Using 1990 Census Public Use Microdata Sample To Estimate Demographic and Automobile Ownership Models

CHARLES L. PURVIS

Disaggregate (household-level) automobile ownership choice models are typically estimated by using large-scale cross-sectional household travel surveys. Automobile ownership choice models typically stratify households into households owning zero, one, or two or more vehicles. This automobile ownership market segmentation is critical in the application of a regional set of disaggregate travel demand models for aggregate forecasting purposes. An alternative regional data set for estimating disaggregate automobile ownership choice models is the 1990 Census Public Use Microdata Sample (PUMS). PUMS consists of two disaggregate files of individual 1990 census records (household and population characteristics) of either 1 percent of an area's households or 5 percent of an area's households (the 1 percent and the 5 percent samples). Disaggregate workers in household and automobile ownership choice (logit) models were estimated on the basis of PUMS data files for the nine-county San Francisco Bay Area and the one-county San Diego region. These models were also compared with disaggregate models on the basis of the 1990 Metropolitan Transportation Commission household travel survey. The strengths and weaknesses of both approaches—PUMS versus household travel surveys—are discussed. The primary weakness of PUMS is the lack of data on neighborhood characteristics, such as land use density or accessibility measures, at a fine enough geographic level (i.e., regional travel analysis zone) for model estimation purposes. The transferability of the model estimation methodology to other metropolitan regions is discussed.

The purpose of this paper is to explore the development of demographic and automobile ownership forecasting models by using data from the 1990 U.S. decennial census and from household travel surveys. Disaggregate (household-level) automobile ownership choice models were estimated by using data from the 1990 census Public Use Microdata Sample (PUMS) and the 1990 San Francisco Bay Area Household Travel Survey. Comparison of the model estimation results from the two data sets shows that the 1990 census PUMS is an appropriate data set for use in updating metropolitan automobile ownership models. The development of PUMS-based automobile ownership models may be appropriate in metropolitan areas and states where current household travel survey data are not readily available.

TECHNIQUES FOR FORECASTING AUTOMOBILE OWNERSHIP

Travel demand forecasting techniques have typically focused on the four-step planning models related to trip frequency choice, destination choice, mode choice, and route choice. Much less attention is typically paid to the development and evaluation of demographic models that feed data into travel demand models. These demographic models include automobile ownership models, labor force participation rate models, household income models, and age cohort survival models. The focus in this paper is on automobile ownership models, although the point to be made is that the other sets of demographic models are of no less importance. The development of robust and credible labor force participation rate models, household income models, and so on is key to successful urban land use, economic, and transportation stimulation modeling. Demographic and other inputs to travel demand modeling have been covered by Hamburg et al. (*1*) and Bajpai (*2*).

Why Is Forecasting Automobile Ownership Important?

Good forecasts of automobile ownership levels are critical in preparing adequate travel demand forecasts. Automobile ownership variables are typically encountered in most travel demand model components, including trip frequency choice, destination choice, and mode choice models.

In terms of trip frequency (trip generation) models, households with no vehicles available take markedly fewer trips than households with one or more vehicles available. Cross-classification or linear regression trip generation (home-based production) models typically include automobile ownership as one of the independent variables used to predict trip frequency choice.

Variables such as the number of automobiles per household, the number of automobiles per worker, and the number of automobiles per licensed drivers have all been used successfully in most if not all work and nonwork mode choice model specifications. Automobile ownership level is less likely to be used in trip destination choice (trip distribution) models, although nested, destination mode choice models invariably include an automobile ownership variable as an independent variable in the mode choice utility.

Understanding of the numbers of automobiles owned or available to a household and household members is critical in defining the captive market and market choice behavior. Households with no automobiles available will be captive to transit, ride-sharing with nonhousehold members, or nonmotorized means of transportation. Households with multiple workers or drivers per household and only one vehicle per household face a partial captivity—which worker (or driver) gets the family car? Households with one or more cars available per licensed driver and faced with infrequent or inaccessible transit services may essentially be cap-

Metropolitan Transportation Commission, 101 Eighth Street, Oakland, Calif. 94607-4700.

tive or forced to use their automobile because of the lack of alternatives.

Underpredicting future automobile ownership levels will have the effect of underpredicting total motorized person trips, perhaps unpredicting average person trip lengths, overpredicting transit patronage levels, and underpredicting congestion, traffic, and air quality emissions. With these considerations in mind it seems important to get the automobile ownership forecasts right rather than assuming no change in automobile ownership levels with respect to base year automobile ownership levels. The "null model" automobile ownership model (i.e., assuming no change from base year automobile ownership levels) may prove to be an undesirable characteristic of future travel demand model forecasting systems.

Aggregate Versus Disaggregate Automobile Ownership Forecasting Models

Simply stated, aggregate automobile ownership forecasting models are estimated on the basis of areawide time series data on automobile ownership per capita or per household and various independent variables; disaggregate automobile ownership forecasting models are statistically estimated on the basis of household-level data and typically stratify households into households by the number of automobiles available (e.g., zero, one, two, or three or more automobiles available). Disaggregate automobile ownership models could also be linear regression in mathematical form and would predict the number of automobiles per household, the number of automobiles per capita, the number of automobiles per licensed driver, or the number of automobiles per worker.

Aggregate automobile ownership models can also be estimated by using aggregate zone-level statistics from decennial census data such as the 1980 census Urban Transportation Planning Package (UTPP) or the 1990 census Transportation Planning Package (CTPP). Pearson (*3*) discusses aggregate automobile ownership models estimated on zone-level data from the 1980 UTPP. Good discussions on aggregate automobile ownership models are included in a publication of the Organization for Economic Cooperation and Development (*4*). Other relevant discussions on automobile ownership trends and saturation levels are included in reports by Lave (*5*) and Pisarski (*6*).

