Title
A Statewide Optimal Resource Allocation Tool Using Geographic Information Systems, Spatial Analysis, and Regression Methods

Permalink
https://escholarship.org/uc/item/2gt23996

Authors
Goulias, Konstadinos G.
Golob, Thomas F.
Yoon, Seo Youn

Publication Date
2008-11-01

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

A Statewide Optimal Resource Allocation Tool Using Geographic Information Systems, Spatial Analysis, and Regression Methods

Konstadinos G. Goulias, Thomas F. Golob,
Seo Youn Yoon

California PATH Research Report
UCB-ITS-PRR-2008-27

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation, and the United States Department of Transportation, Federal Highway Administration.

Final Report for Task Order 5110 & 6110

November 2008

ISSN 1055-1425

A Statewide Optimal Resource Allocation Tool Using Geographic Information Systems, Spatial Analysis, and Regression Methods

FINAL REPORT

Konstadinos G. Goulias

Department of Geography & GeoTrans Laboratory
University of California Santa Barbara
Santa Barbara CA 93106
805-284-1597
Goulias@geog.ucsb.edu

Thomas F. Golob

Institute of Transportation Studies
University of California Irvine

Seo Youn Yoon

Department of Geography & GeoTrans Laboratory
University of California, Santa Barbara

Project: PATH Task Orders 5110 & 6110
A GIS-based Tool for Forecasting the Travel Demands of Demographic Groups within California – An Optimal Resource Allocation Tool

October 2008

Santa Barbara, CA

Table of Contents

Executive Summary

The overall objective of this project is to develop an optimal resource allocation tool for the entire state of California using Geographic Information Systems and widely available data sources. As this tool evolves it will be used to make investment decisions in transportation infrastructure while accounting for their spatial and social distribution of impacts. Tools of this type do not exist due to lack of suitable planning support tools, lack of efforts in assembling data and information from a variety of sources, and lack of coordination in assembling the data. Suitable planning support tools can be created with analytical experimentation to identify the best methods and the first steps are taken in this project. Assembly of widely available data is also demonstrated in this project. Coordination of fragmented jurisdictions remains an elusive task that is left outside the project. When this project begun we confronted some of these issues and embarked in a path of feasibility demonstration in the form of a pilot project that gave us very encouraging results. In spite of this pilot nature aiming at demonstration of technical feasibility, substantive conclusions and findings are also extracted from each analytical step.

In this project we have two parallel analytical tracks that are a statewide macroanalysis (called the *zonal based approach* herein) and an individual and household based microanalysis (called the *person based approach* herein). In the statewide macroanalysis we study efficiency and equity in resource allocation. Resources are intended as infrastructure availability and access to activity participation offered by the combined effect of transportation infrastructure and land use measured by indicators of accessibility. Stochastic frontiers are used to study efficiency and a particular type of inequality measurement called the Theil fractal inequality index is used to study equity in the macroanalysis. The outcome of this analysis are maps identifying places in California that enjoy higher levels of service when compared to the entire state and places which succeeded in allocating resources in a relatively better way than others. In the individual microanalysis we use the accessibility indicators from the macronalysis and expand them by defining a new set of indicators at a second level of spatial (dis)aggregation. Then we use them as explanatory factors of travel behavior with focus on the use of different travel models (e.g., driving alone, use of public transportation and so forth). As expected infrastructure availability and accessibility to activity opportunities has a significant and substantive effect on the use of different modes. Many resource allocation decisions, then, will impact behavior, which in turn influences the optimality and equity conditions. This implies that decisions about where and

when to allocate resources in public and private transportation needs to account for changes in behavior in a dynamic fashion, using scenarios of accessibility provision and assessing their impact by studying activity and travel behavior changes.

There are four distinct work tasks that we describe in this report. First, we assembled statewide spatial US census data at two levels of nested geographic subdivisions that are the tract level and the block group level and merged them with a highway network of the same vintage (year 2000). Each subdivision is considered as a center and around each center we create travel time and travel distance buffers. Within each buffer we compute the amount of persons working in each industry (retail, education, health, manufacturing, and all other activities) to represent the spatial opportunities to participate in activities available to the residents of each virtual center. We also count the number of facility kilometers to represent the supply of infrastructure.

Second, we use the data from the first task to study the ability of each area in providing services to its residents and then we compare all these areas and rank them based on stochastic frontiers, which is a regression method. We named this method the efficiency measurement because it allows to link infrastructure provision (as the input) to the accessibility offered (as the output). Stochastic frontier analysis captures and depicts the complex set of relationships among highways and accessibility showing that providing more roadways is not always better for access to opportunities. This happens either because of competition for space and/or because the spatial distribution of activity opportunities does not follow these roadways but obeys other spatial distribution rules. The regression results also show that the role of roadways depends on the measurement indicator used but also the presence of other surrounding roadways. Overall, however, the presence of primary roadways has a strong positive impact on access. For core access the secondary roadways seem to have a much higher impact and merit attention for investment. Efficiency in the transformation of roadways to access depends on the residents of each tract and depends on the measurement of access (outer ring vs. middle ring).

Third, we demonstrate a method that identifies specific locations in the entire state where resource allocation has succeeded in maximizing benefits to the public. Using a derived factor of accessibility for the population residing in each block group an index for the entire state was computed that measures the disparities in accessibility featured by the block groups in regard to their population. This same index can thus constitute a first tool for policy makers who consider equality as a criterion of allocation of infrastructure investment. Then we implement a fractal (an index based on the nested spatial structure of counties, tracts, and block groups) inequality index (called the Theil index) that gives us a better understanding of the spatial distribution of inequality throughout different geographical scales. This index gives information about the disparities in accessibility between Counties as well as inside the Counties themselves. The Theil index we implemented here constitutes a tool both easy to understand thanks to its intuitive definition, easy to implement since it relies on data that are largely available, and able to give instructive information about the structure of inequality in providing access to residents. It shows which locations in California fail to be equitable and require their residents to travel excessively to pursue the same amount of activities as residents of other locations where travelling enables better time allocation.

Fourth, the wealth of the spatial indicators developed using information from census tracts, census block groups, and the extensive roadway network in California were used as explanatory variables in regression models of travel behavior. Each set of these accessibility capturing variables affects different travel behaviors in different ways. Household density, retail employee density and road infrastructure provided meaningful explanation of the variety in travel behaviors we observe capturing the impact of different dimensions of accessibility such as characteristics of residential area, availability of activity opportunity, and connectivity through road infrastructure. From the model estimation experiments a variety of findings emerge. From the comparisons between the census tract models and the block group models, we see that the variables describing a behavioral aspect can show different levels and patterns of impact on travel behaviors when they are measured using different areal unit sizes. To be more specific, household density measured in census tracts explained better the behavior analyzed here than household density measured using block groups. From the comparisons, we see that census tracts, covering a larger area around a residence, capture the density impact in more informative ways. However, this cannot be the golden rule for every travel behavior indicator. We need to think about the implications that a specific areal unit has on each travel behavior indicator, test its ability to explain behavior, and decide to use the one that is the most informative. In addition, spatial variables involving shortest paths in computation showed better ability of discerning the impacts of each spatial segment and also clearer impact patterns of each variable set when they are computed using smaller unit areas than when they are computed using larger unit areas. Smaller unit areas provide closer approximation of the variables and those variables seem to be less susceptible to measurement error than variables computed using larger geographical units. However, the trade-off between obtaining closely approximated explanatory variables and the computational effort required for smaller areal units has to be considered when we decide which areal unit we want to use. In fact, the improvement in the goodness of fit of some regression models was marginal or even totally absent. Moreover, the two aggregation levels used here have their own inherent advantages and disadvantages. Consequently, we also demonstrate building models using spatial variables from both geographic levels with some clear benefits in explanation and goodness of fit. Overall, however, land use density and supply of roadways are strong and significant explanatory sets of variables and they provide a good candidate for linking land use to travel behavior in policy impact assessments. In terms of efficiency and inequality, the regression models show that even when investments are done to improve efficiency and/or inequality they will impact different behaviors in different ways and their overall impact may not necessarily benefit individuals because different impacts in different facets of behavior may counteract each other. The total effect on the overall daily travel patterns of individuals and groups of individuals exceeds the scope of this project. The only tractable existing method to track these impacts is microsimulation (computer-based synthetic generation of activity and travel patterns of individuals), which is gaining popularity among practitioners.

We believe this project was an immense success as a feasibility pilot study. Existing data sources can be "mined" to extract general useful indications about efficiency and inequality. The same data sources can also be used to gain informative insights about travel behavior and to begin unraveling the complex relationships between infrastructure investments and behavior. Due to pragmatic considerations in the design of the tool presented here many limitations do not allow this tool to be used immediately as a planning support system for statewide policy and decision making. Early during the project design phase we discovered there was no comprehensive clearinghouse of statewide information about transportation projects that tracks them from their inception to the final implementation and impact assessment. Assembly of data from a variety of sources to build a database of all the transportation projects and their impacts would have exceeded the scope and time budget of this project. For this reason we approximated infrastructure supply using an inventory of highways in an existing network database. Similarly, we neglected accounting for public transportation facility and network supply. Moreover, we use as highway speed the reported speed limit for each network link, which we know does not represent prevailing speeds and varies throughout the day, days of the week, and many other seasonal rhythms. These considerations point to one of the next steps, which is to create a project that, on the one hand, builds a data warehouse of public and private investments and related projects and, on the other hand, develops a statewide multimodal network that is updated yearly with additions and added documentation about the quality of the infrastructure components represented by the network. Technology to accomplish both steps exists but institutional support is not readily available at this time.

The entire analysis was done using data from the year 2000. The data are from products such as the Census Transportation Planning Package and a roadway network vintage 2000. The household behavior data span a few months in 2000 and 2001. As a result all the analytical findings are for that period and may not be extendable to other times. This analysis should be expanded to include other years. Opportunities for new data are multiplying due to the American Community Survey, which in 2010 will most likely release its 5-year estimates for areas with a population of less than 20,000, including census tracts and block groups. This may provide an unprecedented opportunity to study the evolution of accessibility in our state and identify the places and their social and demographic groups that benefitted the most by pinpointing geographic areas that increased or decreased residents' accessibility. Comparisons between the year 2000 and 2010 will reveal changes over time and identify areas in California that benefited the most and areas that benefitted the least. If the project information warehousing activity mentioned above is accomplished, we could also distinguish between successful and unsuccessful projects using the tools and ideas in our project.

In the third major area of next steps we can expand the microanalysis to a more comprehensive treatment of travel behavior. This includes activity participation and interactions among household members, trip consolidation in the form of tours, and also the more traditional analysis of trip making. In addition to offering a more detailed picture of the impact that infrastructure and density of opportunities cause on travel behavior, this next step has also the potential to improve the statewide transportation model maintained by Caltrans. This last area of analysis is also a fruitful research direction in developing a next generation of land use transportation integrated models. This is an active area of graduate student and faculty research in the University of California Transportation Center (www.uctc.net).

The tasks in this report involved researchers from the University of California Santa Barbara (UCSB) and University of California at Irvine (UCI). The overall project principal investigator is Kostas Goulias at UCSB. At Irvine Tom Golob with assistance from James Marca extracted a travel behavior database from the California travel survey of 2000 and estimated the first round of travel behavior equations utilizing US Census tract level accessibility indicators. At UCSB Val Noronha and Bryan Krause converted network and US Census data into usable variables at the tract level, computed a first set of accessibility indicators, and developed maps in GIS. During the first part of the project and based on this work a variety of issues were identified, solutions sketched by Kostas Goulias, presented at a series of presentations, and finalized in the second part of the project. A second set of accessibility indicators based on the US census block group data were then computed by Seo Youn Yoon and Kostas Goulias at UCSB that also estimated a new set of travel behavior models. They also estimated the stochastic frontier models used in efficiency measurement. Emmanuel Kemmel, Seo Youn Yoon, and Kostas Goulias also developed the Theil index computations.

The first two sections of this report provide a brief presentation of the study background and design. The third section provides a summary efficiency measurement and computations using US census tract level data and a detailed road network as well as stochastic frontiers. In the fourth section we show the inequality assessment using US census block group data and the Theil computations. This is followed by the fifth section that shows distribution of past allocation of road infrastructure across a variety of sociodemographic segments. The sixth section is dedicated to a variety of model estimation experiments to show the impact of provision of infrastructure and accessibility on travel behavior. This is followed by a seventh section that demonstrates the use of spatial variables calculated at two different but nested geographic levels and the benefit of using them jointly. In the last two sections we provide a brief summary and an outline of three recommended next steps.

1. Introduction

Optimal allocation of resources for infrastructure facilities is a critical issue in planning for development but it is also a critical consideration for the every day life of travelers. In addition to optimal allocation, equally important is also the distribution of benefits in terms of infrastructure facilities (stock) and related quality of service intended here as the ability to reach desired destinations within an acceptable amount of time (service). Different regions of California have received over the years different levels of investment for private or public transportation. The residents at each of these regions are also "investing" time to travel from one location to another. These are inputs to a *production system* that has many outputs including local gross product (e.g., regional gross product) and time allocated by the residents to activities (e.g., time for paid work, time dedicated to leisure and so forth). Depending on local circumstances each region is more or less efficient in maximizing the use of these stock and service resources. Tools exist to judge how efficiently systems work but they focus on economic efficiency and they do not incorporate a comprehensive measure of transportation stock and service offered. Here, we emphasize social efficiency and bring measures of accessibility in the arsenal of resource management and resource allocation to show the degree of efficiency exhibited by different regions in enabling its residents to minimize personal costs and maximize personal benefits. The research findings presented in this report contain a two-component research program as mentioned in the preface above.

The state of California is divided into geographical areas and each is treated as a production unit with its inputs represented by the different types of infrastructure (e.g., lane miles of roadways classified in a finite number of types). The outputs are indicators of the service offered to the unit's residents in terms of the amount of activities the residents of each geographical area can reach. Figure 1 provides a summary of the schema used in this project.

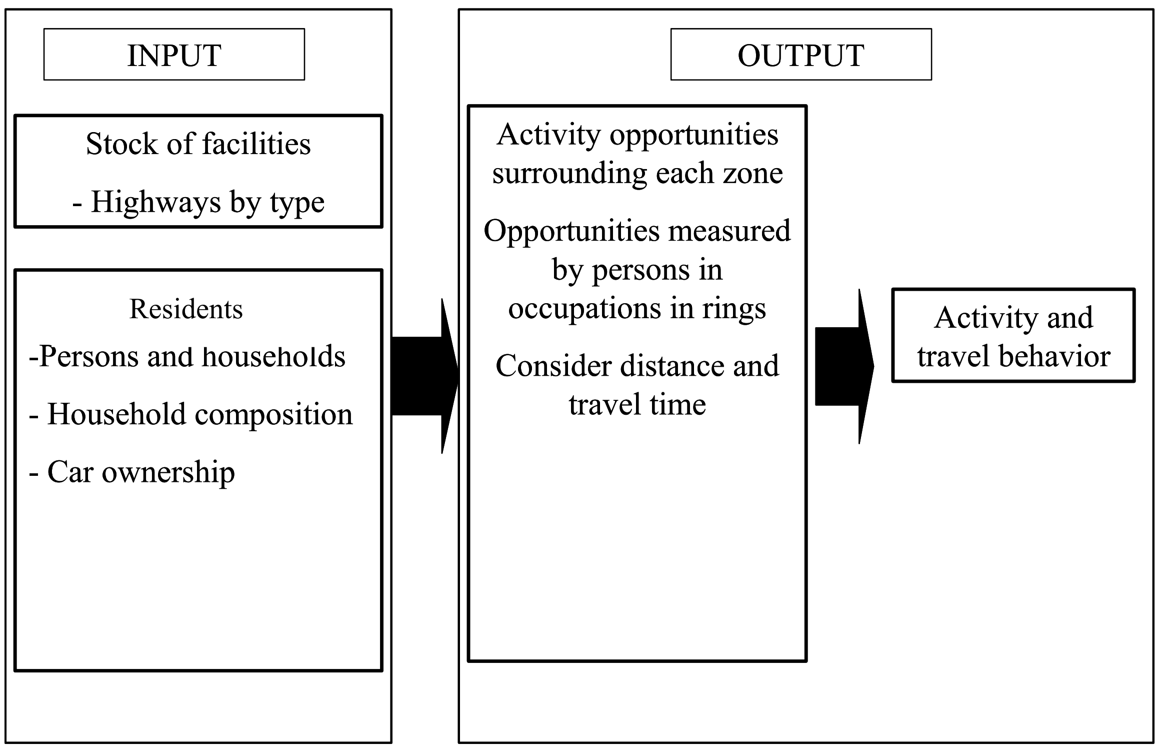


Figure 1 This project's schema

2. Background

Typical studies of transportation investment and economic development are discussed in Berechman (1994), Buffington et al. (1992), Perera (1990), Seskin (1990), and Weisbrod and Beckwith (1992). There are also regional studies addressing the impact of transportation infrastructure on local regional economic development. Assessment of these investments is based on the Gross Domestic (Regional) Product or private output as in Allen et al. (1988) and Wilson et al. (1985), benefit-cost ratios and/or differences as in Buffington et al. (1992) and Weisbrod and Beckwith (1992), property values as in Palmquist (1982) and new business creation or location as in Hummon et al. (1986). Analytical methods in these studies include: a) assessment of the effects of transportation infrastructure investments that compare and contrast the effects of investments among different regions; and b) identification of the important factors that influence and enable economic development. The study here belongs to the first group of analytical methods. Identification of the impacts from transportation infrastructure investment is particularly important when resources are scarce. From the perspective of decision makers, need assessment and accurate measurement of this need allows effective budgeting and financing of projects. It also allows for informed decisions while evaluating individual projects, balanced distribution of resources, and increased efficiency. Considerable research exists in the analysis of investment and optimal allocation of resources. Transportation improvements influence economic development, productivity, and social welfare. "Pure" economic development impacts are usually regional in nature and result from improved access to labor pools or to larger markets. While considering the economic development of different regions of a country, investment in transportation infrastructure as well as in the overall infrastructure system may play significant role in removing regional economic disparities. Within the same country and under the same development policies, significant role for transportation implies that regions with better transportation infrastructure will have better access to the locations of materials and markets making them more productive, competitive and hence more successful than regions with inferior transportation accessibility. Better accessibility and mobility also plays a significant role in human resource development of a region. For a review and an application using Data Envelopment Analysis, see Alam et al. 2004, an example of longitudinal analysis Alam et al. 2005, and a Stochastic Data Envelopment Analysis see Alam et al., 2008, and project by project economic assessment in Gkritza et al., 2008.

One could make similar arguments when considering the time expenditures of individuals and households to paid and unpaid work as well as free time with family and friends. However, transportation investment from a "social efficiency" viewpoint is absent from transportation practice. This is mainly due to the lack of tools capable to assess the role of transportation investment on the efficient allocation of time by the residents of each locality. The tool we aim with the analysis presented here identifies specific locations in the state where resource allocation has succeeded in maximizing benefits to the public. In addition, we aim to develop maps that show which locations in a state fail to be optimal and require their residents to travel excessively to pursue the same amount of activities as other residents of different localities.

More specifically in this report, we answer four key questions:

- Using largely available data, can we develop a small number of variables to describe access to activity opportunities for California residents?
- Are more roadways improving access to these activity opportunities?
- Are these roles different for different types of highways and how?
- Can we identify roadways that are prime candidates for investment?

In this analysis the state of California is divided in 7049 zones using the US Census 2000 tracts. The Census tract (unit of analysis here) is selected as a first order geographical subdivision to make the analysis tractable at the state level and to provide sufficient detail to be meaningful (we will repeat this analysis with a smaller geographic unit and revisit this aspect in the conclusions). We assess each tract in terms of its ability to produce benefits for its residents. Figure 2 provides a schematic representation of the study and Table 1 contains a selection of unit of analysis characteristics.

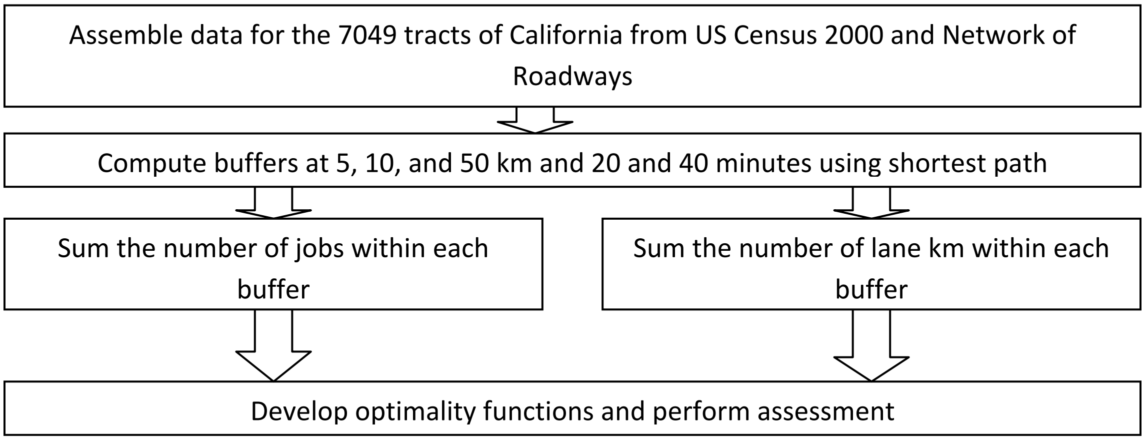


Figure 2: Computation Schema of the Study

Envisioning each tract as a production unit and developing for each tract a production function, we measure access to opportunities, treat them as outputs, and correlate them to the presence of roadways within and surrounding the tract. Access to opportunities for activity participation (e.g., leisure) and services (e.g., health) is the benefit (and output) from each tract that we will assess. Using Geographic Information Systems we compute for each tract the amount of activity opportunities reachable within 5 km, 5 to 10 km, and 10 to 50 km. We repeat the same for 20 minutes and 20 to 40 minutes travel time computed using information about speed limits on the roadway network at hand. Computation of these measures is accomplished by developing an origin-destination network with the origins and destinations as centers (population weighted virtual centroids in each tract). Using the same origin-destination network we also count the number of highways within 5 km, 10 km, and 50 km network distance from each centroid.

Table 1 A selection of Census-tract characteristics

	Mean	Std.Dev.	Maximum*
Tract Square Km	59.0	453.7	20486.8
Tract Population	4805.2	2143.1	36146.0
Tract Households	1631.8	763.0	8528.0
Within a 5 Km Buffer from Tract Centroid			
Workers in Retail (retail)	5031.1	6937.8	54745.0
Workers in Health (health)	2644.0	3524.4	26478.0
Workers in Services but not in Health or Retail (services)	28024.4	44497.0	373127.0
Workers in Manufacturing (manufacturing)	3391.0	5547.7	59059.0
Workers in All Other Occupations (other)	5753.4	6805.7	50287.0
Primary limited access roadways (primary lim)	284.1	448.6	3244.8
Primary without limited access roadways (primary nolim)	77.9	140.6	958.5
Secondary and connecting roadways (secondary)	1867.8	2711.3	17711.4
Rural, local and neighborhood roadways (local)	8549.4	11256.1	71318.1
Special roadways (special)	342.1	591.3	4612.7
All Other types of roadways (other)	778.6	1618.7	10511.1

*The minimum is zero for all variables and tracts

Enjoyment of access is also a function of the tract residents' ability to take advantage of opportunities offered to them. We attempt to capture this by including social and demographic characteristics of the resident population available in the Census tract databases. Transportation investment is often directed to facilities and the striking majority of this investment is allocated to roadways. An indicator of transportation supply (the input in the context of production functions) is the amount of roadways (lane kilometres). Roadways, however, serve different purposes and offer different functions to the users depending on their type (e.g., limited access freeways/motorways, secondary roads connecting limited access roadways, local roads).

Using Geographic Information Systems, we can identify and count the number of kilometres of each roadway in each tract. Roadways, however, form a complex network and the tracts are interconnected. For this reason, we perform a similar task as for activity opportunities and we count the number of roadways by type in a series of concentric rings of 5km, 5 to 10km, and 10 to 50km. We name these rings the *buffers*. We repeat the same operation for travel time using 20 minutes and 40 minutes travel time. The types of roadways we count are: primary highways with limited access (*primary lim* herein), primary roadways without limited access (*primary nolim* herein), secondary and connecting roadways (*secondary* herein), local and rural roads (*local* herein), roads with special characteristics (*special* herein), all other roadways (*other* herein). On the one hand, we have as input a detailed accounting of roadways representing all past investment on highways for each origin (tract centroid). On the other hand, we consider as output the number of workers a resident departing from a centroid can reach. The types of workers that are reachable within each of the buffers are classified into: retail, health, services, manufacturing, and all other. These counts are the indicators capturing access to opportunities to participate in activities and enjoy services.

