ABSTRACT

PAGE, JONATHAN FOSTER. A Model for Deriving Likely Origin Markets for State Park Visitors. (Under the direction of Dr. Gene L. Brothers).

The purpose of this study is to develop a model which can be used in the future to determine likely origin markets for state park visitation in the United States. As state park systems do not typically have funding to conduct system-wide visitor surveys to obtain visitation data, it is imperative to be able to gain some understanding of state park visitors using available secondary data sources. This study uses ArcMap 9.1 spatial software to construct a 50x50 matrix of population origins and park destinations in the United States. The origins in the matrix are centered in each state based upon population densities. The destinations in the matrix are the exact center of each state based upon state park location densities. STATA 11 statistical software package was used to conduct a kmeans cluster analysis in order to develop a clustering of states which serve as the likely origin markets. Two particular weighted matrices are examined, a travel cost matrix and an expected utility matrix. A Calinski-Harabasz pseudo-F index score is used to determine the optimum number of clusters per weighted matrix. The expected utility matrix cluster analysis provides the optimum number and size of clusters to answer this question at 11 clusters with a Calinski-Harabasz pseudo-F score of 44.19. Recommendations are made on this model for future areas of research and other disciplines for which it may be applied.
A Model for Deriving Likely Origin Markets for State Park Visitors

by

Jonathan Foster Page

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Master of Science

Parks, Recreation, and Tourism Management

Raleigh, North Carolina

2010

APPROVED BY:

_______________________________                     _______________________________
Dr. Aram Attarian                     Dr. Chrystos Siderelis

_______________________________
Dr. Gene Brothers

Chair of Advisory Committee
DEDICATION

I would like to dedicate this thesis to all those in my life who had to put up with me for the years that it took for me to get here. Mom, Dad, Mike, Uncle Bill, Stella and Jerome.

A very special thank you to my most understanding and supportive wife Katie. I love you more than words on paper can possibly describe.
BIOGRAPHY

Jonathan Page was born in Raleigh, NC to Bob and Sally Page. After graduating from Sanderson High School he attended NC State University where he earned his B.S. in Parks, Recreation, and Tourism Management with a concentration in Tourism Management. After a couple of years in the workforce he decided to continue his education and re-enrolled at NC State University to earn a Masters of Science degree.
ACKNOWLEDGEMENTS

This thesis would have never been completed without all of the hard work of those who helped me along the way. Drs. Gene Brothers and Chris Siderelis were instrumental to getting this paper completed. I would also like to thank Dr. Aram Attarian for serving on my committee and being a great undergraduate advisor to me as well. Thank you so much for all of your help in answering all of my many questions Anju Singh. I am quite sure if I never bother you again you will be quite pleased. A special thank you goes to Mr. Jeff Essic for all of his help with the ArcMap software that was required to complete this study.
# TABLE OF CONTENTS

LIST OF TABLES................................................................................................. vii
LIST OF FIGURES............................................................................................. viii

CHAPTER 1
INTRODUCTION................................................................................................. 1
   STUDY OBJECTIVE....................................................................................... 3
   ORGANIZATION OF THESIS...................................................................... 3

CHAPTER 2
REVIEW OF LITERATURE.............................................................................. 5
   INTRODUCTION............................................................................................ 5
   TOURISM DEMAND.................................................................................... 6
      Introduction............................................................................................ 6
      Modeling............................................................................................... 8
      Time-Series Modeling........................................................................... 9
      Econometric Modeling......................................................................... 12
      Seasonality Analysis............................................................................ 14
      Conclusions......................................................................................... 15
   TRAVEL COST MODELING........................................................................ 16
      Introduction............................................................................................ 16
      Individual Travel Cost Model............................................................... 17
      Zonal Travel Cost Model..................................................................... 21
      Gravity Model...................................................................................... 24
      Conclusions......................................................................................... 28
   SUMMARY.................................................................................................. 29

CHAPTER 3
METHODOLOGY............................................................................................. 31
   INTRODUCTION............................................................................................ 31
   DATA COLLECTION..................................................................................... 31
      Origins and Destinations..................................................................... 31
   DATA ANALYSIS....................................................................................... 33
      Travel Cost and Expected Utility......................................................... 33
      Cluster Analysis................................................................................... 36
   SUMMARY.................................................................................................. 37

CHAPTER 4
RESULTS.......................................................................................................... 38
   INTRODUCTION............................................................................................ 38
ARCMAP RESULTS

CLUSTER ANALYSIS

SUMMARY

CHAPTER 5
CONCLUSION

INTRODUCTION

METHODOLOGY

RESULTS AND RECOMMENDATIONS

REFERENCES
LIST OF TABLES

Table 1. Travel Cost Matrix Calinski/Harabasz pseudo-F Index Scores .................. 43
Table 2. Travel Cost Matrix: 6 Clusters of State Origin Markets
        for Visitation to State Park Systems............................................................. 43
Table 3. Expected Utility Matrix Calinski/Harabasz pseudo-F Index Scores.............. 44
Table 4. Expected Utility Matrix: 11 Clusters of State Origin Markets
        for Visitation to State Park Systems............................................................. 45
LIST OF FIGURES

**Figure 1.** State Population Density Centroids and State Park Density Centroids ................................................................. 40

**Figure 2.** South Region State Population Density Centroids and State Park Density Centroids .......................................................................................................................... 41
Chapter 1

INTRODUCTION

A common axiom across most business entities is that one must know their customer base in order to serve them best. An example of this would be a travel agent working in Florida trying to sell trips to the Bahamas for those citizens living in Florida versus selling trips to Branson, Missouri. Those are two distinct areas with distinct tourist types, and having an accurate knowledge of the two customer bases is critical to success in selling tours to each location. If a manager of a tourism operation knows that a major disaster has impacted an area which contains the majority of its customer base, it is reasonable for that manager to assume that their business will also be impacted. A resort manager would expect decreased visitation from an economically impacted area of its customer base and make managerial planning decisions accordingly.

The same holds for government agencies. As a government entity, state parks operate in a slightly different manner than when compared to a normal profit generating business. For profit businesses use a marginal cost pricing system to buy low and sell as high as the market will bear for their particular product. In contrast, state park systems do not operate to make profit for the state. Some park budgets are funded only through annual transfer payments from the state legislatures, while others must use the entry fees from users in order to make up for the shortfall in their operating budget (Gladwell, Anderson, & Sellers, 2003). But having an understanding of their customer base is still
important in making their system-wide managerial decisions. In offsetting the inevitable shortfalls in payment amounts, knowledge about the customer base and visitor travel behavior in order to increase revenue streams is of great consequence. Unfortunately, most state park system managers only possess limited information about their visitors (McCool & Reilly, 1993). There are cases where surveys have been used to collect visitor information; however these are only valuable for understanding an individual park’s visitors, and not for an aggregate system-wide analysis, either on a national or a state level. Usually, performing a visitor survey of adequate size and scope to cover an entire state park system is cost prohibitive for most state park systems, so the majority of the data available to state park managers comes from secondary data sources. Since a visitation database does not exist, there is currently no method for determining where the visitor base for a state park system originates. This study addresses that issue by identifying origin markets for state park system visitors through the model it will employ.

Results from this study can be applied in a variety of disciplines, but the direct results can be used by state park system managers to determine their most likely visitor origin market states. These states can then be used for making targeted marketing campaigns, budgeting decisions, as well as many other applications. This study also contributes to the overall body of knowledge concerning travel and tourism demand modeling and forecasting techniques.
STUDY OBJECTIVE

The objective of this study is to determine a natural set of states, grouped on the travel cost, and the expected utility of the average state park system to the average state resident for all 50 states. This is done through the use of a spatial analysis which addresses the likelihood of state park visits from population origins to state park destinations. A matrix consisting of the origins or mean center of each state, based on population density, and destinations or mean center of each state’s park locations will be constructed to help in this analysis.

The main purpose of this study is to develop a method for deriving the likely origin markets for state park visitation. The research question for this study is to determine whether or not there are distinct groupings of destination state park systems for visitors from the 50 state origins as determined by the Calinski-Harabasz stopping rule in cluster analysis (Calinski & Harabasz, 1974). Results from this analysis can also be used as a template for future research on other resource management agencies which serve visitors. In addition, managers of tourism destinations or regions can use this modeling technique (in the absence of visitor studies) for developing more accurate profiles of likely visitors and to enhance management decisions based on the results.

