BRINGING NATIONAL PARK VISITATION MODELS INTO THE FUTURE: INVESTIGATING THE UTILITY OF GOOGLE TRENDS DATA

A THESIS

Presented to

The Faculty of the Department of Economics and Business

The Colorado College

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Arts

By

Anil Jergens

Graduation May 2023

BRINGING NATIONAL PARK VISITATION MODELS INTO THE FUTURE: INVESTIGATING THE UTILITY OF GOOGLE TRENDS DATA

Anil Jergens

May 2023

Mathematical Economics

ABSTRACT: Ever-increasing visitation to American national parks puts strain on the land and people that use it while bringing economic benefits to the surrounding area. Many papers that investigate these economic benefits use models for national park visitation to estimate the current and future economic impact of these lands. To estimate consumer demand, they use disposable per capita income and gas prices. This paper investigates the possibility of using Google Trends search data both in combination with and as an alternative to these consumer demand estimators. In combination, multicorrelation between the three variables makes them difficult to use. Separately, each is similarly useful, and it may be possible to substitute one for the other in circumstances that demand it.

Keywords: National Parks, Park Visitation, Internet Search Data

ON MY HONOR, I HAVE NEITHER GIVEN NOR RECEIVED UNAUTHORIZED AID ON THIS THESIS

Signature

TABLE OF CONTENTS

ABSTRACT

1. INTRODUCTION……………………………………………………………….….4
2. LITERATURE REVIEW………………………………………………………….6
3. THEORY…………………………………………………………………………….…8
4. DATA……………………………………………………………………………..…….9
5. MODEL……………………………………………………………………………….13
6. RESULTS AND DISCUSSION………………………………………….…….13
7. CONCLUSION……………………………………………….…………………….16

**Introduction**

Public lands visitation has changed drastically since their inception (Figure 1) (Google Trends, 2022). The creation of the first national parks caused a stir of interest in the first two years of existence, but the number of visitors cratered over the next half decade. Never ones to be deterred, more and more Americans visited public lands over the following century. Today, American public lands are an international attraction. In 2022, locals who live near national parks and public lands in the Rocky Mountain West feel that visitation has exploded (Jergens, 2022). Locals enjoy the positive, economic impact of public lands while lamenting the issues that crowding brings. National parks and the lands surrounding them are also of great value to those who visit them. The continual increase in public lands visitors means that management of these lands becomes ever more important to both locals and visitors (Jergens, 2022). Research that is vital for understanding public lands and their economic value relies upon models of visitation. This paper investigates the possibility of improving these models with modern Google Trends data.

Figure 1: Visitation to Public Lands

Source: NPS Stats Annual Summary Report, 2022 <https://irma.nps.gov/>

Google Trends data is “search interest over time for a topic […] as a proportion of all searches on all topics on Google at that time and location” (Rogers, 2016). In other words, Google Trends takes a sample of searches at a certain time and determines the number of searches for the given topic relative to the total searches. Then, it takes this ratio and compares it to ratios over different periods to assign each a score between 0 and 100. Figure 2 plots the Google Trends data for the topic “public lands” in the United States from 2004-2022.

Figure 2: Google Trends Data

Source: Google Trends, 2022 <https://trends.google.com/trends/>

From this plot, it is clear that interest in the topic has generally increased over the past two decades, with a dip during the Great Recession. Is it possible that this data can add to our understanding of what drives national park visitation?

**Literature Review**

A limited number of papers investigate the economic value of public lands. They generally find that public lands are positively correlated with economic health (Loomis and Richardson, 2001; Kim and Marcoullier, 2021; Duffy-Deno, 1998; and Rasker, 2006). For example, Kim and Marcoullier find that public lands significantly explain local economic growth but are less conclusive in explaining population and employment growth. Visitation can drive this economic health because visitors bring money to areas with public lands. They purchase gas, groceries, dinners, and accommodations, all of which boost local economies. Papers which investigate public lands visitation in particular include Cline, et al. (2011), Weiler and Seidl (2004), Weiler (2006), and McIntosh and Wilmot (2011). McIntosh and Wilmot (2011) create a large model for all public lands that includes data on acreage and real disposable income. They, along with Weiler (2006) find that public lands are inferior goods where visitation increases with stagnating incomes. Along with Weiler and Seidl (2004) and Weiler (2006), Cline, et al. (2011) found that the redesignation explained an increase in visitation. Cline, et al. (2011) investigate the explanatory power of public lands designations and the economic value they bring by comparing visitor numbers before and after public lands were redesignated as national parks from national monuments. They use gas price to quantify consumer demand, acreage, and state and neighboring state population to estimate public lands visitation. They then use this model to predict visitation and multiply it with the value of each additional visitor. They find that public lands bring significant economic value, while national parks are especially valuable to surrounding areas.