The 1960s state of the practice in disaggregate automobile ownership models is best described by Deutschman (*7*). These are typically linear regression models predicting automobile ownership rates: the number of automobiles per household or the number of automobiles per capita. Independent variables include average household size, mean or median household income (or log transformations of income), residential density, and single-family versus multifamily dwelling units. Independent variables not analyzed by Deutschman included the numbers of workers in the household and the relative transit accessibility of the residence area with respect to working and shopping opportunities.

Disaggregate automobile ownership rate models (typically linear regression models) can be contrasted with disaggregate automobile ownership level models (typically cross-classification or multinomial logit models). The former predict the number of automobiles per household or the number of automobiles per capita; the latter stratify households by the number of automobiles (or vehicles) owned (or available), say, into categories of zero-vehicle, one-vehicle, and two-or-more vehicle households.

Current examples of market-segmented automobile ownership rate models are provided by Prevedouros and Schofer (*8*). They provide some good exploratory research that may prove to be useful in the formulation of operational, practice-oriented automobile ownership models.

Cross-Classification Automobile Ownership Models

A good example of a cross-classification automobile ownership model is the 1982 version of the Honolulu metropolitan area model (*9*). The dependent variable is the number of households stratified by three vehicle ownership levels (zero, one, or two or more vehicles per household). Three independent variables are used in the final Honolulu model specification: households by household size (four groups), households by income level (three groups), and households by geographic area type (three groups). Each of the 36 cells in the cross-classification matrix is assigned three values to split out the shares of households with zero, one, and two or more vehicles. Two other independent variables were examined in the Honolulu analysis: households by number of workers in the household and housing type (single-family versus multifamily units). These two variables were not included in the final model specification, basically to keep the cross-classification model tractable to users. An independent variable not examined in Honolulu included a transit accessibility variable, although one could argue that the area type stratification is perhaps a suitable surrogate for generalized transit accessibility.

Disaggregate Choice Models for Automobile Ownership

Theoretical developments in travel behavior modeling led to the incorporation of nested multinomial logit models to represent automobile ownership choice as a distinct yet integrated element of a "mobility block" of travel demand models [see Lerman (*10*) and Lerman and Ben-Akiva (*11*)]. Lerman and Ben-Akiva critiqued the 1970s state of the practice of automobile ownership forecasting as being a "side calculation made with simple models that rely on trend extrapolations or correlations made between 1 and 2 variables and car ownership rather than on a strong causal theory." (A comprehensive review of metropolitan area forecasting models may unfortunately reveal that automobile ownership forecasting is still treated as a "side calculation.")

Two examples of multinomial logit automobile ownership models, in practice, are the Portland, Oregon (*12,13*), and Bay Area (*14,15*) travel demand models. Both the Portland and the Bay Area models include a series of mobility block models that first predict the distribution of households by the number of workers in households and then second predict the distribution of households by the number of vehicles in the household. Both the Portland and Bay Area model sets use multinomial logit model specifications to predict the number of workers in households and automobile ownership choice.

The Portland workers-in-household model includes four alternative choices: zero-worker, one-worker, two-worker, or three-or-more-worker households. The utility equations use household size, four income categories, and four categories for age of head of household as independent variables. The Bay Area nonworking household (NWHH) model is a binomial logit model that splits households into households without workers and households with

workers. The independent variables included in the Bay Area NWHH model include household size, household income, and special variables to indicate very low income households and low numbers of people per household.

The Portland household automobile ownership model includes four alternative choices: zero-vehicle, one-vehicle, two-vehicle, and three-or-more-vehicle households. Independent variables include the number of households by four household size categories, the number of households by four workers in household categories, the number of households by four income categories, and a generalized transit accessibility variable. This last variable is an "average zonal value of employment accessible within 30 minutes total travel time by transit." Recent revisions to the Portland automobile ownership models, done as part of the 1000 Friends of Oregon Land Use Transportation Air Quality Study, added two variables: the number of retail employees working within 1 mi of the zone of residence and a "pedestrian environment factor." This last factor is essentially a score assigned to each regional travel analysis zone describing the topography, sidewalk continuity, local street pattern, and ease of crossing streets within each zone. These urban form variables—employment accessibility and the pedestrian score—help in explaining the lower automobile ownership levels in the central Portland neighborhoods.

The Bay Area has two automobile ownership models—a nonworking household automobile ownership (NWHHAO) model and a working household automobile ownership (WHHAO) model. Both Bay Area automobile ownership models split households into the number of households with zero, one, or two or more vehicles available. Independent variables in the current Bay Area NWHHAO model include average household size, average household income, and population density. (The original NWHHAO model included a log sum-based off-peak transit accessibility variable in the model specification). Independent variables in the current Bay Area WHHAO model include average household size, average household income, single-family dwelling unit dummy variable, employment density, and a log sum-based peak transit accessibility variable (essentially a ratio of the exponentiated transit and automobile utilities from the work trip mode choice model). For aggregate model application the Bay Area models are applied to zone-level households market segmented (split) by three household income levels.

The output of the Portland set of worker/automobile ownership choice models is a prediction of the number of households in a travel analysis zone by income groups (four groups), household size (four groups), age of head of household (four groups), numbers of workers in the household (four groups), and number of vehicles available in household (four groups), or essentially up to 1,024 potential market segmentations per zone. The output of the Bay Area set of worker/automobile ownership choice models is a prediction of the number of households in a travel analysis zone by income groups (three groups), number of workers in household (two groups), and number of vehicles available in the household (three groups), or essentially 18 market segmentations per zone. Some of these market segments are likely to be very small in magnitude (e.g., high income, working households with no vehicles available) if not excluded as a potential alternative choice (e.g., three workers in a two-person household).