ORGANIZATION OF THESIS

This thesis is organized into five chapters beginning with this introduction. The introduction chapter gives a background to the purpose of the research, an overview of the
objective, purpose, and research question of this study, as well as a brief introduction into the methods that were used. The second chapter reviews the scholarly literature and describes studies on a combination of topics revolving around tourism crises, tourism demand, and travel cost modeling. In the third chapter the study methodology and procedures are described, and data collection techniques, data analysis procedures and methods are explained. The fourth chapter consists of a discussion of results and findings with suggestions for future applications. The final chapter discusses implications and recommendations based on the results and findings.
CHAPTER 2
REVIEW OF LITERATURE

INTRODUCTION

This chapter consists of a review of the relevant literature including the theory as well as the research, methods and findings of the topics of tourism demand and travel cost modeling as they relate to origin markets. This study measures expected utility and travel costs, but as these travel cost models have their roots in demand modeling it is important to note the many different demand modeling techniques as well as travel cost modeling. As the purpose of this study is to develop a method for deriving the likely origin markets for state park visitation, each of these areas of research will serve as the foundation for the concept of the instrument’s design. It is important to gain an understanding of the different techniques that are currently being used tourism demand and travel cost modeling so that the theory behind this model can be fully appreciated. Without this, the model will fall short of its intended application.

Areas of consideration in the application of this model are not just limited to tourism demand and travel cost modeling. A major example for the future use of this model applies to making management decisions when a major origin market has been impacted by a crisis. Impacts result from such events as a major natural disaster (Huang & Min, 2002), to an economic recession (Law, 2001 Papatheodorou, Rosello, & Xiao, 2010; Ritchie, Molinar, & Frechtling, 2010; Sheldon & Dwyer, 2010; Smeral, 2010), to terrorism attacks (Du Preez & Witt, 2003).
TOURISM DEMAND

*Introduction*

Tourism demand has been an area of study due to its importance when making management decisions (Min, 2005; Plog, 2006; Smeral, 1988). This particular study does not measure demand, but the travel cost models upon which this study is based upon have their roots in demand modeling so it is important to note the many different techniques being used today in this discipline. The demand by visitors for any outdoor recreation site is a key concern to managers for a variety of reasons. Some of these areas of concern include but are not limited to: environmental protection and stewardship, staffing, infrastructure improvements and maintenance, use fees, and other budgetary decisions (Emmert, 1999; Eugenio-Martin, Sinclair, & Yeoman, 2005; Fleming & Cook, 2008; Zawacki, Marinsko, & Bowker, 2000).

From a strictly environmental standpoint, knowing how many visitors expected to visit a site in a given season or year is extremely important in being able to make a variety of policy decisions ranging from restricting access to sensitive areas or allowing full access to environmentally resilient areas, to trash collection. “Resource managers are often faced with the problem of trying to manage vulnerable heavily visited recreational sites” (Fleming & Cook, 2008, p. 1197). For example, knowing which parks receive the most visitor use will inform managers which parks need the most resources to help control the impacts to the site. Keeping the environmental impacts on state parks to an acceptable minimum is of the utmost importance if managers hope to keep these parks available to future generations (Fleming & Cook, 2008).
Being able to plan for the number of visitors to a park may also help managers in staffing decisions. It stands to reason that the more people that visit a site in a given time period, the more staff will be needed to keep the park operating. This can range from educational programs, to policing protected wildlife areas. It all depends on what attractions the park contains, but knowing how many visitors can be expected will help keep the staff of the park from being overwhelmed.

Maintenance needs on an annual basis can range from periodic maintenance to major renovations. For instance, some remote areas of Alaska provide rustic shelters which need maintenance perhaps only once every couple of years. But there are also park facilities which require major annual upkeep. All state parks require annual maintenance and upkeep of their facilities, but some parks will need more upkeep than others. This can include buildings as well as roads and equipment. The number of visitors using a park in a given year is just one example of the many factors that contribute to facility degradation. Social and cultural issues, geography and topography, weather, soil and vegetation are other factors that impact facility degradation, but many of these are much harder to forecast. Knowing the number visitors a park will see in a given year is just one tool that park managers can use when trying to determine annual upkeep budgetary concerns, as more visitors means more fees revenue, which can help subsidize any budgetary shortfalls.

Most of the literature on tourism demand focuses on forecasting and the modeling attributed to making these forecasts (Box & Jenkins, 1970; Chan, Lim, & McAleer, 2005; Cho, 2001; Coshall, 2005). During the past twenty years what was once a collection of a
few academic journals on tourism has ballooned to more than 70 different journals on the subject (Song & Li, 2008). With this increased interest in tourism research, tourism demand forecasting and modeling has also seen its share of increased interest. In fact, a review by Li, Song, and Witt (2005) found 420 studies published on the subject from 1960 to 2002 (Song & Li, 2008). Another review by Song and Li (2008) found 119 articles published during 2000-2006. This suggests the importance placed on being able to make accurate forecasts of tourism demand. Unfortunately, one of the conclusions drawn from this review is that the performance of these models varies greatly and therefore there has been no real solution to determining the best demand model.

**Modeling**

Tourism demand modeling is used in a wide variety of settings. Because of the diverse research objectives set forth for each individual study, the independent variables for demand modeling are quite extensive and vary from demographic data to something as arbitrary as what size shoes one wears. However, the dependent variable for the majority of these studies has been much less widely varied. Song and Li point out that in their analysis of articles published from 2000-2006 that “the tourist arrivals variable is still the most popular measure of tourism demand” (Song & Li, 2008, p.204). The authors also note that the tourist arrival variable could be more specific, for example; holiday tourist arrivals; business tourist arrivals; tourist arrivals for visiting friends and relatives (VFR) purposes (Turner & Witt, 2001a and 2001b; Kulendran & Wong, 2005); and tourist arrivals by air (Coshall, 2005; Rossello (2001). In addition, there were other variables used to measure
Expenditures at the tourism destination has been used as the demand variable in several studies, including Li, Song, and Witt (2004 and 2006), as well as Li, Wong, Song, and Witt (2006). Other studies considered tourism spending on specific product categories like food expenditures, shopping, and sightseeing (Au & Law, 2002). The literature also contains other tourism demand variables including tourism revenues (Akal, 2004), tourism employment (Witt, Song, & Wanhill, 2004), and tourism import and export (Smeral, 2004).

Almost as varied as the dependent variables in a demand model are the actual model frameworks themselves. The bulk of the published studies used quantitative methods to forecast tourism demand (Song & Turner, 2006). Of these quantitative methods, non-causal time-series models and causal econometric approaches are the two main models noted in the literature (Song & Li, 2008). The main variation in these two types of models is whether there is any recognized causal relationship between the tourism demand variable and its manipulating factors. Out of the 121 post-2000 studies examined by Song and Li (2008), 72 employed the time-series methods to model tourism demand. Of those 72 studies, 68 made either ex post (after the fact) forecasts or ex ante (beforehand) forecasts and the remaining four made no forecast. From the 121 original studies, 71 produced a variety of different econometric models. Of those, 30 focused on the identification of the connection between tourism demand and its influencing factors while 41 assessed the forecasting performance of the econometric models in addition to the identification of the causal relationships (Song and Li, 2008). Of the 71 econometric studies, 30 of them applied the time-series and econometric methods in determining the demand
models and compared the forecasting performance of these models.

**Time-Series Modeling**

A time-series model uses a variable’s own past history to make forecasting more accurate. It can be used to determine how a past influence on a variable will affect the variable in the future should the same influence/disturbance happen again. “Naïve 1 (or no-change), Naïve 2 (or constant-growth-rate), exponential smoothing models, and simple autoregressive models have also appeared frequently in the post-2000 studies, but as in earlier tourism forecasting studies, they are usually used as benchmarks for forecasting accuracy evaluation” (Song & Li 2008, p.210). Auto-Regressive Integrated Moving-Average (ARIMA) models designed by Box and Jenkins (1970) have been widely used to forecast tourism demand. Different variations of the ARIMA have been used in the recent literature such as the Simple ARIMA or Seasonal ARIMA (SARIMA). With seasonality being a prevailing force in the tourism industry the Seasonal ARIMA has been gaining in popularity (Alleyne, 2006; Du Preez & Witt, 2003; Goh & Law, 2002; Kim & Moosa, 2005). Unfortunately, there have been many contradictory studies as to the effectiveness of the forecasting results from these models (Cho, 2001; Goh & Law, 2002; Smeral & Wuger, 2005).

