analogy in estimating the model. The reader would note that the multiple levels identified here also allows for the joint correlation across the two dependent variables (arrivals and departures). For instance, at observational level (airport-year-quarterly), this η_k will be different across the observations but same across the two dependent variables which implies that the unobserved factors that increase the propensity for arrivals for a given reason, also increase the propensity for departures. Thus, the proposed framework by allowing for additional flexibility allows the analyst to avoid conflation of unobserved effects on quarterly arrivals and departures at an airport for different years.

To complete the model structure of the Equation 2.1 and Equation 2.2, it is necessary to define the structure for the unobserved vectors γ_{qr} and η_k. In this paper, we assume that the three vectors are independent realizations from normal distributions as follows: $\gamma_{qr} \sim N(0, \sigma_r^2)$ and $\eta_k \sim N(0, \varrho^2)$.

With these assumptions, the probability expressions for the air travel demand category may be derived. Conditional on γ_{qr} and η_k the probability for airport q to have arrivals and departures in category *j* in year, *t* and quarter, *l* is given by:

$$P\left(D_{qrtl}\right)|\gamma, \eta = \Lambda\left[\frac{\psi_j - \left((\alpha_r' + \gamma_{qr}')x_{qrtl} + (\eta_k)x_{qrtl} + \rho_{jr}'\right)}{\lambda_{Dr}}\right] - \Lambda\left[\frac{\psi_{j-1} - \left((\alpha_r' + \gamma_{qr}')x_{qrtl} + (\eta_k)x_{qrtl} + \rho_{j-1,r}'\right)}{\lambda_{Dr}}\right] \quad (2.2)$$

where Λ (.) is the cumulative standard normal distribution. The complete set of parameters to be estimated in the bivariate model system of Equations 2.2 are α_r, τ_r and θ_r vectors and the following standard error terms: σ_r and ϱ. Let Ω represent a vector that includes all the standard error parameters to be estimated. Given these assumptions the joint likelihood for airport level quarterly arrivals and departures is provided as follows:

$$L_q|\Omega = \prod_{t=1}^{T} \prod_{l=1}^{L} \prod_{r=1}^{2} \prod_{j=1}^{J} \left[P\left(D_{qrtl}\right)|\gamma, \eta\right]^{d_{qrtlj}} \quad (2.3)$$

where d_{qrtlj} are dummy variables taking a value of 1 if an airport q has the demand dimension, r within the j^{th} category for year, t and quarter, l and 0 otherwise. Finally, the unconditional likelihood function may be computed for airport q as:

$$L_q = \int_{\Omega} (L_q|\Omega) d\Omega \tag{2.4}$$

Now, we can express the likelihood function as follows:

$$\text{LL} = \sum_{q=1}^{Q} \ln L_q \tag{2.5}$$

The likelihood function in Equation 2.5 involves the evaluation of a multi-dimensional integral of size equal to the number of rows in Ω. We apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function (see Bhat, 2001; Yasmin & Eluru, 2013 for more details).

2.3.2 Model Prediction

In the preceding discussion we presented the model estimation approach. In this sub-section, we outline the formula for generating the demand prediction from the proposed model. The approach recognizes that the continuous latent propensity score (D^*_{qrtl}) generated serves as the estimate of airline demand. However, in the presence of alternative specific variables (ρ_{jr}), the latent propensity score needs to be adjusted accordingly. The resulting equation for continuous demand from the proposed model is expressed as follows:

$$p_{qrtl} = (\alpha'_r + \gamma'_{qr})x_{qrtl} + (\eta_k)x_{qrtl} + \sum_{j=2}^{J}(\alpha'_r x_{qrtl} > (\psi_j - \rho_{jr})) \times \rho_{jr} \tag{2.6}$$

where, p_{qrtl} represents the total airline demand for dimension *r*, at airport q, for the year *t* and quarter *l* and x_{qrtl} is a matrix of attributes that influence passenger arrivals and departures. p_{qrtl} generated will allow us to estimate all measures of comparison applicable for linear regression such as squared residuals, R^2 and adjusted R^2.

2.3.3 Equivalent Log-Likelihood Generation Using Linear Regression

The adjusted R^2 measure represents the squared error in the model. However, it is worth noting that the squared error might not penalize the error in observations adequately. To develop a more reliable comparison metric to investigate the model performance, an equivalent linear regression log-likelihood was generated. The reader would note that linear regression model log-likelihood represents the probability density function of the difference between the observed and predicted value. However, in the proposed model, we do not differentiate between any values within each category. Thus, a direct comparison of log-likelihoods is not appropriate. Hence, we present an equivalent log-likelihood that allows for an appropriate comparison. The probability for airport *q* to have arrivals and departures in category *j* in year, *t* and quarter, *l* using linear regression model is given by:

$$P(D_{qrtl}) = \Lambda\left[\frac{\psi_j - (\omega'_r x_{qrtl})}{\kappa_r}\right] - \Lambda\left[\frac{\psi_{j-1} - (\omega'_r x_{qrtl})}{\kappa_r}\right] \tag{2.7}$$

where, ω and κ^2 represent the vector of coefficients and the error variance respectively estimated from the linear regression model and ψ is same as defined earlier in Equation 2.1. The probability thus generated is employed to compute the likelihood function following same equations as presented in 2.3, 2.4 and 2.5.

2.4 Dataset Description

The airport demand data are sourced from the airline origin and destination survey conducted by Bureau of Transportation Statistics (BTS). BTS provides detailed information about 10% of the tickets collected from domestic and international airlines operating in the US. For our current analysis, we confined our attention to the domestic air travelers from 2010 to 2018 across the 51 states in US covering five major regions including South, West, North-East, Mid-West and Pacific regions. Further, we consider both arrivals and departures at an airport for every quarter over the study period. Hence, passenger trips in origin and destination survey are aggregated at quarters and airports and scaled appropriately (as they represent 10% of the total domestic trips) to estimate the quarterly airport level travel demand. In the airport selection process, our focus was to consider all of the public-use airports located in the US. In this effort, we consider 510 airports for which itinerary information are available in origin and destination survey. We ignored the smaller airports that do not have itinerary information available. For the selected airports, we extract the demand data for every two years interval (2010, 2012, 2014, 2016 and 2018). The reader would note that some airports did not have all 20 records for various reasons (such as airports that were opened for passengers at a later time or closed in the time frame). After cleaning the data, we obtain a total of 8,477 observations for estimation.

In preparation of dependent variables, we performed log transformation of arrivals and departures, and then considered 14 categories (≤3, >3-4, >4-5, >5-6, >6-7, >7-8, >8-9, >9-10, >11-12, >12-13, >13-14, >14-15, >15) of the transformed variables. Distribution of the dependent variables are shown in Figure 2.1. The transformed variable reasonably represents a normal distribution.

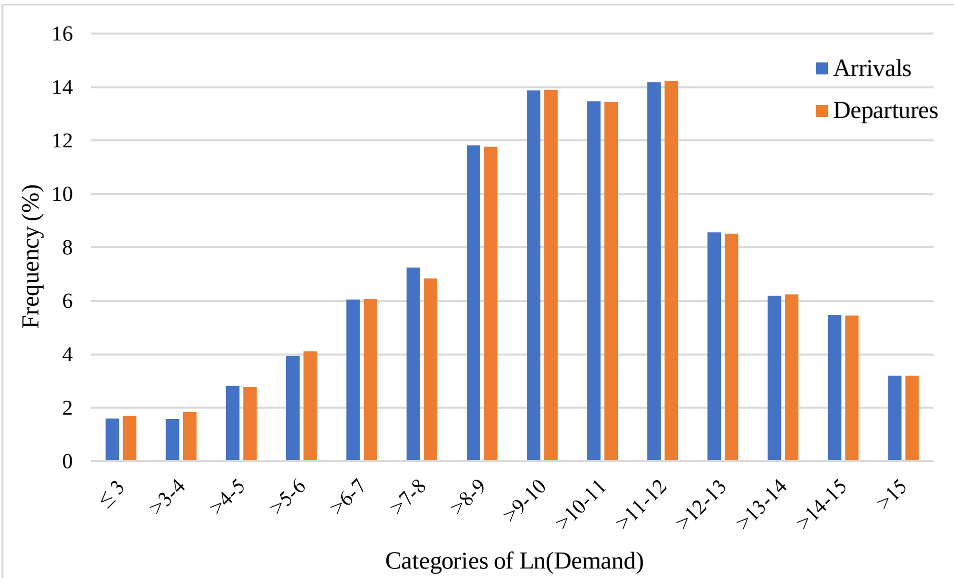


Figure 2.1 Distribution of the Dependent Variables

The BTS airline data is also augmented with a host of independent variables. These variables are sourced from American Community Survey (ACS) and other secondary sources (County health ranking and roadmaps (Roadmaps, 2020) for crime data; Insider, 2020). Independent variables are grouped into four broad categories, namely, demographic characteristics, built environment characteristics, spatial and temporal factors. Demographic factors include population, household median income, employment, out of state employment and education level of the residents in the corresponding metropolitan statistical area (MSA). Built environmental characteristics include number of airports in 50mile buffer area around the airport of interest and tourism ranking of the corresponding state. Spatial factors include location of the airport in terms of region including south, north-east, west, mid-west and pacific region. Temporal factors include year and quarter of the analysis. Detailed descriptions of functional form and summary statistics of the independent variables are provided in Table 2.2 for categorical and continuous variables.

Table 2.2 Description of the Independent Variables

Categorical Independent Variables			
Variables	**Definition**	**Frequency**	**Percentage**
Demographic characteristics			
Education Status			
High	Percentage of adults not having high school degree in the MSA <=12%	4713	55.597
Low	Percentage of adults not having high school degree in the MSA >12%	3764	44.403
Built environment factors			
Tourist attraction			
Top10	The state is among top 10 tourist attraction states	2252	26.566
Bottom10	The state is among bottom 10 tourist attraction states	948	11.183
Others	The state is other than top and bottom tourist attraction states	5277	62.251
Spatial Factors			
Region			
South		2465	29.100
North-East		1079	12.700
West		2176	25.700
Mid-West		1958	23.100
Pacific		799	9.426
Temporal factors			
Quarter			
Quarter 1	January-March	2101	24.785
Quarter 2	April-June	2142	25.268
Quarter 3	July-September	2128	25.103
Quarter 4	October-December	2106	24.844
Continuous Independent Variables			
Variables	**Definition**	**Mean**	**Min/Max**
Socio-demographic factors			
Population	Population in million in corresponding MSA	1.101	0.013/20.031

Continuous Independent Variables			
Variables	**Definition**	**Mean**	**Min/Max**
Median Income	Median income in 100K in corresponding MSA	0.541	0.276/1.147
Employment	Ln(number of workers in thousands in corresponding MSA)	4.848	2.029/9.166
Out of State Employment	Fraction of job holders in corresponding MSA working out of state	0.029	0.000/0.273
Built environment factors			
Number of airports	Ln(Number of airports in 50 mile buffer area)	1.711	0.000/3.664
Ordinal Independent Variables			
Temporal factors			
Year	Ordinal year variable with 2010 as the base year	3.900	0.000/8.000

2.5 Model Selection

The empirical analysis begins with comparing the performance of the proposed group generalized ordered probit (GGOP) model with the performance of a linear regression model. The reader would note that the two model systems are generally estimated using different approaches. The linear regression model is estimated using the least squares estimator (and evaluated based on adjusted R^2) and the GGOP model employs a log-likelihood maximization procedure (evaluated using log-likelihood). In our effort to compare the two frameworks, we build equivalent measures for the two models from both approaches i.e. generate adjusted R^2 and log-likelihood for both models (equations presented in section 2.3.2 and 2.3.3).

The linear regression model for arrivals (departures) with 12 (12) parameters resulted in an adjusted R^2 value of 0.401 (0.397). For the GGOP arrivals (departures) model with 15 (16) parameters resulted in an adjusted R^2 value of 0.408 (0.405). The reader would note, even after accounting for the additional parameters in the GGOP framework, we observe that GGOP model outperforms the linear regression model structure.

The log-likelihood and Bayesian Information Criterion (BIC) value for the equivalent linear regression framework appropriately aggregated to reflect the GGOP structure is -37,363.3 (with 24 parameters) and 74,876.2, respectively. The log-likelihood and BIC value for the proposed GGOP system is -37,128.0 (with 31 parameters) and 74,449.3, respectively. The comparison of the adjusted R^2, log-likelihood and BIC measures clearly illustrate the superiority of the proposed model structure for the present empirical case study.

After establishing the superiority of the GGOP framework (versus the linear regression approach), we estimate advanced model structures in the GGOP regime to account for the presence of two dependent variables and repeated measures. Prior to doing this, we recognized that the arrivals and departures models have similar coefficients for a substantial number of parameters. Hence, to arrive at a parsimonious specification, we restrict the variables with close parameter values and re-estimate the model. The re-estimated model offers no significant loss of fit. Finally, with this specification we estimate the joint panel GGOP model. The fit measures - log-likelihood (parameters) - for the three models are as follows: Independent GGOP model: -37,128.0 (with 31 parameters); 2) Restricted GGOP model: -37,128.2 (with 19 parameters) and 3) Joint Panel GGOP model: -30,175.2 (with 20 parameters). We also compute the BIC value for these three frameworks to determine the best model. The BIC values for the three models are as follows: a) 74,449.3, b) 74,374.9 and c) 60,475.1. Based on the BIC values, the joint panel model that accommodates for the presence of unobserved heterogeneity significantly outperforms the respective independent models highlighting the importance of accommodating for the influence of common unobserved factors affecting the two dependent variables (and their repeated measures). For the sake of brevity, only the joint panel GGOP model results are presented in the dissertation.

2.6 Estimation Results

In the model estimation process, we explored various transformations of the independent variables and chose the best transformation based on model fit. Table 2.3 shows the effects of exogenous variables on passenger arrivals and departures. Positive (negative) coefficients in the model indicate that an increase (decrease) of a variable increases (decreases) the propensity for higher demand. From Table 2.3, the reader would note the variables for arrivals and departures offer identical parameters as they were restricted to be the same based on initial estimations that offered very close values across the two variables. Given the similarity, we will discuss the variable effects for both arrivals and departures together by variable groups.

2.6.1 Demographic Characteristics

Among the various demographic characteristics considered in the model, population, median income in an MSA, out of state employment and education status offer significant impact on the quarterly demand. As evident from Table 2.3, we can see that population - a surrogate for exposure is positively associated with increased demand (arrivals and departures) (please see Zhou et al., 2018 and Grosche et al., 2007 for similar findings). Further, results show that the air travel demand is positively associated with median income in an MSA. Increased income, in general, corresponds to increased affordability for personal travel and higher business activity in the region. Thus, it is possible that airports in MSA’s with higher median income are likely to have higher demand profiles.

The variable specific to out of state employment represents the percentage of employees working out of state and reveals a negative association with the air travel demand. This may indicate that as out of state workers are not actively present in the MSA, consequently, increase of such population may reduce total number of passenger arrivals and departures. Further, from the results it appears that education status in an MSA is an important determinant influencing the air travel demand. Results show that if percentage of adults without high school degree is more than 12%, then air travel demand decreases.

2.6.2 Built Environment Characteristics

The variable number of airports in a 50 mile buffer represents the number of available airports in close proximity (50 mile radius) of an airport. Interestingly, we found that increased number of airports in 50 mile buffer results in higher air travel demand in a MSA. Further, we considered the tourism status of an MSA in our analysis as demand for travel to these destinations can increase air travel demand. For this purpose, we identify the top and bottom 10 desirable states with respect to tourism activity and use that indicator variables as predictors in our model system. As expected, we find that the likelihood of higher air travel demand is greater in an airport located in top 10 tourists' attraction states while a reduced propensity for air demand is observed for an airport located in the least 10 visiting states.

2.6.3 Spatial Factors

Location of the airports in terms of US region has a significant effect on the total number of arrivals and departures through those airports. In general, compared to the airports in the west and the mid-west region, the demand is observed to be higher for an airport in the south region. On the other hand, airports in the north-east and pacific regions experience lower level of demand.

2.6.4 Temporal Factors

Quarterly effects are found to be significant in the model and the results indicate that travel demand is lowest for quarter 1 (January – March) and highest for quarter 3 (July – September). These trends can be attributed to presence of seasonal variation in air travel demand.

2.6.5 Category Specific Deviations

The proposed model also allows for category specific deviations on various predefined thresholds. In our air passenger arrivals and departures estimation, we consider various category specific deviations based on model fit and sample sizes across each trip count categories. The estimation results of these parameters are reported in the third-row panel of Table 2.3. These deviation parameters are similar to a constant in discrete choice models and do not have an interpretation after incorporating other variables.

2.6.6 Effect of Unobserved Factors

In our proposed model, we estimated unobserved effects at multiple levels: airports, year, quarter, airport – year and airport – quarter. Among different levels we considered, we found that the airport – year and airport – quarter level effects have significant influence on air travel demand. The estimation results of these standard deviations are presented in last row panel of Table 2.3. The significant standard deviation parameters at different repetition measures provide evidence toward supporting our hypothesis that it is necessary to incorporate these unobserved effects in examining air travel demand. These variables indicate that the air passenger arrivals and departures may vary for different airports based on the unobserved effects specific to different levels.

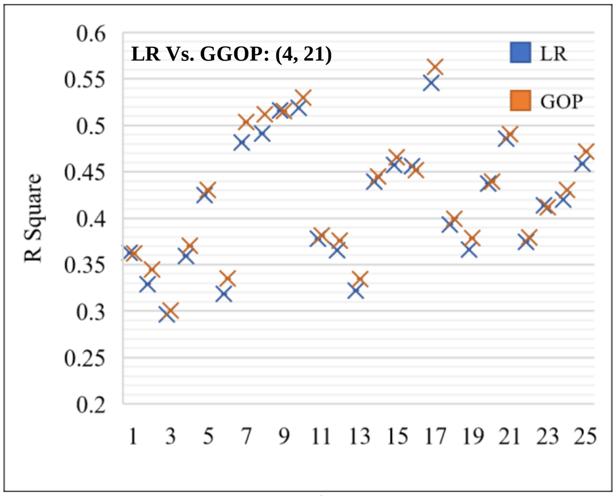
Table 2.3 Model Estimation Results

Variables	Arrivals		Departures	
	Estimate	**t-statistic**	**Estimate**	**t-statistic**
Propensity Components				
Constant	5.6422[a]	48.4330	5.6235	48.2630
Demographic Factors				
Population	0.2681	32.0980	0.2681	32.0980
Median income	3.5463	17.1210	3.5463	17.1210
Out of state employment	-0.6236	-1.7920	-0.6236	-1.7920
Education Level (Base: High (% of adults not having high school degree <=12%))				
Low	-0.6030	-15.0460	-0.6030	-15.0460
Built Environment Factors				
No. of airports	1.3622	41.6110	1.3622	41.6110
Tourist's Attraction (Base: Others)				
Top10	0.8160	15.5210	0.8160	15.5210
Bottom10	-0.4552	-6.7810	-0.4552	-6.7810
Spatial Factors				
Region (Base: West and Mid-West)				
South	1.1928	20.9990	1.1928	20.9990
North-East	-1.4536	-21.4080	-1.4536	-21.4080
Pacific	-2.9293	-36.1590	-2.9293	-36.1590
Temporal Factors				
Quarter (Base: Quarter 1)				
Quarter 2&4	0.1161	2.8440	0.1161	2.8440
Quarter 3	0.2044	4.5600	0.2044	4.5600
Variance Components				
Constant	0.3767	42.6490	0.3855	43.6470
Threshold Specific Constant				
Threshold 11	-0.1275	-5.7110	-0.1309	-5.8620
Threshold 13	-0.4185	-7.7620	-0.4282	-7.9420
Threshold 14	-1.6344	-17.2810	-1.6498	-17.3010
Unobserved Effects				
Variables	**Estimate**		**t stat**	
Airport-Year specific effect	1.9572		38.3520	
Airport-Quarter specific effect	0.3668		19.8320	

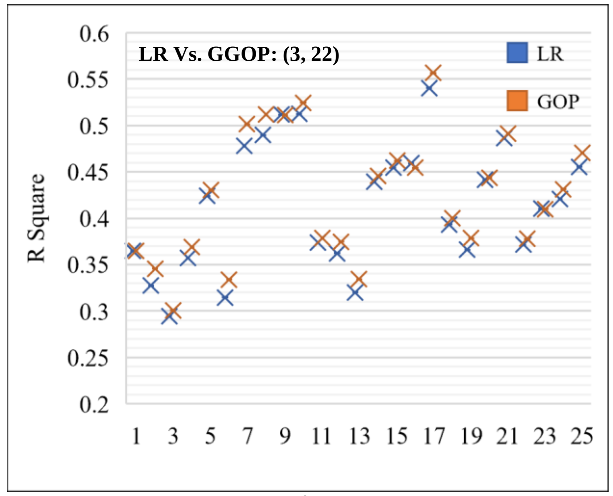
[a] = Significant at 90% confidence level

2.7 Model Validation

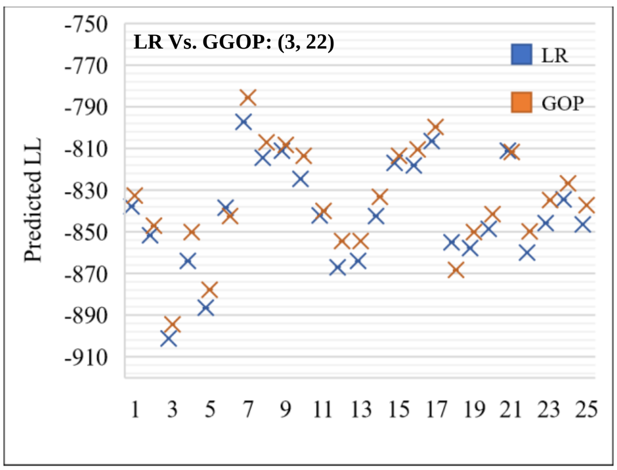
The holdout sample with quarterly passenger arrivals and departures for year 2017 is used to perform the validation test. The validation set consists of 1,609 observations for 415 airports. To test the predictive performance of the proposed model, a validation exercise is performed in this study following the same procedures outlined in Section 2.5. First, we compared the performance of the traditional linear regression with the independent GGOP model. To perform the validation analysis, 25 data samples of 100 airports each, are randomly generated from the hold out validation sample consisting of 415 airports. Predicted R^2 and Log-likelihood values for linear regression model and GGOP model are plotted in Figure 2.2. Figure 2.2 clearly highlights the enhanced performance of the GGOP model over LR across most of the samples for both arrival and departure rate. Specifically, for the arrival model, the GGOP model performs better than LR model in 43 out of 50 cases (R^2: 21 and LL: 22) while for the departure model, the GGOP model performs better in 45 cases (R^2: 22 and LL: 23). While the improvements in predicted R^2 might be small, the consistency of the improved performance of the GGOP model indicates its superiority over the LR model. Further, we compare the distribution of the residuals for linear regression model and the proposed model in Figure 2.3. From Figure 2.3, we can see that the proposed model performs better than traditional linear regression model in terms of RMSE measure. Subsequently, we compared the performance of the three GGOP model systems (LL and BIC): (1) independent GGOP: -6972.12 and 14,131.12, (2) restricted GGOP: -6972.13 and 14,058.80 and (3) joint panel GGOP: -5868.40 and 11,857.37. The LL and BIC values computed using the validation dataset also clearly highlights the superiority of the joint panel GGOP model relative to the other two systems.