What are the pros and cons of cross-classification automobile ownership models versus logit choice automobile ownership models? The positive aspects of cross-classification automobile ownership models are their tractability; their ease of specification, estimation, and application; and their ability to satisfactorily handle the highly nonlinear relationships between household income and automobile ownership and between household size and automobile ownership. Readily available data sources for the estimation of cross-classification automobile ownership models include standard census products such as the 1990 CTPP and the 1990 census PUMS. Household travel surveys can also be used for estimating these cross-classification models.

The negative aspects of cross-classification models include a practical (tractable) limitation to two or three independent variables, and aggregation errors related to grouping of what can be considered continuous variables such as household income or residential density. For example a 5 percent increase in mean or median household income in a low-income cohort has no impact on automobile ownership levels in the context of a standard cross-classification automobile ownership model application. In areas with cross-classification automobile ownership models, changes in labor force participation rates, major transit capital investments, or increased development of mixed-use developments and multifamily dwelling units have no impact on automobile ownership forecasts. An alternative to the standard two- or three-dimensional cross-classification model is a more complex—and less tractable—cross-classification automobile ownership model that could contain four or five independent variables, say, household size, household income level, the number of workers in the household, dwelling unit structure type, and area type or "accessibility class." Large data sets such as the PUMS are ideal for this sort of cross-classification model.

Positive aspects of logit choice automobile ownership models include tractability, ease of estimation and application, and ability to incorporate many of the independent variables that might influence automobile ownership choice. Independent variables that have been included in logit automobile ownership choice models include household size, household income, the number of workers in the household, structure type, employment density and accessibility, transit accessibility to employment, combined transit and highway impedance, population density, and urban design factors. Household travel surveys are the traditional data sources for the estimation of logit choice automobile ownership models. This paper explores the use of the 1990 census PUMS in estimating simple worker/automobile logit choice models.

Negative aspects of logit choice automobile ownership models include the challenges related to model specification, especially with respect to the treatment of the nonlinear relationships between several significant independent variables (e.g., household income and household size) and the dependent variable (the number of households by automobile ownership level). In general logit choice models are less satisfactory in addressing these nonlinear relationships than cross-classification models. In comparison with cross-classification models, logit choice automobile ownership models can be structured to be sensitive to such issues as changes in labor force participation rates, major transit capital investments, and increases in mixed-use land use patterns and multifamily dwelling units.

PUMS AND ITS USE IN TRANSPORTATION PLANNING ANALYSIS

PUMS is a standard Bureau of the Census data product that was first introduced in 1960. The 1990 census version of the microdata sample includes what are called the 1 percent sample and the 5

percent sample as well as a sample of households with elderly householders (*16*). The PUMS data are basically individual census records for a sample of households and people who answered the census "long form." For example in a region the size of the Bay Area, with 2.246 million households and 6.024 million people, the 5 percent PUMS file for the Bay Area includes disaggregate records on 108,491 households and 292,451 people. This amounts to 4.8 percent of the households and 4.9 percent of the total Bay Area population in 1990.

The smallest geographic area for which PUMS data are available is at the Public Use Microdata Area (PUMA). PUMAs may not be less than 100,000 people in total population in 1990. This large geographic restriction protects the confidentiality of census respondents by *not* providing precise enough geographic information with which to locate and identify the individual respondent. In 1991 the boundaries of the 1990 census PUMAs were defined by regional census data center staffs as part of the state census data center program. In the nine-county Bay Area, 48 PUMAs each with an average population of 125,000 people were defined.

The PUMS household records include all housing unit data from the 1990 census long form plus recoded variables such as the number of people in the family and the presence of people age 65 years and over. "Allocation" flag variables are included to denote if data values were imputed or allocated by the Census Bureau.

The PUMS person records include person information from the 1990 census long form as well as recoded variables (e.g., recode of place of birth and recode of person's total earnings) and allocation (imputation) flags.

Bay Area transportation planners required 1980 census PUMS data for market segmentation adjustments in the aggregate application of disaggregate choice models (*17*). Conversion factors were derived from 1980 census PUMS data to convert demographic characteristics of total households into characteristics of households with workers. For example adjustments are needed for four sets of demographic variables included in the Bay Area regional work trip mode choice model TW:

- Income per working household/income per total household;
- Number of people per working household/number of people per total household;
- Number of workers per working household/number of workers per total household; and
- Number of automobiles per working household/number of automobiles per total household.

Income per working household was 14 percent higher than income per total household, according to the 1980 census PUMS for the Bay Area. Household size in working households was 8 percent higher than household size in total households. The number of workers per working households was 26 percent higher than the number of workers per total household, and the number of automobiles per working household was 11 percent higher than the number of automobiles per total household.

Other PUMS data were used as supplementary data inputs to the Bay Area travel model system to adjust demographic inputs by market segment, namely households by three automobile ownership levels, households by three income levels, and working versus nonworking households.

Summary tabulations for the 1990 census PUMS for the San Francisco Bay Area and the San Diego region are included in Tables 1 through 3, which show the number and characteristics of households by three household income levels and three automobile ownership levels, stratified by total households, working households, and nonworking households, respectively. Information extracted from the 1990 census PUMS is critical for market segmentation adjustments in travel forecasting model systems. These data are also used for the aggregate validation of the workers-in-household model.