In a 2001 study, Cho investigated three time-series models (exponential smoothing, univariate ARIMA, and an adjusted ARIMA using influential economic indicators that are highly correlated with travel demand) to predict travel demand (# of arrivals) from 6 different countries to Hong Kong (the United States, the United Kingdom, Singapore, Japan, Taiwan, and Korea). He found that the univariate ARIMA model worked best for forecasting
demand from Japan, while the adjusted ARIMA model worked best for the United States and United Kingdom. Both the univariate ARIMA and adjusted ARIMA behaved similarly for Singapore, Taiwan, and Korea, thus concluding that the ARIMA model in either form was the better forecasting method vs. exponential smoothing.

However, the SARIMA models outperformed eight other time-series models when modeling demand for tourist arrivals to Hong Kong using ten different arrival series (Goh & Law, 2002). Another study found that neither the ARIMA nor the SARIMA could outperform the Naïve 1 (no-change) model when attempting to forecast seasonal and calendar effects as well as unknown special effects on Austrian tourist arrivals (Smeral & Wüger, 2005). In an effort to combat this inconsistency, there have been recent attempts to improve the forecasting performance of the ARIMA-based models by using alternative time-series approaches. Goh and Law (2002) proposed a multivariate SARIMA (MARIMA) model which included an intervention function to capture the potential spill-over effects of the corresponding demand series on a particular tourism demand series. Their study showed that the MARIMA model significantly improved the forecasting performance of the simple SARIMA as well as other univariate time-series models. However, Gustavsson and Nordström (2001), in a similar previous study found that their MARIMA model could not outperform its univariate equivalent.

Another version of the univariate time-series analysis, the Generalised Autoregressive Conditional Heteroskedastic (GARCH) model, has been used in the financial modeling world to examine the instability of time-series models. A study done by Chan,
Lim, and McAleer (2005) used three multivariate GARCH models (the symmetric CCC-MGARCH, symmetric vector ARMA-GARCH, and the asymmetric vector ARMA-GARCH) to analyze the conditional mean and conditional variance of the logarithm of the monthly tourist arrival rate from the four leading tourist source countries to Australia (Japan, New Zealand, the UK, and the USA). “This will assist in evaluating the impact of tourism demand fluctuations and in working together closely for the successful management of the tourism industry” (Chan et al., 2005, p.470). The results of this study found possible interdependent effects in the conditional variances between the four source markets, and asymmetric effects of shocks in two of the four countries which are important as it emphasizes interdependencies between major tourism source markets (Chan et al., 2005).

**Econometric Modeling**

One of the major advantages of the econometric approaches over the time-series models lies in their ability to analyze the causal relationships between the dependent tourism variable and its independent variable. “Recent econometric studies of tourism demand have shown that tourists’ income, tourism prices in a destination relative to those in the origin country, tourism prices in the competing destinations, and the exchange rates are the most important determinants of tourism demand” (Song & Li 2008, p.210). This quote shows the importance being placed on the study of origin markets when modeling demand and making forecasts. Being able to estimate the effects of these relationships effectively is something that managers in the tourism industry strive to attain.

Some of the main econometric forecasting methods arising today are the
Autoregressive Distributed Lag Model (ADLM), the Vector Autoregressive Model (VAR), the Error Correction Model (ECM), and the Time Varying Parameter Models (TVP) (De Mello & Fortuna, 2005; Lim & McAleer, 2001; Kulendran & Witt, 2003; Song & Wong, 2003). Aside from the VAR model, these models use a single-equation modeling approach and assume that the independent variables are exogenous. The VAR model assumes all variables are endogenous and each variable is stated as a linear relationship of the others. The VAR model was used in Song and Witt (2006), Witt et al. (2004), as well as Shan and Wilson (2001). The ECM model was utilized in studies by Lim and McAleer (2001) to analyze tourism demand for Australia by Hong Kong and Singapore, Kulendran and Witt (2003) to model demand for international business tourism, and Kulendran and Wilson (2000) to model business travel. Song, Wong, and Chon (2003) in addition to Song, Witt, and Jensen (2003) utilized the ADLM in their studies. Li, Wong et al. (2006) have also fused the TVP model with the ECM to create a more sophisticated TVP-ECM (Song & Li, 2008).

One model that has seen an increase in use today is the Almost Ideal Demand System (AIDS). This model was developed by Deaton and Muellbauer (1980), and is a system-of-equations approach. This method uses tourism expenditure shares as the dependent variable and is usually used to examine tourism demand in a number of adjacent destinations to a source market. While only five studies used the static AIDS model before 2000 (Fujii, Khaled, & Mark 1985; O’Hagan & Harrison 1984; Papatheodorou 1999; Syriopoulos & Sinclair 1993; and White 1985), there have been several studies done recently that have applied and further developed this method. Durbarry and Sinclair (2003)
used the AIDS method to examine the magnitudes and determinants of changes in destinations’ shares of a major tourist origin market. Specifically, they used this method to quantify the responsiveness of French tourism demand in Italy, Spain, and the UK to changes in relative prices, exchange rates, tourists’ expenditure budget, and external events. What they found was that effective price competitiveness is a key variable driving changes in market shares. Li et al. (2004), De Mello and Fortuna (2005), as well as Mangion, Durbarry, and Sinclair (2005) all used a linear AIDS model (LAIDS) and combined it with an ECM. In all, eleven studies have used different versions of the AIDS model since 2000.

**Seasonality Analysis**

Being able to account for seasonality in tourism is incredibly beneficial to park managers. As most state parks are in climates with distinct seasons, these weather patterns can influence daily/monthly visitation rates greatly. Annual data counts and averages still seem to be the predominant rate at which study data are analyzed, due mainly to the fact that most of the independent variable data are collected at yearly intervals. However, this can be misleading and not show the volatility with which park visitors make their impacts on the park. “In the Tourism demand forecasting literature, seasonality is often treated as a deterministic component or a stochastic component in the time-series” (Song & Li 2008, p.215). Determining how to treat seasonality has been addressed differently in several studies. The HEGY test developed by Hylleberg, Engle, Granger, and Yoo (1990) is widely used to test for seasonal unit roots if using the time-series method. But, whether or not this is the more accurate method of determining seasonality has not been demonstrated.
Kim and Moosa (2001) found that the HEGY test is quite lacking when sample size is small. They developed an alternative called the Caner Test, but that also produces contradictory results. For example, according to the Caner test, stochastic seasonality is appropriate in most cases of their Australian tourist arrivals series. But in an ensuing forecasting comparison exercise, Kim and Moosa found that the stochastic treatment of seasonality did not improve the forecast accuracy. The most accurate forecasts of seasonality they found resulted from a time-series model in which the seasonality was treated as a deterministic component (Kim & Moosa, 2001).

Gustavsson and Nordström (2001) suggested that instead of using the unit root test, a more elastic model structure should be used that can handle the changing seasonal patterns and trends (Song & Li, 2008). One other approach to modeling seasonality is the Periodic Autoregressive (PAR) model. This model lets parameters vary according to the seasons of the year and can therefore be more accurate in considering the effects of seasonal decision making (Osborn & Smith, 1989).

**Conclusions**

Being able to measure demand and forecast future trends for tourism destinations is critical for managers for a variety of reasons. As such, knowing what the main origin markets are can show managers the areas they need to focus on when it comes to analyzing what is going on in those markets that can affect the visitation they may receive in the near future. Just as varied as the possible applications of accurate forecasts for tourism managers are the models in which these forecasts are generated. While each have their
particular strengths and weaknesses, there is no consensus on which method is the most accurate from a general standpoint. For the purposes of this study, these studies will serve as the foundation for the modeling techniques we will use as travel cost modeling is based upon these demand modeling methods.

TRAVEL COST MODELLING

Introduction

Travel cost modeling has been a valuable method for analyzing tourism demand for the last several decades. This method was first presented by Hotelling (1949) as a zonal model and later refined by Clawson and Knetsch (1966). The dependent variables in the early travel cost models were the number of trips originating from a zone, divided by the population of that zone (i.e. trips per capita). Unfortunately, the data available for estimating demand functions for outdoor recreation facilities has historically been rather poor, as most studies have been confined to limited surveys or secondary data sources that were incapable of providing the data required for the specification and estimation of a complete travel cost demand model (Smith & Kopp, 1980). There are essentially two different types of travel cost model; the Individual model, where the dependent variable is trips per year (or season) by individual users of a recreation site, and the Zonal, where the dependent variable is the number of trips taken to the site by the population of a particular region or zone (Fleming & Cook, 2008). It is important to consider for these studies that as
the travel cost models depend not only the implicit price of the travel cost for a visitor to travel to their own state parks, they also depend on the implicit prices of the remaining substitutes.

