Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Profiling commuters' travel behavior in the pacific states of the continental U.S.

Tayo Fabusuyi[a,*], Robert Hampshire[a], Zhen (Sean) Qian[b]

[a]*University of Michigan Transportation Research Institute, Ann Arbor, MI 48109*
[b]*Carnegie Mellon University, Pittsburgh, PA 15213*

Abstract

We develop a more nuanced understanding of commuters' travel mode choices by generating detailed profiles that capture the travel behavior of commuters in the Pacific states of the continental US. These profiles are created by utilizing the US Census Public Use Microdata Sample (PUMS) data. The microdata sample set allows for the estimation of fine-grained models that showcase how individual commuters make travel mode choices. Our results show appreciable locational variation in mode choices and statistically significant differences in commuting profile across and within population segments. A key revealing finding demonstrates that across the three states analyzed, the total number of vehicles driven for any day of the week could be reduced by up to 10 million assuming the commuting patterns observed in San Francisco applies to the rest of the states. We conclude with insights and policy implications provided by the study on making transportation-related infrastructure decisions.

Keywords: travel profile; travel mode; public use microdata sample (PUMS); balanced repeated replication (BRR); multinomial probit (MNP); marginal effects

1. Introduction

The factors that explain commuters' travel mode have been the subject of numerous studies starting from the seminal work of Daniel McFadden on urban travel demand in 1974 (McFadden, 1974). Of late, there has been a resurgence in this line of research with much of the attention devoted to understanding why most commuters drive alone (McKenzie, 2013). Key drivers of this interest include the linkage between travel mode and greenhouse gas emissions (United Kingdom Department of Transport, 2008); the smart cities concept and the shifting demographic towards cities (Frey, Washington D.C., 2010). Of key relevance to the discourse is what this trend portends specifically with regards to travel mode choices and the attendant demand for parking spaces the prevailing travel behavior creates

* Corresponding author: Fabusuyi@umich.edu

(Cortright, 2015), (Fabusuyi T. , Hampshire, Hill, & Sasanuma, 2014).

Modeling travel mode choice has historically been carried out using the measurable attributes of the commuter and the attributes of the travel mode alternatives (Ben-Akiva & Lerman, Discrete Choice Analysis: Theory and Application to Travel Demand, 1985). An appreciable degree of modification has since been made to this approach with the impetus coming largely from the travel behavior research community. The rationale behind the rethinking is the realization that travel mode choice are influenced by lifestyles, preferences and an individual's values (Ben-Akiva, Walker, Bernardino, Gopinath, & Morikave, 1999). These often manifest as unobserved heterogeneities of preferences across individuals, and a concern in econometric analysis is how to address the attendant assumption violations they create.

Our approach to examining what factors motivate an individual to drive to work alone builds on the aforementioned albeit with a slightly different method. Using the US Census Public Use Microdata Sample (PUMS) data from the Pacific region of the US, specifically the states of California, Oregon and Washington, we generate travel mode profiles for representative individuals for each geographical area of interest. The dataset provides an excellent resource for understanding commuting patterns for segments of the population at sub-city level, a feat that was not possible in the past given the relatively small sample of existing surveys. In addition to determining the variations in travel behavior across demography and geography, a key objective of this research is in examining how sensitive different segments of the commuting population are to travel mode choices given changes in the covariates of interest.

A unique contribution of our study are the fine-grained travel profiles generated for subsets of the commuting population. These profiles are created by computing marginal effects of key explanatory variables not only at the mean but also by graphing the probabilities of a representative individual choosing a specific travel mode based on all the possible values that could be assumed by the control variables of interest. Compared to existing literature, this approach presents a more visual and readable option and vividly shows the differences across the subsets of commuters by stacking their travel mode profiles on the same graph. More importantly, the detailed profiles offer rich insights on commuting patterns that could be localized to distinct population segments thus informing the design of targeted policy measures that may be superior to broad based policy prescriptions.

The rest of the paper is organized as follows. The following section provides a review of the existing literature. Section three identifies the sources and provides descriptive statistics of the data employed for the study. The empirical strategy and model specifications are addressed in the fourth section while section five presents the results from the study and discusses the findings. The concluding section recaps and provides insight on the caveats associated with the study.

2. Review of existing literature

The theoretical basis for most travel choice models is the random utility model (RUM) (Manski & Lerman, 1977). The model examines how changes in socio-economic and demographic composition impact commuting patterns. Empirical work on subsets of the population include those on gender differentials in commuting patterns (Rosenbloom & Burns, 1993), (Rosenbloom & Burns, 1994), (Weinberger, 2007); travel mode choices of immigrants (Blumenberg & Smart, 2014), (Blumenberg & Evans, 2010), (Kim, 2009); and the works of Guiliano (Giuliano, 2003), Blumenberg and Smart (Blumenberg & Smart, 2009) that examine differences across racial and ethnic identities. Some of these analyses have been directed to addressing changing demographics, specifically with respect to immigration and the aging population and how these trends will impact aggregate travel mode choices.