3. Optimality Assessment

The literature on optimal assessment of decision making units is largely populated by Data Envelopment Analysis methods (a review on a related topic can be found in Alam et al. 2004, 2005, and 2008) and Stochastic Frontiers (Greene, 1980). Considering the possible measurement errors in the data used, the presence of outliers, and spatial correlation, we opt for stochastic frontiers that can handle some of these possibly undermining issues. However, an additional step is required in our analysis before estimating stochastic frontier production functions. The output of the number of workers that a resident departing from a centroid can reach is depicted by 25 indicators (number of workers in retail, health, services, manufacturing, and other within 5km, within the ring of 5 to 10 km, within the ring of 10 to 50 km, within 20 minutes of travel time, and within the ring of 20 to 40 minutes travel time). To reduce the data into a few variables we use factor analysis using the principal components method, extraction based on correlations, and the varimax method. This yields three components explaining 93% of the variation in the output variables used here. Each component captures a different aspect of access to opportunities surrounding each centroid and the three components are derived in such a way to be uncorrelated. Table 2 provides a summary of the component scores (high scores indicate high correlation between the output variable and the component extracted). The first component represents access of opportunities in the outermost ring between the radius of 50 km and the radius of 10 km but also within the ring defined by the radii of 20 and 40 minutes and for this named the *outer ring access* in this study. One variable, the number of workers in manufacturing within 20 minutes travel time, is more correlated with the first component than the second reflecting the predominant location of manufacturing in the outskirts of cities and closer to high speed roadways. The second component represents access to opportunities in the second ring and it is most correlated with variables defined in the ring between a radius of 5 km and a radius of 10 km (named *middle ring access* herein) and variables of within 20 minutes of travel time. Access to opportunities that are the closest to the centroid is represented by the third component (named *core access* herein), which is most correlated with the remaining variables. For each California tract we compute each of the three components (corresponding to three concentric regions around each centroid – core, middle ring, outer ring) using the scores of Table 2 and the value for each variable used to extract them. These three components replace the 25 variables and are used as the dependent variables in stochastic frontier analysis.

Table 2 The three principal components extracted from 25 output variables and their scores

	Components		
	Outer Ring Access	**Middle Ring Access**	**Core Access**
Number of Workers in Retail (20 to 40 min)	0.945	0.276	0.139
Number of Workers in Services (20 to 40 min)	0.941	0.250	0.128
Number of Workers in Other (20 to 40 min)	0.941	0.275	0.150
Number of Workers in Manufacturing (20 to 40 min)	0.939	0.245	0.130
Number of Workers in Health (20 to 40 min)	0.936	0.287	0.140
Number of Workers in Retail (10 to 50 km)	0.927	0.330	0.159
Number of Workers in Manufacturing (10 to 50 km)	0.926	0.311	0.129
Number of Workers in Other (10 to 50 km)	0.925	0.329	0.157
Number of Workers in Services (10 to 50 km)	0.924	0.326	0.163
Number of Workers in Health (10 to 50 km)	0.919	0.343	0.169
Number of Workers in Manufacturing (0 to 20 min)	0.665	0.625	0.265
Number of Workers in Services (5 to 10 km)	0.234	0.878	0.296
Number of Workers in Retail (5 to 10 km)	0.322	0.868	0.275
Number of Workers in Other (5 to 10 km)	0.380	0.841	0.289
Number of Workers in Health (5 to 10 km)	0.267	0.817	0.350
Number of Manufacturing in Services (5 to 10 km)	0.438	0.766	0.220
Number of Workers in Services (0 to 20 minutes)	0.504	0.703	0.430
Number of Workers in Health (0 to 20 minutes)	0.532	0.688	0.421
Number of Workers in Retail (0 to 20 minutes)	0.585	0.680	0.389
Number of Workers in Other (0 to 20 minutes)	0.605	0.672	0.345
Number of Workers in Services (0 to 5 km)	0.071	0.198	0.955
Number of Workers in Retail (0 to 5 km)	0.139	0.226	0.942
Number of Workers in Other (0 to 5 km)	0.190	0.325	0.871
Number of Workers in Health (0 to 5 km)	0.075	0.308	0.839
Number of Workers in Manufacturing (0 to 5 km)	0.289	0.354	0.699

Stochastic frontiers were developed for models of production. A production function is the ideal amount a unit can produce for a given set of inputs. In empirical settings observed outputs are not ideal (maximum) for reasons that are due to unknown random factors and measurement error (v) that are specific to each observed unit and due to productive inefficiency that also varies with each observed unit (u). To examine the relationship between output variables (access to opportunities) and input variables (highways) a regression model is created with dependent variable (y) the indicator of output and independent variables the highway lane kilometers (x). The model we use here takes the following form:

$$y_i \quad = \quad \alpha + x'_i \beta + v_i - u_i$$

Index i represents each tract, i=1,…, 7049.

We estimate three regression equations that are one for each of the three components of Table 2 (core access, middle ring access, outer ring access). In each equation y is the logarithm of the component values for each tract. The xs are number of highways of each type in each geographic subdivision. The vector β contains the regression coefficients we seek. Variable v is the usual random error term capturing measurement error and variable u is a positive valued offset between observed access and the ideal maximum possible given the input combination of roadways within each tract. The random error term v is assumed to be normally distributed with zero mean and constant variance across observations. The random positive valued term u is specified as a function of other explanatory variables. In the terminology of production functions the values u_i are the measures of inefficiency for each tract i in transforming lane kilometers of roadways into access to opportunities. Creating the $\exp(-u_i)$ we obtain a measure of tract specific efficiency.

Estimation of the three models presented here is carried out using LIMDEP (Greene, 2002). Table 3 shows the regression coefficients associated with each input variable (number of lane kilometers of roadway types in the core, the middle ring, and the outer ring). The correlation between the y variable and its predicted values using the estimated model coefficients is 0.895 for the outer ring, 0.731 for the middle ring, and 0.744 for the core, representing excellent goodness of fit between data and the production function derived here.

Table 3 Stochastic Frontier Regression Coefficients

	Outer Ring		Middle Ring		Core	
	Coeff.	**t ratio**	**Coeff.**	**t ratio**	**Coeff.**	**t ratio**
Constant	-0.413	-3.13	0.857	13.80	1.685	17.89
Log(primary lim in core)	-0.094	-1.71	0.203	11.17	0.443	13.01
Log(primary lim in core) ^{2}	-0.053	-2.29	0.070	8.48	0.135	8.95
Log(primary nolim in core)	0.016	0.23	-0.181	-8.11	0.477	10.25
Log(primary nolim in core) ^{2}	0.001	0.05	-0.039	-5.26	0.137	9.17
Log(secondary in core)	0.035	0.94	-0.195	-13.96	0.748	25.71
Log(secondary in core) ^{2}	-0.072	-5.71	-0.011	-2.07	0.172	19.97
Log(local in core)	-0.101	-3.75	0.091	8.28	-0.160	-7.86
Log(local in core) ^{2}	0.021	2.89	0.020	6.55	-0.100	-20.05
Log(special in core)	0.068	1.21	-0.190	-10.05	-0.145	-4.59
Log(special in core) ^{2}	0.045	2.02	-0.050	-5.91	-0.103	-6.92
Log(other in core)	-0.004	-0.22	-0.010	-1.53	-0.058	-5.36
Log(other in core) ^{2}	-0.003	-0.47	-0.001	-0.59	-0.024	-6.66
Log(primary lim in middle ring)	0.098	2.33	-0.020	-1.07	-0.115	-3.70
Log(primary lim in middle ring) ^{2}	0.077	5.83	-0.036	-6.60	-0.055	-6.56
Log(primary nolim in middle ring)	0.048	3.44	0.039	9.50	-0.082	-9.54
Log(primary nolim in middle ring) ^{2}	0.028	5.13	0.003	1.69	-0.047	-13.76
Log(secondary in middle ring)	-0.155	-3.18	0.146	6.08	-0.249	-6.01
Log(secondary in middle ring) ^{2}	-0.065	-6.13	0.044	9.09	-0.062	-7.06
Log(local in middle ring)	0.025	0.63	-0.014	-0.69	0.059	1.69
Log(local in middle ring) ^{2}	0.015	2.14	-0.020	-6.01	0.058	10.11
Log(special in middle ring)	-0.083	-1.78	0.085	4.37	-0.009	-0.25
Log(special in middle ring) ^{2}	-0.071	-5.22	0.061	11.10	0.012	1.28
Log(other in middle ring)	0.034	1.76	-0.005	-0.69	0.042	3.13
Log(other in middle ring) ^{2}	0.021	4.80	0.006	3.93	0.023	8.69
Log(primary lim in outer ring)	0.077	1.47	-0.012	-0.36	0.002	0.03
Log(primary lim in outer ring) ^{2}	-0.051	-2.56	0.025	2.71	-0.003	-0.20
Log(primary nolim in outer ring)	-0.071	-2.18	0.045	3.16	0.041	1.70
Log(primary nolim in outer ring) ^{2}	0.007	0.75	-0.018	-3.85	0.000	-0.06
Log(secondary in outer ring)	-0.041	-0.66	0.006	0.17	-0.008	-0.13
Log(secondary in outer ring) ^{2}	0.030	2.27	-0.019	-2.65	0.010	0.82
Log(local in outer ring)	0.066	1.40	-0.062	-1.73	0.006	0.12
Log(local in outer ring) ^{2}	0.007	0.90	0.003	0.62	-0.010	-1.33
Log(special in outer ring)	-0.090	-1.80	0.058	1.97	0.009	0.19
Log(special in outer ring) ^{2}	0.093	6.18	-0.019	-2.72	0.006	0.51
Log(other in outer ring)	0.012	0.47	0.005	0.42	0.018	0.82
Log(other in outer ring) ^{2}	-0.025	-5.63	0.002	0.74	-0.008	-2.18
Constant for u	-0.718	-8.06	-17.693	-14.36	-0.144	-3.18
Household density	-0.578	-69.66	1.059	10.34		
Tract perimeter (km)					-1.375	-22.05
σ_u / σ_v	3.797	28.05	13.069	17.89	2.612	45.34
$\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$	0.680	150.31	1.359	17.65	0.468	77.43

The signs, size, and significance of the regression coefficients show how the presence and amount of different types of roadways impact the ability of each geographical tract to provide access to opportunities. A negative sign associated with roadways in the same region (core, middle ring, outer ring) of the dependent variable is more likely to indicate competition for space with businesses and establishments providing services. A positive coefficient is more likely to indicate a clustering of establishments around those roadway types.

Positive coefficients associated with variables in different regions than the dependent variable indicate a supportive relationship with access. For example, access to the outer core may be achieved by driving over local roads in the core, secondary roads in the middle ring, and again local roads in the outer ring. Different establishments however, may be reached by different combinations of roadways. As a result we obtain a variety of significance levels, signs, and sizes of coefficients that may not all correspond to intuition.

As expected, access to the outer core is influenced by roadway quantity in the core, the middle ring and the outer core. However, lower speed facilities in the core (local and secondary roadways) seem to have a stronger influence than the higher speed (primary roadways). The middle ring primary roadways have a strong positive impact on access in the outer ring. These two indications are a reflection of the routes leading to the outer core with high presence of opportunities. However, if there are many primary roadways in the outer core they compete for space with the establishments were opportunities locate and this is reflected in a few negative coefficients associated with roadways in the outer ring (primary nolim and secondary). Access to the middle ring is even more heavily influenced by the amount and type of roads in the core (positively by high speed roadways and negatively by lower speed roadways).

The core access is not influenced by roadways in the outer ring, i.e., a driver does not need to go into the outer core when reaching places within the 5 km radius around a tract centroid and this is reflected in the lack of significance for most of the outer ring variables. In contrast, primary roadways in the middle ring seem to decrease access to the core in a significant way. This is a reflection of the spatial organization of California's roadway network and the spatial distribution of activity opportunities adjacent to the network's roadways. Unfortunately, all this is also masked by the use of the summary indicators (i.e., the principal components) as dependent variables that contain variables from all three regions (i.e., core, middle ring, and outer ring).

When aiming at improving access to opportunities around the core, however, provision of primary and secondary roadways appears to be a worthwhile investment. When we examine the other two components that are heavily influenced by variables that include travel time, the picture is not as clear and may be pointing out to the need for improving travel times in local and secondary roadways in regions that lead to the middle and outer rings.

The bottom portion of Table 3 contains the estimates of variables influencing inefficiency. Exp ($-u_i$) is a measure of technical efficiency and it is the ratio between achieved access over the maximum possible access achieved for the given inputs. The outer ring and middle ring efficiencies (and their opposite inefficiencies) are significantly different among tracts of different household densities (households per square kilometers). The core efficiency is a function of the perimeter of the tract indicating a possible problem with the use of tract as a unit of analysis. In a series of other specifications not shown here we also find that multi-car (>4) households live in tracts with lower efficiency presumably because they are able to combat lack of access with automobility. Other variables considered such as number of households by household size did not exhibit a clear trend. The median efficiency indicators are fairly high at 84%, 92%, and 81% for the outer ring, middle ring, and core respectively. The tenth lower percentiles are 72%, 83% and 62% for the outer ring, middle ring, and core respectively indicating a fairly good efficiency for a system that evolved without a major plan targeting high efficiency. However, considering the large size of many tracts access to opportunities may be quite different among the residents within these tracts (see also the inequality section below).

The final examination we perform for these computed efficiencies here is by mapping them for the entire state. Figure 3 shows the three efficiency indicators for Los Angeles, California, using as cutoff points the 10% percentiles. The first quadrant shows the Los Angeles total lane kilometers of roadways. Each efficiency estimate captures a different aspect of access to locations and shows clearly that providing more lane kilometers does not make a geographical area more accessible for any of the three efficiency measures.

These same efficiency estimates were also computed for the entire state. Figure 4a shows the core efficiency map at 10% percentile increments. Figure 4b shows the middle ring efficiency and Figure 4c shows the outer ring efficiency.

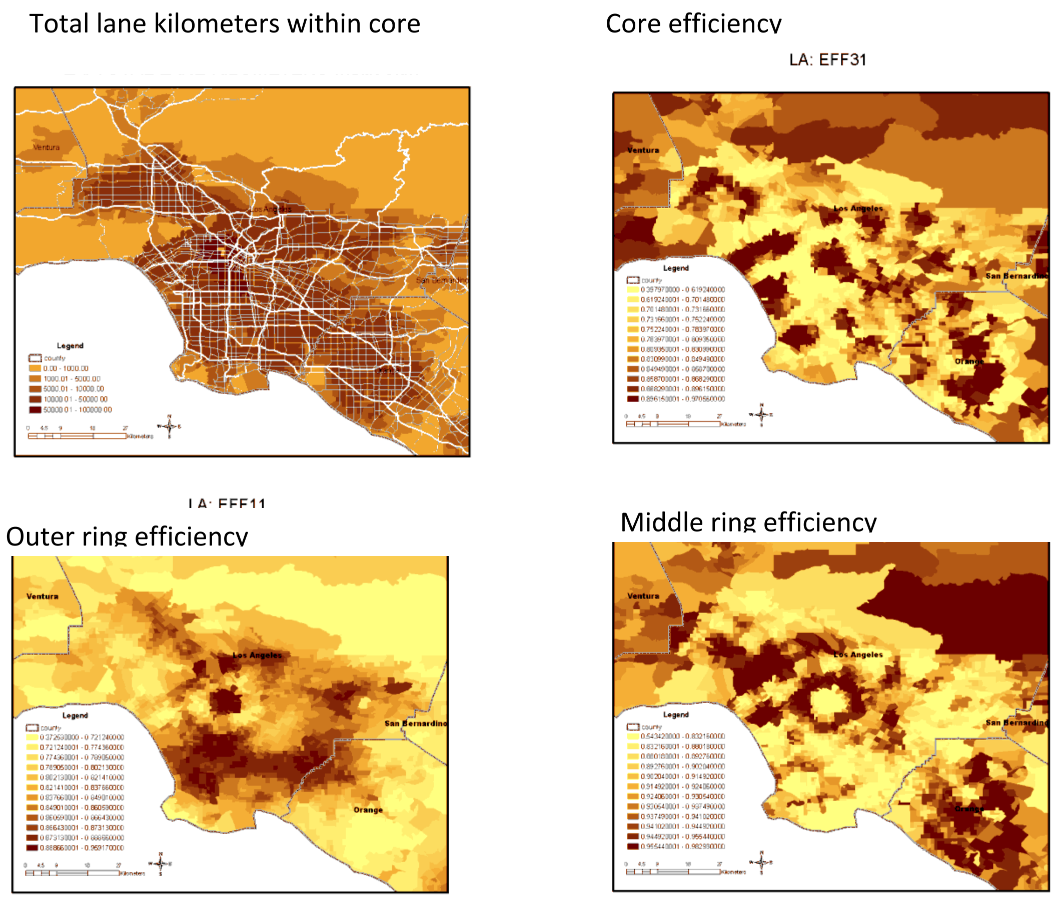


Figure 3: Maps of lane kilometers and efficiency measures in Los Angeles, California

Core efficiency

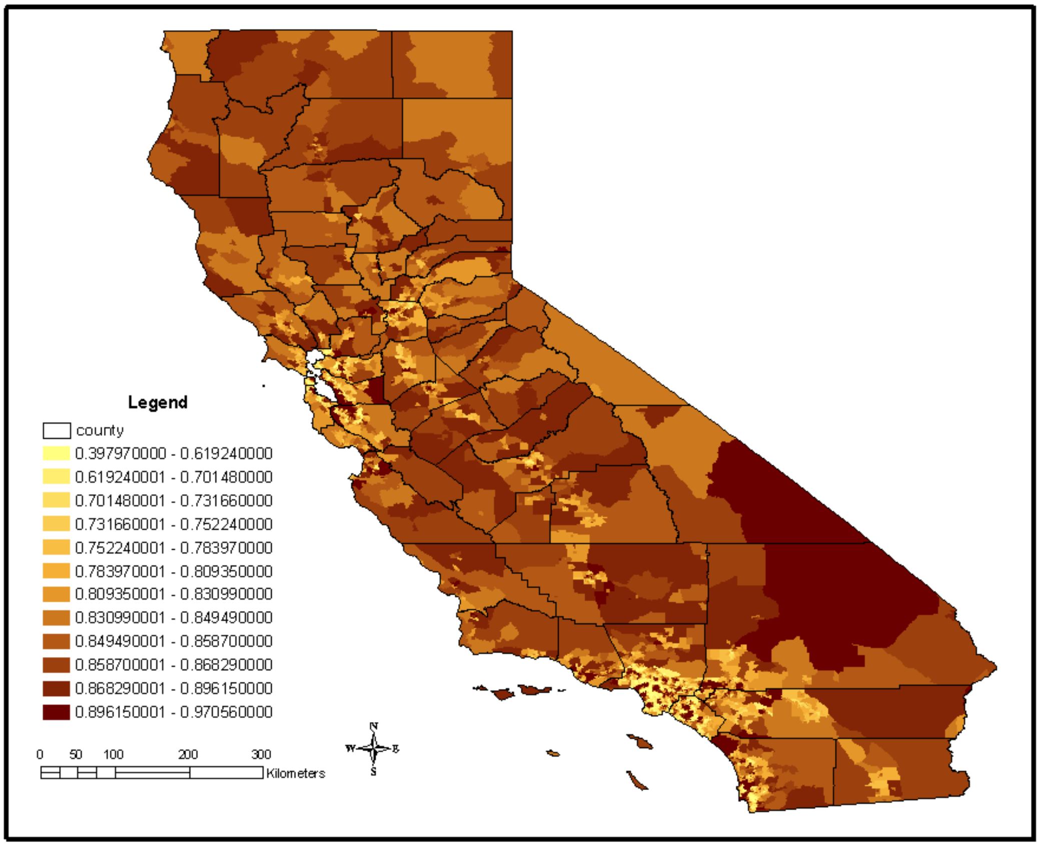


Figure 4a Core Efficiency Estimates

Middle ring efficiency

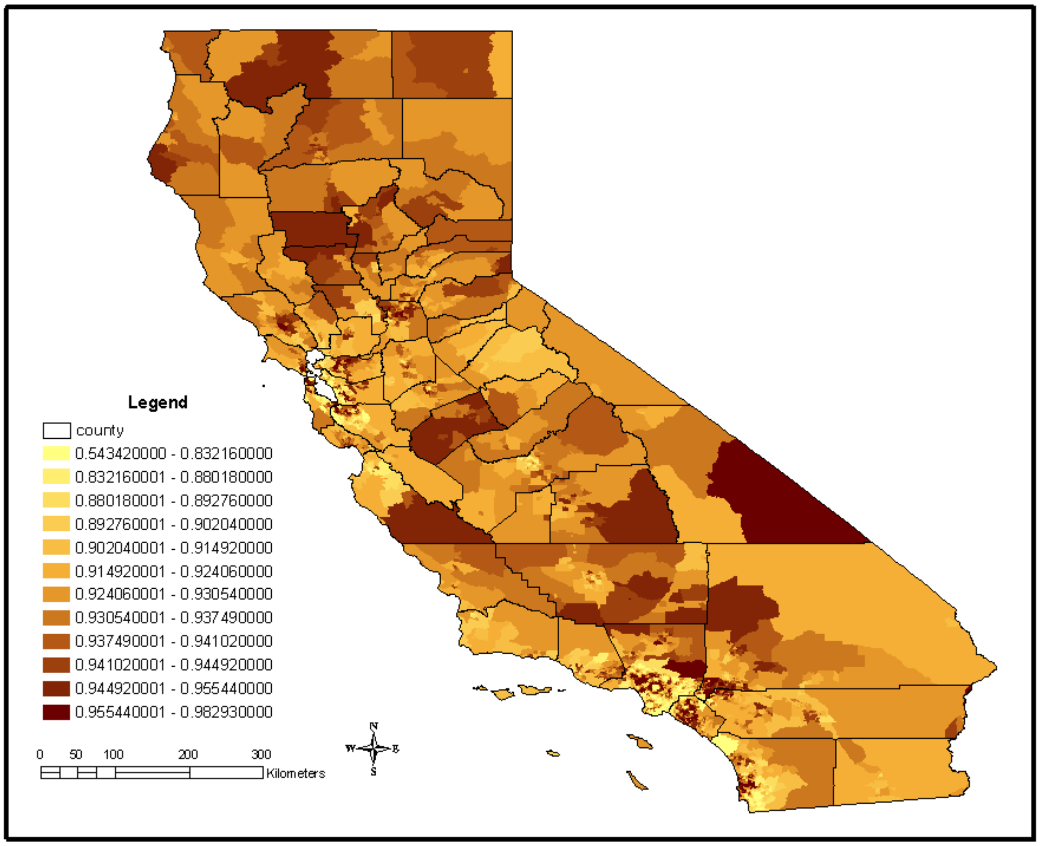


Figure 4b Middle Ring Efficiency Estimates

Outer ring efficiency

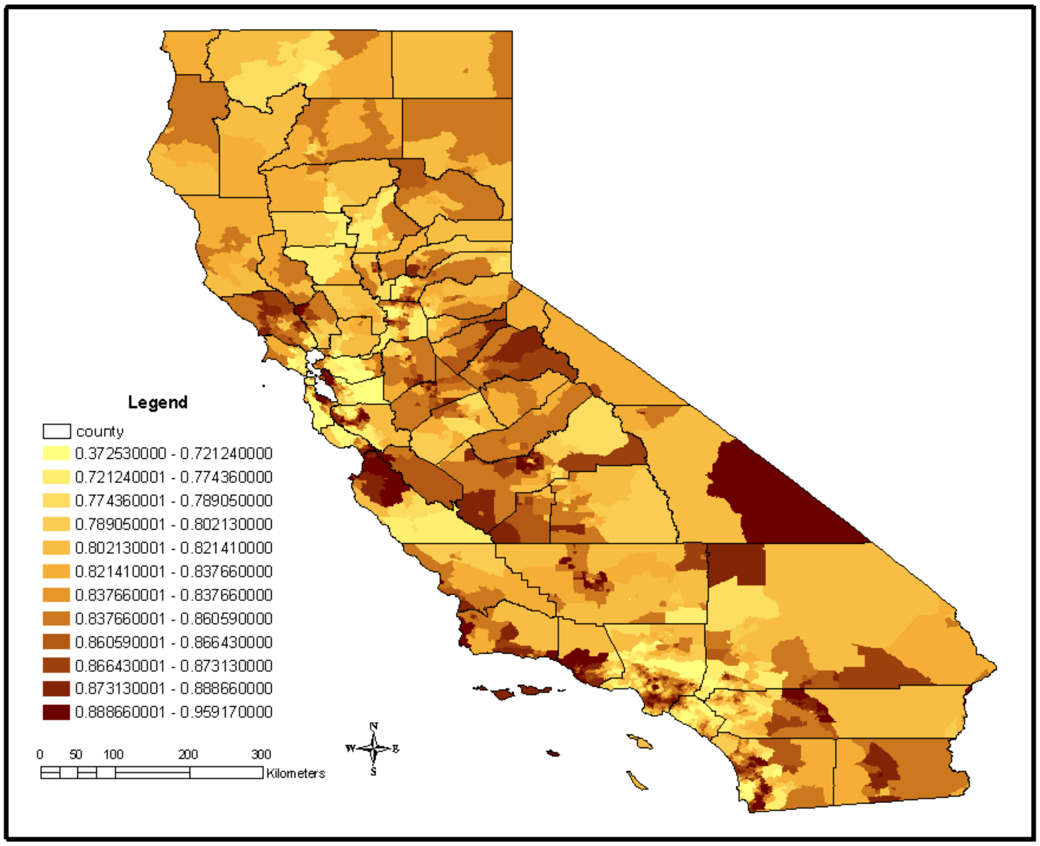


Figure 4C Outer Ring Efficiency Estimates

4. Inequality Assessment

In this section a method to highlight the mismatch that exists between the distribution of the population and the allocation of roads and activity access in California is presented. The tool we aim with the analysis presented here identifies specific locations in the state where resource allocation has succeeded in offering a uniform spatial spread of benefits to the public. In addition, we aim to develop maps that show which locations in a country (a state in our study) fail to be equitable, requiring their residents to travel excessively to pursue the same amount of activities as other residents of different localities. In this section, we answer a few key questions:

- Using largely available data, can we develop a small number of variables to describe access to activity opportunities for California residents?
- Is it possible to capture the structure of inequality in accessibility through a multi-scale analysis?
- Can we identify areas that are prime candidates for investment?