The Individual Travel Cost Model (ITCM) uses survey data from individual visitors to link the demand for tourism to its determinants. These determinants include how far the visitor travels to get to the site, the amount of time spent on-site, travel and on-site expenses, how often they have visited the destination in the past, as well as income and other socioeconomic characteristics (Ha, 2007). “The advantages of the ITCM are that it 1) imitates the conventional methods used by economists to estimate economic values based on market prices; and 2) it is based on what people actually do rather than on what people say they would do in a hypothetical situation” (Ha, 2007, p. 53). Since state park systems do not have comprehensive survey data to use, this model is not the right fit, but it is important to detail its benefits and contributions to the current body of literature.

The Zonal Travel Cost Model (ZTCM) is useful in situations where visitor data is from secondary data sources, for instance recreation permits or fee receipts (Poor & Smith, 2004). There are essentially two methods in the ZTCM, the first of which is Clawson and Knetsch’s (1966) two-step method and the gravity model. In the two-step model the first step creates a visitor demand function, and the second uses this function to derive a total demand curve for the site and illustrates the estimated total visits from all zones. “The gravity model framework represents a departure from the predominant approach taken in the existing literature by examining income in the destination as a determinant of tourism
demand” (Archibald et al., 2008, p. 7). The major limitation seen of the ZTCM is that by aggregating the data by zones, some of the individual data variation is averaged out (Loomis et al., 2009). Since this study is taking an aggregated approach rather than an individualized approach, this is not seen as a limitation.

**Individual Travel Cost Model**

In the ITCM, the dependent variable is the number of trips a visitor makes during each season or year, with each visitor being the unit of observation. This method is currently becoming the most popular in literature, but obtaining individual survey data can be quite cost prohibitive for those entities working under a tight budget so, as with state park systems, the ITCM is not always a viable option. For those who can use this method, the results can be quite invaluable.

A study by Heberling and Templeton (2009) used an ITCM to estimate the value of a recreational trip to Great Sand Dunes National Park and Preserve. The data were collected by the Visitor Services Project to do their analysis and followed the standard approach for estimating the ITCM. This particular model is defined as:

\[
\text{Trips}_i = \int (s_i; \beta) + \varepsilon_i,
\]

Where: \(\text{Trips}_i\) is the number of observed trips that the \(i\)th visitor would take to the national park in a specific time period; \(s_i\) is the vector of explanatory variables for the \(i\)th visitor including travel costs to the site, income, age, type of trip, group size, travel costs to substitute sites, etc.; \(\beta\) is a vector of unknown parameters; and \(\varepsilon_i\) is the error term.”
(Heberling & Templeton 2009, p.621). They chose not to include the opportunity cost of time in the travel cost nor as a separate variable in the model. This approach was taken because of multicollinearity and the difficulty of determining the mode of transportation to get to the park. They used three different count data models to test their findings. The Poisson model, a standard negative binomial model where $\alpha$ does not vary with the visitors’ characteristics, and a generalized negative binomial approach where $\alpha$ is a function of demographic variables. Findings suggested that the estimate of annual consumer surplus per visitor for Great Sand Dunes National Park and Preserve as the primary destination to be $89.00, and the consumer surplus per year related to multi-destination trips and unplanned trips to be $256 and $238 respectively. Very few valuation studies have been done that examine US National Parks, which makes this particular study a great contribution to the existing body of literature.

Another study used the individual travel cost model to estimate a recreation demand function to estimate the economic value of ecosystems for tourist areas in northwest Florida, specifically beaches (Ha, 2007). The demand curve was used to estimate the consumer surplus for the beach amenities. Data were collected from the four visitor information centers in the greater Pensacola area of northwest Florida between September 1999 and April 2002. The survey found that vacation trips had slightly decreased from 67.2% to 62.6% between the pre-and post-9/11 attack time periods. It also found that trips by air decreased significantly from 12.1% to 10.1%. Visitors stayed less nights (5.28 vs. 4.99) and spent less ($203.55 vs. $190.68). It was also determined that visitors arriving by air are
more likely to stay longer in the area. Accordingly, the data also show that the greater distance travelled the longer the stay. The author found that total per day spending was a significant factor in length of stay. Having a negative coefficient suggests that the higher the daily costs, the shorter the visit. Interestingly enough, annual gross income was not a significant factor in determining length of stay. By using the demand equation data from the survey, this study was able to estimate a consumer surplus of $69.90 per day per visitor when the average travel group size is 3.05. By the same token, it was determined that the consumer surplus for a business traveler to be $78.58 and $49.84 for those visiting friends and relatives.

State parks are the primary topic for this study. The travel cost model was used to examine Indiana’s Prophetstown State Park is of particular interest. The study addressed three specific questions, 1) how strong is the effect of distance traveled on visitation at state parks; 2) does the higher opportunity cost of time for higher income individuals deter those individuals from visiting parks; and 3) when Prophetstown State Park is completed, will visitors decrease visitation to other parks in order to visit Prophetstown (Emmert, 1999)? The study used a series of linear and log-linear regressions to measure the effects of distance and other factors on the reported number of visits. The number of annual visits reported by respondent served as the dependent variable. Four sets of independent variables were used in the analysis. These variables were round-trip distance, household income, an income and distance interaction term, and a measure of opportunity cost. The functional form of this model is defined as:
Visits = \int (\text{distance}, \text{income}, \text{opportunity cost})

Where: short and long regressions are estimated in linear and logarithmic forms. Survey results found that the average number of visits was 1.7. Hiking was the most common reason for a visit to the park (69%), and the average opportunity cost of travel in the sample was approximately $468. Results from the linear and log-linear regressions were significant for several variables. A significant negative relationship was found between distance travelled and number of visits, holding true to the notion that as distance increases the number of visits will decrease. “When distance is the sole explanatory variable, the elasticity of demand is -0.3 in the linear regression and -0.28 in the log-linear regression. This means a 1% increase in travel distance results in a 0.30% reduction in annual number of visits” (Emmert, 1999, p.1334). Opportunity cost was also found to be negative and significant at a 95% confidence level demonstrating that as opportunity cost increases, visits will decline. An interesting result noted that as income increases demand for recreation at state parks decreases, suggesting that recreation at state parks may be an inferior good.

**Zonal Travel Cost Model**

The zonal model approach is very useful in situations where visitor data is from secondary sources. This is of particular importance to an analysis of state park demand because managers rarely have the resources to undertake an exhaustive study of visitors to their parks, so being able to use available secondary data to answer travel cost demand questions is very important.

Smith and Kopp (1980) found that there are spatial limits to the zonal travel cost
model. “The spatial limits to the travel cost model result from the assumptions necessary to use the available secondary data to estimate the representative individual’s demand for a given recreational site’s services. The most important of these assumptions involves: (a) the objective of the trip to the recreational site; (b) the amount of time spent on the site during each trip; and (c) the mode of travel and corresponding travel costs required to reach the site” (Smith & Kopp 1980, p. 64). They found that the first assumption was the most important, because demand for a destination can vary greatly from a unique natural setting to a simple urban park. Through their test application to the Ventana Wilderness Area, they were able to determine that there is a significant change in demand after a distance of 672 miles from the site. Now while this is for one specific area with some unique natural qualities, the findings of this study can be applied in future travel cost demand model studies. This study is of particular importance to a national analysis of state park systems as travel will include distances from one side of the country to the other.

Poor and Smith (2004), used the ZTCM to estimate the consumer surplus welfare measure of a cultural heritage site. Specifically, the study looked at St. Mary’s City located in southern Maryland. Because of its uniqueness and remote location the authors felt that this would be the method best suited to perform their analysis. They used zip codes as their zones and included household income, ethnicity, and average age as their demographic variables. They used three different functional forms to estimate the visitor demand to St. Mary’s City (Linear, Semi-Log, and Log-Log). They found that the semi-log form gave the best fit for their data. This particular functional form has advantages in reducing
heteroskedasticity, as well as eliminating negative trip prediction. They found that the quantity of visitors per 1000 zonal population was negatively related with travel cost (as one would expect). They also found that there was a negative relationship between income and the quantity of visitors per 1000 zonal population which was contradictory to previous cultural heritage site valuation studies. They also found that depending on the functional form of demand used, the average consumer surplus measures ranged greatly from $8.00 to $19.26. These results show how data can vary from one functional form to another, but using goodness-of-fit criteria the log-log form was supported.