Data from the 1990 census PUMS can be charted to show nonlinear relationships between the share of the region's households with no workers in comparison with household income, household size, and age of head of household (Figures 1 through 3, respectively). A graphical exploratory analysis of these demographic relationships assists the model developer in setting up model specifications to properly treat the nonlinear relationships that may appear. Households with a 1989 mean household income of less than $40,000 have a much higher likelihood of having no workers. One- and two-person households also have a higher likelihood of having no workers. As the age of the head of the household approaches and exceeds 60 years, the likelihood that the household has no workers present increases dramatically.

TABLE 1 Characteristics of Total Households in Bay Area and San Diego 1990 Census PUMS

Total Households by Three Household Income Levels

	Number and Share of Households				Veh/HH	Veh/HH	P/HH	P/HH	Emp/HH	Emp/HH
	Bay Area	%	San Diego	%	Bay Area	San Diego	Bay Area	San Diego	Bay Area	San Diego
Low Income	783,977	35.0%	378,160	42.7%	1.092	1.236	2.060	2.299	0.737	0.860
Medium Income	791,942	35.3%	316,911	35.8%	1.842	1.995	2.708	2.879	1.482	1.541
High Income	666,635	29.7%	190,503	21.5%	2.439	2.510	3.140	3.171	1.962	1.891
Total	2,242,554	100.0%	885,574	100.0%	1.757	1.782	2.610	2.694	1.364	1.326

Note: "Low Income" = less than $30,000. "Medium Income" = $30,000 - $60,000. "High Income" = greater than $60,000.

Total Households by Three Vehicle Ownership Levels

	Number and Share of Households				Income	Income	P/HH	P/HH	Emp/HH	Emp/HH
	Bay Area	%	San Diego	%	Bay Area	San Diego	Bay Area	San Diego	Bay Area	San Diego
Zero-Vehicle	235,568	10.5%	71,585	8.1%	$19,299	$15,446	1.873	2.117	0.588	0.532
One-Vehicle	729,040	32.5%	296,776	33.5%	$33,379	$27,601	1.932	2.039	0.929	0.863
Two-plus Vehicles	1,277,946	57.0%	517,213	58.4%	$64,603	$54,594	3.132	3.150	1.755	1.701
Total	2,242,554	100.0%	885,574	100.0%	$49,693	$42,384	2.610	2.694	1.364	1.326

TABLE 2 Characteristics of Working Households in Bay Area and San Diego 1990 Census PUMS

Working Households by Three Household Income Levels

	Number and Share of Households				Veh/HH	Veh/HH	P/HH	P/HH	Emp/HH	Emp/HH
	Bay Area	%	San Diego	%	Bay Area	San Diego	Bay Area	San Diego	Bay Area	San Diego
Low Income	441,614	24.8%	235,687	34.2%	1.292	1.424	2.344	2.620	1.308	1.380
Medium Income	707,365	39.7%	278,915	40.4%	1.885	2.052	2.808	3.003	1.659	1.751
High Income	632,065	35.5%	175,180	25.4%	2.483	2.576	3.210	3.279	2.069	2.057
Total	1,781,044	100.0%	689,782	100.0%	1.950	1.971	2.836	2.942	1.717	1.702

Note: "Low Income" = less than $30,000. "Medium Income" = $30,000 - $60,000. "High Income" = greater than $60,000.

Working Households by Three Vehicle Ownership Levels

	Number and Share of Households				Income	Income	P/HH	P/HH	Emp/HH	Emp/HH
	Bay Area	%	San Diego	%	Bay Area	San Diego	Bay Area	San Diego	Bay Area	San Diego
Zero-Vehicle	100,411	5.6%	27,052	3.9%	$28,029	$21,706	2.277	2.758	1.380	1.408
One-Vehicle	517,763	29.1%	200,106	29.0%	$37,835	$30,221	2.058	2.222	1.308	1.280
Two-plus Vehicles	1,162,870	65.3%	462,624	67.1%	$67,278	$56,498	3.231	3.265	1.929	1.902
Total	1,781,044	100.0%	689,782	100.0%	$56,506	$42,384	2.836	2.942	1.717	1.702

It is anticipated that transportation planners of the 1990s will use the 1990 census PUMS for various policy and planning analyses, including the following:

- Describing the characteristics of commuters in corridors targeted for congestion pricing programs or major transit or highway capital investments;
- Describing the characteristics of commuting submarkets, including carpoolers, transit passengers, people who work at home, bicycle commuters, disabled people, elderly people and so on;
- Analyzing market segmentation for travel demand forecasting models; and
- Describing the commuting habits by household life cycle stage, occupation of worker, industry, educational attainment, and so on.

The PUMS files are treasure chests of disaggregate household, person, and commuter characteristics that are waiting to be mined by adventurous transportation planners and policy analysts. Needed are case studies to explore conventional and nonconventional ways of using PUMS data to advance transportation planning practice.

MODEL ESTIMATION RESULTS FOR BAY AREA AND SAN DIEGO: 1990 CENSUS PUMS AND 1990 BAY AREA TRAVEL SURVEY

Three different choice models were estimated in this research:

- Nonworking household (NWHH) model,
- Nonworking household automobile ownership (NWHHAO) model, and
- Working household automobile ownership (WHHAO) model.