To answer these questions the state of California is divided in 22,133 zones using the US Census 2000 block groups. The Census block group (unit of analysis here) is selected as a first order geographical subdivision to make the analysis tractable at the state level and to provide sufficient detail to be meaningful. We assess each block group in terms of its ability to produce benefits for its residents and compare each block group with other block groups within a census tract. We will repeat the same comparison using tracts within counties, and counties within the state. Figure 5 provides a schematic representation of the study

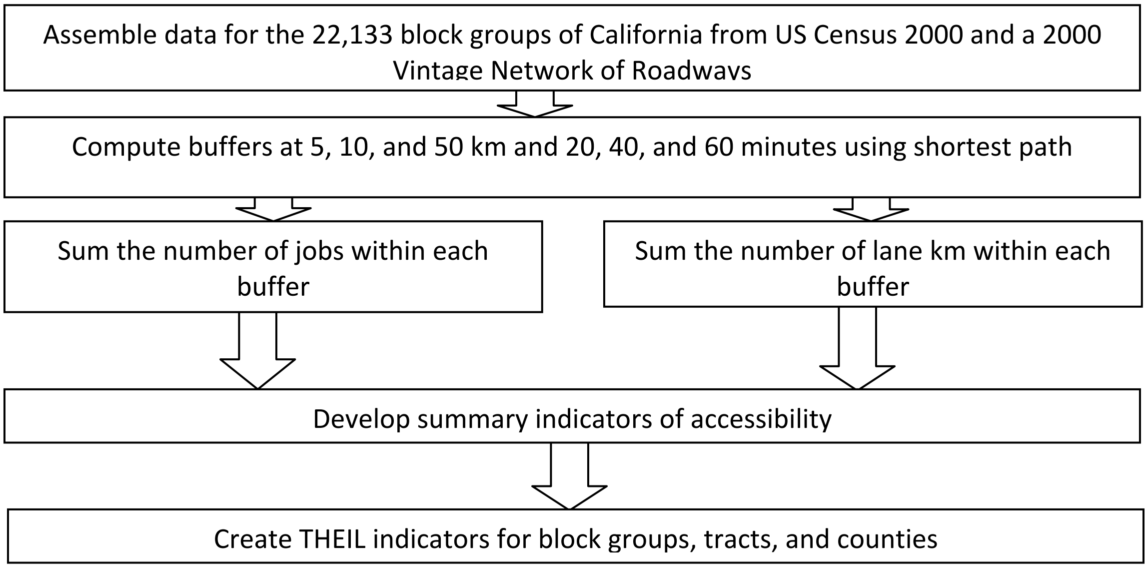


Figure 5: Computation Schema of the Inequality Study

Table 4 contains a selection of unit of analysis characteristics. Access to opportunities for activity participation (e.g., leisure) and services (e.g., health) is the benefit (and output) from each tract that we will assess. As indicators of available opportunities in a block group, numbers of workers classified according to the North American Industry Classification System (NAICS) were used. The original NAICS classification of fourteen types of industries was aggregated into five types: retail, health, services, manufacturing, and all other considering the types of activity in which people can participate related to the industries.

Using Geographic Information Systems (Network Analyst in ArcGIS 9.1), we identified the areas reachable within 20 minutes, 40 minutes, and 60 minutes travel time using information about speed limits on the roadway network at hand. The network data we used for the analyses have information about types of road network, segment length, speed limit, turning restrictions, and one way street enabling a somewhat realistic modeling of the travel environment. Identification of the reachable areas is accomplished by developing two sets of shortest path networks for the origin-destination matrix of the block group centroids using travel time and travel distance as travel cost respectively, and querying the block groups by the travel costs. Combining the reachable areas with the numbers of workers in each block group, accessibility to activity participation was calculated as enumeration of workers of each industry within each reachable area.

Table 4 A selection of block group characteristics

	Mean	Std.Dev.	Maximum*
Block Group Square Km	18.51	179.59	12219.12
Block group Population	1530.3	1008.48	36146
Block group Households			
Within a 20 min travel time buffer from block group Centroid			
Workers in Retail (retail)	56324.49	48926.91	202513
Workers in Health (health)	96664.34	89718.16	389816
Workers in Services but not in Health or Retail (services)	23812.89	23757.93	87798
Workers in Manufacturing (manufacturing)	80640.04	88937.65	339848
Workers in All Other Occupations (other)	75843.44	68947.56	270979
Primary limited access roadways (primary lim)	266.53	206.05	885.86
Primary without limited access roadways (primary nolim)	78.4	82.01	552.42
Secondary and connecting roadways (secondary)	650.52	425.51	2333.31
Rural, local and neighborhood roadways (local)	2561.13	1782.39	12545.59
Special roadways (special)	23.2	39.44	483.4
All Other types of roadways (other)	223.78	275.34	1984.31

*The minimum is zero for all variables and tracts

In a similar way as was done for the tract level, transportation supply is represented by the amount of roadways (lane kilometers) by type (e.g., limited access freeways/motorways, secondary roads connecting limited access roadways, local roads) but this time measured at the level of a US census block group. Using Geographic Information Systems, we can identify and count the number of kilometers of each roadway type in each block group. Roadways, however, form a complex network interconnecting the block groups and through the roadway network the block groups provide activity opportunities to others and also get benefits from others. For this reason, we perform a similar task as for activity opportunities and we sum up the length of roadway segments by type in a series of concentric areas that are accessible in 20 minutes, 40 minutes, and 60 minutes of travel time to quantify roadways that are available from an origin that is considered here as a virtual center of the block group (named centroid). We name these areas the *buffers* (similarly to the process followed in the previous section). The types of roadways we count are: primary highways with limited access (*primary lim* herein), primary roadways without limited access (*primary nolim* herein), secondary and connecting roadways (*secondary* herein), local and rural roads (*local* herein), roads with special characteristics (*special* herein), all other roadways (*other* herein).

On one hand, we have as input a detailed accounting of roadways representing all past investment on highways for each origin and the number of workers a resident departing from a centroid can reach. These counts are the indicators capturing access to opportunities to participate in activities and enjoy services. On the other hand, the main beneficiaries of transportation policies are the number of persons residing in an origin block group. One objective in transportation is to maximize accessibility for most persons. However, some segments of the population receive lower benefits than others. Inequality assessments are needed then to make comparisons.

The assessment of inequality is very often limited to a few disadvantaged population segments (Blumenberg, 2008 - http://www.opportunitycars.com/articles/documents/20051205_Blumenberg.pdf) and they do not encompass an entire state or country in their assessment. In contrast, inequality is a very popular subject in other fields (Krugman and Venables, 1995, Schneider et al., 2002, Ghose, 2004). Considering the strong spatial correlation among accessibility indicators (due to the connectivity of highway network and the agglomeration of businesses) we opt for an index of inequality that has a "fractal" nature (i.e., decomposable geographically) and that can handle multiple output variables.

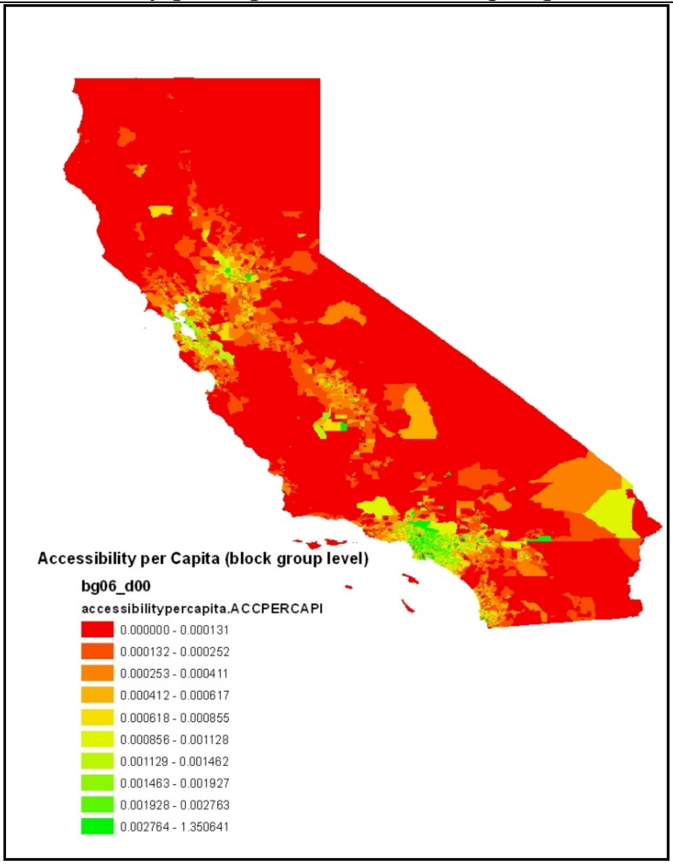
The output of the number of workers that a resident departing from a centroid can reach is depicted by 25 indicators that are: number of workers in retail, health, services, manufacturing, and other employment within 5km, within 10 km, within 50 km, within 20 minutes of travel time, within 40 minutes of travel time, and within 60 minutes of travel time. To reduce the data into a few variables we use factor analysis using the principal components method and extraction based on correlations in the same manner we did for the tracts in the efficiency analysis. During a first stage using all the variables this method produced a few variables that were only marginally informative (as expected due to the strong relationship among the 25 variables considered here) and they were eliminated from further analysis. The reduced set of variables considered in this analysis produced only one factor that captures 90.03% of the variation in the variables used here. Table 5 provides a summary of the component scores (high scores indicate high correlation between the output variable and the component extracted). For each California block group we compute this "accessibility " factor (a_i, i=1,…, 22,133). Figure 6 shows the ratio

a_i/n_i with n_i the resident persons in each block group. These figures show the disparities that exist in providing accessibility at each block group. The figures, however, do not reflect the relationship of accessibilities between block groups and do not provide an indicator that compares them directly with the overall accessibility of the state and its spatial structure.

Table 5 The factor created using a reduced set of the 25 output variables and their scores

Variable	Loading for accessibility factor
NUMBER OF WORKERS IN MANUFACTURING INDUSTRY (WITHIN 20 MINUTE BUFFER)	0.8669
NUMBER OF WORKERS IN RETAIL INDUSTRY (WITHIN 20 MINUTE BUFFER)	0.9233
NUMBER OF WORKERS IN EDUCATION/HEALTH SERVICE INDUSTRY (WITHIN 20 MINUTE BUFFER)	0.8957
NUMBER OF WORKERS IN OTHER INDUSTRY (WITHIN 20 MINUTE BUFFER)	0.8595
NUMBER OF WORKERS IN MANUFACTURING INDUSTRY (WITHIN 40 MINUTE BUFFER)	0.9675
NUMBER OF WORKERS IN RETAIL INDUSTRY (WITHIN 40 MINUTE BUFFER)	0.9881
NUMBER OF WORKERS IN EDUCATION/HEALTH SERVICE INDUSTRY (WITHIN 40 MINUTE BUFFER)	0.9828
NUMBER OF WORKERS IN PUBLIC ADMINISTRATION INDUSTRY (WITHIN 40 MINUTE BUFFER)	0.9538
NUMBER OF WORKERS IN OTHER INDUSTRY (WITHIN 40 MINUTE BUFFER)	0.9757
NUMBER OF WORKERS IN MANUFACTURING INDUSTRY (WITHIN 60 MINUTE BUFFER)	0.9640
NUMBER OF WORKERS IN RETAIL INDUSTRY (WITHIN 60 MINUTE BUFFER)	0.9719
NUMBER OF WORKERS IN EDUCATION/HEALTH SERVICE INDUSTRY (WITHIN 60 MINUTE BUFFER)	0.9700
NUMBER OF WORKERS IN PUBLIC ADMINISTRATION INDUSTRY (WITHIN 60 MINUTE BUFFER)	0.9490
NUMBER OF WORKERS IN OTHER INDUSTRY (WITHIN 60 MINUTE BUFFER)	0.9704
PRIMARY HIGHWAY WITH LIMITED ACCESS(WITHIN 20 MINUTE BUFFER)	0.9313
LOCAL, NEIGHBORHOOD, and RURAL ROAD(WITHIN 20 MINUTE BUFFER)	0.9000
PRIMARY HIGHWAY WITH LIMITED ACCESS(WITHIN 20 MINUTE BUFFER)	0.9852
LOCAL, NEIGHBORHOOD, and RURAL ROAD(WITHIN 20 MINUTE BUFFER)	0.9798
PRIMARY HIGHWAY WITH LIMITED ACCESS(WITHIN 40 MINUTE BUFFER)	0.9570
SECONDARY and CONNECTING ROAD(WITHIN 40 MINUTE BUFFER)	0.9478
LOCAL, NEIGHBORHOOD, and RURAL ROAD(WITHIN 40 MINUTE BUFFER)	0.9439
SECONDARY and CONNECTING ROAD(WITHIN 40 MINUTE BUFFER)	0.9760

Accessibility per capita in each block group



Theil index contribution by each block group

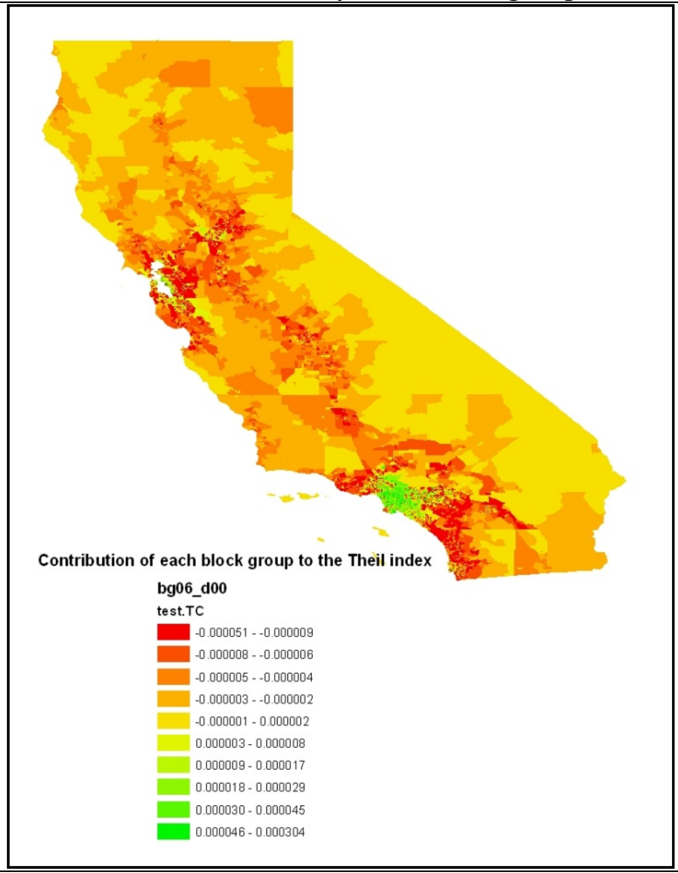


Figure 6 Accessibility and Theil Maps

Under ideal data availability we would like to identify every resident of California, compute an accessibility index associated with each resident and then perform a comparative analysis to assess who enjoys higher accessibility and who does not. Although this is not an impossible task with today's modeling and simulation capabilities, it violates one of the initial requirements of this study of using largely available data to explore new techniques. In addition, accessibility of one location is related to the accessibilities of its neighbors. We start with block group subdivisions and compute an indicator that accounts for the distribution of accessibility. We then consider increasingly larger geographical areas to illustrate the use of the Theil index. The following equation shows the Theil index computed using the block group data in California.

$$T = \sum_{i=1}^{22,133} \frac{a_i}{A} \log \left(\frac{\frac{a_i}{A}}{\frac{n_i}{N}} \right)$$

Where A is the sum of factor values for the entire state of California ($A=\Sigma a_k$) and N the population of the entire state of California ($N= \Sigma n_k$). For each block group i, we name respectively accessibility share and population share the ratios a_i/A and n_i/N .

An important advantage of this index over other measures of inequality is its composition. Each component of the sum in the equation above is a weighted log ratio of the accessibility over the resident population in the block group. Each component in the Theil index is then a weighted measure of the mismatch between its accessibility share and its population share. Thus, our interest will focus on each term of the sum, which we name contribution of the block group to the Theil index, or Theil contribution.

The right hand side of Figure 6 displays these Theil contributions for each block group. This map is more instructive than the left hand side one since the block groups are compared to each other which allows to identify the relative status of each area as compared to the entire state in possible mismatches. The block groups colored in yellow are those that bring little of no contribution the Theil index. That means they can enjoy accessibility to roads and activities opportunities in the right proportion with respect to their population. On the other hand, the green colored areas are those that have an accessibility share higher that their population share offering excess advantage. This "over accessibility" is on the detriment of the red color areas for which the accessibility share is smaller than the population share. Consequently the inhabitants of those block groups may have to spend more travel time to accomplish the same amount of every day activities than their counterparts residents who live in advantaged areas. As far as infrastructure investment is concerned, a public policy aiming at an homogenous development of the state of California should consider the red colored areas as prime candidates for roadway connectivity funding allocation (of course other factors are usually taken into account in allocating resources). Figure 6 shows that major metropolitan areas such as Los Angeles are particularly advantaged in accessibility, but it seems that this over-accessibility was built at the detriment of the block groups that compose their outskirts and the ones that are situated in the central part of the State. It should be noted, however, that travel time here is computed based on the speed limit of roadways and therefore does not account for congestion. As a result this "advantage" of the urban core is somewhat exaggerated in this analysis.

A fractal version of Theil's index enables assessment of inequality across larger regions as well as within larger regions to account for highway and land use connectivity. This is indeed the main characteristic that made us prefer the Theil index to all the other indexes developed in the economics literature. It is decomposable through different levels (e.g., geographical scales) and considers, for each scale unit, a between unit component and an intra unit component. In this way we can also account for heterogeneity within a larger area. As already mentioned above, a better way to measure inequality would be to consider each resident, but since, as most of analysts, we are dealing with groups, we have to study how inequality emerges between and inside these groups. Moreover this fractal approach gives us a deeper understanding of the spatial structure of inequality through the different levels we study.

In our case study, the different geographical units we consider are the following: the County, the Tract and the Block group. Figure 7 displays the tree structure of the recursive calculation here. The state is composed of 58 counties. Each county contains tracts and each tract contains block groups. The general definition of the fractal Theil index is the following (Conceicao and Ferreira, 2000).

$$T_r = \sum_{i=1}^{\text{Branches}} \frac{a_i}{A} . \log\left(\frac{\frac{a_i}{A}}{\frac{n_i}{N}} \right) + \sum_{i=1}^{\text{Branches}} \frac{a_i}{A} . T_{(r,\,i)}$$

Where ai is the accessibility of the branch i of the root r, A the total accessibility of the root r, ni the population of the branch i, N the total population of the root and T(r,i), the Theil index of the branch i.

Applied to our case study, the formula becomes:

$T_{CA} = \sum_{i=1}^{NC} \frac{a_i}{A_{CA}} \cdot \log\left(\frac{\frac{a_i}{A_{CA}}}{\frac{n_i}{N_{CA}}}\right) + \sum_{i=1}^{NC} \frac{a_i}{A_{CA}} \cdot T_{\text{County } i}$	Decomposition of the state by each county
$T_{\text{County}i} = \sum_{\text{Tract } j \text{ of Counnty}i} \frac{a_j}{A_{\text{Couty}i}} \cdot \log\left(\frac{\frac{a_j}{A_{\text{Couty}i}}}{\frac{n_j}{N_{\text{County}i}}}\right) + \sum_{\text{Tract } j \text{ of County}i} \frac{a_j}{A_{\text{Couty}i}} \cdot T_{\text{Tract } j}$	Decomposition of each county by each tract within the county
$T_{\text{Tract } j} = \sum_{\text{Blockgroup } k \text{ of Tract } j} \frac{a_k}{A_{\text{Tract } j}} \cdot \log\left(\frac{\frac{a_k}{A_{\text{Tract } j}}}{\frac{n_k}{N_{\text{Tract } k}}}\right)$	Decomposition of each tract by block group within the tract

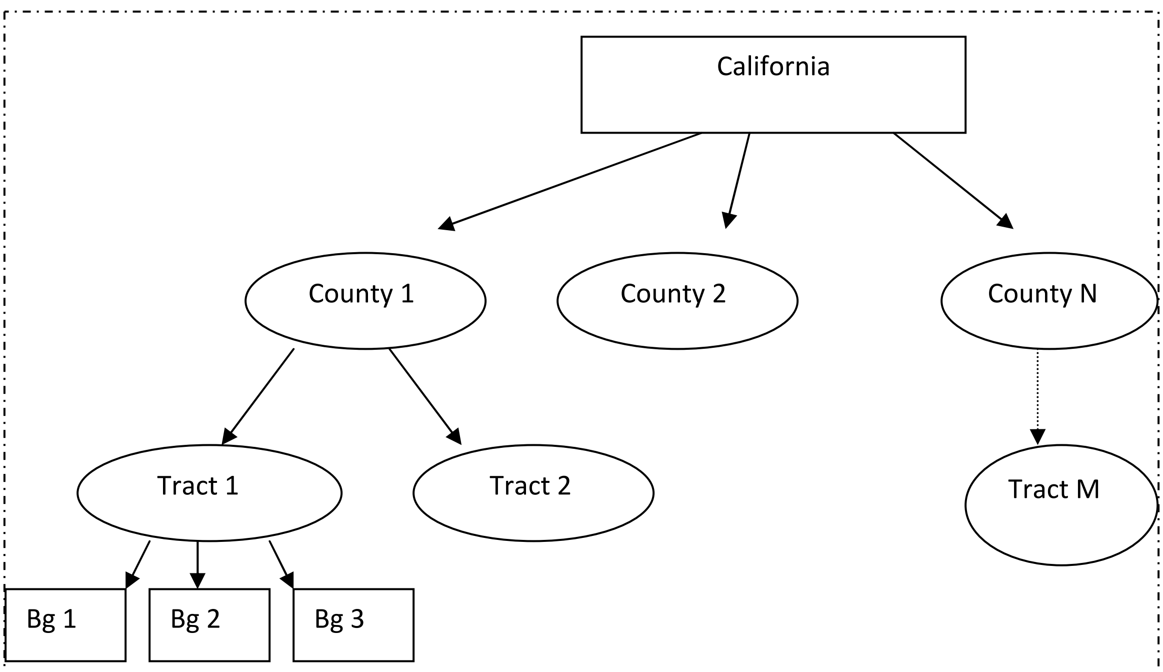


Figure 7: Tree structure used for the computation of the fractal Theil index.

Bringing all these components into one equation leads to the following.

$$T_{CA}=\sum_{i=1}^{NC}\frac{a_i}{A_{CA}}.\log\left(\frac{\frac{a_i}{A_{CA}}}{\frac{n_i}{N_{CA}}}\right)+\sum_{i=1}^{NC}\frac{a_i}{A_{CA}}.\left[\sum_{\text{Tract } j}\frac{a_j}{A_{\text{Couty } i}}.\log\left(\frac{\frac{a_j}{A_{\text{Couty } i}}}{\frac{n_j}{N_{\text{County } i}}}\right)+\sum_{\text{Tract } j}\frac{a_j}{A_{\text{Couty } i}}.\left(\sum_{\text{Blockgroup } k}\frac{a_k}{A_{\text{Tract } j}}.\log\left(\frac{\frac{a_k}{A_{\text{Tract } j}}}{\frac{n_k}{N_{\text{Tract } k}}}\right)\right)\right]$$

T_{CA} = Between Countiescontribution	Intra county contribution	
	Between tracts contribution	Intra-tract contribution

Figures 8, 9 and 10 display the results from this equation. Figure 10 is a statewide summary that displays two kinds of information. First, the contribution of the County to the Theil index, i.e the measure of the mismatch that exists between its accessibility share and its population share toward the other Counties. The other information is an “intra County” contribution that is actually its own Theil index and measures the inequality that exists between and inside its own tracts. Consequently, this map allows us to see, not only how advantaged or disadvantaged a County can be in regard to the others but also if its resources have been equally or unequally allocated showing the main advantage of the Theil index. It allows to understand the structure of inequality and its distribution through different geographic levels, and can thus constitute a decision making tool for public policies. Indeed, this map enables a policy maker to identify both what are the areas that need the most transport infrastructure for an egalitarian development of the State, and which regions have allocated their investments to projects that grant an homogeneous development of their own territory. The map allows to decide if a statewide equality will be emphasized and investments need to be made accordingly or if combating local inequality is more important and investments need to be made at a more local and focused way.

