DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 24-07

Survey Says: Who Needs It? Travel Cost Valuation Methods with Administrative Data

Chaz Simmons
University of Colorado Boulder

October 25, 2024

Department of Economics

University of Colorado Boulder
Boulder, Colorado 80309

Survey Says: Who Needs It? Travel Cost Valuation Methods with Administrative Data

Chaz Simmons

October 25th 2024

Abstract

Travel cost modeling is a valuable tool for valuing environmental goods, helping policymakers make informed decisions regarding funding allocation and resource management. Traditionally, the travel cost literature has relied on survey data to value camping trips and leisure activities, raising concerns about external validity and limiting the analysis of larger-scale policy questions. Despite this, little attention has been given to the potential of administrative data in conducting travel cost modeling. This paper estimates the willingness to pay for campgrounds across the State of Arizona using a novel administrative travel cost dataset comprising over 1.1 million observations from more than 187,000 camping reservations. I leverage camper willingness to pay to conduct an analysis of the budgets for the National Park Service and the United States Forest Service. My findings align with existing survey literature, demonstrating that administrative data can yield comparable results to traditional survey data. Furthermore, my dataset enables a level of analysis that is often cost-prohibitive with survey methods, offering new insights into visitor behavior and resource allocation.

1 Introduction

Understanding the value of environmental goods is crucial for policy bodies such as forest and park services, as it enables them to make informed decisions about conservation, resource management, and funding allocation. One effective method for gauging the economic worth of these public goods, like wilderness areas and park lands, is travel cost modeling. This approach assesses individuals' willingness to pay for access to these environmental assets by analyzing the expenses incurred in traveling to them. However, the scope and impact of the travel cost modeling field have been limited by the field's reliance on survey data, which can constrain model applications and inhibit the extrapolation of results to broader contexts or different locations. Survey data can be time consuming and costly to collect, resulting in only a survey sample comprised of relatively few respondents and a small number of sites. In contrast, administrative data offer an inexpensive alternative that facilitates travel cost modeling spanning multiple locations across a larger sample size. Modeling travel costs through administrative data could significantly enhance the field's ability to quantify the benefits of natural resources and better guide policy and management strategies. To that effect, administrative data and survey data can complement one another when modeling travel costs.

In this paper, I compile the largest dataset present in the recreation demand travel cost literature to estimate the willingness to pay for overnight camping trips across the State of Arizona. This novel dataset combines recreation.gov administrative data with hand-collected campsite characteristics data, manually-computed trip data, and zip code-level demographic data to estimate the willingness to pay for a camping trip and the marginal willingness to pay for specific campsite amenities. The dataset includes over 1.1 million observations across the years 2007 through 2016, covering more than 1,400 campsites in 34 parks and facilities across the state. Because I observe camper zip codes and median zip code incomes rather than exact addresses and personal incomes, a central focus of this paper is evaluating the necessity of individual-level data for accurate travel cost modeling. Ultimately, the computed willingness to pay estimates align with findings from survey-based recreation demand studies, indicating that my administrative data perform comparably to survey data. On average, the per-person-per-night willingness to pay for a campsite in Arizona is approximately \$26, though camping at the Grand Canyon elicits a marginal willingness to pay of approximately \$46 per person per night. These estimates are consistent with results from the survey travel cost literature such as those found in Benson et al. (2013). After estimating willingness to pay, I provide a guide to modeling travel costs using administrative data, with the goal of facilitating the replication or application of my methods to other goods and services.

Because recreation.gov includes sites across the United States, my data-generating process can be replicated at the state or national level. Travel cost modeling with administrative data opens up opportunities for state-by-state comparisons, regional analysis, and national recreational site management. Furthermore, the methodology could extend beyond environmental goods, providing insights into areas such as municipal economic development or tourism by measuring willingness to pay for goods and services through travel costs.

The travel cost literature has traditionally relied on small-scale survey data to estimate willingness to pay (Morey 1981, Feather et al. 1995). Initially, the absence of big data precluded any analysis beyond a handful of recreation sites. However, the field has continued to rely on survey data even as administrative data have become increasingly available. For example, Lupi, Phaneuf, and von Haefen (2020) note that secondary data sources in recreation demand travel cost modeling are of questionable quality and must be rigorously evaluated, while survey data remain the gold standard for data quality. Although primary survey data afford researchers a degree of control over data quality, administrative data can complement the existing literature by allowing for greater geographic and demographic heterogeneity in travel cost modeling. In the context of recreation demand, survey data are constrained by the small number of observations and recreation sites they cover. Thus, the literature's reliance on small datasets has notable drawbacks, especially concerning external validity. This paper provides an exploratory analysis of administrative data use to answer questions where survey data are infeasible. The results indicate that secondary data in travel cost modeling can be more useful and applicable than previously thought. In this way, larger secondary datasets can serve as a valuable complement to the survey data traditionally employed in the travel cost literature.

To illustrate the usefulness of administrative data, consider that the United States Forest Service's (USFS) budget has been increasingly allocated toward wildfire mitigation due to the growing frequency and severity of fires. Consequently, appropriations for campground upkeep and maintenance have diminished, necessitating the use of private companies to manage campsites. As a result, campground use fees have generally risen. To understand how use fees should be set, policymakers must know what campers are willing to pay for campsite access based on site amenities. While survey data could only feasibly address this question at the level of individual sites or small groups of sites, administrative data enable willingness-to-pay analysis at the state or national level. Because my paper examines campsites across Arizona, I am able to provide state-level pricing analysis. I find that current campsite reservation prices are well below the average willingness to pay for a camping trip. Modest price increases could thus significantly subsidize Forest Service campground operations and maintenance costs.

I utilize logit modeling to estimate the willingness to pay for Arizona campsites and selected amenities. Campsite attributes were selected based on having adequate variation in the dataset. My model's amenities include proximity to a locally well-known geographic or geological landmark excluding the Grand Canyon, the presence of dedicated equestrian trails within or adjacent to a camping facility, and the presence of RV parking inclusive of at least one type of hookup. Two of the camping facilities in my dataset lie within two miles of the Grand Canyon rims and offer walking trails to the rims. Because the canyon is a global natural wonder, simply classifying it as a "well-known geological landmark" would be a gross understatement. Instead, I define a separate national park indicator variable for those two facilities.

I define the choice set of each camper to be a camper's chosen site and a site from each of the next four closest camping facilities. While camping facilities in my dataset are scattered across Arizona, many facilities are clustered around certain areas such as the Mogollon Rim or Big Lake in the White Mountains. Grouping facilities and sites by location and proximity is appropriate in this setting since a camper will often choose a site based on an area they would like to visit. For example, it is reasonable to assume that a camper looking to visit the alpine, wet terrain of the White Mountains will not consider low-lying sites in the arid Sonoran Desert. Similarly, a Grand Canyon camper will not want to camp in the White Mountains hours away from the canyon. I avoid forming choice sets based on sites from the same camping facilities because sites within a facility share amenities and characteristics. To accurately assess willingness to pay and avoid correlation among choices, I group choices at the camping facility level rather than the site level.

I find that the raw willingness to pay for a camping trip is approximately \$400 on average. This figure does not account for the number of people on the trip or the length of stay. Because the average camping stay is approximately 3.66 days and the average number of people on a trip is about 4.21, the average per-person to pay for a campsite is about \$26 per day. The marginal willingness to pay for a campsite near the Grand Canyon is approximately \$46 per person per day. Other site amenities, including lake access, equestrian trails, RV parking, and nearby geographic features, elicit marginal willingness to pay values ranging from about \$9.13 to approximately \$22.13 per person per day. Neither the addition or subtraction of a campsite to the choice sets changes the trip cost coefficient, though the campsite amenity coefficients are somewhat sensitive to the number of facilities included in the choice sets. Interacting ZCTA demographic characteristics with campsite amenities reveals that campers from zip code tabulation areas with higher median ages are slightly more willing to pay for sites with lake access, RV parking, and sites near the Grand Canyon, while campers from younger ZCTAs are slightly more willing to pay for equestrian trails and geographic landmarks aside from the Grand Canyon. Campers from predominately Native American zip codes significantly favor Grand Canyon sites relative to campers from predominately Caucasian zip codes.

The remainder of this paper is broken down as follows. Section Two provides a literature review of travel cost modeling. Section Three outlines the paper's methodology, including my data compilation, logit models, and robustness checks. Section Four provides results for my logit specifications and robustness checks. Finally, Section 5 discusses my results in the context of USFS and NPS campground pricing and funding across Arizona.

2 Literature Review

The literature on recreation demand and travel cost models is substantial and dates back decades, with a common theme being the use of relatively small datasets. Notable works in this field include Morey (1981), Train (1998), Feather et al. (1995), and Murdock (2006). Morey (1981) constitutes one of the first papers to extend McFadden's random utility model into a discrete choice maximum likelihood framework for recreation demand, using a dataset of 1,453 ski trips. Train (1998) advanced this model by accounting for heterogeneous consumer tastes through random parameter logit estimation, using data from 962 fishing trips. Feather et al. (1995) focused on a discrete-count model to account for both trip costs and site environmental quality, emphasizing that welfare estimates are directly tied to site quality. Finally, Murdock (2006), while relying on a much larger dataset of 676,000 simulated fishing trips, introduced alternative-specific constants to mitigate bias from unobserved site characteristics. A recurring pattern in this body of work is the reliance on relatively small datasets for empirical testing. In addition to Morey (1981) and Train (1998), Feather et al. (1995) used a dataset of 1,488 anglers. While Murdock's large dataset stands out as an exception, recreation demand is estimated via simulated, not observed, choices. Historically, large-scale datasets were difficult to obtain or prohibitively costly, and concerns about data quality in larger datasets have led researchers to favor surveys, which offer better control over the data generation process but are often limited in scope and scale.