The NWHH model is a binomial logit choice model predicting whether a household has zero or one or more workers. The NWHHAO model is a multinomial logit choice model further splitting nonworking households into households with zero, one, or two or more vehicles available. The WHHAO model is also a multi-

TABLE 3 Characteristics of Nonworking Households in Bay Area and San Diego 1990 Census PUMS

Non-Working Households by Three Household Income Levels

	Number and Share of Households				Veh/HH	Veh/HH	P/HH	P/HH	Emp/HH	Emp/HH
	Bay Area	%	San Diego	%	Bay Area	San Diego	Bay Area	San Diego	Bay Area	San Diego
Low Income	342,363	74.2%	142,473	72.8%	0.835	0.924	1.693	1.768	0.000	0.000
Medium Income	84,577	18.3%	37,996	19.4%	1.480	1.578	1.865	1.970	0.000	0.000
High Income	34,570	7.5%	15,323	7.8%	1.636	1.755	1.852	1.940	0.000	0.000
Total	461,510	100.0%	195,792	100.0%	1.013	1.116	1.737	1.820	0.000	0.000

Note: "Low Income" = less than $30,000. "Medium Income" = $30,000 - $60,000. "High Income" = greater than $60,000.

Non-Working Households by Three Vehicle Ownership Levels

	Number and Share of Households				Income	Income	P/HH	P/HH	Emp/HH	Emp/HH
	Bay Area	%	San Diego	%	Bay Area	San Diego	Bay Area	San Diego	Bay Area	San Diego
Zero-Vehicle	135,157	29.3%	44,533	22.7%	$12,814	$11,644	1.573	1.729	0.000	0.000
One-Vehicle	211,277	45.8%	96,670	49.4%	$22,458	$22,178	1.624	1.660	0.000	0.000
Two-plus Vehicles	115,076	24.9%	54,589	27.9%	$37,566	$38,458	2.136	2.178	0.000	0.000
Total	461,510	100.0%	195,792	100.0%	$23,401	$24,321	1.737	1.820	0.000	0.000

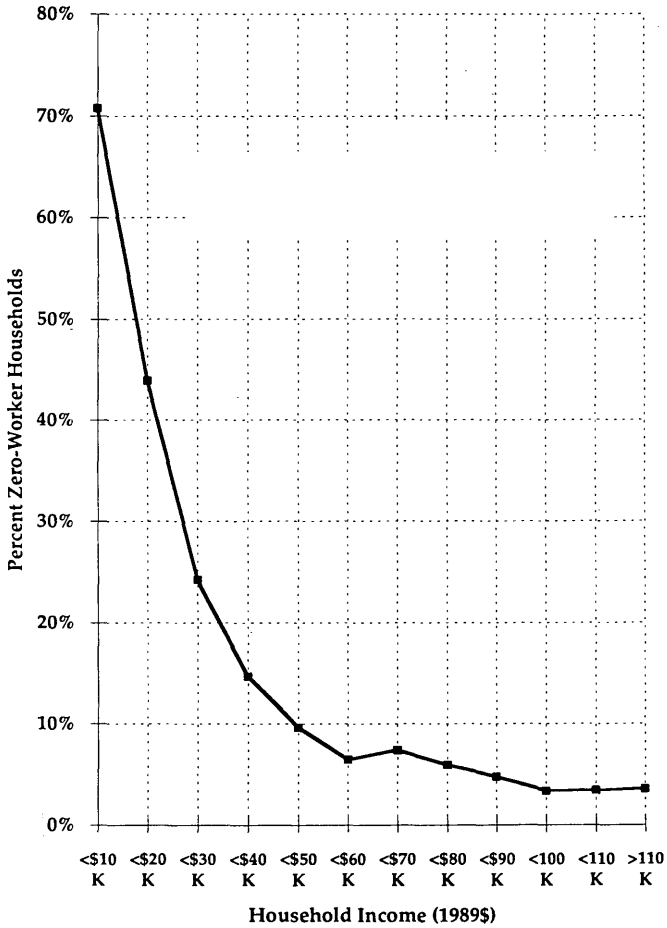

FIGURE 1 Percent zero-worker households by household income ranges, 1989 dollars, 1990 census PUMS, San Francisco Bay Area.

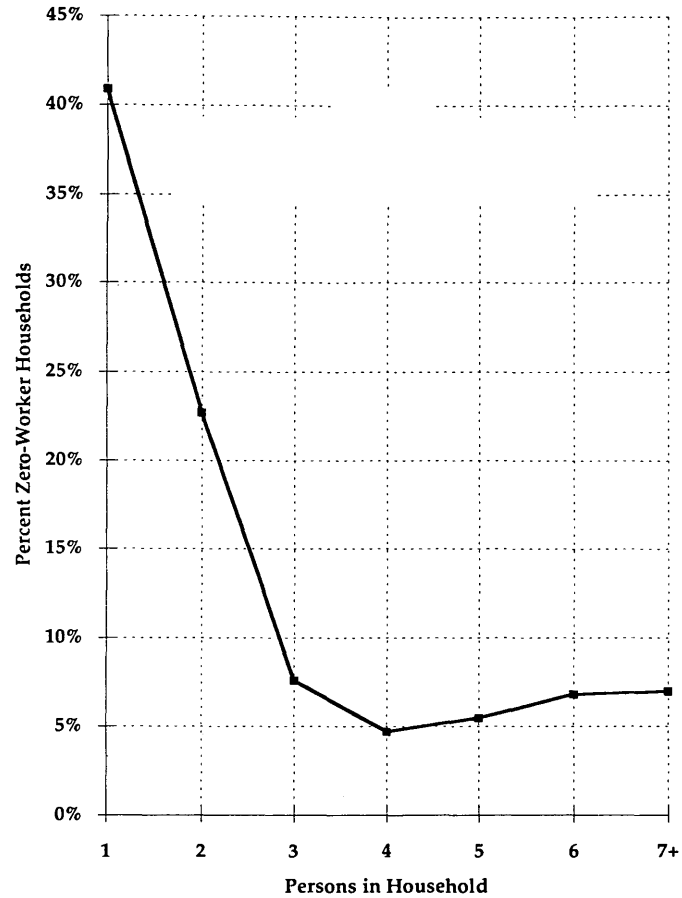

FIGURE 2 Percent zero-worker households by number of people in household, 1990 census PUMS, San Francisco Bay Area.

nomial logit choice model that splits households with workers into households with zero, one, or two or more vehicles available (Figure 4). The models as estimated are similar to previous versions of Bay Area travel demand models, with simplifications and enhancements as noted.

