Estimating Park Choice and Access in Alameda County, California, using Passive Origin-Destination Data.

Gregory Macfarlane[a,*], Teresa Tapia, Carole Turley-Voulgaris[b]

[a] *Civil and Environmental Engineering Department, 430 Engineering Building, Provo, Utah 84602*
[b] *Some Other Place*

Abstract

Parks provide important benefits to those who live near them, in the form of improved property values, health outcomes, etc.; nevertheless, measuring and understanding who lives near a park is an open research question. In particular, it is not well understood which park individuals will choose to use when given a choice among a set of nearby parks of varying sizes and at varying distances from their home. In this paper we present a park activity location choice model estimated from a passive origin-destination dataset — supplied by StreetLight Data, Inc. — representing trips to parks and green spaces in Alameda County, California. The estimated model parameters reveal heterogeneous preferences for park size and willingness-to-travel across block-group level socioeconomic segmentation: Specifically, high-income block groups appear more positively attracted to larger parks, and block groups with a high proportion of ethnic minority individuals are more likely to select nearby parks. The findings have importance for understanding recreational access among different populations, and the methodology more generally supplies a potential template for using passive data products within travel modeling.

Keywords: Accessibility Passive Data Location Choice

---

1. Introduction

Parks and other green spaces generate immense value for the public who are able to access them. The City Parks Alliance [6] categorizes the observed benefits of urban parks as encouraging active lifestyles [1], contributing to local economies, aiding in stormwater management and flood mitigation, improving local air quality, increasing community engagement [21], and enhancing public equity.

*Corresponding Author

Email addresses: gregmacfarlane@byu.edu (Gregory Macfarlane), teresa.tapia@streetlightdata.com (Teresa Tapia), cat@example.com (Carole Turley-Voulgaris)

Preprint submitted to Submitted to Transportation Research Part C: Emerging TechnologiesNovember 25, 2020

Nevertheless, understanding and quantifying these benefits depends in many cases on identifying who lives near the parks and is therefore able to access them. Many previous studies [e.g., 34] rely on comparison of total greenspace across metropolitan areas; this methodology may not adequately control for city-level fixed effects and it may ignore the potentially inequitable distribution of park space within a region. Studies focusing on access within metropolitan areas typically assume that people living within a certain distance or travel time threshold have access to a park, or examine the quantity of park space within one's own arbitrarily defined "neighborhood" [26, Stark2014]. But these methods do not account for the fact that some people will travel to other parks to perform recreational activities. A more holistic measure that continuously measures access across multiple preference dimensions is desirable.

An appealing solution would be to examine and model the activity location choices of park users. Such a model would help researchers understand how individuals of different backgrounds and preferences value different park amenities. Further, the logsums of a location choice model provide a continuous measure of accessibility that explicitly accounts for such variation [8]. Unfortunately, park choice models of this form are rare in the literature. Travel demand models built for infrastructure forecasting are a common way to generate such accessibility logsums, but these models group many different kinds of social and recreational trips together [29]. Further, the attraction term for such trip purposes is commonly a function of the retail or service employment or the number of households at the destination; a typical park or green space has neither employees nor residents. Finally, many regional household travel surveys are oriented towards an average weekday travel pattern, and many park trips occur irregularly or on weekends.

In this paper we present a park destination choice model where individuals living in Alameda County, California choose among parks in the same county. The individuals are constructed from passive data that was derived from mobile devices and processed using algorithms developed by StreetLight Data, Inc. The origin location points are inferred residence block groups for unique devices and the destination points are geofenced polygons representing green and open spaces. The individuals' choice of park location is conditioned on the distance from the block group to the parks in the choice set as well as the size of each park; market segmentation allows for heterogeneous responses between ethnic groups and income strata.

The paper proceeds in the following manner: A discussion of prior attempts to study park choice and employ passive origin-destination data in the literature is given directly. The Methodology section presents the data gathering and cleaning efforts as well as the econometric location choice model. The Results section presents the estimated model coefficients and a discussion of the findings, as well as a model validation exercise. After presenting limitations and associated avenues for future research, a final Conclusions section outlines the contributions of this study for recreational trip modeling and location choice modeling more generally.

2. Literature

Understanding who has access to parks is a long-standing question across multiple scientific disciplines. Researchers specializing in recreation management, public health, urban planning, ecology, and civil engineering have all played a role in shaping our collective understanding of park design, access, and use. A complete review of all of these fields is not warranted for the scope of this paper, but some recent findings are worth discussion.

A popular measure of park accessibility is the Trust for Public Land's "ParkScore" statistic [38]. ParkScore considers the share of the population that resides within a 10-minute walk of a green space using a sophisticated network routing algorithm, in combination with the total city green space, investment, and amenities weighted against the socioeconomic characteristics of the population outside of the 10-minute walk threshold. The resulting score is a convenient quantitative tool in estimating the relative quality of green space access across cities [35]. It may be less useful at identifying the comparative quality of access within a city, particularly as more than 95% of residents in many large metropolitan areas like San Francisco and New York live within the binary 10-minute walk threshold. The Centers for Disease Control and Prevention (CDC) has developed an "Accessibility to Parks Indicator" along similar lines [40], calculating the share of the population living within a half-mile of a park for each county in the U.S.

There is recognition that park access should in some way be linked with park use. After all, a park that has many visitors must by definition be accessible to those visitors. McCormack et al. [22] provide a comprehensive review of this literature; it is sufficient here to note that most studies find a complicated interplay between park size, maintenance, facilities, and travel distance. Many of these attributes are incorporated into ParkIndex [17], which estimates the resident park use potential within $100m^2$ grid cells, based on a household park use survey in Kansas City.

