To Cite: Park, S., & Pan, B. (2018). Identifying the next non-stop flying market with a big data approach. Tourism Management, 66: 411-421.

Identifying the Next Non-Stop Flying Market with a Big Data Approach

So Young Park, M.S.
Department of Recreation, Park, and Tourism Management
School of Health and Human Development
Pennsylvania State University, University Park, USA
Telephone: 1-814-862-8212
Fax: 1-843-953-5697
E-mail: soyoungpark@psu.edu

Bing Pan, Ph.D.
Associate Professor
Department of Recreation, Park, and Tourism Management
School of Health and Human Development
Pennsylvania State University, University Park, USA
Telephone: 1-814-867-2900
Fax: 1-843-953-5697
E-mail: bingpan@psu.edu

May 9, 2017

Acknowledgements

The authors sincerely thank Perrin Lawson, the Deputy Director and Vice President, and the Charleston Area Convention and Visitors Bureau for providing data and funding for this research.

Author Bios

So Young Park, M.S., is pursuing a Ph.D. degree in the Department of Recreation, Park, and Tourism Management at Pennsylvania State University, University Park. Her research interests include data-driven analysis in tourism, tourism economics, and child abuse in tourism.

Bing Pan, Ph.D., is Associate Professor in the Department of Recreation, Park, and Tourism Management at Pennsylvania State University, University Park. His research interests include data analytics, tourism big data, destination marketing, and benefits of travel.

Identifying the Next Non-Stop Flying Market with a Big Data Approach

ABSTRACT

Destination Marketing Organizations (DMOs) strive to increase visitor volume through targeting potential markets and eliminating barriers to travel, such as the lack of direct non-stop flights. This study develops a comprehensive model to help identify the next direct flight route for a destination by deploying buying funnel theory and gravity modeling. In addition to the geographical and economic characteristics of each market of origination, web traffic at the destination's Convention and Visitors Bureau (CVB) website—a proxy for the market's interest in the destination—is used to determine the cities that would exhibit the most potential to generate visitors if a non-stop flight route was opened. The model estimates each market's potential, using multiple gravity models, and compares it to the market's interest in the destination based on buying funnel theory. The present study then empirically tests the model using the actual data of Charleston, South Carolina, where five potential cities were identified.

Keywords: Gravity model, buying funnel theory, direct flight, big data, web traffic

Identifying the Next Non-Stop Flying Market with a Big Data Approach

1. Introduction

Attracting more visitors is one of the major responsibilities of a Destination Marketing Organization (DMO). A DMO can increase visitor volumes through targeted marketing efforts and by eliminating possible barriers to visiting a destination. A series of literature on attracting more visitors has tried to identify potential customers and pinpoint target markets through market segmentation (e.g., Andereck & Caldwell, 1994; Dolnicar, 2008; Jang, Morrison & O'Leary, 2002; Müller & Hamm, 2014; Smith, 1956; Tkaczynski, Rundle-Thiele, & Beaumont, 2009). Another part of increasing visitor volume is to identify and eliminate barriers to travel (Heung, Kucukusta, & Song, 2011), including no direct flight.

The lack of a non-stop flight route can be a barrier for visitors from potential markets to reach a destination, but our research was unable to locate any study that attempted to generate visitors by creating a direct flight route. This paper introduces a model for targeting potential markets and eliminating the lack of a non-stop flight as a possible barrier for a tourist destination. It introduces multiple gravity models based on market characteristics and actual visitation, and it tries to identify the markets that would produce the most significant increase in visitor volume if a non-stop flight route were to open for those points of origin.

The dependent varibles for gravity models in this study are the number of flight passengers, the numbers of mobile devices used by visitors, and the number of hotel guests. They are three different representations of different types of visitors. Flight passengers include visitors arriving via air; the number of mobile devices represents all visitors with a mobile device, and hotel guests correspond to overnight visitors. The examination of different visitor types allows

the study to provide a single model that can investigate heterogeneous visitor types. The explanatory variables are proxies for a market of origin's geographical and economic characteristics: population, income, geographical distance, and the time and cost to reach the destination.

Regardless of how favorable the geographical and economic characteristics are, if a market of origin is not interested in the destination, targeting such a market would be to no avail. Hence, the model examines each market's interest in the destination as well. The level of interest is represented by the maket's web traffic at the destination's Convention and Visitors Bureau (CVB) website.

The present study adopts buying funnel theory as a theoretical framework for using gravity models to estimate a market's potential to generate visitors, and for including non-stop flight availability and web traffic as well. In marketing, buying funnel theory explains customers' decision process (Clow, 2013; Jerath, Ma, & Park, 2014). The process consists of four steps: awareness, research, decision, and purchase. In the research and decision steps, prospective visitors may search online for destinations with desired features. Web traffic for the CVBs of such destinations represents visitor interest. It can be considered a proxy for potential interest as travelers search information online prior to visiting a city.

The availability of a direct flight would come into the decision-making process between research and decision, since people tend to pursue faster and more comfortable ways to travel (Gronau, 1970). This study considers the non-availability of direct flights as a barrier for selecting a destination. Hence, creating a direct flight can function as a mechanism to increase the number of passengers for airlines and the number of visitors to the destination (Fuji, Im, & Mak, 1992; Tveteras & Roll, 2014).

With the estimation models and the actual data, the present study empirically tests the theoretical framework with a tourist destination, and, specifically, Charleston, South Carolina. The paper locates markets with great potential for travel to Charleston and a high interest in the city as a destination, but without the number of visits they would generate if there were direct flights available. This study identifies the top five potential markets for Charleston's next non-stop air route. The model looks at three different representations of potential visitors: flight passengers, the number of mobile devices used by visitors, and hotel guests. The model could serve as a means for any destination to target its potential markets for generating inbound travel.

2. Literature Review

This paper proposes a model for each destination in searching for potential markets using economic and geographical characteristics, availability of a non-stop flight route, and web traffic at the local DMO's website. In this section, the paper introduces previous studies that used gravity modeling to provide the basis for including economic and geographical factors. It also examines literature that incorporated the buying funnel theory to provide the rationale for interpreting a lack of direct flights as a barrier, as well as the level of web traffic as an indicator of destination interest.