More recently, another paper has utilized administrative data in travel cost modeling. Lloyd-Smith and Becker (2020) estimates welfare losses from campground closures using a Kuhn-Tucker travel cost model. Their dataset includes 187,964 trips made by approximately 71,000 individuals over the year 2015. My paper builds on Lloyd-Smith and Becker (2020) in several ways. While Lloyd-Smith and Becker (2020) evaluate welfare losses associated with campground closures, I estimate marginal willingness to pay for campsite amenities and apply my estimates to Forest Service operations and maintenance costs. My dataset is larger, encompasses a greater number of campsites, and includes more years of analysis. Though my dataset is larger, it is less granular; I do not observe individual-level data such as name, personal address, or personal income. Rather, I observe de-identified camper information at the zip code level, including median zip code income. Lloyd-Smith and Beck possess a rare administrative dataset that includes individual information in a similar manner to survey data; such granularity is not typically afforded to administrative data. Given this limitation, a key aspect of my paper is determining whether accurate travel cost models can be constructed without the individual-level data which are typically present in survey datasets. One advantage of my data is that I can control for temperature in my model, while Lloyd-Smith and Becker's Kuhn-Tucker approach precludes weather analysis. Unsurprisingly, temperature is a significant determinant of camping trips in the hot and arid State of Arizona. Finally, the multi-year nature of my dataset also facilitates temporal trend analysis.

One branch of the recreation demand literature pays special attention to the demand model's travel and time cost components. Smith et al. (1983) notes that vehicle costs are not the sole determinant of travel costs - there exists a travel time component that significantly influences demand estimates. The authors also account for wages when evaluating the value of time, though income did not consistently have a statistically significant effect on recreation demand. My dataset includes median income per zip code, so controlling for wages in my recreation demand model is ultimately warranted given income could still easily influence the willingness to drive to and pay for a campsite. Hagerty and Moeltner (2005) specify alternative estimation strategies for vehicular travel costs aside from the typical constant per-mile cost strategy utilized by most recreation papers. One strategy involves including an automotive cost term estimated from engineering models that captures vehicular wear-and-tear from trips. The second strategy refines the first specification by adding car-specific attributes into the travel cost model. Ultimately, the addition of specific vehicle attributes did not produce statistically significant differences in travel cost calculations. However, vehicular wear-and-tear accrued from trips is an important consideration in travel cost estimation – one which my model will account for.

In general, travel cost and recreation demand literature are limited by a reliance on survey data. In addition to the historical scarcity of large recreation datasets, economists tend to utilize survey data in recreation demand models for accuracy and control over the data generating process. In Lupi, Phaneuf, and von Haefen's "Best Practices for Implementing Recreation Demand Models" (2020), the authors warn readers that administrative data quality must be scrutinized and representative approaches such as surveys should be used when possible for data quality control purposes. However, surveys constitute a costly and time-consuming method of data collection. Thus, the drawback of survey reliance is that travel cost papers generally cannot be extrapolated beyond their limited sample and sampling area. Large administrative data are not faultless; Lupi et al. (2020) correctly asserts that secondary data require careful cleaning. Although both survey and secondary data present distinct challenges, well-managed administrative datasets offer the advantage of addressing broader research questions with potentially greater external validity when appropriately cleaned and verified.

Survey limitations are not unique to the travel cost literature. Initially, hedonic price models were significantly fettered by a lack of larger data. The hedonic model's origins can be traced back to Andrew Court in 1939 (Goodman 1998), but the first instances of large-scale data in the hedonic setting did not occur until the late 1970s and early 1980s. Rosen (1974) expanded upon the hedonic framework with a market analysis component, and the existence of larger datasets in the 1970s allowed papers such as Harrison and Rubinfeld (1978) and Palmquist (1984) to forge ahead in the literature with housing datasets spanning entire metropolitan areas and even multiple cities. Prior to this advancement, hedonic modeling relied on survey data in works such as Ridker and Henning (1967) and Kain and Quigley (1970). As with travel cost modeling, survey data limited hedonic models to small sample areas, small sample sizes, or both. Unlike travel cost modeling, hedonic modeling has expanded beyond only using survey data. To be clear, this paper is not arguing against the use of survey data. A well-executed survey design and implementation strategy can offer a high degree of flexibility and precision in data collection. Rather, I argue that administrative data can compliment the existing survey literature by expanding the size and scope of travel cost analysis beyond micro-level data.

My paper contributes to the body of recreation demand travel cost literature by constructing and successfully employing large administrative data in travel cost count model analysis. In contrast to previous work in travel cost literature, I have a dataset spanning over 1.1 million observations across 1,433 campsites within 34 large camping facilities and parks across the entire State of Arizona over a multi-year period. Moreover, my methodology can be replicated across any State in the country. My results indicate that precise individual-level data may not be required to accurately estimate willingness to pay. Just as Harrison and Rubinfeld (1978) expanded hedonic pricing to an entire metropolitan area, my paper constitutes one of the first attempts to conduct travel cost modeling on a much larger scale than previous works. Accounting for the fact that Lloyd-Smith and Becker use a Khun Tucker travel cost model, my paper is the first paper to use employ a large-scale count model in the travel cost literature. Because of my extensive data, my results can draw larger conclusions on the funding of the Forest Service and national park campsites where previous literature could not answer statewide or nationwide research questions.

Table 1
Comparing the Survey Literature to Administrative Data

Paper	**Year**	**Sample Size**	**Number of Sites**	**Analysis Timeframe**
Morey	1981	1,453	15	1967-1968
Train	1998	962	59	1992-1993
Feather et al.	1995	1,488	Unspecified	1989
Benson et al.	2013	903	5	2006
English et al.	2018	41,708	83	2012-2013
This paper	2024	1,127,784	1,433	2007-2016

Notes: Administrative data allow a substantially greater sample size, site count, and analysis time frame than survey data. English et al. (2018) appears to produce the largest survey dataset in the literature, but this paper still produces a significantly larger dataset at a lower time with no financial cost.

3 Methodology

To estimate the willingness to pay for campsite characteristics, I employ a large administrative dataset comprised of campground reservations, trip cost data, weather data, and demographics. In my dataset, campsite characteristics are shared at the camping facility level. To ensure sufficient variation in amenity characteristics within each choice set, I define a camper's choice set as the chosen campsite and the four geographically closest alternative campsites, with the condition that no two campsites in the set belong to the same camping facility. Next, I estimate a random parameter logit model where the indirect utility function accounts for the campsite amenities and reservation data. Finally, marginal willingness to pay is calculated as the quotient of the coefficient on a given campsite characteristic over the coefficient on the trip cost term.

3.1 Data

In contrast to other papers in the travel cost literature that rely on small-scale surveys encompassing few sites, I utilize a comprehensive administrative dataset for my analysis. My dataset contains over 1.1 million observations across 34 unique camping facilities, representing a total of 1,433 individual campsites. Compared to survey-based approaches, this administrative data strategy minimizes extrapolation by including hundreds of campsites, thereby enhancing external validity. To estimate willingness to pay, I collected data on campsite reservations, characteristics, travel costs, weather, and ZCTA-level demographic information. I elaborate on the dataset in the following subsections.

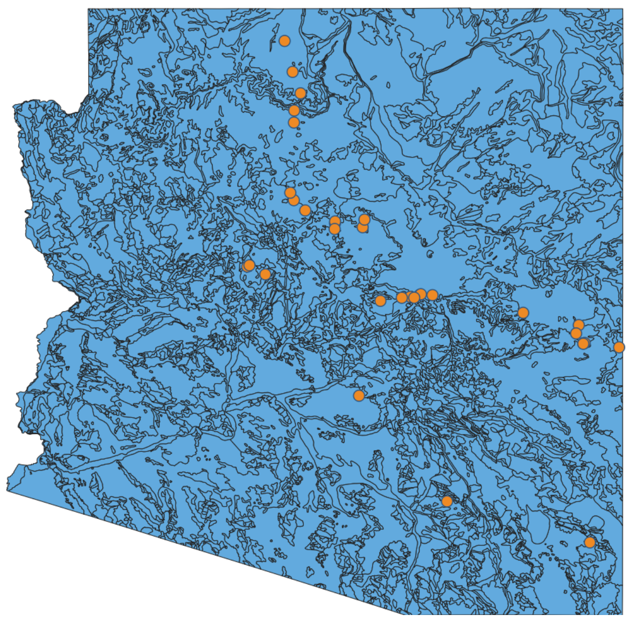
3.1.1 Campsite Data

Data on reservations and campsite use fees from years 2007 through 2016 come from the Recreation.gov API. Each observation in this dataset contains a customer zip code as well as the latitude and longitude of a reserved site. I do not observe the mode of transportation a camper takes to arrive at a campsite. To exclude trips that may involve both flight and driving, I exclude all customer zip codes that lie outside of Arizona. Additionally, I exclude non-camping reservations listed in the original dataset, including reservations for guided tours and cabins. Finally, I limit my analysis to non-group sites, which tend to be significantly more expensive due to their ability to accommodate large groups of people. In sum, my analysis is focused on Arizona residents accessing Arizona campsites by car. Although not all campsites in the state require a Recreation.gov reservation, nearly all can be found within the regions included in my sample.