Recent papers have included social media or search engine trends data to improve the predictive power of visitor prediction models. Wood, et al. (2020) find that social media posts are illustrative of visitor numbers, even on unmonitored public lands. Most importantly for this paper, Clark, et al. (2019) create two competing models for predicting visitors; one autoregressive model that uses the previous years’ visitor numbers and another model based on a park dummy variable estimate and previous Google Trends data. Though the result varies by park, the Google Trends model outperforms the autoregressive model overall.

This Google Trends data can be combined with past models for visitation to investigate the use of this new data. If it is found to be useful, it would provide a new, better basis for estimating the economic value of public lands.

**Theory**

The theory of National Park Visitation is based upon maximizing the utility of a composite good Y and visiting the national park, X.

Equation 1

The division of one’s disposable income between the cost of visiting a national park and all alternative goods is the constraint of this utility maximization problem. Visiting a national park has several costs associated with it: the park entrance fee, gas, travel time, food, and living accommodations. We do not have individual-level data on travel times; therefore, the cost of visiting a national park can be approximated by the fee price () and the price of gas (). The price of Y is equal to one because Y is the numeraire. This gives the constraint Equation 2.

Equation 2

We will use the quadratic utility function Equation 3. In this function, we assume that our coefficients and are positive, as our utility increases by consuming goods and . The coefficients and are negative by the strict concavity enforced by diminishing marginal utility. The coefficient is zero because neither good in this model affects the utility derived from the other.

Equation 3

We can now use the Lagrangian method for deriving a demand function from a utility function.

Equation 4

After solving this system of equations we are left with the demand function Equation 5. See Appendix A for details of the solution and signage.

Equation 5

**Data**

The panel dataset used in this study contains annual observations on 51 different national parks from 2004-2021, yielding 844 observations. Table 1 defines the variables used in the model. Visitation is the dependent variable. It is the annual number of people visiting the park, provided by the NPS stats database at <https://irma.nps.gov/STATS/>.

Table 1: Independent Variable Names, Description, and Source

|  |  |  |
| --- | --- | --- |
| Independent Variables  (Variable Name) | Variable description (units) | Source |
| State Population  (spop)  Neighboring State  Population  (nspop)  Gas Price  (gas)  Acreage  (acre)  Disposable Income  Per Capita  (dipc)  Google Trends  (uspl) | Annual population of state(s) where national park is located (people)  Annual population of state(s) bordering state(s) with national park (people)  U.S. Regular Conventional Gas Price in August of each year, CPI adjusted (dollars per gallon)  Size of national park’s gross area acres, (acres)  Annual Disposable Income per Capita in state(s) with national park, CPI adjusted (dollars)  Annual relative interest in the topic “public lands” in November in the U.S. (scale from 0-100) | Federal Reserve Economic Data, <https://fred.stlouisfed.org/>  United States Census Bureau, <https://www.census.gov/data.html> via dataset provided by Professor Aju Fenn  U.S. Energy Information Administration, <https://www.eia.gov/petroleum/gasdiesel/>  National Park Service Visitor Use Statistics, <https://irma.nps.gov/STATS/>  Bureau of Economic Analysis, <https://www.bea.gov/> , via dataset provided by Professor Aju Fenn  Google Trends, <https://trends.google.com/trends/> |

Gas prices are August averages, per Cline et al. (2011). Parks that crossed into multiple states created problems for data on state population, neighboring state population, and disposable income per capita. To combat this, each state’s population numbers were weighted by proportion of the park in each state and summed. Weighting was determined by U.S. National Park Service (2022) and Wadzinski (2019). For example, Yellowstone National Park is 96% in Wyoming, 3% in Montana, and 1% in Idaho. Each state’s percentage is multiplied by each state’s population and summed, giving the final state population for the park. Neighboring state population numbers include the populations of all states neighboring the occupied states; they were similarly weighted and summed. Disposable income per capita was weighted by proportion of each park’s area and summed.