2.1 Gravity Model in Social Science and Tourism Research

Various disciplines of social science have adopted the gravity model. The gravity model originated from Newton's law of universal gravitation in physics (Newton, 1966). The law posits distance and mass as gravitational forces between two objects. Many researchers have used the model to empirically explain tourism phenomena (e.g., Balli, Balli, & Louis, 2016; Durbarry, 2008; Eryiğit, Kotil, & Eryiğit, 2010; Genç, 2013; Kaplan & Aktas, 2016; Keum, 2010; Khadaroo & Seetanah, 2008; Morley, Rosselló, & Santana-Gallego, 2014). In a simple gravity model, distance as well as each market's attraction factors are used to explain the gravitational pull that determines visitor volume. Greater distance between the destination and market of visitor origin is expected to decrease gravitational pull, whereas a greater number of attractions is expected to increase it. Hence, the gravity model uses the interaction of distance and attractions to explain the flow of visitors as follows:

$$X_{ijt} = G\frac{\left(A_{it}^{\alpha}A_{jt}^{\beta}\right)}{Distance^{\gamma}}$$

X_{ijt} represents the number of visitors between location *i* and location *j*, at time *t*. A_{it} and A_{jt} represent attraction factors at location *i* and location *j*, respectively, at time *t* (e.g. population and income are widely used as attraction factors). *G* stands for the gravitational constant, and Latin superscripts indicate the degree of impact each factor has.

Many studies have added various factors to the simple gravity model, in order to explain visitor behavior more thoroughly. Some have used socio-institutional variables, such as tourism climate index, tourism price index, earthquake, a shared border, or one-off events such as the Iraq War and the September 11 terrorist attacks (Eryiğit et al., 2010). Genç (2013) and Balli et al. (2016) both included immigration information in their gravity models, explaining tourism flow for New Zealand and bilateral traveler flows between 34 OECD countries, respectively. Khadaroo and Seetanah (2008) adjusted the gravity equation to account for the role of transportation infrastructure in inbound tourists.

Some used economic variables in their gravity models. Durbarry (2008) included the real effective price of tourism products, along with common language and EU dummy variables, to understand tourism inflows to the United Kingdom. Similarly, Hanafiah and Harun (2010) include consumer price index as a consideration of price sensitivity, along with economic crisis as a possible factor that could affect Malaysia's tourism industry. Socio-institutional and price index factors, such as these, play a big role in explaining tourism flow at an international scale.

The gravity model can explain how a market’s economic and geographical factors affect visitor flow to a destination. However, it is hard to explain the role of non-stop flights and web traffic in determining visitor volume. To do this, buying funnel theory provides insights by looking at the individual-level decision process.

2.2 Buying Funnel Theory in Tourism Research

In tourism literature, buying funnel theory is often used to understand a visitor's process of choosing a destination (Clow, 2013; Jerath et. al, 2014; Sirakaya & Woodside, 2005; Yoo & Chon, 2008). The theory dissects a consumer's decision-making process into four stages: awareness, research, decision, and purchase. At the awareness stage, a visitor is aware of the existence of the destination. The next stage is research, during which the visitor has a decision task and starts searching for information and accumulates knowledge about a number of destinations that satisfy the visitor's goals and objectives. For example, the visitor may realize a need to choose a destination for an upcoming vacation. At the third stage, which is the decision stage, the visitor forms a choice set of alternative destinations and compares the destinations to make a decision. The last stage is the purchase, when the visitor decides on a single destination and carries out the steps needed to book travel and reach the destination (Jasen & Schuster, 2011; Sirakaya & Woodside, 2005; Yoo & Chon, 2008).

Many studies have shown that internal and external factors affect the visitor's formation of a choice set during the research and decision steps (Fodness & Murray, 1999; Jun, Vogt, & MacKay, 2010; Pan & Fesenmaier, 2006; Yoo & Chon, 2008). Internal factors are connected to personal experiences, such as attitudes, beliefs, and lifestyles. External factors account for the outside forces that affect visitors, such as word-of-mouth from the Internet, media, family, and friends. For example, recent studies assert that the Internet is the most important external information source for a visitor's decision process (Jun et al., 2010; Murphy & Chen, 2016; Xiang, Wang, O'Leary, & Fesenmaier, 2015).

With data collected in 2001 by the Canadian Tourism Commission (CTC), Jun et al. (2010) were able to confirm that visitors—especially those who purchased accommodations online—were likely to search for information online and visit a destination's official websites while planning travel. Murphy and Chen (2016) observed and surveyed 19 participants to analyze visitors' online search behavior while planning visits. They were able to confirm that the participants visited online information sources ranging from search engines to destinations' official websites. With surveys conducted over six years, Xiang et al. (2015) found that approximately 40 percent of people looked at destination websites to plan their travel. These results provide a rationale for interpreting web traffic on CVB websites as an indicator of interest in a destination.

While the review of previous studies provided a rationale to employ a gravity model and include web traffic in the analysis, our research was unable to locate any literature that used direct flights and web traffic information in a gravity model or in buying funnel theory. Both of these factors are critical in examining visitor behavior. Web traffic from a particular market may indicate a high level of interest in a destination, but a lack of non-stop flights can be a barrier for travelers visiting it. Without direct flights, travelers may eliminate a destination from their choice set during the research or decision stage.

In the next section, this study describes its theoretical background in merging buying funnel theory and gravity modeling. Incorporating direct flights and web traffic into the model will also be examined in detail.

3. Theoretical Framework

Due to the availability of data, this study adopts Charleston, South Carolina, USA as an example. The theoretical framework is demonstrated in the following figures (Figure 1 and 2). In

Figure 1, the double circle in the middle represents the Charleston Metropolitan Statistical Area (MSA), and grey dots in the orbit are visitors' markets of origin as MSAs. The size of the grey dots is determined by the population and income of the originating market, and it represents the amount of possible travel volume. If an MSA has greater income and population, the residents in the MSA are considered to have a higher possibility of traveling to Charleston. The black dashed circles indicate distance. The distance between Charleston and the originating MSAs, on the same dashed circle, is identical. The black solid-line orbits indicate the barriers that exist in reaching Charleston. Each barrier decreases the number of visitors from the MSAs, and some are completely obstructed even though they have great potential to travel. The barriers can be explained using buying funnel theory (Figure 2).