Recreation.gov also provided my campsite characteristics data. I manually constructed a facility-level characteristics dataset by searching each camping facility on Recreation.gov. I used a combination of the site's written descriptions and site pictures to verify the presence of various campground amenities at each facility.

Figure 1 displays a map of all camping facilities within my dataset. The campgrounds in the sample are distributed across a substantial elevation range, from 3,200 feet to over 9,000 feet above sea level. This dataset includes campsites located in Arizona's White Mountains, along the Grand Canyon, throughout red rock country, within Tonto National Forest, and among the southern Arizona sky island peaks. Overall, the dataset effectively represents Arizona's varied terrain, further mitigating challenges associated with statewide extrapolation.

Figure 1
Camping Facilities in Arizona



Notes: This figure displays all camping facilities within the sample. Because some of the facilities have nearly identical coordinates, six of the facilities appear to overlap with other campgrounds displayed in the figure.

3.1.2 Travel Cost Data

Because individual camper addresses are not available, I estimate travel costs by calculating the centroid coordinates for each Zip Code Tabulation Area (ZCTA) in Arizona. For each reservation-holding camper, the calculated car route represents the optimal route from the centroid of their associated ZCTA to the campground coordinates. To calculate over 1 million routes, I use a locally-hosted instance of Open Route Service, an open-source routing software developed by Heidelberg University and used by organizations such as The New York Times. Although Google Maps uses a proprietary routing algorithm, Open Route Service generates very similar routes, distances, and travel times, as both services optimize routes based on the road network.

To calculate camper travel costs, I utilize the following metric from English et al. (2018):

$$c_{ijt} = \frac{p_{it}^{D} D_{ij}}{n} + p_{it}^{T} T_{ij}$$

p_{it}^{D} accounts for the per-kilometer vehicle depreciation cost and per-kilometer maintenance cost. p_{it}^{T} constitutes the hourly cost of travel and is equal to 1/3 of the median zip code tabulation area income divided by 2,080 hours worked in a year (English et al. 2018). D_{ij} represents the travel distance in kilometers, while T_{ij} represents the hours spent traveling. Finally, n represents the number of people on the trip. Per-kilometer vehicle depreciation cost and per-kilometer maintenance costs are calculated from American Automobile Association (AAA) data. Because I do not observe the type of vehicle being driven in my dataset, I use AAA's average depreciation and maintenance cost figures when computing per-kilometer costs. The overall trip cost for reservation i at site j at time t is simply the sum the of the per-person site use fee, p_{ijt}^{U}, and the incurred travel cost:

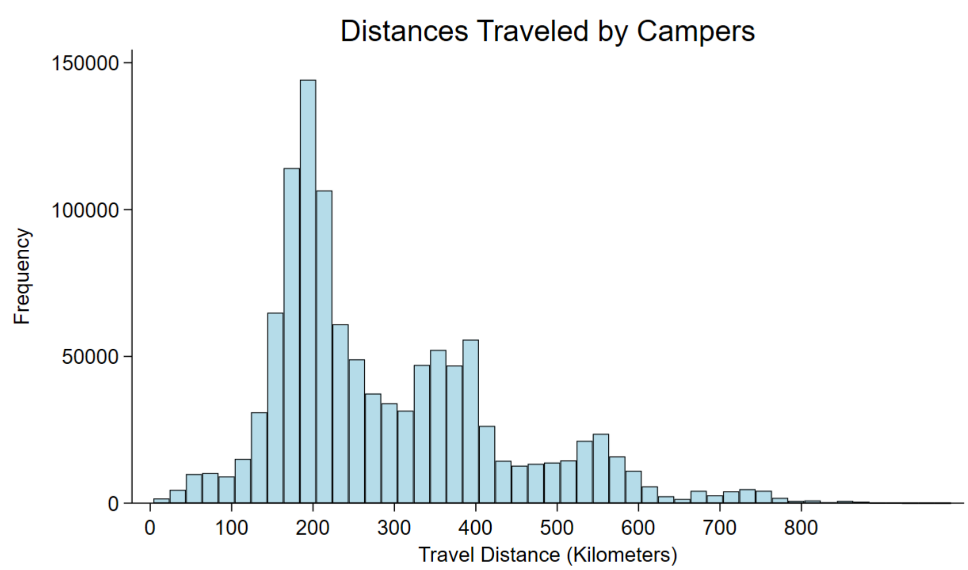
$$p_{ijt}^{Trip} = \frac{p_{ijt}^{U}}{n} + c_{ijt}$$

One caveat of my dataset involves price variation within camping facilities and even within same sites. Many of the campsite reservations vary in price with each other at the same campsite due to unobserved or possibly random factors. Much of the same-site price variation is insignificant - on the order of cents or a few dollars. This may suggest special price discounts or special campsite requirements that would slightly lower or raise the price of a site. However, such unexplained variation still impedes the assignment of prices to non-chosen alternative sites. Camper addresses do not influence or determine reservation prices, indicating that the price variation is not correlated with travel cost. To be certain, I compute the correlation between the site fee and an individual's travel cost and find an insignificant value of -0.0587. Figure 2 shows a scatter plot of observed prices the travel costs associated with those observations. The lack of a trend in the scatter plot combined with the low correlation value lends strength to the assumption that site prices do not vary based on travel costs.

Figure 2
A Scatter Plot of Reservation Prices and Travel Costs

Notes: No strong association between site prices and travel costs exist within the sample.

Figure 3
The Distribution of Travel Distances



Notes: This figure displays the distribution of distances traveled by campers within the sample.

In order to assign prices to the non-chosen alternatives, I compute and impute the mean per-person site fee for each camping facility. In any given choice set, the chosen site's fee is represented by an observed per-person site fee, while the non-chosen alternatives take the mean per-person site fee for the facilities in which they are located.

My travel cost computation rests on the assumption that ZCTA-level income data can be utilized to accurately estimate willingness to pay. Certainly, a degree of measurement error is introduced when data granularity is lost, and most travel cost models utilize individual-level income data rather than median zip code incomes. My computed willingness to pay values found in the results section align with the existing survey literature and, consequently, make a strong argument that the degree of measurement error introduced by ZCTA-level income data is small. I attribute my paper's results to the fact that ZCTA-level income data are still granular enough to capture spatial income variation, while county- or State-level income data lose a significant amount of that variation. I discuss this at greater length in the results section.

3.1.3 Other Data

My model's weather data are sourced from the NOAA. Campsites were matched with weather stations based on proximity and elevation. As such, the weather at each campsite is drawn from the corresponding weather station that is closest in both distance and elevation to the campsite. Unfortunately, most of the weather stations only recorded temperature data, without recording precipitation or wind speeds. However, temperature is the most important weather metric in a region like Arizona, which receives little rainfall and experiences extreme temperatures. Furthermore, temperature can serve as a proxy for inclement weather, as significantly cooler temperatures than normal in Arizona are nearly always associated with rainfall.

Finally, I collect demographic data from the Census. Because I only observe reservations at the zip code level, my demographic information is collected at the Zip Code Tabulation Area (ZCTA) level. My demographic data include ZCTA median income, the percent of residents within each ZCTA that possess a bachelors degree or higher, and ZCTA median age.

In sum, I have 1,127,784 observations. Each observation represents either a chosen campsite or an alternative campsite and includes campsite characteristics, trip cost data, temperature data, and demographic information about the reservation holder's associated ZCTA. Table 2 displays summary statistics for both ZCTA demographic data and data related to site fees, travel costs, trip costs, stay length, and the number of people on each trip. I include information and summary statistics on my chosen campsite characteristics in the estimation strategy subsection.

The raw site fees listed in the summary statistics are not per-person site fees; trip costs, however, are calculated based on per-person site fees. Notably, the average number of people per trip is approximately four, suggesting that the average camper in Arizona may choose to travel with family or friends. The average stay length exceeds two days, indicating that many individuals in the sample chose to take time off of work to camp. Analyzing trip start dates reveals that the average start date falls between Thursday and Friday, with Friday incurring more camping reservations than Thursday. The ZCTA demographic summary statistics indicate a rich variation in median income and education levels. The median ZCTA age suggests that many of the campers in the dataset may be middle-aged, though the absence of micro-level consumer data limits further conclusions.

Table 2
Summary Statistics - Demographic Information and Trip Characteristics

	Mean	Std. Dev.
Percentage of Residents with Bachelors Degrees or Higher	36.612	14.135
Median Income	$77,738	$24,359
Median Age	41.19	9.52
Raw Site Fees (Total Paid)	$46	$31.95
Travel Cost	$159.27	$87.59
Trip Cost	$174.17	$89.20
Average Trip Length	3.66	1.82
Average Number of People	4.21	2.39

Notes: This table reports summary statistics for demographic information and trip characteristics. All demographic data are at the zip code tabulation area (ZCTA) level. Raw site fees do not account for the number of people per trip nor the trip length. Notably, the mean ZCTA-level median income is relatively high with considerable variation. Relatively high travel costs are reflective of campsites generally being located far away of the State's major population centers.