From a transportation engineering perspective, the park use potential measured by ParkIndex is not dissimilar from a park trip production potential. Along these lines, the question of park use is a destination choice problem, where trip makers consider which park is most attractive to accomplish their recreation activity. The Institute of Transportation Engineers (ITE) Trip Generation Manual [2017] contains trip attraction rates for public parks that use as attraction terms the park acreage, number of picnic tables, employees, and other variables. As with many land uses in Trip Generation, the provided trip generation rates are based on a limited number of observational samples and may not represent large-sample behavior [25]. Moreover, regression-based attraction rates isolated from trip production and travel behavior ignore the geographical and behavioral contexts in which people actually make trips to parks [2]: Though more people may come to larger parks, a park cannot attract more people simply by becoming bigger.

There are limited examples of researchers using a destination choice model to predict recreation attractions. Kinnell et al. [18] apply a choice model to a survey of park visitors in New Jersey, and estimate the relative attractiveness of park attributes including playgrounds, picnic areas, and park acreage weighed against the travel disutility and the relative crime rate at the destination. In a similar study, Meyerhoff et al. [24] model the urban swimming location choice for a surveyed sample. In both studies, the researchers were attempting to ascertain which attributes of a recreation generated the most positive utility, and therefore which attributes should be prioritized for improvement. These studies have not to our knowledge been previously referenced in discussions of park accessibility.

2.1. Passive Location Data

The advent of large-scale mobile networks and the seemingly perpetual association of unique devices with unique users has given researchers a new opportunity to observe the movements and activity location patterns for large subsets of the population [28]. Such passively collected movement data — sometimes referred to as “Big Data” — is passively collected as a by-product of other systems including cellular call data records [e.g., 4, 5], probe GPS data [15], and more recently Location Based Services (LBS) [36, 19]. LBS use a network of mobile applications that obtain the users’ physical location. A variety of commercial vendors repackage, clean, and scale these data to population or traffic targets and sell origin-destination matrices to researchers and practitioners at relatively low prices. Monz et al. [27], for example, demonstrate how passive device data can be used to accurately estimate trips to natural recreation areas.

Passive origin-destination matrices are beginning to inform trip distribution model development more directly as well. Kressner [20] proposes one methodology, where passive origin-destination matrices serve as a probabilistic sampling frame for a simulated trip destination choice. Bernardin et al. [3] employ a passive origin-destination matrix as a shadow price reference in an activity-based location choice model, iteratively adjusting the parameters of the choice utilities to minimize the observed error between the matrix and the modeled predictions. A similar method developed by Zhu and Ye [44] uses the passive dataset directly, sampling 10,000 random trips from GPS traces of taxi trips in Shanghai and estimating a destination choice model. Employing the passive data set in this way provides the authors an opportunity to examine the choices of a large sample of a small population (taxi passengers) as well as sufficient data to estimate a “constants-rich” destination choice model with uniquely estimated coefficients for each origin-destination pair. The Zhu and Ye methodology suggests that a similar approach should apply in other contexts, including park choice.

3. Methodology

We constructed a dataset on which to estimate park activity location choices for a sample of observed trips in Alameda County, California. Alameda County is one of the seven counties that constitutes the San Francisco Bay Area metropolitan region in California. Alameda is the seventh most populous county in California with a population of 1.5 million [39], and has 14 incorporated cities and several unincorporated communities. It is an economically and ethnically diverse county and hence it was an attractive area to use for this study. The racial makeup of Alameda County was (49.7%) White, (11.2%) African American, (1.0%) Native American, (38.7%) Asian, (1.0%) Pacific Islander, and (22.4% Hispanic or Latino (of any race). Alameda County has a diverse set of parks, ranging from local small community parks, urban and transit-accessible parks like the Lake Merritt Recreational area, accessible coastal access, and suburban recreational areas like Lake Chabot.

3.1. Data

We constructed an analysis dataset from a publicly-available parks polygons layer, a commercially acquired passive device origin-destination table representing trips between the parks and home block groups, and American Community Survey data for the home block groups.

We obtained a polygons shapefile layer representing open spaces in Alameda County, California from the California Protected Areas Database [12]. This dataset was selected because it included multiple different types of open space including local and state parks, traditional green spaces as well as wildlife refuges and other facilities that can be used for recreation. We removed facilities that did not allow open access to the public (such as the Oakland Zoo) and facilities whose boundaries conflated with freeway right-of-way – this prevents trips through the park from being conflated with park trips in the passive origin-destination data.

From this initial parks polygons dataset, we obtained park attribute information through OpenStreetMap (OSM) using the `osmdata` package for R [30]. Three attribute elements are considered in this analysis. First, we identify playgrounds using OSM facilities given a `leisure = playground` tag. The tagged facilities can be either polygon or point objects; in the former case we use the polygon centroid to determine the location of the playground.

Second, we consider sport fields of various kinds identified with the OSM `leisure = pitch` tag. This tag has an additional attribute describing the sport the field is designed for, which we retain in a consolidated manner. Soccer and American football fields are considered in a single category, and baseball and softball fields are combined as well. Basketball, tennis, and volleyball courts are kept as distinct categories, with all other sport-specific fields combined into a single "other." Golf courses are discarded. As with playgrounds, polygon field and court objects are converted into points at the polygon centroid.

Finally, we identified trails and footpaths using the `path`, `cycleway`, and `footway` values of the `highway` tag. A visual inspection of the resulting data revealed that the large preponderance of sidewalks and cycling trails within parks in Alameda County are identified properly with these variables. Trails are represented in OSM as polylines, or as polygons if they form a complete circle. In the latter case, we converted the polygon boundary into an explicit polyline object.