Google Trends allows one to choose either topics or search terms. The term “national park” and the topics “public land” and “national park” in the U.S. and internationally were tested. The best choice was determined by the most significant p-value in the base model. The topic “public land” in the United States proved to be the most significant indicator of national park visitation. Next, the data for this variable, which is provided on a monthly basis, was summarized in five different ways: a yearly average, a yearly minimum, a yearly maximum, the difference between the yearly minimum and maximum, and each year’s value in November. Data from November was chosen as it was the most common yearly maximum. This indicates that November is the most important month for Google searches and will yield the most significant results. The yearly value in November was the most significant indicator of the five methods and was the chosen variable for the final model.

Not every national park has complete entries between 2004 and 2021 because of redesignations. New River Gorge National Park was redesignated a national park and preserve in 2020 from a national river. Gateway Arch National Park was redesignated a national park in 2018 from a national memorial. Indiana Dunes National Park was redesignated a national park in 2019 from a national lakeshore. White Sands National Park was redesignated a national park in 2019 from a national monument. Pinnacles National Park was redesignated a national park in 2013 from a national monument. Data on these parks begins the year after they were designated because designations happen partway through the year.

Data entries on Alaska, the United States Virgin Islands, and Hawai’i were dropped because they do not border other states, similarly to Cline et al. (2011).

Table 2: Variable Sample Statistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Standard Deviation | Minimum | Maximum |
| Visitation | 1,411,618 | 695,977.5 | 1,814,511 | 6493 | 14,200,000 |
| State Population | 12,700,000 | 5,839,077 | 13,500,000 | 509106 | 42,600,000 |
| Neighboring State Population | 23,000,000 | 17,600,000 | 15,600,000 | 3,316,009 | 65,800,000 |
| Gas Price | 1.23 | 1.19 | 0.27 | 0.81 | 1.74 |
| Acreage | 408,533.8 | 199,108.3 | 615,200.9 | 192.83 | 3,408,407 |
| Disposable Income Per Capita | 17,301.34 | 17,335.3 | 2,393.57 | 12,952.89 | 23,985.68 |
| Google Trends | 55.98 | 49 | 17.97 | 38 | 100 |
|  |  |  |  |  |  |

**Model**

The base national park visitation model is Equation 6.

Equation 6

Based on Cline, et al. (2011), expected signs are: negative, positive, positive, negative, negative, and positive. From McIntosh and Wilmot (2011) is expected to be negative. From Clark, et al. (2019) is expected to be positive.

The Breusch-Pagan and Cook-Weisberg Test reveals that this model suffers from heteroskedasticity. Stata calculates the probability of constant variance at approximately 0.000, so we may reject the null hypothesis. To test for autocorrelation on a panel dataset in Stata we will use the Breusch-Godfrey test. From this we find that we have autocorrelation to the second lag. To correct our model for both heteroskedasticity and serial correlation, we will use a regression with Newey-West standard errors and a lag of 2.

**Results and Discussion**

Table 3: Base Regression Model

|  |  |  |
| --- | --- | --- |
| Dependent variable=Annual Visitation  Observations: 844 | Coefficient  (N-W Standard Error) | P>|t| |
|  |  |  |
| Constant | -199,346.2  (894,944.5) | 0.824 |
| State Population | -0.071  (0.024) | 0.003 |
| State Population Squared | 1.08e-09  (6.43e-10) | 0.093 |
| Neighboring State Population | 0.136  (0.048) | 0.004 |
| Neighboring State Population Squared | -2.11e-09  (7.51e-10) | 0.005 |
| Acreage | 1.017  (0.208) | 0.000 |
| Gas Prices | -170,049.4  (282,657.7) | 0.548 |
| Disposable Income | 3.93  (42.36) | 0.926 |
| Google Trends | 6,106.4  (6,464.7) | 0.345 |
| F-statistic: 8.79 |  |  |

The results of our initial regression reveal that state population squared, gas prices, disposable income, and Google Trends data are all insignificant, while all other variables are significant at the 5% level. All signs except for disposable income follow past research. Disposable income follows our theory instead of McIntosh and Wilmot’s findings. This previous paper includes most public lands in the United States while this paper focuses on national parks exclusively. It seems that public lands are an inferior good while national parks are a normal good.

It appears that our Google Trends data has done little to improve this model. However, a multicorrelation table shows that Google Trends searches for public lands are closely related to both Disposable Income and Gas Prices (Table 4).