3.2 Estimation Strategy

To estimate willingness to pay for campsite amenities, I begin by defining a camper's choice set. It is unrealistic to assume that a consumer will consider all possible Arizona campsites as potential camping options; a camper might only consider staying in a particular region or at a specific landmark. For example, an individual seeking to camp near the Grand Canyon is unlikely to consider a site that is hours away from the canyon in southern Arizona. To address this issue, I define each camper's choice set as their observed chosen site as well as one site each from the next four closest campsites to the chosen site. Adding additional campsites to the choice sets would unnecessarily force faraway campsites into some of the choice sets. For example, if a camper's chosen site and the next four closest campsites are all located on Mount Lemmon, a potential sixth campsite would lie in an entirely different mountain range located a few hours away from the other campsites. Sensitivity analysis is included in the results section of this paper, where I test whether the addition or subtraction of a campsite from each choice set impacts the results. I compute the distances between campsites using the Haversine distance equation. There are 34 unique choice sets within my dataset.

Ultimately, I chose the following campground characteristics for my logit model:

1. The presence of horseback trails at a campsite.
2. The presence of a lake at campsite.
3. The presence of RV parking at a campsite with at least one type of RV hookup.
4. The presence of a significant landmark near a campsite, excluding the Grand Canyon.
5. The presence of the Grand Canyon near a campsite.

For a lake to be present at a campsite, the camp must have a means of accessing the lake, such as a trail or road, and the lake must be within two miles of the campsite. All lakes in the sample are sufficiently large to support fishing and boating activities. Similarly, notable landmarks must be within two miles of the campsite and include trails or roads providing access to the landmark. Campgrounds near significant landmarks are defined as sites located on or within two miles of the Mogollon Rim, Humphreys Peak, Mount Baldy, and Mount Lemmon. The Mogollon Rim is a significant and very well-known geological formation that forms the Colorado Plateau and includes the popular red rock peaks and canyons of Sedona, Arizona. Humphreys Peak and Mount Baldy are the State's two highest mountains and are very popular hiking destinations; Humphreys Peak offers Arizona's only legally climbable alpine tundra. Finally, Mount Lemmon is the most well-known and prominent sky island peak near Tucson, featuring a number of amenities on the mountain including the country's southernmost ski area near the peak and desert cacti forests in the lower regions. These areas see disproportionate use relative to the other areas of the State. Horseback trails are carefully defined as trails that are either dedicated to or primarily used for horseback riding. Similarly, I constrain the definition of RV parking to sites that have dedicated spots at campsites with at least one type of hookup available.

The Grand Canyon requires its own attribute variable. Because the canyon is both a national park and one of the seven natural wonders of the world, Grand Canyon campgrounds are highly desired. The canyon's campsites in my sample are so close in proximity to the park that they have walking trails to the rims of the canyon. Therefore, simply including the Grand Canyon with the other defined landmarks would define the landmark variable too broadly and produce an upward bias in the landmark coefficient. To avoid this issue, I create a separate characteristic indicator variable for the campsites located near the canyon.

Campsite attributes were selected and defined based on their variation within and between choice sets. Within my sample years, campsites in Arizona nearly universally permitted campfires and included campfire amenities such as fire rings. Additionally, nearly all sites feature bathroom facilities and hiking trails, while many include grills and stoves. To ensure adequate variation for logit analysis, I selected characteristics that exhibit sufficient variation within each choice set as well as across choice sets. Table 3 displays summary statistics for the chosen campsite characteristics. Note that all characteristics occur in less than fifty percent of the sites, thereby allowing for sufficient levels of characteristic variation in the choice sets. Notably, none of the campsites in my sample feature all five characteristics simultaneously.

Table 3
Summary Statistics - Campsite Amenities

	Mean	Std. Dev.
Horseback Trails	0.2828	0.4504
RV Parking	0.4312	0.4953
Lake Access	0.4241	0.4942
Landmark Near	0.4738	0.4993
Grand Canyon Site	0.0776	0.2675

Notes: This table presents summary statistics for all chosen campsite amenities in the dataset. As the campsite amenities are represented by indicator variables, each mean in the table reflects the proportion of camping facilities that feature a given characteristic. None of the selected characteristics are present in more than 47.38% of the facilities in the sample.

To estimate the marginal willingness to pay for campsite characteristics, I utilize a logit model. A given camper's indirect utility function can be represented as the following base specification:

$$V_{ijt} = \beta X_j + \alpha p_{ijt}^{Trip} + \omega W_{jt} + \gamma_y + \nu_m + \xi_{ijt} \tag{3.2.1}$$

Here, i represents a particular reservation holder, while j represents a given campsite and t represents the start date of a campsite reservation. Each reservation holder's indirect utility is a function of campsite characteristics (X_j), trip cost (p_{ijt}^{Trip}), weather (W_{jt}), and an individual-specific error term (ξ_{ijt}) that is assumed to be normally distributed. Additionally, I include year and month fixed effects (γ_y and ν_m, respectively). This base-form specification does not include ZCTA-level demographic information; for a specification that includes the demographic information, see Section 3.3.

From the camper's indirect utility function, the conditional probability that the camper chooses site j can be represented as follows:

$$L_{ijt}(\beta|\xi_{ijt}) = \frac{exp(V_{ijt})}{\Sigma_{K_n} exp(V_{ikt})} \tag{3.2.2}$$

The main difference between equation 3.3.2 and the conditional probability of a typical logit model with a single choice set is the presence of different possible choice sets, dependent on a camper's chosen destination.

Rather than j representing the chosen site within a singular choice set, j denotes a chosen site from one of 34 possible choice sets. Let $1 \leq n \leq 34$ denote a choice set index. Then the set of all campsites in a choice set K must be indexed by n in the conditional probability equation.

Finally, the marginal willingness to pay (MWTP) for a characteristic C can be written as follows:

$$WTP_c = -\frac{\beta_c}{\alpha} \tag{3.2.3}$$

In other words, the MWTP for a campsite characteristic is the ratio of the coefficient on that characteristic to the coefficient on trip cost. Logit estimation was selected over other modeling strategies, such as probit, due to the relatively straightforward interpretation of model coefficients and the ease of deriving willingness to pay. Because all trips are overnight stays and most trips involve multiple campers, I subsequently compute the per-person, per-day MWTP for the selected characteristics. The average daily MWTP per person can be represented as the equation below:

$$WTP_{cd} = -\frac{\beta_c}{\alpha dp} \tag{3.2.4}$$

Here, d denotes the average stay length in days, and p represents the average number of people per trip. Thus, the average daily MWTP for a campsite characteristic is approximated by dividing the MWTP for that characteristic by the trip length and number of people. Finally, the overall willingness to pay for a camping trip is denoted as follows:

$$WTP_{trip} = -\frac{1}{\alpha} \tag{3.2.5}$$

3.3 Heterogeneous Impacts & A Robustness Check Overview

Although individual-level demographic information is not available, I can conduct heterogeneity analysis at the Zip Code Tabulation Area (ZCTA) level. The camper's indirect utility function can be represented as follows:

$$V_{ijt} = \beta X_j + \zeta Z_i X_j + \alpha p_{ijt}^{Trip} + \omega W_{jt} + \gamma_y + \nu_m + \xi_{ijt} \tag{3.3.1}$$

The primary distinction between equations 3.3.1 and 3.2.1 lies in the inclusion of interactions between ZCTA demographic variables and site characteristics, denoted as $Z_i X_j$. Given the issues with multicollinearity and the potential bias introduced by ZCTA income and education interactions, I exclude both income and education from the heterogeneity model. Including income in the model would bias the travel cost term due to its direct relationship with income, while education is correlated with income, potentially influencing the model in a similar way. Thus, changes in marginal willingness to pay (MWTP) based on demographic information are estimated using ZCTA median age and majority ethnicity data. To account for ZCTA-level ethnic composition, I create a series of dummy variables, assigning a value of one if a ZCTA is either majority-Hispanic or majority-Native American, with majority-White ZCTAs serving as the baseline. All ZCTAs in my sample reflect majority ethnicities within these three categories. The change in MWTP for a campground characteristic, given a specific ZCTA demographic variable, can be calculated as follows:

$$WTP_c = -\frac{\zeta_{ci}}{\alpha} \tag{3.3.2}$$

The interpretation of the marginal willingness to pay (MWTP) variable is contingent on the unit of the demographic variable. For instance, if we interact ZCTA median age with the lake access variable and equation 3.3.2 yields a value of 0.4, the model indicates that an increase of 1 year in ZCTA median age would raise the WTP for lake access by $0.40. This approach enhances the robustness of the main findings by providing insights into the influence of demographic factors on camper preferences. While the analysis relies on secondary data, it can still offer meaningful insights into demographic heterogeneity, contributing to a more comprehensive understanding of the factors influencing willingness to pay in the context of camping in Arizona. Furthermore, many ZCTAs in Arizona tend to be relatively ethnically homogeneous; thus, interactions with ZCTA majority ethnicity may reveal distinct patterns in MWTP across different ethnic groups.

A potential criticism of my outlined estimation strategy concerns the definition of the choice set. It is reasonable to assume that a camper aiming to visit a specific destination will primarily consider sites in close proximity to that destination; however, the choice set is defined somewhat arbitrarily. In my base specification and heterogeneity analysis, each camper's choice set includes their selected campsite and the next four nearest facilities. This choice set assignment could raise the following question: might a camper realistically consider only three alternatives, or is there a fifth nearest site that could be deemed a plausible alternative? While some researcher discretion is necessary in defining a choice set, a well-defined set should yield a trip cost coefficient that is relatively insensitive to alterations. To assess the robustness of my choice set definition, I estimate two additional base-specification models: the first excludes the furthest alternative from every choice set, while the second includes the fifth closest site in every choice set. While we may anticipate some sensitivity in the coefficients related to campsite amenities, the trip cost coefficients should remain fairly consistent across a robust choice set definition.

