# Auto Ownership Revisited: A Review of Methods Used in Estimating And Distributing Auto Ownership

HAROLD D. DEUTSCHMAN, Transportation Engineer,
Tri-State Transportation Commission

The intent of this study is to examine some of the methods used for estimating and distributing auto ownership. The importance of auto ownership is described for the generation of person trips and for the modal split decision. The key variables in the auto ownership estimation process used by some transportation studies and analysts are reviewed for their logic and predictive power. The preliminary findings from the Tri-State Transportation Commission are used to check out the logic of the variables used in this process.

In a test of the forecasting capability of these variables, auto registrations (by county for the years 1950 and 1960) for the New York Metropolitan Area were extracted from State Vehicular Records and used as a data base along with census data describing the social and economic factors of the predictors. Equations were derived from the 1950 data, applied to the 1960 data, and checked for accuracy with the auto registration figures for 1960.

In addition, techniques for deriving total auto registrations for an area are examined as well as the methodology employed to factor up the small area predictions to this control total. Finally, recommendations are offered for the process of setting up techniques for estimating and distributing auto ownership.

•ONE of the most important factors used in the trip generation process of forecasting person and vehicle trips is auto ownership. For example, the Chicago Area Transportation Study (CATS) found an excellent correlation between autos owned per dwelling place and destinations per dwelling place. The Pittsburgh Area Transportation Study (PATS) derived similar results from this relationship (Figs. 1, 2).

Auto ownership is also a significant determinant of mode choice. The households that do not have an auto available to them are in part captive to the service that transit supplies. In addition, those households in which another member of the household has a more pressing need for the family car (e.g., the housewife in the suburbs) are also constrained as to mode choice in their journey to work. PATS data point up the relationship of auto ownership and residential density on transit trips. Figure 3 shows the descreasing rate of transit trips to the central business district (CBD) with increasing auto ownership, consistent through the range of density readings.

While it is generally agreed that auto ownership rates must be studied and considered in any predictive trip generation equations and/or models, there does not seem to be this concurrence on the methodology employed in predicting and distributing auto ownership. The techniques used by some transportation studies and analysts in the field were reviewed, and the references used by this author are given in Table 1.

Paper sponsored by Committee on Origin and Destination and presented at the 46th Annual Meeting.

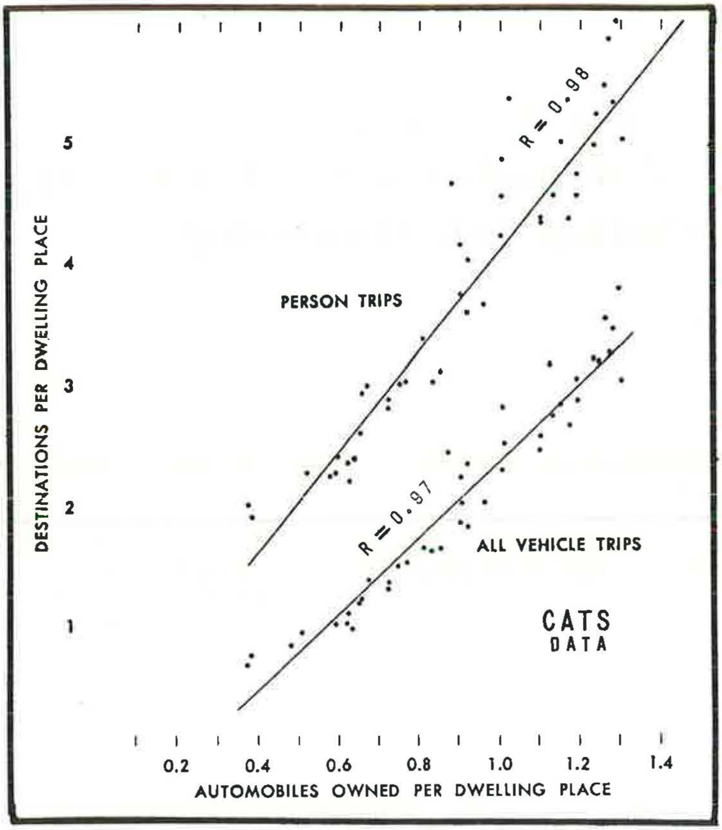


Figure 1. Person and vehicle trip destinations per dwelling place related to automobile ownership.

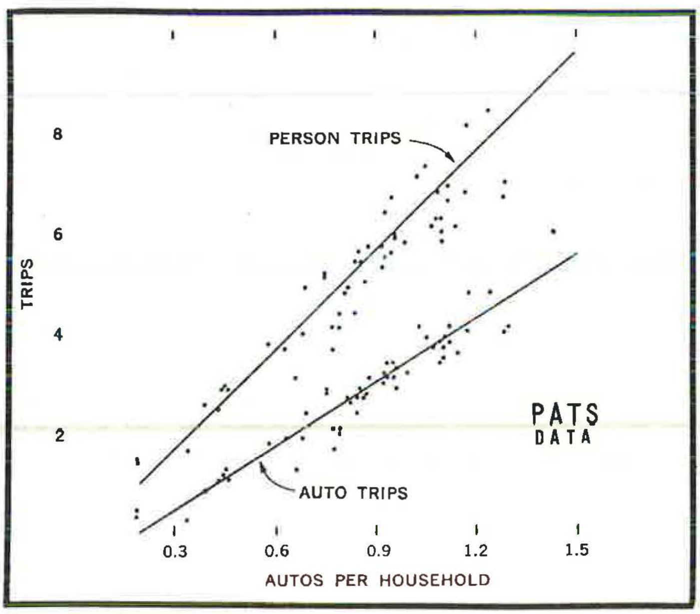


Figure 2. Person and auto trips per household related to auto ownership.

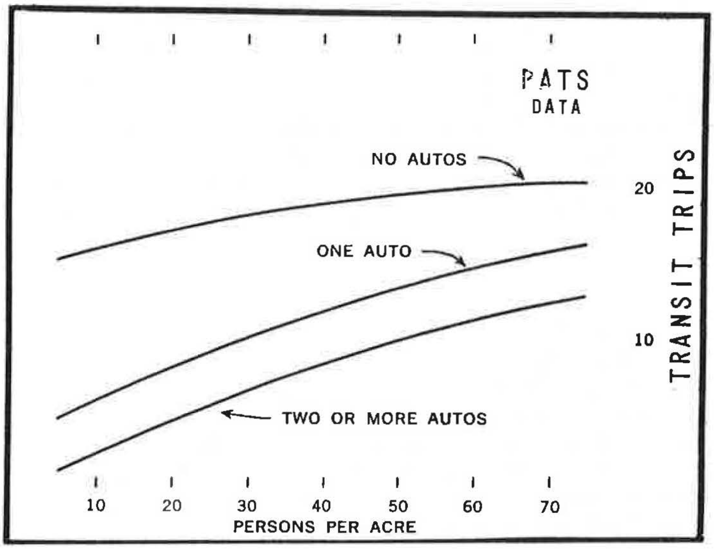


Figure 3. Golden Triangle transit trips per hundred population by auto ownership related to net residential density.

TABLE 1

SELECTED TECHNIQUES USED FOR ESTIMATING AUTO OWNERSHIP

| Procedure Classification | Variable(s) Used | References |
|---|---|---|
| Pittsburgh Area Transportation Study | Dependent variable: persons/auto<br>Independent variable: persons/residential acre | PATS Volume I and II<br>Technical Paper No. 14, Distributing Future Car Ownership, Nov. 1961<br>Technical Paper No. 3, Vehicle Registration Forecast, June 10, 1960 |
| Nathan Cherniack, Economist, The Port of New York Authority | Dependent variable: autos/acre<br>Independent variable: households/acre | Critique of Home-Interview Type O-D Surveys in Urban Areas, HRB Bulletin 253, 1960 |
| Puget Sound Regional Transportation Study | Dependent variable: autos/household<br>Independent variable: persons/household | Staff Report No. 9, 1985 Forecasts of Trucks and Passenger Vehicles Owned at Households, April 1964<br>Staff Report No. 13, Forecasting Household Characteristics for Determining Trip Production-Generation Rates, June 1964 |
| Social Statistics | Dependent variable: autos/household<br>Independent variables: persons/household socioeconomic status | The Use of Social Statistics in Estimating Auto Ownership (Abridgment), by William Michelson, Highway Research News No. 16, Dec. 1964 |
| Chicago Area Transportation Study (Procedure used for Auto Ownership Forecast for Fox River Valley Study) | Dependent variable: autos/household<br>Independent variable: household income | CATS Car Ownership Forecast: A New Approach, by S. V. Ferrera, CATS Research News, June 1965 |
| Penn-Jersey Transportation Study | Dependent variable: autos/household<br>Independent variables: log median household income<br>log households/residential acre | P-J Memo from E. O. Fichtner to Richard Hubbell, Acting Director, April 5, 1965 |

## VARIABLES USED IN ESTIMATING AUTO OWNERSHIP

The variables most commonly used for estimating auto ownership are residential density, household income, and persons per household. These variables have been used individually and also in linear combinations as determinants of auto availability. (For the purposes of this report, auto availability and auto ownership are used interchangeably.)