Buying funnel theory connects the aggregated gravity model with an individual visitor's decision making. At the research and decision stage, the web traffic at a CVB website can occur as the consumer searches for features and activities at certain destinations. Hence, web traffic at a CVB website can be considered as the potential visitors' level of interest. The consumer would also decide on the maximum distance to travel due to time and budget constraints at both stages. In doing so, the consumer would encounter transportation options, and consumers prefer to travel in a more risk-free, comfortable and less time-consuming way (Gronau, 1970; De Vany, 1974; Anderson & Kraus, 1981). Availability of a direct flight can function as one of the positive factors to keep the destination in one's choice set, given that a direct flight can decrease travel time, uncertainty, and discomfort (Nicolau & Màs, 2006). Having no direct flights can make the travel less comfortable and more time-consuming, and thus deter visitors. Accordingly, the study includes distance, length of flight, availability of a direct flight, and transportation cost in the

model. An individual traveler's decision process (Figure 2) impacts the number of total visitors to a destination due to these barriers of travel (Figure 1).

---------- Insert Figure 1 and Figure 2 Here ----------

Ultimately, the gravity model explains the movement between the markets and the destination. Population, income, distance, and availability of a direct flight all work as factors affecting gravity between the market and destination. The gravity between them represents the pulling force from the destination to attract visitors from the market. Closer distance and availability of a direct flight are expected to increase gravity, while gravitational pull can be obstructed by barriers posed during the purchasing decision process. These barriers include time-related constraints, such as distance and a lack of direct flights, as well as financial and physical constraints.

4. Methods

Based on the proposed theoretical framework, this study fits three gravity models using Poisson distribution. The analytical dataset contains only non-negative values and many zeros. Since a log-normal gravity model can suffer from a bias with such datasets, this study employs the generalized linear model (GLM), with its family distribution as a Poisson distribution, to analyze Charleston's data (Appendix A).

The expected value of visitors, $E[V_{ijt}|X]$, follows the Poisson probability distribution and can be explained by the matrix of independent variables, X, which includes population, income, distance, flight time, direct flight, and relative price between the transportation modes:

$$ln(E[V_{ijt}|X]) = \beta_1 Pop_{it} + \beta_2 Pop_{jt} + \beta_3 GDP_{it} + \beta_4 GDP_{jt} + \beta_5 D_{ij} + \beta_6 DF_{ijt} + \beta_7 S_{ijt} + \beta_8 fTime_{ij} + \beta_9 fTime_{ij} * D_{ij} + \epsilon_{ijt}$$

Where

a) V_{ijt} represents the number of visitors from origin *i* to destination *j* at time *t*;

b) Pop_{it} represents the population of origin *i* at time *t*, and Pop_{jt} the population of destination *j*, at time *t*;

c) GDP_{it} and GDP_{jt} represent income level of origin *i* and destination *j*, at time *t*, respectively;

d) D_{ij} represents the physical distance between origin *i* and destination *j*;

e) DF_{ijt} represents the dummy variable for the availability of a direct flight. DF_{ijt} is 0 if there is no direct flight, and it is 1 if there is a direct flight available;

f) S_{ijt} represents the relative price of flying and driving between origin *i* and destination *j*, at time *t;*

g) $fTime_{ij}$ represents the flight time required to reach destination *j* from origin *i*;

h) The interaction term between D_{ij} and $fTime_{ijt}$ is included, since distance and transportation cost is expected to be highly correlated.

The present study estimates the above model using the pseudo-maximum-likelihood method. The model was run three times with three dependent variables, including the numbers of flight passengers, mobile device users, and hotel guests. In the latter two models, the relative cost of the flight was removed because these models do not look exclusively at visitors by air travel.

The present study then compares the fitted values and the actual data to search for the markets with the highest potential but no direct flight. Those with fitted passenger numbers much greater than actual passengers are considered as origin MSAs with the greatest possibility to increase the number of flight passengers if there were a non-stop route. Flight passengers are most likely to be visitors from more distant MSAs because they are flying, rather than driving, to reach the destination. The difference between the fitted and actual number of mobile devices

indicates the potential to increase total visitors via all transportation modes. Mobile device users would represent the most diverse group of visitors, since 85 percent of U.S. vacationers bring their mobile devices while traveling (TripAdvisor, 2013). This includes those with any purpose of visit or length of stay. A greater gap between the fitted and actual number of hotel guests denotes the potential to increase overnight visitors. The present study ranks the three gaps according to the volume of web traffic. Accordingly, MSAs with the greatest potential to visit and with the highest interest level will emerge.

---------- Insert Table 1 and Table 2 here ----------

5. Data Description

There are two sources of data adopted in this study. The first type is the data from actual visitors: these include hotel guests and airline traffic. The airline traffic data is from Seabury APG, a management consulting company specializing in aviation planning and management. The data is an annual average number of flight passengers between 2013 and 2015. The hotel guest data was derived from the guests' zip code data from 13 hotel management companies in Charleston. The number of hotel guests is a three-year average between 2013 and 2015.

The second type of adopted data is big data from digital traces left by visitors. Web traffic data is gleaned from the traffic log of the area's CVB website. The number of mobile devices is reported from a commercial company, AirSage, Inc. AirSage uses cell phone signals from two major mobile carriers, covering more than a third of U.S. mobile device users, to estimate the visitor volume at a specific location. The company looks where a person's cell is most frequently used to determine that person's residential area. If the device stayed in the

Charleston area for less than half of the time period while it's tracked, the holder of the mobile device is considered a visitor. The paper used the data from four single-month periods in 2014 and 2015. The web traffic data represents the annual average number of web sessions on the CVB main website, from 2013 to 2015 and gleaned from Google Analytics.

---------- Insert Table 3 here ----------

The paper used a 2015 estimate by the Census Bureau for the population of each MSA. For income data, a 2015 median family income estimate by the Federal Financial Institutions Examination Council (FFIEC) was used. For MSAs that were missing income information, this study used a 2015 estimate by the Census Bureau and 2016 Bureau of Economic Analysis statistics. Flight distance and driving distance were measured separately. Flight distance was calculated using the Mileage Calculator on WebFlye.com. CDXzipstream was used for measuring the driving distance. This program is a Microsoft Excel add-in that helps calculate distance between zip codes.

The paper used Google Maps to calculate the flight time between Charleston and the originating MSAs. Of all the flight options that appeared on Google Maps, the paper used the shortest time. There were 85 MSAs that were missing flight time information. Among them, the 30 MSAs that did not have average flight ticket prices were dismissed. For the remaining 45 MSAs, 11 had flight passenger data. Flight times for the 11 MSAs were estimated using the airports suggested by TripAdvisor, Orbitz and Cleverlayover. For the final regression, data from 229 MSAs was employed.

