DETERMINATES OF CLUSTERING ACROSS AMERICA'S NATIONAL PARKS: AN APPLICATION OF THE GINI COEFFICIENT

R. Geoffrey Lacher
Department of Parks, Recreation & Tourism Management
Clemson University
rlacher@clemson.edu

Matthew T.J. Brownlee
Clemson University

Abstract.—The changes in the clustering of visitation across National Park Service (NPS) sites have not been well documented or widely studied. This paper investigates the changes in the dispersion of visitation across NPS sites with the Gini coefficient, a popular measure of inequality used primarily in the field of economics. To calculate the degree of clustering nationally, the researchers used visitation data provided by the NPS Public Use Statistics Office for 41 parks covering 34 years. Two ordinary least squares regressions were performed to test whether a time trend, total visits to the parks in question, and economic indicators have a significant impact on clustering of NPS visitation. Results indicated that visitation and visitation density tend to cluster in the more popular parks over time when vacations are more expensive (i.e., when gas prices and the exchange rate increases).

1.0 INTRODUCTION

The total visitation to national parks and within specific park units has been a focal point since the inception of the National Park Service (NPS) (Runte 1987). Researchers have been continually interested in the trends of public land visitation and outdoor recreation participation for decades (Cordell et al. 2008). Often public officials use visitation data, such as total annual visits and visitation trends across time, to elicit public support for conservation initiatives, to understand visiting populations, and to inform new policies (Cessford and Muhar 2003, Hendee and Dawson 2002, Manning 1999). Additionally, visitation data are often directly tied to the appropriation of federal funds to an agency and subsequently within an agency (National Park Service 2010). Furthermore, assessing and understanding park visitation trends allows for an increased collective understanding of human behavior within natural environments, such as visitor impacts to the resource or the overall visitor experience (Loomis 2000).

The methods used to monitor annual national park visitation have been inconsistent. For example, the variety of techniques employed to count and monitor total visitation at a particular park site (e.g., vehicle counts, survey data, permit allocations) vary across sites, sometimes by year, and often within individual sites (Cessford and Muhar 2003). This creates problems for researchers and managers alike when assessing visitation trends.

This limitation aside, a wealth of data exists about visitation trends to national parks. For example, overall visitation to national parks peaked in 1987, when more than 287 million people reportedly visited a national park, and then steadily declined in the early 1990s (Cordell et al. 2008). After rebounding slightly in the late 1990s, national park visits have remained relatively stable since 2001.

However, assessing only overall visitation trends does not capture the full story of visitation to NPS sites. Although overall park visitation may remain stable with limited fluctuations during a decade (e.g., 2001-2010), the degree of clustering of visits across sites may vary widely. For instance, the National Park Service may experience the same number of overall visits in 2006 and 2007. However, visits to specific National Park (NP) units (e.g., Everglades NP and the Great Smoky Mountains NP) may change considerably from 2006 and 2007. This is important because understanding fluctuations in visitation between parks may inform managers (at a national and regional level) about park popularity in relation to other parks, or the effect of infrastructure near the parks on destination choice. Additionally, if only overall visitation trends are assessed, plans for appropriate park staffing, park concessionaire contracting, or park zoning may not be fully informed. Thus, to understand visitation behavior at the macrolevel, assessing the level of dispersion across sites and across years becomes a relevant topic, especially in the current era of stagnation in total park visitation.

2.0 PURPOSE OF THE RESEARCH

This research sought to understand how national park visits across sites have changed annually, with specific interest in the level of annual inequality (i.e., dispersion) of visits across sites. The first step was assessing the degree of inequality of visits across sites annually. Second, the researchers investigated what factors may account for the degree of inequality in annual visits over time.

3.0 METHODOLOGY

To measure the changing distribution of visitation across the NPS, the researchers used the Gini coefficient. The Gini coefficient is a frequently used measure of inequality in the field of economics. The Gini coefficient can range from 0 to 1 with 0 representing perfect equality and 1 representing perfect inequality. It is most commonly used to measure income inequality but has been used in numerous applications. In parks and and tourism research, the Gini coefficient had been used to quantify seasonality (Fernadez-Morales 2003) and industry agglomeration (Urtasun and Gutierrez 2006).