In the following pages, each of the foregoing variables is studied to determine (a) how well these independent variables reproduce the survey data for auto availability; (b) the logic of the variable and the methodology used as a predictive device; and (c) the results of predictions that have been made using these variables vs the results derived from control totals (a trend of the dependent variable, autos per household). Finally, recommendations are made on the desirability of using each of the variables to forecast autos. In addition, recommendations for changes (or additional efforts) are made in order to render the equations or models used operational in the sense of producing reliable future estimates.

## Density

The measures usually employed to represent density in predicting auto ownership are net residential density (persons per residential acre), gross density (persons per acre), and percent single unit structures. In addition, households per acre may be used in lieu of persons per acre.

The correlation results of density vs autos are usually very good for the base or study year. For example, in the Tri-State New York Metropolitan Study, using preliminary home interview results with 278 zones as data points (in expanding the home interview survey from a 1 percent sample to its representative universe, the study area was divided into 278 expansion areas or zones), and fitting the data to a best-fit straight line, the results are as follows (see also Figs. 4, 5):

| Dependent Variable | Independent Variable | R[b] | $S/\overline{X}$[c] |
|---|---|---|---|
| Vehicles/household[a] | (Log of) Gross density (living quarters/sq mile) | 0.92 | 23% |
| Vehicles/household | Percent single unit structures | 0.93 | 22% |

[a]Vehicles/household includes private autos, rented cars and trucks and taxis available to the household.
[b]R = Coefficient of correlation.
[c]S = Standard error of estimate; $\overline{X}$ = Mean of dependent variable.

The results from the Chicago Area Transportation Study also showed good correlations for the survey year using a measure of density to estimate autos. Fitting a parabolic curve to the data from 77 districts in the study area (Fig. 6), the standard error was ±15 percent for the relationship of autos per acre vs households per acre. (This relationship was derived by Nathan Cherniack—see Table 1.)

The Pittsburgh Area Transportation Study, using persons per residential acre as a measure of density, produced a good fit for the relationship of autos per population vs density. The standard error of estimate was 19 percent, using a total of 220 zones in the study area for this analysis.

The variable density has reproduced autos available very well for the survey year. However, if one thinks about using a measure of density for estimating future auto ownership, a careful look at a curve of vehicle availability vs household income, stratified by number of housing units (HU) in the structure (Fig. 7), should indicate that residential density (measured by number of units in the structure), when used as the sole criterion or function for predicting autos, will underestimate autos by a significant number. The areas that are presently at a capacity such that no new growth is expected (or, in other words, when the density will remain constant) must also maintain their constant rate of auto ownership according to the stated relationship (autos vs density). This appears false, since as income changes (rises) with a constant density, the car ownership rate will also change (increase). To illustrate, a change in income from \$6,000 to \$8,000 in ≥ 5 HU/structure will yield an increase in auto availability from 0.46 autos/household to 0.62 autos/household. A change in income of \$8,000 to \$12,000 in

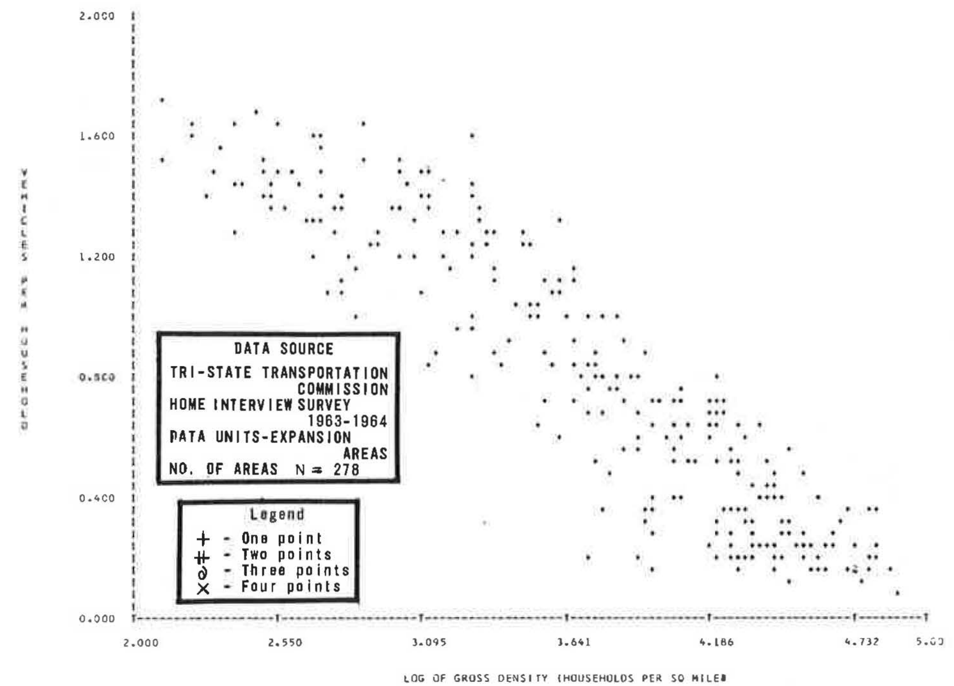


Figure 4. Vehicles per household vs log of gross density.

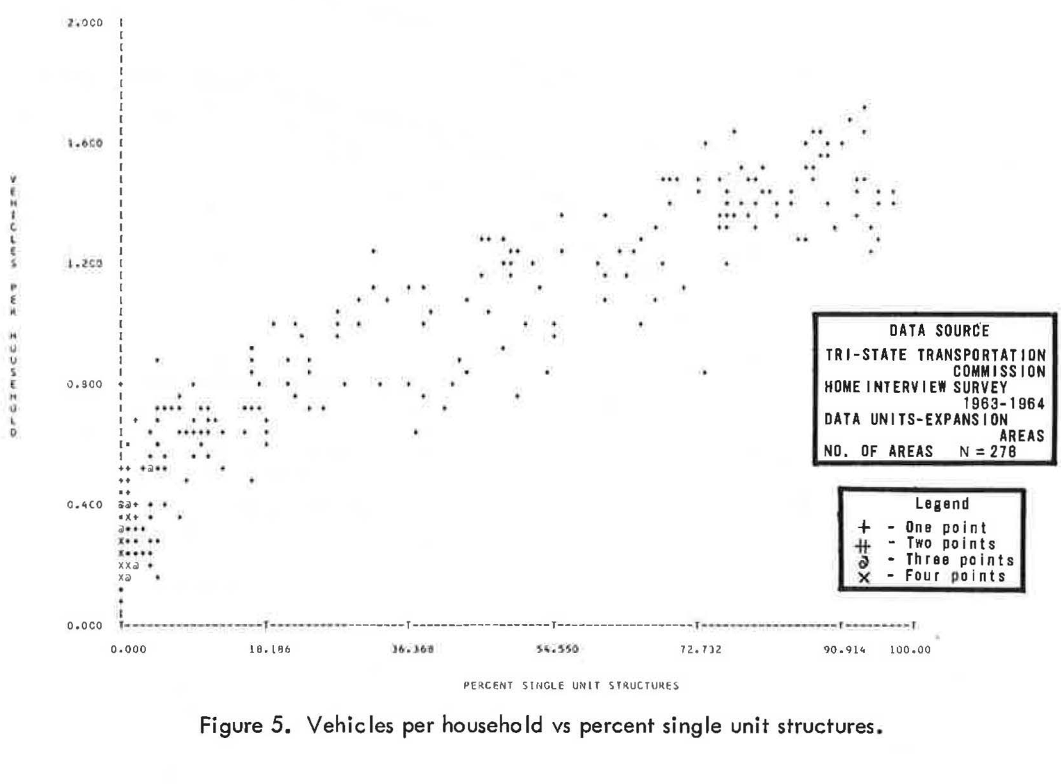


Figure 5. Vehicles per household vs percent single unit structures.