It is possible for each of these facilities to exist outside the context of a public park. For example, many private apartment complexes have playgrounds and

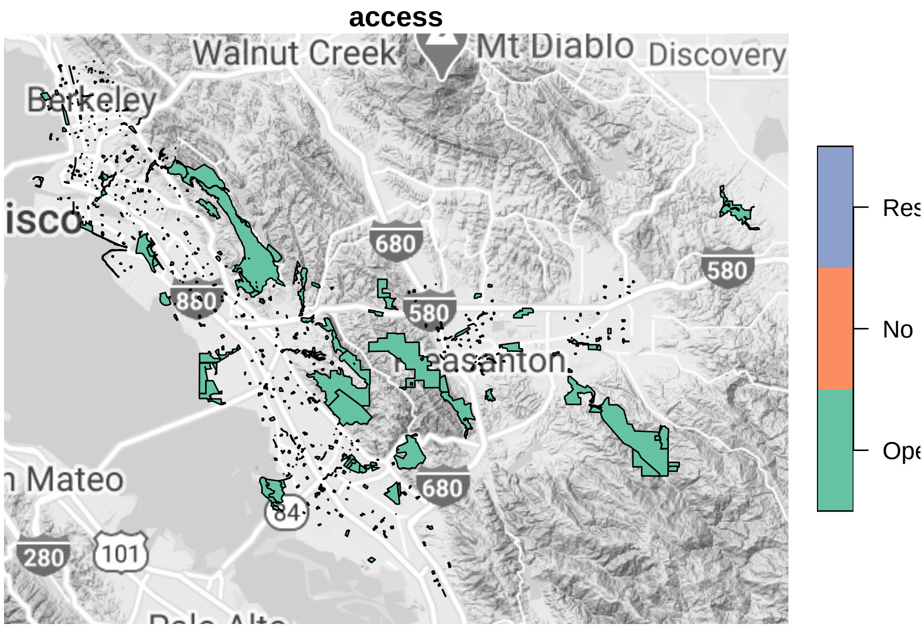

Figure 1: Location of parks in Alameda County.

high schools will have sports facilities that are not necessarily open to the general public. We spatially matched the OSM amenities data and retained only those facilities that intersected with the CPAD open spaces database identified earlier.