To investigate the impact of different levels of income data on travel cost computation, I modify my travel cost calculation strategy by substituting ZCTA-level income data with county- and state-level income data. Significant deviations in results, particularly concerning the travel cost term, could suggest that losing too much income information adversely affects travel cost modeling. Although the absence of individual-level survey income data may not hinder my analysis — given that ZCTA-level income data still reflect rich income variation - aggregating income data beyond the zip code level may hinder accurate willingness to pay calculations by substantially diminishing income variation and introducing considerable measurement error.

Another potential concern regarding my outlined estimation strategy pertains to campground capacities. If certain camping facilities are routinely crowded or consistently reserved, my model may understate the true willingness to pay for those sites and their amenities. For instance, the demand for campsites near the Grand Canyon may significantly exceed the available supply. To address this issue, I collect data on camping facility capacities and compute the total number of reservations at each facility on a daily basis. I then generate a capacity variable that represents the ratio of reservations to available sites at each camping facility. Finally, I estimate specifications that exclude camping facilities exceeding certain capacity thresholds. For example, a facility that is 90% reserved on one day would be excluded from the analysis, whereas it would be included if only 50% of its sites were reserved on another day.

4 Results

The results section is organized as follows: Subsection 3.1 presents the main findings, Subsection 3.2 covers the heterogeneity analysis, and Subsection 3.3 examines the robustness checks. Overall, my findings suggest that the administrative data approach to travel cost modeling is as effective as the traditional survey-based approach.

4.1 Main Results

I begin by estimating willingness to pay without incorporating heterogeneous elements into the logit specification. In this model, the probability that a camper selects a particular campsite over alternatives depends on the site's amenities, trip costs, and weather conditions, without yet accounting for demographic information at the ZCTA level. Table 4 presents the results of the baseline model's logit estimation. For each campsite characteristic, the coefficients represent the average marginal utility a camper derives from the presence of that feature. Positive coefficients suggest that a characteristic is valued by campers, while negative coefficients indicate disutility associated with that feature. As expected, the selected characteristics exhibit positive and significant coefficients—indicating, for instance, that amenities such as proximity to a lake or important landmarks are desirable. Additionally, higher trip costs result in disutility; all else being equal, campers prefer closer, less expensive sites over more distant, costlier alternatives.

Grand Canyon sites are sought after and incur high and positive marginal utilities compared to similar campgrounds located further away from the canyon. In this context, sites are classified as being near the Grand Canyon if they lie within two miles of one of the canyon's rims. Given their proximity to views and trails of a globally recognized natural landmark, it is unsurprising that campers derive considerable utility from staying close to the Grand Canyon. Additionally, increases in average daily temperatures generate positive utility, as most campsites in the sample are located in Arizona's higher-altitude regions. Overall, the results in Table 4 suggest that the administrative data perform well in the random parameter logit estimation. All selected characteristics are statistically significant, with very small standard errors. Standard errors are clustered at the reservation ID level.

Table 4
First Stage Results - Marginal Utilities

	Coefficients	Std. Err.
Horseback Trails	0.4770	0.00796
RV Parking	0.6537	0.00648
Lake Access	0.8421	0.00738
Landmark Near	0.3526	0.00722
Grand Canyon Site	1.7978	0.01200
Trip Cost	-0.0025	0.00004
Avg. Temperature	0.0082	0.00046
Stay Length	0.0454	0.00152
Site Elevation	-0.0002	0.000003

Notes: This table reports the coefficients and associated standard errors of the baseline logit model. The campground amenity coefficients in this table will be divided by the negative of the trip cost coefficient to compute average willingness to pay for each of the campground amenities. See the following table for all willingness to pay estimates.

Marginal willingness to pay values are displayed in Table 5. The first column provides the raw MWTP values without adjusting for trip length or group size. The second column reports bootstrapped standard errors for the raw MWTP values. Finally, the third column reports the per-person, per-day MWTP values, calculated using an average trip length of 3.66 days and an average group size of 4.21 people. On average, an individual camper is willing to pay approximately \$26 per day to camp. Proximity to the Grand Canyon significantly increases this amount to nearly \$47 per day, holding other factors constant. The MWTP values for other campsite characteristics appear plausible, ranging from roughly \$9 to \$22 per day. On average, lake access is valued higher than amenities such as horseback trails or dedicated RV parking, likely due to the additional recreational opportunities, including boating and fishing, available at all lakes in the sample.

Table 5
Second Stage Results - Willingness to Pay

	Raw WTP	Bootstrap Std. Errors	Per-Person Per-Day WTP
Horseback Trails	\$189.38	\$4.52	\$12.29
RV Parking	\$261.38	\$4.65	\$16.96
Lake Access	\$339.60	\$7.17	\$22.03
Landmark Near	\$139.37	\$4.07	\$9.04
Grand Canyon Site	\$723.58	\$12.74	\$46.96
Total Trip	\$400.25	\$6.66	\$25.99

Notes: The first column of this table reports willingness to pay (WTP) values without accounting for reservation length or the number of people. The second column reports the bootstrapped standard errors of the raw WTP estimates. Finally, the third column divides the first column values by average trip length and the average number of people on a reservation. All WTP values should be treated as averages.

The willingness to pay (WTP) estimates derived from my model align closely with those found in the existing survey and recreation travel cost literature. Menz and Mullen (1981) report a per-person, per-day WTP of approximately \$21.60, or \$56.32 in 2015 dollars. Similarly, Lloyd-Smith and Becker (2020) estimate that the average welfare value of a camping trip generally falls between \$40 and \$89 per camper, with specific estimates ranging from \$23 to \$46 and \$63 to \$144, depending on the site. In my model, without adjusting for trip length, the average WTP for a camping trip is approximately \$95 per person. Benson et al. (2013) report a per-person, per-trip benefit of \$276 for visitors to Yellowstone National Park, with an average stay of 8.21 days. This translates to a per-person, per-day WTP of \$33.62, with values ranging from \$12.54 to \$86.97, depending on the sample subgroup. My WTP estimates for camping near the Grand Canyon are comparable: the per-person, per-trip WTP is approximately \$171.16, while the per-person, per-day WTP is around \$46.76. Overall, the estimates from my secondary dataset are consistent with those reported in the broader travel cost literature.

4.2 Heterogeneity Analysis Results

The heterogeneity analysis extends the base specification by incorporating ZCTA-level demographic dummy variables. All campsite characteristics are interacted with demographic factors, specifically income, age and race. Although my data include information on ZCTA income and education levels, I exclude interactions with these variables due to concerns about multicollinearity and potential confounding effects with the travel cost variable. The majority ethnic groups in each ZCTA are either White, Hispanic, or Native American. Given that most ZCTAs are predominantly White, I create dummy variables to indicate whether a ZCTA is primarily Hispanic or primarily Native American. Table 6 presents the marginal utilities with all demographic interactions included.

Table 6
Marginal Utilities with Demographic Interactions

	Coefficients	Std. Err.
Horseback Trails	0.8757	0.03547
RV Parking	0.4829	0.03010
Lake Access	0.6968	0.02706
Landmark Near	0.4932	0.02639
Grand Canyon Site	1.4936	0.04906
Trip Cost	-0.0026	0.00004
Avg. Temperature	0.0057	0.00030
Stay Length	0.0451	0.00153
Site Elevation	-0.0002	0.000003
Lake x Age	0.0039	0.00062
Horse x Age	-0.0094	0.00082
RV x Age	0.0041	0.00070
Canyon x Age	0.0075	0.00112
Landmark x Age	-0.0037	0.00061
Lake x Hispanic	-0.1720	0.01934
RV x Hispanic	0.0908	0.02142
Horse x Hispanic	-0.1037	0.02588
Canyon x Hispanic	-0.0474	0.04085
Landmark x Hispanic	0.0456	0.01742
Lake x Native	0.3362	0.08757
RV x Native	-0.2482	0.10849
Horse x Native	-0.7638	0.12541
Canyon x Native	0.6535	0.13460
Landmark x Native	-0.1450	0.10080

This table reports the marginal utility coefficients for campground amenities and interactions between campground amenities and demographic information. The age interactions measure changes in the marginal utility of an amenity for each one year increase in ZCTA median age.

Most interactions yield small but statistically-significant results. On average, customers from predominately Hispanic ZCTAs incur greater utility for sites with RV parking and natural landmarks than their customers from predominately White ZCTAs. Meanwhile, customers from predominately Native American ZCTAs tend to favor lake access and Grand Canyon sites relative to customers from predominately White ZCTAs. Customers from predominately White ZCTAs seem to greatly prefer campsites with horseback trails compared to customers from either predominately Hispanic or predominately Native ZCTAs. Campers from older ZCTAs appear to favor campsites with RV parking and lake access, while campers from younger ZCTAs incur greater utility from horseback trails and local landmarks.

The travel cost coefficient shows minimal variation; in the base specification, I estimate a coefficient of -0.0025, while the heterogeneity model yields a coefficient of -0.0026. I exclude both income and education interactions from this model due to the significant bias they introduce. Since income is inherently part of the trip cost term, reintroducing income into the logit model would bias the trip cost coefficient downward. Although the inclusion of education interactions does not affect the trip cost parameter as profoundly as income interactions, the inherent correlation between education and income leads to biased estimates of the education interaction coefficients.