6. Results

6.1 Regression Model Results

This paper sought to identify the potential MSAs expected to supply the most visitors to a destination by establishing new direct flight routes. Table 4 shows the descriptive statistics of the data used in the three estimations. Table 5 shows the GLM regression results for the three estimations. Since the data used GLM with the family distribution as Poisson, and the link function as the logarithm, the coefficients should be exponentiated for interpretation.

---------- Insert Table 4 here ----------

Model (1) in Table 4 shows the result of the most basic gravity model, which includes only population, income, and distance. Population and income were divided by 10,000 and 1,000, respectively, to adjust to the scale. The population and income are positively correlated with the number of flight passengers. Income shows a greater degree of impact than the population. This might be due to the fact that Charleston is a high-end destination where it is rather costly to stay and dine. Distance is negatively correlated with the number of passengers—with significance, as expected.

---------- Insert Table 5 here ----------

Models (2) to (4) gradually increased the number of variables included in the regression to show the robustness of the model with the non-stop flight dummy variable. The pseudo R-squared values increase with the inclusion of the direct flight and relative cost variable. Model (2) includes flight time and its interaction with the distance. One unit of increase in flight time—a one-hour increase—is expected to decrease the number of passengers by approximately 49.9%.

Unlike Model (1), where the distance showed a negative correlation, distance shows a positive correlation in Model (2) to Model (4). This could be due to the inclusion of the interaction of distance and flight time. Also, for MSAs within a close distance, the distance is

likely to increase the number of flight passengers, because people are more likely to fly if the destination is outside driving distance.

Model (3) includes the direct flight dummy variable and the relative cost. The direct flight is expected to increase the number of passengers by approximately 203.9%. The degree of impact is not surprising considering the difference in the amount of air traffic by direct flights. The result is in accordance with Fuji et al. (1992) and Tveteras and Roll (2014).

Models (3) and (4) include the relative cost and its interaction term with the distance. Relative cost was computed by dividing average flight ticket prices by average driving cost. The average driving cost was computed by the average driving distance multiplied by the cost of gas per mile. Since the cost of gas per mile varies only slightly across MSAs, it was considered as a constant. When the relative cost is high, the flight cost is higher compared to the driving cost, and it decreases the number of flight passengers significantly. The coefficient of relative cost showed the greatest degree of impact. The number of passengers is expected to decrease 78.9 percent, on average, if there is a unit increase in the relative cost, according to Model (4). The interaction term between the relative cost and the distance shows a negative coefficient. This indicates that increase in the relative cost would negatively impact the distance's effect on the number of passengers.

Models (5) and (6) show the regression results with the dependent variable as the number of mobile devices and hotel guests, respectively. The results for a gradual increase in the variables are not provided for the mobile devices and hotel guests regressions. This is because the main focus is the flight passengers, and the robustness of the regression was already checked previously. The number of mobile devices best represents the total population of visitors. It includes all visitors with a mobile device that reached the destination via all transportation

modes. Moreover, it includes leisure tourists, business tourists, and family members visiting friends and relatives (VFRs)—in which case they stayed for any number of days. The number of hotel guests, on the other hand, may not account for VFRs. Rather, hotel guests are more likely to represent overnight visitors. The results of Model (5) and (6) mostly coincide with the findings with the flight passengers. However, distance shows a significantly negative effect on the number of total visitors and hotel guests. This is because road travelers are included in both data. Road travel would be more affected by distance than air travel. As the number of hotel guests is less affected by distance and more by flight time, it is reasonable to suspect that many hotel guests are arriving at Charleston by air.

In the next section, the paper locates potential markets by comparing the regression results with web traffic data.

6.2 Potential Market Identification

The fitted values from the regressions are compared to the actual values to identify potential markets. First, for the regression with the number of flight passengers as its dependent variable, the fitted value is the expected number of flight passengers given its income, population, distance, flight time, relative cost of flying and availability of a non-stop flight route. Hence, the difference between the fitted value and the actual number is the market's potential to fly to Charleston:

market's potential to fly = fitted value – actual flight passenger number

Those with greater negative values could be considered overachievers, whereas those with greater positive values could be seen as underachievers with a potential to visit by air.

Creating a non-stop flight would help realize such potential. However, if the market is not interested in the destination, there will be no increase in the number of visitors, even with high

potential. The markets' potential to fly should be compared to their interest in the destination to locate final, potential markets. The level of interest is the market's annual average web sessions at the destination's CVB website (Figure 3). The MSA with the highest potential increase in the number of flight passengers is Portland-South Portland, ME. The MSA with the highest interest in Charleston is Allentown-Bethlehem-Easton, PA-NJ (Figure 3).

---------- Insert Figure 3 here ----------

The comparison between the level of interest and the potential increase in the total number of visitors (mobile device users) yield a few common MSAs for the first figure: Allentown-Bethlehem-Easton, PA-NJ; Evansville, IN-KY; Lancaster, PA; and York-Hanover, PA (Figure 4). These four MSAs are likely to show an increase in both flight passengers and total visitors if there was a direct flight available. The three MSAs other than Allentown-Bethlehem-Easton, PA-NJ reappear in the third figure as well (Figure 5). Hence, Charleston would see an increase in the number of flight passengers, total visitors, and hotel guests from the three MSAs if it creates a non-stop flight route to and from any of those MSAs.

---------- Insert Figure 4 here ----------

Another overlap between the hotel guest and the flight passenger analysis is Springfield, MA (Figure 5). Though Springfield may not see the most increase in the total number of visitors, hotel guests and flight passengers would increase with a direct flight between Charleston and Springfield. There are many overlaps between figure 4 and 5: Charleston, WV; Charlottesville, VA; Columbus, GA-AL; Huntington-Ashland, WV-KY-OH; Huntsville, AL; Jacksonville, NC; Montgomery, AL; and Tallahassee, FL. These MSAs represent those with the potential to increase both total visitors and hotel guests.