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=37.69427,-122.130033&zoom=
```

We provided the park boundaries layer to a commercial firm, StreetLight Data Inc., which develops and resells origin-destination matrices derived from passive device location data. The provider employs a proprietary data processing engine (called Route Science) to algorithmically transform observed device location data points (the provider uses in-vehicle GPS units and mobile device LBS) over time into contextualized, normalized, and aggregated travel patterns. From these travel patterns, the Route Science processing algorithms infer likely home Census block group locations for composite groups of people and converts raw location data points into trip origin and destination points [31, 11].

For each park polygon, the firm returned a population-weighted estimate of how many devices were observed by home location block group over several months in the period between May 2018 and October 2018. We transformed this table such that it represented the weighted unique devices traveling between block groups and parks. We discarded home location block groups outside of Alameda County; the imputed home locations can be far away from the study area for a small amount of trips and are unlikely to represent common or repeated park activities.

Table 1: Park Summary Statistics

		Local Park (N=441)		Local Recreation Area (N=57)		State Recre	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std
Acres		59.8	370.5	115.2	509.6	421.0	271
Mobile Devices		1450.0	6685.4	2749.5	6250.4	80.5	113
		N	%	N	%	N	%
Access	Open Access	441	88.2	57	11.4	2	0.4
	No Public Access	0	0.0	0	0.0	0	0.0
	Restricted Access	0	0.0	0	0.0	0	0.0
Playground	FALSE	225	45.0	42	8.4	2	0.4
	TRUE	216	43.2	15	3.0	0	0.0
Any Sport Field	FALSE	277	55.4	37	7.4	2	0.4
	TRUE	164	32.8	20	4.0	0	0.0
Football / Soccer	FALSE	415	83.0	49	9.8	2	0.4
	TRUE	26	5.2	8	1.6	0	0.0
Baseball	FALSE	364	72.8	43	8.6	2	0.4
	TRUE	77	15.4	14	2.8	0	0.0
Basketball	FALSE	342	68.4	50	10.0	2	0.4
	TRUE	99	19.8	7	1.4	0	0.0
Tennis	FALSE	387	77.4	51	10.2	2	0.4
	TRUE	54	10.8	6	1.2	0	0.0
Volleyball	FALSE	434	86.8	55	11.0	2	0.4
	TRUE	7	1.4	2	0.4	0	0.0
Trail	FALSE	156	31.2	22	4.4	0	0.0
	TRUE	285	57.0	35	7.0	2	0.4

Table **??** presents descriptive statistics on the 500 parks assembled for this study, grouped according to the park type as defined on CPAD. A little more than half of the parks have identified paths, while around 40% of the identified parks have playgrounds and sport fields.

In order to understand the demographic makeup of the home block groups and potentially the characteristics of the people who make each trip, we obtained 2013-2017 five-year data aggregations from the American Community Survey using the `tidycensus` [41] interface to the Census API for several key demographic and built environment variables: the share of individuals by ethnic group, the share of households by income level, household median income, and the housing unit density. An important attribute of the choice model is the distance from the home block group to the park boundary. Because we have no information on where in the block group a home is actually located, we use the population-weighted block group centroid published by the Census Bureau as the location for all homes in the block group. We then measured the network-based distance between the park and the home block group centroid using a walk network derived from OpenStreetMap using the `networkx` package for Python [13],

	Unique (#)	Missing (%)	Mean
Density: Households per square kilometer	1040	0	1714.5
Income: Median tract income	971	3	93246.3
Low Income: Share of households making less than $35k	979	0	18.4
High Income: Share of households making more than $125k	1004	0	32.5
Black: Share of population who is Black	927	0	12.5
Asian: Share of population who is Asian	1010	0	27.8
Other: Share of population who belong to other minority groups	612	0	1.5

3.2. Model

In random utility choice theory, if an individual living in block group n wishes to make a park trip, the probability that the individual will choose park i from the set of all parks J can be described as a ratio of the park's measurable utility V_{ni} to the sum of the utilities for all parks in the set. In the common destination choice framework we apply a multinomial logit model [23, Recker and Kostyniuk [33]],

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_{j \in J} \exp(V_{nj})} \tag{1}$$

where the measurable utility V_{ni} is a linear-in-parameters function of the destination attributes.

$$V_{ni} = X\beta \tag{2}$$

where β is a vector of estimable coefficients giving the relative utility (or disutility) of that attribute to the choice maker, all else equal. It is possible to add amenities of the park or the journey to the utility equation. However, as the number of alternatives is large, it is impractical to consider alternative-specific constants or coefficients and therefore not possible to include attributes of the home block group or traveler n directly. We can, however, segment the data and estimate different distance and size parameters for different segments to observe heterogeneity in the utility parameters between different socioeconomic groups.

The logarithm of the sum in the denominator of Equation 1 (called the logsum) provides a measure of the consumer surplus of the choice set [42],

$$CS_n = \ln \sum_{j \in J} \exp(V_{nj}) + C \tag{3}$$

where C is a constant indicating an unknown absolute value. But comparing the relative logsum values across choice makers, $CS_n - CS_{n-1}$ gives an indication of which choice maker has a more valuable choice set. Or, in this case of a park destination choice model, which choice maker has better access to parks. Such a "utility-based" accessibility term is thus continuously defined, dervied directly from choice theory, and can contain multiple dimensions of the attributes of the choice [14, 9].

In the most typical cases, researchers estimate the utility coefficients for destination choice models from household travel surveys. As we have no knowledge of an appropriate survey on park access, we need to synthesize a suitable estimation data set. We do this by sampling 2×10^4 random discrete device origin-destination pairs from the commercial passive data matrix, weighted by the volume of the flows. This corresponds to a 4.3% sample of all the observed device origin-destination pairs.