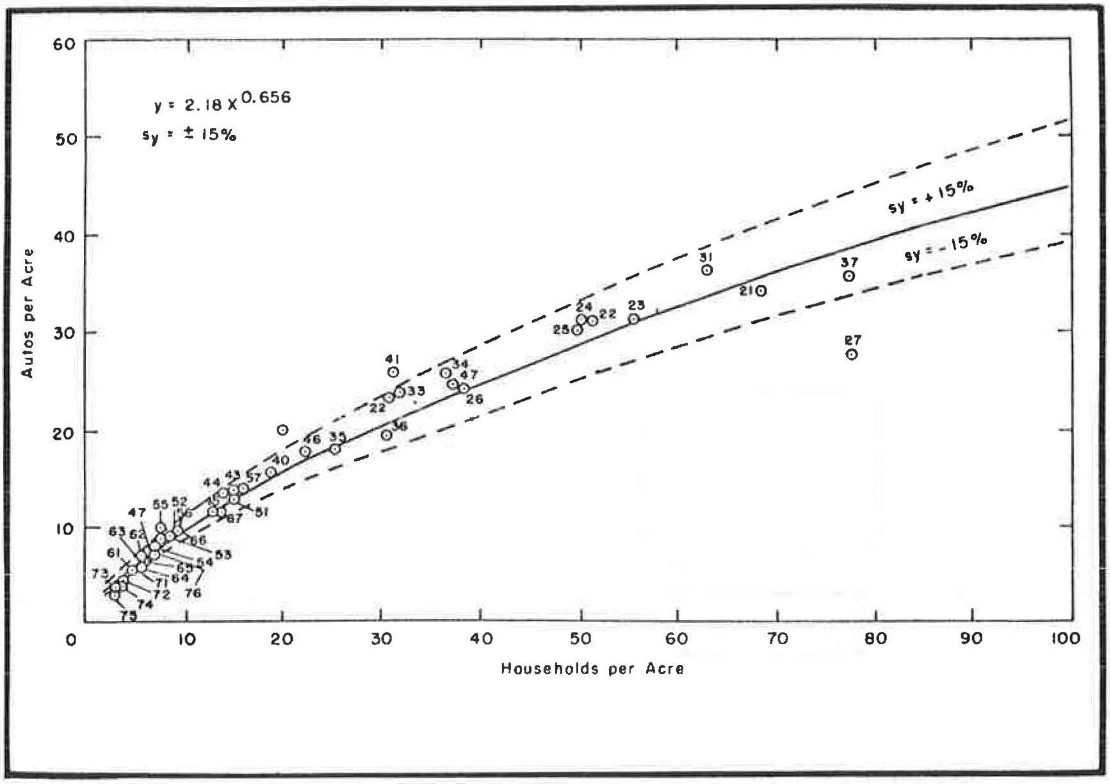


Figure 6. Relation between autos per acre and households per acre for 77 districts in Chicago, 1956–57.

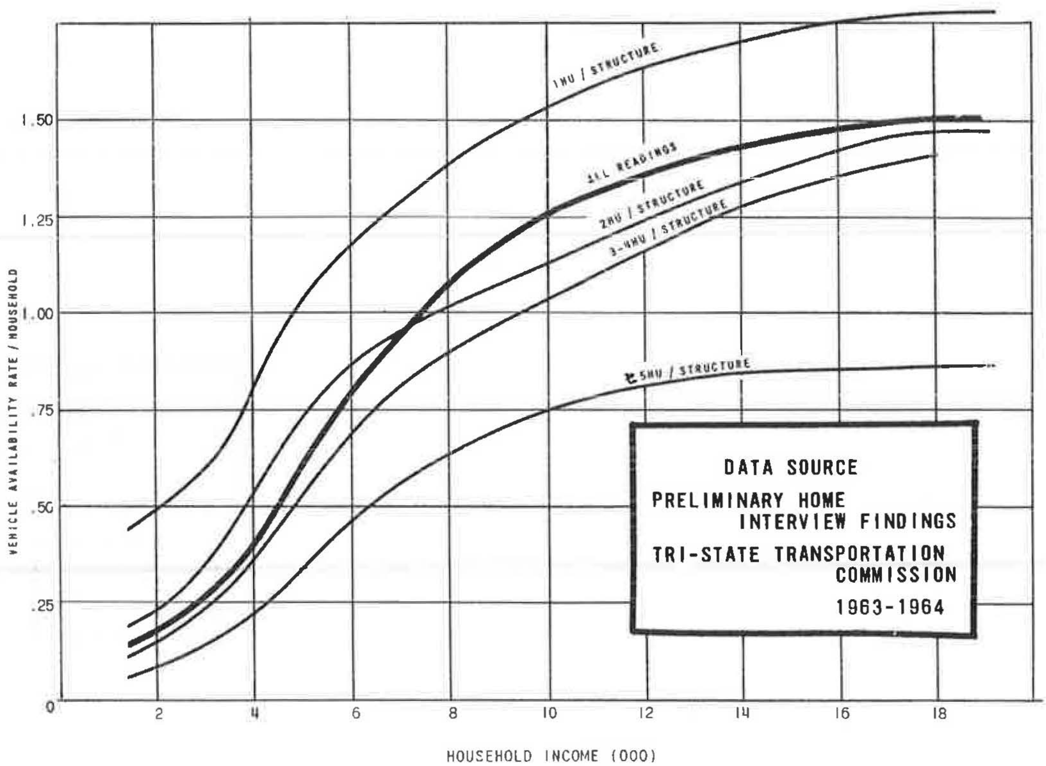


Figure 7. Vehicle availability vs household income, stratified by number of housing units in structure.

single unit structures will yield an increase in auto availability from 1.38 autos/household to 1.65 autos/household.

The studies that used density as the sole variable to forecast autos found that their prediction underestimated the expected number of autos derived from control totals. In distributing future auto ownership, PATS used the best-fit straight line of persons per residential acre vs autos per person. PATS estimated the independent variable for the forecast year and then predicted autos per person using the regression results. However, this method distributed only 55 percent of the "expected" total auto increase. The expected increase in autos was derived from a trend analysis of the dependent variable (autos per person) and a comparison of PATS data to that of the United States. PATS then had the task of distributing the remaining 45 percent of the autos. This was completed by distributing them as a direct function of the population of each zone.

In a general conclusion, the cross-sectional type analysis of density vs autos is not valid for predictive purposes. The reasoning that when any zone B reaches the density of a zone A it will have the same auto availability rate as zone A appears to be false unless the element of time is introduced to the solution. This element of time refers to the natural growth of autos per household in zone A due to the effect of increased real household income, more leisure time, etc.

Household Income

Preliminary results from the Tri-State Transportation Study have shown that the relationship between income and autos is approximately a straight line in the low- and middle-income range and then flattens out (or is parabolic) for the higher incomes (Fig. 8). Using zonal data (n = 278 zones), and fitting a straight-line relationship between median household income and vehicles per household, the correlation results were only fair with a coefficient of correlation of 0.68 and a standard error of 48 percent (Fig. 9).

Whereas household income has been used in combination with other independent variables in predicting vehicles, it has not been used by transportation studies as a sole

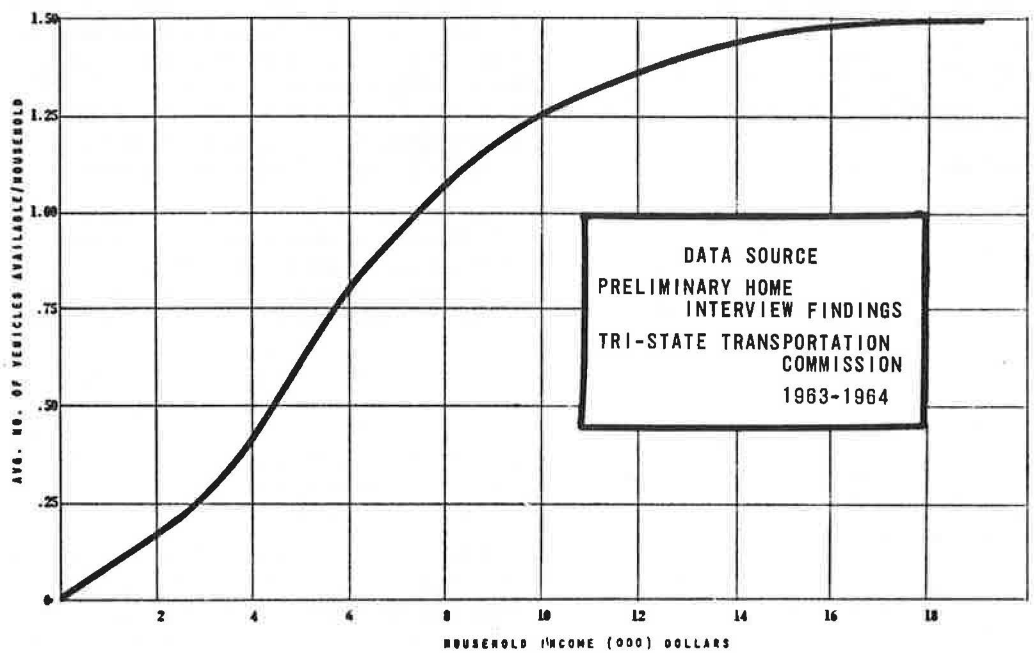


Figure 8. Vehicle availability vs household income.