Six to 10 model specifications were tested for each component model. Only the best model is reported here for the sake of brevity. Three data sets were used in the research project:

- 1990 Bay Area Household Travel Survey,
- 1990 census PUMS 5 percent sample for the Bay Area, and
- 1990 census PUMS 5 percent sample for the San Diego region.

The 1990 Bay Area Household Travel Survey was a telephone-based trip diary survey of 10,838 households conducted during the spring and fall of 1990. Households that refused or did not answer the household income question on the 1990 survey (approximately 30 percent of survey respondents) were excluded from the model calibration file. The 5 percent PUMS for the nine-county Bay Area contains 108,491 household records. The 5 percent PUMS for the one-county San Diego region contains 41,987 household records. All PUMS records, including households for which income was imputed, were included in the model specification tests.

Commercially available software was used for preprocessing the rather immense PUMS data sets on a mainframe computer before downloading the calibration files to a microcomputer. Logit models were estimated by using a commercially available logit estimation package.

All of the coefficients were reviewed for reasonableness in terms of coefficient magnitude and sign. All of the *t*-statistics for all of the coefficients in the summary tables are significant (>2.0) although some of the coefficient signs are counterintuitive. For the present purposes a magnitude of less than a 10-fold difference in model coefficients between data sets is considered a reasonably consistent result (i.e., in the same "ballpark").

The nonworking household model was one of the more difficult models to estimate, although the final rho-bar squared statistics (>0.40) are acceptable for disaggregate choice models (Table 4). The most troublesome variables were the household size variables. The persons per household coefficient for the San Diego PUMS model has the incorrect (negative) coefficient. The low household size variable (dummy variable of 1 if one-person household and 0 if two-or-more-person household) apparently misbehaves in all models. The income coefficients are quite well behaved and are correct in sign and magnitude. The low-income variable is necessary to correct for the nonlinear relationship between zero-worker household shares and household income. A strong cross-

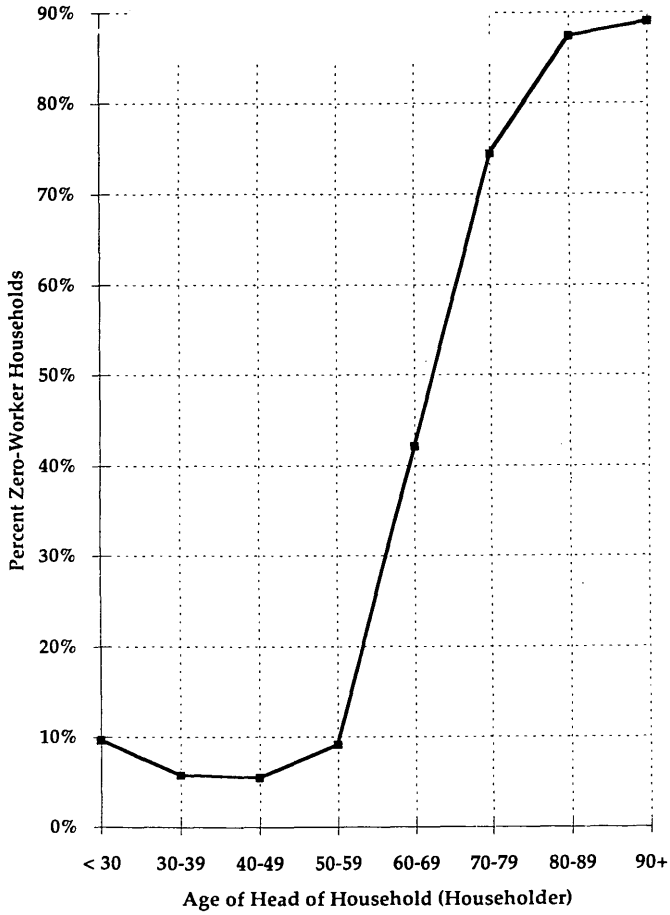

FIGURE 3 Percent zero-worker households by age of head of household, 1990 census PUMS, San Francisco Bay Area.

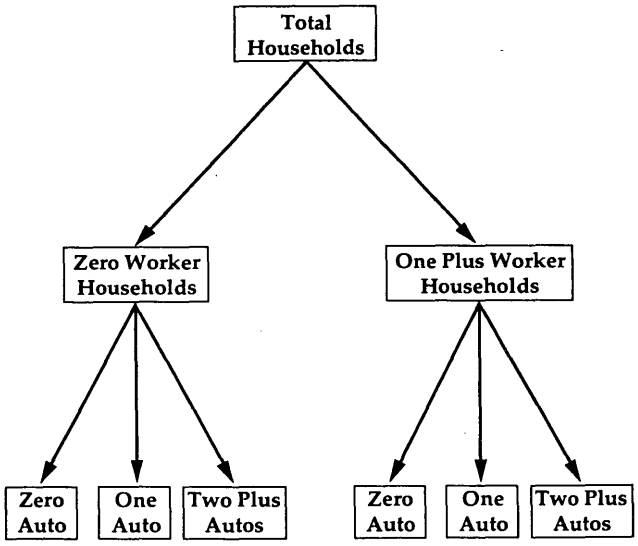

FIGURE 4 Structure of Bay Area automobile ownership models.