Table 7 provides the computed willingness to pay (WTP) values for the statistically significant interactions.Several interactions yield notably large effects. In particular, customers from predominately Native American ZCTAs demonstrate a significantly higher willingness to pay for campsites near the Grand Canyon compared to those from predominately White ZCTAs. This heightened valuation may stem from the Grand Canyon's partial management and operation by Native American tribes, as well as its designation as a sacred site for several tribes across Arizona. Conversely, customers from predominately Native American ZCTAs exhibit a markedly lower willingness to pay for campsites featuring dedicated horseback trails when compared to their White counterparts. This discrepancy may be attributed to the fact that many tribes, including the Navajo Nation, maintain their own horseback trail systems, often providing low-cost or free access to tribal members. Consequently, it seems unlikely that consumers would opt for a pricier trip to a horseback trail campsite when more affordable alternatives are readily available nearby.

Table 7
Willingness to Pay Interaction Computations

	Change in Raw WTP	Bootstrap Std. Errors	Change in PPPD WTP
Lake x Age	$1.51	$0.24	$0.10
Horse x Age	-$3.71	$0.32	-$0.24
RV x Age	$1.22	$0.18	$0.08
Canyon x Age	$1.46	$0.27	$0.09
Landmark x Age	-$0.89	$0.15	-$0.06
Lake x Hispanic	-$38.02	$4.96	-$2.47
RV x Hispanic	$36.41	$5.19	$2.29
Canyon x Hispanic	-$18.99	$9.00	-$1.23
Landmark x Hispanic	$18.85	$4.40	$1.22
Lake x Native	$89.98	$18.90	$5.84
Horse x Native	-$143.59	$28.15	-$9.32
Canyon x Native	$144.39	$29.27	$9.37

This table reports changes in raw and per-person per-day willingness to pay to all statistically significant interaction terms from Table 6. All interactions should be interpreted at the Zip Code Tabulation Area level.

4.3 Robustness Checks

In this subsection, I first test the robustness of my choice set creation strategy before testing my value of time definition. We expect that the trip cost parameter should not significantly vary under a robust choice set definition, even with slight alterations to the choice set. To test my choice set creation strategy, I remove the furthest alternative from my choice sets in one specification and add a fifth alternative to the choice sets in a second specification. Table 8 displays my model results after subtracting the fourth alternative from each choice set; Table 9 showcases my results after adding a fifth alternative to each choice set. While the campsite characteristics are somewhat sensitive to my choice set definition, the coefficient on trip cost does not significantly change. Finally, I investigate whether campground capacity influences my results by examining instances where campgrounds reach or approach full capacity based on reservations and excluding campgrounds on specific days when they are nearly full.

Table 8
Marginal Utilities with the Furthest Alternative Removed from Each Choice Set

	Coefficients	Std. Err.
Horseback Trails	0.5894	0.00835
RV Parking	1.0029	0.00739
Lake Access	0.2993	0.00879
Landmark Near	0.4564	0.00719
Grand Canyon Site	0.83556	0.01423
Trip Cost	-0.0022	0.00004
Avg. Temperature	0.0028	0.00031
Year	-0.0127	0.00108
Month	-0.0078	0.00164
Stay Length	0.0388	0.00160
Site Elevation	-0.0001	0.000003

This table presents the marginal utility coefficients for campground amenities following the removal of the furthest alternative from each choice set. The trip cost coefficient is estimated at -0.0022, demonstrating minimal change compared to the base model, indicating that the choice set alteration has only a marginal impact on the results.

Table 9
Marginal Utilities with a 5th Alternative Added to Each Choice Set

	Coefficients	Std. Err.
Horseback Trails	0.0815	0.00723
RV Parking	0.6050	0.00647
Lake Access	0.6018	0.00739
Landmark Near	0.1073	0.00678
Grand Canyon Site	2.1740	0.01215
Trip Cost	-0.0029	0.00004
Avg. Temperature	0.00978	0.00029
Year	0.0040	0.00101
Month	-0.0154	0.00155
Stay Length	0.05748	0.00143
Site Elevation	-0.0002	0.000003

This table presents the marginal utility coefficients for campground amenities following the addition of a fifth alternative to each choice set. Compared to an initial trip cost coefficient of -0.0025, the inclusion of a fifth alternative does not significantly change the trip cost coefficient.

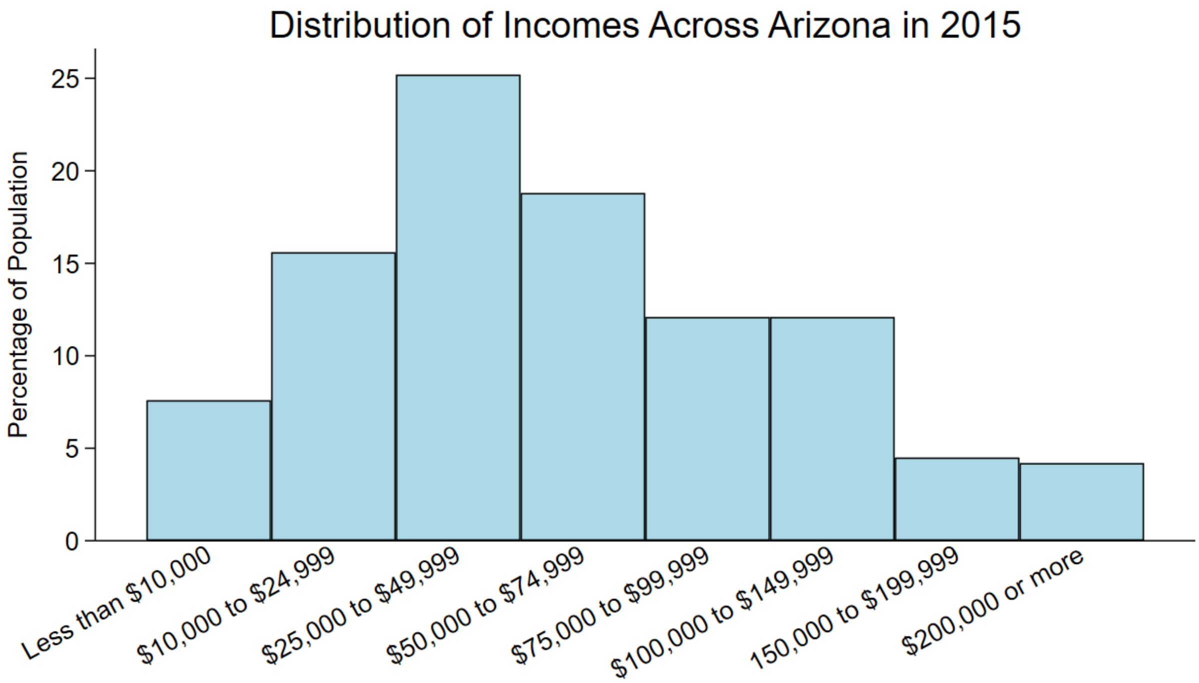
In my base model, the trip cost coefficient is -0.0025. Reducing the number of alternatives in each choice set results in a marginal decrease in the coefficient's magnitude to -0.0022. Conversely, adding an additional site leads to a slight increase in magnitude to -0.0029. These minor fluctuations are expected, as the alternatives added or removed are determined based on their distance to the chosen site. The per-person per-day willingness to pay for a campsite initially averages approximately $26; removing a site from each choice set raises this value to about $29, while adding a site decreases it to approximately $22. Therefore, altering the choice set creation strategy has only a marginal impact on the trip cost coefficient and affects the average individual's willingness to pay for a campsite by $3 to $4 per day.

Next, I test the effect of altering the value of time parameter in my model's trip cost. I hypothesize that county- and State-level income data will significantly alter travel cost computation. Thus far, trip costs have been calculated using a value of time equal to the median ZCTA income divided by 2080 hours of work in a year and multiplied by the camper's travel time:

$$VOT = \frac{Income_{ZCTA}}{2080} TravelTime_{Hours} \tag{4.3.1}$$

To assess the sensitivity of the trip cost to the aggregation of income data, I replace ZCTA-level median income with county-level and state-level median income. Figure 3 illustrates the distribution of individual household incomes across Arizona in 2015, while Figure 4 shows the frequency of median ZCTA incomes present in my data. Notably, there is a similarity in distribution between the two figures. In contrast, county-level income data fail to adequately capture individual income variation. Although the highest median county income is just over $68,000, a significant percentage of observations in my sample, along with numerous reported incomes, exceed this amount substantially.