Loomis et al. (2009) used a hybrid individual-zonal travel cost model to estimate the price elasticity and consumer surplus for golfing in Colorado. They followed the lead of Brown, Sorhus, Chou-Yang, & Richards (1983) who first suggested the hybrid individual-zonal model. “This model uses the individual visitor as a unit of observation and allows maintaining the individual travel cost, travel time, and demographic variables on the right hand side. However, the dependent variable is calculated by dividing that individual’s visits by his or her share of the of the zones’ population to calculate trips per capita” (Loomis et al., 2009, p. 158). This method takes into account for the likelihood that closer zones will not only have visitors with a higher trip frequency per visitor, but also have higher participation rates from their population. The basic equation for this model is listed as:

\[
\text{IndTrips}_{ij} / (\text{Pop}_z/\text{V}_z) = B_0 - B_1 (\text{TC}_z) - B_2 (\text{Cost of Travel Time}_z) + B_3 (\text{Age}_i) + B_4 (\text{Gender}) + B_5 (\text{Fees}_{ij}) + B_6 (\text{Income}_i) + B_7 (\text{GCMountains})
\]
Where: IndTrips_{ij} is the number of trips by golfer $i$ living in zip code $z$ to golf course $j$. Pop$_z$ is the population of zip code $z$, and $V_{ij}$ is the number of golfers from zip code $z$ that visit course $j$. TC$_{ij}$ is the calculated round trip distance from the golfer’s residence zip code $z$ to the golf course $j$, multiplied by the variable automobile cost per mile. Cost of Travel Time$_{ij}$ is the full wage rate of golfer $i$ times the calculated travel time from the golfer’s residence zip code to the golf course. Fees$_{ij}$ is the cost of green fees and any cart fees the golfer $i$ reported they paid at course $j$. GCMountains is a dummy variable equal to 1 if the golf course is in the scenic Rocky Mountains and equal to 0 if it is on the Front Range or eastern plains of Colorado. They used both a semi-log and linear functional form to analyze their data.

The results of this study found that the demand for golf in Colorado is quite price inelastic with respect to both transportation costs and green fees. They found that the gross willingness to pay for a round of golf in Colorado comes to an average of $75. Automobile travel accounts for $8 of this and $49 is paid for green and cart fees, which leaves a consumer surplus of $18. Based on this data it was estimated that the total net economic value associated with the 7.8 million rounds of golf played annually is $143.8 million (Loomis et al., 2009).

**Gravity Model**

The gravity model is sometimes considered a more elegant version of the Clawson and Knetsch (1966) zonal travel cost model (Fleming & Cook, 2008). “In its most basic form the gravity equation upon which the gravity model is based is an economic variant of the Newtonian gravity equation and assumes that attraction between two countries depends
directly on their economic sizes and inversely on the distance between them” (Archibald, LaCorbiniere, & Moore, 2008, p.7). This model differs from the more popular approaches in existing literature in the way that it examines income in the destination as a determinant of tourism demand. This model also considers the positive correlation between income in the origin market and demand for tourism in the destination market. Since the analysis of state park demand considered distance as a key element to traveler’s willingness to travel, having an understanding of this model and its application is important.

A study by Khadaroo and Seetanah (2008) used a gravity model to evaluate the importance of transport infrastructure in determining the tourism attractiveness of destinations. They selected 28 countries (based upon data availability) and treated them as both origin and destination, with the data set later being “disaggregated into continent-wise sub-panels for a deeper comparative analysis” (Khadaroo & Seetanah, 2008, p. 838). This enabled the researchers to determine that infrastructure was a more sensitive factor when traveling to African or Asian destinations. Of particular importance to their analysis was the importance of transport capital in overall tourism attractiveness. The specific gravity model for this study was written as:

\[ TR_{odt} = \int (GDPO_{ot}, CPI_{dt}, DISTAN_{od}, TOURINF_{dt}, ROAD_{dt}, AIR_{dt}, PORT_{dt}, POPO_{ot}, LANG_{od}, BORD_{od}, PROX_{od}) \].

Where: \( o \) is used to index countries of origin, \( d \) to index countries of destination and \( t \) to index time. \( TR = \) tourist arrival per year, \( GDPO = \) income of origin country, \( CPI = \) Consumer
Price Index of destination country adjusted for the $ exchange rate, DISTAN = distance between the capital cities of origin and destination countries, TOURINF = # of hotel rooms available in the country, ROAD = length of paved road divided by the size of the country, AIR = the total number of international terminals in airports each country, PORT = the number of ports in each country, POP = size of population, LANG = dummy variable that takes the value of 1 if origin and destination country have a common first language and 0 if otherwise, BORD = dummy variable that takes a value of 1 if origin and destination country share a border and 0 if they do not, and PROX = dummy variable that takes a value of 1 if the destination country has a number of alternative destinations in proximity and 0 if it does not. Using a Generalized Method of Moments (GMM) estimator it was determined that transport capital carries a positive and significant coefficient. Airport infrastructure seemed to be a more important transport element which makes sense as international tourism is predominantly air based. The panel data regression resulted in an income elasticity of 0.8, which confirmed for the authors that the tourism product is not a necessity. They also found a price elasticity of -0.7 which showed that cost of living does matter and that tourists are sensitive to price level. The lagged tourist arrivals variable is positive and significant for all cases except for Africa which suggests repeated tourism around the world except for countries in Africa. The specification of this model was log linear and pair-wise correlation between the variables varies in the range of -0.23 to 0.61 suggesting that multicollinearity is not a serious issue. An important finding of this study was that “investing in tourism infrastructure, marketing efforts and liberalizing air access might not be enough
without efficient transportation support infrastructure” (Khadaroo & Seetanah, 2008, p. 838). Regardless of the improvements to a location’s sight specific infrastructure, if the access is limited due to poor transport infrastructure, then tourist levels will still suffer.

Archibald, LaCorbinière, and Moore (2008) presented a study at the 29th Annual Review Seminar for the Central Bank of Barbados on an analysis of tourism competitiveness in the Caribbean using a gravity model approach. The specific model that they employed can be written in log-linear form as:

\[
\ln T_{ijt} = \alpha_i + y_j + \lambda_t + \beta_1 X_{ijt} + \beta_2 X_{it} + \beta_3 X_{jt} + u_{ijt}
\]

“Where: \(i\) and \(j\) are the index of home and source country respectively, \(t\) is a time index, \(\alpha_i\), \(y_j\) and \(\lambda_t\) are local country, source country and time specific effects, \(X\) are explanatory variables that can vary over all three dimensions or just two and \(u_{ijt}\) is an error term which is assumed to have normal properties” (Archibald et al., 2008, p. 9).

The explanatory variables that were used in this regression model were home and source country income, the distance between the home and destination countries, prices at home and relative prices, the bilateral exchange rate of the home and destination country and the population size. The dependent variable chosen for this study is annual tourist arrivals from the US, Canada, Europe, and the Caribbean. Archibald et al. (2008) also chose to employ a dynamic panel data model like the one used by Khadaroo and Seetanah (2008). Their model is also estimated using a GMM estimator, which attempts to explicitly account for the correlation between the lagged regressor and the error term. This model revealed that the coefficients on both home (0.097) and destination (0.173) incomes were highly
significant and positive. “These findings suggest that destination attraction capacity seems to play a larger role in attracting visitors to the region relative to trip generation capacity and implies that as territories in the region experience economic growth, the accompanying increase in tourism competitiveness – through infrastructural development, for example – creates a greater incentive for tourists to travel to the Caribbean” (Archibald et al., 2008, p. 12). The coefficient estimates on the two relative price variables were negative as expected, with a 1% rise in destination prices relative to the source market reducing arrivals by about 0.1%, while a 1% rise in destination prices relative to competing destinations lowering arrivals by approximately 0.2%. This data represents a statistically significant level of destination substitution among the other Caribbean countries. The results on the exchange rate variables were also statistically significant, as the model found that for every 10% of depreciation in the bilateral exchange rate, arrivals were reduced by 1.2%. The population variable was found to be statistically insignificant. In all, it was concluded from this study that the main determinants of arrivals to the region were the destination income, source market income, relative destination-source market prices, relative destination-competitor prices, and exchange rate fluctuations. The importance of this study is that it was one of the first to examine the role that destination income, which is a key indicator of infrastructural development and other aspects of tourism competitiveness, has in determining tourist arrivals.