Comparative assessments of travel mode choices for population segments and the spatial variation in mode choices across geographical areas have been carried out using PUMS dataset. Examples of studies utilizing PUMS data to analyze travel behavior include the work by Blumenberg and Shiki (Blumenberg & Shiki, 2007), which used 2000 PUMS data to examine travel mode choices of the foreign-born population of California. The study also looked at the relationship between transit usage rates and the tenure of U.S. residency. PUMS data has also been used to show differences in commuting patterns across gender (Weinberger, 2007), (Krizek, Johnson, & Tilahun, Nov 18-20, 2004); in the analysis of access to job opportunities (Hu & Giuliano, Jan 23-27, 2011); and for land use policies (Haas, Makarewicz, Benedict, Sanchez, & Dawkins, July 2006). Further credence to the use of PUMS data in examining travel behavior is also provided by the works of Thakuriah et al (Thakuriah, Sriraj, Soot, Liao, & Berman, 2005) on job access and reverse commute; on the transportation infrastructure implications of driving solo for a specific city (Fabusuyi & Hampshire, 2017); and an activity-based demand model by Donnelly et al (Donnelly, Erhardt, Moeckel,

& Davidson, 2010) and Deal et al (Deal, Kim, & Pallathucheril, 2009) that examines the impact of growth management policies on travel mode choice.

Across the empirical studies surveyed, there are significant differences in the type of model used, though there has been movement towards the use of more robust models that can address unobserved variables including latent attitudes and variations in taste and preferences. Huang et al (Huang, Yang, & Bell, March 2000), for example, posit that preferences for autonomy, including self-reliance, independence and privacy, in addition to economic preferences, are contributory factors in determining travel mode choice. Added to this, Charles and Kline) (Charles & Kline, 2006) argue for the need to consider social capital as a key element in determining carpooling. They make the case that while traditional economic factors come into play, social identification and social capital also matter - a view point shared by Blumenberg & Smart (Blumenberg & Smart, 2010) that looks at the carpooling behavior of the immigrant population in Southern California.

Recognizing these limitations and driven by the need to reflect the unique attributes of our dataset, we model commuters' travel mode choices using the multinomial probit (MNP) approach with the travel mode options categorized into four alternatives. This is achieved by assuming a utility function that is linear in parameters and error terms that are Gaussian and not independently distributed from one another. This modification means that the off-diagonal elements for the error terms variance-covariance structure are not constrained to take on only zero values and thus are correlated. The model is robust enough as to be relevant to the transportation field on heterogeneity of taste (Garrido & Mahmassani, 2000) and with providing plausible assumptions with regards sample size and superior estimates as argued by Munizaga et al (Munizaga, Heydecker, & Ortuzar, 2000). In addition, the modification to the error term relaxes the independence assumption and allows for the violation of the independence of irrelevant alternatives (IIA) assumption that has been shown not to hold particularly in transportation (Mayberry, 1973) although it must be said that the estimation is done at the cost of more demanding computing resources (Borsch-Supan & Hajivassiliou, 1993).

3. Geographic area and dataset

The study covers the Pacific states of Washington, Oregon and California, henceforth referred to as the Pacific region. Our geographical area of interest excludes Alaska and Hawaii. The states of California, Oregon and Washington have an estimated combined population of 50.6 million as of July 2016 (United States Census Bureau/American FactFinder, 2013), representing about 15% of the U.S. population. The study used revealed preference data obtained from the 2015 ACS 1-year US Census PUMS dataset with 216,257 observations. This figure represents the total number of commuters who provided information on their travel mode. The travel mode options include individuals who drive solo; those who use public transit, those who carpool and individuals who work at home or within walking distance to home. Economic wellbeing is measured relative to the federal poverty level (FPL)that is state-wide and based on the number of individuals per household. The wellbeing measure uses an income to FPL ratio that is top coded at 501%. Number of children is restricted to respondents' own children that are less than 17 years. We have aggregated the time of day commute into broader time windows using a categorical variable and travel duration is censored at 200 minutes. The PUMS dataset does not have information on individuals' latent characteristics such as attitude or preferences or the cost associated with the travel mode. Variables related to household are not controlled for by household size.

4. Empirical strategy and model specification

The empirical strategy employs a dual pronged approach that includes the following:

- Locational variation using population estimates: Generating the population estimates of travel mode choices from the sample dataset using balanced repeated replication (BRR) approach to examine the variation in travel mode choice across the PUM areas and;
- Econometric Analysis: Determining the impact of potential covariates on travel model choice and inferring differences in travel mode choice across space and population segments using multinomial probit (MNP) models

Using the US Census Public Use Micro-data Sample dataset (United States Census Bureau , 2009-2011), we generated population estimates of commuters' transportation modes. In calculating the estimates, we employ two types of weights: person weight and replicate weights. The person weight is required for the point estimates and both person weight and the replicate weights are needed to calculate the standard errors. Given that the analysis is conducted on small population segmentation, Fay's variant of the Balanced Repeated Replication (BRR) method was utilized in calculating the standard errors (Fay, 1995). Fay's approach, called the Modified Half Sample (MHS), improves on the BRR by addressing the problem o perturbed weights and decreased sample size using an adjustment factor called the Fay coefficient. This coefficient was set to 0.5 for the PUMS data. Building on the sampling variance:

$$\hat{V} = \frac{\sum_{i=1}^{N}(X_i - X)^2}{N}$$

Fay's MHS variance equals:

$$\hat{V}_{mhs=} \frac{1}{(1-m)^2}\hat{V} = \frac{1}{(1-m)^2}\frac{\sum_{r=1}^{N}(X_r - X)^2}{N}$$

where $\boldsymbol{m}$ is the Fay's coefficient, $\boldsymbol{X_r}$ the replicate estimate and $\boldsymbol{X}$, the full sample estimate. Since the PUMS person records file has $\boldsymbol{N}$ = 80 replicates and $\boldsymbol{m}$=0.5, the expression above reduces to:

$$\hat{V}_{mhs=} \frac{\sum_{r=1}^{80}(X_r - X)^2}{20}$$

In determining the contributory factors that explain travel mode choice by commuters, we borrow from the typical discrete choice models (DCM) that are based on random utility theory where the utility is made up of a deterministic part and a random error component shown below:

$$V_{ij} = \beta X_i + \epsilon_{ij}$$

V_{ij} representing the indirect utility function and X_i representing a vector of the individual's characteristics. The individual will choose the travel mode for which V_{ij} is the highest. Therefore, the probability that an individual *i* chooses mode *j* will be:

$$P_{ij} = Pr(V_j > V_k)\forall j \neq k$$

$$= Pr(\beta_j X_i - \beta_k X_i) > \epsilon_{ik} - \epsilon_{ij}$$

The above notation represents a DCM in preference space and it constitutes the basic premise of the model with the pioneering work attributed to McFadden's (McFadden, 1974) seminal article. The appropriate estimation method to be used is typically informed by the statistical properties of the error term. Given a situation of more than two alternative choices, the estimation has often been carried out using multinomial logit regression (MNLR); a method that assumes error terms that are independently and identically distributed with a Gumbel distribution. This assumption is however problematic given that it precludes heterogeneity in scale or taste preference. In addition, a corollary of the zero covariance across the error terms assumption is the independence of irrelevant alternatives (IIA) assumption, a supposition that seems counter intuitive particularly for travel mode choices leading to biased and inconsistent estimates.

The empirical strategy takes into consideration these shortcomings by using multinomial probit (MNP) model in estimating the parameters of the DCM. However, as earlier argued, the coefficient cannot be estimated using a closed form analytic solution given the computation resources required particularly when the number of alternatives is greater than three. To guarantee that the numerical solution converges, a simulated maximum likelihood estimation is used, with the travel mode options categorized into four alternatives. This is achieved by assuming a utility function that is linear in parameters and error terms that are Gaussian and not independently distributed from one another. This modification means that the off-diagonal elements for the error terms variance-covariance structure are not constrained to take only zero values and thus are correlated.

This formulation allows for the incorporation of taste variation and provides for more plausible assumptions particularly for large sample size and superior estimates as argued by Munizaga et al (Munizaga, Heydecker, & Ortuzar, 2000). However, given the computational requirements, the coefficient estimates cannot be solved for using a closed form analytical solution. Thus, to guarantee that the numerical solution converges, a simulated maximum likelihood estimation (SMLE) method is employed (Geweke, Keane, & Runkle, 1997). Modifications to the algorithm and more powerful computing resources have gradually facilitated the use of the SMLE for real world transportation studies (Bolduc, 1999). To address the indeterminacy problem, we selected a reference group and evaluate the coefficients of the other travel modes relative to the reference travel option. This reduces the dimensionality of the variance-covariance matrix from a square matrix of four to three.

Instead of a multidimensional integral, to simplify the exposition and without loss of generalization, we assume a one dimensional integral where the support of the function is contained in $[j, k]$; then the expected value is the expression $\int_j^k \frac{1}{k-j} f(x)dx = E(f(x))$. The numerical equivalence from the SMLE, obtained by taking N random draws from the constant probability distribution, equals $\frac{1}{N}\sum_{i=1}^{N}(f(x_i)) \approx E(f(x))$. Appealing to the laws of large numbers (LLN), the value of the integral converges to $E\big(f(x)\big) * (k-j)$. By the same token, informed by the central limit theorem (CLT), the margins of error for the integral reduces with increase in the number of the draws with the standard deviation shrinking by the reciprocal of $\sqrt{N}$.

The SMLE provides computational feasibility in a manner theoretically consistent with the RUM and how it is implemented using DCM. More importantly, it reflects the realities of the situation being modeled by adequately reflecting taste variation and the violation of the IIA assumption. Beyond this, our approach improves on existing studies by creating rich commuting profiles for each population cohort, thanks to the microscopic census profiles. This is done by examining multiple dimensions of the variation using distinct segments of the population sampled with the objective of ascertaining the responsiveness of each of the subsets to changes in the explanatory variables. We subsequently generate marginal effects of the said explanatory variables not only at the mean but by graphing the probabilities of an individual selecting specific travel modes based on the control variable of interest. This presents a more visual and readable option and shows the differences among the subsets of the sampled population by stacking their travel mode profiles on the same graph.

We recognize the endogeneity between travel time and travel mode choice – for example, the choice of commuting by car determining the travel time and vice versa. We address this concern by reflecting travel time not at the individual level but at PUM area level using the average commuting travel time of all respondents for each PUM area irrespective of the travel mode. Not only does this This approach, not only exogenizes travel time in determining travel mode, it also controls for PUM area idiosyncratic attributes that we do not have data to explicitly capture. Finally, weighted sample survey data and local means for the population segments are used in the regression analysis.