The sampled origin-destination pair gives the home location as well as the "chosen" alternative for a synthetic person. In principle the individual's choice set contains all the parks in our dataset; in practice it can be difficult to estimate choice models with so many alternatives ($|J| = 500$). For this reason we randomly sample 10 additional parks to serve as the non-chosen alternatives for our synthetic choice maker. Such random sampling of alternatives reduces the efficiency of the estimated coefficients but the coefficients remain unbiased [37]. As the model has no alternative-specific constants, the standard likelihood comparison statistic against the market shares model ρ^2 is not computable. We instead use the likelihood comparison against the equal shares model ρ_0^2.

The resulting analysis dataset therefore contains 2×10^4 choice makers that select between 11 parks including the park they were observed to choose; the measured distance between the choice maker's block group and all parks in the choice set; and the acreage of each park in the choice set. We hold out a random sample of approximately 20% of choice makers for validation purposes. We use the `mlogit` package for R [7, 32] to estimate the multinomial logit models.

4. Results

We estimated multinomial logit park activity location choice models including coefficients for the distance between the park and the home block group and the acreage of the park. We applied a Yeo-Johnson transformation [43] to both distance and acreage; the Yeo-Johnson transformation replicates the constant marginal elasticity of a logarithmic transformation while avoiding undefined values ($YJ(0) = 0$). For simplicity, we call this transformation `log()` in the model results tables. Using a constant marginal elasticity is better reflective of how people perceive distances and sizes; a one-mile increase to a trip distance is more impactful to a one-mile trip than a ten-mile trip.

Table 2 presents the model estimation results for a series of models with different utility function definitions, each estimated on the complete set of synthetic choice makers. The first model — named "Network Distance" — only considers the distance to the park and the size of the park in the utility equation. The estimated coefficients are significant and of the expected sign: That is, individuals will travel further distances to reach larger parks. The ratio of the estimated coefficients implies that on average, people will travel 3.3089752 times further to reach a park twice as large.

The second and third models in Table 2 include a vector of park attributes. In the model labeled “Park Attributes,” the presence of any sport field is considered with a single dummy value, and in the “Sport Detail” model this variable is disaggregated into facilities for different sports. The value of the size and distance coefficients change modestly from the “Network Distance” model, with the implied size to distance trade-off rising to 3.3089752. Examining the two amenities models — independently and in comparison with each other — reveals a few surprising findings. First, it appears that playgrounds and sport fields in general contribute *negatively* to the choice utility equation. This is both unintuitive and contradictory to previous findings in this space [e.g., 18]. Considering different sports separately, there is a wide variety of observed response with tennis and volleyball facilities attracting more trips, and football and basketball facilities attracting fewer, all else equal. Trails and walking paths give substantive positive utility in both models. The difference in likelihood statistics between the three models is significant (likelihood ratio test p-value 1.8190643×10^{-4}), and so in spite of the curious aggregate findings, we move forward with this utility specification.

It is worth investigating the heterogeneity in preferences that exist among different demographic groups. Though the income and ethnicity of the synthetic park visitors is not known, we can segment the estimation dataset based on the socioeconomic makeup of the visitors’ residence block group. The models presented in Table 3 were estimated on segments developed in this manner. Models under the “Minority” heading include a race-based segmentation: simulated individuals living in block groups with more than thirty percent Black persons are included in the “>30% Black” model, an analogous segmentation for block groups with high Asian populations are in the “>30% Asian” model, and the “Other” model contains all other block groups. Another set of model segmentation relies on the share of the population in each block group with household incomes above or below certain thresholds. Again, we use the threshold definitions based on in **??**.

The model estimates in Table 3 reveal that there is noticeable heterogeneity in the response among different socioeconomic groups. Park visitors living in block groups with a high proportion of Black and low-income residents show less affinity for trails and other walkways, but appear are also considerably to be more sensitive to the distance of a park. Visitors living in high-income neighborhoods are more attracted to the amenities of a park, or rather these visitors do not show significant negative coefficients for amenities, e.g. basketball courts and football fields.

Seeing that there is a difference in the response in the model segmentation, it is also worth considering the role of our segmentation thresholds in these findings. Figure 2 shows the estimated coefficients and confidence intervals for these different amenities at different threshold levels of segmentation. The threshold level means that at least that percent of the block group’s population falls in that category. The confidence intervals widen as more observations are excluded from the model. The estimated coefficients for the different segmentations are identical when the share equals zero, and simply represent the “Sport Detail”

Table 2: Estimated Model Coefficients

	Network Distance	Park Attributes	Sport Detail
log(Distance)	-1.354***	-1.394***	-1.391***
	(0.022)	(0.022)	(0.022)
log(Acres)	0.409***	0.353***	0.348***
	(0.011)	(0.012)	(0.012)
Playground		-0.436***	-0.550***
		(0.049)	(0.049)
Trail		0.559***	0.567***
		(0.052)	(0.053)
Sport Field		-0.322***	
		(0.050)	
Basketball			-0.226***
			(0.068)
Baseball			0.065
			(0.067)
Football / Soccer			-0.482***
			(0.095)
Tennis			0.234***
			(0.066)
Volleyball			0.613***
			(0.135)
Other Sport			-0.192**
			(0.090)
Num.Obs.	3971	3971	3971
AIC	11600.2	11307.2	11292.8
Log.Lik.	-5798.103	-5648.579	-5636.379
rho2			
rho20	0.391	0.407	0.408