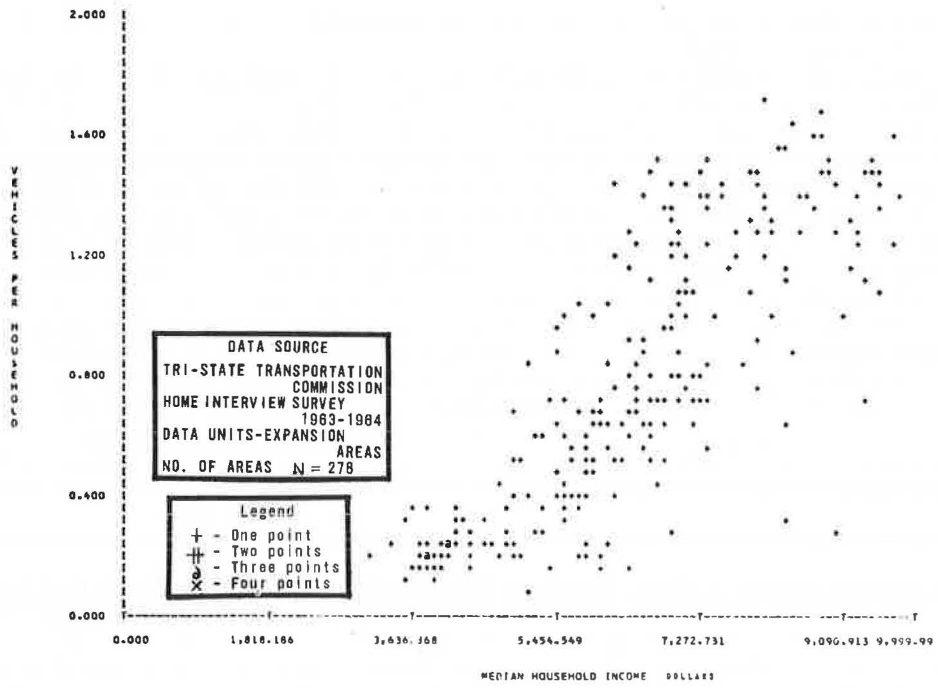


Figure 9. Vehicles per household vs median household income.

TABLE 2

CATS PROCEDURE FOR FORECASTING AUTOS PER HOUSEHOLD

| Income Class ($) | Percent of Households in Survey Year | Change* | Percent of Households in Forecast Year |
|---|---|---|---|
| 0-3,000 | 15 | 0<br>-7.5 | 7.5 |
| 3,000-5,000 | 13 | +7.5<br>-6.5 | 14.0 |
| 5,000-7,000 | 24 | +6.5<br>-12.0 | 18.5 |
| 7,000-10,000 | 27 | +12.0<br>-13.5 | 25.5 |
| 10,000-15,000 | 14 | +13.5<br>-7.0 | 20.5 |
| 15,000-25,000 | 4 | +7.0<br>-2.0 | 9.0 |
| 25,000+ | 3 | +2.0<br>0 | 5.0 |

*Assuming a change in real income per household of 50% for the survey year to the forecast year, the percent in each income class is changed as follows: 50% of the households in each income class are shifted to the next class. Thus, the income class $0-3,000 which initially contained 15% of all households, now has only 7.5% of the households, with 7.5% shifting to the $3,000-5,000 income class. The $3,000-5,000 income group loses 50% x 13% or 6.5% to the next income group and thus has a +7.5% - 6.5% total change or an increase of 1% of households for the forecast year.

Assuming the number of persons per household will remain constant (survey year to forecast year) the total number of autos per zone is determined by multiplying the auto ownership rates (by income class) by the number of households: $\left(\frac{\text{Estimated Pop.}}{\text{Pop./Household}}\right)$ (Auto/Household)

independent variable for a regression-type analysis. Income, however, has been used as the single independent variable in a process for estimating autos per household. This process, recently developed by CATS (see Table 1), uses the following methodology for estimating autos per household and total autos for an area:

1. Establish from survey results the percent of households owning 1, 2 and 3 or more autos per income group;
2. Predict a single figure of percent growth of real income (per household) for the survey year to the forecast year; and
3. Hold the rates from (1) constant with the percent of households in each income group changing at the same rate as the real income change (survey to forecast year). The method used in this procedure is outlined in Table 2.

Critical Appraisal of CATS Procedure—The CATS methodology is limited by its rigid and arbitrary movement from one income class to the next. It is in a sense tied in to the classification of its income classes for its results. For example, for the $10, 00 to $15, 000 income classification, a uniform income increase of 50 percent should propel all of the households in the class into the next one ($15, 000 to $25, 000) and not the 54 percent (7. 5 ÷ 14) suggested by the CATS method. The author suggests that the methodology would have real merit if a uniform or normal distribution is assumed for each income class and the households moved through time as follows:

| Income Class | Percent of Households in Survey Year | Change |
|---|---|---|
| $0-3000 | 15 | 0<br>-5.0% |
| $3000-5000 | 13 | +5.0%<br>-10.8% |

To explain, if everyone's income is increased 50 percent and a uniform distribution is assumed for each income class, then all the households earning $2000 or more in the survey year will be propelled to the next class. Thus, ⅓ of 15 percent or 5 percent of the households move to the $3000-$5000 class and ⅔ of 15 percent or 10 percent of the households remain in the $0-$3000 classification.

The results of the two procedures for predicting household income distributions (CATS vs uniform distribution) are given in Table 3.

Limitations—The CATS procedure, modified by the stated recommendations in the procedure of moving households through and within the income groups, has merit for forecasting autos by studying the changes in real income. One assumption inherent in this procedure is that the growth in any county or zone between the survey year and the forecast year will approximate the density configuration already intact in that area. In other words, the additional households should have approximately the same percentage distribution of single family units and apartment houses. For example, if the growth of a suburban community is expected mainly in two-story garden apartment houses, then the relationship between household income and auto availability when not stratified by density will produce rates on the high side for autos available per household for these new garden-type apartments. In a similar manner, if the growth is expected predominantly in single family units, above and beyond the distribution of single unit structures/ total units for this area for the survey year, then the rates derived from the survey data will produce results on the low side for the additional units (see Fig. 7).

TABLE 3

COMPARISON OF TWO PROCEDURES FOR FORECASTING HOUSEHOLD INCOME DISTRIBUTIONS

| Income Class ($) | Percent of Households in Survey Year | Percent of Households in Forecast Year | |
|---|---|---|---|
| | | CATS Method | Uniform Distribution |
| 0-3, 000 | 15 | 7. 5 | 10. 0 |
| 3, 000-5, 000 | 13 | 14. 0 | 7. 2 |
| 5, 000-7, 000 | 24 | 18. 5 | 8. 5 |
| 7, 000-10, 000 | 27 | 25. 5 | 22. 3 |
| 10, 000-15, 000 | 14 | 20. 5 | 31. 0 |
| 15, 000-25, 000 | 4 | 9. 0 | 14. 6 |
| 25, 000+ | 3 | 5. 0 | 6. 4 |

Assuming a mean income value as the midpoint of each income group and the mean value of the open-ended class of $25,000+ as $40,000, then the mean household income for the survey year was $8,310. Under the assumption that everyone's income increases by 50%, then the mean value of a household income for the forecast year should be about $12,500. The mean income for the CATS procedure of rigidly moving 50% of the households from one group to the next produced a mean income of $10,430 while the recommended shifting of households by assuming a uniform distribution produced a mean income of $12,300.

The procedure also has other recognizable limitations in the assumptions. The average income increase is assumed to be constant for each income class. Thus, if the increase is 50 percent, the household earning $10, 000 will move to $15, 000 while the household earning $3, 000 will move to $4, 500. Data from the New York Tri-State Metropolitan Region for income distributions for the years 1950 and 1960 show that the income increase is not distributed uniformly for each class (Fig. 10). In this time period, the increase for the middle-income class in 1950 was greater than that for the lower and upper classes. In addition, there was a spreading out of the range or flattening out of the curve for these middle-income classes in this 10-year period.