Figure 4
2015 Household Incomes as Reported by the Census

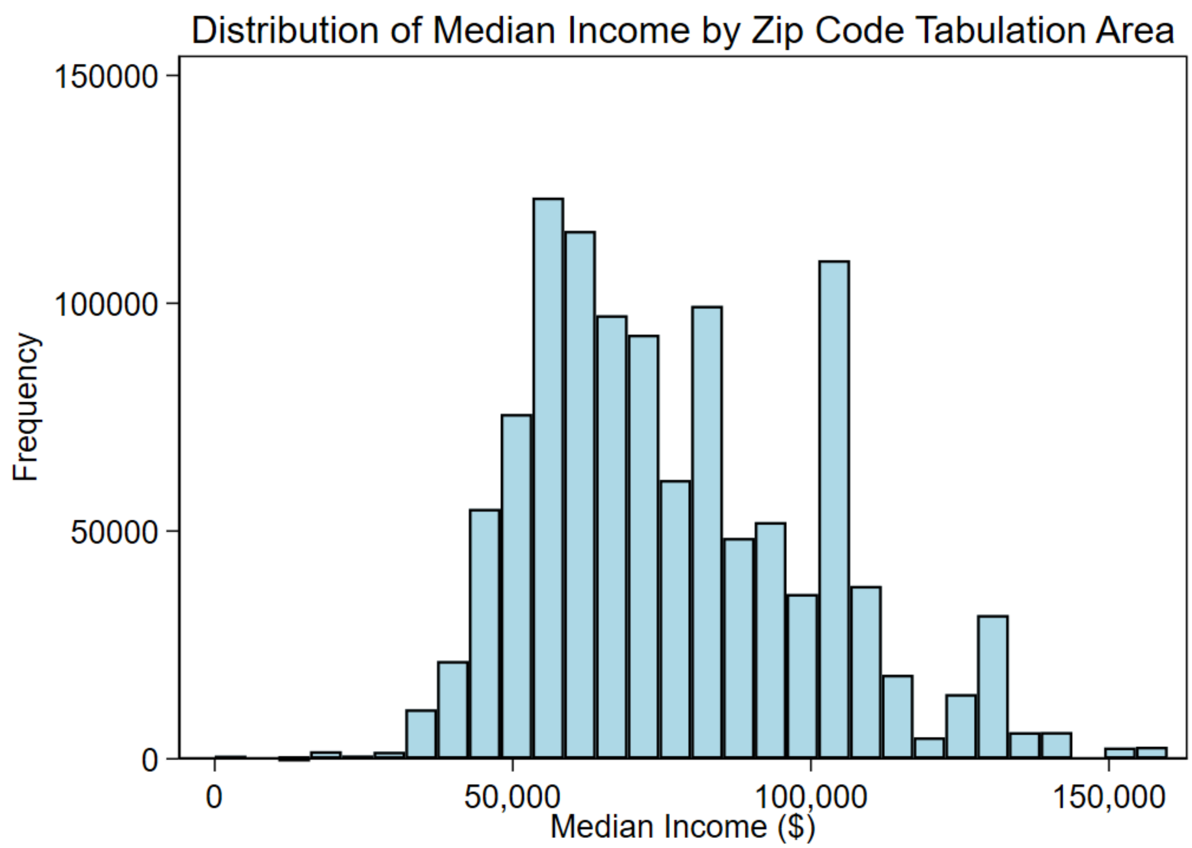


Notes: This figure displays Arizona household incomes as reported by the Census in 2015. A significant minority of households earn more than $75,000 annually. The highest median county income reported in 2015 was $68,000, while the state median income was about $52,248. Thus, income data aggregated at the county or state level cannot accurately approximate individual incomes.

A significant proportion of campers in my dataset originate from high-income zip code tabulation areas (ZCTAs). However, the general distribution of ZCTA-level income data in my sample follows the actual distribution of household incomes across the state. While calculating travel costs based on median ZCTA incomes may introduce some degree of measurement error—compared to using individual-level income data—the similarity in income distribution and the inclusion of high-income ZCTAs mitigates the potential impact of such errors. Aggregating income data beyond the ZCTA level, such as to the county level, would fail to capture a substantial degree of income variation, particularly at higher income levels. As shown in Figure 3, nearly one-third of reported incomes exceed $75,000, while the highest county-level median income in 2015 was $68,000. Moreover, my dataset includes a considerable number of reservations from ZCTAs with median incomes above $100,000, further emphasizing the need for granular income data."

Unsurprisingly, the use of county-level and state-level median income data introduces bias into the trip cost parameter. In the model where travel costs are calculated based on ZCTA median incomes, the esti-

Figure 5
The Median ZCTA Income Distribution in this Paper's Data



Notes: This figure displays the median ZCTA income distribution across my sample. Note the similarity in distribution between ZCTA incomes and individual incomes across Arizona displayed in Figure 4. Relatively wealthier campers appear to form a significant subgroup within the sample.

mated coefficient is approximately -0.0025. However, this coefficient nearly doubles in magnitude, reaching around -0.0051, when using county- or state-level median income data. Lower-resolution income data disproportionately undervalue the time cost for higher-income campers, resulting in biased travel cost estimations.

Finally, I examine the role of campsite capacity and availability in determining willingness to pay. If popular camping facilities frequently reach full capacity, demand may exceed supply, potentially leading to an understatement of willingness to pay. To assess whether campsite availability affects my estimates, I exclude facilities that are over 70%, 80%, and 90% reserved on specific days. Table 11 presents the results after excluding campgrounds nearing full capacity. The first-stage estimates remain largely unchanged. Notably, nearly 95% of the observations involve camping facilities at less than 50% capacity, suggesting that campsite availability and capacity do not significantly influence the results.

Table 10
Marginal Utilities with Low Availability Campgrounds Excluded

	>70% Capacity Excluded		**>80% Capacity Excluded**		**>90% Capacity Excluded**	
	Coefficients	**Std. Err.**	**Coefficients**	**Std. Err.**	**Coefficients**	**Std. Err.**
Horseback Trails	0.4013	0.00818	0.4479	0.00804	0.4657	0.00799
RV Parking	0.6050	0.00647	0.6512	0.00651	0.6527	0.00649
Lake Access	0.6581	0.00655	0.8314	0.00741	0.8381	0.00739
Landmark Near	0.3108	0.00732	0.3364	0.00726	0.3464	0.00723
Grand Canyon Site	1.8293	0.01209	1.8088	0.01203	1.8022	0.01201
Trip Cost	-0.0025	0.00004	-0.0025	0.00004	-0.0025	0.00004
Avg. Temperature	0.0068	0.00046	0.0075	0.00046	0.0079	0.00046
Stay Length	0.0457	0.00153	0.0456	0.00152	0.0456	0.00152
Site Elevation	-0.0002	0.000003	-0.0002	0.000003	-0.0002	0.000003

This table reports marginal utilities while accounting for campgrounds that are heavily reserved. The first two columns report results when observations with campgrounds that are above 70% reserved are excluded.

5 Discussion

The purpose of this section is two-fold. In the first subsection, I apply my willingness to pay computations to policy analysis related to park maintenance and upkeep. My results suggest that both the United States Forest Service (USFS) and the National Park Service (NPS) could generate additional revenue by raising campsite fees, which could help alleviate some of the budgetary pressures these agencies face. Given that a growing portion of the USFS operations budget is being consumed by wildfire mitigation efforts, such revenue increases could provide much-needed financial relief for the department and the public goods it manages. The second subsection offers guidelines and considerations regarding the use of administrative data in travel cost modeling. designed for the effective generation and collection of administrative data and travel cost modeling. While travel cost modeling has traditionally focused on recreation, the manual's instructions and applications are designed to support broader environmental economic and economic development efforts and contribute to evidence-based decision-making.

5.1 Analyzing the Forest Service and National Park Sevice Budgets

My sample includes two National Park Service (NPS)-managed campgrounds near the Grand Canyon: Mather Campground and North Rim Campground. While both campgrounds charge a fee of $18 per night, I find that the average per-person, per-day willingness to pay (WTP) for a campsite in this area is approximately $47. This discrepancy generates a consumer surplus of $29 per person per day on average. Based on the 41,427 campers in my sample, with an average party size of 3.28 and an average stay of 2.92 days, the total consumer surplus for Grand Canyon campsites is approximately $11,506,366.

I am not suggesting that the NPS set their canyon campground prices at $47 per night per person; adverse distributional effects and consumer surplus reductions would likely diminish total welfare. The average willingness to pay does not capture what lower-income households may be willing to pay for camping; thus, prohibitively-high pricing is likely to impede visitation from lower-income campers. However, capturing a portion of this surplus could help fund park maintenance and improvements. For example, increasing the fee from $18 to $22 per person per night could generate an additional $1,587,085, assuming no drop in reservation numbers. This added revenue could cover more than half of the estimated costs for park-wide climate adaptation efforts proposed in the NPS Fiscal Year 2017 Budget Justification Report.

Across Arizona, the U.S. Forest Service (USFS) manages six National Forests: Kaibab, Coronado, Coconino, Tonto, Apache-Sitgreaves, and Prescott. My sample includes 187,964 trips to these forests. On average, campers paid $46 per trip in site fees, which equates to $2.99 per person per day. However, my results indicate a willingness to pay of $26 per person per day, generating an average consumer surplus of $23.01 per person per day. This implies a total consumer surplus of approximately $66,642,991 million across National Forest campers in my sample. By increasing fees to $4 per person per day and assuming no change in behavior, the USFS could generate $11.59 million. Even a modest fee increase of $1.01 per person per day would raise over $2.93 million in additional revenue over ten years compared to current revenue estimates.

At the end of my sample period, the 2017 Forest Service discretionary budget was about $4.732 billion across the entire country, with $2.386 billion allocated to nationwide wildfire mitigation. Of the remaining $2.346 billion, $1.504 billion was allotted for the country's National Forests. While it is unclear out how much of this $1.504 billion was utilized toward Arizona National Forest land, Arizona contains 11.25 million acres of the 188,336,179 acres of America's National Forest land. If we assume that the budget is distributed relatively evenly across USFS land, this would imply that the National Forest discretionary budget for Arizona was about $89,839,351 in 2017. Given the pressures on USFS funding, even small campground fee increases could significantly subsidize operations and maintenance without the need to rely on private management.