5. Estimation results and discussion of findings

5.1. Locational variation analysis

The locational variation section utilized the BRR method in determining differences in commuting patterns across geographical areas. The PUM area constitutes the unit of analysis for the BRR approach with a total number of 352 PUMS units analyzed. The 2015 dataset, the most recent year used in the study, has approximately half a million observations from which we generated a total population estimate of 49.3 million for the three states. Of this figure, 28.2 million are in the workforce with 21.6 million coming from California, 4.2 million from Washington and the balance of 2.4 million coming from Oregon. A detailed breakdown of commuting patterns including vehicle occupancy for individuals that provided information on their travel mode is shown in the table 1.

Table 1. Population Estimates and the associated margins of error (in thousands).

Means of Transportation	California		Oregon		Washington	
	Pop Estimate	%	Pop Estimate	%	Pop Estimate	%
Drive solo	13,074 ± 32.3	73.9	1306 ± 10.8	70.9	2453 ± 14.1	72.6
Car pool	1755 ± 16.1	9.9	185.2 ± 5.86	10.0	319 ± 7.79	9.4
Transit	945.8 ± 13.6	5.3	91.2 ± 3.97	4.9	215.2 ± 6.12	6.4
Non-motorized	1,907 ± 13.8	10.8	261 ± 7.25	14.2	393.3 ± 8.02	11.6
Total	17,681 ± 31.7	100.0	1,843 ± 9.24	100.0	3,381 ± 13.5	100.0

We further document the differences in commuting patterns across more granular geographical areas by carrying out intra-state analyses using the PUM area as the enumeration unit. Of the 352 PUMAs analyzed, only four have less than 30% solo drivers relative to all commuters' ratio and all four are from San Francisco County. A more detailed analysis, looking solely at the seven PUMS area that make up the City of San Francisco shows that only 35% of the workforce drive to work unaccompanied compared to 73% for the three states analyzed.

Even more revealing is the fact that the average San Francisco resident is more than seven times (34.4% to 4.9%) likely to use the public transit system compared to a representative commuter across the three states. In addition, a higher percentage of San Franciscans (23.7%) either work close to home or use some form of non-motorized or bi-pedal travel option compared to 10.7% for the balance of the commuting population. Very little variation in carpooling is observed between the two cohorts with regards to carpooling probability.

Complementing this analysis are thematic maps of Los Angeles County and the study area based on the percentage of commuters that travel to work unaccompanied relative to either all vehicle commuters or all commuters though for brevity, only the latter result is shown in Figure 1. Merely eyeballing the maps in Figure 1 reveals that more densely populated and urban areas have lower percentages compared to the suburbs. However, there is appreciably more variability even across densely populated PUMS areas, revealing that population density does not explain all that was observed. This finding buttresses an earlier result that shows that of the three states analyzed, LA County has four of the five worst performing PUMS area when measured by the ratio of commuters driving alone as a fraction of all private vehicle users.

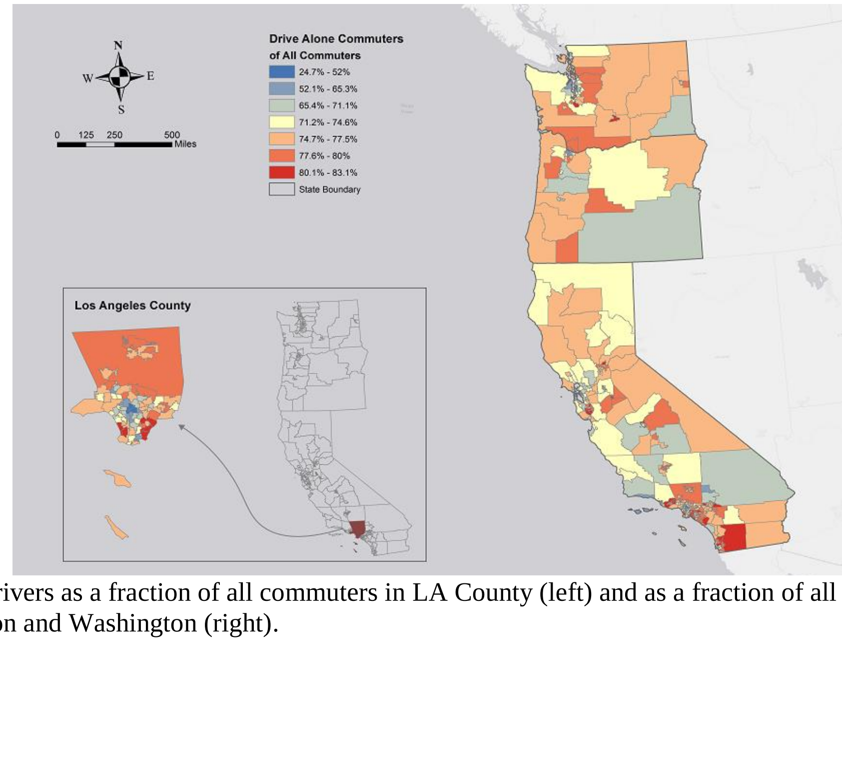

Figure 1. Solo Drivers as a fraction of all commuters in LA County (left) and as a fraction of all commuters in California, Oregon and Washington (right).