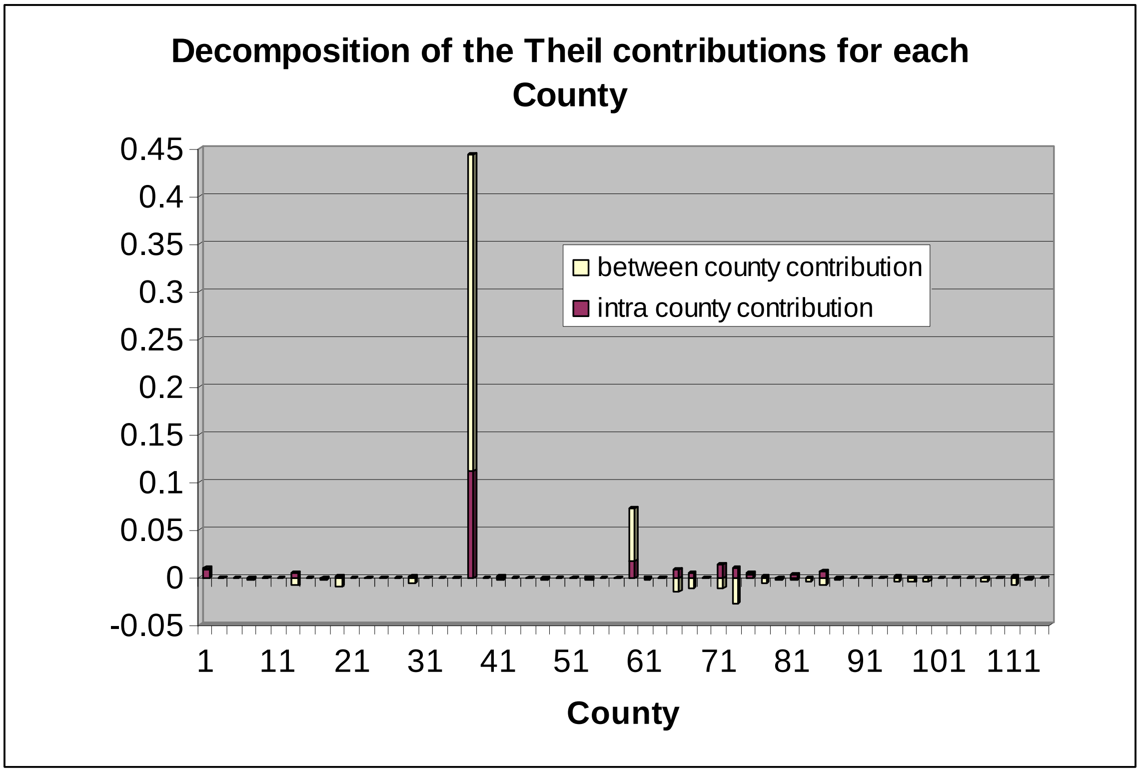


Figure 8 Decomposition of the Theil contribution of each California County

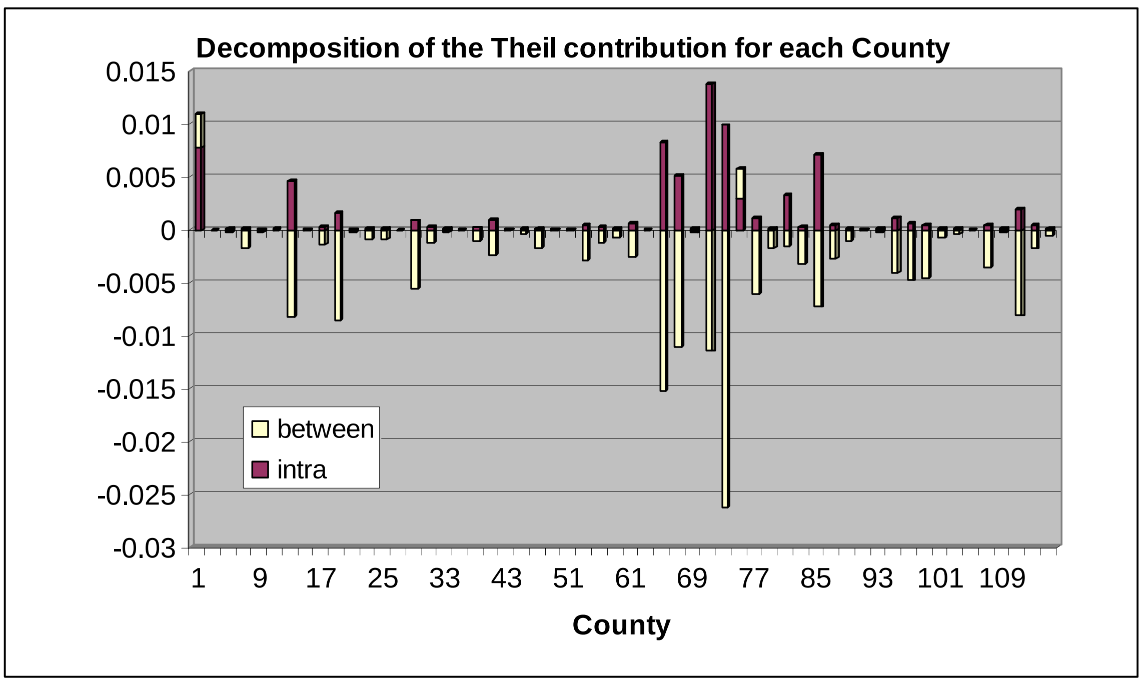


Figure 9 Decomposition of the Theil contribution of each California County without Los Angeles (37) and Orange (59)

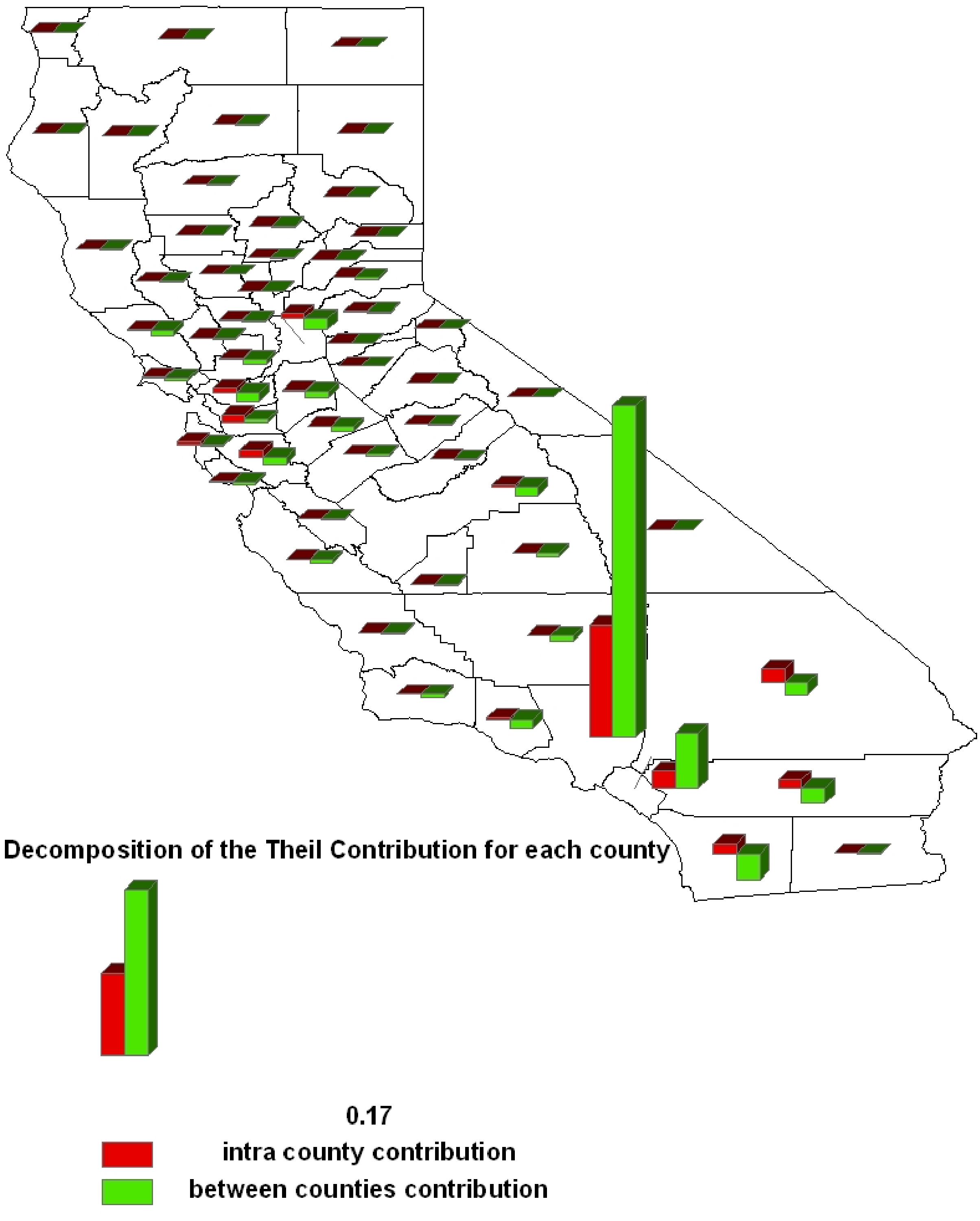


Figure 10 Map of the decomposition of the Theil contribution of each County

In California the most evident phenomenon that appears is the supremacy of the County of Los Angeles and Orange County in terms of both “over accessibility” (with, for LA, a contribution almost 45 times larger than the one of the third most advantaged County) and intra inequality (an intra inequality index about ten times larger than the one of the third most inhomogeneous County). This illustrates a property of the Theil index, its sensitivity to distributional impacts and disparities among the groups considered and in particular to "wealth" transfers from the disadvantaged to the advantaged. The measure of the mismatch is indeed amplified by the accessibility share weight (Conceicao and Ferreira, 2000, pages 12 and 13). Of course there is also a scale effect in all this. The larger a County is, the more likely it is to have internal heterogeneity.

Among the other Counties, there is another trend that is worth noting. Counties that show the most lack of accessibility are also those with the highest intra-county inequality. This points out to the need for a more detailed study to identify those disadvantaged counties that did not benefit from large scale infrastructure investment such that would have allowed them to develop a coherent policy for an homogeneous development of their territory. The findings here show some sort of negative feedback; the less investment a County receives, the more it is likely to suffer from territorial disparities.

5. Microanalysis (Person Based) Analysis

In the development of the microanalysis in this project, we have identified relationships between travel, household sociodemographic characteristics, spatial accessibility, and road infrastructure. When considered separately, sociodemographic characteristics, spatial accessibility, and road infrastructure all influence travel behavior. Dense urban areas make walking trips more feasible; extensive networks of freeways and arterials encourage vehicular trips; large households make more trips per day than small households, and so on. However, in the real world, all of these variables interact simultaneously. Households consider the costs and benefits of different locations and feasible travel modes in light of their circumstances, and choose residential locations accordingly. Indeed, one could argue that households are not merely reacting to their circumstances, but rather are actively trying to improve their lot in any way they can. Adjustment strategies include moving residence, changing jobs, choosing different travel destinations, bundling individual single-occupancy vehicle (SOV) trips into high-occupancy vehicle (HOV) trips, and so on. One cannot merely consider the influence of spatial infrastructure characteristics in isolation and this motivates the development of regression analyses attempting to take into account multiple factors.

One source of information about individuals and their households is the California Statewide Travel Survey, conducted over several months in the years 2000 and 2001. It provides an excellent starting point for disentangling the relationships between space, infrastructure, and sociodemographics. The survey sample, consisting of more than 17,000 households, is a quota sample by county and planning region, rather than a representative sample of California proportional to the population of each county. Each trip destination has been geocoded, usually to the nearest intersection, but sometimes to the approximate census tract centroid. The location (geocodes and census tract) of almost every household can also be determined from the survey data. To this data have been added spatial accessibility variables and roadway infrastructure variables by census tract and block group computed in the efficiency and inequality analysis discussed in previous sections of this report. The relatively even distribution of the sample across all California counties ensures that the data represent a wide variety of spatial environments.

The need to account for (control) for sociodemographics when assessing relationships between travel behavior and spatial factors is revealed when we investigate the residential location patterns of sociodemographic groups. Through a series of statistical tests, we determined that eight categorical sociodemographic variables were paramount in explaining travel behavior. These variables with their categories and distributions by percent of the sample are listed in Table 6. Each of these five variables was strongly related to our spatial variables through residential location.

Table 6 Sociodemographic Variables Used in the Models

Variable	%	Variable	%	Variable	%
Annual Household income		Average age of heads		Highest education of head	
<$10,000	4.3	18-25	5.8	not high school	9.1
$10,000-$24,999	14.2	25.5-35	14.1	high school graduate	24.5
$25,000-$34,999	13.2	35.5-45	20.1	Some college	23.7
$35,000-$49,999	13.9	45.5-55	22.7	associates degree	7.4
$50,000-$74,999	19.9	55.5-65	15.5	bachelors degree	20.9
$75,000-$99,999	10.9	65.5-75	11.8	graduate degree	13.4
$100,000-$149,999	7.4	75.5+	7.5	Unknown	1.1
$150,000+	3.4	Unknown	2.5	Whether any children < 6	
unknown	12.8	Ethnicity of heads		Yes	7.5
Household size		White	75.5	No	89.4
1	26.4	Hispanic	10.2	Whether any children 6-12	
2	40.8	Black	2.3	Yes	9.3
3	14.4	Asian/Pacific Islander	1.9	No	85.6
4	11.2	White & Hispanic	3.1	Whether any children 13-17	
5	4.7	White & Asian	1.3	Yes	9.0
6 or more	2.5	Other or unknown	5.8	No	2.9

The residential location patterns for various demographic groups can be seen by graphing the category means of four key spatial variables for each of our five polychotomous sociodemographic variables, as shown in Figures 11 through 15. These four spatial variables are: (1) housing density, in terms of households per square kilometer, (2) regional retail accessibility, in terms of total retail workers within 10 and 50 kilometers, (3) local road infrastructure, in terms of total kilometers of local, neighborhood, and rural roads within 10 kilometers, and (4) regional non-freeway primary road infrastructure, in terms of total kilometers of primary roads without limited access within 10 to 50 kilometers. Each of these four spatial variables are standardized (zero mean and standard deviation of one) to allow plotting on a single scale.

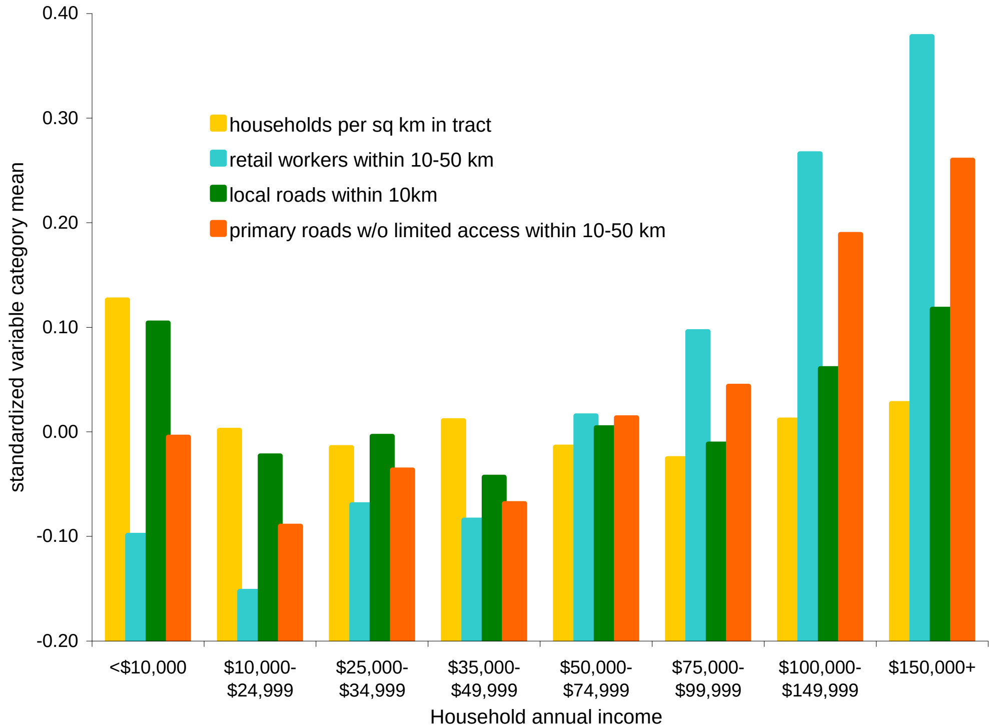


Figure 11 Some Spatial Variable Means by Categories of Household Income

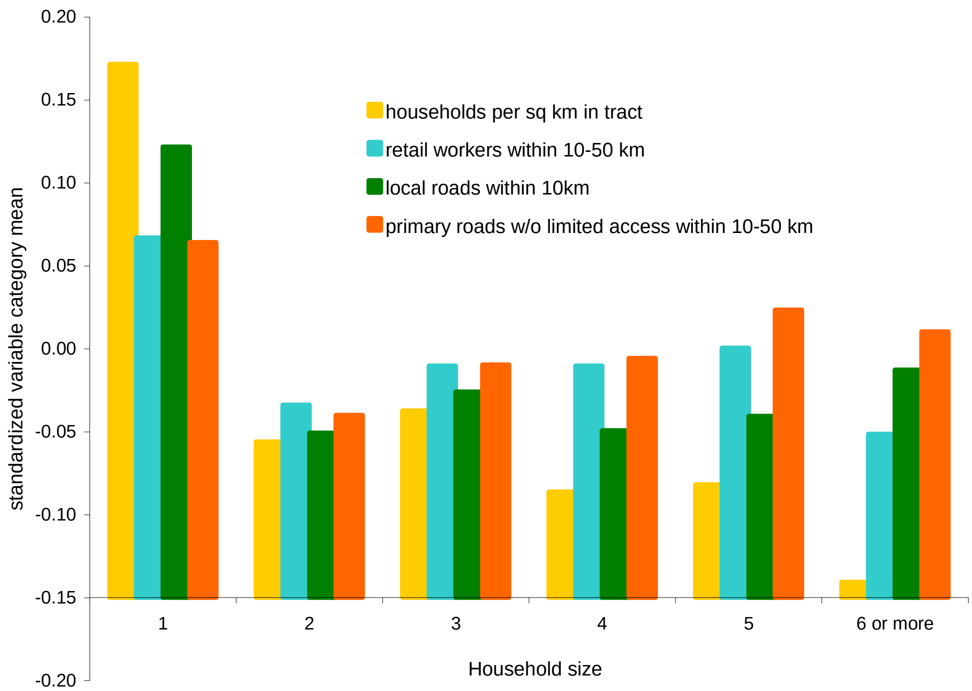


Figure 12 Some Spatial Variable Means by Categories of Household Size

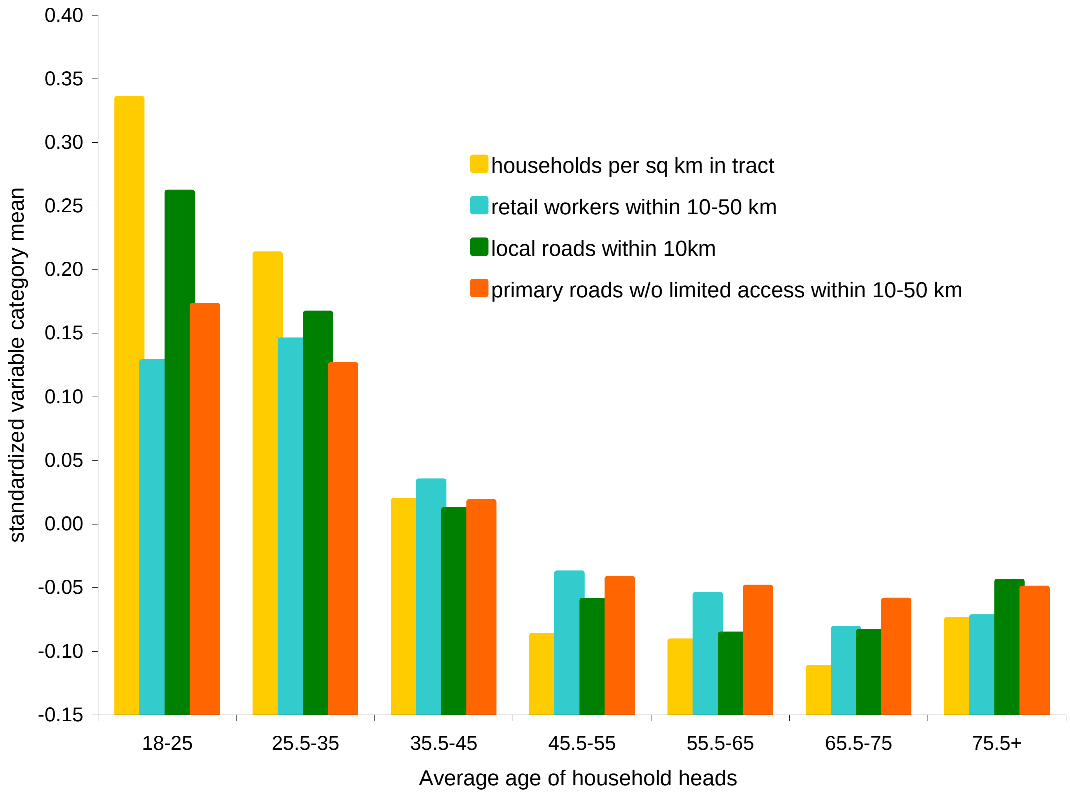


Figure 13 Some Spatial Variable Means by Average Age of Household Heads

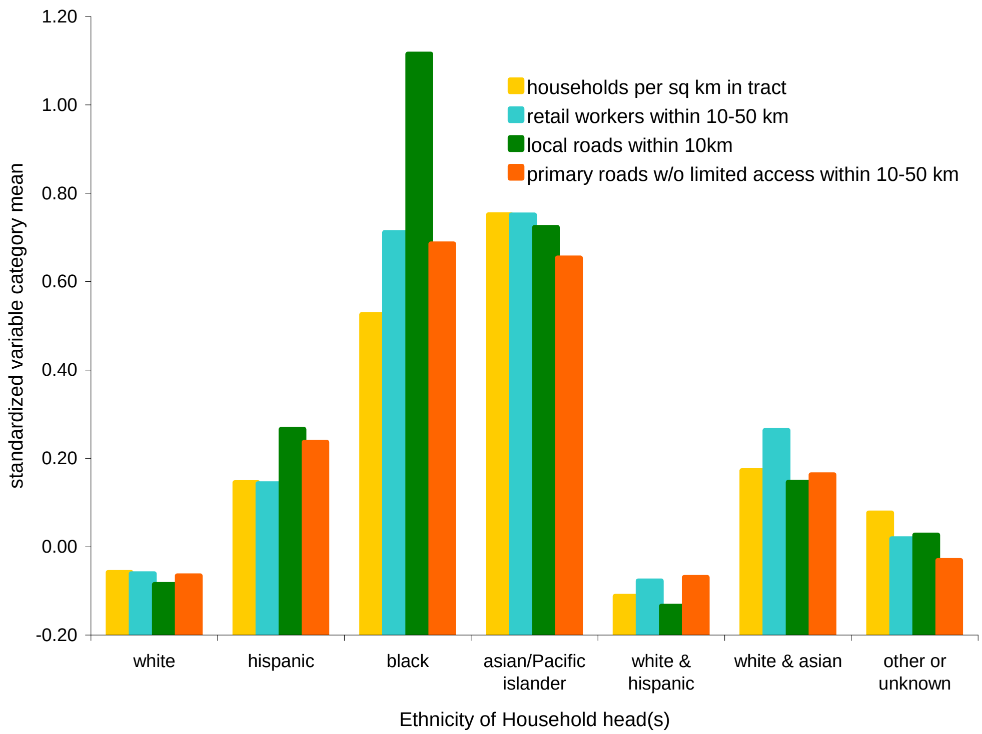


Figure 14 Spatial Variable Means by Ethnicity of Household Heads

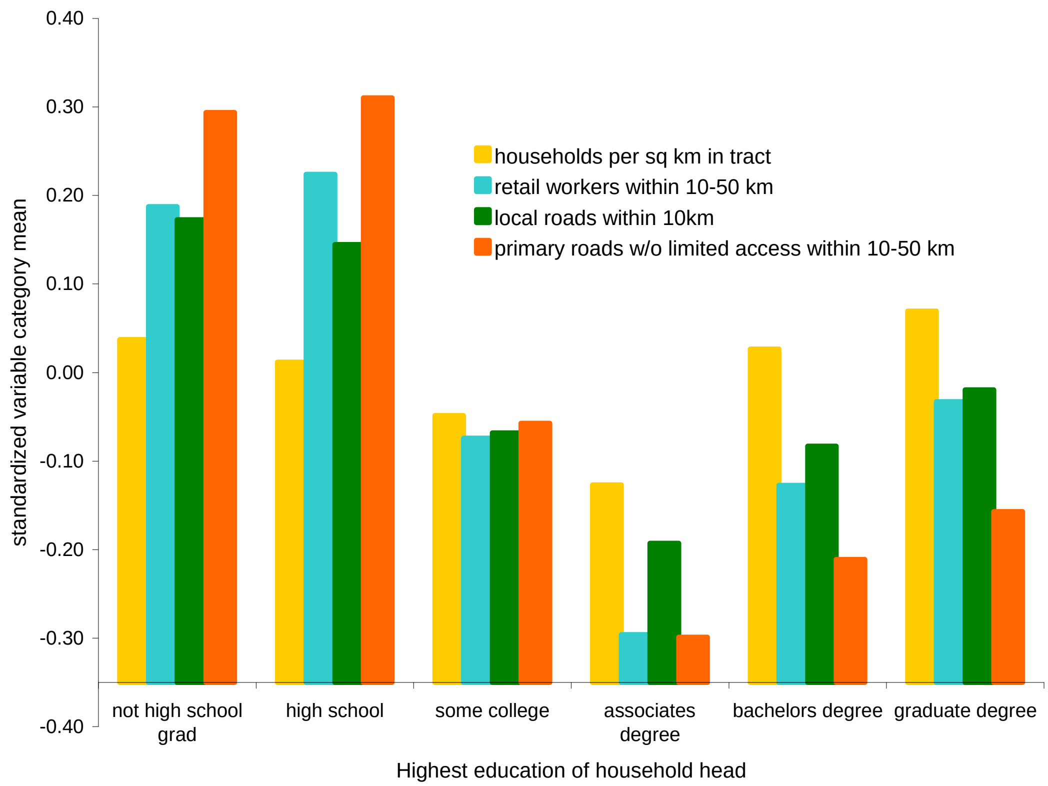


Figure 15 Spatial Variable Means by Education of Household Heads

The income dimension of residential location of households is most strongly related to regional retail accessibility (Figure 11). With the exception of the lowest income groups, higher income households tend to be located in areas surrounded by the highest retail activity. Higher income households (households with incomes of $100,000 or more in 2000 dollars) are also located in regions with the highest levels of primary road infrastructure. With respect to local road infrastructure, households in the highest and lowest income classes tend to be located in areas with the greatest density of local roads; there is no statistically significant difference (p = .01) between local road densities among the six middle income classes. Finally the only statistically significant relationship between income and housing density is that the lowest income households reside in denser census tracts; otherwise income is not a factor in housing density. With regard to household size, all spatial effects involve single-person households as distinguished from multi-person households. For each of the four spatial variables graphed in