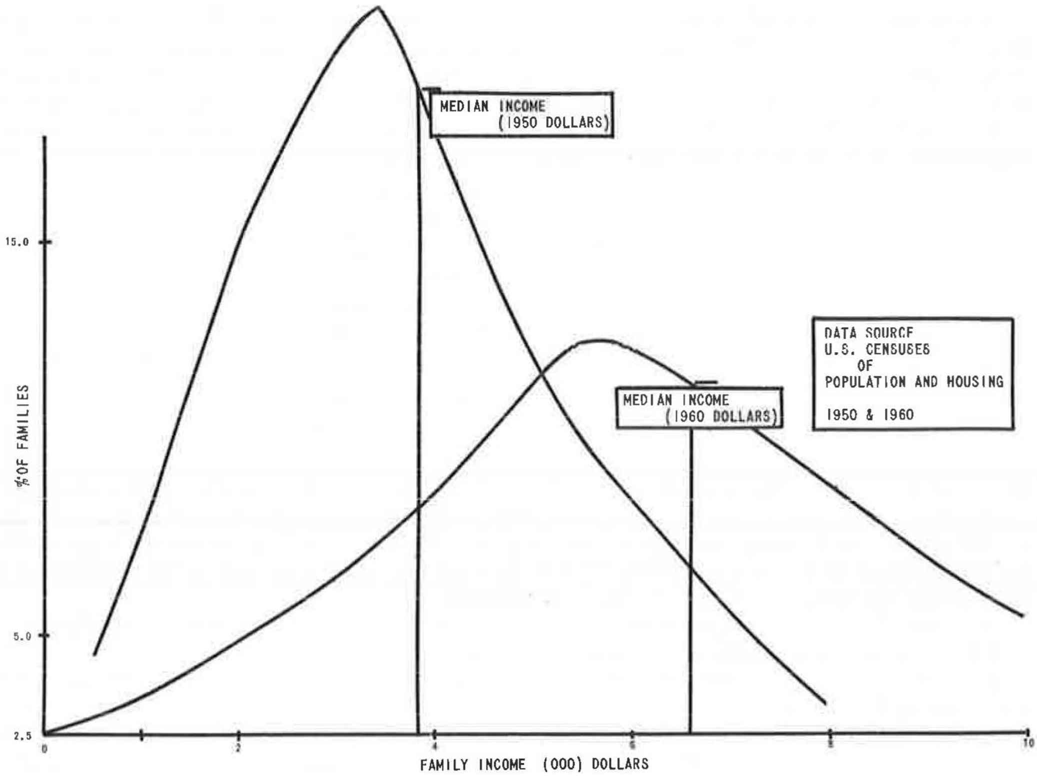


Figure 10. Family income distributions for the New York Metropolitan Area, 1950-1960.

Input to this procedure also includes an estimate of the average change in real income by area, between the survey year and the forecast year. Data are available for this input from the U. S. Department of Commerce in the Survey of Current Business publications, which report personal income per consumer unit by year, by state and also for the United States. In addition, income data are available from census surveys every 10 years (with the possibility in the future that such reporting will be made at 5-year intervals). The Census Bureau reports income data for census tracts, municipalities, and counties.

The use of household income (as specified in the modified CATS procedure) appears valid as a technique for forecasting autos. It is recommended, however, that care be exercised in using the technique in the following areas of concern:

1. Checking that the distribution of new housing units by type is approximately equal to that existing (if a large difference exists, the estimates of autos must be adjusted accordingly);
2. Checking the validity of the assumption that everyone's income increases by a uniform amount; and
3. Making a concentrated effort to insure that the estimated average change in real income from the survey year to the forecast year is reasonable and reliable.

Persons Per Household

The variable of persons per household has been used by the Puget Sound Regional Transportation Study as the sole determinant for distributing autos for a forecast year. This distribution was made on the basis of a linear regression relationship between

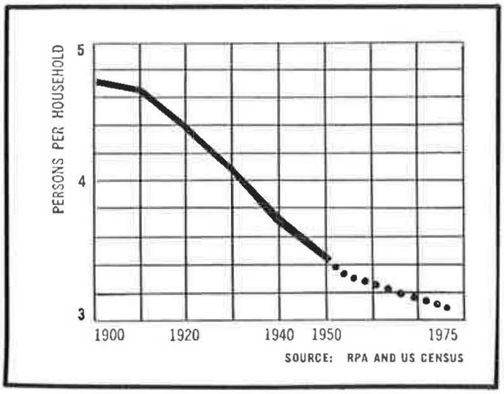


Figure 11. Persons per household vs year for the New York Metropolitan Area.

TABLE 4

REGISTERED AUTOS PER HOUSEHOLD FOR NEW JERSEY COUNTIES IN THE TRI-STATE METROPOLITAN AREA*

| County | Year | | |
|---|---|---|---|
| | 1940 | 1950 | 1960 |
| Bergen | 1.01 | 1.03 | 1.26 |
| Essex | 0.84 | 1.00 | 1.02 |
| Hudson | 0.60 | 0.72 | 0.77 |
| Mercer | 0.90 | 1.05 | 1.18 |
| Middlesex | 0.94 | 0.91 | 1.23 |
| Monmouth | 1.07 | 1.15 | 1.26 |
| Morris | 1.14 | 1.04 | 1.42 |
| Passaic | 0.81 | 1.00 | 1.09 |
| Somerset | 1.08 | 1.12 | 1.02 |
| Union | 1.00 | 1.20 | 1.33 |

*Number of households were abstracted from census data; number of registered autos were abstracted from state registration totals.

average household size and average number of automobiles per household (autos per household increases with increasing household size). Using survey results, this relationship reproduced the survey results (by analysis zone) to within an accuracy of ±0.05 autos per household on the average.

Since autos increased faster than population (independent estimates), Puget Sound underestimated the total number of autos distributed by approximately 28 percent, or 255,666 automobiles, when checked with control totals. The trend total was derived in part by an historical trend of the dependent variable autos per household using an admittedly small number of households for the base.

For the New York Tri-State Metropolitan Area, the household size has been on the decline since 1900, with the number of households increasing about 1.4 times faster than the population (Fig. 11). Furthermore, county data for 1940, 1950 and 1960 for the New Jersey portion of the Metropolitan Area indicate that autos per household are on the increase (Table 4).

Evidence from the data for the New York Metropolitan Area has shown that autos per household increased while persons per household decreased (contrary to the Puget Sound results). It is thus recommended that persons per household should not be used as a sole determinant of autos per household.

## INTRODUCING THE ELEMENT OF TIME IN ESTIMATING AUTOS FOR A FORECAST YEAR

In describing residential density as a predictor of auto availability, it was pointed out that this independent variable would underestimate the number of autos because auto ownership changes at a much more rapid rate with time than does density. Hoch, in a CATS paper (1), introduces time as a function of his analysis procedure. The methodology of this procedure is as follows: Budget data are gathered relating income to autos registered for a number of different years. A linear equation $Y = MX + B + Kt$ is derived where Y is autos per household, X is a measure of household income and $B + Kt$ is the Y axis intercept. The observed intercept $B + Kt$ is then related to the average household income (for the different years in which the data were collected). The forecast year intercept is then derived by extrapolating the curve of intercept value vs average income (Fig. 12).

Another method of introducing time in the predictive function is to relate auto availability to the combined effect of household income and residential density. If the density of an area remains the same, then the auto availability rate per household will increase if the real income per household increases. In addition, a move or shift from one density level to another will produce a real change in auto availability. Preliminary results from the Tri-State Transportation Study show excellent linear correlations between the combination of density and income in estimating autos. Using expansion areas as zones ($n = 278$ zones) for observation points, the results are as follows:

| Dependent Variable | Independent Variables | R | S/$\overline{X}$ |
|---|---|---|---|
| Vehicles/household | Median Household Income<br>Percent Single Unit Structures | 0.94 | 20% |
| Vehicles/household | Median Household Income<br>Log of Gross Density (living quarters/sq mile) | 0.95 | 19% |

Probably more significant than the good correlation for one point in time between the foregoing independent variables and vehicles per household is the seemingly logical reaction of these variables (rate of change with time) with that of auto availability.

## Validity and Limitations of the Procedures

The use of the combined time series and budget study procedure for forecasting autos (Hoch's methodology) is considered valid, although some inherent assumptions must be recognized before employing the procedure. Since data on income vs auto ownership are available only for the United States, one must assume that the relationship derived with national data holds for the study area under consideration. Care must also be exercised in the extrapolation of the relationship between mean income and the intercept value.

The procedure of using household income and density to forecast autos is also a valid and logical process although this technique has some of the same limitations as those discussed previously in the section on income, such as estimating the real income change as well as the distribution of income classes.