5.2. Econometric Analysis

The analysis is based on both the population and household records for the 2015 1-year US Census PUMS dataset. Variables employed for the analysis include the travel mode choice as the dependent variable and the following explanatory variables:

- Continuous and discrete variables: *age* – age of the respondent in years; *schl* – years of formal educations; *avtime* – average travel time to work in minutes; *povpip* – economic wellbeing, measured relative to the FPL; *hveh* – number of household vehicles.
- Dummy variables: *morn* – equaled 1 if the individual commutes to work in the morning and 0 otherwise; *kids* – equaled 1 if respondent has at least, one related kid below the age of 17 and 0 otherwise; *nevmar* – dummy variable that equaled 1 if respondent has never been married and 0 otherwise; *male* – the gender dummy that equaled 1 if the respondent is male and 0 otherwise; *Asian*, *Black* or *White* – dummies which equaled 1 if respondent reported belonging to any of these race/group and 0 otherwise.
- Interaction variables (including squared terms): squared terms for age (*agesq*), years of formal education (*schlsq*) and average travel duration to work (*avtimesq*)

5.3. Estimation results

Table 2 shows the coefficient estimates using multinomial probit with individuals who carpool as the excluded group. In the multinomial case, the coefficients obtained are measured relative to the subset of commuters who carpool, essentially treating each of the model evaluated as a probit regression of the model and the excluded group. The probit models the probability of a travel mode option being observed using a cumulative normal distribution function F(z) evaluated at $Z=X^T\beta$. Thus, a unit change in any of the regressors, say X_i changes the z-score of the dependent variable by β_i - the coefficient estimate.

For example, the z-score for the non-motorized option, relative to individuals who carpool, reduces by 0.275 for commuters from households with automobile. Relative to the excluded group, the effect of belonging to a household with an automobile is even more pronounced for the public transit travel mode where an estimated reduction of 0.395 in the z-score is observed. These changes are estimated with the other predictor variables held constant. Having children reduces the z-score for all the travel mode options relative to the excluded group and all the coefficients are significant at the 99% confidence interval for each parameter estimate provided. Individual who are never married exert a diametrically opposite effect – being single increases the z-score for all the travel mode options relative to individuals who carpool with the same order of magnitude and same level of significance.

Table 2. Regression coefficient estimates

	Non_motor b/se	Pub_Transit b/se	Drive_Solo b/se
agep	-0.011**	-0.000	0.027***
	(0.00)	(0.00)	(0.00)
schl	-0.051***	-0.061***	0.057***
	(0.01)	(0.01)	(0.01)
schlsq	0.003***	0.003***	-0.000
	(0.00)	(0.00)	(0.00)
avtime	0.047**	0.276***	0.054***
	(0.01)	(0.02)	(0.01)
avtimesq	-0.001***	-0.004***	-0.001***
	(0.00)	(0.00)	(0.00)
povpip	-0.000***	0.000	0.001***
	(0.00)	(0.00)	(0.00)
hveh	-0.275***	-0.395***	0.064***
	(0.01)	(0.01)	(0.01)
morn	-0.083***	0.129***	-0.043***
	(0.02)	(0.02)	(0.01)
kids	-0.305***	-0.375***	-0.147***
	(0.03)	(0.03)	(0.02)
Nevmar	0.335***	0.410***	0.107***
	(0.02)	(0.02)	(0.02)
Black	-0.042	0.148***	0.093**
	(0.05)	(0.04)	(0.03)
Asian	-0.152***	0.004	-0.137***
	(0.03)	(0.03)	(0.02)
White	0.159***	-0.030	0.166***
	(0.03)	(0.03)	(0.02)
male	0.128***	-0.107***	-0.022
	(0.02)	(0.02)	(0.01)
_cons	-0.267	-4.415***	-1.063***
	(0.24)	(0.29)	(0.17)
F	197		
df_r	200691		

* p<0.05, ** p<0.01, *** p<0.001
Base travel mode: Carpool

Equally revealing is the years of schooling continuous predictor variables when modeled with quadratic specification. Relative to individuals who carpool, the quadratic specification of education travel mode profile exhibits a convex relationship for both the non-motorized and public transit travel mode options though only a linear relationship survived for the drive solo mode choice. The convex relationship is an indication that at low levels of education, a decrease in z-score is observed for education while the change becomes positive at higher levels of education.

5.3.1. Examining marginal effect differences

To better illustrate the impact of regressors on the probability of exercising a travel mode, we used marginal effect to estimate the change a regressor has on the predicted probability. The computation shows the effect on travel mode probabilities given a marginal change in the covariate of interest while other explanatory variables are held fixed at the group's average with the result shown in Table 3. The table provides information along three dimensions; population of interest categorized by race, continuous predictor variables with explanatory power, and predicted probabilities of travel mode choice, for each population of interest. The predicted probabilities across the travel modes for each race sums up to 1. Irrespective of the population segment, more than three out of every four commuters drive unaccompanied. The marginal changes represent an increase of one standard deviation at the mean and all the estimated marginal changes are significant at the 99% confidence level.