3.1 Calculating the Gini Coefficient

While using the Gini coeffiecent to assess visitation inequality across the park's system is a novel application, this statistic can measure the inequality in any data set. Specifically, the Gini coefficient is the ratio of: 1) the difference between a Lorenz curve and a line of perfect equality; and 2) the total area under the line of perfect equality (see Figure 1). The Lorenz

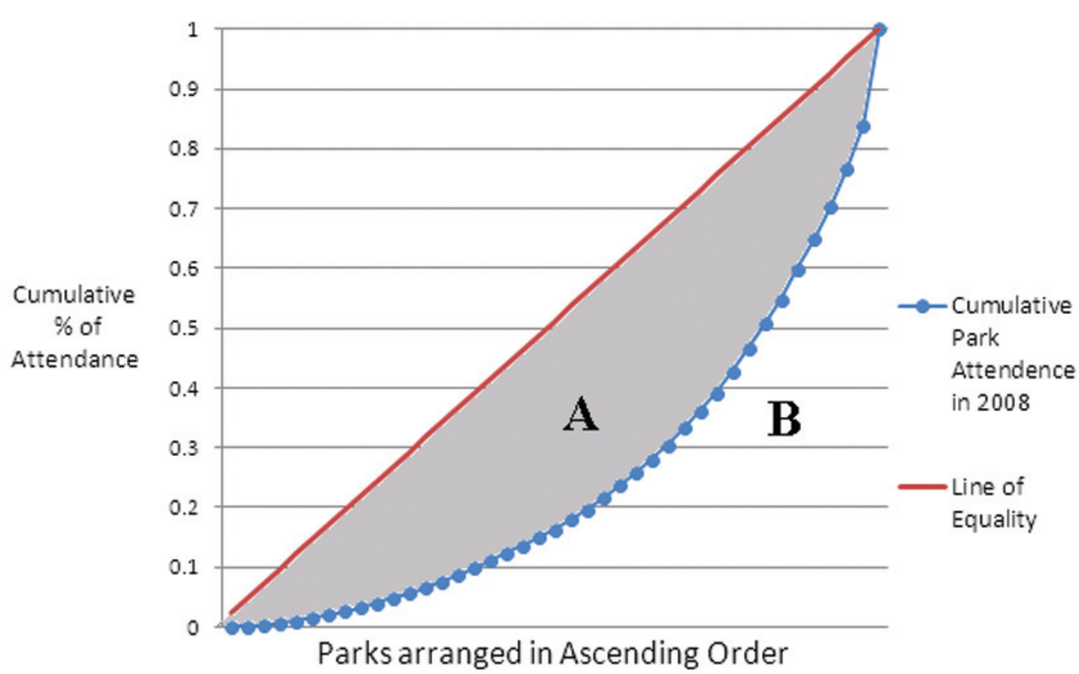

Figure 1.—A visual representation of the Gini coefficient. The Gini coefficient is equal to the ratio of the area between the two lines to the area below the line of equality or Gini = A / (A+B).

curve visually displays the relationship between the cumulative percentage of numbers when the numbers are arranged in ascending order. For our purposes, the x-axis of the Lorenz curve is the cumulative percentage of the parks in the sample (arranged from least to most visited), and the y-axis represents the cumulative percentage of number of visits to these parks.

The Gini coefficient is a preferred method for calculating inequality for a variety of reasons. First, it is scale independent and is not affected by the type of units measured. Additionally, it is easy to interpret (e.g., 0 to 1). Other measurements of inequality, including the logarithmic variance or Herfindahl-Hirschman index, have no upward bound and therefore can be more difficult to interpret (Urtasun and Gutierrez 2006). The Gini coefficient also considers all data points when measuring inequality and is not as influenced by extreme values as other measures of inequality, such as those relying on variance or standard deviation (Wanhill 1980).