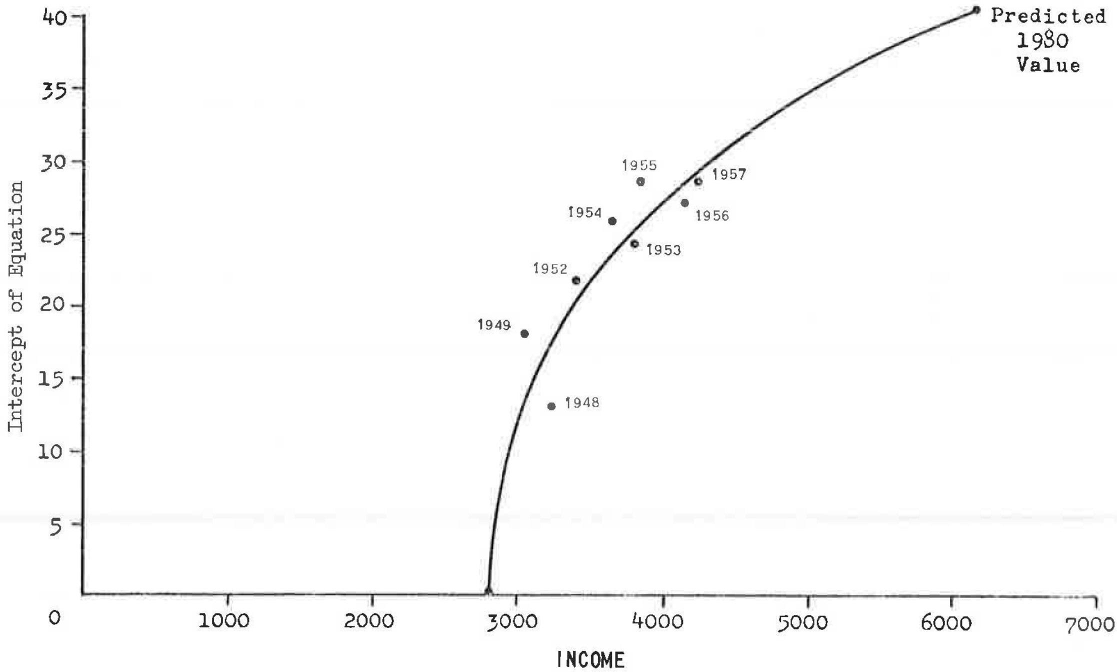


Figure 12. Relation of intercept in equation to average income (using equation to forecast autos, United States data).

## DERIVING CONTROL TOTALS

In conjunction with estimating autos per household or the total number of automobiles for a study area by using independent causal variables, most studies control these estimates by applying a trend analysis to the dependent variable autos per household or autos per person. The basic data available for such control totals are the yearly tabulations from the Automobile Facts and Figures Handbook. These tabulations present the percent of households owning 1 or more cars for the United States by year starting in 1948 to the present and the percent households owning 2 or more autos by years for the years 1954 to the present. One method that transportation studies have used to establish a control total of autos for their study area is to extrapolate the percent of household in each auto ownership class, as developed from United States data. This assumes that the percentage distribution of households in the auto ownership classes (0, 1, 2 or more autos) for the study area approximates that for the United States, and more important, that this relationship (of study area to the U. S.) will hold in the future.

Automobile Facts and Figures also tabulates the number of autos registered by county, by year, for selected counties in the United States. In addition these tabulations include estimates of population and households. The State Motor Vehicle Agencies also publish yearly data by county on auto registrations. A second method of establishing control totals is to draw a trend of autos per person by county, for a number of years. A control total of autos per person could be established by studying these trends as well as data from other areas. A maximum rate of autos per person could be established by determining what portion of the population will most likely own an auto. To illustrate, assume the portion of the population either under 18 or over 65 years (for the forecast year) is 40 percent, and that no one in these two age groups is likely to own an auto;

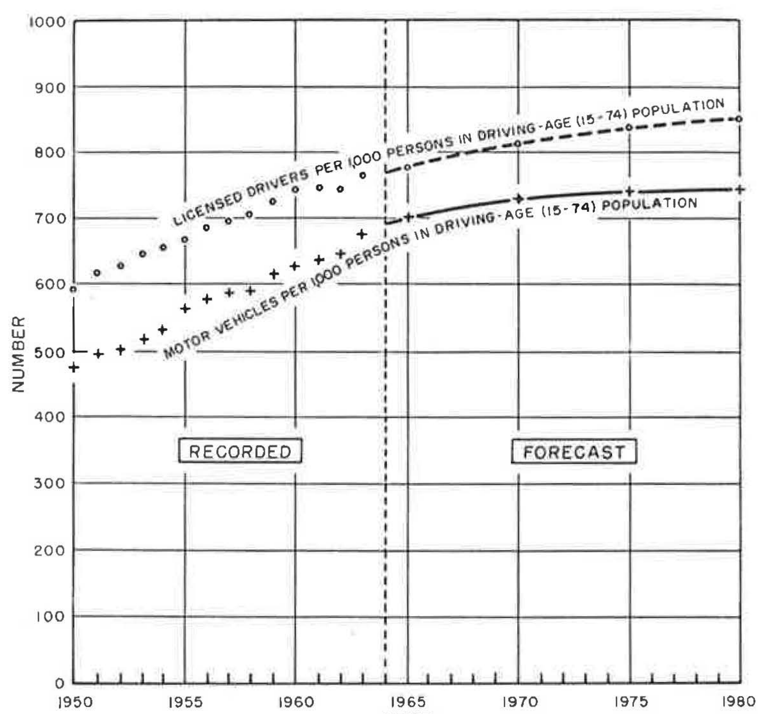


Figure 13. Relationship of the number of licensed drivers and vehicle registrations to population in the vehicle operating age groups.

then the maximum rate of autos per person for this area is 60 autos per 100 persons. This, of course, presumes an average rate of 1 auto per person in the 18-65 age group. The absolute minimum rate would be the present rate of autos per person assuming that autos will increase at the same rate as population. More reasonable control totals of autos per person may be established within these maximum and minimum rates.

A study by the Highway Statistics Division of the Bureau of Public Roads (2) has indicated that the technique of relating the number of licensed drivers to autos owned is very useful for predictive purposes. Data collected for the total population of the United States, the persons of driving age, and the number of licensed drivers produced the relationships shown in Figure 13. This graph shows a good fit of licensed drivers/1000 persons vs year as well as motor vehicles/1000 persons vs year. More important, a stable relationship is also shown between licensed drivers and motor vehicles. The curves (parabolas) are then extrapolated to yield estimates of drivers and vehicles for the future.

Either of the outlined procedures, trend of percent households vs car ownership rate by year, or trend of autos per person by year, is considered valid for establishing a trend of autos per person or of autos per household. The latter method is preferred for the following reasons:

1. Use of United States data presumes that the study area characteristics will be similar in the future to those of the United States;
2. A curve of percent households owning 1 or more autos vs time does not yield reasonable control limits, except that the total should not exceed 100 percent, and it also does not reflect individual household behavior; and
3. A trend of persons per household can be checked for reasonableness. The data for this type of analysis are also readily available at the level of the study area.

### Use of Control Total vs Estimates From Independent Variables

The estimate of total autos derived from a technique or model using independent variables should be reasonably close when compared with that established by the control totals. If the two estimates differ substantially (i.e., greater than ±20 percent), then there is reason to review carefully both procedures. Too often the total derived by a trend of autos/household is held fixed even though the data source to derive these results is not as rich or reliable as that used in the model using independent variables.

The ideal case involves the establishment of reasonable maximum and minimum limits of persons/auto and the acceptance of the results produced by the independent variables or models if they fall in this range.

## TESTING FOR THE PREDICTIVE POWERS OF THE VARIABLES

The process of determining the equation(s) and/or models for use in predicting autos usually consists of formulating a logical hypothesis and then testing it against the survey data. Too many times, however, the variables for the process are chosen by an analysis of the survey data. In other words, the process is often a sophisticated method of curve fitting. The measure of success of a procedure is not primarily how well the curve fits for the present (measured by the coefficient of correlation and the standard error of estimate), but how well the relationship holds up over time. In lieu of a time machine, the analyst must test his procedure by gathering up data for two periods in time. The data may not be too rich in information or source, and they may not be for the specific area under consideration, but nevertheless they can serve as an indication of the predictive power of the variables chosen. This type of analysis coupled with the logic of the variables in describing the change over time is a must for insuring an acceptable performance by the estimating process.