Conclusion

Travel cost modeling has become a very popular method of examining tourism
demand. The two predominant methods detailed here consist of the Individual Travel Cost Model and the Zonal Travel Cost Model which includes a variation of Gravity Modeling. Each method has its own advantages and weaknesses. The method chosen will vary depending upon what kind of data are available and the desired analysis. The state park demand analysis will be taking advantage of secondary data sources, and as it is specifically utilizing distance as a possible variable the Zonal Travel Cost Model is the most appropriate.

SUMMARY

The literature examined in the previous sections lays the theoretical framework for this study. There are a myriad of ways in which to measure the impacts of tourism on a destination or region. Which method one chooses to use is entirely based upon the situation and the data one has available to them. There is no one particular method that trumps all others as most of the current methods being used today are for the most part equally effective. The research question for this study, which is whether or not there are distinct groupings of destination state park systems for visitors from the 50 state origins as determined by the Calinski-Harabasz stopping rule in cluster analysis, has not been addressed in previous literature as it pertains to state parks, so it was important to detail the work that has been done to this point in the areas of tourism demand modeling, measurement, and forecasting. The more knowledge any manager has about his customer base, the better suited they will be to serve them. For this study, a cluster analysis based upon state park visitation data will be utilized, in addition to state population data and the
theories laid down in the ZTCM as well as from Gravity Modeling. The purpose of this cluster analysis is to establish viable “clusters” of states that can logically be identified as origin market areas.
Chapter 3

METHODS

INTRODUCTION

This chapter describes the research design for this model, data collection techniques, and the data analysis. Similar to Archibald et al. (2008), this study will be using secondary data sources to answer the research question. Since this will be done on a state level, the ZTCM (Loomis et al., 2009; Poor & Smith, 2004; Smith & Kopp, 1980) will serve as a theoretical foundation for the model construction. For the purposes of this study, it was necessary to construct different data sets to develop the theoretical model for determining likely origin markets for state park visitation. These will include a 50x50 matrix consisting of origin and destination points, a travel cost equation and an expected utility travel cost equation.

DATA COLLECTION

Origins and Destinations

The first step in constructing the model was to design the 50x50 matrix that contains the exact mean center of each state based on population which would serve as the start point. This matrix would also have the exact mean center of each state based on the locations of its state parks which would serve as an endpoint. The basic premise of these
points being that the matrix would detail the distance an average citizen of origin state \( i \) would have to travel to get to the average state park in destination state \( j \).

In order to construct the distance matrix, ArcMap 9.1 software was utilized. The starting points were easily obtained as the US Census Bureau already has determined the mean center of each state based upon the 2000 census data (US Census Bureau, 2000). These data were readily available on the US Census Bureau website. These points were entered into the map and set-up as a shapefile to become the starting points by using the Add XY coordinate tool. Unfortunately, the mean center of each state based on their state park densities was not available, so this had to be constructed for this specific purpose. This was accomplished by utilizing a data file from the GPS Data Team website (GPS Data Team, 2010), which contains a listing of all of the state parks in the US (minus Alaska) by their GPS coordinates (Alaska Department of Natural Resources, 2010). The Alaskan coordinates were collected independently by obtaining a complete listing of all state parks and recreation areas in the state from the Alaska Department of Natural Resources website. Google Earth was employed to locate the exact GPS coordinates for each park. Once the coordinates were collected they were entered into ArcMap as another shapefile, using the Add XY coordinates tool. In order to get these points classified by state it was necessary to perform a spatial join. This process basically involves laying a map of the US divided by state over the coordinates that were entered into ArcMap and based on the state borders. The state parks were assigned to the states in which they were located. The Mean Center Tool, found in the Measuring Geographic Distributions from the Spatial Statistics Toolbox.
was used to find the mean center of each state based on location of state parks (ArcMap 9.1, 2005).

In order to complete the matrix the distance from the state population centroids to the state park centroids needed to be found. To do this the Point Distance Tool from the Proximity Tool Box in the Analysis Tools was used. The state parks feature was used as the Input Feature Class and the State_FIPS ID as the Case Field. This tool took the 50 mean centers of state by population points and calculated their distance from each of the 50 mean centers of state by state park points in terms of miles (ArcMap 9.1, 2005).

**DATA ANALYSIS**

*Travel Cost and Expected Utility*

Once the 50x50 distance matrix is completed it can then be used to construct the weighted matrices for the cluster analysis. The first of these matrices to be completed is the travel cost weighted matrix. In order to complete this an average travel cost from one aggregate state population to an aggregate state park location needs to be calculated. The travel cost includes the round-trip cost of $0.60 per mile (based on the American Automobile Association fuel gauge report) plus the opportunity costs of travel time (AAA, 2010). The estimation of the opportunity cost of travel time is made up of the following steps. The travel time equaled (a) two times the distance between the state population origin and destination park system divided by 50 miles per hour; (b) multiplied by the hourly wage (derived by taking the median annual income of origin state residents for 2009 divided
by the 2,010 annual work hours); (c) multiplied by the constant of one-third, which is the customary value of travel time used in travel cost studies (Feather and Hellerstein, 1997). Therefore the 50x50 travel cost matrix involved the computations of 2500 prices. In order to create a usable format for the cluster analysis, individual travel costs need to have a normal log distribution applied:

\[
\text{Travel Cost}_{ij} = \ln(2 \times (\text{DIST}_{ij} \times 0.60) + \left(\frac{1}{3}\right) \times (\text{WAGE}_i \times \frac{1}{50} \times \text{DIST}_{ij}))
\]  

(1)

Where: \( \text{DIST}_{ij} \) is the distance from origin state population \( i \) to destination state park \( j \) and \( \text{WAGE}_i \) is the median hourly wage of state \( i \), which is calculated by taking the state median annual income and dividing it by the total 2,040 work hours per year.

The next step after completing the travel cost matrix is to complete the expected utility matrix. This matrix uses the travel cost to determine the probability of someone from the average state \( i \) visiting the average state park in state \( j \). This is computed using the following equation:

\[
\text{EXPECTED UTILITY}_{ij} = 1 - (p \times (\ln(\text{TC}_{ij})))
\]  

(2)

Where: \( p \) is the probability, \( \ln \) is the natural logarithm, and \( \text{TC}_{ij} \) is the travel cost from origin state population \( i \) to destination state park \( j \). Once all of these calculations are complete the cluster analyses can move forward.

From a conceptual standpoint there are \( T \) visits to all of the state park systems in the United States that can be distributed, so that \( y_{ij} = 1 \) if the consumer of a state park visit is from population origin \( i \) and visits state park system \( j \) otherwise \( y_{ij} = 0 \). The utility for each
The consumer’s visit is $u_{ij}$, which equals the utility from residing in population origin $i$ and visiting the destination park $j$. As this problem is presented in an outdoor recreation setting $u_{ij}$ is the expected utility. The expected utilities of visiting state parks will reflect the various prices or travel costs and $u_{ij}$, the utility of a state park visit, is independent of the travel pattern. So that, $u_{ij}$ equals $-c_{ij}$ and the price $c_{ij}$ is thought of as a disutility with a negative correlation. In other words as travel cost goes up, visitation numbers can be expected to decrease. The inherent price is fundamentally made up of the out-of-pocket expense of travel, which is primarily a function of the distance traveled; the opportunity cost of travel time, which reflects the concept that the time spent traveling could be devoted to alternative activities of some worth to the individual; and the physiological and psychological efforts involved in travel. Assuming that the visitor has to pay the prices for the travel costs and expenses if they choose to make park visits between the origin and destination state parks, the net utility becomes $u_{ij} - c_{ij}$. Assuming the visitor wants to maximize net utility, the maximization problem of the visitor then becomes:

$$\sum_i \sum_j y_{ij} = 1, \ y_{ij} \geq 0.$$  

It can be assumed that this model captures the essential elements in trip making behaviors to state parks and that the estimate of the trip pattern $T_{ij} = \sum y_{ij}$ or the sum of the visits that were observed. However, with no on-site survey data regarding the number of visits from each of the population origins $i$ to destination state parks $j$, the problem cannot be computed using individual visitor data, so it will instead have to come from aggregate state park system visitation data.
**Cluster Analysis**