3.2 Visitation Data for the Gini Coefficient

To calculate a Gini coefficient for each year of visitation, the researchers used a data set of the total visitation and park acreage of 41 parks over 34 years (data gathered from the NPS Public Use Statistics Office Web site). The original data set included 48 parks, but after initial review of the data, 7 parks were removed from further analysis due to discrepencies (e.g., after remaining relatively stable for a number of years, total visitation would double in a year, and then stabilize again for the remaining years) that perhaps resulted from inconsistent visitation monitoring methods.

The data set with the remaining 41 parks was used to calculate the Gini coefficients for total visitation and the number of visitors per acre (i.e., visitation density) for every year from 1975 to 2009, yielding 34 observations for both total visitation and visitation density. The two different Gini coefficients measure different types of clustering and will give greater depth to the analysis. The Gini coefficients for total visitation measure the degree of visitation clustering in the more popular parks, while the Gini coefficients for visitation density measure clustering in the more crowded parks. For example, Yellowstone is a park with high visitation but, due to its size, it has low visitation density. The Gini coefficients for total visitation and visitation density were calculated for the 34 years using the trapezoidal method (equation below) in which X_k represents the cumulative proportion of the parks and Y_k represents the cumulative proportion of visitors arrival (again where the parks are arranged by ascending order of visitation). This method calculates the area under the Lorenz curve for each park as a trapezoid and the sum of the trapezoids is the Gini coefficient.

$$G_1 = 1 - \sum_{k=1}^{n} (X_k - X_{k-1})(Y_k + Y_{k-1})$$

3.3 Predictor Variables

An initial set of predictors was created from an examination of previous literature and consulation with colleagues (Table 1).

Two ordinary least square regressions (OLS) were performed using independent variables to predict the Gini coefficient for total visitation and visitation density for the parks. Due to multicollinearity issues within the variable list, several independent variables were eliminated. Using a trial and error approach, models with high R^2, significant predictors, and low multicollinearity were created. The final independent variables included in the models were gas price, gross national product (GDP) growth, United States dollar (USD) exchange rate, and time trend (i.e., year).

4.0 RESULTS

The mean for the Gini Coefficient for total visitation over the 34 observations was 0.55 (SD = 0.012) and for visitation density was 0.62 (SD = 0.019) (see Figure 2). A visual examination of the data reveals similarities between the total visitation inequality in the late seventies and in the late 2000s, and the relationships appear to be negatively correlated.

Table 1.—Variables used in the analysis

Variable	Interpretation	Source
Year	Numbered 0-34 for the years 1975 to 2009	–
Gas price	Inflation adjusted price of a gallon of gas in dollars	EIA[1]
Exchange rate	Relative cost of a dollar to a basket of foreign currencies. The higher the number, the weaker the dollar	Federal Reserve (2009)
Growth	Gross Domestic Product (GDP) growth for the year	BEA[2]
Unemployment*	Average unemployment rate for the year	BLS[3]
1995 Dummy Variable*	0 for all pre-1995 years, 1 for all post-1995 years, based on a change made in National park fee structures in 1995	–
Visits *	Visits to the 41 parks in question	NPS[4]

Notes: *Variable eliminated from final model due to colinearity issues.
[1]Energy Information Administration (2009).
[2]Bureau of Economic Analysis (2009).
[3]Bureau of Labor Statistics (2009).
[4]National Park Service Public Use Statistics Office.

The Gini coefficient for total visitation and the Gini coefficient for visitation density are in fact significantly correlated ($p<0.05$) in a negative manner, meaning that when visitation to the National Park system is more clustered (i.e., the most popular parks are more highly visited), the parks with the highest visitation per acre receive relatively less visitation. However, the variables only explain a small amount of each other's variance ($R^2 = 0.14$).

The model predicting the Gini coefficient for total visits to parks was significant ($p<0.01$) and produced an R^2 of 0.366 (Table 2). The only significant predictors were gas price and exchange rate. This model suggests that as the price of gas increases, clustering of visitation increases as well. The coefficeint for exchange rate indicates that when more funds (USD) are necessary to purchase foreign currency (i.e., when the dollar is weak), more clustering in visitation occurs.