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Estimated Model Coefficients

	Minority			Income	
	> 30% Asian	> 30% Black	Other	> 30% Low income	> 50% High income
log(Distance)	-1.324***	-1.528***	-1.379***	-1.419***	-1.309***
	(0.038)	(0.063)	(0.031)	(0.051)	(0.051)
log(Acres)	0.372***	0.311***	0.341***	0.334***	0.345***
	(0.020)	(0.032)	(0.017)	(0.027)	(0.028)
Playground	-0.509***	-0.376***	-0.641***	-0.525***	-0.491***
	(0.084)	(0.119)	(0.071)	(0.104)	(0.122)
Trail	0.677***	0.361***	0.636***	0.307***	0.903***
	(0.094)	(0.126)	(0.076)	(0.109)	(0.148)
Basketball	-0.012	-0.374**	-0.393***	-0.170	-0.103
	(0.109)	(0.183)	(0.103)	(0.153)	(0.160)
Baseball	0.116	0.193	-0.027	0.025	-0.107
	(0.111)	(0.170)	(0.097)	(0.146)	(0.164)
Football / Soccer	-0.453***	-1.015***	-0.370***	-0.867***	-0.250
	(0.152)	(0.278)	(0.139)	(0.226)	(0.223)
Tennis	0.370***	-0.529**	0.335***	-0.081	0.658***
	(0.108)	(0.210)	(0.095)	(0.163)	(0.146)
Volleyball	0.728***	-0.086	0.433**	0.435	0.450*
	(0.195)	(0.615)	(0.215)	(0.403)	(0.257)
Other Sport	-0.105	-0.417	-0.232*	-0.493**	-0.070
	(0.144)	(0.270)	(0.132)	(0.234)	(0.188)
Num.Obs.	1379	591	2001	777	758
AIC	3991.8	1767.9	5479.7	2396.9	1908.5
Log.Lik.	-1985.903	-873.953	-2729.872	-1188.475	-944.244
rho2					
rho20	0.399	0.383	0.431	0.362	0.481

* p < 0.1, ** p < 0.05, *** p < 0.01

model from Table 2.

Overall, increasing the segmentation threshold level reveals some additional information about user preferences. First, it should be noted that there is some inconsistency: for instance, block groups with at least 40% of low income households show a lower importance of distance than block groups with either 30% or 50% low income households. The increasing width of the confidence interval, however, means it is difficult to make generalized statements. Residents of block groups with a higher share of Asian or high income households both show relatively more affinity for tennis courts and trails. Block groups with high Black populations show a somewhat greater preference for playgrounds, as well appearing to be the most distance-sensitive group.

4.1. Model Application

In this section, we apply the models to examine a policy intervention to create additional parks in Oakland by converting closed streets into effective pedestrian plazas. What is the logsum benefit of this policy, and how is it distributed?

This analysis will be completed in December 2020.

5. Limitations and Future Directions

The ideal dataset for estimating individual choices would be a high-quality, large-sample household travel survey of real individuals. Such a survey would give details on whether an observed trip to a park was actually a recreation trip or rather a different activity entirely. The individual-level demographic data would also be valuable in understanding more clearly the observed heterogeneity in response among different income or ethnic groups. Additionally, the trends and correlations revealed in the presented models may reflect situational inequalities rather than true preferences. For example, the distinct observed parameters on size and distance for minority block groups may indicate that areas with large minority populations tend to have smaller parks that are more geographically distributed relative to other regions of the region. Transit access may also affect park choice and how far people are willing to travel to access a park. Preliminary analysis of our source data indicates a qualitative correlation between good transit access and diverse park use from both a geographic and demographic perspective.

We limited our analysis to home locations and parks in Alameda County, California. It is possible that some Alameda residents visit parks in neighboring counties, just as it is possible that parks in Alameda County attract trips from outside the county borders. This is most likely for block groups and parks on the north and south borders of the county. The lower measured accessibility in the area around Berkeley in the northern part of the county () is likely affected by the ommission of parks and residents in Contra Costa County.

Using Euclidean distance to represent the distance between the block group centroid and the border of the park leaves something to be desired: Depending

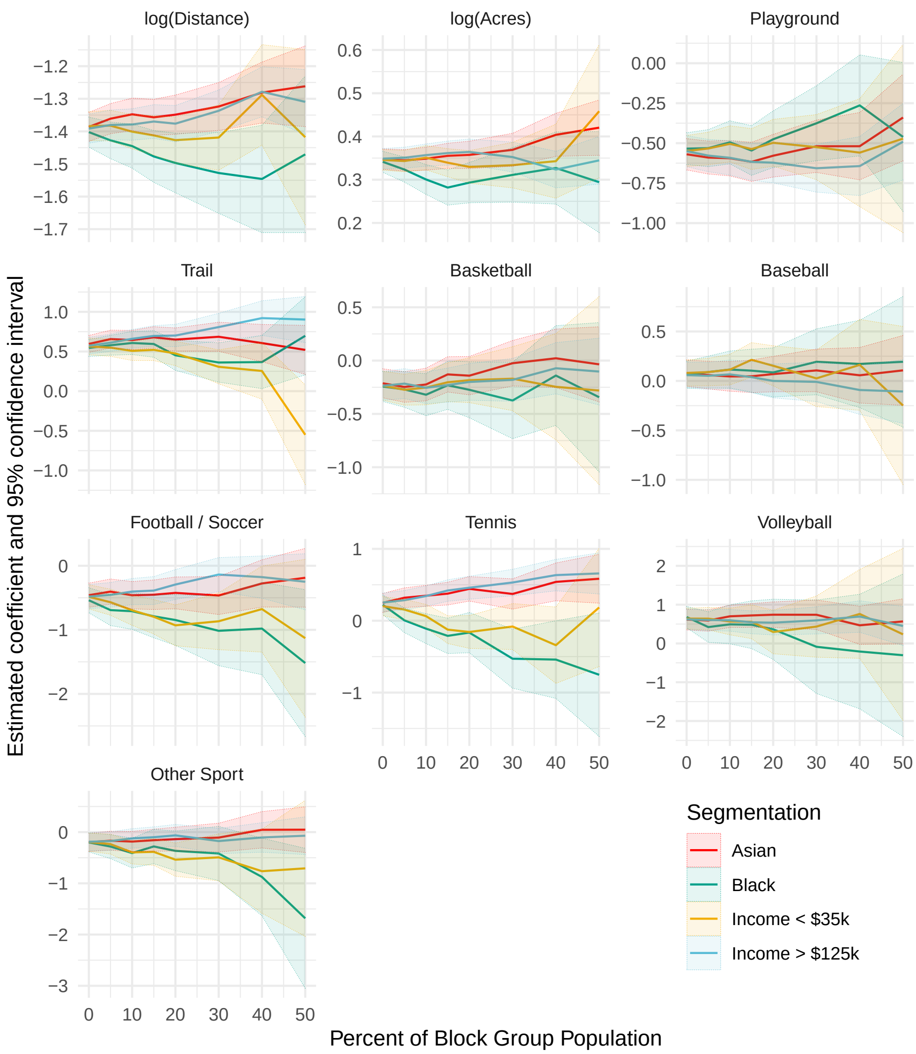

Figure 2: Estimated utility coefficients and 95% confidence intervals for park amenities at different socioeconomic threshold levels.

on network topography and built environment characteristics, there may be a significant variation in perceived travel times between two parks with similar straight-line distances. That said, a more precise network-based measure might not overcome the inaccuracies resulting from our necessarily measuring distances from the block group centroid. As above, an individual-level survey where the home location is explicitly known would be preferable regardless of the distance method employed.

The activity location data used in this specific analysis treats all days of the week and day periods together; it is likely that weekend park choice is substantially different from weekday choice, as the activities performed may be the same. Also recall that the data consider each device-park pair as a unique trip. Repeated trips to the same park may not be properly considered in the data sample. A more precise time-of-day and day-of-week segmentation is warranted.

We applied a naive random sampling of the alternatives in our model estimation and validation; a more considered approach involving hierarchical destination sampling may yield more efficient estimates and therefore a clearer picture of the role of size, distance, and other amenities on the observed choices. The relatively weak predictive power of such a simple model formulation (size and distance only) indicates that there is potential to examine the role that additional park amenities — ball fields, playgrounds, water features, etc. — play in the relative attractiveness of parks for different market segments. The quality of park maintenance is another important feature identified in the recreation literature [10] that is not included here.

6. Conclusions