Figure 12, there is no statistically significant difference among multi-person households of different sizes. As opposed to income, household density at the tract level is most strongly related to household size. The patterns of residential location as a function of age of the household head(s) revolves around decreases in density by age groups up until the 45.5-55 category, after which there are no statistically significant effects (Figure 13). The strongest relationship is that between age and housing density, followed by a moderately strong relationship between age and the density of local road infrastructure. Black and Asian households tend to locate in the highest density areas, in terms of all four spatial measures (Figure 14). In terms of one of these variables, local road infrastructure, Black households reside in areas that are even denser than Asian households. Excluding Black and Asian households, there are still statistically significant differences between the other ethnic groups. Hispanic and mixed White and Asian households live in areas that are denser than those resided by White and mixed White and Hispanic households. The residential location patterns of households according to education of household head, which is also a proxy for occupation, are shown in Figure 15. The strongest relationship is for regional primary road infrastructure. Less educated households tend to reside in areas with the greatest regional coverage of surface arterial primary roads. Households with associated degrees reside in areas with the lowest coverage of regional coverage of surface arterials, and the same low density for those with associate degrees is true for the other dimensions, especially regional retail accessibility. There is a similar, but less pronounced pattern for local roads. Finally, households in the highest education segments reside in areas with high residential density, compared with households in the middle education segments. The presence of young children does not appear to be a major factor in residential location, as there are no significant relationships involving between the indicator variable of young children and any of our four key spatial variables. Households with children aged six through twelve tend to be located in areas with lower housing density. There are no statistical relationships with retail accessibility or local or regional primary arterial road coverage. Households with children aged six through twelve tend to be located in areas with lower housing density, with lower regional retail accessibility, and in areas with lower coverage of roads. This discussion points to the need for further analysis that accounts for these sociodemographics and goes at least one step further in the spatial unit within which these households reside to study the impact of all this on travel. We do the remaining analysis with regression models that account for multiple influences on travel behavior.

6. Microanalysis Using Regression Models

In each of the models that follow, three blocks of variables are tested: (1) the same set of sociodemographic variables, (2) residential and activity site density variables, and (3) any road infrastructure variables found to be significant in explaining the dependent travel behavior variable after controlling for the first two sets of variables. Spatial variables were derived using buffer areas (e.g., around the population centroid of a census tract such as retail employees within 10 km of a census tract). Several such measures were developed, using both time and distance to define the boundary of the buffer. Based on preliminary data exploration only the 10 km and 50 km buffer variables were found to have a substantial effect and are used here. Some shorter time buffers could have been used and would have produced similar results, but the 10km and 50km distances were found to be more effective in capturing the influence of infrastructure provision and access to activity opportunities. The shortest distance buffer zone indicators are tested both in direct and difference (ring) format.

Modeling the contribution of spatial accessibility and infrastructure density was further complicated by the presence of spikes at zero and long positive tails. For example, some rural census tracts in California are extremely large with a very small population concentrated in a small portion of the tract. These need to be modeled together with census tracts that have some of the highest densities of roadway infrastructure in the nation. To overcome this distributional heterogeneity, spatial variables were converted to a scale in which the population was ranked into ten groups of equal frequency (deciles). This relieves the estimation bias caused by outlying observations and restrictions to the positive domain with spikes at zero value. It also facilitates estimation in which the spatial variables can contribute nonlinear and even non-ordinal effects.

We present omnibus tests of each set of variables, but the variable coefficients are shown only for the final complete model. These coefficients are displayed as odds ratios; the raw coefficient can be computed as the natural logarithm of the odds ratio. To aid in interpretation, only statistically significant ($p = .05$) coefficients are listed. All variables are categorical, and the continuous spatial variables are discretized into ten equal categories (deciles).

In the following sections we present the results of six sets of models aimed at assessing the influence of the spatial environment on travel demand in California. The first model identifies which households contain adults (persons eighteen and older) who are non-drivers, as this special group is an important component of passenger, transit, and nonmotorized demand. The second set of models deals with public transport (transit) demand, and we estimate separate models for transit demand by any household member and by adult drivers only. Similar sets of models are then estimated for nonmotorized travel and for high occupancy vehicle (HOV) travel. The latter set also contains a model of the HOV travel time. The final set of models is for solo driving, with one model for household solo driving demand, and one model for solo driving distance.

We also analyze the impact of spatial aggregation (e.g., tract level measurement versus block group measurement) level on power of the models. The same procedure of variable computation was conducted using block groups, which are smaller than census tracts, and the same six sets of models were built using the block group variables. The potentially deleterious impact may arise from the modifiable areal unit problem (MAUP; Openshaw and Alvanides, 1999). MAUP is one of the important issues that should be considered when we use GIS. Artificial boundaries imposed on continuous geographical phenomena, such as accessibility, results in the generation of artificial spatial patterns, and the spatial patterns generated in different levels of spatial aggregation differ from each other. We analyze the existence and the impact of MAUP in the six sets of travel behavior models and show how spatial variables at different aggregation levels can be used in the models to mitigate this artificial spatial resolution considering the impact of unit area sizes. In the models that follow we show estimation results using census tract accessibility variables and sociodemographics and estimation results using combinations of block group level variables.

6.1 Adults Who Do Not Drive

A substantial portion of household travel behavior that does not involve driving is due to the constrained choices of non-driving adults. California has a reputation for being an automobile-oriented state. While this reputation may be somewhat unfair, it is indeed the case that most of the state was developed over the past 50 years, and so it is designed and developed with the automobile in mind. This contrasts sharply with older east coast or European cities. One might presume therefore that households with non-driving adults might choose to locate in denser areas, where walking, bicycling and public transport are well supported. We test this hypothesis in the model presented in this section. This model uses 16,949 observations, or 99.5% of the sample with complete data. In this sample, 11.6% of households have non-driving adults.

6.1.1 Census Tract Model

The contributions of the three variable sets in explaining which households have non-driving adults are captured in the omnibus statistical tests listed in Table 7. All eight of the sociodemographic variables were important, but only one spatial variable, housing density, was found to be significant in describing the residential location of these households, controlling for their socioeconomic characteristics. Road infrastructure was not significantly different than zero.

Table 7 Binary Logit Model of Presence in Household of Non-driving Adult

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	2678.10	35	2678.10	35	.285
Spatial density	51.60	9	2729.70	44	.290
Road infrastructure	(not significant)				

The statistically significant influences of the sociodemographic variables are listed in Table 8. Income and household size display monotonic effects, and age highlights the expected elderly outcome. Non-driving adults are more likely to be found in Hispanic and Black households, and in households in the lowest education groups. Finally, households with children present are less likely to have non-driving adults, regardless of the ages of the children.

Table 8 Logit Model of Presence of Non-driving Adult – Sociodemographic

Independent variable	Significance	Odds ratio
Income (base = unknown)	0.00	
<$10,000	0.00	5.910
$10,000-$24,999	0.00	3.093
$25,000-$34,999	0.00	1.510
$35,000-$49,999		
$50,000-$74,999	0.00	0.653
$75,000-$99,999	0.00	0.491
$100,000-$149,999	0.00	0.366
$150,000+	0.00	0.250
household size (base = 6 or more)	0.00	
1	0.00	0.073
2	0.00	0.217
3		
4	0.00	1.706
5	0.00	4.200
Average age of heads (base = unknown)	0.00	
18-25	0.00	0.702
25.5-35	0.00	0.691
35.5-45	0.00	0.625
45.5-55		
55.5-65		
65.5-75	0.03	1.189
75.5+	0.00	2.998
Ethnicity (base = unknown)	0.00	
White	0.00	0.620
Hispanic	0.00	1.723
Black	0.00	1.482
Asian/Pacific Islander	0.04	0.694
White & Hispanic	0.02	0.726
White & Asian		
Education (base = unknown)	0.00	
not high school graduate	0.00	1.862
high school graduate	0.00	1.269
Some college		
associates degree	0.01	0.766
bachelors degree	0.00	0.713
graduate degree	0.00	0.687
presence of children 0-5 yrs. Old	0.00	0.262
presence of children 6-12 yrs. Old	0.00	0.326
presence of children 13-17 yrs. Old	0.00	0.369

As shown in Table 9, households with non-driving adults are less likely to be located in low density residential areas (e.g, the lowest quartile of residential density), and more likely to be located in the very highest density areas. There is no statistically significant relationship between accessibility or road infrastructure and the likelihood of the presence of non-driving adults. In other words, households with non-driving adults are most likely not choosing where they live to accommodate the travel behavior of their non-driving members. Consequently, in the travel behavior models that follow, the accessibility and infrastructure effects are not attributable to the contribution to travel behavior of non-driving adults.

Table 9 Logit Model of Presence of Non-driving Adult – Spatial Density (Tract)

Ind. variable (all bases = 50th %tile)	Significance	Odds ratio
tract household density	0.00	
<10 %tile	0.03	0.828
10th %tile	0.00	0.752
20th %tile	(0.06)	(0.849)
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile	0.00	1.608

6.1.2 Comparison with Block Group Model

In Table 10, the contributions of sociodemographics, spatial variables measured at the tract level, and spatial variables measured at the block group level are compared in terms of their contribution to goodness of fit. The impact of the sociodemographic variables on the presence of non-driving adult in households is almost identical in the census tract model and the block group model of non-driving adults. Because the measurement of retail employee within a certain travel distance involves shortest distance, and the calculation of household density does not, using smaller spatial unit has different model implications. When we use a smaller spatial unit, it means that we consider a smaller "neighborhood(s)" around home locations for household density computation, but it means closer approximation for measurements involving shortest travel distance. To see the influence of using smaller spatial units on each variable set, the contribution of household density and retail employee density are given separately in Table 10. Only household density has a significant impact on non-driving adults, and it contributes more to the model when it is measured using the larger spatial unit areas, census tracts in this case. Additional estimation details are also offered by Table 11 for spatial density.

Table 10 Binomial Logit Models of Presence in Household of Non-driving Adult

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Census tract	**Sociodemographic**	**2678.10**	**35**	**2678.10**	**35**	**.285**
	Spatial density	**51.60**	**9**	**2729.70**	**44**	**.290**
	Household density	*51.60*	*9*			
	Retail employee	-	-			
	Road infrastructure	(not significant)				
Block group	**Sociodemographic**	**2678.69**	**35**	**2678.69**	**35**	**.285**
	Spatial density	**37.54**	**9**	**2716.22**	**44**	**.289**
	Household density	*37.54*	*9*			
	Retail employee	-	-			
	Road infrastructure	(not significant)				

Table 11 Logit Models of Presence of Non-driving Adult – Spatial Density

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
household density	0.00		0.00	
<10 %tile	0.03	0.828	0.03	0.835
10th %tile	0.00	0.752	(0.06)	(0.850)
20th %tile	(0.06)	(0.849)	0.02	0.816
30th %tile				
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile	0.00	1.608	0.00	1.435

6.2 Transit Usage by Any Household Member

Transit usage is defined as taking any local transit mode, including bus, rail, and light rail, but not including long distance bus trips. School bus trips are also included as household public transport trips. Of the 16,750 households with complete data (98.3% of the sample), 8.1% had a household member who made at least one trip by public transport (transit); the highest concentration of these households being in the San Francisco Bay Area, where 14.4% of households in this sample were transit users.

6.2.1 Census Tract Model

Compared to the previous model for households with non-driving adults, socioeconomic factors are less effective in explaining which households are transit users, but there are three significant spatial factors, and one road infrastructure variable is important as shown in Table 12.

Table 12 Logit Model of Any Household Transit Use and Spatial Density at Tract Level

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	1633.28	35	1633.28	35	.216
Spatial density	177.90	27	1811.19	62	.238
Road infrastructure	81.57	9	1892.76	71	.248

The estimated effects of the sociodemographic variables are reported on Table 13. Transit usage is a decreasing function of income (where statistically insignificant categories are shown to complete the picture), and an increasing function of household size. Transit usage is generally a decreasing function of age of the household head(s), but usage is greatest for the second youngest group, and lowest for the second oldest group. Transit services for the elderly probably increase the likelihood of transit usage for households with the oldest household heads. Education is not an effective predictor of transit usage, and only one ethnicity category is important: black households are 1.6 times more likely to use transit. Regarding children, households with only young children are less likely to use transit, while those with older children are more likely to use transit.

Table 13 Logit Model of Any Household Transit Use – Sociodemographic Variables

Independent variable	Significance	Odds ratio
Income (base = unknown)	0.00	
<$10,000	0.00	2.175
$10,000-$24,999	0.00	1.381
$25,000-$34,999	0.03	1.198
$35,000-$49,999		
$50,000-$74,999	0.00	0.810
$75,000-$99,999	(0.29)	(0.909)
$100,000-$149,999	(0.26)	(0.886)
$150,000+	0.00	0.379
household size (base = 6 or more)	0.00	
1	0.00	0.376
2	0.00	0.491
3		
4	0.00	1.416
5	0.00	1.737
Average age of heads (base = unknown)	0.00	
18-25	0.05	1.267
25.5-35	0.00	1.488
35.5-45	0.00	1.255
45.5-55		
55.5-65		
65.5-75	0.00	0.554
75.5+	0.03	0.691
Ethnicity (base = unknown)	0.09	
White		
Hispanic		
Black	0.00	1.618
Asian/Pacific Islander		
White & Hispanic		
White & Asian		
Education (base = unknown)	0.41	
not high school graduate		
high school graduate		
some college		
associates degree		
bachelors degree		
graduate degree		
presence of children 0-5 yrs. old	0.01	0.775
presence of children 6-12 yrs. Old	0.00	2.363
presence of children 13-17 yrs. old	0.00	3.001

Spatially, as expected, transit-using households are concentrated in the densest 10% of residential areas, and also in the least dense 20% of areas, as shown in Table 14. But excluding areas in the highest 10% of housing density, households located in areas above median density are less likely to use transit. Census tracts with low density housing tend to be located in rural counties. While the presence of school age children in the household coupled with the inclusion of school bus trips as public transit trips may account for some of this effect, this result underscores the importance of rural public transport.

Table 14 Logit Model of Any Household Transit Use – Spatial Density

Ind. Variable (all bases = 50th %tile)	Significance	Odds ratio
tract household density	0.00	
<10 %tile	(0.20)	(1.215)
10th %tile	0.04	1.262
20th %tile		
30th %tile		
40th %tile		
60th %tile	0.01	0.758
70th %tile	(0.43)	(0.919)
80th %tile	0.00	0.725
90th %tile	0.00	1.677
retail employees within 10 km	0.00	
<10 %tile		
10th %tile		
20th %tile	0.02	0.755
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile	0.00	1.795
retail employees within 10 to 50km	0.00	
<10 %tile	(0.19)	(0.764)
10th %tile	0.00	0.594
20th %tile	0.01	0.664
30th %tile		
40th %tile		
60th %tile	0.02	1.381
70th %tile		
80th %tile		
90th %tile	0.00	2.140

Accessibility to retail services, particularly accessibility at the regional level (10 to 50 km), indicates lower transit usage for households located in low accessibility areas, and high transit usage for households located in the highest 10% of retail accessibility. This effect undoubtedly captures the urban core phenomenon. The influence of road infrastructure is complex, as shown in Table 15. Controlling for sociodemographic factors and spatial density, households that live in areas in the lower quartile of regional primary surface road coverage (primary roads without limited access within 10 to 50 km of network distance) exhibit the highest transit usage, together with households in the 80th percentile. However, households above the 90th percentile have very low transit usage. Once again, the importance of rural public transport is picked up by the road infrastructure variable, even when controlling for housing and retail density. In tracts with both low housing density and lower levels of road infrastructure, the likelihood of transit usage is unusually high.

Table 15 Logit Model of Any Household Transit Use – Road Infrastructure

Variable (Bases = 50th %tile)	Significance	Odds ratio
primary roads w/o limited access within 10 to 50 km	0.00	
<10 %tile	0.02	1.678
10th %tile	0.02	1.466
20th %tile	0.04	1.401
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile	0.01	1.483
90th %tile	0.00	0.364

6.2.2 Comparison with Block Group Model

As shown in Table 16, household density contributes slightly more to the model when it is measured at the census tract level, and the other spatial variable sets – retail employee density and road infrastructure - contribute more to the model when they are measured based on block groups. Especially, the road infrastructure in the block group model contributed almost twice as much as in the census tract model in terms of chi-square.

Table 16 Logit Models of Any Household Transit Use

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Census tract	**Sociodemographic**	**1633.28**	**35**	**1633.28**	**35**	**.216**
	Spatial density	**177.90**	**27**	**1811.19**	**62**	**.238**
	Household density	*125.45*	*9*			
	Retail employee	*52.45*	*18*			
	Road infrastructure	**81.57**	**9**	**1892.76**	**71**	**.248**
Block group	**Sociodemographic**	**1633.58**	**35**	**1633.58**	**35**	**.216**
	Spatial density	**180.37**	**27**	**1813.95**	**62**	**.238**
	Household density	*106.50*	*9*			
	Retail employee	*73.87*	*18*			
	Road infrastructure	**158.66**	**9**	**1972.60**	**71**	**.258**

The spatial density variables show similar impact patterns on household transit usage in the block group analysis, too. However, in the block group model, the concentration of transit usage in the highest density area is stronger and the concentration in 10th percentile of household density is not captured. The highest percentile of the block group retail employee density had higher impact in both buffers (0 to 10 km and 10 to 50 km). This can be a typical influence of MAUP. First, different sizes of unit area produce different statistics, household density in this case, and they reveal different patterns of influences. The patterns can have different impact in the models as the variable sets do in the Logit model of household transit use (Table 17). Second, different levels of spatial aggregation lead to different levels of approximation of the explanatory variables. From the comparison between the two models of household transit use, it appears that a better approximation of an explanatory variable by going one level of disaggregation down (from tract to block group) improves the contribution of the independent variables by explaining variation in the dependent variable.

The influence pattern of road infrastructure of the block group model is similar to that of the census tract model, but in addition to primary roads without limited access within 10 to 50km, which was the only road infrastructure variable set significant in the census tract model of household transit usage, local roads variables were found to be significant in the block group model (Table 18). In the block group model, the importance of rural public transportation is also picked up, and the likelihood of transit usage is low in the households which belong to the highest 10% road network areas.

Table 17 Logit Models of Any Household Transit Use – Spatial Density

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
household density	0.00		0.00	
<10 %tile	(0.20)	(1.215)		
10th %tile	0.04	1.262	(0.16)	(1.157)
20th %tile				
30th %tile				
40th %tile				
60th %tile	0.01	0.758		
70th %tile	(0.43)	(0.919)	0.01	0.756
80th %tile	0.00	0.725	0.05	0.810
90th %tile	0.00	1.677	0.00	1.627
retail employees within 10 km	0.00		0.00	
<10 %tile				
10th %tile				
20th %tile	0.02	0.755	(0.08)	(0.806)
30th %tile			(0.09)	(0.821)
40th %tile				
60th %tile			0.02	0.752
70th %tile				
80th %tile				
90th %tile	0.00	1.795	0.00	2.218
retail employees within 10 to 50km	0.00		0.00	
<10 %tile	(0.19)	(0.764)		
10th %tile	0.00	0.594	0.01	0.700
20th %tile	0.01	0.664	0.00	0.668
30th %tile				
40th %tile				
60th %tile	0.02	1.381		
70th %tile				
80th %tile			0.02	1.389
90th %tile	0.00	2.140	0.00	3.294

Table 18 Logit Models of Any Household Transit Use – Road Infrastructure

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
primary roads w/o limited access within 10 to 50 km	0.00		0.00	
<10 %tile	0.02	1.678	0.05	1.249
10th %tile	0.02	1.466		
20th %tile	0.04	1.401		
30th %tile				
40th %tile				
60th %tile			(0.09)	(1.180)
70th %tile			0.00	1.314
80th %tile	0.01	1.483		
90th %tile	0.00	0.364	0.00	0.505
Local roads within 10 km			0.03	
<10 %tile				
10th %tile				
20th %tile				
30th %tile				
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile			0.02	0.681
Local roads within 10 to 50 km			0.00	
<10 %tile			0.03	1.389
10th %tile				
20th %tile				
30th %tile				
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile			0.00	0.400

6.3 Transit Usage by an Adult Driver in the Household

Analyzing transit trips made by *any* household member can be difficult to interpret, as children and non-driving adults may be skewing the results for some households but not others. The next model describes transit usage by adult drivers, being those adults who were either recorded as having a driver's license, or else were observed to have driven at least once. Only 2.7% of the 14,160 households with adult drivers and complete data have an adult driver that makes at least one transit trip.

6.3.1 Census Tract Model

As expected, it is much more difficult to predict which households these are, based on sociodemographic factors, as seen by comparing the goodness-of-fit log-likelihood-ratio model Chi-square statistics and the pseudo-R^2 indices in Table 19. However, spatial density is relatively more important in the case of adult drivers, and the same road infrastructure variable is also significant.

Table 19 Logit Model of Household Transit Use by Adult Drivers

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	216.32	35	216.32	35	.068
Spatial density	282.90	27	499.22	62	.155
Road infrastructure	64.76	9	563.98	71	.175

The sociodemographic predictors of transit usage by adult drivers are shown in Table 20. Such usage is concentrated in low income households, larger households, households in the middle age groups (35 to 55), black households, and more highly educated households. This latter effect probably captures central business district employment. Households less likely to have adult driver transit usage are high and middle income households, small households, households with heads in the 65-75 year range, lower educated households, and households with children.

Tables 21 and 22 show that the effects of rural public transport (tracts with low density housing and road infrastructure) disappear when the focus is restricted to adult drivers. Still, controlling for sociodemographic factors, households that live in areas with the highest residential and retail density are the heaviest transit users. The phenomenon of low relative transit usage households in the 90th percentile of regional primary surface road coverage still prevails, as seen in Table 22. Households in the 90th percentile of regional primary arterial coverage are concentrated in Orange, Los Angeles, and San Mateo, and Alameda Counties, but there are also such households located in San Bernardino, Santa Clara, Riverside and Ventura Counties. An abundance of primary arterials appears to correlate with fewer household transit trips in these areas.

6.3.2 Comparison with Block Group Model

As shown in Table 23, we can see the influence of using smaller unit areas in this comparison, too. Household density contributes more to the model when it is measured at the census tract level, and the other spatial variables contribute more to the model when they are measured at the block group level. Table 24 shows the likelihood of transit usage by adult drivers is relatively low among households in the 90th percentile of primary and local roads coverage as shown in the census tract model. However, the local roads variable set in the block group model still show the effect of rural public transport usage by adults drivers and the 70th percentile of primary road infrastructure had positive impact in the block group model, which couldn't be seen in the census tract model. Table 25 shows the likelihood of transit usage by adult drivers was found to be the highest in the households in the 90th percentile of spatial density as it was in the census tract model. High transit usage in the 40th percentile of household density was marginally significant, which was not found in the census tract model. The impact of the highest deciles of retail employee density was higher and also clearer in the block group model.