It was thus decided to test the predictive power of variables most often used for estimating auto ownership: (a) residential density, (b) household income, and (c) persons per household. These tests were made on the variables taken one at a time and also in combinations. The two points in time for this study were 1950 and 1960 since census data were available by county for these two years. The areas chosen for the analysis

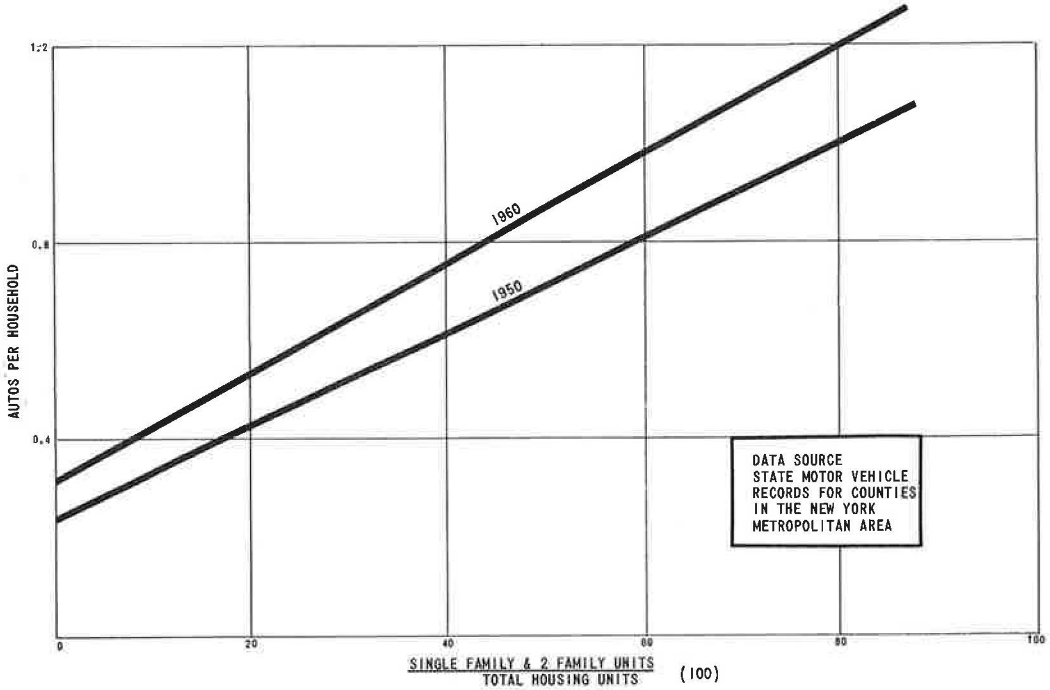


Figure 14. Autos per household vs percent (1 & 2) units per structure for the years 1950 and 1960.

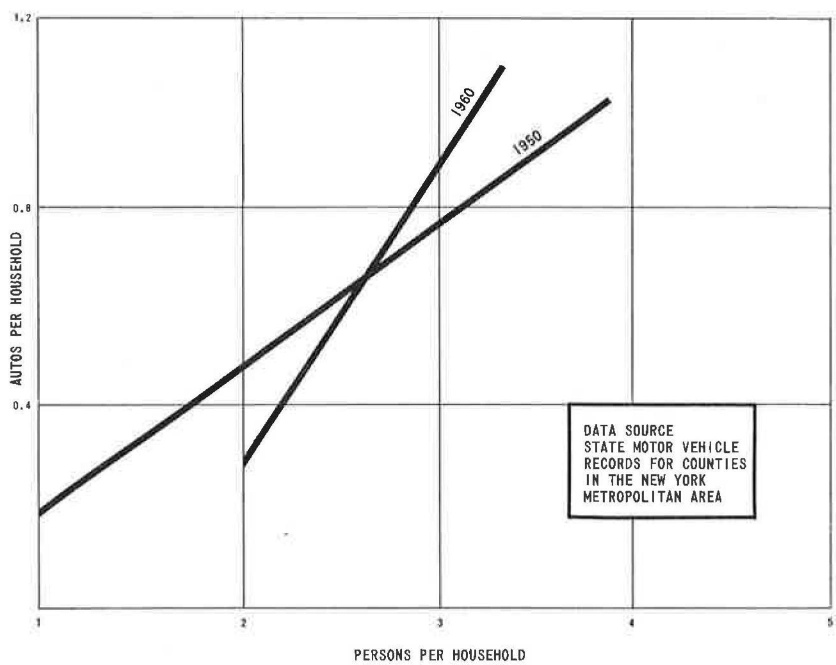


Figure 15. Autos per household vs persons per household for the years 1950 and 1960.

TABLE 5

ACTUAL AND PREDICTED 1960 AUTOS PER HOUSEHOLD

| Zone | Actual 1960 Autos/Household | Predictive Variables: Median Household Income | % (1 & 2) Unit Structures | Persons/ Household | Median Income and % (1 & 2) Unit Structures | Median Income and Persons/Household | % (1 & 2) Unit Structures and Persons/Household |
|---|---|---|---|---|---|---|---|
| 1 Bergen | 1.26 | 1.17* | 1.05 | 0.90 | 1.31 | 1.16 | 1.05 |
| 2 Essex | 1.02 | 1.09 | 0.72 | 0.82 | 0.88 | 0.98 | 0.73 |
| 3 Hudson | 0.77 | 1.06 | 0.64 | 0.81 | 0.77 | 0.93 | 0.65 |
| 4 Mercer | 1.18 | 1.10 | 1.07 | 0.93 | 1.24 | 1.09 | 1.08 |
| 5 Middlesex | 1.23 | 1.12 | 1.09 | 0.96 | 1.29 | 1.16 | 1.10 |
| 6 Monmouth | 1.26 | 1.08 | 1.09 | 0.93 | 1.24 | 1.07 | 1.10 |
| 7 Morris | 1.42 | 1.16 | 1.13 | 0.97 | 1.37 | 1.22 | 1.14 |
| 8 Passaic | 1.09 | 1.08 | 0.94 | 0.85 | 1.08 | 0.99 | 0.95 |
| 9 Union | 1.33 | 1.16 | 1.01 | 0.89 | 1.25 | 1.14 | 1.02 |
| 10 New York City | 0.47 | 1.06 | 0.50 | 0.76 | 0.61 | 0.87 | 0.51 |
| 11 Dutchess | 1.14 | 1.08 | 1.04 | 1.00 | 1.18 | 1.16 | 1.04 |
| 12 Nassau | 1.37 | 1.20 | 1.11 | 1.00 | 1.41 | 1.31 | 1.11 |
| 13 Orange | 1.06 | 1.04 | 1.04 | 0.91 | 1.13 | 1.00 | 1.04 |
| 14 Putnam | 1.59 | 1.09 | 1.14 | 0.91 | 1.30 | 1.06 | 1.15 |
| 15 Rockland | 1.24 | 1.14 | 1.08 | 1.06 | 1.30 | 1.30 | 1.08 |
| 16 Suffolk | 1.31 | 1.10 | 1.16 | 1.03 | 1.33 | 1.22 | 1.16 |
| 17 Westchester | 1.20 | 1.18 | 0.81 | 0.89 | 1.07 | 1.16 | 0.81 |
| Root Mean Square Comparison $\sqrt{\frac{\Sigma(\text{Actual}-\text{Predicted})^2}{N}}$ | | 0.23 | 0.23 | 0.33 | 0.10 | 0.19 | 0.23 |

Mean of Dependent Variable (Autos/HH) = 1.17 for 17 zones

*To illustrate the use of this table, the number 1.17 refers to the number of autos per household predicted by the independent variable Median Household Income.

were the counties in the New York Metropolitan Area, except for New York City which was one reading. Each point was of equal weight, since the desired result was a test of rates and not of totals. Auto ownership data were taken from state registrations. The best-fit linear regression line was derived for 1950 for each independent variable and combinations of the variables. These relationships were then used to estimate autos per household for 1960 and compared to the actual figures for that year. The results of the comparison are given in Table 5.

The results indicate that the combination of median household income and a measure of gross density yielded the smallest root mean square error[1], 0.10, or 8.5 percent. The next best combination of variables produced an error of almost twice this magnitude, 0.19. The variable persons per household produced the worst results, a root mean square error of 0.33, or 28 percent. Selected graphs from this analysis for 1950

TABLE 6

ESTIMATING AUTO OWNERSHIP, 1950

| Dependent Variable | Independent Variable(s) No. | Description | Equation | R Coeff. of Correlation | S Std. Error of Estimate | $S/\bar{X}_1$ (in percent |
|---|---|---|---|---|---|---|
| Auto Ownership ($X_1$) Autos/Household | ($X_2$) | Median household income (000) | $X_1 = 0.05098 X_2 + 0.603$ | 0.14 | 0.21 | 22 |
| | ($X_3$) | % (1 & 2) units/structures | $X_1 = 0.00958 X_3 + 0.231$ | 0.90 | 0.09 | 9 |
| | ($X_4$) | Persons per household | $X_1 = 0.294 X_4 - 0.098$ | 0.33 | 0.20 | 21 |
| | ($X_2$)<br>($X_3$) | Median household income (000)<br>% (1 & 2) units/structures | $X_1 = 0.075 X_2 + 0.00965 X_3 - 0.11$ | 0.92 | 0.09 | 9 |
| | ($X_2$)<br>($X_3$)<br>($X_4$) | Median household income (000)<br>% (1 & 2) units/structures<br>Persons per household | $X_1 = 0.075 X_2 + 0.00965 X_3 - 0.0001 X_4 - 0.11$ | 0.92 | 0.09 | 9 |
| | ($X_2$)<br>($X_4$) | Median household income (000)<br>Persons per household | $X_1 = 0.077 X_2 + 0.313 X_4 - 0.513$ | 0.37 | 0.20 | 21 |
| | ($X_3$)<br>($X_4$) | % (1 & 2) units/structures<br>Persons per household | $X_1 = 0.0097 X_3 - 0.019 X_4 + 0.293$ | 0.90 | 0.09 | 9 |