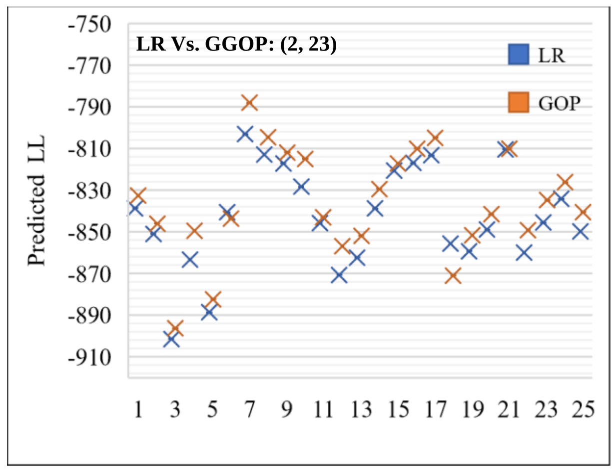
(a) Predicted R^2 for arrivals



(b) Predicted R^2 for departures

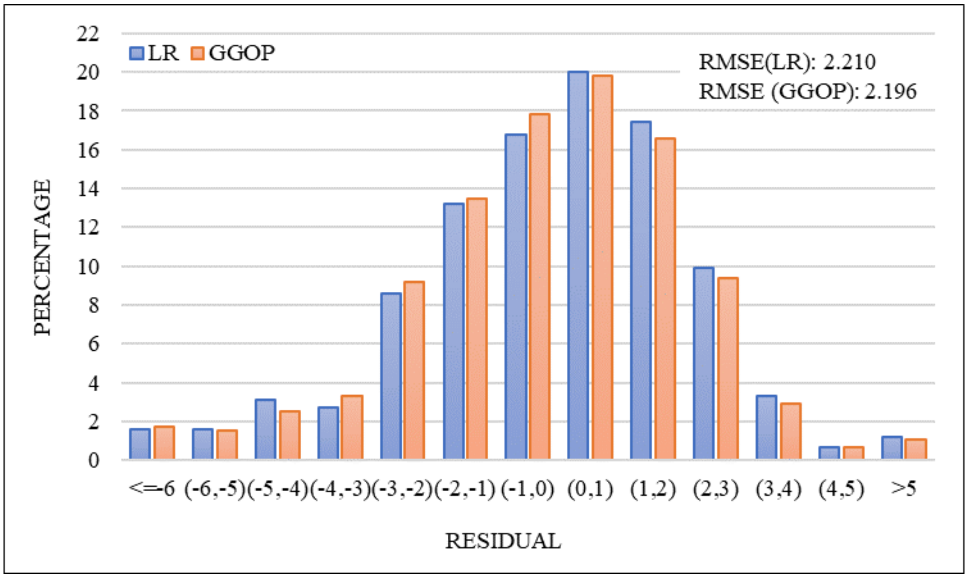


(c) Predicted LL for arrivals

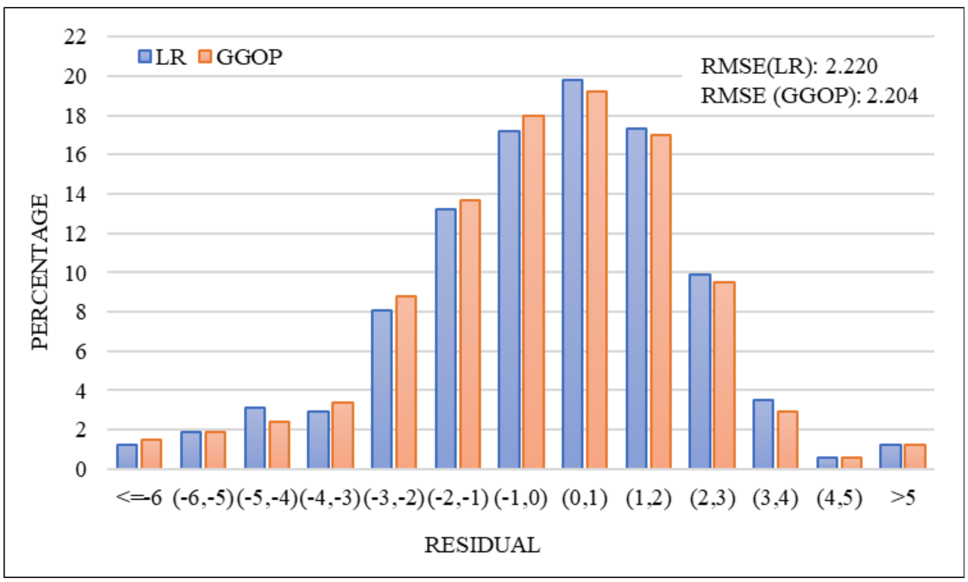


(d) Predicted LL for departures

Figure 2.2 Predicted R^2 and LL Comparison between LR and GGOP Model



(a) Arrival Model



(b) Departure Model

Figure 2.3 Distribution of the Residuals

2.8 Policy Analysis

In order to highlight the effect of various attributes on air passenger arrivals and departures, an elasticity analysis is also conducted (see Eluru & Bhat, 2007 for a discussion on the methodology for computing elasticities). We investigate the change in demand, due to the change in selected independent variables. To elaborate, we compute the percentage change of aggregate probability of the demand categories because of the change in the factors considered. The variables considered include MSA level population, household median income, out of state employment, education status, number of airports in close proximity, tourism related variables and quarter of analysis. The results of elasticity analysis are presented in Table 2.4. Several observations can be made from the results. First, airport location in tourism driven states has a significant impact on air travel demand. Further, we observe that increased air travel demand is associated with number of airports in proximity, population and median income. Second, air travel demand is adversely affected by MSA level education status (higher proportion of adults without high school education) and state's presence in the bottom tier of tourist attractions. These findings illustrate how the proposed approach can be employed to understand how air travel demand is affected by various independent variables. For instance, examining population and median income trends over time will allow planning agencies to expect changes to air travel demand. Further, from our analysis, it is also apparent that the tourism rank of a state has a substantial effect on air travel demand. Thus, marketing the tourist attractions in a state might be a beneficial investment.

Table 2.4 Elasticity Analysis Results

Arrivals categories	Bins													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Population	-1.91[b]	-1.13	-0.74	-0.56	-0.49	-0.46	-0.45	-0.42	-0.39	-0.31	-0.18	-0.24	-0.35	23.66
Median income	-37.41	-30.71	-25.26	-20.36	-15.79	-11.32	-6.82	-2.34	1.97	6.05	9.22	11.26	10.54	19.38
Out of state employment	0.31	0.28	0.24	0.18	0.14	0.10	0.06	0.02	-0.02	-0.06	-0.09	-0.10	-0.03	-0.27
Education Status (Low)	183.12	133.79	101.85	75.75	53.36	34.10	17.83	4.12	-7.75	-18.81	-27.69	-32.32	-25.45	-45.00
No. of airports	-10.54	-12.39	-13.85	-14.58	-14.05	-12.07	-8.82	-4.62	0.30	5.85	10.92	14.87	15.27	32.43
Top10	-113.58	-96.89	-84.41	-72.42	-59.54	-45.27	-29.67	-12.78	5.43	25.34	42.53	52.02	44.11	60.03
Bottom10	122.85	93.18	73.40	56.82	41.78	27.59	14.34	2.54	-7.30	-15.23	-20.49	-22.80	-19.85	-29.87
Quarter 2&4	-23.57	-19.33	-15.94	-12.79	-9.75	-6.77	-3.86	-1.09	1.44	3.71	5.38	6.34	5.67	8.51
Quarter 3	-37.77	-31.63	-26.50	-21.57	-16.70	-11.82	-6.92	-2.14	2.34	6.47	9.58	11.44	10.29	15.33
Departures categories	**Bins**													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Population	-1.84	-1.10	-0.73	-0.55	-0.49	-0.46	-0.45	-0.42	-0.38	-0.31	-0.18	-0.25	-0.24	23.86
Median income	-36.87	-30.26	-24.92	-20.08	-15.57	-11.13	-6.68	-2.24	2.02	6.06	9.21	11.23	10.56	19.49
Out of state employment	0.31	0.28	0.23	0.18	0.13	0.10	0.06	0.02	-0.02	-0.06	-0.09	-0.10	-0.03	-0.27
Education Status	178.94	131.10	99.98	74.39	52.39	33.44	17.42	3.87	-7.89	-18.86	-27.64	-32.17	-25.49	-45.26
No. of airports	-10.69	-12.52	-13.91	-14.56	-13.96	-11.94	-8.69	-4.50	0.38	5.89	10.91	14.82	15.31	32.62
Top10	-112.20	-95.87	-83.59	-71.67	-58.83	-44.64	-29.14	-12.36	5.71	25.46	42.44	51.78	44.15	60.22
Bottom10	120.35	91.52	72.22	55.92	41.07	27.05	13.98	2.34	-7.38	-15.24	-20.45	-22.73	-19.89	-29.96
Quarter 2&4	-23.22	-19.05	-15.72	-12.61	-9.60	-6.65	-3.77	-1.03	1.47	3.72	5.38	6.32	5.68	8.54
Quarter 3	-37.28	-31.21	-26.15	-21.28	-16.46	-11.62	-6.78	-2.05	2.39	6.48	9.57	11.40	10.29	15.40

[b] = percentage change of aggregate probability of the demand categories

2.9 Summary

Understanding the factors affecting airline demand at US airports is important for long-term planning and operational decisions. The current study contributes to the existing literature along multiple directions. The first contribution our study to the literature arises from spatial and temporal data enhancement of airline demand data from BTS. Also, in presence of airport level variables - arrivals and departures, we develop a bivariate framework that recognizes the influence of common unobserved factors. The second contribution of the research is on empirically examining the appropriate hierarchy of unobserved factors that affect airline demand. Finally, to address the inherent limitations of traditional linear models, we employ the generalized response framework for developing a non-linear framework that subsumes the linear regression model system. In summary, the proposed research develops a joint panel group generalized ordered probit model system with observed thresholds for modeling air passenger arrivals and departures. The proposed model is estimated using airline data compiled by Bureau of Transportation Statistics for 510 airports across the US. A host of exogenous variables including demographic characteristics, built environment characteristics, spatial and temporal factors are considered in the model estimation.

The empirical analysis shows that the flexible structure of group generalized ordered probit model (GGOP) allows us to capture the non-linearity between air travel demand and its contributing factors resulting in better data fit compared to linear regression model. To arrive at a parsimonious specification, we estimated a restricted GGOP model without any significant loss of data fit. Finally, the joint panel model that accommodates for the presence of unobserved heterogeneity performs the best in terms of empirical context highlighting the importance of accommodating for the influence of common unobserved factors affecting the two dependent variables (and their repeated measures). Finally, to illustrate how the enhanced demand model allows policy agencies to understand changes to airline demand with changes to independent variables a policy analysis is conducted. The results identify important predictors for airline demand. In particular, they highlight the role of tourism in the state, regional population and median income.

However, this study is not without limitations. Augmenting the data in our research with local economic indicators and airport specific attributes might be an avenue for future research.

CHAPTER 3: EXAMINING THE IMPACT OF COVID-19 ON AIRLINE DEMAND

[3]COVID-19 pandemic has affected every facet of life and airline industry is amongst the worst affected industries. In December 2020, domestic airline demand in US is only 39.1% of the demand in December 2019. As the demand recovery starts, a high-resolution demand prediction framework that accommodates the effect of COVID-19 global and local factors on airline demand is important to build a template of potential demand recovery in the future months. Thus, this chapter presents a linear mixed model of monthly air passenger departures at the airport level that considers the effect of COVID-19 factors. In addition, a validation exercise is undertaken to see how the proposed model captures the actual demand variations. Finally, the chapter presents a potential band of airline demand recovery over time by considering three hypothetical scenarios including expected, pessimistic, and optimistic scenarios.

3.1 Earlier Studies

The literature relevant to the current study context can be categorized into three major streams: a) studies investigating the factors influencing airline demand, b) studies examining the influence of external shocks (such as September 11 attacks) or health pandemics such as Severe Acute

---

[3] 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*, 169-181. https://doi.org/10.1016/j.tra.2022.03.014

Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) and c) studies investigating the impacts of COVID-19.

The first group of studies develop airline demand prediction frameworks considering a host of independent variables. The demand prediction exercise is typically conducted at two spatial resolutions: (a) airport level (Li & Wan, 2019; Loo et al., 2005; Suryani et al., 2010; Wei & Hansen, 2006; Zhou et al., 2018) and (b) regional level (Abed et al., 2001; Chang, 2014; Chen et al., 2009; Chi, 2014; Chi & Baek, 2013; Endo, 2007; Grosche et al., 2007; Grubb & Mason, 2001; Kalić et al., 2014; Matsumoto, 2004; Mostafaeipour et al., 2018; Rengaraju & Arasan, 1992; Tsui et al., 2014; Valdes, 2015). In the former category, studies analyze passenger demand data for individual airports while in the latter category, the analysis is conducted by aggregating demand at a regional level (such as state level or census region level). Across the two spatial resolutions, the factors affecting airline demand include socio-demographic factors (population, education, age distribution), socio-economic factors (income, unemployment rate, GDP), built environment (number of trade centers, tourist attractions), level of service factors (average airfare and distance) and historical demand (considered as lag variables). In terms of mathematical frameworks employed for analyzing demand, prevalent approaches include: (a) prediction methods using data and (b) distribution or assignment methods. The majority of prediction methods focused on trip departures from the spatial unit of interest employing passenger demand models such as regression models and their advanced variants (Abed et al., 2001; Chang, 2014; Chi, 2014; Endo, 2007; Rengaraju & Arasan, 1992; Valdes, 2015), artificial neural networks (Mostafaeipour et al., 2018), Holt–Winters method (Chen et al., 2009; Grubb & Mason, 2001), seasonal autoregressive integrated moving average (Chen et al., 2009; Tsui et al., 2014; Xu et al., 2019) and fuzzy models (Kalić et al., 2014). The second set of studies match the pairwise origin destination demand using approaches such as gravity models (Grosche et al., 2007; Matsumoto, 2004; Zhou et al., 2018), bi-level optimization (Li et al., 2013; Li & Wan, 2019) and continuous equilibrium approach (Loo et al., 2005).

The second group of studies considered include research efforts that examined the impact of external shocks (such as September 11th attacks) or health shocks such as SARS and MERS on airline industry. Ito & Lee, 2005a assessed the influence of September 11 terror attacks on US airline demand using monthly observations of revenue passenger miles. The study found a sudden reduction of about 30% in demand in response to the shock. Further, the authors also found that the reduction in demand took well over 2 years to dissipate while controlling for various independent variables (such as economic and seasonal factors). In a subsequent paper (Ito & Lee, 2005b), the authors extended the work to examine the impact of the terror attack on international airline markets. The subset of studies examining health shocks also developed similar approaches. Chi & Baek, 2013 employed autoregressive distributed lag model to study relationship between economic growth and airline demand while controlling for the impact of SARS outbreak. The results indicate that SARS epidemic decreased US air passenger demand by 6%. Pine & McKercher, 2004 also studied the impact of SARS outbreak on tourism and airline industry and presented a descriptive analysis of reductions induced by the epidemic.

The third group of studies, conducted after the onset of COVID-19 pandemic, can be broadly characterized as preliminary research studying the impact of COVID-19 on airline demand. Maneenop & Kotcharin, 2020 identified three crucial announcements that triggered the airline demand reduction including (a) the first case reported outside China, (b) Italy outbreak and (c) the global pandemic declaration issued by WHO. Nižetić, 2020 performed descriptive analysis to see how COVID-19 affected air transport mobility concluding that the number of flights in the EU region dropped by more than 89% in April 2020 (relative to April 2019). Gudmundsson et al., 2020 analyzed world air transport industry employing time series models to study air traffic volume recovery timeline. The authors developed models employing economic indicators (such as Gross Domestic Product and Oil prices) and COVID-19 indicators and conclude that air transport recovery is likely to take about 2.4 years starting from 2020 with the most optimistic estimate of recovery in latter half of 2021. Gallego & Font, 2021 examined a large data of airline passenger searches and picks to evaluate airline demand and recovery patterns. The analysis using Big Data approaches suggests an L-shaped recovery as the pandemic progresses. Sun et al., 2020 employed data from 150 airlines and 2751 airports across the world to evaluate the impact of COVID-19 on airline industry between January 2020 and May 2020 employing complex network approaches. The study concluded that airport networks in the southern hemisphere experienced more significant disruptions relative to airport networks in the northern hemisphere.

3.2 Contributions of the Current Study

The review of literature highlights the exhaustive research on developing airline demand prediction frameworks. The research on measuring the impact of shocks (external or health) on airline demand focused on a retrospective analysis as opposed to offering insights for the potential recovery of demand in response to the shock. While earlier research provides the building blocks of demand prediction systems and some insights on modeling demand in the presence of shocks, these frameworks have not been employed to study demand recovery patterns.

The proposed research builds on the demand prediction frameworks at the airport level by accommodating for the influence of COVID-19. Specifically, the study contributes to our understanding of the unprecedented drop in air passenger demand by examining airline data from the Bureau of Transportation Statistics (BTS) at the disaggregate resolution of airport using a linear mixed model. The study contributes to the airline demand literature along multiple directions. First, research on COVID-19 impact on airline industry is in the nascent stages and has predominantly focused on global or regional effects. In our research, we examine the influence of COVID-19 at the disaggregate resolution of airport to incorporate the interplay of local and global factors on airline demand. The interaction between local and global factors is considered by considering global and local COVID-19 transmission, temporal indicators of pandemic start and progress, and interactions of airline demand predictors with global and local COVID-19 indicators. In our study, we conduct our analysis considering 380 airports across the country. For these airports, we augmented the airline demand data with a host of independent variables including COVID-19 related factors, demographic characteristics and built environment characteristics at the county level, spatial factors, temporal factors, and adjoining county attributes. Second, the research study employs a robust modeling framework to analyze airline demand variable. The study examines monthly airline demand (transformed to the natural logarithm) for 24 months from January 2019 through December 2020. A linear mixed model system that accommodates for the presence of repeated measures is developed. An exhaustive specification exercise is conducted to evaluate the impact of various COVID-19 factors while controlling for other attributes affecting airline demand. Finally, the proposed model is employed to undertake a scenario analysis that will allow us to provide a blueprint to the path to recovery for airline demand. The research team considers three scenarios – expected, pessimistic and optimistic – to generate the recovery patterns for airline demand. The results at the airport level were aggregated at the state or regional level by adding the demand from all airports in the corresponding state or region. These trends are presented by State and Region to illustrate potential differences across various scenarios.

3.3 Econometric Methodology

The airport level monthly departure variable is a continuous value and can be analyzed using linear regression models. However, the traditional linear regression model is not appropriate for data with multiple repeated observations. In our empirical analysis, we observe monthly airline demand at the same airport for twenty four months. Hence, we employ a linear mixed modeling approach that builds on the linear regression model while incorporating the influence of repeated observations for the same airport (Bhowmik & Eluru, 2021; Bhowmik et al., 2021). The linear mixed model collapses to a simple linear regression model in the absence of any airport specific effects.

Let z = 1, 2, …, Z = 380 be an index to represent each airport, t = 1, 2, …24 be index to represent the month for which data is compiled for each airport. The dependent variable (airport level monthly departures) is modeled using a linear regression equation with the following structure:

$$y_{zt} = \beta X_{zt} + \varepsilon_{zt} \tag{3.1}$$

where y_{zt} is the natural logarithm of monthly airline demand, X is a K×1 column vector of attributes and the model coefficients, β, is a K×1 column vector. The random error term ε_{zt}, is assumed to be normally distributed across the dataset. In our analysis, each airport is repeated 24 times, once for each month. These repetitions over months can result in common unobserved factors affecting the dependent variable. In our model, we used first order autoregressive moving average as the repeated covariance structure. The exact functional form of the covariance structure assumed is shown below:

$$\Omega = \sigma^2 \begin{pmatrix} 1 & \phi\rho & \dots & \phi\rho^{n-1} \\ \phi\rho & 1 & \dots & \phi\rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi\rho^{n-1} & \phi\rho^{n-2} & \dots & 1 \end{pmatrix} \tag{3.2}$$

The covariance structure allows for a dampening relationship over time. The parameters estimated in this correlation structure are σ, ρ and ϕ. The models are estimated in SPSS using the Restricted Maximum Likelihood Approach (REML). The REML approach estimates the parameters by computing the likelihood function on a transformed dataset. The approach is commonly used for linear mixed models (Harville, 1977).

3.4 Data Description

3.4.1 Data Preparation and Summary

In this current study, we consider monthly airline demand for 24 months from January 2019 through December 2020. The dependent variable is sourced from T-100 Domestic Market dataset provided by Bureau of Transportation Statistics (BTS). Flight passenger counts are aggregated over origin airports for each month to generate the dependent variable. For selection of the airports, we consider the top 400 busiest airports in the US. After removing airports with missing records 380 airports remain in the estimation sample. The final dataset consists of 9120 records in total (24 records for each airport). A representation of the monthly demand across the 380 airports is presented in Figure 3.1. The total demand is partitioned by region including South, West and Mid-West and rest of the country. The figure clearly illustrates the shock to the airline industry beginning in March 2020. The demand has started recovering in June 2020. However, the airline demand in December is still only a fraction of the previous year flows.

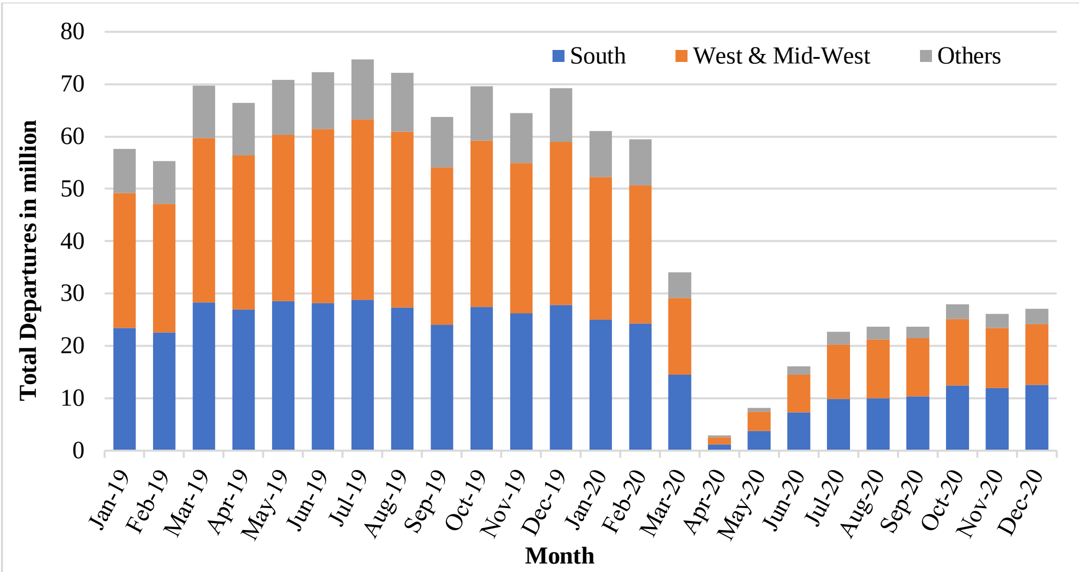


Figure 3.1 Domestic Air Passenger Departure Rate by Month and Region

A more detailed examination of demand during pandemic months (March through December) is presented in Figure 3.2. Specifically, Figure 3.2 presents the monthly percentage change in airline demand relative to the previous year. The results highlight the varying recovery patterns across the various regions. From the figures we can observe that demand in the Southern region is recovering slightly faster than the rest of the country.

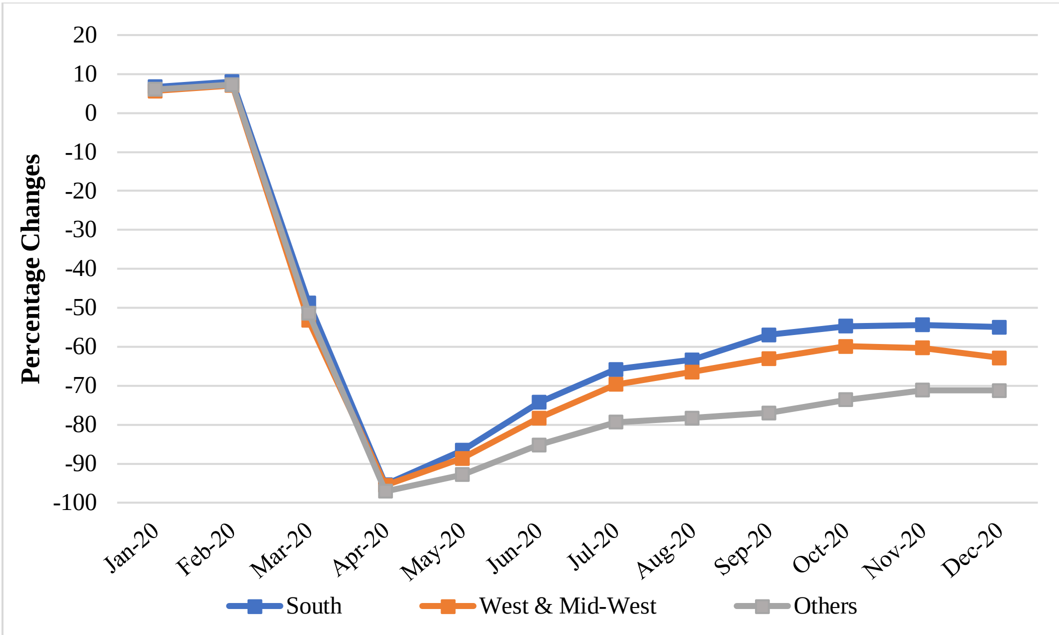


Figure 3.2 Changes of Air Passenger Demand across Different Regions

3.4.2 Independent Variable Compilation

The airline demand variable is augmented with a comprehensive set of independent variables including COVID-19 related factors, county level demographic characteristics, built environment characteristics at the county level, airport specific factors, spatial factors, temporal factors, and adjoining county attributes. COVID-19 related factors include both global and local effects of COVID-19 on airline demand. Global factors capture the change of demand across the months since the pandemic started while controlling for the number of local COVID-19 cases. In our study, we considered several binary variables as global factors including pandemic started month, May 2020 or later, July 2020 or later, October 2020 or later variables and their interactions with other variables. The local effects of COVID-19 represent the impact of county level Covid-19 cases on airline demand. In this study, we consider natural logarithm of past month's new cases at the county level of the airport as the local effect. County level monthly cases are processed from the COVID-19 dataset from Center for Systems Science and Engineering (CSSE) Coronavirus

Resource Center at Johns Hopkins University (CSSE, 2021). Total cases in the US by month from January 2020 to August 2021 is presented in Figure 3. The figure highlights the new surge in COVID-19 cases starting from July 2021.