Table 20 Logit Model of Household Transit Use by Adult Drivers – Sociodemographic

Independent variable	Significance	Odds ratio
Income (base = unknown)	0.00	
<$10,000	0.00	3.144
$10,000-$24,999		
$25,000-$34,999		
$35,000-$49,999		
$50,000-$74,999	0.00	0.614
$75,000-$99,999		
$100,000-$149,999		
$150,000+	0.00	0.454
household size (base = 6 or more)	0.00	
1	0.00	0.413
2	0.00	0.446
3	0.00	0.640
4	0.02	1.379
5	0.00	1.879
Average age of heads (base = unknown)	0.01	
18-25		
25.5-35		
35.5-45	0.01	1.418
45.5-55	0.00	1.419
55.5-65		
65.5-75	0.00	0.487
75.5+		
Ethnicity (base = unknown)	0.03	
White		
Hispanic		
Black	0.01	1.749
Asian/Pacific Islander		
White & Hispanic		
White & Asian		
Education (base = unknown)	0.00	
not high school graduate	0.03	0.610
high school graduate		
some college		
associates degree		
bachelors degree	0.01	1.465
graduate degree	0.00	1.867
presence of children 0-5 yrs. old	0.00	0.382
presence of children 6-12 yrs. Old	0.00	0.425
presence of children 13-17 yrs. old	0.01	0.589

Table 21 Logit Model of Household Transit Use by Adult Drivers – Spatial Density

Ind. variable (all bases = 50th %tile)	Significance	Odds ratio
tract household density	0.00	
<10 %tile		
10th %tile		
20th %tile		
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile	0.00	2.335
retail employees within 10 km	0.00	
<10 %tile		
10th %tile		
20th %tile		
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile	0.02	1.622
90th %tile	0.00	2.148
retail employees within 10 to 50 km	0.06	
<10 %tile		
10th %tile		
20th %tile	0.03	0.499
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile	0.00	2.644

Table 22 Logit Model of Household Transit Use by Adult Drivers – Infrastructure

Variable (base = 50th %tile)	Significance	Odds ratio
primary roads w/o limited access within 10 to 50 km	0.00	
<10 %tile		
10th %tile		
20th %tile		
30th %tile		
40th %tile		
60th %tile	0.02	0.522
70th %tile		
80th %tile		
90th %tile	0.01	0.418

Table 23 Logit Models of Household Transit Use by Adult Drivers

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Census	**Sociodemographic**	**216.32**	**35**	**216.32**	**35**	**.068**
Tract	**Spatial density**	**282.90**	**27**	**499.22**	**62**	**.155**
	Household density	*205.39*	*9*			
	Retail employee	*77.51*	*18*			
	Road infrastructure	**64.76**	**9**	**563.98**	**71**	**.175**
Block	**Sociodemographic**	**216.34**	**35**	**216.34**	**35**	**.068**
Group	**Spatial density**	**297.52**	**27**	**513.86**	**62**	**.159**
	Household density	*180.12*	*9*			
	Retail employee	*117.40*	*18*			
	Road infrastructure	**116.93**	**18**	**630.76**	**80**	**.195**

Table 24 Logit Models of Household Transit Use by Adult Drivers – Infrastructure

Ind. Variable (all bases = 50^{th} %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
primary roads w/o limited access within 10 to 50 km	0.00		0.00	
<10 %tile				
10th %tile			(0.07)	(0.701)
20th %tile				
30th %tile				
40th %tile				
60th %tile	0.02	0.522	(0.10)	(1.317)
70th %tile			0.00	1.905
80th %tile				
90th %tile	0.01	0.418	0.00	0.367
Local roads within 10 to 50 km			0.01	
<10 %tile			0.01	2.272
10th %tile				
20th %tile				
30th %tile				
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile			0.00	0.295

Table 25 Logit Models of Household Transit Use by Adult Drivers – Spatial Density

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
household density	0.00		0.00	
<10 %tile				
10th %tile				
20th %tile			0.02	0.514
30th %tile				
40th %tile			0.05	1.397
60th %tile				
70th %tile				
80th %tile				
90th %tile	0.00	2.335	0.00	1.981
retail employees within 10 km	0.00		0.00	
<10 %tile				
10th %tile				
20th %tile			0.01	0.443
30th %tile			0.02	0.533
40th %tile				
60th %tile				
70th %tile				
80th %tile	0.02	1.622	(0.08)	(1.395)
90th %tile	0.00	2.148	0.00	3.061
retail employees within 10 to 50km	0.06		0.00	
<10 %tile				
10th %tile				
20th %tile	0.03	0.499	(0.09)	(0.649)
30th %tile			0.04	0.593
40th %tile				
60th %tile			(0.07)	(0.641)
70th %tile				
80th %tile			0.00	2.400
90th %tile	0.00	2.644	0.00	5.484

6.4 Nonmotorized Travel by Any Household Member

Of our 16,750 households with complete data (98.3% of the sample), 14.2% had a household member that made at least one trip walking or by bicycle. As in the case of transit, the highest concentration of these households was in the San Francisco Bay Area, where 25.9% of the households in this survey recorded a nonmotorized trip segment, followed by Santa Barbara County, with 19.2% of households.

6.4.1 Census Tract Model

Compared to transit-using households, it is more difficult to explain households that generate nonmotorized travel (Table 26). However, spatial factors are relatively more important in nonmotorized travel demand.

Table 26 Logit Model of Any Household Nonmotorized Travel

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	1065.65	35	1065.65	35	.116
Spatial density	373.08	27	1438.73	62	.147
Road infrastructure	104.31	18	1543.04	80	.158

The sociodemographic predictors of household nonmotorized travel are listed in Table 27. As expected, the presence of children older than 6 increases the likelihood of a household making a nonmotorized trip, while the presence of very young children decreases that likelihood. Lower income and the youngest households are more likely to make nonmotorized trips, but so are the most highly educated households. With regard to influences of the built environment on nonmotorized travel (Tables 28 and 29), the "rural" effect is somewhat different for nonmotorized trips than for public transport trips. Here low housing density produces a lower propensity for nonmotorized trips, confirming that extreme distances among activities inhibit the use of slower modes. It is possible that for some households rural transit trips is taking the place of rural nonmotorized trips. In terms of road infrastructure, Table 29 shows that the lower percentiles have much higher propensity for nonmotorized trips, as is the case for transit. Higher levels of road infrastructure correspond to lower levels of nonmotorized trips. Both of these effects are perhaps related to using nonmotorized trips as a form of recreation, as it is pleasant to walk or bike in less developed, low traffic areas, while it is both unpleasant and dangerous to walk or bike in highly developed, high traffic areas.

6.4.2 Comparison with Block Group Model

The contribution of household density is larger in the census tract model, and the contributions of the other variable sets are larger in the block group model for household nonmotorized travel, too (Table 30 and Table 31). As shown in Table 32, the block group models also show that low household and retail employee density produces a lower propensity for nonmotorized trips, but the impact of retail employee density of 10 to 50 km distance is clearer in the block group models. The influence of primary roads without limited access within 10 to 50 km, which was not significant in the census tract model, is found to be significant in the block group model. Instead of local roads within 10 km which were significant in the census tract model, local roads within 10 to 50 km were found to be more significant in the block group model (Table 31).

Table 27 Logit Model of Household Nonmotorized Travel – Sociodemographic

Independent variable	Significance	Odds ratio
Income (base = unknown)	0.00	
<$10,000	0.00	1.623
$10,000-$24,999	0.00	1.305
$25,000-$34,999		
$35,000-$49,999		
$50,000-$74,999		
$75,000-$99,999		
$100,000-$149,999	0.01	0.803
$150,000+		
household size (base = 6 or more)	0.00	
1	0.00	0.542
2	0.00	0.650
3		
4		
5	0.00	1.557
Average age of heads (base = unknown)	0.00	
18-25	0.00	1.294
25.5-35		
35.5-45		
45.5-55		
55.5-65		
65.5-75	0.00	0.753
75.5+	0.03	0.806
Ethnicity (base = unknown)	0.29	
White		
Hispanic		
Black		
Asian/Pacific Islander		
White & Hispanic		
White & Asian		
Education (base = unknown)	0.00	
not high school graduate		
high school graduate	0.00	0.829
some college		
associates degree		
bachelors degree	0.00	1.239
graduate degree	0.00	1.670
presence of children 0-5 yrs. old	0.00	0.765
presence of children 6-12 yrs. Old	0.00	1.804
presence of children 13-17 yrs. old	0.00	2.234

Table 28 Logit Model of Household Nonmotorized Travel – Spatial Density

Ind. variable (all bases = 50th %tile)	Significance	Odds ratio
tract household density	0.00	
<10 %tile	0.00	0.564
10th %tile	0.00	0.494
20th %tile	0.00	0.620
30th %tile		
40th %tile		
60th %tile		
70th %tile	0.00	1.292
80th %tile	0.00	1.553
90th %tile	0.00	2.264
retail employees within 10 km	0.02	
<10 %tile	0.01	0.640
10th %tile	0.00	0.635
20th %tile		
30th %tile		
40th %tile		
60th %tile	0.00	1.423
70th %tile		
80th %tile		
90th %tile		
retail employees within 10 to 50 km	0.00	
<10 %tile		
10th %tile		
20th %tile		
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile	0.01	1.931

Table 29 Logit Model of Household Nonmotorized Travel – Infrastructure

Variable (base = 50th %tile)	Significance	Odds ratio
local roads within 10 km	0.00	
<10 %tile	0.00	1.859
10th %tile	0.00	1.992
20th %tile	0.01	1.393
30th %tile		
40th %tile		
60th %tile	0.00	0.686
70th %tile	0.00	0.610
80th %tile	0.00	0.619
90th %tile		
local roads within 10 to 50 km	0.00	
<10 %tile	0.00	1.850
10th %tile		
20th %tile		
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile		

Table 30 Logit Models of Any Household Nonmotorized Travel

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Census Tract	**Sociodemographic**	**1065.65**	**35**	**1065.65**	**35**	**.116**
	Spatial density	**373.08**	**27**	**1438.73**	**62**	**.147**
	Household density	*277.78*	*9*			
	Retail employee	*95.30*	*18*			
	Road infrastructure	**104.31**	**18**	**1543.04**	**80**	**.158**
Block Group	**Sociodemographic**	**1065.94**	**35**	**1065.94**	**35**	**.110**
	Spatial density	**325.60**	**27**	**1391.54**	**62**	**.143**
	Household density	*191.22*	*9*			
	Retail employee	*134.38*	*18*			
	Road infrastructure	**156.27**	**27**	**1547.82**	**89**	**.158**

Table 31 Logit Models of Household Nonmotorized Travel – Infrastructure

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
primary roads w/o limited access				
within 10 to 50 km			0.00	
<10 %tile				
10th %tile			0.00	0.735
20th %tile			0.03	0.842
30th %tile				
40th %tile				
60th %tile				
70th %tile			0.03	1.164
80th %tile				
90th %tile				
Local roads within 10 km	0.00		(0.09)	
<10 %tile	0.00	1.859	0.00	1.355
10th %tile	0.00	1.992		
20th %tile	0.01	1.393		
30th %tile				
40th %tile				
60th %tile	0.00	0.686		
70th %tile	0.00	0.610		
80th %tile	0.00	0.619		
90th %tile				
Local roads within 10 to 50 km	0.00		0.00	
<10 %tile	0.00	1.850	0.00	2.231
10th %tile			0.00	1.635
20th %tile			0.01	1.321
30th %tile			0.00	1.367
40th %tile			(0.17)	(1.124)
60th %tile			(0.10)	(0.870)
70th %tile			0.00	0.698
80th %tile			0.00	0.662
90th %tile			0.00	0.312

Table 32 Logit Models of Household Nonmotorized Travel – Spatial Density

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
household density	0.00		0.00	
<10 %tile	0.00	0.564		
10th %tile	0.00	0.494	0.00	0.696
20th %tile	0.00	0.620	0.00	0.618
30th %tile				
40th %tile				
60th %tile				
70th %tile	0.00	1.292	0.02	1.182
80th %tile	0.00	1.553	0.03	1.172
90th %tile	0.00	2.264	0.00	1.474
retail employees within 10 km	0.02		0.00	
<10 %tile	0.01	0.640	0.00	0.563
10th %tile	0.00	0.635		
20th %tile				
30th %tile				
40th %tile				
60th %tile	0.00	1.423	(0.10)	(1.148)
70th %tile				
80th %tile				
90th %tile			0.00	1.996
retail employees within 10 to 50km	0.00		0.00	
<10 %tile			0.00	0.650
10th %tile			0.00	0.698
20th %tile			0.02	0.806
30th %tile			0.00	0.760
40th %tile			0.00	0.570
60th %tile				
70th %tile			(0.06)	(1.203)
80th %tile			0.00	1.961
90th %tile	0.01	1.931	0.00	3.243

6.5 Nonmotorized Travel - by an Adult Driver in the Household

Once again, adult drivers were used as a specialized subset to control for the various walking and biking behaviors of different kinds of household members. Of the 14,160 households with adult drivers, 10.4% had an adult driver that recorded a nonmotorized trip segment. For Bay Area households the split is 23.6% and it is 6.3% for Santa Barbara County.

6.5.1 Census Tract Model

As in the case of transit use by adult drivers, it is much more difficult to predict which households generate nonmotorized trips by adult drivers (Table 33), and spatial factors are more important than sociodemographic factors, as seen in the goodness-of-fit contributions of the three sets of variables (Table 33). Road infrastructure also plays a relatively important role. The sociodemographic predictors of nonmotorized travel demand by adult drivers in the household (Table 34) are similar to the predictors of nonmotorized travel demand by any household member, with the notable exception of the influence of older children. As expected, such children travel by bicycle and walking, but their presence actually decreases the likelihood that adult drivers in the household engage in such trips. Also, there is no statistically significant difference in the propensity for nonmotorized trips among adult drivers in older households, indicating that the lower propensity observed earlier for these households is likely due to the immobility of non drivers in such households.

Table 33 Logit Model of Household Nonmotorized Travel by Adult Drivers

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	251.86	35	251.86	35	.036
Spatial density	306.80	27	558.66	62	.079
Road infrastructure	158.58	27	717.24	71	.101

Table 34 Logit Model of Nonmotorized Travel by Adult Drivers – Sociodemographic

Independent variable	Significance	Odds ratio
Income (base = unknown)	0.00	
<$10,000	0.00	1.766
$10,000-$24,999	0.00	1.279
$25,000-$34,999		
$35,000-$49,999		
$50,000-$74,999		
$75,000-$99,999		
$100,000-$149,999	0.03	0.811
$150,000+		
household size (base = 6 or more)	0.00	
1	0.00	0.726
2	0.00	0.730
3	0.03	0.844
4		
5	0.01	1.397
Average age of heads (base = unknown)	0.31	
18-25	0.02	1.280
25.5-35		
35.5-45		
45.5-55		
55.5-65		
65.5-75		
75.5+		
Ethnicity (base = unknown)	0.00	
White		
Hispanic	0.02	0.776
Black		
Asian/Pacific Islander		
White & Hispanic		
White & Asian		
Education (base = unknown)	0.00	
not high school graduate		
high school graduate	0.00	0.729
some college		
associates degree		
bachelors degree	0.00	1.442
graduate degree	0.00	2.030
presence of children 0-5 yrs. old	0.02	0.767
presence of children 6-12 yrs. Old		
presence of children 13-17 yrs. old	0.00	0.689

The spatial density factors influencing nonmotorized travel by adult drivers (Table 35) indicate a similar pattern as for all nonmotorized travel, with a considerably enhanced positive effect of regional retail accessibility on nonmotorized travel demand by adult drivers. Likewise, the influences of road infrastructure are accentuated in the case of nonmotorized travel by adult drivers (Table 36). This shows that adult drivers are largely responsible for the effects of the built environment on demand for nonmotorized travel discussed in the previous section. It is not surprising that nonmotorized travel by non-drivers is less sensitive to spatial factors because this type of indicators are based on highways and highway speeds and not pedestrian and bicycle facilities.

Table 35 Logit Model of Nonmotorized Travel by Adult Drivers – Spatial Density

Ind. variable (all bases = 50th %tile)	Significance	Odds ratio
tract household density	0.00	
<10 %tile	0.01	0.635
10th %tile	0.00	0.511
20th %tile	0.00	0.676
30th %tile		
40th %tile		
60th %tile	0.04	1.219
70th %tile	0.01	1.274
80th %tile	0.00	1.386
90th %tile	0.00	2.177
retail employees within 10km	0.01	
<10 %tile	0.01	0.553
10th %tile	0.00	0.525
20th %tile		
30th %tile		
40th %tile		
60th %tile	0.01	1.395
70th %tile		
80th %tile		
90th %tile	0.01	1.780
retail employees within 10 to 50km	0.00	
<10 %tile	0.01	0.510
10th %tile	0.02	0.589
20th %tile		
30th %tile		
40th %tile	0.01	0.649
60th %tile	0.00	1.849
70th %tile		
80th %tile		
90th %tile	0.01	2.335

Table 36 Logit Model of Nonmotorized Travel by Adult Drivers – Infrastructure

Variable (base = 50th %tile)	Significance	Odds ratio
local roads within 10 km	0.00	
<10 %tile	0.00	2.484
10th %tile	0.00	2.207
20th %tile	0.01	1.560
30th %tile		
40th %tile		
60th %tile	0.01	0.704
70th %tile	0.00	0.623
80th %tile	0.00	0.547
90th %tile	(.151)	(0.734)
local roads within 10 to 50 km	0.00	
<10 %tile	0.00	2.684
10th %tile	0.02	1.738
20th %tile		
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile		
primary roads w/o limited access within 50 km	0.00	
<10 %tile		
10th %tile		
20th %tile		
30th %tile	0.00	1.689
40th %tile		
60th %tile	0.01	0.656
70th %tile	0.04	0.725
80th %tile		
90th %tile		

6.5.2 Comparison with Block Group Model

Less likelihood of nonmotorized travel by adult driver in higher levels of road infrastructure is also picked up in the block group models. But instead of primary roads without limited access within 50 km, the influence of primary roads without limited access within 10 to 50 km is found to be significant, and the impact is reversed when it was measured using block groups. The impact of local roads within 10 to 50 km is shown more clearly in the block group model. It implies that block group model can be better for discerning different impact of smaller segment of the space.

Table 37 Logit Models of Household Nonmotorized Travel by Adult Drivers

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Census Tract	**Sociodemographic**	**251.86**	**35**	**251.86**	**35**	**.036**
	Spatial density	**306.80**	**27**	**558.66**	**62**	**.079**
	Household density	*204.38*	*9*			
	Retail employee	*102.42*	*18*			
	Road infrastructure	**158.58**	**27**	**717.24**	**71**	**.101**
Block Group	**Sociodemographic**	**251.86**	**27**	**251.86**	**35**	**.036**
	Spatial density	**272.49**	**27**	**524.34**	**62**	**075**
	Household density	*131.98*	*9*			
	Retail employee	*140.51*	*18*			
	Road infrastructure	**182.18**	**27**	**706.52**	**89**	**.100**

Table 38 Logit Models of Nonmotorized Travel by Adult Drivers – Spatial Density

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
household density	0.00		0.00	
<10 %tile	0.01	0.635		
10th %tile	0.00	0.511	0.01	0.751
20th %tile	0.00	0.676	0.00	0.675
30th %tile				
40th %tile				
60th %tile	0.04	1.219		
70th %tile	0.01	1.274		
80th %tile	0.00	1.386	(0.07)	(1.176)
90th %tile	0.00	2.177	0.00	1.410
retail employees within 10 km	0.01		0.00	
<10 %tile	0.01	0.553	0.00	0.503
10th %tile	0.00	0.525		
20th %tile				
30th %tile				
40th %tile				
60th %tile	0.01	1.395	0.02	1.281
70th %tile				
80th %tile				
90th %tile	0.01	1.780	0.00	2.417
retail employees within 10 to 50km	0.00		0.00	
<10 %tile	0.01	0.510	0.00	0.566
10th %tile	0.02	0.589	0.00	0.633
20th %tile			0.03	0.789
30th %tile			0.00	0.691
40th %tile	0.01	0.649	0.00	0.516
60th %tile	0.00	1.849	0.01	1.292
70th %tile			(0.14)	(1.203)
80th %tile			0.00	2.016
90th %tile	0.01	2.335	0.00	3.824

Table 39 Logit Models of Nonmotorized Travel by Adult Drivers – Infrastructure

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
primary roads w/o limited access within 50 km	0.00		0.00	(within 10 to 50km)
<10 %tile				
10th %tile			0.00	0.653
20th %tile			0.01	0.767
30th %tile	0.00	1.689	0.00	1.292
40th %tile				
60th %tile	0.01	0.656		
70th %tile	0.04	0.725	0.00	1.282
80th %tile				
90th %tile				
Local roads within 10 km	0.00		0.00	
<10 %tile	0.00	2.484	0.00	1.557
10th %tile	0.00	2.207		
20th %tile	0.01	1.560	(0.09)	(1.192)
30th %tile				
40th %tile			0.05	0.821
60th %tile	0.01	0.704		
70th %tile	0.00	0.623		
80th %tile	0.00	0.547		
90th %tile	(.151)	(0.734)	0.04	0.731
Local roads within 10 to 50 km	0.00		0.00	
<10 %tile	0.00	2.684	0.00	3.006
10th %tile	0.02	1.738	0.00	1.665
20th %tile			0.00	1.465
30th %tile			0.00	1.622
40th %tile				
60th %tile				
70th %tile			0.00	0.693
80th %tile			0.00	0.625
90th %tile			0.00	0.247

6.6 High Occupancy Vehicle (HOV) Demand (Driving with Anyone as a Passenger)

Household high occupancy vehicle travel is more easily predicted than other modes, due to the fact that most of this travel involves the chauffeuring of children. Spatial density factors are marginally significant, and there is one significant road infrastructure variable, but sociodemographic factors are paramount (Table 40).

6.6.1 Census Tract Model

All dimensions of family size contribute to HOV demand, as seen in the results of Table 41. However, controlling for household size and composition, Black and Hispanic households, are less likely to generate HOV trips. Spatial density influences are relatively weak (Table 42), but, there is evidence that HOV demand benefits from a high degree of freeway accessibility (Table 43). HOV demand is highest for households that are located in areas that are above the 80th percentile in freeway coverage, controlling for socioeconomic factors and spatial density. Such households are typically located in San Francisco, Santa Clara, Contra Costa, and Marin Counties in the north and in Los Angeles and Orange Counties in the south. Such residential areas can also be found in San Diego, San Mateo, San Bernardino, Alameda, Riverside, Solano, and Ventura Counties. It is likely that this is capturing an HOV Lane provision effect.

Table 40 Logit Model of Driving an HOV

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	3489.30	35	3489.30	35	.255
Spatial density	49.47	27	3538.77	62	.258
Road infrastructure	27.25	9	3566.02	71	.260

Table 41 Logit Model of Any Household Member Driving HOV – Sociodemographic

Independent variable	Significance	Odds ratio
Income (base = unknown)	0.00	
<$10,000	0.00	0.704
$10,000-$24,999	0.00	0.858
$25,000-$34,999		
$35,000-$49,999	0.01	1.128
$50,000-$74,999	0.00	1.152
$75,000-$99,999	0.02	1.124
$100,000-$149,999	0.01	1.174
$150,000+		
household size (base = 6 or more)	0.00	
1	0.00	0.225
2	0.00	0.848
3	0.00	1.281
4	0.00	1.724
5	0.00	1.674
Average age of heads (base = unknown)	0.00	
18-25	0.04	0.862
25.5-35		
35.5-45	0.02	1.106
45.5-55		
55.5-65		
65.5-75	0.00	1.238
75.5+		
Ethnicity (base = unknown)	0.00	
White	0.00	1.155
Hispanic	0.03	0.871
Black	0.00	0.676
Asian/Pacific Islander		
White & Hispanic		
White & Asian		
Education (base = unknown)	0.00	
not high school graduate	0.00	0.696
high school graduate		
some college		
associates degree	0.00	1.198
bachelors degree	0.01	1.127
graduate degree	0.00	1.236
presence of children 0-5 yrs. old	0.00	1.884
presence of children 6-12 yrs. Old	0.00	1.993
presence of children 13-17 yrs. old	0.00	1.623

Table 42 Logit Model of Any Household member Driving HOV – Spatial Density

Ind. Variable (all bases = 50th %tile)	Significance	Odds ratio
tract household density	0.19	
<10 %tile		
10th %tile		
20th %tile		
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile		
retail employees within 10 km	0.03	
<10 %tile		
10th %tile		
20th %tile		
30th %tile	0.00	1.183
40th %tile	0.03	1.132
60th %tile		
70th %tile		
80th %tile		
90th %tile		
retail employees within 10 to 50 km	0.02	
<10 %tile		
10th %tile		
20th %tile	0.04	1.227
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile	0.02	0.683
90th %tile	(0.27)	(0.800)

Table 43 Logit Model of Any Household Member Driving HOV – Infrastructure

Variable (base = 50th %tile)	Significance	Odds ratio
primary roads with limited access within 50 km	0.00	
<10 %tile		
10th %tile		
20th %tile		
30th %tile	0.01	0.769
40th %tile		
60th %tile	0.02	0.806
70th %tile		
80th %tile	0.00	1.602
90th %tile	(0.076)	(1.446)

6.6.2 Comparison with Block Group Model

Household density doesn't have a significant impact on HOV demand as a passenger when it is measured using block group units. Among the retail employee density variable sets, only retail employees within 10km have an impact on HOV demand as a passenger and the total contribution of the retail employee variable set influence is lower in the block group model than in the census tract model. However, the impact of road infrastructure slightly increased in the block group model.