---------- Insert Figure 5 here ----------

The volume of hotel guests and mobile devices appears to be smaller than the flight passengers due to measurement methods. The hotel guests are data from 13 hotels in Charleston. As of 2016, there are more than 150 hotels in the Charleston MSA (STR, 2016). The total number of hotel guests would be more than 10 times the number used in the analysis. For mobile device users, the number should be multiplied by approximately 12 since it reflects four-month data from only two mobile carriers, which represent about 33 percent of the total mobile traffic (Statstica, 2016). By comparing these three analyses, this study was able to locate the top five potential MSAs for the next non-stop flight route.

---------- Insert Table 6 here ----------

7. Conclusion and Discussion

The present study has developed a theoretical framework that incorporates buying funnel theory and gravity modeling, which helped identify the most promising markets for a new direct flight route. Using a gravity model, the paper empirically showed the impact of direct flight route, and it was able to validate that the availability of a non-stop route has a strong positive relationship with the number of visitors. From the regression results, this study also found the potential markets from which the number of visitors would increase, if there was a direct flight available.

The paper looked at three types of visitors: flight passengers, those using mobile devices, and hotel guests. Flight passengers represent visitors via air; the number of mobile devices represents general visitors; and hotel guests represent overnight visitors. By comparing the numbers of flight passengers and mobile devices, it is possible to see the ratio of visitors flying to the destination. Comparison between mobile devices and hotel guests may show the ratio of

VFRs and day-trip visitors. The flight passenger and the hotel guest data can indicate what percentage of flight passengers are visitors staying overnight in hotels.

The three types of visitors provide three different types of visitor flow for more accurate examination of the visitor pattern. At the same time, the paper observes an additional flow: the flow of information. Each MSA's web traffic at the website of a destination's CVB represents an information flow. The paper interprets the information flow as interest from future visitors. Therefore, examining visitor flow helps understand the current visitor pattern, and examining the information flow helps understand the future visitor pattern. The paper's model provides a comprehensive analysis of current and future visit patterns by including both visitor flow and information flow.

The analyses of flight passengers, mobile devices, and hotel guests identified several MSAs in common: Evansville, IN-KY; Lancaster, PA; and York-Hanover, PA. If the purpose is to increase overall visitor volume, targeting these three MSAs would be beneficial. Springfield, MA was also a common denominator for flight passengers and hotel guests. Targeting Springfield would be the most beneficial to the hospitality industry in Charleston, as it would bring in tourists that would stay in hotels. Lastly, Portland-South Portland, ME was added as a potential market because it had the greatest expected increase in flight passenger volume along with very high interest. Portland-South Portland would be the most profitable for the airlines.

With the assumption that all five MSAs are targeted, two strategies are recommended: Portland-South Portland and Springfield are closely located, and Lancaster and York-Hanover are less than one hour apart. Hence, if Charleston were to open a single non-stop route, Portland-South Portland or Springfield is recommended. With the assumption that potential visitors from Springfield will fly from Portland-South Portland, this non-stop flight is expected to bring

approximately 574 passengers weekly, or 29,930 yearly. If another direct flight route can be added, either Lancaster or York-Hanover would bring the most visitors. This would bring an additional 100 passengers per week, or 35,144 per year.

The airlines, however, may not agree with the potential markets because this study selected them in the interest of a destination. Important factors for the airlines, such as current flight load, competition, and network effect were not considered when forming the model. For the airlines, the number of passengers on a flight is the sole focus. Therefore, they might possibly create a new route for which there is already a large number of passengers. However, this does not mean that the number of visitors will increase. The number of passengers for an airline can increase while the number of total visitors remains the same. The airlines would also take into account how many competing airlines exist. Moreover, the airlines consider network effect when creating new flight routes, since forming networks can result in higher competitiveness and more customers (Putsay, 1980). Further research could be done incorporating the interests of the airlines for forecasting the best potential markets for the airlines.

References

Anderson, J. E., & Kraus, M. (1981). Quality of service and the demand for air travel. The Review of Economics and Statistics, 533-540.

Andereck, K. L., & Caldwell, L. L. (1994). Variable selection in tourism market segmentation models. *Journal of Travel Research*, 33(2), 40-46.

Balli, F., Balli, H. O., & Louis, R. J. (2016). The impacts of immigrants and institutions on bilateral tourism flows. *Tourism Management*, 52, 221-229.

Clow, D. (2013, April). MOOCs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 185-189). ACM.

Creel, M. D., & Loomis, J. B. (1990). Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. *American journal of agricultural economics*, 72(2), 434-441.

De Vany, A. (1974). The Revealed Value of Time in Air Travel. *The Review of Economics and Statistics*, 56(1), 77-82.

Dolnicar, S. (2008). Market segmentation in tourism. *Tourism management, analysis, behaviour and strategy*, 129-150.

Durbarry, R. (2008). Tourism taxes: Implications for tourism demand in the UK. *Review of Development Economics*, 12(1), 21-36.

Eryiğit, M., Kotil, E., & Eryiğit, R. (2010). Factors affecting international tourism flows to Turkey: a gravity model approach. *Tourism Economics*, *16*(3), 585-595.

Flowerdew, R., & Aitkin, M. (1982). A method of fitting the gravity model based on the Poisson distribution. *Journal of regional science*, 22(2), 191-202.

Fodness, D., & Murray, B. (1999). A model of tourist information search behavior. *Journal of travel research*, *37*(3), 220-230.

Fujii, E., Im, E., & Mak, J. (1992). The economics of direct flights. *Journal of Transport Economics and Policy*, 185-195.

Genç, M. (2013). Migration and tourism flows to New Zealand. In *Quantitative methods in tourism economics* (pp. 113-126). Physica-Verlag HD.

Gronau, R. (1970). The value of time in passenger transportation: The demand for air travel. *NBER Books*.

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

Hanafiah, M. H. M., & Harun, M. F. M. (2010). Tourism demand in Malaysia: A cross-sectional pool time-series analysis. *International Journal of Trade, Economics and Finance*, *1*(2), 200.

Hellerstein, D. M. (1991). Using count data models in travel cost analysis with aggregate data. *American journal of agricultural economics*, 73(3), 860-866.

Hellström, J. (2006). A bivariate count data model for household tourism demand. *Journal of Applied Econometrics*, 21(2), 213-226.

Heung, V. C., Kucukusta, D., & Song, H. (2011). Medical tourism development in Hong Kong: An assessment of the barriers. *Tourism Management*, *32*(5), 995-1005.

Jang, S. (2005). The past, present, and future research of online information search. *Journal of Travel & Tourism Marketing*, *17*(2-3), 41-47.