A marginal increase in education attainment reduces the probability of choosing any of the travel mode options excerpt driving solo where more than a 0.05 increase in probability is reported irrespective of the commuter's race. A similar effect, though with smaller magnitudes is calculated for a standard deviation increase at the mean for income using the FPL figures. A preference for driving solo is also noticed with increase in age evaluated at the mean where probability increase exceeds 8% across the races.

Table 3. Marginal Effects

	Asian	**Black**	**White**
Probability (NON-MOTORIZED travel mode) Asian = 0.046, Black = 0.052, White = 0.058			
Age	-0.021	-0.025	-0.029
Income Effect	-0.006	-0.008	-0.008
Education Attainment	-0.017	-0.015	-0.022
Probability (PUBLIC-TRANSIT travel mode) Asian = 0.079, Black = 0.100, White = 0.047			
Age	-0.023	-0.030	-0.016
Income Effect	-0.004	-0.005	-0.003
Education Attainment	-0.031	-0.029	-0.019
Probability (CAR-POOL travel mode) Asian =0.122, Black = 0.097, White = 0.095			
Age	-0.036	-0.031	-0.033
Income Effect	-0.008	-0.007	-0.007
Education Attainment	-0.018	-0.010	-0.016
Probability (DRIVE-SOLO travel mode) Asian = 0.753, Black = 0.752, White = 0.800			
Age	0.081	0.086	0.077
Income Effect	0.018	0.020	0.018
Education Attainment	0.066	0.054	0.057

5.3.2. Commuting profile

We further show the differences across the population segments by examining the marginal effects not only at the mean as was done above, but by graphing the probabilities of an individual choosing a mode choice based on the control variable of interest and using the result to create granular commuting profiles. This presents a more readable option and shows the stark differences among the population segments by portraying on the same graph the responsiveness of each of the cohort to changes in the explanatory variables. Merely from looking at the gradient of the marginal effect functions, we can determine how responsive commuters are to choosing a mode choice as a result of a marginal change in an explanatory variable of interest. We explore multiple dimensions of this approach though for conciseness, we only show differences in the probability associated with the public transit travel mode based on race with respect to changes in formal years of education in Figure 2.

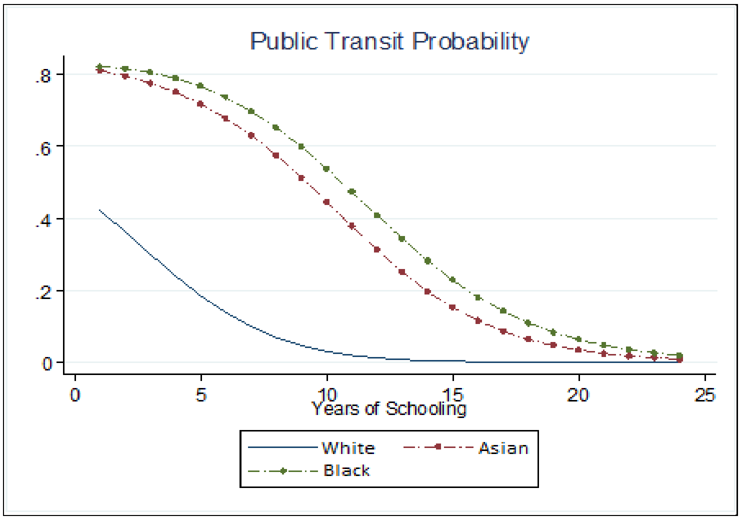

Figure 2. Commuting profile based on race with respect to years of formal education.

Figure 2 demonstrates that the commuting profile of Whites with respect to public transit deviates appreciably from Blacks or Asian commuters. The public transit probability profile shows that once the average White has the equivalent of a high school diploma, the probability of her taking the public transit travel mode becomes almost negligible. This finding corroborates an earlier referenced article (McKenzie, 2013) that showed that the differences in commuting behavior are pronounced based on analyses associated with the characteristics of different races. The article, for example, revealed that Non-Hispanic White workers drive to work solo at a rate almost 10% points higher compared to other racial groups though our predicted probabilities only support half of this magnitude.

6. Discussion of findings

The results from the study revealed that irrespective of the population segment, approximately three out of every four commuters drive solo. Within specific cohorts, appreciable variation in predicted probabilities for travel mode options were observed. Among Asians for example, where the predicted probability for carpooling is 12%, the predicted probabilities varies significantly based on the respondent's education level – ranging from 16% for 3 years of formal education to only 2% for individuals with 24 years of formal education. Finer segmentation is possible by creating smaller population segments by using interaction terms for the dummy variables.

A couple of observations are particularly worth mentioning. For one, the public transit probability profile shows that once an average White has the equivalent of a high school diploma, the probability of her taking the public transit travel mode becomes almost negligible. Secondly, our findings show that commuters are quite resistant to carpooling given the minimal locational variation observed in the percentage of commuters who carpool. However they may be amenable to using a public transit system where a functional one exists. As a policy tool, it will be more effective getting commuters who drive to work to shift to public transit than it is to get them to carpool. Presumably this has a lot to do with the decrease in flexibility that accompanies carpooling. For an individual who carpools, her home and home and work departure times are necessarily fixed, a constraint that is appreciably relaxed with the public transit given that she has multiple time-based options for leaving for work and home.