The analysis of these data determines the origin market clusters of states. Cluster analysis is an exploratory data-analysis technique that attempts to determine the optimal groupings of prices among states in this study. In order to do this, a STATA software package was employed (STATA Corp, 2010). The clustering variable is an aggregate price index. The logarithm of the aggregate price index was the share-weighted sum of the log of the destination site’s price. The partition method as applied with this software broke the substitute price index observations into distinct non-overlapping kmean groups. The kmeans clustering is a repetitive process, where the number of clusters, k, were specified. Each of the price index observations is assigned to the group whose mean is closest, and then based on that categorization, new group means are determined. These steps are continued until no observations change groups, hence the repetition. In order to determine the optimal number of clusters, the cluster command was executed in five-step cluster increments (5,10,15…35 groups) with the kmean partition method. A separate algorithm estimated the Calinski-Harabasz pseudo-F index at each cluster solution. The Calinski-Harabasz index is a method that is used in cluster analysis when the number of clusters is not pre-determined. This index score is used to determine the optimum number of clusters for a given matrix, the larger the value, the more distinct the clustering. This “pseudo-F” term refers to the outcome that any p-value computed from the statistics would not be valid because the cluster analysis searches for structure and the p-values normally associated with the statistic would always end up statistically significant (Calinski and
Harabasz, 1974). Using this method two different weighted matrices are examined. The first of which is a weighted travel cost matrix. This was chosen because from a theoretical standpoint, it makes sense to try and determine origin markets based solely on the concept that as travel costs increase a visitor’s willingness to travel will decrease. The second weighted matrix used is an expected utility matrix, which is essentially the normal probability distribution of people paying the travel costs to travel from origin $i$ to destination $j$.

**SUMMARY**

This chapter offered a summary of the framework utilized for developing the model used to determine a cluster of likely origin markets for state park visitation. First, ArcMap software was used to construct a 50x50 matrix which determined the travel distances from the population centroid starting points to the state park destination centroids which served as the endpoints. A calculation of 2500 travel costs from this matrix created a travel cost matrix that when run though a kmeans cluster analysis provided a distinct set of states. Another weighted matrix was calculated using the expected utility function as its weighting factor to produce another distinct set of state groupings. Chapter 4 will review the results from the actual models.
Chapter 4

RESULTS

INTRODUCTION

This chapter will detail the results produced from the ArcMap procedures to develop the 50x50 matrix of origins and destinations, as well as the outcomes of the kmeans cluster analyses run in the STATA software package for the travel cost and expected utility weighted matrices. The results as well as interpretation of the findings will be detailed in the section which follows.

ARC MAP RESULTS

ArcMap software was employed in order to construct a 50x50 matrix containing origin and destination points. The origin points are the state population centroids. The destination points are the state park centroids. As this is an aggregate examination, it was necessary to aggregate all of the population and state park data, as the underlying theory being that this study is developing a travel cost model for the average citizen of origin state $i$ to the average state park in destination state $j$. To develop as accurate an instrument as possible, it was important to note that for most states, the majority of the population and state parks do not all reside in the exact center of the state, so simply devising a matrix from the center of each state to the center of all other states would be completely inaccurate from a theoretical standpoint.
In order to account for this, it was necessary to find the center of each state based upon current population densities as well as the center of each state based upon state park densities. To complete the first step of this process already existing data from the US Census Bureau which gives the exact center of each state based upon population densities from the 2000 census was used (US Census Bureau, 2000). The second step was not as simple and required ArcMap software to enter the locations of all state parks in the US. With all of the parks identified by latitude and longitude, a spatial join was performed, which assigned all state parks to the state in which they reside (ArcMap 9.1, 2005). Next, the Mean Center Tool was used to find the exact mean center of each state based upon the densities of state park locations (ArcMap 9.1, 2005). This produced a map with the origin points as well as the destination points for the 50x50 matrix (Figure 1). As an example, Figure 2 details the southern region of the United States in order to give a better understanding of the differences in locations for the state park centroids versus the state population centroids.
Figure 1. State Population Density Centroids and State Park Density Centroids
Figure 2. South Region State Population Density Centroids and State Park Density Centroids
Once this was completed the Point Distance Tool was used to calculate the distance from each of the state population centroids to the state park centroids based on mileage producing the complete 50x50 distance matrix (ArcMap 9.1).

CLUSTER ANALYSIS

Two different matrix weights utilizing STATA 11 were used to analyze the data. First, a Travel Cost Matrix was analyzed to determine the clusters. Travel Cost is calculated as the round-trip travel at $0.60 per mile plus the opportunity costs of travel time which consists of (a) two times the distance between the state population origin and destination park system divided by 50 miles per hour, (b) multiplied by the hourly wage, which is the median annual income of origin state residents for year 2009 divided by the 2,010 annual work hours, (c) multiplied by the constant of one-third (Feather & Hellerstein, 1997). Using the Calinski-Harabasz pseudo-F index, clusters of 5, 10, 15, and 20 were analyzed (Table 1). For the Calinski-Harabasz pseudo-F index, the higher the score calculated, the more distinct the clustering, and therefore the better result. From this it was determined that 5 clusters seemed to be the best fit. With this information a kmeans cluster of 6 was calculated to determine if this would be a better fit than 5. It received a pseudo-F score of 20.64 (Table 1) improving the Calinski-Harabasz pseudo-F index score for 5 clusters. Further calculation to optimize clusters using a cluster of 7 was investigated to determine if it would produce a better result, but found that this solution scored a 15.96 (Table 1), so it was surmised that the 6 cluster groupings was the best fit based upon this matrix (Table 2).
Table 1
*Travel Cost Matrix Calinski- Harabasz Scores*

<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>Calinski- Harabasz peudo-F Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>20.08</td>
</tr>
<tr>
<td>10</td>
<td>11.18</td>
</tr>
<tr>
<td>15</td>
<td>12.89</td>
</tr>
<tr>
<td>20</td>
<td>12.84</td>
</tr>
<tr>
<td>6</td>
<td>20.64</td>
</tr>
<tr>
<td>7</td>
<td>15.96</td>
</tr>
</tbody>
</table>

Table 2
*Travel Cost Matrix: 6 Clusters of State Origin Markets for Visitation to State Park Systems*

<table>
<thead>
<tr>
<th>Cluster</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>South</td>
<td>Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, Tennessee, Texas</td>
</tr>
<tr>
<td>Pacific</td>
<td>Alaska, Hawaii, Oregon, Washington</td>
</tr>
<tr>
<td>Mid America</td>
<td>Indiana, Kentucky, Maryland, Michigan, North Carolina, Ohio, South Carolina, Virginia, West Virginia</td>
</tr>
<tr>
<td>West</td>
<td>Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming</td>
</tr>
<tr>
<td>Mid West</td>
<td>Illinois, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, Oklahoma, South Dakota, Wisconsin</td>
</tr>
</tbody>
</table>

A cluster analysis was also conducted on the expected utility matrix. Expected utility can be expressed as:

$$\text{EXPECTED UTILITY}_{ij} = 1 - (p(\ln(TC_{ij})))$$ (1)

Where: $p$ is the probability, $\ln$ is the natural logarithm, and $TC_{ij}$ is the travel cost from origin state population $i$ to destination state park $j$.

This expected utility will give the probability distributions of people paying the travel
costs to get to the destination parks. Again, the clusters were generated in groupings of five, so cluster analyses were performed on clusters of 5, 10, 15, and 20 (Table 3). This matrix produced much higher Calinski-Harabasz pseudo-F index scores. With 5 and 10 being the best scores (36.23 and 37.66 respectively) a 9 cluster analysis was calculated (Table 3) to find a better fit (32.63). A cluster analysis was conducted on clusters of 11 and 12 respectively (Table 3). Since 11 spiked to 44.19 and 12 dropped to 37.50, it was determined that for this matrix the result of 11 clusters provided the best result (Table 4).