TABLE 4 NWHH Model Specifications

Model #4

Altern. 0	Altern. 1+	Variable	Bay Area Survey	Bay Area PUMS	San Diego PUMS
✔		Constant	-5.546	-5.619	-6.310
	✔	Income	1.32E-05	1.36E-05	9.20E-06
	✔	Persons/HH	0.2035	0.2777	-0.1442
✔		Low Income	1.16E-04	1.05E-04	1.15E-03
✔		Low Pers/HH	-0.108	-0.0539	-0.2478
✔		Age of Head	0.08612	0.08959	0.09252
rho-bar squared			0.427	0.469	0.455

where:

Income = mean household income in 1989 constant dollars.

Persons/HH = persons in household

Low Income = MAX((30000 - Income), 0)

Low Pers/HH = MAX((2 - Persons/HH), 0)

Age of Head = age of head of householder, or householder

correlation between household size and household income is a main culprit in explaining the incorrect signs for the household size coefficients.

The age of the head of household, as well as the average age in the household, was tested in the nonworking household model. The age of the head of household is an extremely powerful variable and basically doubles the rho-bar squared statistics from about 0.24 to 0.43 and above. Inclusion of the age of the head of household variable in the NWHH model raises an important issue: this variable clearly reduces model specification error. On the other hand including the age of the head of household in the model increases model measurement error. How can demographers accurately and confidently forecast the age of the head of household at a zonal level for 20 years into the future? Extensive further disaggregate validation checks are required to determine the value of including age in this demographic model.

With the exception of the household size variables, the estimation results for the NWHH model are quite encouraging when comparing results from the household travel survey with estimation results from the PUMS files for the two California regions.

The nonworking household automobile ownership models show excellent consistency between the Metropolitan Transportation Commission travel survey-based model and two PUMS-based models (Table 5). The household size variable for the one-automobile alternative is rather unstable and probably should be dropped from any final model. The model based on the natural logarithmic transformation of household income seems to work slightly better than that based on mean household income. The single-family dwelling unit dummy variable (1 for single-family, 0 for multifamily) is a strong, intuitive variable that suggests that automobile ownership increases as the share of single-family housing units in a neighborhood increases. The rho-bar squared statistics are quite low (0.16) but are characteristic of multinomial logit choice automobile ownership models.

The working household automobile ownership models are structured similarly to the NWHHAO models (Table 6). The household size variable in the one-automobile utility was dropped because of counterintuitive (negative coefficient) results. Mean household income was used instead of the logarithm of household income. The number of workers per working household variable was added to the two-or-more automobile utility equation to show

TABLE 5 NWHHAO Model Specifications

Model #5

Alternative 0	1	2+	Variable	Bay Area Survey	Bay Area PUMS	San Diego PUMS
✓			Constant (0)	6.453	5.439	5.861
		✓	Constant (2+)	-9.988	-9.513	-9.673
	✓		Log Income	0.8108	0.5779	0.6900
		✓	Log Income	1.567	1.297	1.445
	✓		Persons/HH	-0.2815	0.0453	-0.0862
		✓	Persons/HH	0.3622	0.4723	0.3703
	✓		Single Family DU	0.7166	0.9588	0.8685
		✓	Single Family DU	1.796	2.290	1.934
rho-bar squared				0.158	0.163	0.161

where:

Log Income = natural logarithm of mean household income in 1989 constant dollars.

Persons/HH = persons in household.

Single Family DU = single family dwelling unit dummy variable, where:

1, if single family dwelling unit.

0, if multiple family dwelling unit.

the impact of multiworker households on increasing the probability of owning two or more vehicles. All coefficient signs are correct in direction. Coefficient magnitudes tend to fluctuate more in this model than in the NWHHAO model, but all coefficients are in the same ballpark.

The model estimation results are basically encouraging and show the utility of using unconventional data sources, that is, the 1990 census PUMS, in statistically estimating selected demand models for a regional travel forecasting system. Results are generally consistent when comparing survey-based models with PUMS-based models and when comparing PUMS-based models between different metropolitan areas. The prospects for the transferability of these models and methodologies to other metropolitan areas are quite good. The San Diego PUMS models are quite similar to the Bay Area PUMS models. Prospects for the ease of access to PUMS data will increase as the PUMS files are released by the Census Bureau in CD-ROM format.

The basic and critical weakness of these PUMS-based automobile ownership models is the lack of sensitivity to land use density, urban form, and transit accessibility characteristics. These types of variables cannot be estimated from PUMS because the lowest geographic area is the PUMA, or a district of 100,000-plus population. Transit accessibility has been used successfully in the Bay Area and Portland model systems and should probably be incorporated (or at least attempted to be incorporated) into travel forecasting models in large metropolitan areas with significant transit ridership levels and significant shares of zero-automobile households.

The NWHH model is essentially a demographic model that splits households in a travel analysis zone into households by the

TABLE 6 WHHAO Model Specifications

Model #7.5

Alternative 0	1	2+	Variable	Bay Area Survey	Bay Area PUMS	San Diego PUMS
✓			Constant (0)	-1.384	-0.736	-1.139
		✓	Constant (2+)	-3.451	-2.951	-2.973
	✓		Income	4.22E-05	2.12E-05	3.03E-05
		✓	Income	6.02E-05	3.95E-05	5.60E-05
		✓	Persons/HH	0.3902	0.2908	0.2232
	✓		Single Family DU	1.2740	0.7576	0.2742
		✓	Single Family DU	2.234	2.070	1.341
		✓	Workers/HH	0.998	0.812	1.015
rho-bar squared				0.221	0.245	0.235

where:

Income = mean household income in 1989 constant dollars.