Table 4: Correlation Table for Base Regression

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | visit | uspl | spop | acre | gas | dipc | npop | spop2 | nspop2 |
| visit | 1 |  |  |  |  |  |  |  |  |
| uspl | 0.08 | 1 |  |  |  |  |  |  |  |
| spop | -0.12 | 0.04 | 1 |  |  |  |  |  |  |
| acre | 0.23 | -0.01 | 0.25 | 1 |  |  |  |  |  |
| gas | -0.06 | -0.60 | -0.03 | 0.01 | 1 |  |  |  |  |
| dipc | 0.001 | 0.55 | 0.37 | 0.20 | -0.33 | 1 |  |  |  |
| npop | 0.06 | 0.08 | -0.29 | -0.23 | -0.04 | -0.43 | 1 |  |  |
| pop2 | -0.12 | 0.05 | 0.98 | 0.24 | -0.03 | 0.40 | -0.31 | 1 |  |
| npop2 | 0.03 | 0.07 | -0.30 | -0.28 | -0.04 | -0.40 | 0.98 | -0.31 | 1 |

We will compare our Google Trends data to combinations of Gas Prices and Disposable per Capita Income in our model, separately, to determine how each is related to national park visitation (Table 5). For each of these models, State Population, State Population Squared, Neighboring State Population, Neighboring State Population Squared, and Acreage remain, while the variables in the top row of the table are added in.

Table 5: Comparing Models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | uspl | dipc | gas | dipc, gas |
| Model F-Stat | 12.47 | 13.60 | 12.21 | 10.77 |
| Variable P>|t| | 0.124 | 0.181 | 0.158 | 0.376, 0.268 |

Table 5 compares the models of each of these variables by their model F-statistics and their individual p-values. The explored variables remain insignificant.

This begs the question as to why the disposable income model fits better while the variable itself has a worse p-value than the Google Trends model. The answer to this is in their correlation tables (Table 6, 7).

Table 6: Disposable Income per Capita Model Correlation Row

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | visit | spop | spop2 | nspop | nspop2 | acre |  |
| dipc | 0.0014 | 0.37 | 0.40 | -0.40 | -0.40 | 0.20 |  |

Table 7: Google Trends Model Correlation Row

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | visit | spop | spop2 | nspop | nspop2 | acre |  |
| uspl | 0.08 | 0.04 | 0.05 | -0.08 | 0.07 | -0.01 |  |

These tables reveal that multicollinearity is again the culprit behind the result. As disposable income is multicollinear with other variables, the model’s accuracy increases while our ability to trust individual variable statistics decreases.

With this knowledge, we can firmly say that Google Trends data is on par with both Disposable Income and Gas Prices for explaining national park visitation.

**Conclusion**

In this research, Google Trends, Disposable Income, or Gas Prices are all insignificant in explaining national park visitation, yet theory dictates they remain in models. Gas Prices and Disposable Income are included in these models because they are parts of a theoretical consumer demand function. As we found earlier, Google Trends data is multicollinear with Gas Prices and Disposable Income. This indicates that Google Trends might be used as a stand-in for Gas Prices and Disposable Income in quantifying consumer demand.

It is likely that a researcher would need to replace Gas Price or Disposable Income data as they may be untrustworthy or unavailable in other locations with parks. In these cases, it may be possible for researchers to use Google Trends data to fill this gap. While working on this paper the author discovered that the U.S. Virgin Islands contain a national park, but there is no widely available data on their Disposable Income. If one were creating a model for national park visitation in the U.S. Virgin Islands, Disposable Income could be replaced with Google Trends data that is easily accessible and could be used over the years that it has been collected.

However, there are limitations to the use of Google Trends data for predicting public lands visitation. In many countries, Google is unavailable or not the main search engine. Even in the U.S., Google only became the leader in search engine market share in 2003. Search engines have grown massively in their impact on individual lives over the period of this study, so it may be that Google Trends data is more explanatory in recent years than it was in the 2000’s. Different subsets of people may use search engines differently, so Google Trends may be better at explaining certain segments of the population. For example, younger people or urban people may be more likely to use Google when considering a national park visit.

Future research could break down Google Trends data on a state-by-state level, similarly to our population data. This might improve the explanatory power of the new data greatly.

Google Trends data holds a trove of information on interest in an unlimited variety of subjects. It could be used to indicate consumer interest in international travel by country or even state. It could be used to show how many people are interested in quitting drugs. The internet knows more about us than we do, and this resource has great potential for improving our understanding of the world. In this particular case, it appears that this new data may be a useful stand-in for previous methods of quantifying consumer demand.