[1]Root mean square error $= \sqrt{\frac{\Sigma(\text{Actual} - \text{Predicted})^2}{N}}$

TABLE 7

ESTIMATING AUTO OWNERSHIP, 1960

| Dependent Variable | Independent Variable(s) No. | Independent Variable(s) Description | Equation | R Coeff. of Correlation | S Std. Error of Estimate | $S/\bar{X}_1$ (in percent |
|---|---|---|---|---|---|---|
| Auto Ownership ($X_1$) Autos/Household | ($X_2$) | Median household income (000) | $X_1 = 0.167\ X_2 + 0.006$ | 0.52 | 0.22 | 19 |
| | ($X_3$) | % (1 & 2) units/structures | $X_1 = 0.011\ X_3 + 0.31$ | 0.88 | 0.13 | 11 |
| | ($X_4$) | Persons per household | $X_1 = 0.611\ X_4 - 0.94$ | 0.66 | 0.20 | 17 |
| | ($X_2$)<br>($X_3$) | Median household income (000)<br>% (1 & 2) units/structures | $X_1 = 0.086\ X_2 + 0.010\ X_3 - 0.207$ | 0.91 | 0.11 | 9 |
| | ($X_2$)<br>($X_3$)<br>($X_4$) | Median household income (000)<br>% (1 & 2) units/structures<br>Persons per household | $X_1 = 0.10\ X_2 + 0.013\ X_3 - 0.284\ X_4 + 0.446$ | 0.93 | 0.10 | 8 |
| | ($X_2$)<br>($X_4$) | Median household income (000)<br>Persons per household | $X_1 = 0.10\ X_2 + 0.498\ X_4 - 1.25$ | 0.72 | 0.19 | 16 |
| | ($X_3$)<br>($X_4$) | % (1 & 2) units/structures<br>Persons per household | $X_1 = 0.013\ X_3 - 0.17\ X_4 + 0.75$ | 0.88 | 0.13 | 11 |

and 1960 indicate the best-fit straight line relationships for the variables (Figs. 14-16). The equations for these graphs are given in Tables 6 and 7.

The graph of autos per household vs percent (1 + 2) units/structure (Fig. 14) shows an almost constant slope between these two variables with an increasing Y intercept with time. If a methodology is to be developed using this independent variable as a predictor of autos, then the (increasing) intercept factor must be established for the future. The graph of persons per household vs autos per household (Fig. 15) reveals the unstable

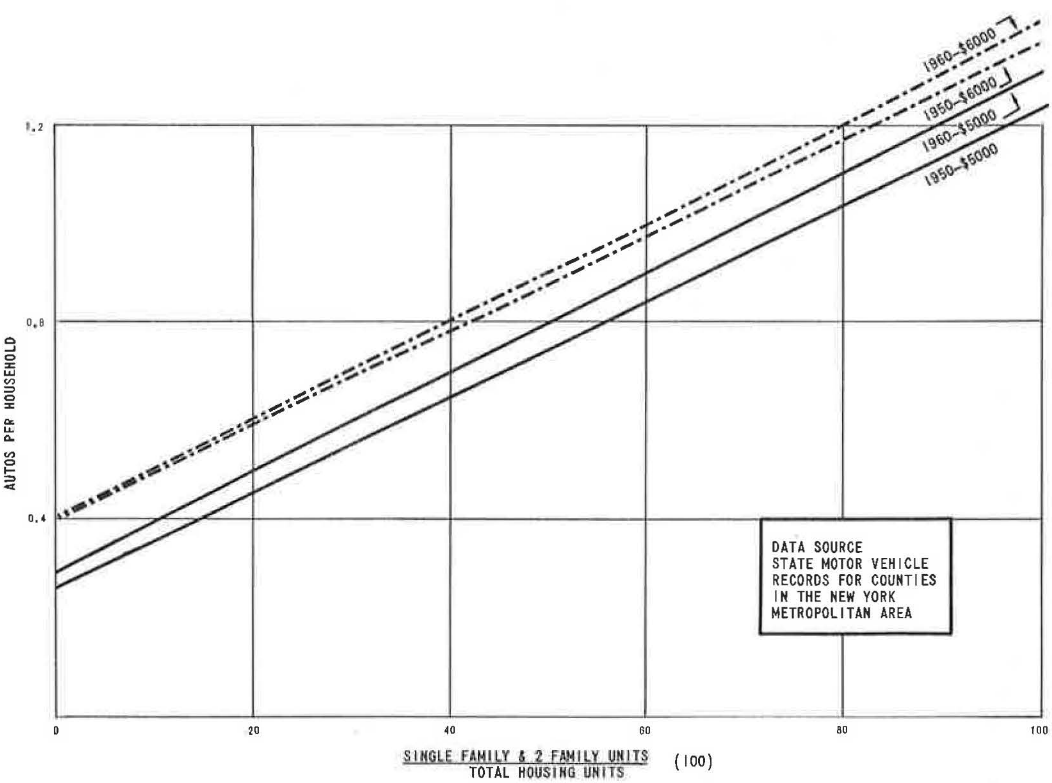


Figure 16. Autos per household vs percent (1 & 2) units per structure stratified by household income for the years 1950 and 1960.

relationship between the two variables. A decline of auto availability from 1950 to 1960 for 1 and 2 person households is indicated from the graph, which seems to be unreasonable. The graph of percent (1 + 2) units/structure (Fig. 16) stratified by median household income shows excellent correlation between the 1950 and 1960 curves for the $5000 and $6000 household income levels plotted.

## FORECAST CAPABILITY

A technique selected for forecasting must be based not only on the variables' and/or model's capabilities as predictors but also on the dependability of the estimates of these independent variables.

A purpose of this report is to isolate those techniques that would yield reasonable forecasts, regardless of the difficulties of estimating the independent variables. For example, it was pointed out that an estimation of household income is considered essential in all of the recommended techniques for forecasting autos, even though the estimation of household income is thought of as somewhat of an arduous task for the analyst. Nevertheless, the results of this report indicate a need for more concentrated efforts in an analysis of this variable. Guidelines for needed work in this area include:

1. Study of the changing shape of the income distribution curve (rate of change by income groups);
2. Relationship of change of median income to change in each income group; and
3. Relationship of household income vs auto availability (are the rates constant over time or are they changing?).

### Need for Evaluation

Perhaps the greatest need pointed out by this report is that of a continued evaluation of procedures for estimating auto availability. Data are needed for an area for two points in time to establish whether the procedures currently in use produce acceptable results.

The trip generation procedures and modal split models in use today are very much dependent on a measure of auto availability. Measuring the reliability and sensitivity of the techniques used in forecasting autos is essential for the effective use of these procedures.

## CONCLUSIONS

1. The use of residential density measures as the sole determinant for estimating autos is not valid for predictive purposes. Autos forecast by this procedure will generally be significantly lower than the totals established by control totals.
2. Persons per household, when used as the only parameter, is not a good indicator of autos per household. In the New York Metropolitan Tri-State Region, autos per household have increased in the 10-year period between 1950 and 1960 while persons per household have decreased in this interval. This relationship has shown a positive slope in other areas (autos per household rising with an increasing persons per household), which indicates an unstable relationship between these two variables.
3. Household income may be used as the single independent variable for forecasting autos per household. Care must be exercised in using the relationship of average income changes leading to average auto availability changes since a substantial change in residential density for a zone will yield auto availability rates significantly different from those rates predicted by income alone. If the incremental growth of residential development forecast does not approximate the density configuration already in place, then the preferred methodology for forecasting autos per household is the use of the combined effect of household income and residential density.
4. The recommended procedure for establishing control totals for autos available is to develop a trend of persons per auto by county in conjunction with the setting up of maximum and minimum limits for persons per auto.

ACKNOWLEDGMENT

The preparation of this report was financed in part through federal funds made available by the U.S. Bureau of Public Roads, a federal grant from the Urban Renewal Administration of the U.S. Department of Housing and Urban Development under the planning assistance program authorized by Section 701 of the Housing Act of 1954 as amended, and in cooperation with the states of Connecticut, New Jersey and New York.

## REFERENCES

1. Hoch, Irving. Automobile Registration Forecast. CATS Paper 3.6.2.00, Feb. 1959.
2. Cope, E. M., and Mundy, A. R. 139 Million Drivers in 1980. Public Roads, Vol. 33, No. 4, pp. 68-79, Oct. 1964.