If we were to assume a shift in travel mode from the average commuter across the three states analyzed to the representative commuter who is resident in San Francisco, we estimate, for the three states analyzed, a reduction in excess of 10 million vehicles on the road on any working day, each of which will require a parking spot when it reaches its destination. The analysis was carried out using three forms of vehicle occupancy – individuals who drive alone; 2-person carpools and 3 or more persons carpools and subsequently applying the proportionate breakdown of the forms of transportation and vehicle occupancy to the relevant segment of the population – the representative San Francisco resident commuter and the average commuter for the balance of the population. The equivalent number of vehicles was obtained by dividing the number of vehicles by the occupancy number while taking a conservative view by assuming that maximum occupancy for any vehicle is three.

This finding shows that there is appreciable variability across geographical areas. We acknowledge that differences in numerous factors including the built environment, state and municipal transportation policies are relevant in shaping this outcome. Policy measures may particularly be relevant in shaping supply side issues – accessibility, quality and cost of the public transit systems, transportation infrastructure and land use management to mention just a few while some may be oriented towards the demand side. For example, San Francisco's experiment with SFpark (Pierce & Shoup, 2013), (Millard-Ball, Weinberger, & Hampshire, 2014), (Fabusuyi & Hampshire, 2018) was designed in part, with the objective of achieving a more societal friendly travel behavior by pricing parking spaces correctly and using the revenue to subsidize public transit.

7. Conclusions

The findings from our study have yielded crucial insights on the factors influencing travel mode choices individuals make including detailed commuting profiles associated with fine grained segments of the working population. the impact on their mode choices assuming these factors are varied, and we have shown how commuting behavior varies either across distinct population segments and/or across geographical areas. An advantage of the present study is the use of the PUMS dataset that allowed for the estimation of a disaggregated model which showcased how representative commuters, endowed with some characteristics, make travel mode choices. The theoretical basis utilized naturally lends itself to this approach in that each commuter is assumed to exercise a choice that maximizes his utility. The empirical strategy implements this by computing probabilities for each mode choice using an approach that maximizes the likelihood that the predicted travel option is actually the mode utilized in real life by the commuter.

Our results show that significant differences on travel mode options were observed across distinct population segments. The City of San Francisco provided the most socially optimal travel behavior. The total number of vehicles driven for any day of the week could be reduced by up to 10 million assuming the commuting patterns observed in San Francisco applied to the rest of the three states analyzed. While these findings are illuminating, our approach is not without limitations. The travel choices, for example, are assumed made with perfect information, which may not be in real life. New movers may not have the requisite information on the available commuting options and tenured commuters may exhibit their choices in a myopic manner – a plausible trait given that most of us are creatures of habit. Consequently, the differences in travel mode preferences we are observing may be influenced by asymmetric information.

References

Ben-Akiva, M., & Lerman, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* Boston: MIT Press.

Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., & Morikave, T. (1999). Integration of Choice and Latent Variable Models. MIT Working Paper.

Blumenberg, E., & Evans, A. (2010). Planning for Demographic Diversity: The Case of Immigrants and Public Transit. *Journal of Public Transportation*, *13*(2), pp. 23-45.

Blumenberg, E., & Shiki, K. (2007). Transportation Assimilation : Immigrants, Race and Ethnicity, and Mode Choice. *86th Annual Meeting of the Transportation Research Board*.

Blumenberg, E., & Smart, M. (2009). Travel in the'hood: Ethnic neighborhoods and mode choice. *88th Annual Meeting of the Transportation Research Board*.

Blumenberg, E., & Smart, M. (2010). Getting by with a little help from my friends… and family: immigrants and carpooling. *Transportation*, *37*(3), pp. 429-446.

Blumenberg, E., & Smart, M. (2014). Brother can you spare a ride? Carpooling in immigrant neighbourhoods. *Urban Studies*, *51*(9), pp. 1871-1890.

Bolduc, D. (1999). A Practical Technique to Estimate Multinomial Probit Models in Transportation. *Transportation Research Part B: Methodological*, *33*(1), pp. 63-79.

Borsch-Supan, A., & Hajivassiliou, V. (1993). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics, 58*, pp. 347-368.

Charles, K., & Kline, P. (2006). Relational Costs and the Production of Social Capital: Evidence from Carpooling. *The Economic Journal, 116*(511), pp. 581-604.

Cortright, J. (2015, February). *Surging City Center Jobs.* Retrieved February 18, 2015, from Cityobservatory.org: http://cityobservatory.org/wp-content/uploads/2015/02/Surging-City-Center-Jobs.pdf

Deal, B., Kim, J., & Pallathucheril, V. (2009). Growth Management and Sustainable Transport: Do Growth Management Policies Promote Transit Use? *Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington D.C.*

Donnelly, R., Erhardt, G., Moeckel, R., & Davidson, W. (2010). NCHRP Synthesis 406: Advanced Practices in Travel Forecasting. *Transportation Research Board of the National Academies, Washington D.C.*

Fabusuyi, T., & Hampshire, R. C. (2017). The Mode Most Traveled: Transportation Infrastructure Implications and Policy Responses. In S. Geertman, A. Allan, C. Pettit, & S. J., *Planning Support Science for Smarter Urban Futures* (pp. 279-295). Springer, Cham.