Jang, S. C., Morrison, A. M., & T O'Leary, J. (2002). Benefit segmentation of Japanese pleasure travelers to the USA and Canada: selecting target markets based on the profitability and risk of individual market segments. *Tourism Management*, 23(4), 367-378.

Jansen, B. J., & Schuster, S. (2011). Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research*, *12*(1), 1.

Jerath, K., Ma, L., & Park, Y. H. (2014). Consumer click behavior at a search engine: The role of keyword popularity. *Journal of Marketing Research*, 51(4), 480-486.

Jun, S. H., Vogt, C. A., & MacKay, K. J. (2010). Online information search strategies: A focus on flights and accommodations. *Journal of Travel & Tourism Marketing*, *27*(6), 579-595.

Kaplan, F., & Aktas, A. R. (2016). The Turkey Tourism Demand: A Gravity Model. *The Empirical Economics Letters*, *15*(3), 265-272.

Keum, K. (2010). Tourism flows and trade theory: a panel data analysis with the gravity model. *The Annals of Regional Science*, *44*(3), 541-557.

Khadaroo, J., & Seetanah, B. (2008). The role of transport infrastructure in international tourism development: A gravity model approach. Tourism management, 29(5), 831-840.

Morley, C., Rosselló, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: theory and use. *Annals of Tourism Research*, *48*, 1-10.

Müller, H., & Hamm, U. (2014). Stability of market segmentation with cluster analysis–A methodological approach. *Food Quality and Preference*, 34, 70-78.

Murphy, H. C., & Chen, M. M. (2016). Online Information Sources Used in Hotel Bookings: Examining Relevance and Recall. *Journal of Travel Research*, *55*(4), 523-536.

Newton, I. (1966). *Philosophiae naturalis principia mathematica*, London 1687. English translation: Sir Isaac Newton's Mathematical Principles, Univ. of Calif. Press, Berkeley.

Nicolau, J. L., & Màs, F. J. (2006). The influence of distance and prices on the choice of tourist destinations: The moderating role of motivations. *Tourism Management*, 27(5), 982-996.

Pan, B., & Fesenmaier, D. R. (2006). Online information search: vacation planning process. *Annals of Tourism Research*, *33*(3), 809-832.

Parsons, G. R. (2003). The travel cost model. In *A primer on nonmarket valuation* (pp. 269-329). Springer Netherlands.

Pustay, M. W. (1980). Airline competition and network effects. *Transportation Journal*, 63-72.

Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4), 641-658.

Sirakaya, E., & Woodside, A. G. (2005). Building and testing theories of decision making by travellers. *Tourism management*, *26*(6), 815-832.

Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of marketing*, 21(1), 3-8.

Statistica. (2016). Wireless carrier/operator subscriber share in the U.S. 2011-2016. Retrieved from https://www.statista.com/statistics/199359/market-share-of-wireless-carriers-in-the-us-by-subscriptions/

STR. (2016). *College of Charleston.*

Tkaczynski, A., Rundle-Thiele, S., & Beaumont, N. (2009). Destination segmentation: A recommended two-step approach. *Journal of Travel Research.*

TripAdvisor, (2013). TripBarometer Mobile & Social Survey finds 85% of US travelers bring their smartphones on vacation & 61% report using social media while available at: https://www.tripadvisor.com/TripAdvisorInsights/n2006/tripbarometer-mobile-social-

survey-finds-85-us-travelers-bring-their-smartphones-vacation-61-report. Accessed May 1, 2017

Tveteras, S., & Roll, K. H. (2014). Non-stop flights and tourist arrivals. *Tourism Economics*, 20(1), 5-20.

Westerlund, J., & Wilhelmsson, F. (2006). Estimating the Gravity Model Without Gravity Using Panel Data. *Applied Economics*, 7(9), 1466-4283.

Woodside, A. G., Moore, E. M., & Etzel, M. J. (1980). Vacation travel behavior and perceived benefits of home states residents. *Business and Economic review*, 26(5), 28-34.

Xiang, Z., Wang, D., O'Leary, J. T., & Fesenmaier, D. R. (2015). Adapting to the internet: trends in travelers' use of the web for trip planning. *Journal of Travel Research*, *54*(4), 511-527.

Yoo, J. J. E., & Chon, K. (2008). Factors affecting convention participation decision-making: Developing a measurement scale. *Journal of Travel Research*, *47*(1), 113-122.

Tables and Figures

Table 1. Regression Estimation Steps

Step 1	Step 2	Step 3	Step 4	Step 5
Regress different dependent variables	Compare fitted values to actuals	Estimate the potential numbers	Compare to web traffic volumes	Identify the candidate cities

Table 2. Three Regression Models

Model	Dependent Variable	Independent Variables	Potential Market
1	Flight passengers	Income, population, distance, flight time, relative cost of flight, and availability of direct flight	Potential market for airlines
2	Hotel guests	Income, population, distance, flight time, and availability of direct flight	Potential market for the hospitality industry
3	Mobile devices	Income, population, distance, flight time, and availability of direct flight	Potential market for total visitor volume

Table 3. Variables and Data Sources

Variable	Data Source
MSA Population	U.S Census Bureau 2015 estimates
MSA Median Family Income	Federal Financial Institutions Examination Council (FFIEC) 2015 estimates
Nearest Airport	http://www.travelmath.com/nearest-airport/
Flight Distance	http://www.webflyer.com/travel/mileage_calculator/
Driving Distance	CDX zipstream
Flight Time	http://flights.google.com/
Number of Hotel Guests	Guests' zip code data from13 hotel management companies, in Charleston, SC, 2013-2015
Web Traffic	Average unique web sessions from CVB websites, 2013-2015, gleaned from Google Analytics
Mobile Devices	Four months' Sprint and Verizon mobile devices from AirSage Inc.