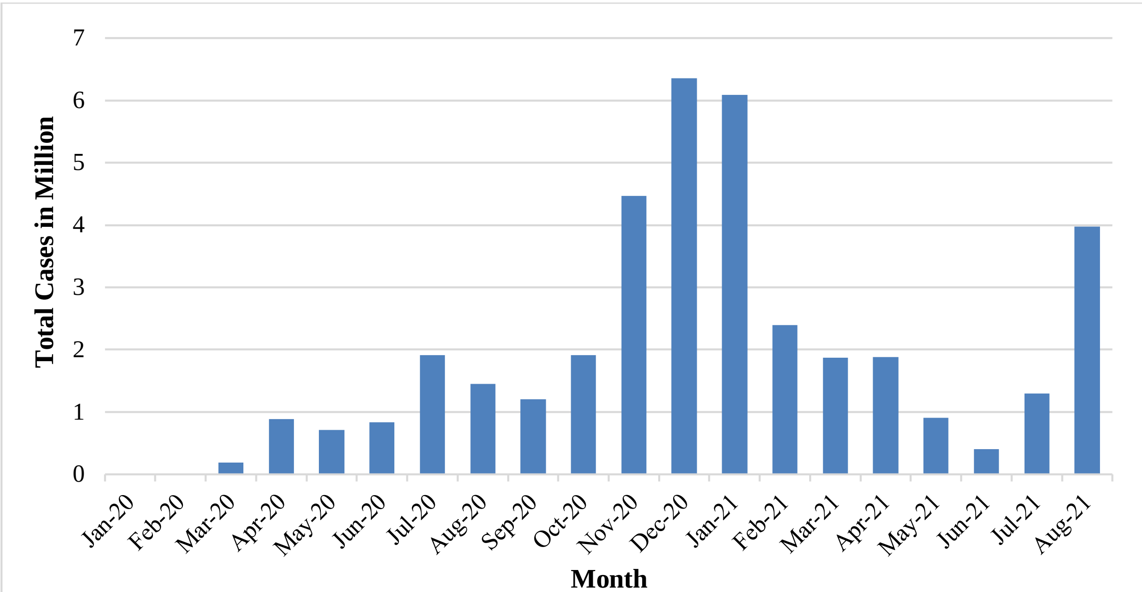


Figure 3.3 Total COVID-19 Cases by Month

County level demographic characteristics considered include population, median income, unemployment rate, percentage of senior residents and percentage of households with 2 or more vehicles. Demographic data are sourced from American Community Survey (ACS) data. Built environment characteristics tested include number of airports in 50-mile buffer area and state level tourism ranking (Insider, 2020). Airport specific factors include the type of airport. In this study, we consider a binary classification of the airport categorized as large and small airports. Operational Evolution Partnership (OEP 35) airports are marked as the large airports and remaining airports are marked as the small airports. Spatial factors include location of the airport in terms of US regions. The regions include South, North-East, West, Mid-West, and Pacific regions. Temporal factors include quarters and month of the year. Finally, we consider the effect of attributes of adjoining counties (spillover effect) on airline demand. Spillover attributes include mean of different attributes of the neighboring counties such as population, median income, unemployment rate, vehicle ownership level and new COVID-19 cases in the preceding month. A descriptive analysis summary of the independent variables is presented in Table 3.1.

Table 3.1 Descriptive Analysis of the Independent Variables

Categorical Variables			
Variables	**Description**	**Frequency**	**Percentage**
Built Environment Characteristics			
State level tourism			
Top 10	The state is among top 10 tourists' attraction state	109	28.7
Bottom 10	The state is among bottom 10 tourists' attraction state	39	10.3
Others	The state is not among top 10 or bottom 10 states	232	61.1
Airport Specific Factors			
Operational Evolution Partnership (OEP) airports			
Yes		35	9.2
No		345	90.8
Spatial Factors			
Region			
South	The airport is located in South region	122	32.1
North-East	The airport is located in North-East region	45	11.8
West	The airport is located in West region	91	23.9
Mid-West	The airport is located in Mid-West region	84	22.1
Pacific	The airport is located in Pacific region	38	10.0
Temporal Factors			
Month			
June 2019		380	4.2
July 2019		380	4.2
November 2019		380	4.2
December 2019		380	4.2
Other months		7600	83.3
COVID-19 Related Factors			
Pandemic started			
Yes	Month is March 2020 or later	3800	41.7
No	Month is before March 2020	5320	58.3
May or later			
Yes	Month is May 2020 or later	3040	33.3
No	Month is before May 2020	6080	66.7
July or later			
Yes	Month is July 2020 or later	6840	75.0
No	Month is before July 2020	6912	25.0

Categorical Variables			
Variables	**Description**	**Frequency**	**Percentage**
October or later			
Yes	Month is October 2020 or later	1140	12.5
No	Month is before October 2020	7980	87.5
Continuous Variables			
Variables	**Description**	**Mean**	**Min/Max**
County Level Demographic Characteristics			
Population	Population in million	0.518	0.000/10.160
Median income	Ln(Median income in thousands)	10.944	10.350/11.820
Unemployment	County level unemployment rate	4.346	2.000/19.900
Senior population	% of population having age 65 and over	15.658	5.877/39.444
Vehicle 0	% of HH with 0 vehicle	8.982	1.700/87.800
Vehicle 1	% of HH with 1 vehicle	33.329	10.000/47.800
Vehicle 2	% of HH with 2 vehicles	37.034	2.100/48.200
Vehicle 3+	% of HH with 3 or more vehicles	20.658	0.100/38.100
Built Environment Characteristics			
Ln(Airport)	Ln(No. of airports in 50-mile buffer area)	1.842	0.000/3.740
COVID-19 Related Factors			
Ln(COVID-19 cases)	Ln(County level new COVID-19 cases in the past month)	2.138	0.000/11.670
Adjoining County Attributes (Spillover Effects)			
Population	Average population in neighboring counties in million	0.207	0.000/4.520
Median Income	Ln(average median income in neighboring counties in thousand)	3.935	0.000/4.690
Unemployment	Unemployment rate	4.612	0.000/16.470
Vehicle 0	% of HH with 0 vehicle	8.102	0.000/68.400
Vehicle 1	% of HH with 1 vehicle	29.723	0.000/57.400
Vehicle 2	% of HH with 2 vehicles	35.989	0.000/44.450
Vehicle 3+	% of HH with 3 or more vehicles	24.084	0.000/44.700
Ln(COVID-19 cases)	Ln(average new COVID-19 cases in the past month in neighboring counties)	1.801	0.000/10.680

3.5 Analysis and Results

In our study, we analyzed airport level monthly air passenger departures using a linear mixed model. A host of independent variables were considered in the model development process. As the main focus of our study is on understanding the impact of COVID-19, variables related to COVID-19 and various interactions were tested in the model specification. However, we also included different factors that have been identified as important determinants of airline demand. In summary, the model estimation process was guided by earlier research, variable interpretability and parameter statistical significance.

The final model results are presented in Table 3.2. The positive (negative) value of the parameter estimates indicates increase of a parameter increases (decreases) the airline demand. The results are discussed in detail in the following subsections by the attribute levels.

3.5.1 County Level Demographic Characteristics

Demographic characteristics are expected to serve as controls for airline demand. As expected, counties with larger population are likely to have higher airline demand as population serves as a surrogate measure for demand (please see Grosche et al., 2007; Zhou et al., 2018 for similar results). On the other hand, a higher percentage of senior population is found to be negatively associated with the air passenger demand. The parameter for county level unemployment rate highlights the negative association of unemployment rate with airline demand. The result is plausible as increased unemployment rate, in general, corresponds to decreased affordability for personal travel and fewer business activity in the region.

3.5.2 Built Environment Characteristics

The variable "number of airports in a 50-mile buffer" represents the number of available airports in close proximity (50-mile radius) of an airport. We found that an increased number of airports in the 50-mile buffer results in higher air travel demand at an airport. The presence of additional airport(s) in close proximity reflects higher demand in the region. Further, we considered the tourism status of the state in our analysis by identifying the top and bottom 10 desirable states with respect to tourism activity. As expected, we find that air travel demand is higher (marginally significant) in an airport located in top 10 tourist attraction states while a reduced air demand is observed for an airport located in the bottom 10 visiting states (see Sivrikaya & Tunç, 2013 for similar results). The reader would note that tourism ranking in our analysis is considered at a state resolution. Ideally, county level tourism measures such as expenditures or hotel beds would be preferred variables. However, access to such data across the country is not readily available and is a direction for future research.

3.5.3 Airport Specific Factors

In this study, we consider the type of airport as an airport specific factor. We classified the airports as Operational Evolution Partnership (OEP) airports and other airports. OEP airports capture approximately 70% of the total domestic airline demand in US and are identified as large airports in the analysis. The positive coefficient of binary OEP airport variable indicates that air passenger departure rate is higher in OEP airports compared to other airports. The result reflects the higher demand in OEP airports in the US.

3.5.4 Spatial Factors

Location of the airports across various US regions has a significant effect on the total number departures from those airports. In general, compared to the airports in the other regions, the demand is observed to be higher for an airport in the South region. An examination of the airports in the South region reveals that some of the busiest airports in the US (3 of the top 10 busiest airports (Travel, 2021)) are from this region. Further, we also find that airports in the South are located further away from one another relative to airports in the North-East and West. It is possible that these airports have much larger catchment areas, and the South indicator variable possibly serves as a surrogate for the larger catchment size.

3.5.5 Temporal Factors

Monthly and Quarterly indicator variables were tested in the model to allow for seasonal effects. In our model estimation, the results indicate that travel demand was higher in the months of June, July and December 2019 and lower in November 2019 compared to other months while controlling for other factors. These results can be attributed to presence of seasonality in air travel demand.

3.5.6 COVID-19 Related Factors

COVID-19 related factors considered in this study include both global and local effects of COVID-19 on airline demand. Global factors were considered in the model in various functional forms including continuous (such as linear, square and other polynomial) and indicator variables (such as month indicator for pandemic, pandemic from May or later and Pandemic from July or later). Local COVID-19 factors considered include the natural logarithm of county level total new COVID-19 cases in the preceding month. The reader would note that the net effect of COVID-19 is a sum of the global effect and the local case specific effect.

As expected, the pandemic variable (set to 1 for all months from March 2020) has a negative coefficient indicating that airline demand dropped significantly after the pandemic started. The positive coefficient of May or later variable indicates that airline demand recovered after May (while controlling for other variables). The positive coefficient of July or later and October or later variables indicate that airline demand increased further since these time periods. However, airline demand was negatively influenced by local COVID-19 data in the airport county for these months. The result indicates that the air travel is likely to reduce in the presence of increasing COVID-19 cases in the preceding month. It should be noted that while some recovery has happened as reflected in May or later, July or later and October or later indicator variables, the net change in airline demand relative to the corresponding month in 2019 has been negative across the country.

In addition to the main effects described, we also tested for several interaction effects of COVID-19 variables with other factors affecting airline demand. The positive coefficient of the interaction of pandemic variable and population indicates that the initial drop of demand in March and April of 2020 due to COVID-19 was lower in the airports located in a county with higher population. The interaction analysis also found that the larger airports (OEP airports) exhibit slightly different trends. Specifically, we found that the initial drop of demand in March and April of 2020 is slightly lower in OEP airports. The coefficient for May or later at OEP airports further highlights higher recovery in these airports. A negative coefficient for July or later variable indicates a reduced differential with other airports from July. Finally, interaction of south region and pandemic started variable is found significant. The positive coefficient of the interaction term indicates that the initial drop in airline demand in the airports in the south region is lower compared to the airports in other regions. The finding might be attributed to reduced adherence to public health guidelines in many states from this region.

3.5.7 Adjoining County Attributes (Spillover Effects)

The parameter estimates indicate that airline demand is also influenced by the attributes of adjoining counties. We found that mean population, median income, and new COVID-19 cases in the neighboring counties influence airline demand at the airport level in an intuitive manner. The effects of population and median income indicate that increased population and median income in the neighboring counties increase airline demand. The effect of neighboring county COVID cases indicates that increased new COVID cases in the adjoining counties significantly decreases airline demand.

3.5.8 Covariance parameters

The last row panel of Table 3.2 present the results for the covariance parameters (σ^2, ρ and ϕ). As expected, these parameters are significant and highlight the presence of common unobserved factors affecting the repeated airline demand data for each airport.

Table 3.2 Parameter Estimates for Liner Mixed Model

Parameter	**Estimates**	**Std. Error**	**t stat**
Fixed Effects			
Intercept	9.293	0.696	13.354
County Level Demographic Characteristics			
Population in million	0.379	0.088	4.304
Senior population	-0.070	0.019	-3.696
Unemployment rate	-0.304	0.036	-8.411
Built Environment Characteristics			
Ln(No. of airports in 50 mile buffer)	0.448	0.119	3.750
State level tourism (Base: Others)			
Top10	0.353	0.193	1.829
Bottom10	-0.593	0.266	-2.232
Airport Specific Factors			
OEP airports (Base: No)			
Yes	3.082	0.310	9.930
Spatial Factors			
Region (Base: Other regions)			
South	0.552	0.184	2.998
Temporal Factors			
Month (Base: Other months)			
June 2019	0.053	0.027	2.008
July 2019	0.114	0.027	4.293
November 2019	-0.105	0.027	-3.972
December 2019	0.059	0.027	2.241
COVID-19 Related Factors			
Pandemic started (Base: No)			

Parameter	Estimates	Std. Error	t stat
Fixed Effects			
Yes	-0.957	0.039	-24.352
May or later (Base: No)			
Yes	1.621	0.035	46.143
July or later (Base: No)			
Yes	0.958	0.032	29.886
October or later (Base: No)			
Yes	0.183	0.030	6.105
Ln(County Level Covid-19 Cases in the last month)	-0.304	0.012	-25.356
Population × Pandemic started	0.059	0.028	2.070
OEP airports × Pandemic started	0.181	0.113	1.601
OEP airports × May or later	0.171	0.104	1.641
OEP airports × July or later	-0.292	0.104	-2.817
South × Pandemic started	0.213	0.064	3.322
Adjoining County attributes (spillover effects)			
Average population (million)	0.572	0.240	2.379
Ln(average median income in thousand)	0.299	0.131	2.277
Ln(average COVID-19 cases in past month)	-0.107	0.015	-7.350
Covariance Parameters			
$\boldsymbol{\sigma}^2$	3.201	0.175	18.252
ρ	0.965	0.003	349.007
ϕ	0.940	0.003	272.200

3.6 Model Performance

The performance of the linear mixed model was compared with the performance of the traditional linear regression model using log-likelihood and Bayesian Information Criterion (BIC)[4]. The log-likelihood (BIC) values for the models are as follows: linear regression model: -17247.95

[4] The reader would note that due to the inherent structure of linear mixed models, traditional goodness of fit measures such as R^2 are not readily applicable and require more involved approaches to computing the measure (see Nakagawa & Schielzeth, 2013 for more details).

(34505.01) and linear mixed model: -8720.29 (17467.92). From the comparison, it is evident that the linear mixed model offers improved fit in our data.

We also evaluate the performance of the proposed model in predicting the demand. Specifically, we compare the total observed demand and predicted demand in US (see Figure 3.4). An examination of Figure 3.4 plot illustrates that the proposed model represents the demand trends before and after the pandemic. The model successfully captures the demand drops after the start of the pandemic and the slow continuing recovery after the initial months. The reader would note that the airline demand data is available only till December 2020. In our prediction exercise, we also generate demand for January 2021 through September 2021 by employing COVID-19 data available up to August 21, 2021. For COVID-19 cases in the full month of August (as required to predict demand for September), we assumed same infection rate in the remaining days of August as it was in the first 21 days of the month. The figure shows that the demand may decrease in the future months, especially in September, due to recent increase of COVID-19 cases.

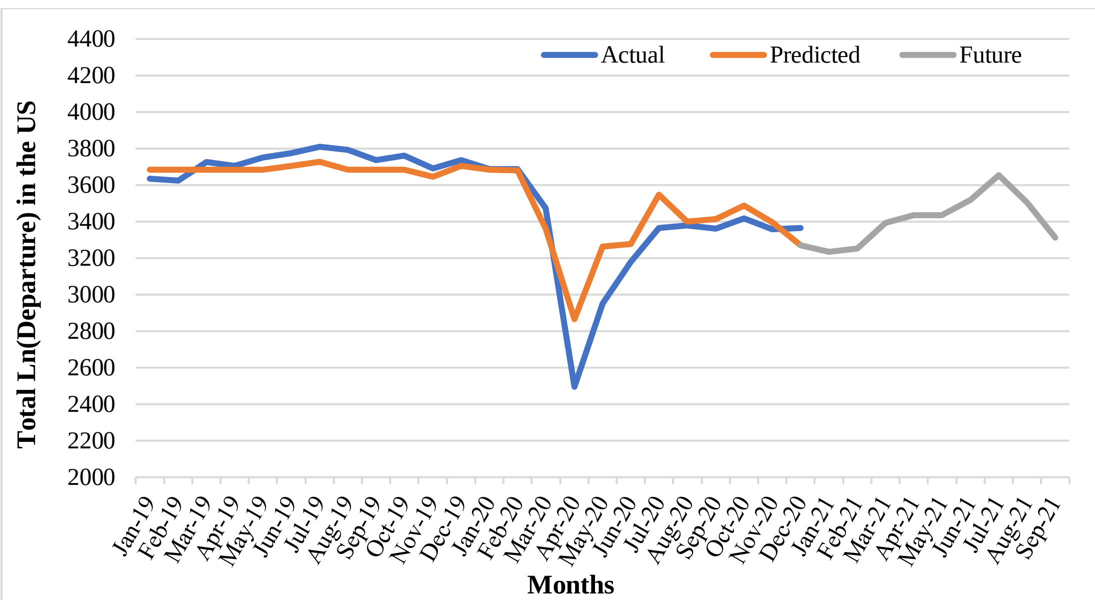


Figure 3.4 Predictive Performance of the Proposed Model

3.7 Policy Analysis

As discussed in the study objectives, the model development exercise was motivated by the need to present a blueprint for airline demand recovery. To illustrate the model applicability for generating monthly airline demand estimates, we consider three hypothetical scenarios of COVID-19 transmission rates – expected, pessimistic and optimistic. In these scenarios, a rate of increase or decrease for COVID-19 is considered. While the rate considered is uniform across the country, the actual change in cases will depend on current transmissions in the counties. Thus, we will be accommodating for a spatially varying COVID-19 case load in the country. The exact assumptions for the scenarios are described below:

1) Expected Scenario: The scenario is based on the expected increase in vaccinations of the approved vaccines such as Pfizer BionTech and Moderna across US. Hence, in this scenario COVID-19 cases are likely to increase marginally in September 2021 (August is at a high and plateauing). As increased proportions of the population are vaccinated, we expect the transmissions to drop in the subsequent months as follows: October 2021 (15%), November 2021 (20%), December 2021 (20%), and January 2022 (20%). The reader may see IHME, 2021 to find similar expected scenario of COVID-19 transmission rate.
2) Pessimistic Scenario: The proportion of vaccinated population does not increase significantly, and infection rate keeps increasing following the trend in recent months. We assume the infection rate will increase by 20% in September 2021 and October 2021 followed by 10% increases in each of the months from November 2021 through January 2022.
3) Optimistic Scenario: The scenario assumes a better-than-expected impact of vaccination due to rapidly increased vaccination rate and possible emergence of booster doses. In this

scenario, we assume that the infection rate will increase by 5% in September and booster doses will be available by end of September causing 25% decrease in COVID cases in October 2021. Finally, cases will decrease by 35%, 50% and 50% in the following months.

Table 3.3 Percentage Changes in New COVID-19 Cases Compared to the Preceding Month

Month	Expected	Pessimistic	Optimistic
Sep-21	10%	20%	5%
Oct-21	-15%	20%	-25%
Nov-21	-20%	10%	-35%
Dec-21	-20%	10%	-50%
Jan-22	-20%	10%	-50%

The rate of change of COVID-19 cases for different scenarios by month are summarized in Table 3.3. Based on these assumptions, the airline demand is predicted using the proposed linear mixed model and the demand is aggregated to identify the total airline demand for the months of interest. Then, we perform a month-by-month comparison of airline demand for months September 2021 through February 2022 with the corresponding months in 2019. To facilitate the understanding of recovery process across the country, the numbers are aggregated by US region. The results of the analysis are then presented in Figure 3.5. Figure 3.5 shows the future percentage changes in airline demand in the US by month across different regions. From Figure 3.5, the following observations can be made for the three scenarios considered:

1) Expected Scenario: In this scenario, airline demand will decrease in September 2021 and October 2021, followed by a small increase in November 2021. Airline demand will further decrease in December 2021. Finally, the demand will start recovering starting from January 2022. If we compare the demand changes for the four regions, we can see that south region

experienced the lowest drop in demand and may also recover faster than the other regions in the US.

2) Pessimistic Scenario: In this scenario, airline demand will continue decreasing and drop by 87% by February 2022. Airports across all regions may experience similar decrease in demand but airports in West region may have the highest decrease estimated at 90%.

3) Optimistic Scenario: In this scenario, airline demand may keep decreasing till October 2021. But in response to the improved COVID-19 condition, the recovery will start from November 2021 and accelerate in the latter months. By February 2022, airline demand in the US may reach 38% of the typical demand. If we compare the recovery rate for regions, we can see that South region will recover faster and the demand may be 50% of original demand by February 2022.

To offer further insights on the predictions generated, we aggregate the demand at the State level based on all airports in the state and present the estimates for all scenarios (see Figure 3.6). Figure 3.6 shows percentage change of airline demand from September 2021 through February 2022 compared to the usual demand from 2019 at the state level. The results follow expected scenario specific trends. We recognize that the policy assumptions are unlikely to be matched exactly. The objective of this exercise is to illustrate the insights that can be generated from the model. These plots generated can be customized with more up to date information on COVID-19 cases to arrive at accurate expected demand.

Moreover, we compare our demand prediction for the future (2021) based on expected scenario with terminal area forecasts provided by FAA. Recently available demand data for year

2021 is considered as a benchmark for the comparison[5]. We consider 2021 air passenger departure prediction provided by 2020 Terminal Area Forecast (TAF) model for the selected airports. In demand prediction, TAF model first predicts O-D air passenger flows using a regression model. Then, segment pair demand is calculated based on O-D pair predictions and T-100 data. Finally, yearly airport level demand is forecasted by aggregating segment pair demands (see FAA, 2020 for the detailed process). From the comparison, it is evident that both proposed model and TAF model underpredict the demand for 2021. But TAF model shows superior prediction accuracy to the proposed model. Such result is expected as high dimensions data (O-D pair demand data) is employed by TAF model. Such O-D pair-based demand modeling approach can be an avenue for future research.

[5] As of February 14, 2022, T-100 domestic marketing carrier data is available up to November 2021. Demands at the airports for full year of 2021 are estimated by assuming demand in December 2021 to be same as average monthly demand observed for other months in 2021.