As transportation professionals seek to improve access to parks and better coordinate transportation and land use efforts — and as researchers more generally try to understand the role parks and open spaces play in public health and society — it is increasingly important to better understand how, when, and why individuals travel to parks. This intersection between recreation and transportation has received relatively little exploration, partially because travel survey data emphasizes weekday travel and because the role of parks in daily activities can be more complicated than with other land uses. This study contributes to the understanding of recreation access by presenting a method to develop access measures explicitly based on the observed choices of individuals. The resulting access measure is continuously defined and incorporates multiple dimensions of access, including the travel necessary to reach all nearby parks as well as the amenities of each of those parks. Further, the measure we have presented reveals heterogeneous preferences for travel and park size across market segments, a heterogeneity that could perhaps be incorporated into an understanding of accessibility.

With the growing availability of passive transportation data, there is a correspondingly increased opportunity to explore such data to develop a better understanding of travel patterns in more careful detail than is possible with household travel surveys. Capturing a sufficiently large survey to study trip patterns to a single park is an enormous undertaking, and doing such an exercise for an entire park system is prohibitively expensive and time-consuming. Passive data sets therefore enable analyses that would be unlikely or impossible by other means. Challenges to the representativeness and comprehensiveness of passive data products are in many cases fair, but this should not preclude their use in cases where traditional techniques are not practicable.

References

[1] Carolyn Bancroft, Spruha Joshi, Andrew Rundle, Malo Hutson, Catherine Chong, Christopher C. Weiss, Jeanine Genkinger, Kathryn Neckerman, and Gina Lovasi. Association of proximity and density of parks and objectively measured physical activity in the United States: A systematic review. *Social Science & Medicine*, 138:22–30, aug 2015. ISSN 02779536. doi: 10.1016/j.socscimed.2015.05.034. URL https://linkinghub.elsevier.com/retrieve/pii/S0277953615003160.

[2] P. O. Barnard and R. E. Brindle. A review and critique of current methods used to predict traffic generation with some accompanying suggestions on alternative approaches. *Transportation Planning and Technology*, 11(4): 273–288, jun 1987. ISSN 0308-1060. doi: 10.1080/03081068708717349. URL http://www.tandfonline.com/doi/abs/10.1080/03081068708717349.

[3] Vincent L Bernardin, John L Bowman, Mark Bradley, Jason Chen, Nazneen Ferdous, and Yuen Lee. Incorporating Big Data in an Activity-Based Travel Model: The Chattanooga Case Study. In *TRB Annual Meeting*. Transportation Research Board, 2018. URL https://trid.trb.org/view/1495930.

[4] R. Bolla and F. Davoli. Road traffic estimation from location tracking data in the mobile cellular network. In *2000 IEEE Wireless Communications and Networking Conference. Conference Record (Cat. No.00TH8540)*, pages 1107–1112. IEEE, 2000. ISBN 0-7803-6596-8. doi: 10.1109/WCNC.2000.904783. URL http://ieeexplore.ieee.org/document/904783/.

[5] Francesco Calabrese, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. Estimating Origin-Destination Flows Using Mobile Phone Location Data. *IEEE Pervasive Computing*, 10(4):36–44, 2011. doi: 10.1109/mprv.2011.41. URL http://dx.doi.org/10.1109/mprv.2011.41http://hdl.handle.net/1721.1/101623http://creativecommons.org/licenses/by-nc-sa/4.0/.

[6] City Parks Alliance. Why City Parks Matter, 2019. URL https://cityparksalliance.org/about-us/why-city-parks-matter/.

[7] Yves Croissant. *mlogit: Multinomial Logit Models*, 2019. URL https://cran.r-project.org/package=mlogit.

[8] Gerard de Jong, Andrew Daly, Marits Pieters, and Toon van der Hoorn. The logsum as an evaluation measure: Review of the literature and new results. *Transportation Research Part A: Policy and Practice*, 41(9):874–889, nov 2007. ISSN 0965-8564. doi: 10.1016/J.TRA.2006.10.002. URL https://www.sciencedirect.com/science/article/pii/S0965856407000316.

[9] Xiaojing Dong, Moshe E. Ben-Akiva, John L. Bowman, and Joan L. Walker. Moving from trip-based to activity-based measures of accessibility. *Transportation Research Part A: Policy and Practice*, 40(2):163–180, feb 2006. ISSN 0965-8564. doi: 10.1016/J.TRA.2005.05.002. URL https://www.sciencedirect.com/science/article/pii/S0965856405000820?via{%}3Dihub.

[10] Donna Fletcher and Harold Fletcher. Manageable predictors of park visitor satisfaction: Maintenance and personnel. *Journal of Park and Recreation Administration*, 21:21–37, 2003. URL https://pdfs.semanticscholar.org/f5cb/11ef839c6366e90654d1a0110f11cbf65334.pdf.

[11] Markus Friedrich, Katrin Immisch, Prokop Jehlicka, Thomas Otterstätter, and Johannes Schlaich. Generating Origin-Destination Matrices from Mobile Phone Trajectories. *Transportation Research Record: Journal of the Transportation Research Board*, 2196:93–101, 2010. doi: 10.3141/2196-10. URL https://journals.sagepub.com/doi/pdf/10.3141/2196-10.

[12] GreenInfo Network. California Protected Areas Database, 2019. URL https://www.calands.org/cpad/.

[13] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[14] S L Handy and D A Niemeier. Measuring Accessibility: An Exploration of Issues and Alternatives. *Environment and Planning A*, 29(7):1175–1194, jul 1997. ISSN 0308-518X. doi: 10.1068/a291175. URL http://journals.sagepub.com/doi/10.1068/a291175.

[15] Arthur Huang and David Levinson. Axis of travel: Modeling non-work destination choice with GPS data. *Transportation Research Part C: Emerging Technologies*, 58:208–223, sep 2015. ISSN 0968-090X. doi: 10.1016/J.TRC.2015.03.022. URL https://www.sciencedirect.com/science/article/pii/S0968090X15001072.

[16] Institute of Transportation Engineers. *Trip Generation Manual*. Institute of Transportation Engineers, Washington, D.C., 10th editi edition, 2017. ISBN 9781933452647.

[17] Andrew T. Kaczynski, Jasper Schipperijn, J. Aaron Hipp, Gina M. Besenyi, Sonja A. Wilhelm Stanis, S. Morgan Hughey, and Sara Wilcox.

ParkIndex: Development of a standardized metric of park access for research and planning. *Preventive Medicine*, 87:110–114, jun 2016. ISSN 0091-7435. doi: 10.1016/J.YPMED.2016.02.012. URL https://www.sciencedirect.com/science/article/pii/S0091743516000517.

[18] J. C. Kinnell, M. F. Bingham, A. F. Mohamed, W. H. Desvousges, T. B. Kiler, E. K. Hastings, and K. T. Kuhns. Estimating Site Choice Decisions for Urban Recreators. *Land Economics*, 82(2):257–272, may 2006. ISSN 0023-7639. doi: 10.3368/le.82.2.257. URL http://le.uwpress.org/cgi/doi/10.3368/le.82.2.257.