Fabusuyi, T., & Hampshire, R. C. (2018). Rethinking performance based parking: A Case Study of SFpark. *Transportation Research Part A: Policy and Practice, 115,* 90-101.

Fabusuyi, T., Hampshire, R., Hill, V., & Sasanuma, K. (2014). Decision Analytics for Parking Availability in Downtown Pittsburgh. *Interfaces, 44*(3).

Fay, R. (1995). *VPLX: Variance Estimation for Complex Surveys, Program Documentation.* Bureau of US Census.

Frey, W. (Washington D.C., 2010). Population and Migration. *State of Metropolitan America - On the Frontlines of Demographic Transformation*.

Garrido, R., & Mahmassani, H. (2000). Forecasting freight transportation demand with the sapce-time multinomial probit model. *Transportation Research Part B: Methodological, 34*(5), pp. 403-418.

Geweke, J. F., Keane, M. P., & Runkle, D. E. (1997). Statistical Inference in the Multinomial Multiperiod Probit Model. *Journal of Econometrics, 80*(1).

Giuliano, G. (2003). Travel, location and race/ethnicity. *Transportation Research Part A: Policy and Practice, 37*(4), pp. 351-372.

Haas, P., Makarewicz, C., Benedict, A., Sanchez, T., & Dawkins, C. (July 2006). Housing and Transportatioin Cost Trade-Offs and Burdens of Working Households in 28 Metros, Center for Neighborhood Technology, Chicago, IL.

Hu, L., & Giuliano, G. (Jan 23-27, 2011). Beyond the Inner City: A New Form of Spatial Mismatch. *Proceedings of the 90th Annual Meeting of the Transportation Research Board*.

Huang, H., Yang, H., & Bell, M. (March 2000). The models and economics of carpools. *Annals of Regional Science, 34*(1), pp. 55-68.

Kim, S. (2009). Immigrants and Transportation: An Analysis of Immigrant Workers' Work Trips. *Cityscape: A Journal of Policy Development and Research, 11*(3).

Krizek, K., Johnson, P., & Tilahun, N. (Nov 18-20, 2004). Gender Differences in Bicycling Behavior and Facility Preferences. *Conference Proceedings 35 Research on Women's Issues in Transportation*.

Manski, C. F., & Lerman, S. R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica: Journal of the Econometric Society, 45*(8), pp. 1977-1988.

Mayberry, J. (1973). Structural requirements for abstract-mode models of passenger transportation. In R. Quandt, *The Demand for Travel: Theory and Measurement.* Lexington, Mass: D.C. Heath and Co.

McFadden, D. (1974). The Measurement of Urban Travel Demand. *Journal of Public Economics, 3*(4), pp. 303–328.

McKenzie, B. (2013). *Who Drives to Work? Commuting by Automobile in the U.S. in 2013.* American Community Survey Report, ACS-32 Bureau of US Census, 2015.

McKenzie, B. (2013). *Who Drives to Work? Commuting by Automobile in the U.S. in 2013.* American Community Survey Report, ACS-32 Bureau of US Census, 2015.

Millard-Ball, A., Weinberger, R., & Hampshire, R. C. (2014). Is the curb 80% full or 20% empty? Assessing the impacts of San Francisco's parking pricing experiment. *Transportation Research Part A, 63*, pp. 76-92.

Munizaga, M., Heydecker, B., & Ortuzar, J. (2000). Representation of Heteroskedasticity in Discrete Choice Models. *Transportation Research Part B, 34*(3), pp. 219-240.

Pierce, G., & Shoup, D. (2013). Getting the Prices Right. *Journal of the American Planning Association, 79*(1), pp. 67-81.

Rosenbloom, S., & Burns, E. (1993). Gender differences in commuter travel in Tucson: Implications for travel demand management programs. *Transportation Research Record of the National Academies*, pp. 82-90.

Rosenbloom, S., & Burns, E. (1994). Why working women drive alone: Implications for travel reduction programs. *Transportation Research Board*, pp. 39-45.

Thakuriah, V., Sriraj, P., Soot, S., Liao, Y., & Berman, G. (2005). Activity and Travel Changes of Job Access Transportation Service Users: Analysis of a User Survey. *Transportation Resarch Record: Journal of the Transportation Research Board*, pp. 55-62.

United Kingdom Department of Transport. (2008). *2008 Annual Report.* Retrieved March 28, 2015, from www.gov.uk/government/publications/department-for-transport-annual-report-2008

United States Census Bureau . (2009-2011). *Public Use Microdata Sample (PUMS) Dataset*. Retrieved May 13, 2013, from http://www.census.gov/acs/www/data_documentation/pums_data/

United States Census Bureau/American FactFinder. (2013). *American Community Survey (ACS).* Retrieved March 28, 2015, from B01003: Total Population. 2009-2013: http://factfinder2.census.gov

Weinberger, R. (2007). Men, Women, Job Sprawl and Journey to Work in the Philadelphia Region. *Public Works Management and Policy, 11*(3), pp. 177–193.