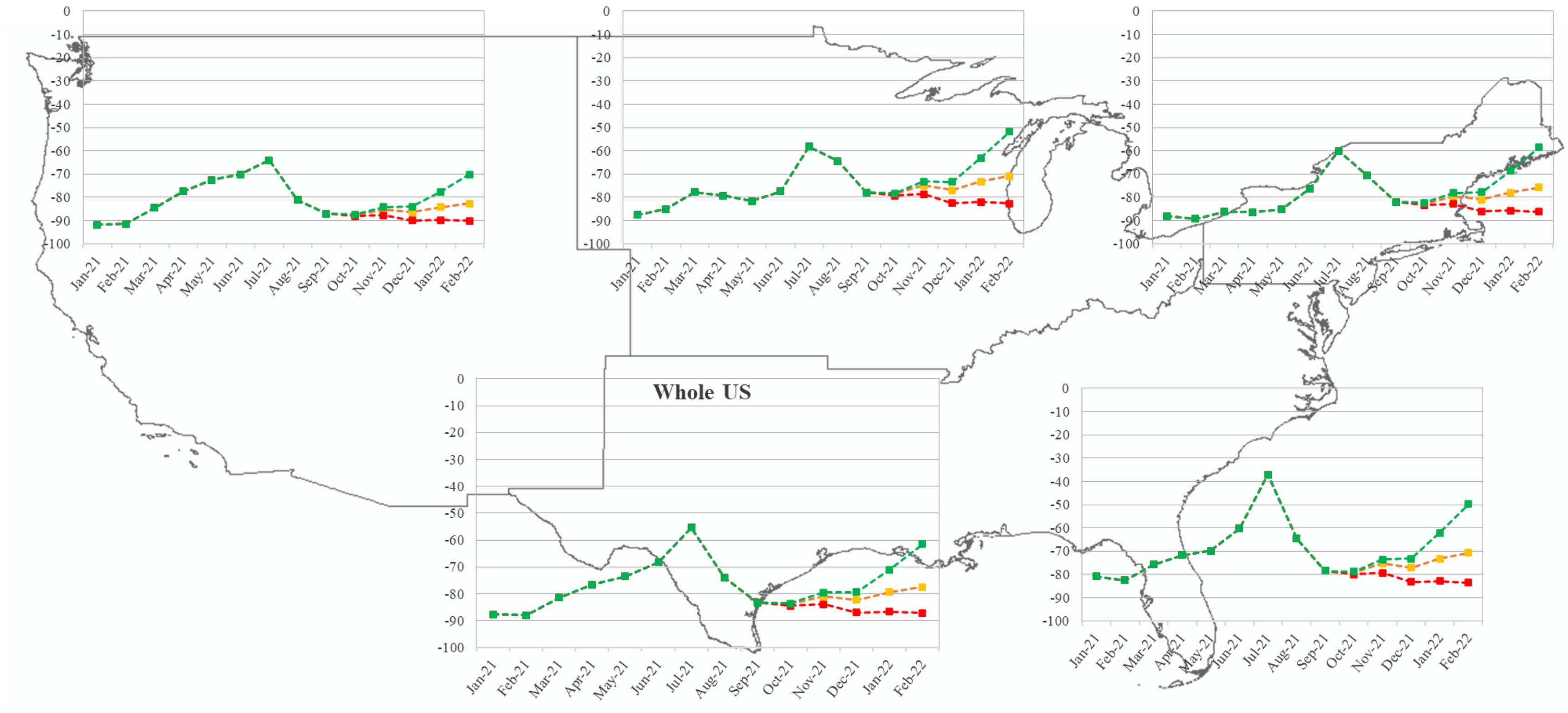
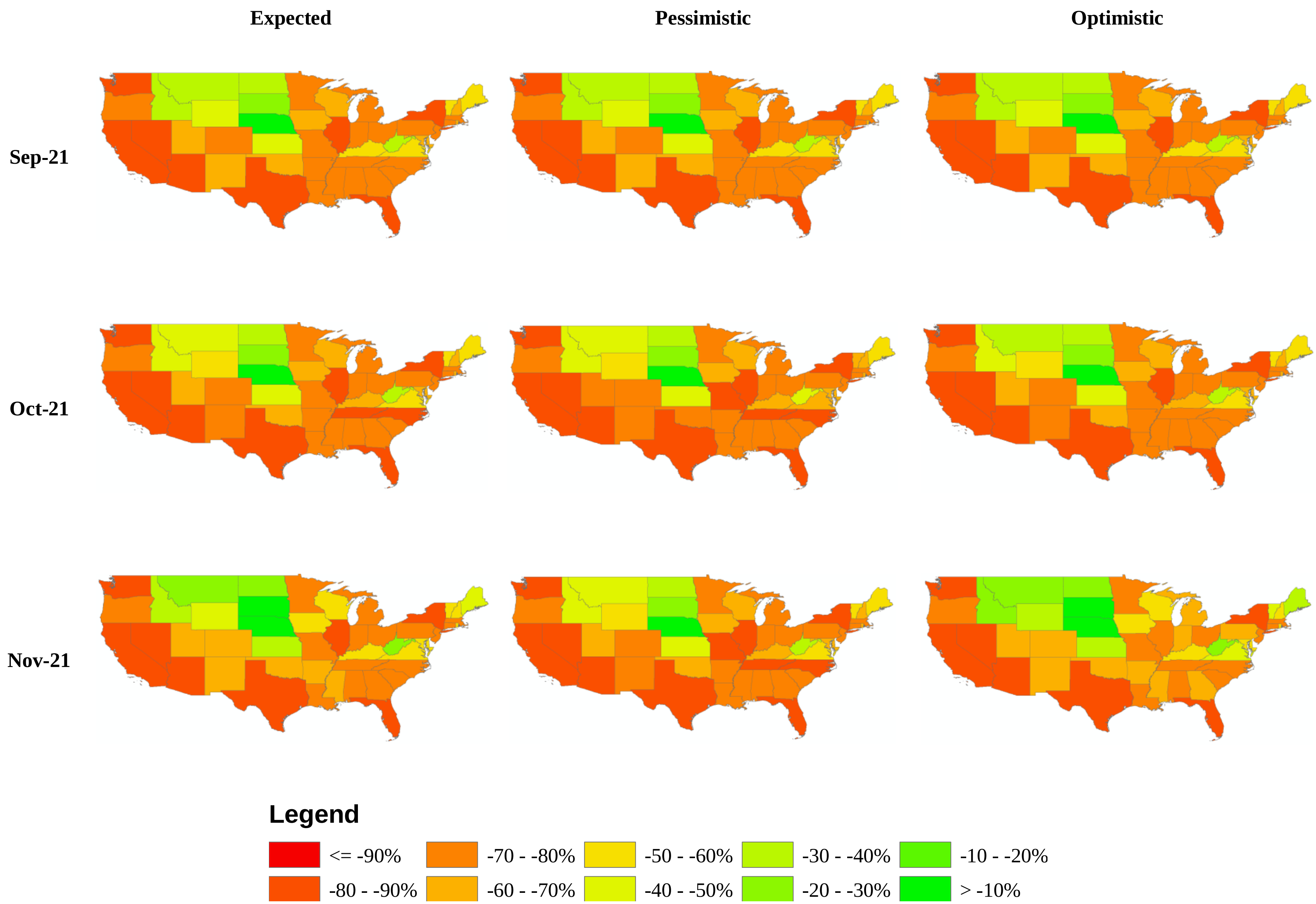


Figure 3.5 Future Demand Based on Hypothetical Scenarios



Expected
Pessimistic
Optimistic
Sep-21
Oct-21
Nov-21
Legend
<= -90%
-80 - -90%
-70 - -80%
-60 - -70%
-50 - -60%
-40 - -50%
-30 - -40%
-20 - -30%
-10 - -20%
> -10%

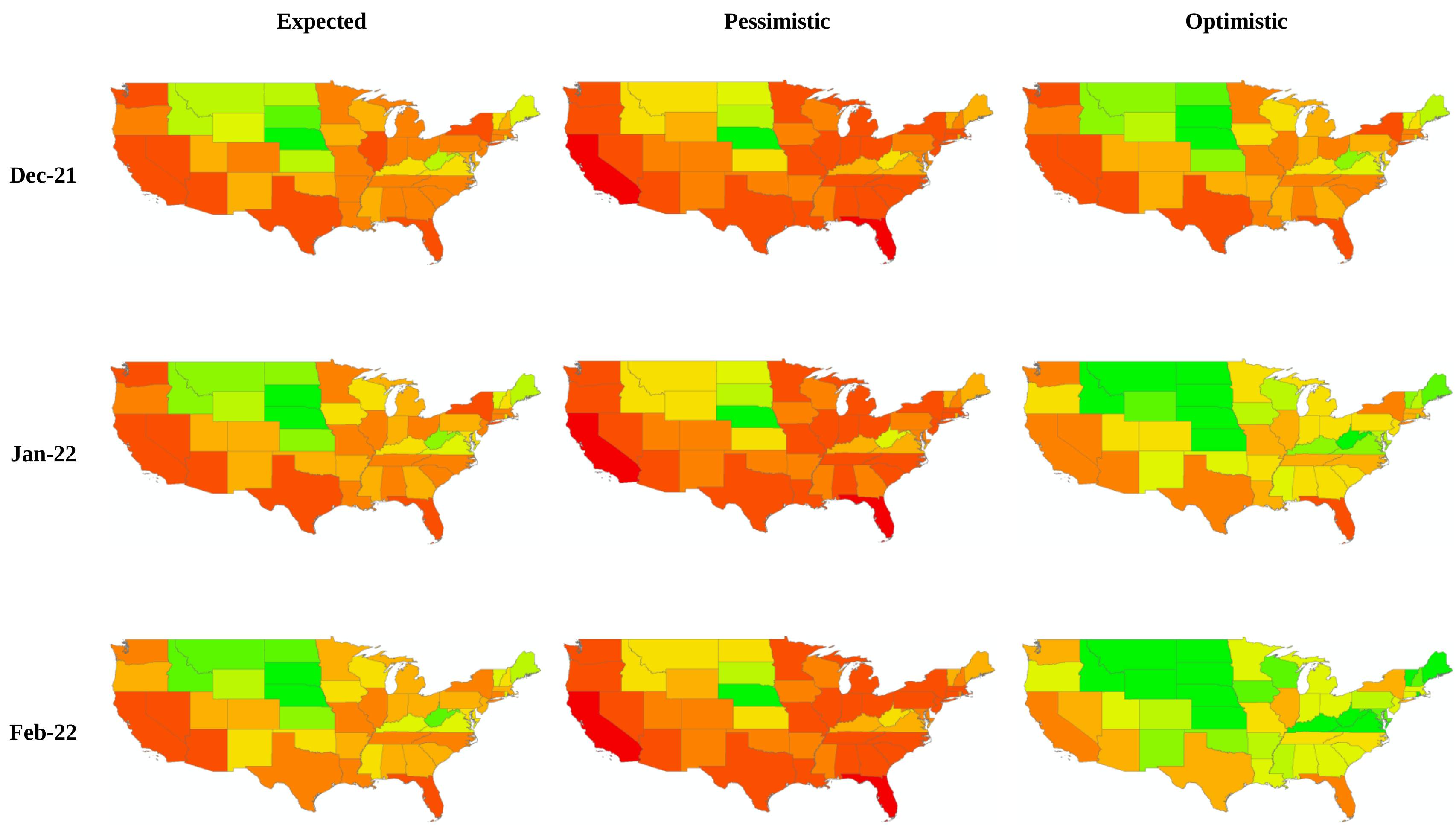


Figure 3.6 Future Airline Demand at the State Level

3.8 Summary

The COVID-19 pandemic has affected every facet of life in the world significantly burdening social, health and economic systems. Among these affected industries, airline industry ranks as one of the worst affected industries. The emergency use authorization of vaccines offers promise in curbing the pandemic and supporting the recovery. As the recovery begins airlines and airports would need to address supply side shortages with growing demand. In this context, the primary focus of our proposed research effort is to develop a framework that provides a blueprint for airline demand recovery at a high resolution as COVID-19 cases evolve over time. In our study, we conduct our analysis considering 380 airports across the country. Airline data employed in this study is sourced from Bureau of Transportation Statistics (BTS) for 24 months from January 2019 through December 2020 which is augmented with a host independent variables including COVID-19 related factors, demographic characteristics and built environment characteristics at the county level, airport specific factors, spatial factors, temporal factors, and adjoining county attributes. COVID-19 related factors include both local and global factors by considering global and local COVID-19 transmission, temporal indicators of pandemic start and progress, and interactions of airline demand predictors with global and local COVID-19 indicators. We employ a linear mixed model system that accommodates for the presence of repeated measures for modelling airline demand.

The linear mixed model identifies several important determinants of airline demand while also capturing the impact of global and local COVID-19 effects on demand. The performance of the model is examined by comparing observed and predicted demand for all airports across the US. The result indicates that model successfully captures the demand drops after the start of the pandemic and the slow continuing recovery after the initial months. Subsequently, we present a blueprint for airline demand by considering three hypothetical scenarios of COVID-19 transmission rates – expected, pessimistic and optimistic. The results at the airport level from these scenarios are aggregated at the state or regional level by adding the demand from all airports in the corresponding state or region. These trends are presented by State and Region to illustrate potential differences across various scenarios. The result from the expected scenario presents a path to slow recovery as COVID-19 cases reduce. The various scenarios clearly illustrate how the proposed model can be employed to generate airline demand estimates at the airport level, state, region or country level.

The study is not without limitations. In our analysis, data was generated at the airport county level. Thus, when the same county has multiple airports, the model includes substantially similar information for these airports (except OEP 35 indicator and number of airports in a 50-mile buffer). While only 22 of the 354 counties in our data had multiple airports, it might be interesting to explore how aggregation of the demand for these airports affects the findings. Moreover, the airline demand data is available only till December 2020 which restricted us from employing linear and non-linear functions of continuous temporal variables. Given the data availability for the next few months, continuous temporal variables could be considered to enhance the current model. Further, COVID-19 pandemic is an evolving situation, and it is appropriate to consider updating the models with newer airline demand (as they become available), local vaccination data and local COVID-19 cases. Finally, the airport level analysis conducted in the paper can be augmented by examining airport level actions/strategies (such as changes to fare, priority for freight movement) in response to COVID-19 pandemic. The research might have to be conducted for a subset of airports where such data is available.

CHAPTER 4: ACCOMMODATING SPATIAL DEPENDENCY IN AIRLINE DEMAND MODELING

Air passenger demand is important to understand overall health of airline industry. While earlier research efforts identified the factors affecting airline demand, spatial interactions between air passenger demand at multiple airports have not been adequately considered. Thus, the current study develops novel spatial group generalized ordered probit models of monthly air passenger departures at the airport level that accommodate spatial dependency between proximally located airports. Specifically, we employ spatial error and spatial lag models of airport level air passenger departures in this study. Further, we compare the predictive performance of alternative models using a validation exercise.

4.1 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 (please see Chapter 2 for the detailed review of the relevant studies). 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 (Rahman et al., 2021; Faghih-Imani & Eluru, 2016), impact of transportation infrastructure on regional/agricultural growth (Chen & Haynes, 2015; Tong et al., 2013; Yu et al., 2013), land use modeling (Ferdous & Bhat, 2013; Chakir & Parent, 2009; Carrión-flores et al., 2009; Xiaokun & Kockelman, 2006), crash injury severity modeling (Castro et al., 2013), recreational activity modeling (Bhat et al., 2010), and airfare analysis (Daraban & Fournier, 2008). Dependent variables in such studies can be categorized as both continuous variable (Rahman et al., 2021; Faghih-Imani & Eluru, 2016; Chen & Haynes, 2015; Tong et al., 2013; Yu et al., 2013; Daraban & Fournier, 2008), and categorical variables (Castro et al., 2013; Ferdous & Bhat, 2013; Bhat et al., 2010; Chakir & Parent, 2009; Carrión-flores et al., 2009; Xiaokun & Kockelman, 2006). The aforementioned studies employ different variants of spatial models to capture spatial correlations including spatial lag or spatial autoregressive model (SAR) (Rahman et al., 2021; Faghih-Imani & Eluru, 2016; Chen & Haynes, 2015; Ferdous & Bhat, 2013; Castro et al., 2013; Lee & Yu, 2010; Chakir & Parent, 2009; Carrión-flores et al., 2009; Daraban & Fournier, 2008; Xiaokun & Kockelman, 2006), spatial intermediate model (Castro et al., 2013), spatial error model (SEM) (Rahman et al., 2021; Faghih-Imani & Eluru, 2016; Chen & Haynes, 2015; Castro et al., 2013; Bhat et al., 2010), and spatial Dublin model (SDM) (Chen & Haynes, 2015; Tong et al., 2013; Yu et al., 2013). Most of the aforementioned modeling approaches require spatial weight matrix representing the spatial arrangements of the analysis units to understand spatial correlation among themselves. 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 (Rahman et al., 2021; Faghih-Imani & Eluru, 2016; Yu et al., 2013), inverse of distance squared (Ferdous & Bhat, 2013; Daraban & Fournier, 2008), inverse of distance cubed (Castro et al., 2013), and inverse of exponential of distance (Ferdous & 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, Xiaokun & 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 Castro et al., 2013; Ferdous & Bhat, 2013; Bhat et al., 2010) and Markov chain Monte Carlo (MCMC) (see Chakir & Parent, 2009) method.

4.2 Contributions of the Current Study

While earlier studies in airline literature examined the impact of key factors on airline demand, 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 group 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. 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., 2022a, 2022b). 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 Xiaokun & Kockelman, 2006 for 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.

In this study, airline demand data is sourced from T-100 marketing 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).

4.3 Econometric Methodology

In this section, we first present the details of group 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.

4.3.1 Group Generalized Ordered Probit Model

Let k (k= 1, 2,…, K) be an index to represent airports, t (t = 1, 2, 3,…, $T = 5$) represent the different years, m(l=1, 2, 3,…., $M = 12$) represent 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 presentation, we express each observational unit as an unique combination of airport k, year t, and month m, using q (q = 1, 2, 3,…., Q=k*t*m). Then, the equation system for modeling demand may be written as follows:

$$y_q^* = \alpha' x_q + \varepsilon_q, y_q = j \;\; if \; \psi_{j-1} < y_q^* \leq \psi_j \tag{4.1}$$

In Equation 4.1, 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[6]. Finally, to allow for alternative specific effects, we also introduce threshold specific deviations in the model as $\rho_j = \tau'_j x_q$.

The probability for airport *k* to have departures in category *j* in year, *t* and month, *m* is given by:

$$P(y_q = j_q) = \Lambda\left[\frac{\psi_{q,j} - (\alpha' x_q + \rho'_{q,j})}{\lambda_q}\right] - \Lambda\left[\frac{\psi_{q,j-1} - (\alpha' x_q + \rho'_{q,j-1})}{\lambda_q}\right] \tag{4.2}$$

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

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

4.3.2 Spatial Lag GGOP Model

The spatial lag formulation includes spatial correlation in the latent propensity of airline demand presented in Equation 4.1 as follows (Castro et al., 2013):

$$y_q^* = \delta \sum_{q'=1}^{Q} w_{qq'} y_{q'}^* + \alpha' x_q + \varepsilon_q, y_q = j \ \ if \ \psi_{j-1} < y_q^* \leq \psi_j \tag{4.3}$$

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 2018. Therefore, absolute value of month difference between the observations is 29 and we add 1 to the absolute difference (=30) for computational advantage. 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 diagonal elements of **W** to be zero[7]. Then, we normalize each column of **W** matrix using row total 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 parameters of δ'. The latent demand propensity presented in Equation 4.3 can be re-written using vector notation as follows:

---

[7] We replace off-diagonal zero values with large values in the distance matrix to avoid strong correlations between same airports at different time points.

$$\boldsymbol{y}^{*} = \boldsymbol{\delta}\mathbf{W}\boldsymbol{y}^{*} + \mathbf{x}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{4.4}$$

Now, the Equation 4.4 can be re-written as follows (Castro et al., 2013):

$$\boldsymbol{y}^{*} = \boldsymbol{S}(\mathbf{x}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}) \tag{4.5}$$

where $\mathbf{S} = \left[\mathbf{I}_Q - \delta\mathbf{W}\right]^{-1}$ 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, $y^{*} \sim MVN_Q\left(\boldsymbol{B_{lag}}, \boldsymbol{\Sigma_{lag}}\right)$. We represent $\boldsymbol{B_{lag}}$ and $\boldsymbol{\Sigma_{lag}}$ as follows:

$$\boldsymbol{B_{lag}} = \mathbf{Sx}\boldsymbol{\alpha} \text{ and } \boldsymbol{\Sigma_{lag}} = \mathbf{SI_Q}\boldsymbol{S}' \tag{4.6}$$

4.3.3 Spatial Error GGOP Model

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

$$y_q^{*} = \delta \sum_{q'=1}^{Q} w'_{qq'} \varepsilon_{q'} + \alpha' x_q + \varepsilon_q, y_q = j \;\; if \; \psi_{j-1} < y_q^{*} \leq \psi_j \tag{4.7}$$

Now, vector representation of Equation 4.7 is as follows:

$$\boldsymbol{y}^{*} = \boldsymbol{\delta}\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{x}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{4.8}$$

We can re-write Equation 4.8 as follows:

$$\boldsymbol{y}^{*} = \mathbf{x}\boldsymbol{\alpha} + \boldsymbol{S\varepsilon} \tag{4.9}$$

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:

$$\boldsymbol{B_{error}} = \mathbf{x}\boldsymbol{\alpha} \text{ and } \boldsymbol{\Sigma_{error}} = \mathbf{SS}' \tag{4.10}$$

4.3.4 Model Estimation

The vector of parameters to be estimated in both spatial lag and spatial error GGOP model is θ = $(\alpha', \rho_j', \lambda, \delta')$. 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):

$$L_{CML}(\theta) = \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} \Pr\left(y_q = j_q, y_q' = j_q'\right)$$

$$= \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} [\Phi\left(\varphi_q, \varphi_{q'}, \nu_{qq'}\right) - \Phi\left(\varphi_q, \mu_{q'}, \nu_{qq'}\right) - \Phi\left(\mu_q, \varphi_{q'}, \nu_{qq'}\right) + \Phi\left(\mu_q, \mu_{q'}, \nu_{qq'}\right)] \tag{4.11}$$

Where, $\varphi_q = \frac{\psi_{q,j} - ([\boldsymbol{B}]_q + \rho'_{q,j})}{\sqrt{\lambda_q * [\boldsymbol{\Sigma}]_{qq}}}$, $\mu_q = \frac{\psi_{q,j-1} - ([\boldsymbol{B}]_q + \rho'_{q,j-1})}{\sqrt{\lambda_q * [\boldsymbol{\Sigma}]_{qq}}}$, $\nu_{qq'} = \frac{[\boldsymbol{\Sigma}]_{qq'}}{\sqrt{[\boldsymbol{\Sigma}]_{qq} * [\boldsymbol{\Sigma}]_{q'q'}}}$

In computing marginal likelihood function presented in Equation 4.11, we need to calculate $Q(Q-1)/2$ numbers of joint probabilities. In Equation 4.11, $\nu_{qq'}$ 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 considerable within a distance band and a time threshold. Therefore, we employed different values of distance band and time threshold and select the best combination. Based on model fit and significance of spatial autoregressive parameter, we select 100 miles distance band and 36 months threshold for our study. Therefore, we can re-write Equation 4.11 as follows:

$$L_{CML}(\theta) = \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} \Pr\left(y_q = j_q, y_q' = j_q'\right)$$

$$= \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} [\Phi\left(\varphi_q, \varphi_{q'}, R_{qq'} \nu_{qq'}\right) - \Phi\left(\varphi_q, \mu_{q'}, R_{qq'} \nu_{qq'}\right) - \Phi\left(\mu_q, \varphi_{q'}, R_{qq'} \nu_{qq'}\right)$$
$$+ \Phi\left(\mu_q, \mu_{q'}, R_{qq'} \nu_{qq'}\right)] \quad (4.12)$$

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

In above Equation 4.12, $R_{qq'}$ 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 (1960) sandwich information matrix (see Bhat et al., 2010 and Castro et al., 2013 for the details of covariance matrix).

4.4 Dataset Description

The airline demand data is sourced from T-100 Domestic Marketing Carrier dataset compiled by Bureau of Transportation Statistics. The marketing carrier dataset contains number of passengers carried by all 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 51 states in 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. 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 4.1. The figure shows that the dependent variable is approximately normally distributed.

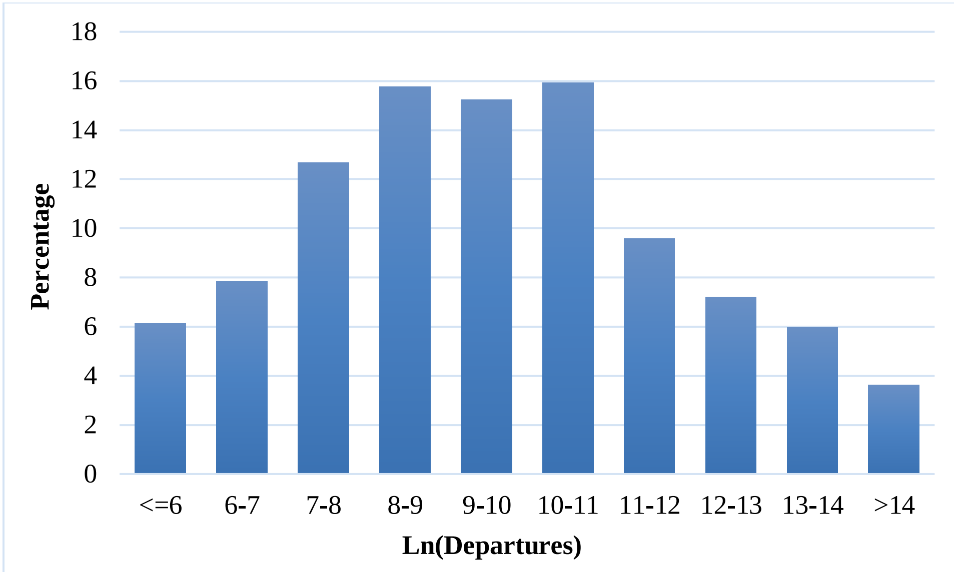


Figure 4.1 Distribution of The Dependent Variable

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 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 (Insider, 2020). 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 4.1. Table 4.1 includes mean, minimum and maximum values for continuous variables and frequency and percentage for categorical variables.