[19] Anurag Komanduri, Laura Schewel, Dan Beagan, and Dale Wong. Using Big Data to Develop Comparative Commercial Vehicle Metrics for Port Traffic at Major Ports in the U.S. In *TRB Annual Meeting*. Transportation Research Board, 2017. URL https://trid.trb.org/View/1438797.

[20] Josephine D Kressner. Synthetic Household Travel Data Using Consumer and Mobile Phone Data. Technical report, Transportation Research Board, 2017.

[21] Juliana Madzia, Patrick Ryan, Kimberly Yolton, Zana Percy, Nick Newman, Grace LeMasters, and Cole Brokamp. Residential Greenspace Association with Childhood Behavioral Outcomes. *The Journal of Pediatrics*, dec 2018. ISSN 0022-3476. doi: 10.1016/J.JPEDS.2018.10.061. URL https://www.sciencedirect.com/science/article/pii/S0022347618315683.

[22] Gavin R. McCormack, Melanie Rock, Ann M. Toohey, and Danica Hignell. Characteristics of urban parks associated with park use and physical activity: A review of qualitative research. *Health & Place*, 16(4):712–726, jul 2010. ISSN 1353-8292. doi: 10.1016/J.HEALTHPLACE.2010.03.003. URL https://www.sciencedirect.com/science/article/pii/S1353829210000316.

[23] Daniel L. McFadden. Conditional Logit Analysis of Qualitative Choice Behavior. In Paul Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974.

[24] Jürgen Meyerhoff, Alexandra Dehnhardt, and Volkmar Hartje. Take your swimsuit along: the value of improving urban bathing sites in the metropolitan area of Berlin. *Journal of Environmental Planning and Management*, 53 (1):107–124, jan 2010. ISSN 0964-0568. doi: 10.1080/09640560903399863. URL http://www.tandfonline.com/doi/full/10.1080/09640560903399863.

[25] Adam Millard-Ball. Phantom trips: Overestimating the traffic impacts of new development. *Journal of Transport and Land Use*, 8(1):31–49, 2015. doi: 10.5198/jtlu.2015.384. URL https://www.jstor.org/stable/pdf/26202700.pdf?refreqid=excelsior{%}3A6bf53b09919c223966f1dc8468372150.

[26] Richard Mitchell and Frank Popham. Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet*,

372(9650):1655–1660, nov 2008. ISSN 0140-6736. doi: 10.1016/S0140-6736(08)61689-X.

[27] Christopher Monz, Milan Mitrovich, Ashley D'Antonio, and Abigail Sisneros-Kidd. Using Mobile Device Data to Estimate Visitation in Parks and Protected Areas: An Example from the Nature Reserve of Orange County, California. *Journal of Park and Recreation Administration*, 0 (0), sep 2019. ISSN 21606862. doi: 10.18666/JPRA-2019-9899. URL https://js.sagamorepub.com/jpra/article/view/9899.

[28] Diala Naboulsi, Marco Fiore, Stephane Ribot, and Razvan Stanica. Large-Scale Mobile Traffic Analysis: A Survey. *IEEE Communications Surveys & Tutorials*, 18(1):124–161, 2016. ISSN 1553-877X. doi: 10.1109/COMST.2015.2491361. URL http://ieeexplore.ieee.org/document/7299258/.

[29] National Academies of Sciences Engineering and Medicine. *Travel Demand Forecasting: Parameters and Techniques*. NCHRP 716. National Academies Press, Washington, D.C., 2012. ISBN 978-0-309-28255-0. doi: 10.17226/14665. URL http://www.nap.edu/catalog/14665.

[30] Mark Padgham, Bob Rudis, Robin Lovelace, and Maëlle Salmon. osmdata. *The Journal of Open Source Software*, 2(14), jun 2017. doi: 10.21105/joss.00305. URL https://doi.org/10.21105/joss.00305.

[31] Changxuan Pan, Jiangang Lu, Shan Di, and Bin Ran. Cellular-Based Data-Extracting Method for Trip Distribution. *Transportation Research Record: Journal of the Transportation Research Board*, 1945(1):33–39, jan 2006. ISSN 0361-1981. doi: 10.1177/0361198106194500105. URL http://journals.sagepub.com/doi/10.1177/0361198106194500105.

[32] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

[33] Wilfred W. Recker and Lidia P. Kostyniuk. Factors influencing destination choice for the urban grocery shopping trip. *Transportation*, 7(1):19–33, mar 1978. ISSN 0049-4488. doi: 10.1007/BF00148369. URL http://link.springer.com/10.1007/BF00148369.

[34] Elizabeth A Richardson, Richard Mitchell, Terry Hartig, Sjerp de Vries, Thomas Astell-Burt, and Howard Frumkin. Green cities and health: a question of scale? *Journal of Epidemiology and Community Health*, 66(2): 160 LP – 165, feb 2012. doi: 10.1136/jech.2011.137240.

[35] Alessandro Rigolon, Matthew Browning, and Viniece Jennings. Inequities in the quality of urban park systems: An environmental justice investigation of cities in the United States. *Landscape and Urban Planning*, 178:156–169, oct 2018. ISSN 0169-2046. doi: 10.1016/J.LANDURBPLAN.2018.05.026. URL https://www.sciencedirect.com/science/article/pii/S0169204618304316.

[36] Josh Roll. Evaluating Streetlight Estimates of Annual Average Daily Traffic in Oregon. Technical report, Oregon Department of Transportation, 2019. URL https://trid.trb.org/View/1630252.

[37] Kenneth E. Train. *Discrete Choice Methods with Simulation, 2nd Edition*. Cambridge University Press, Cambridge, mar 2009. doi: 10.1016/S0898-1221(04)90100-9. URL http://linkinghub.elsevier.com/retrieve/pii/S0898122104901009.

[38] Trust for Public Land. 2019 ParkScore Index, 2019. URL https://www.tpl.org/parkscore.

[39] U.S. Census Bureau. QuickFacts: Alameda County, California, 2019. URL https://www.census.gov/quickfacts/fact/table/alamedacountycalifornia/PST045218.

[40] Emily Neusel Ussery, Leah Yngve, Dee Merriam, Geoffrey Whitfield, Stephanie Foster, Arthur Wendel, and Tegan Boehmer. The National Environmental Public Health Tracking Network Access to Parks Indicator: A National County-Level Measure of Park Proximity. *Journal of park and recreation administration*, 34(3):52, 2016. doi: 10.18666/JPRA-2016-V34-I3-7119. URL http://www.ncbi.nlm.nih.gov/pubmed/28868528http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5580831.

[41] Kyle Walker. *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*, 2019. URL https://cran.r-project.org/package=tidycensus.

[42] H C W L Williams. On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A: Economy and Space*, 9(3):285–344, mar 1977. ISSN 0308-518X. doi: 10.1068/a090285. URL http://journals.sagepub.com/doi/10.1068/a090285http://ideas.repec.org/a/pio/envira/v9y1977i3p285-344.html.

[43] In-Kwon Yeo and Richard A Johnson. A new family of power transformations to improve normality or symmetry. *Biometrika*, 87:954–959, 2000. URL https://www.jstor.org/stable/pdf/2673623.pdf?refreqid=excelsior%3A7f84252acd9abceb732e404db81fe53f.

[44] Jiayu Zhu and Xin Ye. Development of destination choice model with pairwise district-level constants using taxi GPS data. *Transportation Research Part C: Emerging Technologies*, 93:410–424, aug 2018. ISSN 0968-090X. doi: 10.1016/J.TRC.2018.06.016. URL https://www.sciencedirect.com/science/article/pii/S0968090X18306454?via{%}3Dihub{#}b0085.