Table 44 Logit Models of Driving an HOV

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Census	**Sociodemographic**	**3489.30**	**35**	**3489.30**	**35**	**.255**
Tract	**Spatial density**	**49.47**	**27**	**3538.77**	**62**	**.258**
	Household density	*12.63*	*9*			
	Retail employee	*35.84*	*18*			
	Road infrastructure	**27.25**	**9**	**3566.02**	**71**	**.260**
Block	**Sociodemographic**	**3488.66**	**35**	**3488.66**	**35**	**.255**
Group	**Spatial density**	**22.39**	**9**	**3511.05**	**44**	**.256**
	Household density	-	-			
	Retail employee	*22.39*	*9*			
	Road infrastructure	**34.24**	**16**	**3545.29**	**60**	**.258**

Table 45 Logit Models of Any Household member Driving HOV – Spatial Density

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
household density	0.19			
<10 %tile				
10th %tile				
20th %tile				
30th %tile				
40th %tile				
60th %tile	0.04	1.116		
70th %tile				
80th %tile				
90th %tile				
retail employees within 10 km	0.03		0.01	
<10 %tile				
10th %tile				
20th %tile				
30th %tile	0.00	1.183	0.00	1.189
40th %tile	0.03	1.132	0.01	1.144
60th %tile				
70th %tile				
80th %tile				
90th %tile				
retail employees within 10 to 50km	0.02			
<10 %tile				
10th %tile				
20th %tile	0.04	1.227		
30th %tile				
40th %tile				
60th %tile				
70th %tile				
80th %tile	0.02	0.683		
90th %tile	(0.27)	(0.800)		

In contrast to the census tract model, primary roads without limited access within 10km and local roads within 10 to 50 km had a significant impact on HOV demand as a passenger instead of primary roads without limited access within 50km. The influence pattern of road infrastructure also contradicts the result of the census tract model. The 20th percentile of primary roads without limited access within 10km and the 10th percentile of local road within 10 to 50km has a positive relation with household HOV driving. We cannot be sure about this only from this comparison, but the block group variables seem to discern different impacts of spatial segments better than the census tract variables do.

Table 46 Logit Models of Any Household Member Driving HOV – Infrastructure

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
primary roads w/o limited access within 10 km			0.01	
<10 %tile				
10th %tile				
20th %tile			0.02	1.170
30th %tile				
40th %tile			0.02	0.858
60th %tile				
70th %tile				
80th %tile				
90th %tile				
primary roads w/o limited access within 50 km	0.00			
<10 %tile				
10th %tile				
20th %tile				
30th %tile	0.01	0.769		
40th %tile				
60th %tile	0.02	0.806		
70th %tile				
80th %tile	0.00	1.602		
90th %tile	(0.076)	(1.446)		
Local roads within 10 to 50 km			0.04	
<10 %tile				
10th %tile			0.04	1.127
20th %tile				
30th %tile				
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile				

6.7 Adult Driver as a Passenger in an HOV

The dependent variable in this section is the event of a driving adult being a passenger in an HOV vehicle during the observation period.

6.7.1 Census Tract Model

When we limit ourselves to HOV passenger travel by adult drivers, the predicted model is no longer driven by the presence of children, and there are no statistically significant spatial influences (Tables 47 and 48). Households in which adult drivers are more likely to be passengers in household vehicles include low income households, large households, and young households, which all imply car sharing. However, households with highly educated heads also are more likely to generate HOV travel by adult drivers. Small households, households in the \$100,000 - \$150,000 income range (in year 2000 USD), and households with either younger or older children are less likely to generate such travel.

Table 47 Logit Model of Household Adult Driver as HOV Passenger

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	1739.48	35	1739.48	35	.197
Spatial density	(Not significant)				
Road infrastructure	(Not significant)				

6.7.2 Comparison with Block Group Model

The road infrastructure variable set, which was not significant in the census tract model, found to be significant in the block group model (Table 49). Only the 60th percentile of local road infrastructure within 10 to 50 km had positive impact on adult driver's traveling as a passenger in an HOV (Table 50).

Table 48 Logit Model of Household Adult Driver as HOV Passenger

Independent variable	Significance	Odds ratio
Income (base = unknown)	0.00	
<$10,000	0.00	1.766
$10,000-$24,999	0.00	1.279
$25,000-$34,999		
$35,000-$49,999		
$50,000-$74,999		
$75,000-$99,999		
$100,000-$149,999	0.03	0.811
$150,000+		
household size (base = 6 or more)	0.00	
1	0.00	0.726
2	0.00	0.730
3	0.03	0.844
4		
5	0.01	1.397
Average age of heads (base = unknown)	0.31	
18-25	0.02	1.280
25.5-35		
35.5-45		
45.5-55		
55.5-65		
65.5-75		
75.5+		
Ethnicity (base = unknown)	0.00	
White		
Hispanic		
Black		
Asian/Pacific Islander		
White & Hispanic		
White & Asian		
Education (base = unknown)	0.00	
not high school graduate		
high school graduate	0.00	0.729
some college		
associates degree		
bachelors degree	0.00	1.442
graduate degree	0.00	2.030
presence of children 0-5 yrs. Old	0.02	0.767
presence of children 6-12 yrs. Old		
presence of children 13-17 yrs. old	0.00	0.689

Table 49 Logit Models of Household Adult Driver as HOV Passenger

	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Census Tract	**Sociodemographic**	**1739.48**	**35**	**1739.48**	**35**	**.197**
	Spatial density	(Not significant)				
	Road infrastructure	(Not significant)				
Block Group	**Sociodemographic**	**1732.04**	**35**	**1732.04**	**35**	**.197**
	Spatial density	(Not significant)				
	Road infrastructure	**17.64**	**9**	**1749.67**	**44**	**.198**

Table 50 Logit Models of Household Adult Driver as HOV Passenger – Infrastructure

Ind. Variable (all bases = 50^{th} %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
Local roads within 10 to 50 km			0.04	
<10 %tile				
10th %tile				
20th %tile				
30th %tile				
40th %tile				
60th %tile			0.05	1.146
70th %tile				
80th %tile				
90th %tile				

6.8 Adult HOV Passenger Travel Time

Another aspect of HOV demand is the amount of time (and indirectly distance) traveled by the survey participants. This also reflects destination chosen but also the home to work distance and it is influenced by housing-jobs balance, which is an important policy issue in our state.

6.8.1 Census Tract Model

An ordered Logit model was next used to explain the total time that household driving adults spend as HOV passengers, where that time was divided into deciles. The model results are listed in Tables 51 and 52. As this is an ordered Logit model, as opposed to a binary Logit model, Table 52 lists regression coefficients, rather than odds ratios. In this case, with a dependent variable that reflects space travelled, spatial density is more important than in the other two HOV models. It is also more important than sociodemographics, as seen in the goodness-of-fit statistics listed in Table 51. Road infrastructure also explains some of the variation albeit in a lesser fashion.

Table 51 Ordered Logit Model of Total Household Adult Driver HOV Passenger Time

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	79.16	35	79.16	35	.035
Spatial density	119.64	27	198.80	62	.085
Road infrastructure	20.55	9	219.35	71	.093

Socioeconomic factors explain very little of total household passenger travel time of adult drivers (Table 52). The only trend is that adults in younger households spend less time as HOV passengers. No other sociodemographic variable categories are statistically significant, but all variables are included in the model, in order to control for sociodemographic effects when assessing the influences of spatial variables. In contrast, the influence of spatial density factors is clearly of paramount importance (Table 53). Passenger travel time is highest for households located in residential areas with the lowest densities, although this effect is imprecisely estimated for two of the density categories. Likewise, adult passenger travel time is higher for households located in areas with the lowest accessibility to retail opportunities. This retail accessibility effect is stronger at the local, as opposed to the regional level. Finally, driving adults from households located in areas where there is the lowest level of accessibility to regional local roads in the region spend less time as HOV passengers than those located elsewhere, all else being equal (Table 54). Such residential areas tend to be located in the most rural of California's counties.

Table 52 Ordered Logit Model of Adult Driver HOV Pass. Time – Sociodemographic

Independent variable	Significance	Coefficient
Income (base = unknown)		
<$10,000		
$10,000-$24,999		
$25,000-$34,999		
$35,000-$49,999		
$50,000-$74,999		
$75,000-$99,999		
$100,000-$149,999		
$150,000+		
household size (base = 6 or more)		
1		
2		
3		
4		
5		
Average age of heads (base = unknown)		
18-25	0.02	-0.701
25.5-35	0.00	-0.978
35.5-45	0.04	-0.537
45.5-55	0.01	-0.682
55.5-65		
65.5-75		
75.5+		
Ethnicity (base = unknown)		
White		
Hispanic		
Black		
Asian/Pacific Islander		
White & Hispanic		
White & Asian		
Education (base = unknown)		
not high school graduate		
high school graduate		
some college		
associates degree		
bachelors degree		
graduate degree		
presence of children 0-5 yrs. Old		
presence of children 6-12 yrs. Old		
presence of children 13-17 yrs. Old		

Table 53 Ordered Logit Model of Adult Driver HOV Pass. Time – Spatial Density

Ind. variable (all bases = 50th %tile)	Significance	Coefficient
tract household density		
<10 %tile	0.00	0.682
10th %tile	(0.13)	(0.302)
20th %tile	(0.16)	(0.251)
30th %tile	0.05	0.321
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile		
retail employees within 10 km		
<10 %tile	0.00	0.975
10th %tile	0.00	0.900
20th %tile	0.00	0.556
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile		
retail employees within 10 to 50 km		
<10 %tile	(0.16)	(0.502)
10th %tile	(0.18)	(0.417)
20th %tile	0.04	0.556
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile		

Table 54 Ordered Logit Model of Adult Driver HOV Pass. Time – Infrastructure

Variable (base = 50th %tile)	Significance	Coefficient
local roads within 10 to 50 km		
<10 %tile	0.01	-1.016
10th %tile	0.00	-1.097
20th %tile	0.03	-0.615
30th %tile	0.00	-0.744
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile		

6.8.2 Comparison with Block Group Model

Except for the household density variable set, the block group variables contributed slightly better to the model than the census tract variables (Table 55).

Table 55 Ordered Logit Models of Total Household Adult Driver HOV Passenger Time

Model	Variable set	Contribution of set: Chi-square	Contribution of set: Degrees of freedom	Cumulative model: Chi-square	Cumulative model: Degrees of freedom	Cumulative model: Nagelkerke R^2
Census Tract	**Sociodemographic**	**79.16**	**35**	**79.16**	**35**	**.035**
	Spatial density	**119.64**	**27**	**198.80**	**62**	**.085**
	Household density	*76.31*	*9*			
	Retail employee	*43.33*	*18*			
	Road infrastructure	**20.55**	**9**	**219.35**	**71**	**.093**
Block group	**Sociodemographic**	**79.39**	**35**	**79.39**	**35**	**.035**
	Spatial density	**121.78**	**27**	**201.17**	**62**	**.086**
	Household density	*74.96*	*9*			
	Retail employee	*46.82*	*18*			
	Road infrastructure	**23.29**	**9**	**224.46**	**71**	**.095**

As shown in the result of the census tract model, HOV passenger travel time was lower for households located in residential areas with lower densities in the block group model, too, but the impact of the block group household density was weaker than the impact of the census tract household density. The impact of local retail employee density (within 10 km) was stronger in the block group model, and the highest 30% of the regional retail employee density (within 10 to 50km) had a positive impact in the block group model instead of the 20th percentile (Table 56).

Households with lower percentile of local roads within 10 to 50 km showed lower HOV passenger time in the block group model, too, but the impact was weaker than in the census tract model (Table 57).

Table 56 Ordered Logit Models of Adult Driver HOV Pass. Time – Spatial Density

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Coefficient	Significance	Coefficient
household density				
<10 %tile	0.00	0.682	(0.09)	(0.380)
10th %tile	(0.13)	(0.302)	0.01	0.452
20th %tile	(0.16)	(0.251)	(0.06)	(0.314)
30th %tile	0.05	0.321	(0.16)	(0.221)
40th %tile			0.01	0.411
60th %tile				
70th %tile				
80th %tile				
90th %tile				
retail employees within 10 km				
<10 %tile	0.00	0.975	0.00	1.525
10th %tile	0.00	0.900	0.00	0.849
20th %tile	0.00	0.556	0.00	0.548
30th %tile			0.08	0.308
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile				
retail employees within 10 to 50km				
<10 %tile	(0.16)	(0.502)		
10th %tile	(0.18)	(0.417)		
20th %tile	0.04	0.556		
30th %tile				
40th %tile				
60th %tile				
70th %tile			0.01	0.490
80th %tile			(0.07)	(0.330)
90th %tile			0.02	0.493

Table 57 Ordered Logit Models of Adult Driver HOV Pass. Time – Infrastructure

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Coefficient	Significance	Coefficient
Local roads within 10 to 50 km				
<10 %tile	0.01	-1.016	0.05	-0.397
10th %tile	0.00	-1.097	0.03	-0.397
20th %tile	0.03	-0.615	(0.28)	(-0.197)
30th %tile	0.00	-0.744	(0.17)	(0.242)
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile				

6.9 Solo Driving Demand - Household Solo Driving

Here we estimate a binomial choice model of whether or not a household engages in any amount of solo driving (i.e., whether the household generates single occupant vehicle, or SOV, travel). As in the previous case of passenger travel, we then model the level of solo driving in a separate model. Of the 16,750 households in our sample, 74.2% recorded an SOV trip segment.

6.9.1 Census Tract Model

The model using spatial opportunities computed at the tract level contains variables from all three sets, but the explanatory power was largely concentrated in the sociodemographic factors (Table 58).

Table 58 Logit Model of Household Solo Driving (SOV use)

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	2083.85	35	2083.85	35	.172
Spatial density	53.91	18	2137.76	53	.176
Road infrastructure	46.74	9	2184.50	62	.180

Whether or not household members engage in solo driving is predicted by almost all of the sociodemographic variables (Table 59). The propensity for solo driving increases as a function of income until the second-highest income category, and solo driving demand peaks with age of the household head(s) at the 45-55 category. Also, mixed White and Asian households are more likely to generate solo driving trips, while Black and Hispanic households are less likely. Among the spatial density variables (Table 60), household density at the census tract level is not significantly related to the probability that a household engages in solo driving, and consequently it is not included in the model. In terms of accessibility to retail activity sites, on both a local and regional level, SOV usage peaks at densities slightly below the median. SOV use is lowest for households residing in areas with the highest retail accessibility. Of all the road infrastructure variables, primary surface arterial coverage within 50 km is the most effective explanatory variable (Table 60). Controlling for sociodemographic characteristics and accessibility to retail services, a household located in an area with the 90th percentile of regional primary arterial coverage will be more likely to generate solo driving trips. These residential areas are concentrated in Los Angeles, Orange, and San Mateo Counties, but such areas can also be found in Alameda, Riverside, San Bernardino, San Francisco, Santa Clara, and Ventura Counties.

6.9.2 Comparison with Block Group Model

In the block group model, only one set of retail employee variables (retail employees within 10km) is found to be significant to the household solo driving behavior (Table 62). Households in residence area with the 20th – 60th household density are more likely to drive alone (Table 63). The impact of primary roads in the census tract model was divided into the impacts from two different segments (within 10km and within 10 to 50km) and they show different pattern. The 20th percentile of primary roads without limited access within 10km, and the 40th and 90th percentile of primary roads within 50km have positive impact on household solo driving (Table 64).

Table 59 Logit Model of Household Solo Driving – Sociodemographic

Independent variable	Significance	Odds ratio
Income (base = unknown)	0.00	
<$10,000	0.00	0.315
$10,000-$24,999	0.00	0.588
$25,000-$34,999		
$35,000-$49,999	0.00	1.264
$50,000-$74,999	0.00	1.487
$75,000-$99,999	0.00	1.649
$100,000-$149,999	0.00	1.749
$150,000+	0.04	1.239
household size (base = 6 or more)	0.00	
1	0.00	0.688
2	0.00	0.696
3		
4	0.00	1.437
5	0.00	1.324
Average age of heads (base = unknown)	0.00	
18-25	0.00	1.308
25.5-35	0.00	1.439
35.5-45	0.00	1.459
45.5-55	0.00	1.521
55.5-65		
65.5-75	0.00	0.674
75.5+	0.00	0.391
Ethnicity (base = unknown)	0.00	
White		
Hispanic	0.01	0.829
Black	0.00	0.670
Asian/Pacific Islander		
White & Hispanic		
White & Asian	0.04	1.446
Education (base = unknown)	0.00	
not high school graduate	0.00	0.702
high school graduate	0.00	0.864
some college		
associates degree	0.00	1.342
bachelors degree	0.00	1.228
graduate degree	0.00	1.403
presence of children 0-5 yrs. Old	0.00	0.721
presence of children 6-12 yrs. Old		
presence of children 13-17 yrs. Old		

Table 60 Logit Model of Household Solo Driving – Spatial Density

Ind. variable (all bases = 50th %tile)	Significance	Odds ratio
retail employees within 10 km	0.00	
<10 %tile	0.00	0.793
10th %tile		
20th %tile		
30th %tile	(0.09)	(1.109)
40th %tile	0.02	1.150
60th %tile		
70th %tile		
80th %tile		
90th %tile	0.01	0.812
retail employees within 10 to 50 km	0.00	
<10 %tile		
10th %tile		
20th %tile		
30th %tile		
40th %tile	0.00	1.292
60th %tile		
70th %tile		
80th %tile	0.00	0.729
90th %tile	0.00	0.615

Table 61 Logit Model of Household Solo Driving – Infrastructure

Variable (base = 50th %tile)	Significance	Odds ratio
primary roads without limited access within 50 km	0.00	
<10 %tile		
10th %tile		
0th %tile		
30th %tile	0.00	0.752
40th %tile	0.01	0.831
60th %tile		
70th %tile		
80th %tile		
90th %tile	0.00	2.050

Table 62 Logit Models of Household Solo Driving (SOV use)

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Census Tract	**Sociodemographic**	**2083.85**	**35**	**2083.85**	**35**	**.172**
	Spatial density	**53.91**	**18**	**2137.76**	**53**	**.176**
	Household density	-	-			
	Retail employee	*53.91*	*18*			
	Road infrastructure	**46.74**	**9**	**2184.50**	**62**	**.180**
Block Group	**Sociodemographic**	**2084.57**	**35**	**2084.57**	**35**	**.172**
	Spatial density	**39.05**	**9**	**2123.62**	**44**	**.175**
	Household density	-	-			
	Retail employee	*39.05*	*9*			
	Road infrastructure	**43.10**	**16**	**2166.73**	**60**	**.178**

Table 63 Logit Models of Household Solo Driving – Spatial Density

Ind. Variable (all bases = 50^{th} %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
retail employees within 10 km	0.00		0.00	
<10 %tile	0.00	0.793	0.01	0.851
10th %tile				
20th %tile			0.07	1.113
30th %tile	(0.09)	(1.109)	0.02	1.150
40th %tile	0.02	1.150		
60th %tile			0.01	1.181
70th %tile				
80th %tile				
90th %tile	0.01	0.812	0.00	0.761
retail employees within 10 to 50km	0.00			
<10 %tile				
10th %tile				
20th %tile				
30th %tile				
40th %tile	0.00	1.292		
60th %tile				
70th %tile				
80th %tile	0.00	0.729		
90th %tile	0.00	0.615		

Table 64 Logit Models of Household Solo Driving – Infrastructure

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Odds ratio	Significance	Odds ratio
primary roads w/o limited access within 10 km			0.02	
<10 %tile				
10th %tile			0.08	2.127
20th %tile				
30th %tile				
40th %tile				
60th %tile			0.00	0.746
70th %tile				
80th %tile				
90th %tile				
primary roads w/o limited access within 50 km	0.00		0.00	(within 10 to 50 km)
<10 %tile			(0.06)	(0.894)
10th %tile			(0.08)	(0.903)
20th %tile				
30th %tile	0.00	0.752		
40th %tile	0.01	0.831	0.02	1.151
60th %tile				
70th %tile			0.00	0.843
80th %tile				
90th %tile	0.00	2.050	0.00	1.287

6.10 Adult Solo Driving Time

The amount of solo driving time by adults in the household is also analyzed here to identify any correlational patterns with spatial density and infrastructure supply.

6.10.1 Census Tract Model

An ordered Logit model is used for the total time that household adults spend solo driving, where that time was divided into deciles. The model results are listed in Tables 65, 66, and 67 using census tract level spatial explanatory variables and listed in Tables 68, 69 are model results using block group level spatial explanatory variables.

In terms of Sociodemographics (Table 66), total household solo driving time is a monotonically increasing function of household income, and of household size. The presence of children also has a positive effect, which is greater for younger children. The only influence of age is that the oldest two categories of households generate less solo driving time, and there are three ethnic groups that also generate less than average solo driving time, controlling for all other variables that are Asian households, mixed White and Asian households, and Hispanic households. The spatial density effects (Table 67) are particularly revealing. Controlling for sociodemographic differences, households located in the lowest quintile of residential density spend more time solo driving, while those in the highest quintile of residential density spend less time. This implies that policies aimed at densification of residential areas will likely reduce solo driving time, *ceteris paribus*. Similarly, households located in areas with lower local (within 10 km) retail accessibility spend more time solo driving than households located in the highest level of retail accessibility. But, the opposite is true for regional retail accessibility: Households located in areas with above median regional retail accessibility travel more, while households located in areas with below-median regional retail accessibility travel less. We can surmise that the availability of local retail services reduces solo driving time, while the availability of services further from home increases such time, and conversely.

6.10.2 Comparison with Block Group Model

For adult solo driving time, the block group model worked slightly better and the pattern of the impact of the spatial variables did not show substantial difference in the two models (Tables 68 and 69).

Table 65 Ordered Logit Model of Total Household Solo Driving Time by Adults

Variable set	Contribution of set		Cumulative model		
	Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Sociodemographic	1811.24	35	1811.24	35	.121
Spatial density	237.64	27	2048.88	62	.136
Road infrastructure	(none significant)				

Table 66 Ordered Logit Model of Household Solo Driving Time – Sociodemographic

Independent variable	Significance	Coefficient
Income (base = unknown)		
<$10,000	0.00	-0.422
$10,000-$24,999	0.00	-0.386
$25,000-$34,999	0.04	-0.124
$35,000-$49,999		
$50,000-$74,999	0.00	0.200
$75,000-$99,999	0.00	0.431
$100,000-$149,999	0.00	0.484
$150,000+	0.00	0.456
household size (base = 6 or more)		
1	0.00	-1.458
2	0.00	-1.121
3	0.00	-0.570
4	0.00	-0.310
5		
Average age of heads (base = unknown)		
18-25		
25.5-35		
35.5-45		
45.5-55		
55.5-65		
65.5-75	0.00	-0.442
75.5+	0.00	-0.653
Ethnicity (base = unknown)		
White		
Hispanic	0.03	-0.178
Black		
Asian/Pacific Islander	0.03	-0.272
White & Hispanic		
White & Asian		
Education (base = unknown)		
not high school graduate		
high school graduate		
some college		
associates degree		
bachelors degree		
graduate degree		
presence of children 0-5 yrs. Old	0.00	0.269
presence of children 6-12 yrs. Old	0.00	0.192
presence of children 13-17 yrs. Old	0.00	0.140

Table 67 Ordered Logit Model of Household Solo Driving Time – Spatial Density

Ind. variable (all bases = 50th %tile)	Significance	Coefficient
tract household density		
<10 %tile	0.05	0.185
10th %tile	0.05	0.154
20th %tile		
30th %tile		
40th %tile		
60th %tile		
70th %tile		
80th %tile	(0.06)	(-0.125)
90th %tile	0.01	-0.208
retail employees within 10 km		
<10 %tile	0.05	0.191
10th %tile	0.00	0.234
20th %tile	0.01	0.183
30th %tile	0.02	0.158
40th %tile		
60th %tile		
70th %tile		
80th %tile		
90th %tile	0.01	-0.212
retail employees within 10 to 50 km		
<10 %tile	0.00	-0.487
10th %tile	0.00	-0.273
20th %tile	0.03	-0.145
30th %tile	0.01	-0.168
40th %tile		
60th %tile	0.01	0.179
70th %tile	0.00	0.391
80th %tile	0.00	0.432
90th %tile	0.00	0.512

Table 68 Ordered Logit Models of Total Household Solo Driving Time by Adults

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Census	**Sociodemographic**	**1811.24**	**35**	**1811.24**	**35**	**.121**
Tract	**Spatial density**	**237.64**	**27**	**2048.88**	**62**	**.136**
	Household density	*11.89*	*9*			
	Retail employee	*225.75*	*18*			
	Road infrastructure	(Not significant)				
Block	**Sociodemographic**	**1811.80**	**35**	**1811.80**	**35**	**.121**
Group	**Spatial density**	**267.61**	**27**	**2079.42**	**62**	**.138**
	Household density	*26.83*	*9*			
	Retail employee	*240.78*	*18*			
	Road infrastructure	(Not significant)				

Table 69 Ordered Logit Models of Household Solo Driving Time – Spatial Density

Ind. Variable (all bases = 50th %tile)	Census tract		Block group	
	Significance	Coefficient	Significance	Coefficient
household density				
<10 %tile	0.05	0.185	(0.11)	(0.140)
10th %tile	0.05	0.154	0.05	0.148
20th %tile			0.00	0.196
30th %tile				
40th %tile				
60th %tile				
70th %tile				
80th %tile	(0.06)	(-0.125)		
90th %tile	0.01	-0.208	(0.09)	(-0.123)
retail employees within 10 km				
<10 %tile	0.05	0.191	0.00	0.351
10th %tile	0.00	0.234	0.00	0.262
20th %tile	0.01	0.183	0.19	0.096
30th %tile	0.02	0.158	0.01	0.186
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile	0.01	-0.212	0.00	-0.244
retail employees within 10 to 50km				
<10 %tile	0.00	-0.487	0.00	-0.495
10th %tile	0.00	-0.273	0.00	-0.296
20th %tile	0.03	-0.145	0.01	-0.166
30th %tile	0.01	-0.168	0.01	-0.184
40th %tile				
60th %tile	0.01	0.179	0.01	0.166
70th %tile	0.00	0.391	0.00	0.362
80th %tile	0.00	0.432	0.00	0.403
90th %tile	0.00	0.512	0.00	0.503

7. Models Combining Sociodemographics and Spatial Variables from Tracts and Block Groups

In section 6 the comparison between the census tract models and the block group models implies that the census tract household density explains many travel behavior indicators better than the block group household density but that the other block group level variables perform better than the census group level variables in most regression models. Therefore, an attempt to build better models by combining the census tract household density and the other variables computed at the block group level was made, and three out of a total of ten models are improved in terms of their goodness of fit by the combination of the variables. They are the models of nonmotorized travel by any household member, nonmotorized travel by an adult driver in the household and adult HOV passenger travel time. Below are the summaries of the model estimation.