Table 4.1 Descriptive Statistics of the Independent Variables

Continuous Variables			
Variables	**Description**	**Mean**	**Min/Max**
Demographic Characteristics			
Population	Population in million in corresponding MSA	1.179	0.013/20.031
Median Income	Median income in 100K in corresponding MSA	0.544	0.312/1.147
Employment	Ln(number of workers in thousands in corresponding MSA)	0.464	0.290/0.607
Out of state employment	Fraction of job holders in corresponding MSA working out of state	0.029	0.000/0.269
Built Environment Characteristics			
No. of airports	Ln(Number of airports in 50 mile buffer area)	1.753	0.000/3.664
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

Categorical Variables			
Variables	**Description**	**Freq.**	**Percent**
November		156	8.455
December		169	9.160

4.5 Analysis and Results

In model development, we first estimate a simple group 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. Log-composite likelihood (LL) at convergence and Bayesian Information Criteria (BIC) values of GGOP models are -3510.43 and 7156.22, respectively. In the second step, we estimate a series of spatial lag and spatial error models considering various formulation of $w_{qq'}$ as discussed in section 4.3.2. Based on the data fit and significance of spatial autoregressive parameter, 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 LL and BIC values of the proposed spatial lag GGOP model are -3395.49 and 6933.86, respectively. The LL and BIC values of the proposed spatial error GGOP model are -3378.82 and 6900.52, respectively. Therefore, both spatial lag model and spatial error model offer improved data fit compared to simple GGOP model and spatial error GGOP model offers the best fit in terms of the BIC measure. For the sake of brevity, only the spatial error GGOP model results are presented in this dissertation.

4.5.1 Estimation Results

The proposed spatial error GGOP model is presented in Table 4.2. 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:

4.5.1.1 <u>Demographic Characteristics</u>

Estimation results indicate that airline demand is significantly influenced by MSA level demographics. From the result, it is evident that airport level passenger departure rate is positively associated with MSA level population. Thus, an increase in MSA population increases the propensity for higher monthly airline demand (see Tirtha et al., 2022c 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.

4.5.1.2 <u>Built Environment Factors</u>

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 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. In addition to 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.

4.5.1.3 Airport Specific Factors

In the 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 4.2, 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.

4.5.1.4 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 4.2, 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.

4.5.1.5 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.

4.5.1.6 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 4.2. The deviation parameter is similar to a constant in discrete choice models and does not have an interpretation after incorporating other variables.

4.5.1.7 Variance Components

In the proposed model, we estimate and parameterize error variance. Variance components are presented in third-row panel of Table 4.2. 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.

4.5.1.8 Spatial Correlation

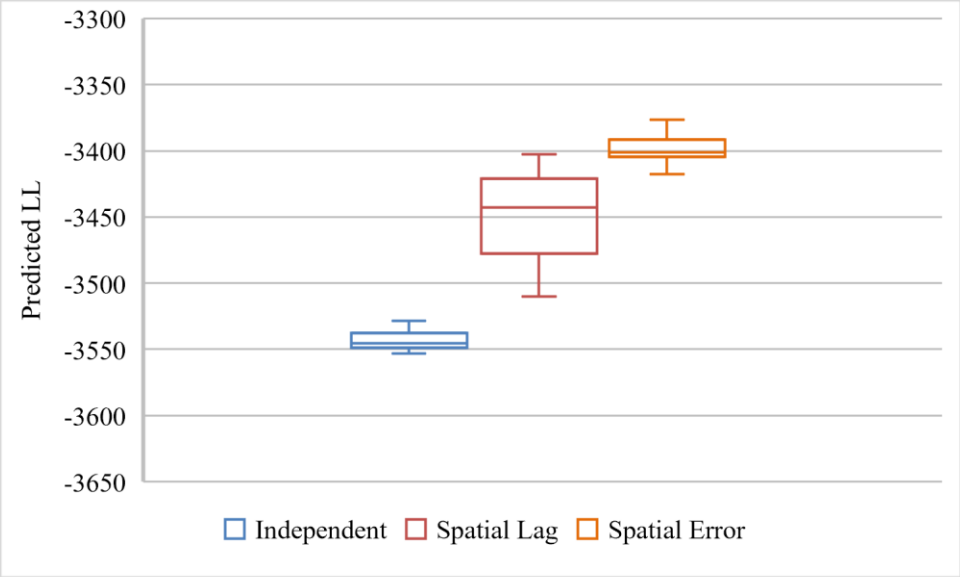
The main contribution of this chapter arises from consideration of spatial dependency in the airline demand modelling. From the analysis results, we found spatial autocorrelation parameter as strong in magnitude (the value is 3.943) and highly significant (t stat is 109.558). The significance of the 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 also conclude that such spatial correlation varies significantly over time.

Table 4.2 Estimation Results for Spatial Error GGOP Model

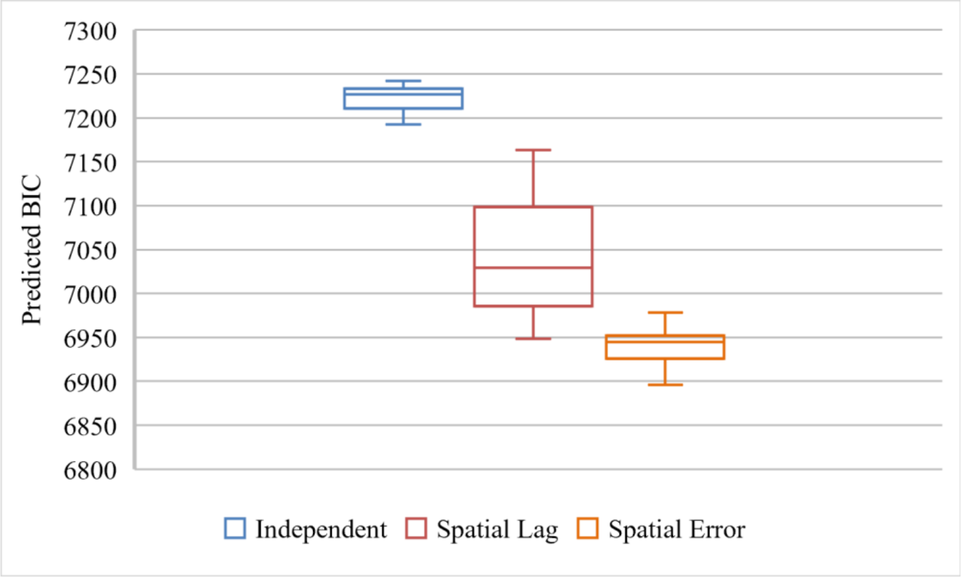
Variables	Estimates	t stat
Propensity Components		
Constant	4.102	12.826
Demographic Factors		
Population in million	0.130	11.729
Median income in 100K	1.932	5.610
Ln(Employment in thousands)	5.591	6.708
Built Environment Factors		
Ln(no. of airports in 50 miles)	0.894	17.431
Tourism Ranking (Base: other states)		
Top 10	0.509	6.932
Bottom 10	-0.508	-4.826
Airport Specific Factors		
Core Airports (Base: No)		
Yes	2.898	29.266
Spatial Factors		
Region (Base: West and Mid-West)		
South	0.615	8.562
North-East	-0.867	-9.757
Pacific	-1.840	-12.932
Temporal Factors		
Month (Base: other month of the year)		
January	-0.425	-4.027
July	0.398	4.274
Threshold Specific Effects		
Threshold 2	-0.403	-4.980
Threshold 3	-0.289	-5.194
Threshold 4	-0.172	-4.767
Variance Components		
Region (Base: other regions)		
South	-0.145	-5.336
North-East	0.136	3.884
Spatial Autoregressive Parameters		
Constant	3.943	109.558

4.6 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, spatial lag 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 Bayesian Information Criteria (BIC) measures for the three model systems. The results from 20 samples are compiled to generate the average and range of the model performance measures across the three 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: -3543.94 [-3549.61, -3538.27] and 7223.24 [7211.90, 7234.59], (2) spatial lag GGOP model: -3456.19 [-3479.38, -3432.99] and 7055.26 [7008.87, 7101.65], (3) spatial error GGOP model: -3398.88 [-3403.02, -3394.74] and 6940.64 [6932.36, 6948.92]. The results from the validation exercise confirm that both spatial models perform considerably better than the independent model that does not consider for spatial correlations between the observation units. Further, spatial error model is found to be superior to independent GGOP model and spatial lag GGOP model. 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 4.2 Comparison between Three Model Systems

4.7 Summary

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 group 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 the BIC measure. Importantly, both spatial models are found to be superior to the independent GGOP model that does not consider any spatial dependency between the observations. 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 autoregressive 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 lag GGOP and spatial error GGOP models compared to independent GGOP model. The result of validation exercise indicates the superiority of both spatial models relative to the independent model. Among the two spatial models, spatial error GGOP model offered improved data fit.

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. 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.

CHAPTER 5: A FLIGHT LEVEL ANALYSIS OF DEPARTURE DELAY AND ARRIVAL DELAY

Airline delay has become a recurrent event in the US airports causing both direct and indirect costs to the industry. In 2019, 21.03% of all flights operated in the US arrived late by 15 minutes or more. Understanding the factors influencing airline delay is important to improve airline on-time performance or mitigate the delays. In this dissertation, we develop a novel copula-based group generalized ordered logit (GGOL) model of departure and arrival delay at a disaggregate resolution of flight. Further, we compare predictive performance of the proposed model relative to independent models of flight departure and arrival on a holdout sample. Finally, we conduct a model application analysis to present the policy implications of the current research.

5.1 Earlier Studies

In airline literature, airline delay can be considered as a departure and/or an arrival delay. According to BTS, departure/arrival delay can be defined as the time difference between scheduled and actual gate departure/arrival time. Traditionally, earlier studies identified the factors affecting airline delays and developed prediction models. A summary of previous studies examining airline delay is provided in Table 5.1 with information on the delay measure of interest, spatial resolution of analysis, number of airports considered, study objectives, methodology employed, and independent variables considered. From Table 5.1, we can make several observations. *First*, earlier studies on airline delay study three types of delay measures: (a) departure delay, (b) arrival delay and (c) both departure and arrival delay. From the review, a majority of earlier research analyzed either departure or arrival delay. The studies, modeling both departure and arrival delays, modelled the two delay categories independently. *Second*, earlier research on airline delay is conducted at three resolutions: (a) flight, (b) airport and (c) national airspace system (NAS) level. In the first resolution, studies analyzed airline delay for individual flights while in the latter two resolutions, delay is analyzed at an aggregate level of airport or network as an average daily delay. The review also shows that earlier studies analyzed airline delay data mostly employing a limited set of airports[8]. *Third*, the factors considered in modeling airline delays vary across the studies and include traffic conditions (average queuing delay, average arrival delay, total operations), trip specific factors (carrier, route, distance), weather conditions (visibility, wind speed, thunderstorm, precipitation, snow depth), spatial factors (location of origin and destination airports), and temporal factors (season, weekday/weekend, time of the day). *Fourth*, several mathematical models were employed in literature to predict airline delays and they can be broadly classified as (a) discrete outcome and (b) continuous outcome models. In discrete outcome models, the dependent variable is characterized as a binary outcome (flight delayed or not based on the BTS threshold of 15 minutes) or a categorical variable (for example, Gui et al., 2020 categorized flight arrival delay in 4 groups). Among discrete outcome models, binary/multinomial logit models are generally employed to determine the factors affecting airline delay. Among continuous outcome models, where delay is measured in minutes, commonly employed models include: (a) linear regression model, (b) time series analysis, (c) machine learning approaches, (d) survival model,

[8] 35 Operational Evolution Partnership Airports (OEP-35) are the largest set of airports considered by the airport level studies (Hao et al., 2014; Nayak & Zhang, 2011). However, flight level studies considered flights operated in most of the major airports across the US.

(e) piecewise regression model, and (f) optimization methods. *Finally*, discrete outcome models are more commonly employed in flight level analysis while continuous outcome models are employed in both disaggregate and aggregate level analysis.

5.2 Contributions of the Current Study

In this study, our goal is to model departure and arrival delays in a joint framework at the disaggregate resolution of flights.

A major contribution of this study to literature arises from data enhancement for flight delay analysis. The variables processed from 2019 BTS marketing carrier on time performance data are augmented with a comprehensive set of independent variables sourced from secondary data sources including Automated Surface Observing System (ASOS) dataset (sourced from Iowa Environment Mesonet) and FAA's Aviation System Performance Metrics (ASPM). We prepare weather variables – wind speed, hourly precipitation, thunderstorm proportion and visibility - from ASOS dataset. The data compilation is achieved by charting the potential airline flight route to identify weather conditions near the flight's origin airport, along the route, and at the destination airport. Towards processing this weather data, we divide the continental US into a latitude longitude grid of 5 degrees and compile hourly weather data from all weather stations within each grid while estimating the flight path and its intersection with the grid system (more details in Data Section). The detailed process allows us to generate weather conditions for the entire duration of the flight. Subsequently, we employ ASPM data to determine air traffic conditions at the origin and destination airports in the hours preceding the flight's departure and arrival, respectively.

Table 5.1 Summary of Literature Review

Study	Dependent Variable	Spatial Resolution	No. of Airports	Objective	Method	Independent variables
Hao et al., 2014	Average daily arrival delay (continuous)	Airport level	New York airports and OEP 32 airports	Estimating impact of NY airports' delay on other airports	2SLS regression model	Air traffic condition such as total operations and average queuing delay, weather factors including portion of thunderstorms in different regions in the US
Nayak & Zhang, 2011	Average daily arrival delay (continuous)	Airport level	OEP 34 airports and other airports in NAS	Estimating impact of single airport delay on NAS	Multivariate simultaneous regression model	Air traffic condition such as queuing delay, observed arrival delay at other airports and NAS, weather factors (thunderstorms and IMC condition), temporal factors including seasonal and year
Schaefer & Millner, 2001	Average arrival and departure delay per flight (continuous)	Airport level	3 sample airports	Modeling propagation of delay	Air traffic simulation	Weather factors (IMC duration)
Klein et al., 2010	Average daily arrival delay (continuous)	Airport level	Major airports in US	Estimating airport delay using weather data	Regression model	NAS and airport weather conditions including wind speed, snow depth, IMC condition, queuing delay
Markovic et al., 2008	Average daily punctual flights (continuous)	Airport level	1 airport in Germany	Identifying weather impact on arrival delays	Hybrid regression/time series modelling	Weather factors such as wind speed, snow depth, the traffic flow, and the airport system state (strikes, air traffic control failures, roadworks, or safety related shutoffs)
Abdel-Aty et al., 2007	Average daily arrival delay and flight arrival delay (continuous)	Airport and flight level	1 airport – MCO	Identifying periodicity in arrival delays	Multinomial logit model	Temporal factors, weather factors (precipitation)
Choi et al, 2016	Arrival delay (binary)	Flight level	45 major airports in US	Identifying weather factors of arrival delay	Machine learning approach	Temporal factors, and weather factors such as wind speed, visibility, precipitation, snow depth, and weather intensity code
Pérez-Rodríguez et al., 2017	Arrival/departure delay (binary)	Flight level	All US airports	Estimating the daily probabilities of delay in aircraft arrivals.	Bayesian model	Trip specific factors including distance and airlines, temporal factor such as day of the week
Gui et al., 2020	Arrival Delay (categorical)	Flight level	--	Flight delay prediction	Machine learning method	Air traffic condition, weather condition, temporal factors, spatial factors including origin and destination airport

Study	**Dependent Variable**	**Spatial Resolution**	**No. of Airports**	**Objective**	**Method**	**Independent variables**
Arora & Mathur, 2020	Departure delay (binary)	Flight level	All US airports	Identifying the impact of airline choice and temporality on flight delays	Binary logit model	Trip specific factor (carrier) and Temporal factors
Wong & Tsai, 2012	Flight delay propagation (continuous)	Flight level	--	To study relationship between flight delays and the causes	Survival Model	Trip specific factors such as delay cause, aircraft type, air traffic condition (turnaround buffer time), temporal factors
V. N. Bhat, 1995	Arrival delay (binary)	Flight level	--	Identifying operating and financial factors of airline delays	Binary logit model	Operating and financial variables such as capital ratio and current ratio
Xu et al., 2008	Arrival delay (continuous)	Airport level	34 OEP airports	To predict flight delays at airports in 15-min epochs	Piecewise linear regression model	Delay cause, Departure delay, Time, GDP holding time
Wong et al., 2002	Arrival and departure delay (continuous)	Flight level	1 – Taipei airport	Identifying the factors and predict airline delays	Optimization model	Departure and arrival patterns, number of departure and arrival routes
Mueller & Chatterji, 2002	Average daily arrival and departure delay (continuous)	Airport level	10 airports in the US	Examining relation between airline demand and flight delay	Least Squares method	Traffic demand related factors such as number of departures, number of arrivals, time of the day, casual factors
Kim, 2016	Arrival delay (continuous)	Flight level	1– Denver International Airport	Forecasting flight arrival time	Nonparametric additive techniques	Arriving and departing airport capacity, weather and airline, temporal factors including day of the month and month
Deshpande & Arikan, 2012	Truncated block time (continuous)	Flight level	All airports in US	Identifying the impact of scheduled block time on arrival delay	Ordinary least square regression	Route, carrier, temporal and spatial factors, traffic condition
Lee & Zhong, 2016	Arrival delay (continuous)	Flight level	1 airport – Singapore	Studying the correlation between weather condition and flight delay	Linear regression and square root regression	Weather factors such as rainfall and thunderstorm duration
Allan et al., 2001	Arrival delay type (categorical)	Airport level	1 airport – Newark airport	Determining the delay cause and delay type based on weather data	Descriptive analysis	Weather factors including wind speed ceiling, visibility, and thunderstorm
Greenfield, 2014	Arrival delay per flight (continuous)	Carrier and route level	Top 100 airports in US	To study the effects of market competition on airline delay	Regression analysis	Weather condition, airport traffic and market structure market structure, airline demand

Finally, we perform spatial data enhancement in our study by considering all flights between 67 airports across the US to capture the effects of spatial factors on flight level delay. The selected 67 airports are a subset of ASPM 77 airports and include all operational evolution partnership (OEP-35) airports in the US. The data for our analysis is augmented with other independent variables including (a) trip specific factors (carrier and flight distance), (b) spatial factors (region of origin and destination airports) and (c) temporal factors (season, day of the week and time of the day). The reader would note that the current study is the first effort to consider the influence of high resolution spatio-temporal weather conditions along the entire flight on flight delay.

Employing the data prepared, the current research contributes to airport departure and arrival delay analysis by developing a novel copula-based group generalized ordered logit (GGOL) model. The proposed framework recognizes that delay measure in minutes is not exclusively a categorical variable or a continuous variable. A cursory examination of delay variable would indicate the presence of clusters of data points as delay increases i.e., as delay increases, it is likely to be rounded to larger time bins (such as 5 minutes or 15 minutes). For analyzing such data, the application of a purely discrete outcome model system while feasible, does not allow the estimation of a continuous measure in prediction (without any strong assumptions). On the other hand, employing a continuous variable representation is not appropriate with rounded values. Thus, in our proposed research we employ a hybrid framework that ties the continuous delay measure to a categorical variable allowing us to estimate the model as a discrete outcome system with the inherent ability to predict as a continuous variable (Tirtha et al., 2020; Tirtha et al., 2022a; Yasmin & Eluru, 2018) (see more details in the section 5.3).

Our proposed model system also recognizes that it is very plausible that there might be some common unobserved factors influencing both delay categories. Given the obvious interactions between two types of delay variables, we develop a copula-based group generalized ordered logit model framework that accommodates for the influence of common observed and unobserved effects on flight departure and arrival delays. In this study, we also estimate and parameterize the error variance of the delay component to account for heteroscedasticity. The two GGOL model components are then stitched together as a joint distribution using the flexible copula-based approach. In our analysis, we employ six different copula structures – the Gaussian copula, the Farlie-Gumbel-Morgenstern (FGM) copula, and set of Archimedean copulas including Frank, Clayton, Joe and Gumbel copulas (see Bhat & Eluru, 2009 for a detailed discussion). The value of the proposed model system is illustrated by comparing predictive performance of the proposed model relative to independent models of flight departure and arrival on a holdout sample (records not used in estimation). Finally, we conduct an application analysis to present the policy implications of the current research. The illustration provides a mechanism for employing the proposed model as a tool for airline carrier level or airport level delay prediction analysis using weather forecasts.

5.3 Econometric Methodology

In this section, econometric formulation of the copula-based group generalized ordered logit model (GGOL) model is presented.

5.3.1 Flight Delay Model

Let q *(q=1,2,…,Q)*, and k *(k=1,2,…,K;K=2)* be the indices to represent flight and the corresponding delay type (departure/arrival), respectively. Let j_k *(=1,2,…J;J=6*) be the index for the discrete outcome that corresponds to delay levels for delay type k. In the group ordered response model, the discrete incident duration levels (y_{qk}) are assumed to be associated with an underlying continuous latent variable (y_{qk}^{*}). This latent variable is typically specified as follows:

$$y_{qk}^{*} = (\alpha_k + \eta_{qk})\, z_{qk} + \varepsilon_{qk},\, y_{qk} = j_k \text{ if } \psi_{j_k} < y_{qk}^{*} < \psi_{j_k+1} \quad (5.1)$$

Where, z_{qk} is a vector of exogenous variables for delay type k for a flight q, α_k is row of unknown parameters, η_{qk} is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element of z_{qk}, ψ_{j_k} and ψ_{j_k+1} are the observed lower bound threshold and upper bound threshold, respectively for time interval level j_k for delay type k. In this study, ψ takes a value from -α, 5, 10, 15, 30, 60, +α. ε_{qk} captures the idiosyncratic effect of all omitted variables for delay type k. The error terms are assumed to be independently logistic distributed with variance λ_{qk}^{2}. The variance vector is parameterized as follows:

$$\lambda_{qk} = exp\big(\rho_k g_{qk}\big) \quad (5.2)$$

Where, g_{qk} is a set of exogenous variables (including a constant) associated with delay type k for a flight q and ρ_k is the corresponding parameters to be estimated. g_{qk} accommodates for the potential presence of heteroscedasticity within the grouped ordered framework. Finally, to allow for alternative specific effects, we also introduce threshold specific deviations in the model as $\sigma_{j_k} = \tau_{j_k} z_{qk}$. The probability for delay type k for time interval in category j_k is given by:

$$Pr(y_{qk} = j_k) = \Lambda\left(\frac{\psi_{j_k+1} - ((\alpha_k + \eta_{qk})\, z_{qk} + \sigma_{j_k})}{\lambda_{qk}}\right) - \Lambda\left(\frac{\psi_{j_k} - ((\alpha_k + \eta_{qk})\, z_{qk} + \sigma_{j_k})}{\lambda_{qk}}\right) \quad (5.3)$$

Where, $\Lambda(.)$ is the cumulative standard logistic distribution.

5.3.2 Bivariate Copula Model

In examining the grouped time intervals across two delay types simultaneously, the levels of correlations between two dimensions of interests depend on the type and extent of dependency among the stochastic terms (ε_{qk}) of Equation 5.1. The joint probability function of involving departure delay level j_{q1} and arrival delay level j_{q2} for flight q can be expressed as (Laman et al., 2018):

$$Pr\left(y_{q1} = j_{q1}, y_{q2} = j_{q2}\right) = Pr(\psi_{j_{q1}} < y_{q1}^{*} < \psi_{j_{q1}+1}, \psi_{j_{q2}} < y_{q2}^{*} < \psi_{j_{q2}+1}) \quad (5.4)$$

Now, the Equation 5.4 can be written as follows (Laman et al., 2018):

$$Pr\left(y_{q1} = j_{q1}, y_{q2} = j_{q2}\right)$$
$$= \sum_{a_1=1}^{2} \sum_{a_2=1}^{2} (-1)^{a_1+a_2} \left[Pr(y_{q1}^* < \psi_{j_{q1}+a_1-1}, y_{q2}^* < \psi_{j_{q2}+a_2-1})\right] \tag{5.5}$$

The copula is a device or function that generates a stochastic dependence relationship (*i.e.*, a multivariate distribution) among random variables with pre-specified marginal distributions (Bhat & Eluru, 2009), and can be defined as:

$$C_\theta(u_1, u_2, u_3, \dots, u_{\mathrm{I}}) = Pr(U_1 < u_1, U_2 < u_2, U_3 < u_3, \dots, U_I < u_I) \tag{5.6}$$

where θ is a parameter vector of the copula commonly referred to as the dependence parameter vector. The Equation 5.5 can be written within a Copula system as (Laman et al., 2018):

$$Pr\left(y_{q1} = j_{q1}, y_{q2} = j_{q2}\right)$$
$$= \sum_{a_1=1}^{2} \sum_{a_2=1}^{2} (-1)^{a_1+a_2} \left[C_{\theta_q}(u_{j_{q1}+a_1-1}, u_{j_{q2}+a_2-1})\right] \tag{5.7}$$

To allow for the dependency structure to vary across flights, the dependence parameter θ_q is parameterized as a function of observed attributes as follows:

$$\theta_q = fn(\boldsymbol{\gamma s}_q) \tag{5.8}$$

Where, $\boldsymbol{s}_q$ is a column vector of exogenous variables, $\boldsymbol{\gamma}$ is a vector of unknown parameters (including a constant) and fn represents the functional form of parameterization. Based on the dependency parameter permissible ranges, alternate parameterization forms for the four copulas are considered in our analysis. For the Clayton and Frank copulas we employ $\theta_q = exp(\boldsymbol{\gamma s}_q)$, and for Joe and Gumbel copulas we employ $\theta_q = 1 + exp\ (\boldsymbol{\gamma s}_q)$ (see Eluru et al., 2010; Wang et al., 2015; Yasmin et al., 2014 for a similar approach). In our analysis we employ Gaussian copula, Farlie-Gumbel-Morgenstern (FGM) copula and four Archimedean copulas Frank, Clayton, Joe and Gumbel copulas (Bhat & Eluru, 2009).