<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>Calinski-Harabasz pseudo-F Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>36.23</td>
</tr>
<tr>
<td>10</td>
<td>37.66</td>
</tr>
<tr>
<td>15</td>
<td>33.45</td>
</tr>
<tr>
<td>20</td>
<td>28.2</td>
</tr>
<tr>
<td>9</td>
<td>32.63</td>
</tr>
<tr>
<td>11</td>
<td>44.19</td>
</tr>
<tr>
<td>12</td>
<td>37.5</td>
</tr>
</tbody>
</table>
Table 4
*Expected Utility Matrix: 11 Clusters of State Origin Markets for Visitation to State Park Systems*

<table>
<thead>
<tr>
<th>Cluster</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gulf Coast</td>
<td>Alabama, Florida, Louisiana, Mississippi, Texas</td>
</tr>
<tr>
<td>South</td>
<td>Georgia, Kentucky, North Carolina, South Carolina, Tennessee, West Virginia</td>
</tr>
<tr>
<td>Mid West</td>
<td>Illinois, Indiana, Michigan, Ohio</td>
</tr>
<tr>
<td>Mid Atlantic</td>
<td>Delaware, Maryland, New Jersey, Pennsylvania, Virginia</td>
</tr>
<tr>
<td>North Central</td>
<td>Iowa, Minnesota, Nebraska, Wisconsin</td>
</tr>
<tr>
<td>Dakotas</td>
<td>North Dakota, South Dakota</td>
</tr>
<tr>
<td>Tornado Alley</td>
<td>Arkansas, Kansas, Missouri, Oklahoma</td>
</tr>
<tr>
<td>Rockies</td>
<td>Colorado, Idaho, Montana, New Mexico, Utah, Wyoming</td>
</tr>
<tr>
<td>West</td>
<td>Arizona, California, Nevada, Oregon, Washington</td>
</tr>
<tr>
<td>New England</td>
<td>Connecticut, Maine, Massachusetts, New Hampshire, New York, Rhode Island, Vermont</td>
</tr>
<tr>
<td>Pacific</td>
<td>Alaska, Hawaii</td>
</tr>
</tbody>
</table>

**SUMMARY**

This chapter provided the results of the ArcMap and STATA software applications that were used to develop the model for determining the likely origin markets for state park visitation. ArcMap was employed to develop the 50x50 matrix of travel distances from state population origins to the state park destinations. Using STATA software, an optimum group of 6 clusters of states was produced when looking purely at the travel cost matrix. While an optimum group of 11 clusters of states was produced when examining the expected utility matrix. The implications of these two clusters will be examined in the next chapter.
Chapter 5

CONCLUSION

INTRODUCTION

It was the purpose of this study to develop a method for deriving the likely origin markets for state park visitation. State park systems do not typically have the resources to conduct in depth surveys of all of the visitors to all of their parks. Occasionally they can perform a survey of an individual park here and there, but not for the entire system. The lack of information on visitors to parks makes it very difficult for managers to know who is visiting the parks and what origin markets are contributing to park system visits. While businesses and government agencies have different objectives, it is still vital to the continued success to know where one’s customers/visitors origin markets are located. In order to alleviate the management issues arising from this lack of visitor information, secondary data is used to derive origin markets for state park systems.

METHODOLOGY

This study utilized ArcMap 9.1 spatial software to construct a 50x50 matrix of travel distances. The origin points were the mean center of every state based upon population densities, while the destination points the mean center of every state based upon state park location densities. Once this was constructed these data were used along with STATA
statistical software in order to develop a set of clusters that would serve as the likely origin markets. There were two different cluster analyses done, one using travel costs as the variable, and the other using the expected utility of travel as the variable. Travel cost is calculated as the round-trip travel at $0.60 per mile plus the opportunity costs of travel time which consists of (a) two times the distance between the state population origin and destination park system divided by 50 miles per hour, (b) multiplied by the hourly wage, which is the median annual income of origin state residents for year 2009 divided by the 2,010 annual work hours, (c) multiplied by the constant of one-third (Feather & Hellerstein, 1997). By examining the results of the Calsinki-Harabasz pseudo-F Index scores for the two analyses the cluster grouping that will work best as the likely origin markets for state park visitation was determined to be the expected utility of travel matrix grouping of 11 clusters.

RESULTS AND RECOMMENDATIONS

The major outcome of this study was that it was able to develop two different cluster groupings depending on which determinant was chosen to use as the clustering variable. For this study it appears that the expected utility matrix of 11 clusters is the best for establishing acceptable likely origin markets for state park visitation. Choosing the grouping of 11 clusters over the grouping of 6 clusters seems to make more sense since it was established that there is a negative relationship between willingness to travel and distance, so the smaller the number of states in each cluster will give the more optimal solution. This determination is also supported by the Calinski-Harabasz pseudo-F index
scores. The optimum number of clusters from the travel cost matrix is 6 clusters, which produced a pseudo-F level of 20.64, while the optimum number of clusters from the expected utility matrix is 11 clusters, which produced a pseudo-F level of 44.19 more than twice the score of the optimum travel cost matrix clusters. As the clusters were examined, it was apparent to see why this model is more reliable. The largest cluster group is the New England cluster, which has seven states included in the cluster, but this cluster contains many small northeastern states including, Rhode Island, Vermont, and New Hampshire. There is a tie for the smallest cluster group at 2 states per cluster. One of these is the Dakotas cluster which contains North and South Dakota, and the Pacific cluster which contains the two non-contiguous states in the union in Alaska and Hawaii. From a theoretical standpoint, these clusters make the most sense as opposed to some of the larger clusters from the travel cost matrix whose smallest clustering of states was 4 and the 2 largest clusters both contained 10 states each. As we saw previously in the review of current literature, there are a myriad of studies modeling tourism demand. These studies used many different modeling techniques to answer their specific research question. The type of question the study was attempting to answer, as well as the types of data used for the particular study dictated which modeling technique would work best. For the purposes of this study it was necessary to use a technique not previously used in tourism demand modeling as the data available are from secondary sources. The results of this study suggest that the expected utility weighted matrix provided the best results. As this matrix measured the probability visitors would be willing to pay the associated travel costs to
attend a state park in a specific cluster of states, this makes the most sense when we are trying to determine likely origin markets.

The recommendations that this research leads to is based from a theoretical standpoint. In theory, it would be most ideal for state park system budgets to allow for the occasional system-wide visitor survey analysis. That way, managers could simply do an analysis of the visitation data collected from these surveys to develop their own state system specific origin markets. But until the funds are allocated for state park system managers to conduct such a survey, this theoretical framework will serve as a way to obtain this information by utilizing the available secondary data. This framework can also be used as a model for deriving likely origin visitor markets for other similar entities. National Parks operate more like destination locations where a visitors willingness to travel is higher due to the perceived value of the uniqueness of the destination. Therefore, National Parks may not be able to gain the same insights with the same confidence level as this model applies to the state park system. But a collection of small tour operators in a centralized location may not have the budgetary discretion to perform an in-depth visitor survey either, therefore tour operators may be able to use this model as a framework for conducting their own cluster analysis to derive likely origin markets for their visitation numbers. State park system managers can use this model to obtain a greater understanding of their visitor’s origin market. The implication of this being that this model can help park system managers know if they will see increased or decreased attendance based upon impacts in that market area. It can also be used as a guide in determining the annual entry fee for a given park.
system. For instance, if a manager knows that they will see a decrease in visitors in a given year due to an impact in one of the park’s main visitor origin states then they can determine whether or not to raise the entry fees to the park in order to recoup the lost revenue that the decrease in visitation will have on the park. This model can be used not only by individual state park managers, but also by state park system managers. For those state parks that are heavily based upon the recreation activities offered on their site (like golf courses), in park systems where entry fees make up the majority of their operating budget, knowing where your visitors come from can help managers to help increase attendance numbers so that they can increase revenue not just for the individual park but also for the overall system budget.

A study by Steven Davis (2008) detailed the different management strategies for state park systems in the United States. He detailed the levels to which each state land management agency is oriented towards preservation, resource extraction, and recreation. From this list a future study could be performed to determine possible spillover effects of visitors choosing to attend a state park in a neighboring state because of its orientation towards recreation vs. resource extraction. The complete list of applications of this model cannot be summarized here, but it is clear to see the wide assortment of possibilities.

This is a theoretical model, based upon many different assumptions of the travel, tourism, and recreation disciplines. It should therefore be used as a tool to guide policy makers in their decision-making processes when it comes to the many different aspects of state park system management or whatever business entity is employing this framework.
BIBLIOGRAPHY


StataCorp, LP (2010). STATA 11 [Computer Software]. College Station, TX.