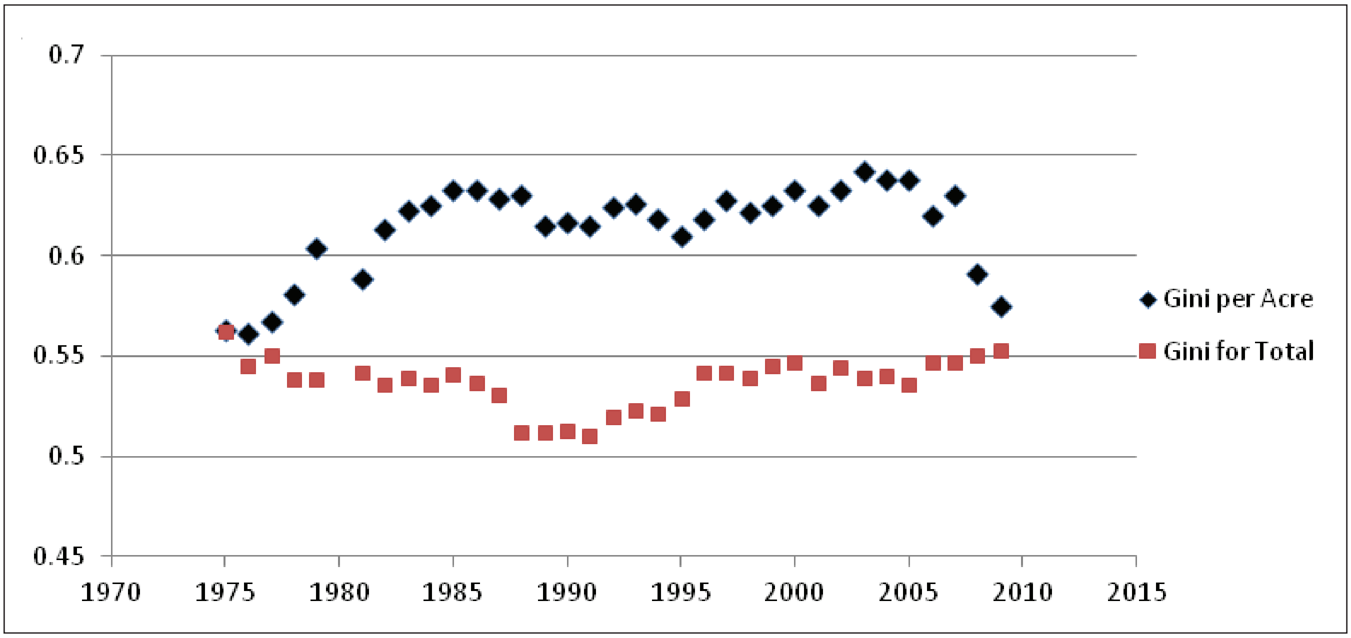

Figure 2.—The Gini coefficients across time for the 41 parks.

Table 2.—Results of regression on Gini coefficient for total visitation to parks

	B	Std. Error	p
Constant	.5299*	.005360	.000
Year	.000278	.000192	.159
Growth	.633	.001095	.568
Gas price	.135*	.000038	.001
Exchange	.409*	.000173	.025

*indicates $p \leq 0.05$.

The model predicting the Gini coefficient for density of park visitation was significant (p<0.01) and produced an R^2 of 0.462 (Table 3). The significant predictors were the time trend (i.e., year) and exchange rate. The coeffiecent for year indicates that as time has increased, visitation has become increasingly clustered. The coefficient for exchange rate demonstrates that when the dollar is weak, more clustering in visitation occurs.

5.0 DISCUSSION

The Gini coefficient for visitation density indicates a tendency towards clustering over time. This indicates a trend of visitors seeking less isolated and less remote parks. The parks with denser visitation tend to have a greater emphasis on easy-to-access, short-duration activities while the parks with less dense visitation tend to emphasize multiday wilderness experiences. Although this trend is slow in developing, it has some potential implications. For example, near-park communities may economically benefit from a higher concentration of visitors (e.g., Gatlinburg, TN outside the Great Smoky Mountains NP). However, as more visitors converge in one area, increases in visitor-caused environmental impacts and/or social crowding and loss of solitude may occur.