In examining the model structure of flight delay across two delay types, it is also necessary to specify the structure for the unobserved vector η_{qk} represented by $\boldsymbol{\Omega}$. In this paper, it is assumed that η_{qk} is drawn from a normal distribution: $\Omega \sim N(0, \boldsymbol{\pi}_k{}^{\boldsymbol{2}})$. Thus, the conditional likelihood function for flight q based on the joint probability expression in Equation 5.7 can be expressed as:

$$L_q|\Omega = \prod_{j_1=1}^{J}\prod_{j_2=1}^{J} Pr\big(y_{q1} = j_{q1}, y_{q2} = j_{q2}\big)^{w_{qj_1j_2}} \tag{5.9}$$

where $w_{qj_1j_2}$ is a dummy indicator variable. For a flight q, $w_{qj_1j_2}$ takes a value of 1 if departure delay level is j_1 and arrival delay level is j_2, and 0 otherwise. The unconditional likelihood function for flight q can be constructed as:

$$L_q = \int_{\Omega} (L_q|\Omega)\mathrm{d}\Omega \tag{5.10}$$

Now, we can express the log-likelihood function as follows:

$$LL = \sum_{q=1}^{Q} \ln(L_q) \tag{5.11}$$

The parameters to be estimated in the copula model are $\alpha_k, \tau_{j_k}, \rho, \boldsymbol{\gamma}, \boldsymbol{\pi}_k$. All the parameters are estimated by maximizing the log-likelihood function presented in Equation 5.11. The reader would note that the proposed discrete outcome model system can be employed to predict a continuous measure of delay by generating the estimate of y_{qk}^* based on model results. Thus, the proposed hybrid approach allows us to handle the presence of rounded delays (see Chapter 2 for implementation details).

5.4 Dataset Description

The main data for our study is drawn from the BTS 2019 non-stop domestic marketing carrier on time performance dataset. Marketing on time performance dataset includes departure and arrival data for 10 marketing carriers who market flights for themselves and their regional code share partners. On-time performance dataset offers flight level information including scheduled and actual gate departure/arrival date and time, departure/arrival delay in minutes, delay cause, cancellation and diversion indicator, origin and destination airports, marketing carrier and operating carrier. Initially, we started our analysis considering all the 77 ASPM airports. However, 10 of these airports do not report any considerable operations and hence, we excluded these airports from the dataset. The final dataset consists of all the flights operated in 2019 between 67 selected airports in the US. After excluding all cancelled and diverted flights, the final dataset results in a total 5,053,375 observations.

For our estimation sample, we randomly sample 200 flights departing from each of the selected 67 airports, resulting in a dataset of 13,400 records. For a validation sample, we sampled 100 flights departing from each airport amounting to 6,700 records. The dependent variables, departure delay and arrival delay are categorized (in minutes) into 6 groups (0-5, 5-10, 10-15, 15-30, 30-60, >60 minutes). Distributions of departure and arrival delay categories are presented in Figure 5.1. From the figure, we observe that 18.12% of the domestic flights in 2019 departed late and 17.97% flights arrived late by more than 15 minutes.

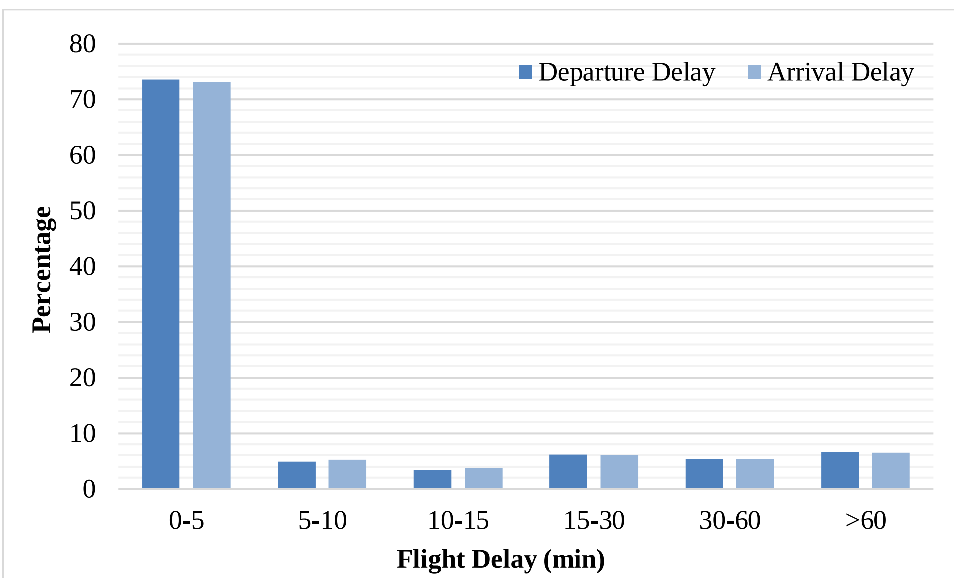


Figure 5.1 Distribution of Flight Departure and Arrival Delays

5.4.1 Independent Variables

Airline delay variables are augmented with a host of independent variables. Detailed description of the variable generation process by variable group follows.

5.4.1.1 Airport Level Traffic Conditions

Airport level traffic conditions includes air traffic and delay variables at the origin and destination airports. FAA's ASPM dataset provides hourly air traffic and delay information at the airport level. In this study, we aggregate hourly level data in the preceding 6 hours before scheduled departure and arrival time of a flight at the origin and destination airports. Airport level traffic condition at the origin (destination) airport includes scheduled number of departures (arrivals), percentage of on time gate departures (arrivals), percentage of on time airport departures, average gate departure (arrival) delay, average taxi out (in) delay, and average airport departure delay.

5.4.1.2 Trip Level Attributes

Trip level attributes are mainly sourced from BTS airline on time performance dataset and includes distance and operating carrier. In case of operating carrier, we consider 7 major operating carriers including Southwest Airlines, American Airlines, Delta Air Lines, United Air Lines, SkyWest Airlines, JetBlue Airways, and other airlines based on the distribution.

5.4.1.3 Weather Factors

We compile a comprehensive set of weather variables including thunderstorm occurrence, hourly precipitation, visibility, and wind speed at the origin, destination and along the route sourced from ASOS dataset from Iowa Environmental Mesonet (Iowa State University, 2021). The weather variable data generation process includes series of steps. First, the airline route is generated for every origin destination pair considering the shortest geodesic path between the origin and destination[9]. Second, we divide continental US into a latitude longitude grid of 5 degrees (see Figure 5.2) and compile hourly weather data from all weather stations within each grid. Third, we identify weather conditions at the origin airport during flight departure by aggregating weather data from multiple stations during departure hour and preceding 2 hours at the origin grid. Similarly, we identify weather conditions at the destination airport considering weather conditions during arrival hour and preceding 2 hours. Third, we identify the sequence of exact grid units along a route allowing us to generate the time when a flight passes through a grid and record its corresponding weather condition based on weather stations in the grid. To find the intermediate grid, we first identify the shortest route between origin and destination airports considering geodesic distance. Routes between the airports considered in this study are presented in Figure 5.2. Then, we identify direction of a flight in terms of grids using distance between origin airport and centroids of intermediate grids. In our processed dataset, number of intermediate grids between origin and destination airports varies from 0 to 11 (higher number of grids for longer flights). Finally, we allocate flight duration based on the distances between origin airport and grids' cut points to determine the hour of passing and corresponding weather condition[10]. This process allows us to generate weather conditions during the entire flight.

To illustrate the whole process, we describe the weather variable generation process in Figures 5.3 to 5.5 for a flight from John F. Kennedy International Airport (JFK) to Seattle

---

[9] The route generated might not necessarily match the exact proprietary carrier flight path, but it still provides an excellent surrogate route for consideration.

[10] It is important to note that the proposed model system is flexible to accommodate for varying number of intermediate grids for flights.

International Airport (SEA). Consider a non-stop flight that is scheduled to depart at 6:30am Coordinated Universal Time (UTC) and arrive at 12:30pm UTC. First, we identify weather conditions (90 percentile wind speed, 90 percentile precipitation, thunderstorm proportion and 10 percentile visibility across weather stations) in the origin grid at 4am-5am, 5am-6am and 6am-7am. Similarly, we identify weather condition in destination grid for 10am-11am, 11am-12pm and 12pm-1pm. Then, we aggregate weather condition measures of 3 hours to estimate origin and destination weather variables (see Figure 5.3). Second, we identify the shortest route between JFK and SEA and obtain a path of 10 intermediate grids. Now, we rank intermediate grids from 1 to 10 based on distance between JFK and centers of the grids as shown in Figure 5.4. Third, we estimate the distances of grid cut points from JFK and calculate the average distances of the grids. Based on average distance, scheduled departure time, trip length and trip duration, we determine the hour when a flight passes a grid (see Figure 5.5) and identify the weather conditions in each individual intermediate grid.

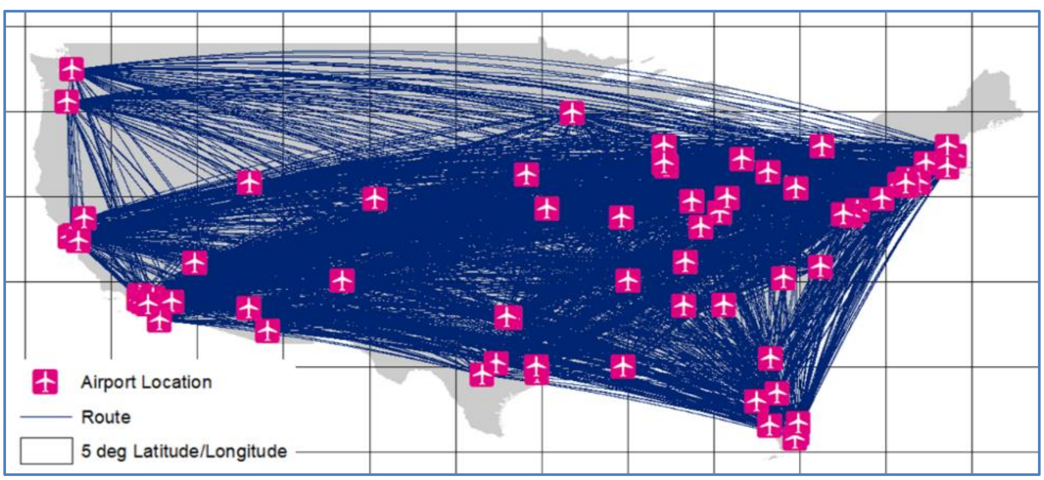


Figure 5.2 Grid System and Routes between the Airports

5.4.1.4 Spatial Factors

We consider the location of origin and destination airports in terms of US regions including South, Northeast, West, and Midwest.

5.4.1.5 Temporal Factors

In this current study, we also investigate presence of any temporal variability in flight delays. We consider different temporal variables including time of the day, day of the week and season.

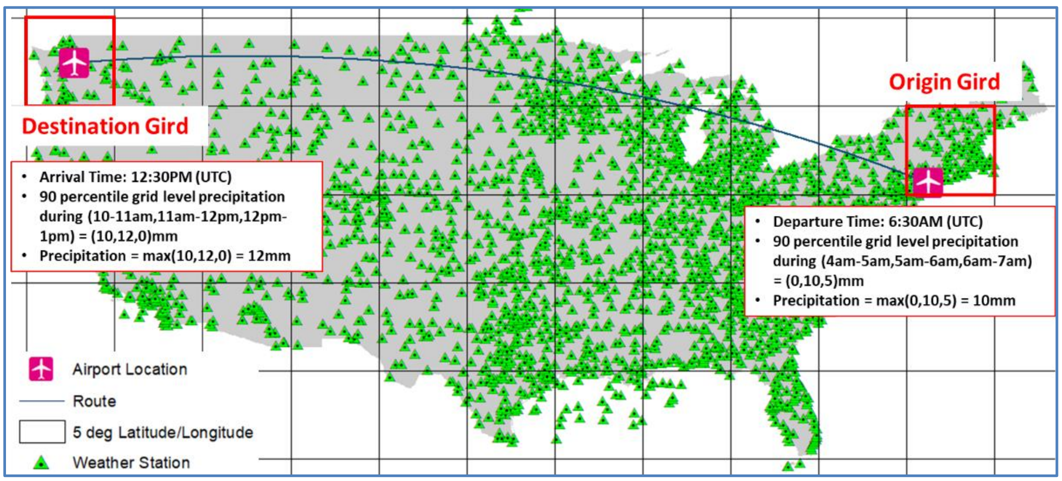


Figure 5.3 Weather Condition at Origin and Destination Airports

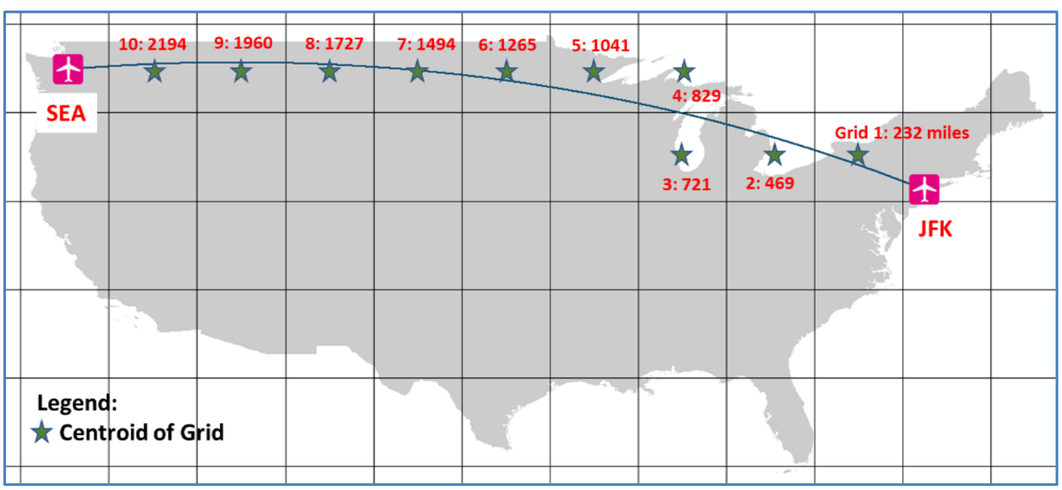


Figure 5.4 Identification of Intermediate Grids and Their Sequence

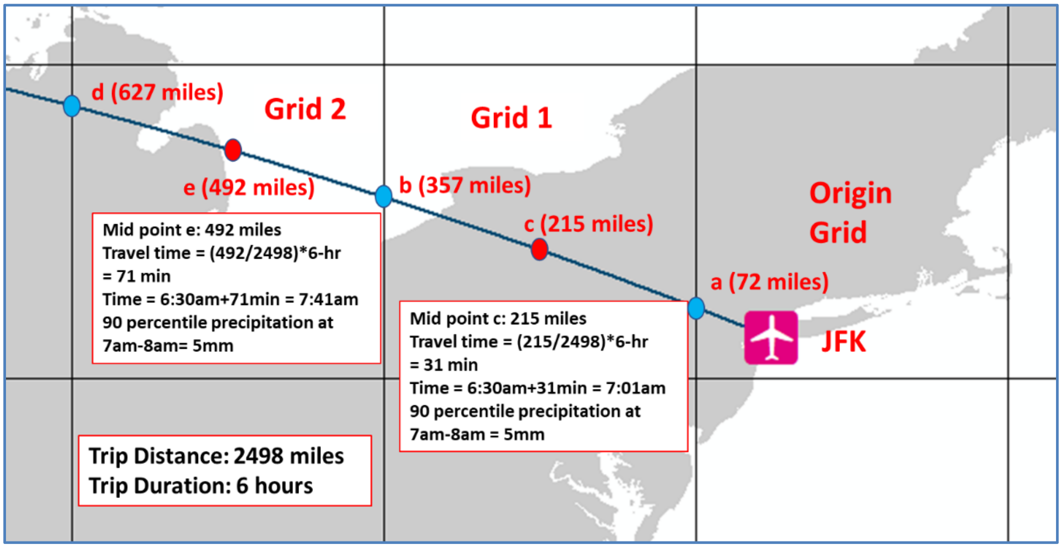


Figure 5.5 Weather Condition Estimation at Intermediate Grid

Table 5.2 offers the summary statistics (minimum, maximum and average values for continuous variables; frequency for categorical variables) of the considered exogenous variables for the estimation sample.

Table 5.2 Descriptive Statistics of Independent Variables

Continuous Variables			
Variable	**Description**	**Mean**	**Min/Max**
Airport Level Traffic Condition			
Origin Airport Level Traffic Condition			
Scheduled departures	Scheduled departures in preceding 6-hrs of flight departure	84.71	0.00/522.00
On time gate departures	% On time gate departures in preceding 6-hrs of flight departure	80.35	0.00/100.00
On time airport departures	% On time airport departures in preceding 6-hrs of flight departure	73.23	0.00/100.00
Gate departure delay	Average gate departure delay (min) in preceding 6-hrs of flight departure	12.68	0.00/344.00
Taxi out time	Average taxi out time (min) in preceding 6-hrs of flight departure	15.80	0.00/86.00
Taxi out delay	Average taxi out delay (min) in preceding 6-hrs of flight departure	5.42	0.00/76.75
Airport departure delay	Average airport departure delay (min) in preceding 6-hrs of flight departure	16.65	0.00/367.00
Destination Airport Level Traffic Condition			

Continuous Variables			
Variable	**Description**	**Mean**	**Min/Max**
Scheduled arrivals	Scheduled arrivals in preceding 6-hrs of flight arrival	152.8	0.00/530.00
On time gate arrivals	% On time gate arrivals in preceding 6-hrs of flight arrival	80.06	0.00/100.00
Taxi in delay	Average taxi in delay (min) in preceding 6-hrs of flight arrival	3.12	0.00/38.99
Block delay	Average block delay (min) in preceding 6-hrs of flight arrival	3.49	0.00/67.61
Gate arrival delay	Average gate arrival delay (min) in preceding 6-hrs of flight arrival	13.51	0.00/211.00
Trip Level Attributes			
Distance	Ln(Trip Distance+1)	6.48	4.22/7.91
Weather Factors			
Origin Grid Level Weather Condition			
Wind Speed	Max(90 percentile wind speed (mph) in origin grid during departure hour, 1 hour before, and 2 hours before departure)	12.92	2.30/35.27
Hourly Precipitation	Max(90 percentile precipitation(mm) in origin grid during departure hour, 1 hour before, and 2 hours before departure)	0.18	0.00/6.96
Thunderstorm proportion	Max(percentage of weather stations recording a thunderstorm event in origin grid during departure hour, 1 hour before, and 2 hours before departure)	1.55	0.00/59.79
Visibility	Min(10 percentile visibility (miles) in origin grid during departure hour, 1 hour before, and 2 hours before departure)	7.09	0.22/10.00
Destination Grid Level Weather Condition			
Wind Speed		13.08	1.38/37.45
Precipitation		0.17	0.00/8.83
Thunderstorm		1.55	0.00/56.67
Visibility		7.41	0.25/10.00
Categorical Variables			
Variable	**Description**	**Freq.**	**Percent**
Trip Level Attributes			
Operating Carrier			
Southwest Airlines		3602	26.88
American Airlines		1719	12.83
Delta Air Lines		1659	12.38
United Air Lines		994	7.42
SkyWest Airlines		919	6.86
JetBlue Airways		714	5.33
Other Airlines	Endeavor Air Inc., Alaska Airlines Inc., Spirit Air Lines, etc.	3793	28.31
Spatial Factors			
Region (Origin Airport)			
South		5000	37.31
Northeast		2400	17.91
West		3800	28.36
Midwest		2200	16.42
Region (Destination Airport)			

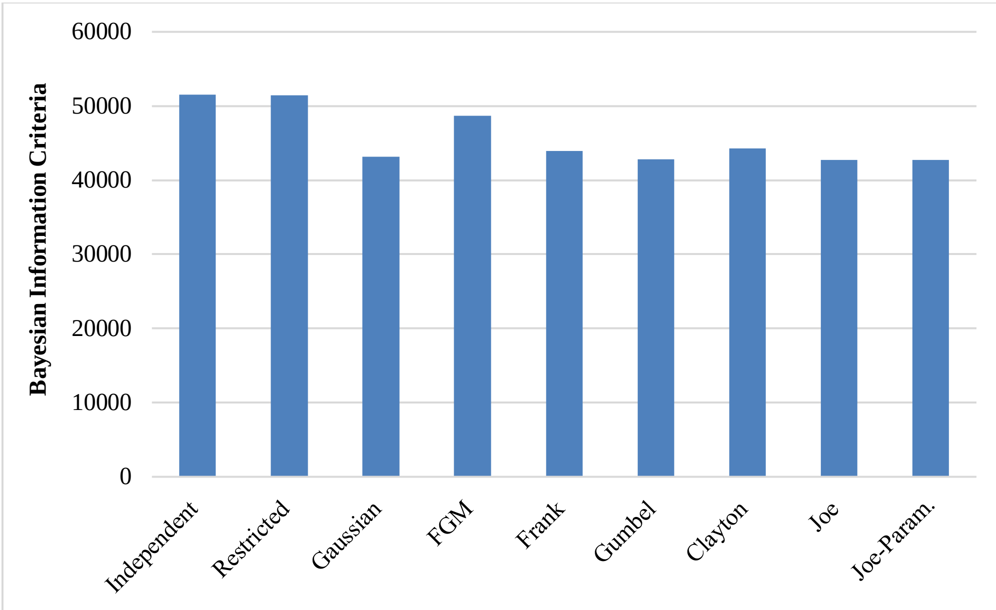
Categorical Variables			
Variable	**Description**	**Freq.**	**Percent**
South		5281	39.41
Northeast		1953	14.57
West		4005	29.89
Midwest		2161	16.13
Temporal Factors			
Time of the Day (Departure)			
Morning	6am – 10am (local time)	28.57	3829
Midday	10am – 4pm (local time)	35.96	4818
Evening	4pm – 8pm (local time)	24.11	3231
Nighttime	8pm – 6am (local time)	11.36	1522
Time of the Day (Arrival)			
Morning	6am – 10am (local time)	2474	18.46
Midday	10am – 4pm (local time)	4748	35.43
Evening	4pm – 8pm (local time)	3189	23.80
Nighttime	8pm – 6am (local time)	2989	22.31
Day of the Week (Departure)			
Saturday		1586	11.84
Other Days		11814	88.16
Day of the Week (Arrival)			
Saturday		1613	12.04
Other Days		11787	87.96
Season			
Spring	March, April, May	3519	26.26
Summer	June, July, August	3367	25.13
Fall	September, October, November	3354	25.03
Winter	December, January, February	3160	23.58

5.5 Analysis and Results

5.5.1 Model Selection

The empirical analysis involves the estimation of models by using six different copula structures: a) FGM, b) Frank, c) Gumbel, d) Clayton, e) Joe and f) Gaussian copulas. A series of models were estimated, and the best data fit is chosen based on Bayesian Information Criterion (see Figure 5.6). First, an independent copula model (separate GGOL models for flight departure delay and arrival delay) is estimated to establish a benchmark for comparison. Second, we recognize that arrivals and departures delay models have similar coefficients for 3 origin and destination grid weather variables (wind speed, precipitation, and thunderstorms). Therefore, we estimate a restricted version of independent copula model where we restrict 3 origin and destination grid weather variables to be same across departure and arrival delays. The restricted model offered improved fit relative to unrestricted model in terms of BIC. Third, six different models considering six copula dependency structures across departure delay and arrival delay are estimated. Based on log-likelihood (LL) and BIC measures, Joe copula dependency structure provides the best fit. Subsequently, the copula profile of selected Joe model has been parameterized (see Equation 8). Parameterized Joe copula model shows improved data fit in terms of the BIC measure. Further, the log-likelihood ratio test yields a statistics value of 20.64 which is substantially larger than the critical value (= 9.21) with 2 degrees of freedom at 99% confidence level. Therefore, Joe copula model with parameterization of the copula profile is selected as the final model[11].