The analysis thus far has rested on the assumption that small fee increases will not reduce campsite demand. As such, the estimated revenue increases should be interpreted as slightly liberal estimates. While a $1.04 per person per day fee increase is unlikely to significantly reduce camping reservations, let's suppose that 10% of campers in my current sample would choose to stay home when faced with the fee increase, resulting in approximately 169,168 reservations rather than 187,964. Assuming the average number of people per trip were still 4.21 for USFS campground trips and the average stay length were still 3.66 days, the USFS would still generate $10.43 million over the sample period — an increase of $1.86 million compared to current revenues. My WTP estimates align with findings in the recreation demand literature and suggest that current campsite fees are below market value.

Both the USFS and NPS could increase revenues while still pricing campsites below the average willingness to pay found in this study. While pricing campsites at full willingness to pay would reduce access for lower-income campers, modest increases could generate substantial benefits for park and forest maintenance. Though these findings are based on data from Arizona, similar strategies could apply to other areas in the American Southwest, such as Southern Utah, Nevada, and inland California. Because recreation.gov database includes data that span the entire country, this paper's approach could be replicated across any US State.

5.2 Travel Costs with Administrative Data - Guidelines and Considerations

Travel cost modeling and willingness to pay analysis extend beyond recreational settings like campgrounds. Below, I outline several potential applications of travel cost modeling:

1. State or local governments can use travel cost models to estimate how residents value public parks and facilities, even when access is free.

2. Economic development planners may apply these models to identify which businesses drive consumer activity in specific areas.

3. Transportation specialists can leverage travel cost insights to inform decisions on future transit developments and infrastructure improvements.

4. Private businesses, such as amusement parks, can use travel cost analysis to optimize admission pricing and maximize revenue.

To apply my administrative data modeling approach to U.S. Forest Service (USFS) land management and recreational site planning, several key data elements are necessary. First, count data derived from purchased permits, reservations, or other collection methods should include consumer location information, such as visitor addresses or zip codes; while individual-level demographic data can be beneficial, they are not essential for general travel cost modeling. Second, it is crucial to identify characteristics of the evaluated goods or services, such as site amenities, accessibility measures, and visitor capacities, which should be selected based on the researcher's objectives and potential factors influencing visitor travel. Depending on the good or service being evaluated, quality markers – such as environmental quality metrics – could provide useful information in visitation analysis. Third, approximate data on the vehicles used by individuals in the sample is required, focusing on factors that estimate vehicle gas and depreciation costs. Vehicle make and model data are not necessary – travel costs can be approximated from averages across vehicles. The AAA data utilized in my model serve this purpose effectively. Furthermore, visitor income data, or proxies like ZCTA median income data, are necessary to compute travel costs by valuing the time spent traveling in relation to a visitor's income. For outdoor recreation and environmental goods, weather data should be collected due to the potential influence of weather on consumer decision-making. Permit, reservation, or general use prices should be factored into a visitor's travel cost. Lastly, reservation timing or temporal data, along with the costs of related activities, are vital for understanding demand patterns and capturing the overall expenditure and travel behavior of visitors to USFS recreational sites.

In implementing my administrative data modeling approach, several practical considerations and challenges arose that merit discussion. One significant issue was ensuring data accuracy and completeness. While administrative data sources offer a wealth of information, they often require thorough cleaning and verification to eliminate inconsistencies. For instance, careful verification of the campground amenities was required for this dataset, and certain variables like the campground amenities were recorded in different units of measurement. Establishing robust data validation processes is critical to maintaining data integrity. Another challenge involved integrating various data sources. Combining permit data, reservation records, and vehicle usage statistics necessitated careful alignment of datasets to ensure consistency. Different formats and units of measurement posed additional hurdles, particularly when estimating travel costs from vehicle data. Developing a standardized method for data integration was essential for deriving accurate travel cost estimates. Additionally, the travel cost estimation process in this paper required the calculation of over one million routes, which took several days for one state. Larger-scale applications, such as analyzing multiple states or national datasets, would exponentially increase the processing time and may necessitate more robust computing infrastructure. Finally, while ZCTA median income data served as a useful proxy for individual income, reliance on aggregate data can obscure finer demographic nuances. This limitation highlights the importance of considering the context in which the data were collected, as socio-economic factors may vary significantly within a single zip code. Future researchers should be cognizant of these challenges and consider strategies for addressing data limitations, such as incorporating supplementary qualitative research to enrich quantitative findings.

Despite these challenges, administrative data can complement survey-based travel cost modeling by providing a more comprehensive picture of visitor behavior and preferences. Administrative data enable researchers to analyze patterns over longer time periods and across broader geographical areas at a substantially lower cost than traditional surveys often allow. This broader perspective not only enhances willingness to pay analysis but also supports more informed decision-making for land management and recreational planning. As the U.S. Forest Service and other agencies face increasing budgetary constraints, leveraging administrative data offers a promising avenue for optimizing resource allocation and the management of environmental goods.

6 Conclusion

In summary, I estimate the average willingness to pay for campsites in Arizona to be about $26 per person per day from 2007 to 2016 using administrative data in a travel cost model. The survey travel cost literature estimates, which typically range from $22-$29, align with my estimates derived from administrative data. While administrative data may lack some of the granularity of survey data, my findings suggest that the two can complement each other—survey data offering detailed insights into specific sites and consumer demographics, while administrative data enables broader policy analysis and generalization. My analysis also demonstrates the potential of using zip code data for travel cost modeling. While individual-level data would improve precision, my findings suggest that zip code tabulation area (ZCTA) income data is a valid proxy. While state- and national-level income data cannot accurately approximate the distribution of individual incomes, ZCTA-level income data offers a reasonable second-best alternative for travel cost computation when individual-level data are not available.

Finally, the scope and temporal range of my dataset offer significant advantages over traditional survey data. A typical survey might capture willingness to pay for a specific site over a limited timeframe, but my administrative data allows for state-wide analysis over a decade. This breadth enables a more comprehensive understanding of travel behavior and demand, making administrative data a valuable tool for policy analysis. Future research could leverage these data for temporal analyses or other policy applications.

References

Anderson Jr, R. J., & Crocker, T. D. (1971). Air pollution and residential property values. *Urban Studies*, *8*(3), 171-180.

Benson, C., Watson, P., Taylor, G., Cook, P., & Hollenhorst, S. (2013). Who visits a national park and what do they get out of it?: A joint visitor cluster analysis and travel cost model for Yellowstone National Park. *Environmental management*, *52*, 917-928.

English, E., von Haefen, R. H., Herriges, J., Leggett, C., Lupi, F., McConnell, K., Welsh, M., Domanski, A., & Meade, N. (2018). Estimating the value of lost recreation days from the Deepwater Horizon oil spill. *Journal of Environmental Economics and Management*, *91*, 26-45.

Feather, P., Hellerstein, D., & Tomasi, T. (1995). A discrete-count model of recreational demand. *Journal of Environmental Economics and Management*, *29*(2), 214-227.

Goodman, A. C. (1998). Andrew Court and the invention of hedonic price analysis. *Journal of Urban Economics*, *44*(2), 291-298.

Griliches, Z. (1961). Hedonic prices for automobiles: An econometric analysis of quality change."The Price Statistics of the Federal Government, General Series No. 73", Columbia Univ. Press for the National Bureau of Economic Research, New York, 137-196.

Hagerty, D., & Moeltner, K. (2005). Specification of driving costs in models of recreation demand. *Land Economics*, *81*(1), 127-143.

Harrison Jr, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, *5*(1), 81-102.

Kain, J. F., & Quigley, J. M. (1970). Measuring the value of housing quality. *Journal of the American Statistical Association*, *65*(330), 532-548.

Lloyd-Smith, P., & Becker, M. (2020). The Economic Value of Camping Using Administrative Data. *Journal of Agricultural and Resource Economics*, *45*(3), 445-461.

Lupi, F., Phaneuf, D. J., & von Haefen, R. H. (2020). Best practices for implementing recreation demand models. *Review of Environmental Economics and Policy.*

Menz, F. C., & Mullen, J. K. (1981). Expected encounters and willingness to pay for outdoor recreation. *Land Economics*, *57*(1), 33-40.

Morey, E. R. (1981). The demand for site-specific recreational activities: a characteristics approach. *Journal of Environmental Economics and Management*, *8*(4), 345-371.

Murdock, J. (2006). Handling unobserved site characteristics in random utility models of recreation demand. *Journal of Environmental Economics and Management*, *51*(1), 1-25.

Palmquist, R. B. (1984). Estimating the Demand for the Characteristics of Housing. *The Review of Economics and Statistics*, 394-404.

Ridker, R. G., & Henning, J. A. (1967). The determinants of residential property values with special reference to air pollution. *The Review of Economics and Statistics*, 246-257.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34-55.

Smith, V. K., Desvousges, W. H., & McGivney, M. P. (1983). The opportunity cost of travel time in recreation demand models. *Land economics*, *59*(3), 259-278.

Train, K. E. (1998). Recreation demand models with taste differences over people. *Land Economics*, 230-239.

Wei, C., Fu, M., Wang, L., Yang, H., Tang, F., & Xiong, Y. (2022). The research development of hedonic price model-based real estate appraisal in the era of big data. *Land*, *11*(3), 334.