Table 3.—Results of regression on Gini coefficient for visitation density to parks

	B	Std. Error	p
(Constant)	.5953*	.008822	.000
Year	.001127*	.000316	.001
Growth	.000247	.001802	.892
Gas price	-.000098	.00062	.126
Exchange	.000796*	.000284	.009

*indicates $p \leq 0.05$.

The Gini coefficient results for total visitation indicate that visitation tends to cluster in the more popular parks when vacations are more expensive (e.g., when gas price and exchange rate both increase). This observation adds a nuance to the previously reported hypothesis that National Park vacations might be an inferior good (i.e. individuals' demand for a product decreases when their income increases) (Weiler and Seidl 2004). Additionally, in this analysis, exchange rate may also be a proxy for international visitation and could indicate that international visitors tend to go to the more popular parks located in proximity (e.g., Yellowstone and Great Smoky Mountains Parks).

6.0 CONCLUSIONS

The Gini coefficient is an efficient and effective method to summarize a changing distribution and can easily be applied by researchers involved in recreation resource management. The Gini coefficient could be applied to many additional types of analyses related to park visitation. For example, visitation data could be analyzed at the regional and state level or assessed across numerous land management agencies (e.g., Burea of Land Management, NPS, U.S. Forest Service) using the Gini coefficient. Thus, the Gini coefficients could be compared across agencies within a given time period (e.g., one year). Additionally, the Gini coefficient may be a useful tool for those investigating issues beyond visitation. For example, total annual expenditures allocated to concessionaires at park sites could be assessed across parks.

As park researchers and resource management professionals move forward with new methods and novel inquiry, the Gini coefficient may assist in understanding and explaining components of visitor behavior. Furthermore, assessing inequalities may lead to increased insight about the role of public lands in society, and about human behavior.

7.0 LITERATURE CITED

Bureau of Economic Analysis. 2009. **U.S. economic accounts.** Available at http://www.bea.gov/. (Accessed October 10, 2009).

Bureau of Labor Statistics. 2009. **Labor force characteristics.** Available at http://www.bls.gov/cps/lfcharacteristics.htm#unemp. (Accessed October 10, 2009).

Cessford, G.; Muhar, A. 2003. **Monitoring options for visitor numbers in parks and natural areas.** Journal for Nature Conservation. 11: 240-250.

Cordell, K.; Betz, C.; Green, G. 2008. **Nature-based outdoor recreation trends and wilderness.** International Journal of Wilderness. 14(2): 7-13.

Energy Information Administration. 2009. **Petroleum.** Available at http://www.eia.doe.gov/. (Accessed October 10, 2009).

Federal Reserve. 2009. **Foreign exchange rates.** Available at http://www.federalreserve.gov/releases/h10/Summary/. (Accessed October 10, 2009).

Fernandez-Morales, A. 2003. **Decomposing seasonal concentration.** Annals of Tourism Research. 30(4): 942-956.

Hendee, J.C.; Dawson, C.P. 2002. **Wilderness management: stewardship and protection of resources and values.** Golden, CO: Fulcrum Publishing.

Loomis, J. 2000. **Counting on recreation use data: A call for long-term monitoring.** Journal of Leisure Research. 32(1): 93-96.

Manning, R. 1999. **Studies in outdoor recreation: search and research for satisfaction.** Corvallis, OR: Oregon State University Press.

National Park Service. 2010. **NPS budget justifications.** Available at http://home.nps.gov/applications/budget2/gbchoose.htm. (Accessed May 13, 2010).

Runte, A. 1987. **National parks: the American experience. 3rd ed.** Lincoln, NE: University of Nebraska Press.

Urtasun, A.; Gutiérrez, I. 2006. **Tourism agglomeration and its impact on social welfare: an empirical approach to the Spanish case.** Tourism Management. 27(5): 901-912.

Wanhill, S. 1980. **Tackling seasonality: a technical note.** Tourism Management. 1(4): 243-245.

Weiler, S.; Seidl, A. 2004. **What's in a name? extracting econometric drivers to assess the impact of national park designation.** Journal of Regional Science. 44(2): 245-262.