[11] We investigated random effects of the variables and we found 1 random parameter offered a statistically significant result. However, the model with the random parameter does not improve BIC value of the model compared to the BIC value of the model without the random parameter. Hence, we did not consider the model with random parameter as our final model.



* Joe-Param. = Joe copula model with parameterization

Figure 5.6 Comparison of Alternative Models

5.5.2 Estimation Results

In this sub-section, we discuss estimation results from the joint copula model with Joe copula dependency (with parameterization).

5.5.2.1 Airport Level Traffic Conditions

Airport level traffic conditions at origin and destination airports are found to be significantly associated with flight departure and arrival delay, respectively. Among the variables considered in the analysis, number of scheduled departures and average gate departure delay at the origin airport during previous 6 hours of a flight affect departure delay while average gate arrival delay at the destination airport during previous 6 hours of flight arrival affects arrival delay. The estimation results show that increased number of scheduled departures and gate departure delay at origin airport increase the likelihood of a flight to be delayed. Similarly, increased average gate arrival delay at the destination airport increases the likelihood of a flight to be delayed. This result is very intuitive in that adverse traffic condition at the origin and destination airports mostly trigger flight delay.

5.5.2.2 Trip Level Attributes

Among trip specific factors, trip distance and operating carrier have significant effect on flight delay. Interestingly, we find the influence of trip distance on the departure delay only. The results indicates that departure delay increases with increased trip distance in general. It is an interesting finding that only departure delay is influenced by trip distance. It is plausible that longer flights have more opportunity to compensate for any initial delay by adjusting their route, a mechanism called “direct routing” (HowStuffWorks, 2019). Given this flexibility, it is possible airports alter the departure times of flights with longer distance more often than other flights. In terms of operating carrier, we find Delta Air Lines to provide the best on time performance as indicated by the negative coefficient on both departure and arrival delay. Further, the parameter estimates also suggest reduced departure delay if the flight is operated by United Air Lines and SkyWest Airlines. In terms of arrival delay, flights operated by American Airlines, JetBlue Airways and other airlines are susceptible to longer delays as indicated by the positive coefficient in Table 5.3.

5.5.2.3 Weather Factors

The results corresponding to the weather level factors highlight the important role of weather in flight’s delay (both departure and arrival). In this current study, we consider three set of weather variables: origin level, along the route and destination level. Origin level weather factors are considered in departure delay component. On the other hand, route level and destination level weather variables are considered in arrival delay component. As discussed earlier, effects of the corresponding origin level and destination level weather variables (same effect for wind speed on departure and arrival delay; similar too for hourly precipitation, and thunderstorm proportion) are restricted to be same on departure delay and arrival delay. All the weather level variables offer expected trends for both departure and arrival delay. For instance, if adverse weather condition exists at/near the origin/destination airports including higher precipitation, higher wind speed and higher frequency of thunderstorm, a flight will be more likely to experience increased departure and arrival delay which is intuitive. Further, our results also underscore the association of visibility with the arrival delay. As expected, decreased level of visibility near destination airport causes increased arrival delay. Under adverse weather conditions, flight operators are unlikely to operate under optimal conditions affecting flight speed and landing operations. It is important to note that effects of intermediate grid level weather variables are accommodated in the arrival delay model. For each weather indicator, we estimate a single effect across all intermediate grids. The results indicate that intermediate grid level hourly precipitation and thunderstorm proportion have significant positive impact on arrival delay indicating the higher likelihood of arrival delay with increased amount of precipitation and thunderstorm along the route (as expected).

5.5.2.4 Spatial Factors

The influence of spatial factors (such as location of origin and destination airports) represent factors specific to these airports that are usually unobserved to the analyst. For example, the airport crew hours and shifts are likely to be similar in a region and thus can positively or negatively affect delay. The exact details of these variables are not easy to obtain. Hence, it is accommodated through regional and/or temporal indicator variables. It is evident from estimation results that flight delay is closely associated with location of origin and destination airports. Flights departing from airports located in Northeast region in the US experience less departure delay compared to flights from other regions in the US (when all other factors are the same). For arrival delay model component, we observe that flights destined to airports in the West region experience increased arrival delay compared to airports in other regions (when all other factors are the same).

5.5.2.5 Temporal Factors

Among the temporal factors considered in this study, time of the day, day of the week and season were significantly associated with flight delays. In general, departure delay is found to be less in the morning time period and higher in the evening time period compared to nighttime and midday even after controlling for scheduled arrivals and departures. On the other hand, arrival delay is found to be lower in morning and midday periods compared to other times of the day. From the parameter estimates, we found effects of day of the week and season consistent across departure and arrival delay. Results show that departure and arrival delays are lower on Saturday compared to other days in a week. It is also evident that both departure delay and arrival delay are more frequent in summer season and less frequent in fall season relative to delays in winter and spring seasons.

5.5.2.6 Threshold Specific Effects

The proposed delay model also accommodates for threshold specific effects on various predefined thresholds. The estimation results of these parameters are reported in the second-row panel of Table 3 and have no substantive interpretation.

5.5.2.7 Variance Components

We estimate variance of delay model components as a function of exogenous variables. From the results, it is evident that the morning time period variable contributes to the variance profiles of both departure and arrival delay models. Specifically, morning time period delay is subject to a higher variance relative to delay in other time periods. Additionally, Northeast region variable affects variance component of the departure delay model. Significance of such factors indicates the presence of heteroscedasticity in the delay data.

5.5.2.8 Dependence Effects

As indicated earlier, the estimated Joe copula based GGOL model with parameterization provides the best fit incorporating the correlation between departure delay and arrival delay. The result of the dependency profile is presented in the last row panel of Table 5.3. The results clearly highlight the presence of common unobserved factors affecting departure delay and arrival delay. Joe dependency is found positive indicating upper tail dependency between departure and arrival delays. Such correlation indicates that unobserved factors modifying the likelihood of higher-level departure delay categories also modify the likelihood of higher-level arrival delay categories. Among the various variables considered, we found that season variable affects dependence structure. Specifically, the results indicate a stronger dependence between departure and arrival delay during Spring and Summer seasons.

Table 5.3 Parameter Estimates of Delay Model

Variables	Departure Delay		Arrival Delay	
	Estimates	**t statistics**	**Estimates**	**t statistics**
Propensity Component				
Constant	-70.194	-15.603	-39.431	-18.244
Airport Level Traffic Condition				
Origin airport's delay condition in previous 6-hr				
Scheduled departures	0.016	3.945	--	--
Average gate departure delay (min)	0.205	6.743	--	--
Destination airport's delay condition in previous 6-hr				
Average gate arrival delay (min)	--	--	0.391	13.950
Trip Level Attributes				
Distance	5.477	9.104	--	--
Operating Carrier (base: Southwest Airlines)				
Delta Air Lines	-11.636	-5.947	-6.282	-3.306
American Airlines	--	--	7.046	5.940
United Air Lines	-9.071	-6.149	--	--
SkyWest Airlines	-6.703	-4.186	--	--
JetBlue Airways	--	--	4.952	2.910
Other Airlines	--	--	7.600	8.150
Weather Factors				
Origin level weather condition				
Wind speed (mph)	0.332	5.345	--	--
Hourly precipitation (mm)	1.083	2.278	--	--
Thunderstorm proportion	0.198	3.842	--	--
Destination level weather condition				
Wind speed (mph)	--	--	0.332	5.345
Hourly precipitation (mm)	--	--	1.083	2.278
Thunderstorm proportion	--	--	0.198	3.842
Visibility (miles)	--	--	-0.468	-3.594
Route level weather condition				
Hourly precipitation (mm)	--	--	1.842	4.953
Thunderstorm proportion	--	--	0.258	6.756
Spatial Factors				
Region (origin airport) (Base: other regions)				
Northeast	-6.937	-3.173	--	--
Region (destination airport) (Base: other regions)				
West	--	--	2.377	2.976
Temporal Factors				
Time of the day (Departure) (base: midday and nighttime)				

Variables	Departure Delay		Arrival Delay	
	Estimates	**t statistics**	**Estimates**	**t statistics**
Propensity Component				
Morning	-21.277	-8.169	--	--
Evening	4.189	4.508	--	--
Time of the day (Arrival) (base: evening and nighttime)				
Morning	--	--	-14.882	-5.786
Midday	--	--	-6.509	-7.017
Day of the week (Departure) (base: other day of the week)				
Saturday	-6.830	-3.726	--	--
Day of the week (Arrival) (base: other day of the week)				
Saturday	--	--	-9.387	-5.394
Season (base: Spring and winter)				
Summer	4.604	3.114	4.329	2.957
Fall	-8.899	-5.667	-8.701	-5.747
Threshold Specific Effect				
Threshold 2	6.930	10.707	8.034	12.490
Threshold 3	2.749	6.724	3.330	8.144
Threshold 5	-3.664	-6.575	-2.724	-5.113
Variance Component				
Constant	3.463	139.902	3.467	148.611
Time of the day (Departure) (base: other time)				
Morning	0.152	3.691	--	--
Time of the day (Arrival) (base: other time)				
Morning	--	--	0.100	2.359
Region of origin airport (Base: Other regions)				
Northeast	0.119	3.067	--	--
Dependence Effect				
Variables	**Estimates**		**t statistics**	
Constant	0.822		24.693	
Season (base: Fall and Winter)				
Spring	0.198		4.064	
Summer	0.177		3.661	

5.6 Model Validation

To test the predictive performance of the proposed model, we perform a validation exercise with the 6700-record holdout sample. For testing the predictive performance of the copula model and its independent counterpart, 25 data samples of 500 records each, are randomly generated from the hold out validation sample. The average log-likelihood and BIC score for the proposed copula model are -807.81 [(-824.98, -790.63)] and 1895.27 [(1860.92, 1929.62)], respectively. The average log-likelihood and BIC score for independent model (with restriction) of departure and arrival delays are -968.54 [(-987.24, -949.85)] and 2235.39 [(2198.01, 2272.77)], respectively. The validation results clearly highlight the superiority of the proposed copula model over independent models (see Figure 5.7). Finally, we compare predicted shares of delay categories with observed shares for the validation sample. The comparison results are presented in Figure 5.8 and Figure 5.9. From these figures, we can clearly see that predicted shares of delay categories are very close to the observed shares.

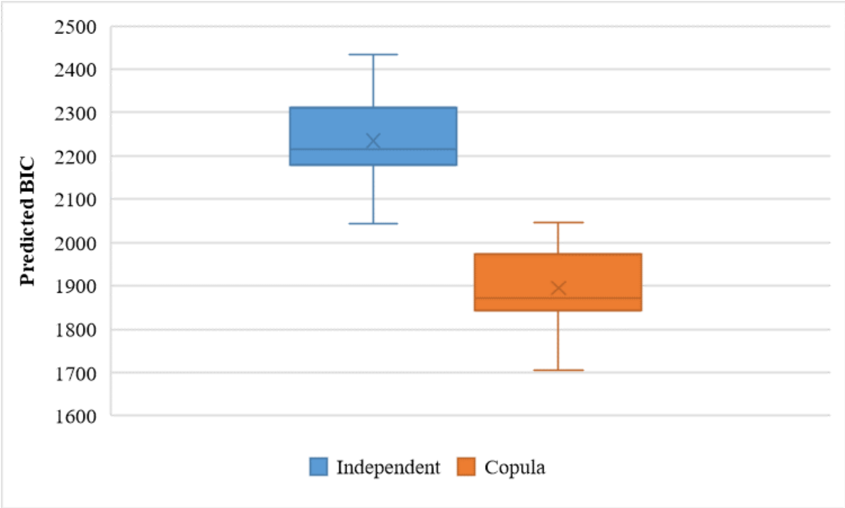


Figure 5.7 Comparison of Predictive Performance of Two Models

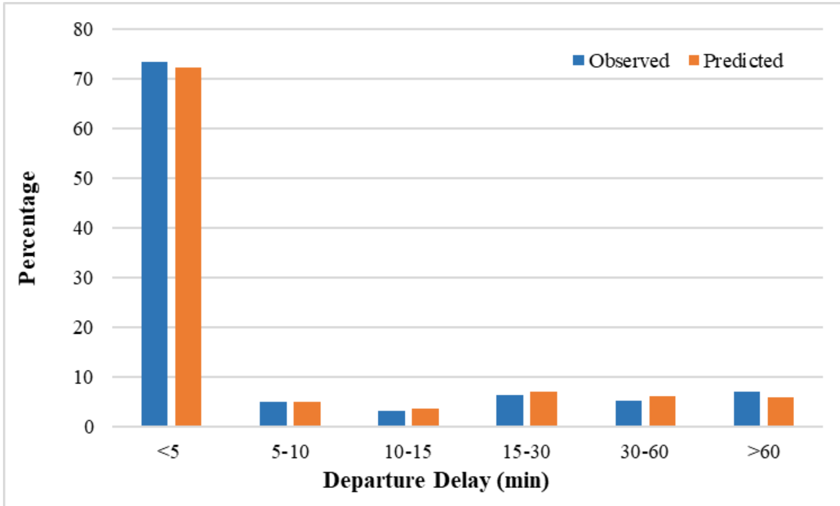


Figure 5.8 Comparison of Predicted and Observed Share of Departure Delay

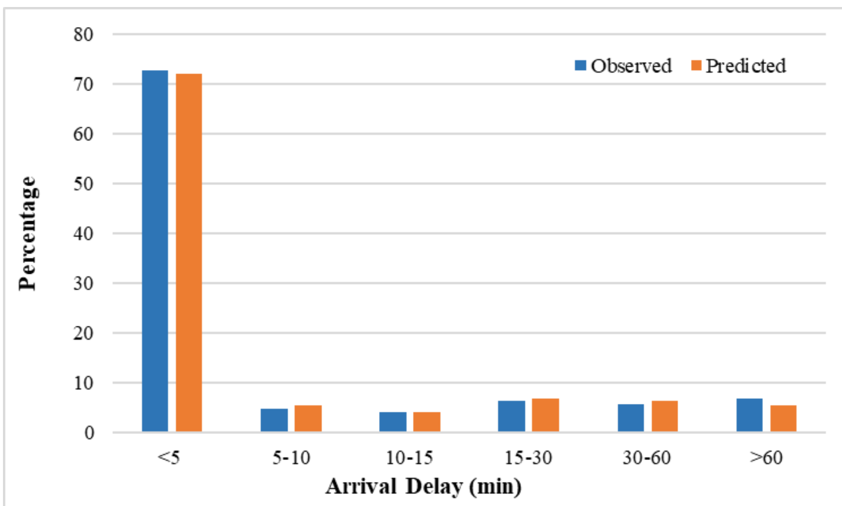


Figure 5.9 Comparison of Predicted and Observed Share of Arrival Delay

5.7 Model Illustration

Parameter estimates from Table 5.3 do not directly provide the magnitudes of the impacts of various independent variables. To illustrate the impact of independent variables, we compute the probability changes of both departure and arrival delay categories for bidirectional flights between an OD pair. We estimate probability of flight delay based on five hypothetical scenarios. For these hypothetical scenarios, we consider different weather condition attributes at the origin grid, intermediate grid, and destination grid level. In generating the probability profile, we consider the following conditions:

Scenario 1: Origin (Destination) precipitation = 0mm, Thunderstorm proportion = 0%, Wind speed = 10 mph, Intermediate grid thunderstorm proportion = 0% for all grids

Scenario 2: Origin (Destination) precipitation = 10mm, Thunderstorm proportion = 0%, Wind speed = 10 mph, Intermediate grid thunderstorm proportion = 0% for all grids

Scenario 3: Origin (Destination) precipitation = 10mm, Thunderstorm proportion = 25%, Wind speed = 10 mph, Intermediate grid thunderstorm proportion = 0% for all grids

Scenario 4: Origin (Destination) precipitation = 10mm, Thunderstorm proportion = 25%, Wind speed = 30 mph, Intermediate grid thunderstorm proportion = 0% for all grids

Scenario 5: Origin (Destination) precipitation = 10mm, Thunderstorm proportion = 25%, Wind speed = 30 mph, 3rd Intermediate grid thunderstorm proportion = 25% and 0% for others

In these scenarios, the remaining variables are considered to be the same. For ease of presentation, we identify flight delay probability as a two-alternative prediction - delay under 15 minutes or delay over 15 minutes. The probability values for delay over 15 minutes based on the above-mentioned scenarios are plotted in Figure 5.10. Departure and arrival delay probabilities are plotted for each airport considering bidirectional flights. For example, departure and arrival delay probabilities are plotted for John F. Kennedy International Airport (JFK) considering flights to and from Los Angeles International Airport (JFK-LAX and LAX-JFK). From the plots, we can clearly see that probability of delay increases with adverse weather conditions with a probability of arrival delay increasing to about 30%. Among the impact of weather variables we consider, precipitation is found to have the highest influence on flight delay while thunderstorm proportion has the least influence. It is also evident that route level weather conditions affect arrival delay, not departure delay. It is important to note that these plots are illustrations for the chosen hypothetical scenarios and can be easily generated for different values of independent variables.

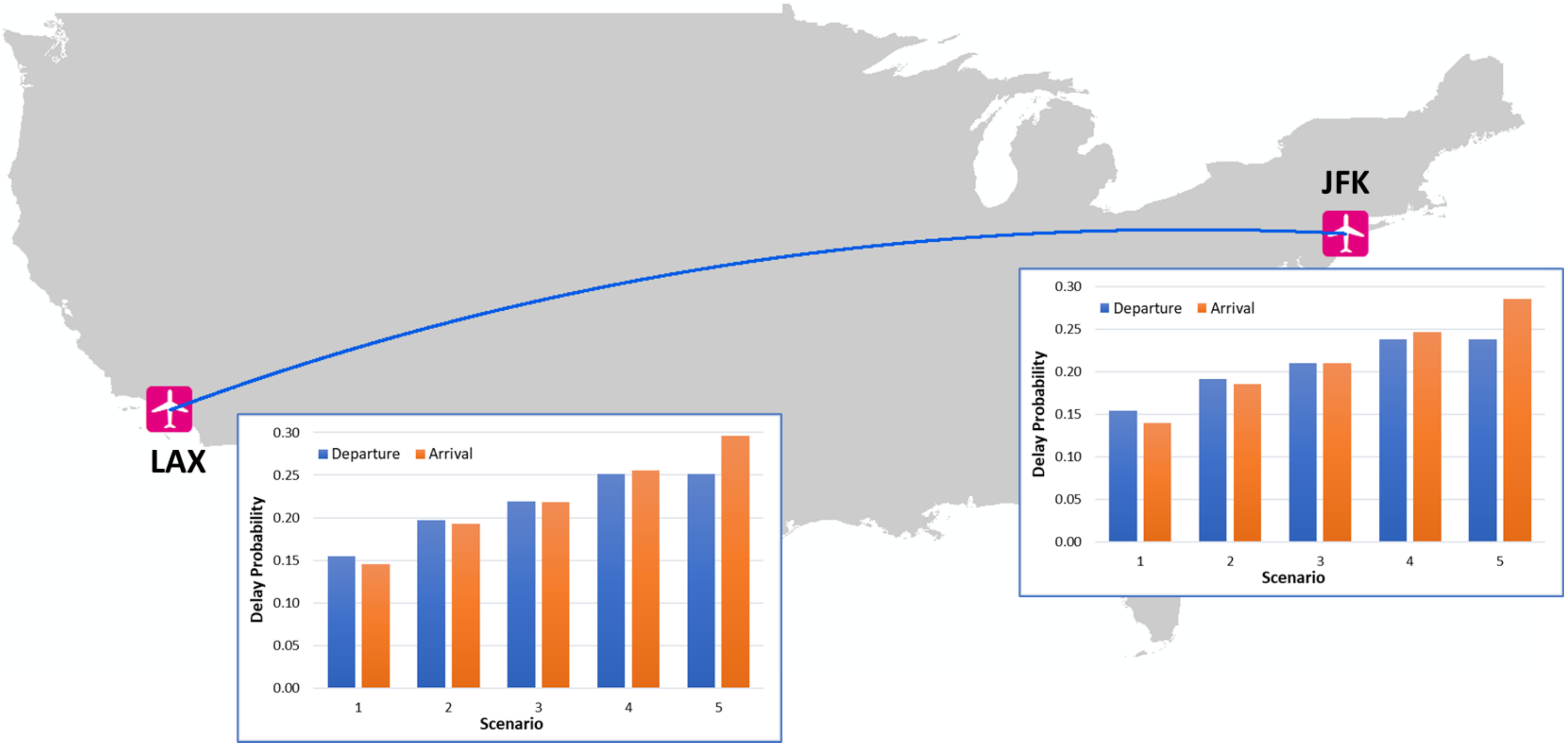


Figure 5.10 Departure and Arrival Delay Probability Based on Hypothetical Scenarios

5.8 Summary

The main focus of the current study is to identify the key factors affecting airline delay by modeling departure and arrival delays at the flight level. This study makes several contributions to airline delay literature. The first contribution of the current study arises from data enhancements for the delay analysis. The main data source of the current study is the 2019 marketing carrier on time performance data compiled by BTS. The variables processed from BTS dataset are augmented with a comprehensive set of independent variables sourced from secondary data sources including ASOS dataset and ASPM dataset. Using ASOS dataset, we prepare a comprehensive set of weather variables for the entire flight duration near the origin airport, along the flight route and the destination airport. Also, we process ASPM data to determine the traffic conditions at the origin and destination airports in the hours preceding the flight departure and arrival. The current research also contributes to airport departure and arrival delay analysis by developing a novel copula-based group generalized ordered logit (GGOL) model. The proposed model accommodates for the influence of common observed and unobserved effects on flight departure and arrival delays. In our analysis, we employ six different copula structures – the Gaussian copula, the Farlie-Gumbel-Morgenstern (FGM) copula, and set of Archimedean copulas including Frank, Clayton, Joe and Gumbel copulas.

We compare the predictive performance of independent models of departure and arrival delays and the proposed joint model with different dependency profiles. Based on the model fit measures, Joe copula model with parameterization provides the best result. The final model indicates that flight delay is significantly influenced by airport level traffic conditions, trip specific factors, weather factors, spatial factors, and temporal factors. We test the predictive performance of the proposed model by performing a validation exercise with a holdout sample. The results illustrate the superiority of the proposed model system. Finally, to illustrate the potential applicability of our model system and illustrate the impact of independent variables, we generate the probabilities for arrival and departure delays under a host of hypothetical scenarios for one bidirectional origin-destination pair. The generated airport level delay probabilities provide a framework for airlines and airports across the nation, to evaluate departure and arrival delay possibilities for their flights based on current weather predictions. The delay analysis can offer potential strategies to improve boarding, deplaning and luggage handling of flights (identified in advance to have a delay) to improve on time departure and/or quick turnaround for the next flight.

To be sure, the current study is not without limitations. In this study, we process weather variables at 5-degree latitude/longitude resolution. It would be interesting to examine if a finer resolution analysis can improve the accuracy of model by considering more localized weather data. The dataset available to us can also be improved with airline carrier specific route information to enhance the weather data collection process and contribute to an improved model. Moreover, a comparison of the developed model with machine learning approaches would be an interesting avenue for future research.

CHAPTER 6: CONCLUSIONS

Understanding the factors affecting airline demand at US airports is important for long-term planning and operational decisions. Given the recent drop in airline demand due to the pandemic, it is also important to understand the potential path to demand recovery to design plans for increasing flight availability and hiring staff for airline and airport operations. Moreover, while earlier studies examined the factors of airline demand, spatial correlations between proximally located airports have not been adequately considered. In addition to the airline demand challenges due to COVID-19, flight delays at airports have become recurrent events in recent years causing significant economic loss to commercial aviation industry. Given the negative impacts of airline delays on the US economy, understanding the factors influencing airline on time performance is important that allows airlines to improve their on-time performance or mitigate the delays by increasing and reallocating their resources such as aircrafts, crews, and staff. Within this broad vision, the dissertation makes multiple contributions. First, we propose a novel modeling approach for modeling airline demand evolution over time. Second, we explore the impact of COVID-19 on domestic airline demand in the US and provide a blueprint of recovery path in the upcoming months. Third, we build on the airline demand modeling framework by accommodating for observed and unobserved spatial and temporal effects. Finally, we study the factors affecting airline departure and arrival delays employing high resolution data at the flight level.