Table 4. Descriptive Statistics*

Variable	Min	Mean	Max
Air traffic (per year)	0	5,272	202,562
Mobile devices (per month)	2	1,442	44,492
Hotel guests (per year)	19	3,798	85,407
Income (thousands of dollars)	37.5	64.3	109.4
Population (10,000)	8.2	99.8	2018.2
Direct flight (0 if no direct flight, 1 if yes)	0	0.0485	1
Flight time (hh.mm)	0.55	4.54	11.45
Distance (100 mile)	0.96	9.62	36
Flight ticket price (dollars)	116	230	663
Web session (session per year)	162	19,613	393,457
Relative cost (flight ticket price/average driving distance)	0.0528	0.2843	1.3324

* *Slight differences per regression model.*

Table 5. Regression Result of Gravity Model of Three Regressions

Variable	**Flight passengers**				**Mobile devices**	**Hotel guests**
Model	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.0352***	0.0219***	0.0178***	0.0139***	0.0223***	0.0149***
Population	0.0019***	0.0014***	0.0010***	0.0010***	0.0013***	0.0010***
Distance	-0.0012***	0.1253***	0.0904***	0.0889***	-0.5255***	-0.1642***
Flight Time		-0.6949***	-0.5015***	-0.3908***	-0.6168***	-0.7282***
Distance*Flight Time		0.0052***	0.0030***	0.0044***	0.0492***	0.0249***
Relative cost			-2.2043***	-1.5554***		
Distance* Relative cost				-0.4885***		
Direct Flight			1.1115***	1.0843***	0.1355***	0.4615***
Constant	5.6791***	7.9856***	8.3630***	8.9778 ***	9.5030***	9.8146***
N	359	285	229	229	224	230
Pseudo R^2	0.6003	0.6995	0.7786	0.7849	0.7228	0.7248

*legend: * p<.1; ** p<.05; *** p<.01*

Table 6. Top Five Potential MSAs for Direct Flights

MSA	Population	Income	Driving Time	Current Flight time	Web traffic	Market potential*
Portland-South Portland, ME	526,295	73,200	16h48	3h35	8,282	537
York-Hanover, PA	442,867	70,300	10h10	2h55	6,608	55
Lancaster, PA	536,624	70,000	10h56	2h55	10,751	45
Evansville, IN-KY	315,693	64,800	10h30	3h25	5,676	40
Springfield, MA	631,982	67,300	14h25	3h15	8,385	37