Appendix A

As noted previously, coefficient because consumption of one good does not affect the utility of the other.

From this, we can tell that the price of visiting the national park has an inverse relationship with an individual consumer’s national park visitation. Disposable income has a linear positive relationship with visitation.

References

Clark, M., Wilkins, E. J., Dagan, D. T., Powell, R., Sharp, R. L., & Hillis, V. (2019). Bringing forecasting into the future: Using Google to predict visitation in U.S. national parks. *Journal of Environmental Management*, *243*, 88–94. <https://doi.org/10.1016/j.jenvman.2019.05.006>

Cline, S. A., Weiler, S., & Aydin, A. (2011). The value of a name: Estimating the economic impact of public land designation. *The Social Science Journal*, *48*(4), 681–692. <https://doi.org/10.1016/j.soscij.2011.06.001>

Duffy-Deno, K. T. (1998). THE EFFECT OF FEDERAL WILDERNESS ON COUNTY GROWTH IN THE INTERMOUNTAIN WESTERN UNITED STATES. *Journal of Regional Science*, *38*(1), 109–136.

*FAQ: Testing for panel-level heteroskedasticity and autocorrelation | Stata*. (n.d.). Retrieved October 4, 2022, from <https://www.stata.com/support/faqs/statistics/panel-level-heteroskedasticity-and-autocorrelation/>

Google Trends. (2022). Google Trends. Google Trends. <https://trends.google.com/trends/>

Jergens, A. (2022). *Report on the State of the Rockies Public Lands Survey 2022*. State of the Rockies at Colorado College. <https://docs.google.com/document/d/1s4xvy4yPNGg0EdkVqIMpKbmGqw-H6mNV/edit?usp=sharing&ouid=102619002613104366368&rtpof=true&sd=true>

Kim, D., & Marcouiller, D. W. (2021). The role of public lands in local economies of the US Lake States: A spatial simultaneous equation approach. *Land Use Policy*, *100*, 104883. <https://doi.org/10.1016/j.landusepol.2020.104883>

Loomis, J. B., & Richardson, R. (2001). Economic Values of the U.S. Wilderness System: Research Evidence to Date and Questions for the Future. *International Journal of Wilderness*, *7*(1), 31–34.

McIntosh, C. R., & Wilmot, N. (2011). An empirical study of the influences of recreational park visitation: The case of US National Park Service sites. *Tourism Economics*, *17*(2), 425–435. <https://doi.org/10.5367/te.2011.0036>

Neuvonen, M., Pouta, E., Puustinen, J., & Sievänen, T. (2010). Visits to national parks: Effects of park characteristics and spatial demand. *Journal for Nature Conservation*, *18*(3), 224–229. <https://doi.org/10.1016/j.jnc.2009.10.003>

Rasker, R. (2006). An Exploration Into the Economic Impact of Industrial Development Versus Conservation on Western Public Lands. *Society and Natural Resources*, *19*, 191–207.

Rogers, S. (2016, July 1). What is Google Trends data—And what does it mean? *Google News Lab*. <https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8>

Schwartz, Z., & Lin, L.-C. (2006). The impact of fees on visitation of national parks. *Tourism Management*, *27*(6), 1386–1396. <https://doi.org/10.1016/j.tourman.2005.12.015>

U.S. National Park Service. (n.d.). *Park Statistics—Great Smoky Mountains National Park*. Retrieved October 4, 2022, from <https://www.nps.gov/grsm/learn/management/statistics.htm>

U.S. National Park Service. (2022). STATS - National Reports. <https://irma.nps.gov/STATS/Reports/National>

Wadzinski, G. (2019, August 8). Where is Yellowstone National Park? *Yellowstone National Park*. <https://www.yellowstonepark.com/park/faqs/where-is-yellowstone-national-park/>

Weiler, S. (2006). A park by any other name: National Park designation as a natural experiment in signaling. *Journal of Urban Economics*, *60*(1), 96–106. <https://doi.org/10.1016/j.jue.2006.02.001>

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.

Wood, S. A., Winder, S. G., Lia, E. H., White, E. M., Crowley, C. S. L., & Milnor, A. A. (2020). Next-generation visitation models using social media to estimate recreation on public lands. *Scientific Reports*, *10*(1), Article 1. <https://doi.org/10.1038/s41598-020-70829-x>