The first objective of this research is to identify the factors of quarterly air passenger arrivals and departures at the airport level and quantify their impact. Towards achieving this goal, the current study develops a joint panel group generalized ordered probit model system with observed thresholds for modeling air passenger arrivals and departures while accommodating for the influence of observed and unobserved effects on airline demand across multiple time periods. The second objective of this research effort is to develop a framework that provides a blueprint for airline demand recovery at a high resolution as COVID-19 cases evolve over time. Towards achieving this broad objective, the current study develops a model for analyzing airport level passenger demand data characterized as monthly departures at the airport level. The third objective of this research is to analyze monthly air passenger departures at an aggregate level of airport considering for spatial and temporal interactions between the proximally located airports. To be specific, we develop two variants of spatial models, namely spatial lag model and spatial error model in our study. Finally, the fourth objective of this research is to model departure and arrival delays in a joint framework at the disaggregate resolution of flights. Towards achieving this goal, we develop a novel copula-based group generalized ordered logit (GGOL) model system of flight departure delay and arrival delay.

Thus, the contributions of the current research are divided into four groups. Section 6.1 through section 6.4 present methodological approaches and summary of the results for objective 1 through objective 4, respectively. Section 6.5 summarizes the contributions of this dissertation. Finally, section 6.6 presents the scope for future research and limitations of the current study.

6.1 Understanding the Factors Affecting Airport Level Airline Demand

In the United States, commercial aviation sector is a significant contributor to the economy. Airline industry is closely intertwined with tourism, hospitality, and related auxiliary business (such as rental cars). An important metric to examine the health of the aviation sector is passenger demand – arrivals and departures - at airports. Understanding the factors affecting airline demand at US airports is important for long-term planning (such as airport runway and terminal design and expansion, intermodal transportation facilities) and operational decisions (such as crew management for airport services).

Thus, we develop a joint panel group generalized ordered probit model system with observed thresholds for modeling air passenger arrivals and departures. The current study contributes to the existing literature along multiple directions. The first contribution our study to the literature arises from spatial and temporal data enhancement of airline demand data from BTS. Also, in presence of airport level variables - arrivals and departures, we develop a bivariate framework that recognizes the influence of common unobserved factors. The second contribution of the research is on empirically examining the appropriate hierarchy of unobserved factors that affect airline demand. Finally, to address the inherent limitations of traditional linear models, we employ the generalized response framework for developing a non-linear framework that subsumes the linear regression model system. The proposed model is estimated using airline data compiled by Bureau of Transportation Statistics for 510 airports across the US. In model estimation, we consider a host of exogenous variables including demographic characteristics, built environment characteristics, spatial and temporal factors.

The empirical analysis shows that the flexible structure of group generalized ordered probit model (GGOP) allows us to capture the non-linearity between air travel demand and its contributing factors resulting in better data fit compared to linear regression model. To arrive at a parsimonious specification, we estimated a restricted GGOP model without any significant loss of data fit. Finally, the joint panel model that accommodates for the presence of unobserved heterogeneity performs the best in terms of empirical context highlighting the importance of accommodating for the influence of common unobserved factors affecting the two dependent variables (and their repeated measures). Finally, to illustrate how the enhanced demand model allows policy agencies to understand changes to airline demand with changes to independent variables a policy analysis is conducted. The results identify important predictors for airline demand. In particular, they highlight the role of tourism in the state, regional population, and median income.

6.2 Examining the Impact of COVID-19 on Airline Demand

Airline industry has experienced a significant shock in air passenger demand worldwide due to the recent outbreak of COVID-19. In the US, airline domestic passenger demand dropped by 476.4 million in 2020 compared to the previous year. Airline demand in the recent months has started to recover from April 2020 lows as precautions at airports, access to testing, mask mandates and finally the emergency use authorization of vaccines have encouraged some air travel. As the recovery begins airlines and airports would need to address supply side shortages with growing demand. Thus, understanding the potential path to recovery will allow airlines, airport management agencies to design plans for increasing flight availability and hiring staff for airline and airport operations.

In this context, the primary focus of our proposed research effort is to develop a framework that provides a blueprint for airline demand recovery at a high resolution as COVID-19 cases evolve over time. In our study, we analyze airport level monthly air passenger departure data for 24 months from January 2019 through December 2020 considering 380 airports across the country. In this study, we consider a host of independent variables including COVID-19 related factors, demographic characteristics and built environment characteristics at the county level, airport specific factors, spatial factors, temporal factors, and adjoining county attributes. COVID-19 related factors include both local and global factors by considering global and local COVID-19 transmission, temporal indicators of pandemic start and progress, and interactions of airline demand predictors with global and local COVID-19 indicators.

In this study, we employ a linear mixed model system that accommodates for the presence of repeated measures for modelling airline demand. In the validation exercise, we examine the performance of the proposed model by comparing observed and predicted demand for all airports across the US. From the result, we found that the proposed model successfully captures the demand drops after the start of the pandemic and the slow continuing recovery after the initial months. Subsequently, we present a blueprint for airline demand by considering three hypothetical scenarios of COVID-19 transmission rates – expected, pessimistic and optimistic. The result from the expected scenario presents a path to slow recovery as COVID-19 cases reduce. The various scenarios clearly illustrate how the proposed model can be employed to generate airline demand estimates at the airport level, state, region, or country level.

6.3 Accommodating Spatial Dependency in Airline Demand Modeling

Given the importance of understanding airline demand, earlier studies examined airline demand at different spatial and temporal resolutions. However, earlier research efforts have neglected to adequately consider for spatial interactions between air passenger demands at multiple airports. To elaborate, there may be some unobserved factors associated with closely linked spatial units that may cause spatial correlations among the airports. Neglecting such spatial correlations in the demand modelling may result in some biased estimates. Thus, the current research analyzes monthly air passenger departures at an aggregate level of airport considering for the spatial and temporal interactions between proximally located airports.

Towards this end, we develop a novel spatial group 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 the BIC measure. Importantly, both spatial models are found to be superior to the independent GGOP model that does not consider any spatial dependency between the observations. 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 autoregressive 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 lag GGOP and spatial error GGOP models compared to independent GGOP model. The result of validation exercise indicates the superiority of both spatial models relative to the independent model. Among the two spatial models, spatial error GGOP model offered improved data fit.

6.4 A Flight Level Analysis of Departure Delay and Arrival Delay

In addition to the airline demand challenges due to COVID-19, flight delays at airports have become recurrent events in recent years causing huge economic loss to commercial aviation industry. Airline delays cause both direct and indirect costs to several components of the industry. Given the negative impacts of airline delays on the US economy, understanding the factors influencing airline on time performance will allow airlines to improve their on-time performance or mitigate the delays by increasing and reallocating their resources such as aircrafts, crews, and staff.

Therefore, we develop a novel copula-based group generalized ordered logit (GGOL) model system of flight departure delay and arrival delay. This study makes several contributions to airline delay literature. The first contribution of the current study arises from data enhancements for the delay analysis. The main data source of the current study is the 2019 marketing carrier on time performance data compiled by BTS. The variables processed from BTS dataset are augmented with a comprehensive set of independent variables sourced from secondary data sources including ASOS dataset and ASPM dataset. Using ASOS dataset, we prepare a comprehensive set of weather variables for the entire flight duration near the origin airport, along the flight route and the destination airport. Also, we process ASPM data to determine the traffic conditions at the origin and destination airports in the hours preceding the flight departure and arrival. The current research also contributes to airport departure and arrival delay analysis by developing a novel copula-based group generalized ordered logit (GGOL) model. The proposed model accommodates for the influence of common observed and unobserved effects on flight departure and arrival delays. In our analysis, we employ six different copula structures – the Gaussian copula, the Farlie-Gumbel-Morgenstern (FGM) copula, and set of Archimedean copulas including Frank, Clayton, Joe, and Gumbel copulas.

In model selection, we compare the predictive performance of independent models of departure and arrival delays and the proposed joint model with different dependency profiles. Based on the model fit measures, Joe copula model with parameterization provides the best result. The final model indicates that flight delay is significantly influenced by airport level traffic conditions, trip specific factors, weather factors, spatial factors, and temporal factors. We test the predictive performance of the proposed model by performing a validation exercise with a holdout sample. The results illustrate the superiority of the proposed model system. Finally, to illustrate the potential applicability of our model system and illustrate the impact of independent variables, we generate the probabilities for arrival and departure delays under a host of hypothetical scenarios for one bidirectional origin-destination pair. The generated airport level delay probabilities provide a framework for airlines and airports across the nation, to evaluate departure and arrival delay possibilities for their flights based on current weather predictions. The delay analysis can offer potential strategies to improve boarding, deplaning and luggage handling of flights (identified in advance to have a delay) to improve on time departure and/or quick turnaround for the next flight.

6.5 Contributions of the Dissertation

The current dissertation makes substantial contributions towards addressing methodological gaps and enhancing data for airline demand and flight delay analysis. The key contributions of the current research include: 1) conducting spatial and temporal data enhancement of airline demand data, 2) examining the appropriate hierarchy of unobserved factors that affect airline demand, 3) employing the generalized response framework for developing a non-linear framework of airline demand, 4) developing a framework that provides a blueprint for airline demand recovery at a high resolution, 5) developing a novel spatial GGOP model system of airline demand considering for the spatial dependencies between the airports, 6) conducting data enhancements for the flight delay analysis, and 7) accommodating for the influence of common observed and unobserved effects on flight departure and arrival delays. In addition to the aforementioned contributions, the current research makes several contributions to aviation literature by undertaking multiple policy analyses. Such studies may act as decision-making tools for the airport officials, planners, and airlines for both short-term and long-term design and operational planning.

6.6 Limitations and Future Research

To be sure, this dissertation is not without limitations. Limitations and future research scopes are discussed below by four research objectives.

6.6.1 Understanding the Factors Affecting Airport Level Airline Demand

Towards modeling airline demand, augmenting the airline demand data in our research with local economic indicators and airport specific attributes might be an avenue for future research.

6.6.2 Examining the Impact of COVID-19 on Airline Demand

In our analysis, data was generated at the airport county level. Thus, when the same county has multiple airports, the model includes substantially similar information for these airports (except OEP 35 indicator and number of airports in a 50-mile buffer). While only 22 of the 354 counties in our data had multiple airports, it might be interesting to explore how aggregation of the demand for these airports affects the findings. Moreover, the airline demand data is available only till December 2020 which restricted us from employing linear and non-linear functions of continuous temporal variables. Given the data availability for the next few months, continuous temporal variables could be considered to enhance the current model. Further, COVID-19 pandemic is an evolving situation, and it is appropriate to consider updating the models with newer airline demand (as they become available), local vaccination data and local COVID-19 cases. Finally, the airport level analysis conducted in the paper can be augmented by examining airport level actions/strategies (such as changes to fare, priority for freight movement) in response to COVID-19 pandemic. The research might have to be conducted for a subset of airports where such data is available.

6.6.3 Accommodating Spatial Dependency in Airline Demand Modeling

It would be useful to accommodate for other socio-economic factors in the proposed model such as MSA specific GDP and business-related indicators. 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.

6.6.4 A Flight Level Analysis of Departure Delay and Arrival Delay

In this study, we process weather variables at 5-degree latitude/longitude resolution. It would be interesting to examine if a finer resolution analysis can improve the accuracy of model by considering more localized weather data. The dataset available to us can also be improved with airline carrier specific route information to enhance the weather data collection process and contribute to an improved model. Moreover, a comparison of the developed model with machine learning approaches would be an interesting avenue for future research.

LIST OF REFERENCES

Abdel-Aty, M., Lee, C., Bai, Y., Li, X., & Michalak, M. (2007). Detecting periodic patterns of arrival delay. *Journal of Air Transport Management*, *13*(6), 355–361.

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.

Allan, S. S., Beesley, J. A., Evans, J. E., & Gaddy, S. G. (2001, December). Analysis of delay causality at Newark International Airport. In *4th USA/Europe Air Traffic Management R&D Seminar* (pp. 1-11).

Arora, S. D., & Mathur, S. (2020). Effect of airline choice and temporality on flight delays. *Journal of Air Transport Management*, *86*, 101813.

Ba-Fail, A. O., Abed, S. Y., Jasimuddin, S. M., & Jeddah, S. A. (2000). The determinants of domestic air travel demand in the Kingdom of Saudi Arabia. *Journal of Air Transportation World Wide*, *5*, 72–86.

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.

Bhat, C. R., & Eluru, N. (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B: Methodological*, *43*(7), 749–765.

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, V. N. (1995). A multivariate analysis of airline flight delays. *International Journal of Quality & Reliability Management*, *12*(2), 54–59.

Bhowmik, T., & Eluru, N. (2021). A comprehensive county level model to identify factors affecting hospital capacity and predict future hospital demand. *Scientific reports*, *11*(1), 1-12.

Bhowmik, T., Tirtha, S. D., Iraganaboina, N. C., & Eluru, N. (2021). A comprehensive analysis of COVID-19 transmission and mortality rates at the county level in the United States considering socio-demographics, health indicators, mobility trends and health care infrastructure attributes. *Plos one*, *16*(4), e0249133.

Bhowmik, T., Yasmin, S., & Eluru, N. (2019). A multilevel generalized ordered probit fractional split model for analyzing vehicle speed. *Analytic Methods in Accident Research*, *21*, 13–31.

BTS. (2020). Flight Cancellations Stabilize in May, but Total Flights Hit Another Record Low. Retrieved January 28, 2022, from https://www.bts.gov/data-spotlight/flight-cancellations-stabilize-may-total-flights-hit-another-record-low

BTS. (2022a). Passengers All Carriers - All Airports. Retrieved January 28, 2022, from https://www.transtats.bts.gov/Data_Elements.aspx?Data=1

BTS. (2022b). Airline On-Time Statistics and Delay Causes. Retrieved April 15, 2022, from https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?20=E

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

Carson, R. T., Cenesizoglu, T., & Parker, R. (2011). Forecasting (aggregate) demand for US commercial air travel. *International Journal of Forecasting*, *27*(3), 923–941.

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.

CDC. (2022). CDC COVID Data Tracker. Retrieved January 29, 2022, from https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days

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

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.

Chi, J., & Baek, J. (2013). Dynamic relationship between air transport demand and economic growth in the United States: A new look. *Transport Policy*, *29*, 257–260.

Choi, S., Kim, Y.J., Briceno, S., Mavris, D. (2016). Prediction of weather-induced airline delays based on machine learning algorithms. In 2016 *IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 1–6.

Coldren, G. M., Koppelman, F. S., Kasturirangan, K., & Mukherjee, A. (2003). Modeling aggregate air-travel itinerary shares: Logit model development at a major US airline. *Journal of Air Transport Management*, *9*(6), 361–369.

CSSE. (2021). COVID-19 Map - Johns Hopkins Coronavirus Resource Center. Retrieved August 21, 2021, from https://coronavirus.jhu.edu/map.html

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.

Deshpande, V., & Arikan, M. (2012). The impact of airline flight schedules on flight delays. *Manufacturing and Service Operations Management*, *14*(3), 423–440.

Eluru, N., & Bhat, C. R. (2007). A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis and Prevention*, *39*(5), 1037–1049.

Eluru, N., Paleti, R., Pendyala, R., & Bhat, C. (2010). Modeling injury severity of multiple occupants of vehicles: Copula-based multivariate approach. *Transportation Research Record*, *2165*(1), 1–11.

Endo, N. (2007). International trade in air transport services: Penetration of foreign airlines into Japan under the bilateral aviation policies of the US and Japan. *Journal of Air Transport Management*, *13*(5), 285–292.

FAA. (2019). Cost of Delay Estimates. Retrieved from https://www.faa.gov/data_research/aviation_data_statistics/media/cost_delay_estimates.pdf

FAA. (2020). Forecast Process for 2020 TAF. Retrieved February 15, 2022, from https://taf.faa.gov/Downloads/ForecastProcessfor2020TAF.pdf

FAA. (2022). Air Traffic By The Numbers. Retrieved January 27, 2022, from https://www.faa.gov/air_traffic/by_the_numbers/

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.

Gallego, I., & Font, X. (2021). Changes in air passenger demand as a result of the COVID-19 crisis: Using Big Data to inform tourism policy. *Journal of Sustainable Tourism*, *29*(9), 1470-1489.

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

Greenfield, D. (2014). Competition and service quality: New evidence from the airline industry. *Economics of Transportation*, *3*(1), 80–89.

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

Grubb, H., & Mason, A. (2001). Long lead-time forecasting of UK air passengers by Holt-Winters methods with damped trend. *International Journal of Forecasting*, *17*(1), 71–82. https://doi.org/10.1016/S0169-2070(00)00053-4

Gudmundsson, S. V., Cattaneo, M., & Redondi, R. (2020). Forecasting Recovery Time in Air Transport Markets in the Presence of Large Economic Shocks: COVID-19. *SSRN Electronic Journal*, 1–17.

Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., & Zhao, D. (2020). Flight delay prediction based on aviation big data and machine learning. *IEEE Transactions on Vehicular Technology*, *69*(1), 140–150.

Hao, L., Hansen, M., Zhang, Y., & Post, J. (2014). New York, New York: Two ways of estimating the delay impact of New York airports. *Transportation Research Part E: Logistics and Transportation Review*, *70*(1), 245–260.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*(358), 320–338.

HowStuffWorks. (2019). How Do Pilots Make Up Time in the Air? Retrieved October 20, 2021, from https://science.howstuffworks.com/transport/flight/modern/do-pilots-make-up-time-in-air.htm

Hwang, C. C., & Shiao, G. C. (2011). Analyzing air cargo flows of international routes: an empirical study of Taiwan Taoyuan International Airport. *Journal of Transport Geography*, *19*(4), 738–744.

ICAO. (2022). Economic Impacts of COVID-19 on Civil Aviation. Retrieved January 28, 2022, from https://www.icao.int/sustainability/Pages/Economic-Impacts-of-COVID-19.aspx

Insider, B. Accessed on July 18, 2020, URL: https://www.businessinsider.com/the-most-popular-us-states-for-tourism-2014-10

Institute for Health Metrics and Evaluation (IHME). (2021). COVID-19 Projections. Retrieved August 25, 2021, from https://covid19.healthdata.org/united-states-of-america?view=cumulative-deaths&tab=trend

Iowa State University. (2021). Iowa Environmental Mesonet. Retrieved from https://mesonet.agron.iastate.edu/request/download.phtml?network=FL_ASOS

Ito, H., & Lee, D. (2005a). Assessing the impact of the September 11 terrorist attacks on U.S. airline demand. *Journal of Economics and Business*, *57*(1), 75–95.

Ito, H., & Lee, D. (2005b). Comparing the impact of the September 11th terrorist attacks on international airline demand. *International Journal of the Economics of Business*, *12*(2), 225–249.

Kalić, M., Kuljanin, J., & Dožić, S. (2014). Air travel demand fuzzy modelling: Trip generation and trip distribution. *Advances in Intelligent Systems and Computing*, *223*, 279–289.

Kim, M. S. (2016). Analysis of short-term forecasting for flight arrival time. *Journal of Air Transport Management*, *52*, 35–41.

Klein, A., Craun, C., & Lee, R. S. (2010). Airport delay prediction using Weather-Impacted Traffic Index (WITI) model. *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, 1–13.

Laman, H., Yasmin, S., & Eluru, N. (2018). Joint modeling of traffic incident duration components (reporting, response, and clearance time): A copula-based approach. *Transportation Research Record*, *2672*(30), 76–89.

Lee, L. F., & Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, *154*(2), 165–185. https://doi.org/10.1016/j.jeconom.2009.08.001

Lee, Y. X., & Zhong, Z. W. (2016). A study of the relationship between adverse weather conditions and flight delay. *Journal of Advances in Technology and Engineering Research*, *2*(4), 113–117.

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*, 267–291.

Li, T., Baik, H., & Trani, A. A. (2013). A method to estimate the historical US air travel demand. *Journal of Advanced Transportation*, *43*, 249–265.

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.

Maneenop, S., & Kotcharin, S. (2020). The impacts of COVID-19 on the global airline industry: An event study approach. *Journal of Air Transport Management*, *89*, 101920.

Markovic, D., Hauf, T., Röhner, P., & Spehr, U. (2008). A statistical study of the weather impact on punctuality at Frankfurt Airport. *Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling*, *15*(2), 293–303.

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.

Mueller, E. R., & Chatterji, G. B. (2002). Analysis of aircraft arrival and departure delay characteristics. *AIAA's Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum*, 1–14.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142.

Nayak, N., & Zhang, Y. (2011). Estimation and comparison of impact of single airport delay on national airspace system with multivariate simultaneous models. *Transportation Research Record*, *2206*(1), 52–60.

Nižetić, S. (2020). Impact of coronavirus (COVID-19) pandemic on air transport mobility, energy, and environment: A case study. *International Journal of Energy Research*, *44*(13), 10953–10961.

Pérez–Rodríguez, J. V., Pérez–Sánchez, J. M., & Gómez–Déniz, E. (2017). Modelling the asymmetric probabilistic delay of aircraft arrival. *Journal of Air Transport Management*, *62*, 90–98.

Pine, R., & McKercher, B. (2004). The impact of SARS on Hong Kong's tourism industry. *International Journal of Contemporary Hospitality Management*, *16*(2), 139–143.

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. *Transportmetrica A: Transport Science*, *15*(2), 1081–1101.

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.

Rengaraju, V. R., & Arasan, V. T. (1992). Modeling for air travel demand. *Journal of Transportation Engineering*, *118*(3), 371–380.

Roadmaps, C. H. R. Accessed on July 18, 2020, URL: https://www.countyhealthrankings.org/

S & P global. (2020). Retrieved December 8, 2020, from https://www.spglobal.com/marketintelligence/en/news-insights/blog/industries-most-and-least-impacted-by-covid19-from-a-probability-of-default-perspective-september-2020-update

Schaefer, L., & Millner, D. (2001). Flight delay propagation analysis with the detailed policy assessment tool. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, *2*, 1299–1303.

Sivrikaya, O., & Tunç, E. (2013). Demand forecasting for domestic air transportation in Turkey. *The Open Transportation Journal*, *7*(1).

Sun, X., Wandelt, S., & Zhang, A. (2020). How did COVID-19 impact air transportation? A first peek through the lens of complex networks. *Journal of Air Transport Management*, *89*, 101928.

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.

Suzuki, Y. (2000). The relationship between on-time performance and airline market share: A new approach. *Transportation Research Part E: Logistics and Transportation Review*, *36*(2), 139–154.

Tirtha, S. D., Bhowmik, T., & Eluru, N. (2022a). Understanding the Factors Affecting Airport Level Demand (Arrivals and Departures) Using a Novel Modeling Approach. *Working Paper, University of Central Florida.*

Tirtha, S. D., Bhowmik, T., & Eluru, N. (2022b). A Flight Level Analysis of Departure Delay and Arrival Delay Using Copula-based Joint Framework. *Working Paper, University of Central Florida*.

Tirtha, S. D., Bhowmik, T., & Eluru, N. (2022c). An Airport Level Framework for Examining the Impact of COVID-19 on Airline Demand. *Transportation Research Part A: Policy and Practice*, *159*, 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., & Ugarte, D. D. L. T. (2013). Evaluating the spatial spillover effects of transportation infrastructure on agricultural output across the United States. *Journal of Transport Geography*, *30*, 47-55.

Travel, M.F. (2021). The Busiest Airports in the US. Retrieved August 25, 2021, from https://myfunkytravel.com/the-busiest-airports-in-the-us.html

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

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

Wang, K., Yasmin, S., Konduri, K. C., Eluru, N., & Ivan, J. N. (2015). Copula-Based Joint Model of Injury Severity and Vehicle Damage in Two-Vehicle Crashes. *Transportation Research Record*, *2514*(1), 158–166.

Wei, W., & Hansen, M. (2006). An aggregate demand model for air passenger traffic in the hub-and-spoke network. *Transportation Research Part A: Policy and Practice*, *40*(10), 841–851.

Wong, J. T., & Tsai, S. C. (2012). A survival model for flight delay propagation. *Journal of Air Transport Management*, *23*, 5–11.

Wong, J. T., Li, S. L., & Gillingwater, D. (2002). An optimization model for assessing flight technical delay. *Transportation Planning and Technology*, *25*(2), 121–153.

Worldometer. (2022). COVID-19 CORONAVIRUS PANDEMIC. Retrieved January 29, 2022, from https://www.worldometers.info/coronavirus/#countries

Xiaokun, W., & Kockelman, K. M. (2006). Tracking land cover change in mixed logit model: Recognizing temporal and spatial effects. *Transportation Research Record*, *1977*(1), 112–120.

Xu, N., Sherry, L., & Laskey, K. B. (2008). Multifactor model for predicting delays at U.S. airports. *Transportation Research Record*, *2052*(1), 62–71.

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.

Yasmin, S., & Eluru, N. (2013). Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accident Analysis and Prevention*, *59*, 506–521.

Yasmin, S., & Eluru, N. (2018). A mixed grouped response ordered logit count model framework. *Analytic Methods in Accident Research*, *19*, 49–61.

Yasmin, S., Eluru, N., Pinjari, A. R., & Tay, R. (2014). Examining driver injury severity in two vehicle crashes – A copula based approach. *Accident Analysis & Prevention*, *66*, 120–135.

Yu, N., de 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. (Cecilia), Luo, Q., Nikolova, G., Sun, J., Hughes, B., … 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*, 91–103.