Persons/HH = persons in household

Single Family DU = single family dwelling unit dummy variable, where:

1, if single family dwelling unit.

0, if multiple family dwelling unit.

Workers/HH = employed residents in household

number of workers in the household. Neighborhood variables such as accessibility or density are probably appropriate for exclusion from a worker choice model. This means that the PUMS file is an appropriate data set for developing final workers-in-household models for metropolitan areas. Automobile ownership models, especially for working households, are probably better estimated by using household travel survey data and incorporating zone-level density and accessibility measures.

CONCLUSIONS

The study described here shows the ease and utility of developing demographic and automobile ownership models by using the 1990 census PUMS data sets. PUMS-based demographic and automobile ownership models can be developed for metropolitan areas and states that do not have access to recent household travel survey data. PUMS-based automobile ownership models can be developed either as cross-classification or as logit choice models. Without further disaggregate validation tests it is arguable whether automobile ownership logit choice models are superior to automobile ownership cross-classification models. These PUMS-based models would exclude potentially important independent variables such as density, urban form, and transit accessibility, but are improvements on the rudimentary cross-classification automobile ownership models that are typically limited to just one or two independent variables.

Forecasting the independent variables included in automobile ownership models—household income, household size, age, workers in household, accessibility, density, and so on—is arguably as important as the automobile ownership model specifications. Credible automobile ownership forecasts must be based on credible forecasts of the necessary demographic input variables. Research on the development of forecasting models for demographic variables is equally as important as travel behavior research.

Final, revised Bay Area travel demand models will be based on models estimated from the 1990 Bay Area Household Travel Survey rather than the 1990 census PUMS. Final Bay Area travel demand models will build on the insight gleaned from this PUMS-based analysis. A basic conclusion is that the 1990 census PUMS is a "second-best" data set for demographic and automobile ownership model development and is no substitute for a comprehensive household travel behavior survey.

REFERENCES

1. Hamburg, J. R., G. T. Lathrop, and E. J. Kaiser. *NCHRP Report 266: Forecasting Inputs to Transportation Planning.* TRB, National Research Council, Washington, D.C., 1983.
2. Bajpai, J. N. *NCHRP Report 328: Forecasting the Basic Inputs to Transportation Planning at the Zonal Level.* TRB, National Research Council, Washington, D.C., 1990.
3. Pearson, D. F. Disaggregating Zonal Households by Size, Income and Auto Ownership. Presented at Third National Conference on Transportation Planning Methods Applications, Dallas, Tex., April 1991.
4. Organization for Economic Cooperation and Development. *Forecasting Car Ownership and Use.* OECD Road Research Group, Paris, 1982.
5. Lave, C. *Things Won't Get a Lot Worse: The Future of U.S. Traffic Congestion.* UCI-ITS-WP-90-2. Institute of Transportation Studies, University of California, Irvine, 1990.
6. Pisarski, A. E. *Commuting in America: A National Report on Commuting Patterns and Trends.* Eno Foundation for Transportation, Incorporated, Westport, Conn., 1987.
7. Deutschman, H. D. Auto Ownership Revisited: A Review of Methods Used in Estimating and Distributing Auto Ownership. In *Highway Research Record 205*, HRB, National Research Council, Washington, D.C., 1967, pp. 31–49.
8. Prevedouros, P. D., J. L. Schofer. Factors Affecting Automobile Ownership and Use. In *Transportation Research Record 1364*, TRB, National Research Council, Washington, D.C., 1992, pp. 152–160.
9. Schimpeler Corradino Associates. *Oahu Model Update Study: Final Report: Volume II.* Oahu Metropolitan Planning Organization, Honolulu, Hawaii, 1982.
10. Lerman, S. R. Location, Housing, Automobile Ownership, and Mode to Work: A Joint Choice Model. In *Transportation Research Record 610*, TRB, National Research Council, Washington, D.C., 1976, pp. 6–11.
11. Lerman, S. R., and M. E. Ben-Akiva. Disaggregate Behavioral Model of Automobile Ownership. In *Transportation Research Record 569*, TRB, National Research Council, Washington, D.C., 1976, pp. 34–51.
12. Lawton, T. K. *Travel Forecasting Methodology Report (Regional Models).* Metropolitan Service District, Portland, Oreg., 1989.
13. Cambridge Systematics, Incorporated. *Making the Land Use Transportation Air Quality Connection: Model Modifications: Volume 4.* 1000 Friends of Oregon, Portland, Oreg., November 1992.
14. Ruiter, E. R., and M. E. Ben-Akiva. Disaggregate Travel Demand Models for the San Francisco Bay Area: System Structure, Component Models, and Application Procedures. In *Transportation Research Record 673*, TRB, National Research Council, Washington, D.C., 1978, pp. 121–128.
15. Kollo, H. P. H. *Home-Based Work Trips Models—Final Disaggregate Version: Travel Model Development with 1980/81 Data Base.* Working Paper 2. Metropolitan Transportation Commission, Oakland, Calif., 1987.
16. *Public Use Microdata Sample Technical Documentation.* Census of Population and Housing, 1990. U.S. Bureau of the Census, Washington, D.C., 1992.
17. *A Disaggregate Work-Trip Mode Choice Model for Aggregate Forecasting (Model TW): Technical Summary: Travel Model Update with 1980/81 Data Base.* Metropolitan Transportation Commission, Oakland, Calif., April, 1988.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.