7.1 Nonmotorized Travel by any Household Member

The first quadrant of Table 70 shows the chi-square contribution of spatial variables measured at the block group level. In the second quadrant we show the impact of combining in the model specification variables measured at the census tract with variables measured at the block group level. The combination provides a slightly better fit using a smaller amount of degrees of freedom. It is also important to note the unaltered chi-square contribution of sociodemographics between the two specifications.

Table 70 Logit Models of Any Household Nonmotorized Travel

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Block Group	Sociodemographic	1065.94	35	1065.94	35	.110
	Spatial density	325.60	27	1391.54	62	.143
	Road infrastructure	156.27	27	1547.82	89	.158
BG variables with CT HH density	Sociodemographic	1065.54	35	1065.54	35	.110
	Spatial density	382.99	27	1488.53	62	.148
	Road infrastructure	126.67	19	1575.20	81	.161

Table 71 Logit Models of Household Nonmotorized Travel – Spatial Density

Ind. Variable (all bases = 50th %tile)	Block group		BG variables with CT HH density	
	Significance	Odds ratio	Significance	Odds ratio
household density	0.00		0.00	
<10 %tile			0.00	0.690
10th %tile	0.00	0.696	0.00	0.532
20th %tile	0.00	0.618	0.00	0.611
30th %tile			0.04	0.854
40th %tile				
60th %tile			0.05	1.161
70th %tile	0.02	1.182	0.01	1.236
80th %tile	0.03	1.172	0.00	1.447
90th %tile	0.00	1.474	0.00	2.036
retail employees within 10 km	0.00		0.00	
<10 %tile	0.00	0.563	(0.09)	(0.804)
10th %tile				
20th %tile			(0.07)	(1.171)
30th %tile				
40th %tile			(0.09)	(0.874)
60th %tile				
70th %tile			(0.09)	(0.868)
80th %tile			0.01	0.795
90th %tile	0.00	1.996	0.00	1.525
retail employees within 10 to 50km	0.00		0.00	
<10 %tile	0.00	0.650	0.00	0.678
10th %tile	0.00	0.698	0.00	0.749
20th %tile	0.02	0.806		
30th %tile	0.00	0.760	0.00	0.780
40th %tile	0.00	0.570	0.00	0.575
60th %tile				
70th %tile	0.06	1.203	(0.10)	(1.174)
80th %tile	0.00	1.961	0.00	1.866
90th %tile	0.00	3.243	0.00	2.816

Table 72 Logit Models of Household Nonmotorized Travel – Infrastructure

Ind. Variable (all bases = 50th %tile)	Block group		BG variables with CT HH density	
	Significance	Odds ratio	Significance	Odds ratio
primary roads w/o limited access within 10 to 50 km	0.00		0.00	
<10 %tile				
10th %tile	0.00	0.735	0.00	0.745
20th %tile	0.03	0.842	0.02	0.829
30th %tile				
40th %tile				
60th %tile				
70th %tile	0.03	1.164	0.04	1.160
80th %tile				
90th %tile				
Local roads within 10 to 50 km	0.00		0.00	
<10 %tile	0.00	2.231	0.00	2.091
10th %tile	0.00	1.635	0.00	1.606
20th %tile	0.01	1.321	0.01	1.275
30th %tile	0.00	1.367	0.00	1.359
40th %tile				
60th %tile				
70th %tile	0.00	0.698	0.00	0.692
80th %tile	0.00	0.662	0.00	0.697
90th %tile	0.00	0.312	0.00	0.343

7.2 Nonmotorized Travel by an Adult Driver in the Household

The impact of combining spatial variables from the two levels of spatial aggregation is similar to the model of Table 70 but this time the degrees of freedom are the same between the model that uses only block group level independent variables and the model that uses a combination of block group level with census tract level variables due to the inclusion of more variables.

Table 73 Logit Models of Household Nonmotorized Travel by Adult Drivers

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Block Group	Sociodemographic	251.97	35	251.86	35	.036
	Spatial density	272.49	27	524.34	62	075
	Road infrastructure	182.18	27	706.52	89	.100
BG variables with CT HH density	Sociodemographic	251.97	35	251.86	35	.036
	Spatial density	312.70	27	564.67	62	.080
	Road infrastructure	164.57	27	729.24	89	.103

Table 74 Logit Models of Nonmotorized Travel by Adult Drivers – Spatial density

Ind. Variable (all bases = 50th %tile)	Block group		BG variables with CT HH density	
	Significance	Odds ratio	Significance	Odds ratio
household density	0.00		0.00	
<10 %tile				
10th %tile	0.01	0.751	0.00	0.568
20th %tile	0.00	0.675	0.00	0.656
30th %tile				
40th %tile				
60th %tile				
70th %tile				
80th %tile	(0.07)	(1.176)	0.00	1.347
90th %tile	0.00	1.410	0.00	1.938
retail employees within 10 km	0.00		0.00	
<10 %tile	0.00	0.503	0.02	0.691
10th %tile				
20th %tile				
30th %tile				
40th %tile				
60th %tile	0.02	1.281	(0.08)	(1.199)
70th %tile				
80th %tile			0.06	0.775
90th %tile	0.00	2.417	0.00	1.855
retail employees within 10 to 50km	0.00		0.00	
<10 %tile	0.00	0.566	0.00	0.591
10th %tile	0.00	0.633	0.00	0.681
20th %tile	0.03	0.789		
30th %tile	0.00	0.691	0.00	0.714
40th %tile	0.00	0.516	0.00	0.531
60th %tile	0.01	1.292	0.04	1.247
70th %tile				
80th %tile	0.00	2.016	0.00	1.902
90th %tile	0.00	3.824	0.00	3.363

Table 75 Logit Models of Nonmotorized Travel by Adult Drivers – Infrastructure

Ind. Variable (all bases = 50th %tile)	Block group		BG variables with CT HH density	
	Significance	Odds ratio	Significance	Odds ratio
primary roads w/o limited access within 10 to 50 km	0.00		0.00	
<10 %tile				
10th %tile	0.00	0.653	0.00	0.658
20th %tile	0.01	0.767	0.00	0.752
30th %tile	0.00	1.292	0.00	1.276
40th %tile				
60th %tile				
70th %tile	0.00	1.282	0.00	1.270
80th %tile				
90th %tile				
Local roads within 10 km	0.00		0.01	
<10 %tile	0.00	1.557	0.00	1.476
10th %tile				
20th %tile				
30th %tile				
40th %tile	0.05	0.821	0.04	0.815
60th %tile				
70th %tile				
80th %tile				
90th %tile	0.04	0.731	(0.06)	(0.753)
Local roads within 10 to 50 km	0.00		0.00	
<10 %tile	0.00	3.006	0.00	2.828
10th %tile	0.00	1.665	0.00	1.589
20th %tile	0.00	1.465	0.01	1.398
30th %tile	0.00	1.622	0.00	1.576
40th %tile				
60th %tile				
70th %tile	0.00	0.693	0.00	0.703
80th %tile	0.00	0.625	0.00	0.665
90th %tile	0.00	0.247	0.00	0.270

7.3 Adult HOV Passenger Travel Time

When examining the amount of travel time by HOV the impact of using a combination of variables from different geographic levels is small in terms of the improvement in goodness-of-fit as Table 76 shows.

Table 76 Ordered Logit Models of Total Household Adult Driver HOV Passenger Time

Model	Variable set	Contribution of set		Cumulative model		
		Chi-square	Degrees of freedom	Chi-square	Degrees of freedom	Nagelkerke R^2
Block Group	Sociodemographic	79.39	35	79.39	35	.035
	Spatial density	121.78	27	201.17	62	.086
	Road infrastructure	23.29	9	224.46	71	.095
BG variables with CT HH density	Sociodemographic	79.39	35	79.39	35	.035
	Spatial density	128.68	27	208.07	62	.088
	Road infrastructure	22.91	9	230.98	71	.098

Table 77 Ordered Logit Models of Adult Driver HOV Pass. Time – Infrastructure

Ind. Variable (all bases = 50^{th} %tile)	Census tract		BG variables with CT HH density	
	Significance	Coefficient	Significance	Coefficient
<10 %tile	0.05	-0.397	(0.12)	(-0.313)
10th %tile	0.03	-0.397	0.05	-0.369
20th %tile				
30th %tile				
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile				

Table 78 Ordered Logit Models of Adult Driver HOV Pass. Time – Spatial Density

Ind. Variable (all bases = 50th %tile)	Block group		BG variables with CT HH density	
	Significance	Coefficient	Significance	Coefficient
household density				
<10 %tile			0.00	0.653
10th %tile	0.01	0.452		
20th %tile	(0.06)	0.314	0.02	0.392
30th %tile			0.03	0.362
40th %tile	0.01	0.411		
60th %tile				
70th %tile				
80th %tile				
90th %tile				
retail employees within 10 km				
<10 %tile	0.00	1.525	0.00	1.350
10th %tile	0.00	0.849	0.00	0.711
20th %tile	0.00	0.548	0.02	0.466
30th %tile	0.08	0.308	(0.18)	(0.239)
40th %tile				
60th %tile				
70th %tile				
80th %tile				
90th %tile				
retail employees within 10 to 50km				
<10 %tile			0.04	-0.406
10th %tile				
20th %tile				
30th %tile				
40th %tile				
60th %tile				
70th %tile	0.01	0.490	0.01	0.483
80th %tile	(0.07)	(0.330)	(0.09)	(0.308)
90th %tile	0.02	0.493	0.02	0.471

8. Summary and Conclusions

The critical issues of optimal allocation of resources for infrastructure facilities is addressed in this report from two different viewpoints that are both covering the entire state of California and play the role of pilot tests for the creation of a Geographic Information System based tool. The first viewpoint is a macronalysis that considers the entire state and makes relative assessments of efficiency and inequality. US census tracts and US census block groups are used as the units of analysis to compute a variety of indicators and then study efficiency and equity providing a practical implementation using Geographic Information Systems. The models thus derived are capable to identify places where investment will lead to gains and places where equity is more likely to be advanced. The second viewpoint is at the microanalytical level. Households and the individuals within households use the allocated facilities in their neighborhoods to participate in activities. In this analysis we examine residential location spatial distributions and focus on the choice of modes and identify through regression models the impact of infrastructure supply on behavior and the impact of spatial densities. In this way we can also examine if the impact of optimal allocation of resources will also translate in different behaviors that may amplify the benefits of allocations or may even inhibit them.

In the efficiency analysis access to activity opportunities in a variety of environments for the entire state of California is analyzed. First, three principal components are used to derive summaries of 25 variables describing the diverse access patterns. These three components represent access to opportunities in a sequence of concentric regions around a virtual origin, i.e. a centroid, in each of 7049 tracts used to subdivide California geographically. The first region is a circle of 5 km network distance radius around each centroid. The second region is a ring between 5 km and 10 km around the centroid, and the third is an outer ring between 10 km and 50 km network distance from each centroid. Using the derived principal components as the dependent variables and lane kilometers of roadways as the independent variables we employ stochastic frontier analysis to identify a complex set of relationships showing that more roadways is not always better for access to opportunities, either because of competition for space or because of the spatial distribution of activity opportunities that does not follow these roadways but obeys other spatial distribution rules. The regression results also show that the role of roadways depends on the indicator considered but also the presence of other surrounding roadways. Overall, however, the presence of primary roadways has a strong positive impact on access. For core access the secondary roadways seem to have a much higher impact and merit attention for investment. Efficiency in the transformation of roadways to access depends on the residents of each tract and depends on the measurement of access (outer ring vs. middle ring). This begs for a more detailed analysis possibly using much smaller geographical areas. Preliminary tests using Santa Barbara County as the pilot case indicate substantial differences in the findings when the Census-tracts are large (e.g., in rural environments). Repetition of the analysis here using the smaller geographical units can be used to reveal within tract potential for optimal allocation of resources but other project priorities did not allow us to pursue this further.

Although the data analysis offers unique and unprecedented insights at a statewide level, our study here unavoidably suffers from a variety of limitations. Employment of the principal components as a dependent variable does not allow a clear linkage between access to specific opportunities (e.g., retail, health, and so forth) and their relationship to highway types. In addition the interconnectedness of the highways makes identification of specific optimal investment segments very hard when aggregation at the level of a tract is used. The effect of data transformations to express variables in logarithms may also add approximations. In a continuation of the research here, as mentioned above, we will examine smaller geographical regions but from the inequality viewpoint applied to nested geographical areas. In addition, efforts are also directed towards a better description of the highway quality and performance and the incorporation of access provided by other modes. The parallel microanalysis also examines the mode choice of individual traveler data to continue the assessment and correlation between facilities and optimal level of service provision.

In the equity analysis in this report access to activity opportunities in a variety of environments for the entire state of California is analyzed using a hierarchy of geographic subdivisions starting from the US Census block group level. First, factor analysis is used to derive a summary of 25 variables describing the diverse access patterns. One factor emerged as sufficient describing the variation of accessibility among the 22,133 block groups used to subdivide California geographically. Using this derived factor values and the population residing in each block group an index for the entire state was computed that measures the disparities in accessibility featured by the block groups in regard to their population. This same index can thus constitute a first help for policy makers who consider equality as a criterion of allocation of infrastructure investment. Finally we implement a fractal inequality index that gives us a better understanding of the spatial distribution of inequality throughout different geographical scales. This index gives information about the disparities in accessibility between Counties as well as inside the Counties themselves.

For these reasons, and despite the limitations of this case study (e.g., the congestion effects are not taken into account, highways and implicitly roadway travel is the only transportation system accounted for) we can already conclude that the Theil index we implemented here constitutes a tool both easy to understand thanks to its intuitive definition, easy to implement since it relies on data that are largely available, and able to give instructive information about the structure of inequality in providing access to residents.

Nevertheless, this study constitutes only a first step in the utilization of this tool. In fact, the Theil index can reveal much more opportunities. First of all, it could give more instructive results if applied to more complete data, such as data that include congestion effects and public transportation information. But it could also constitute an approach to inequality dynamics analysis, notably by predicting the evolution of inequality as a consequence of public policies or demographic changes. For instance, the study of each loading that composes the accessibility factor can offer a deeper understanding of the variables a public policy should focus on to reduce inequality. As well, one could study the sensitivity of the accessibility index to demographic dynamics in order to choose the investments that can follow demographic trends with a goal of equal development of the territory.

The wealth of the spatial indicators developed for the efficiency and equity analysis is then used as supplemental information when moving to the microanalysis, which contains an analysis of residential location choice and more detailed analysis of travel behavior. The residential household location analysis simply validates that in California different social, demographic, and ethnic groups reside in places of wide variety of transportation supply and land use environments and for this reason enjoy different potential for activity participation.

The accessibility variables generated for the ten travel behavior models contributed significantly in explaining the selected travel behavior. We analyzed how each set of the accessibility variables affect different travel behaviors in different ways. Household density, retail employee density and road infrastructure provided meaningful explanation of various travel behavior facets with description of different dimensions of accessibility such as characteristics of residential area, availability of activity opportunity, and connectivity through road infrastructure. This complicates the study of land use and travel behavior because policy actions may create counteracting effects on travel behavior. One way to resolve lack of clear indications about land use policy impacts is to employ microsimulation and capture direct and indirect effects of land use policy in a way that comes closer to real behavioral impact assessment.

We also analyzed the impact of modifiable areal unit problem (MAUP) on the travel behavior models. MAUP is one of the important issues that have to be considered in spatial analysis, and arises from imposing artificial unit areas on continuous geographical phenomena. From among the many possible ways of generating spatial variables using GIS we select a small number and illustrate how those variables can be used in travel behavior models to build better models. From the model estimation experiments a variety of findings emerge.

First, from the comparisons between the census tract models and the block group models, we demonstrate the difference between the two. Household density measured at census tracts explained better the behavioral indicators used here than household density measured using block groups. Census tracts cover a larger area around a residence and therefore capture the density impact in more informative ways. However, this cannot be the golden rule for every travel behavior indicator. We need to think about the implications that a specific areal unit has on each travel behavior indicator, test its ability to explain behavior, and decide to use the one that is the most informative.

Second, spatial variables involving shortest paths in computation showed better ability of discerning the impacts of each spatial segment and also clearer impact patterns of each variable set when they are computed using smaller unit areas than when they are computed using larger unit areas. Smaller unit areas provide closer approximation of the variables and those variables seem to be less susceptible to measurement error than variables computed using larger geographical units. However, the trade off between obtaining closely approximated explanatory variables and the computing demand of using smaller areal units has to be considered when we decide which areal unit we want to use. In fact, the improvement in the goodness of fit for some regression models was marginal or even totally absent. Moreover, the two aggregation levels used here have their own inherent advantages and disadvantages. Consequently, we also demonstrate building models using spatial variables with both levels with some clear benefits. A better comparison among spatial independent variables can be achieved by moving one level lower in the disaggregation and consider accessibility indicators computed at the individual level. In addition, instead or examining mode choice alone we can expand the analytical envelope and use an activity-based approach to the analysis here. This is a task of a project in the University of California Transportation Center.

9. Next Steps

Evidence from this project shows the tremendous potential a GIS tool has for optimal resource allocation of resources. Unavoidably for pragmatic reasons many approximations remain and require attention. In this section we focus, however, on project activities that have the highest potential to help us meet the overall objective and they are: a) a longitudinal analysis between year 2000 and year 2010; b) project tracking and accessibility impact assessment; and c) expand the analysis to include comprehensive treatment of travel behavior.

The entire analysis was done using data from the year 2000. The data are from products such as the Census Transportation Planning Package and a roadway network vintage 2000. The household behavior data span a few months in 2000 and 2001. As a result all the analytical findings are for that period and may not be extendable to other times. The macroanalysis (efficiency and equity) should be expanded to include other years as opportunities for new data are multiplying due to the American Community Survey, which in 2010 will release its 5-year estimates for areas with a population of less than 20,000, including census tracts and block groups. This may provide an unprecedented opportunity to study the evolution of accessibility in our state and identify the places and their sociodemographic groups that benefitted the most by pinpointing geographic areas that increased or decreased residents' accessibility.

A major barrier in the early stages of this research project has been the lack of suitable databases of specific infrastructure building tracking statewide. Armed with experience and knowledge from our research we can design a procedure to rebuild the past ten-year history of infrastructure development, correlate projects with improvements in accessibility, and repeat the analysis here to more clearly show the impact of specific projects on improving efficiency and equity for the resident population.

In the third major area of next steps we can expand the microanalysis to a more comprehensive treatment of travel behavior. This includes activity participation and interactions among household members, trip consolidation in the form of tours, and also the more traditional analysis of trip making. In addition to offering a more detailed picture of the impact infrastructure and density of opportunities causes on travel behavior, this next step has also the potential to improve the statewide transportation model maintained by Caltrans. This last area of analysis is also a fruitful research direction in developing a next generation of land use transportation integrated models.

References

Alam, J. B., S. H. Sikder and K. G. Goulias (2004) Role of transportation in regional economic efficiency in Bangladesh: Data Envelopment Analysis, *Transportation Research Record 1864, Journal of the Transportation Research Board*, pp.112-120.

Alam J. B., S. H. Sikder and K. G. Goulias (2005) Assessing time lag between transportation investment and economic development by the Data Envelopment Approach. *Transportation Research Record 1932, Journal of the Transportation Research Board,* pp. 79-88.

Alam Md. J. B., J. W. Polak, and K.G. Goulias (2008) Regional efficiency and accessibility demand assessment using Stochastic Data Envelopment Analysis. Paper presented at the 87th Transportation Research Board Annual Meeting, January 13-17, 2008, and included in the CD ROM proceedings.

Allen, B.L., D.W. Butterfield, A. Kazakov, M.L. Kliman, A.A. Kubursi, and J.D. Welland (1988) Measuring economic stimulation from capital investment in transportation, *Transportation Research Record 1197*, TRB, National Research Council, Washington, D.C., pp49-55.

Berechman, J. (1994) Urban and regional economic impacts of transportation investment: A critical assessment and proposed methodology, *Transportation Research A, 28(4)*, 1994, pp351-362.

Buffington, J.L., L.M. Crane, B. Clifton, and J.R. Speed (1992) Methodology for estimating the economic impacts of highway improvements: Two case studies in Texas, Presented at the 71st Annual Transportation Research Board Meeting, Washington, D.C..

Conceicao P. and P. Ferreira (2000) The Young Person's Guide to the Theil Index. UITP Working Paper, No.14. University of Texas Inequality Project. University of Texas, Austin, TX.

Ghose A. K. (2004) Global inequality and international trade. *Cambridge Journal of Economics 28*:229-252.

Gkritza K., K.C. Sinha, S. Labi, and F.L. Mannering (2008) Influence of highway construction projects on economic development: an empirical assessment. *Annals of Regional Science*, 42:545–563.

Greene, W. (1980) Maximum likelihood estimation of econometric frontier functions, *Journal of Econometrics*, Volume 13(1), pp. 27-56.

Greene, W.H. (2002) *LIMDEP Version 8.0 Econometric Modeling Guide*. Econometric Software Inc. Plainview, NY.

Hummon, N.P., L. Zemotel, A.G.R. Bullen, and J.P. DeAngelis (1986) Importance of transportation to advanced technology companies. *Transportation Research Record,* 1076, TRB, National Research Council, Washington, D.C., pp. 1-7.

Krugman P. and A. J. Venables (1995) Globalization and the Inequality of Nations. *The Quarterly Journal of Economics*, Vol. 110, No. 4, pp. 857-880

Openshaw S. and S. Alvanides (1999) Applying geocomputation to the analysis of spatial distributions. In Longley P, Goodchild M, Maguire D, Rhind, D (eds) *Geograhpic Information Systems: Principles and Technical Issues*. Vol 1, 2nd ed. New York: John Wiley and Sons Inc.

Palmquist, R.B. (1982) Impact of highway investment on property values, *Transportation Research Record,* 887, TRB, National Research Council, Washington, D.C., pp. 22-29.

Perrera, M.H. (1990) Framework for classifying and evaluating economic impacts caused by transportation improvement, *Transportation Research Record,* 1274, TRB, National Research Council, Washington, D.C., pp. 41-52.

Seskin, S.N. (1990) Comprehensive framework for highway economic impact assessment: Methods and results, *Transportation Research Record,* 1274, TRB, National Research Council, Washington, D.C., pp. 24-34.

Schneider M.C. , C Castillo-Salgado, E Loyola-Elizondo, J Bacallao, O J Mujica, M Vidaurre and G A O Alleyne (2002) Trends in infant mortality inequalities in the Americas: 1955–1995. *Journal of Epidemiology and Community Health*, 56, pp.538-541.

Weisbrod, G.E. and J. Beckwith (1992). Measuring economic development benefits for highway decision-making: The Wisconsin case, *Transportation Quarterly 46(1)*, pp. 57-79.

Wilson, F. R., G.M. Graham, and M. Aboul-Ela (1985) Highway investment as a regional development policy tool, *Transportation Research Record,* 1046, TRB, National Research Council, Washington, D.C., pp.10-14.