** Market potential represents the expected increase in the number of flight passengers per week.*

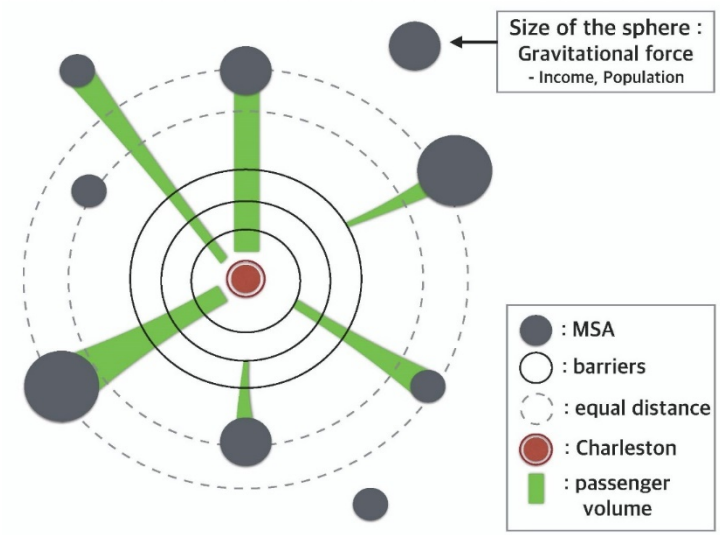

Figure 1. Markets of Origin and the Destination City

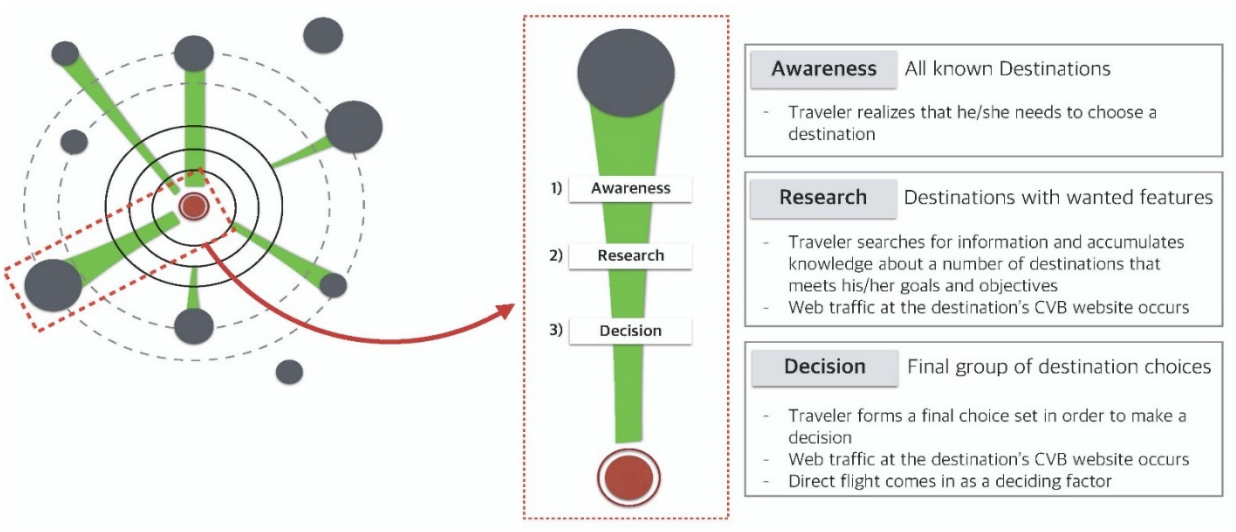

Figure 2. Buying Funnel and Gravity Model

Figure 3. Market Potential to Increase Flight Passengers versus Web Traffic

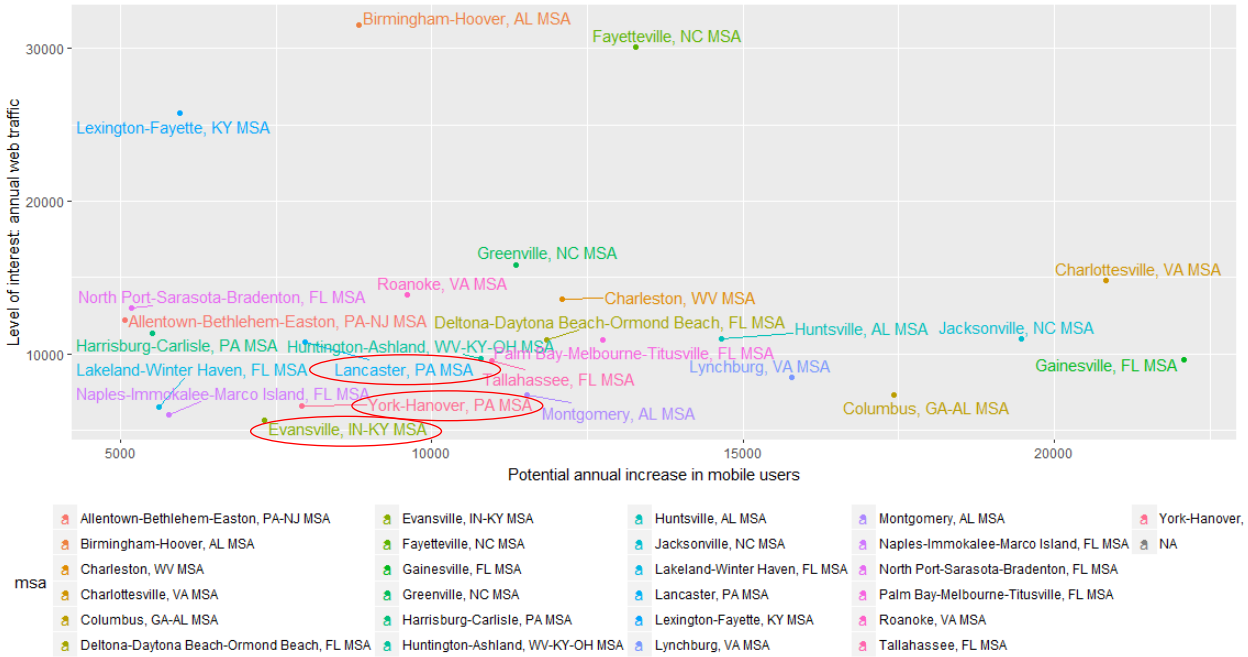

Figure 4. Market Potential to Increase Mobile Devices versus Web Traffic

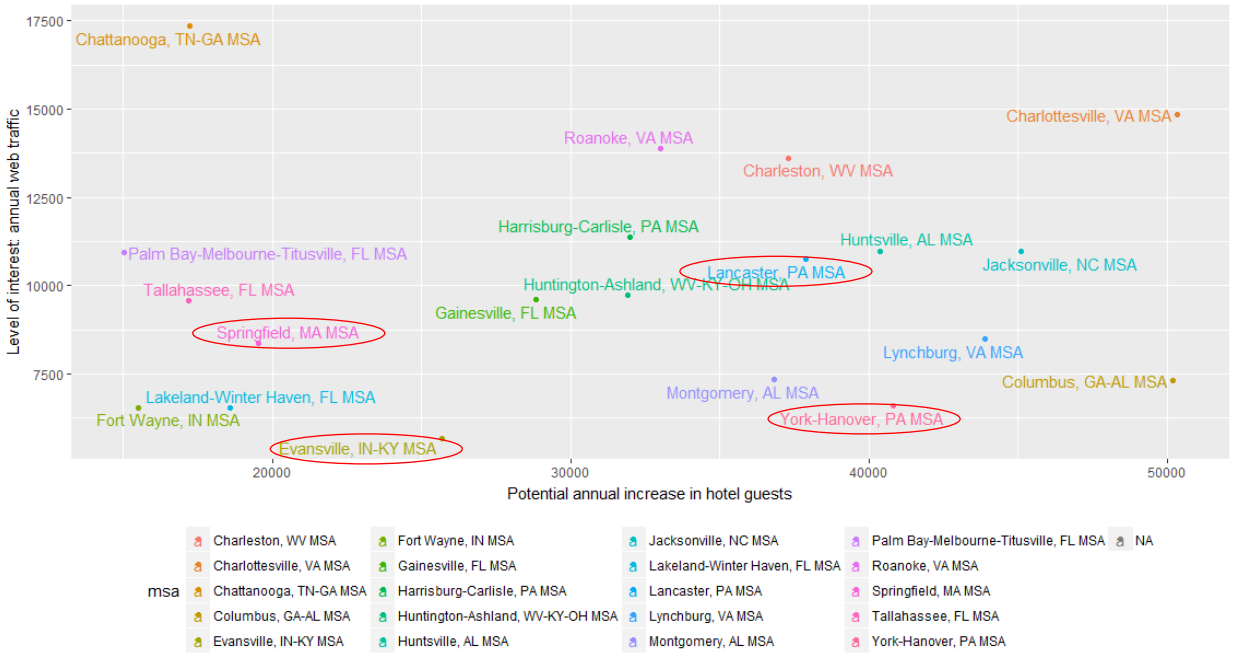

Figure 5. Market Potential to Increase Hotel Guests versus Web Traffic

Appendix A. Rationale for the use of general linear method (GLM) with Poisson distribution

Though the log-normal model is often used in gravity models, it tends to have serious disadvantages due to its strong assumptions of normality and homogeneous variances (Flowerdew and Aitkin, 1982; Silva and Tenreyro, 2006). First, the log-normal model transforms the data into the logarithm format, which can lead to under-estimation or biased estimation of the observed values (Silva & Tenreyro, 2006). The bias is more severe when there is heteroscedasticity, since the log-normal model assumes homogeneous variance. Second, there is a high possibility that the logarithmic form of positive-valued integers would not follow normal distribution (Flowerdew & Aitkin, 1982), violating one of the main assumptions. The third and most commonly known drawback is its innate inability to process zero values.

These problems with log-normal models are true not only for panel data, but also for cross-sectional data (Westerlund & Wilhelmsson, 2006). Hence, our dataset is vulnerable to such biased estimates. There are a few possible ways of processing zero values, such as adding a certain small number to zeros, or treating them as missing variables. Both of them can result in biased estimates. Many studies have also suggested using a generalized linear model with Poisson as its family distribution to solve heteroskedasticy and zero-value problems (Flowerdew & Aitkin, 1982; Silva & Tenreyro, 2006; Westerlund & Wilhelmsson, 2006).

Tourism researchers have incorporated the Poisson measure to analyze tourism demand and travel costs, often stating the method as the Count model (Creel & Loomis, 1990; Feather et al., 1994; Hellerstein, 1991; Parsons, 2003; Hellström, 2006). For example, Hellström (2006) employed a bivariate Poisson log-normal model to efficiently and accurately look at the household tourism demand.