UCI-ITS-WP-77-1

A Disaggregate Model of Auto-Type Choice

UCI-ITS-WP-77-1

Charles A. Lave [1]
Kenneth Train [2]

[1] School of Social Sciences and Institute of Transportation Studies
University of California, Irvine

[2] Cambridge Systematics, Inc.
Berkeley, California

January 1977

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

In press for *Transportation Research*

ACKNOWLEDGEMENTS

We wish to acknowledge the financial support of the U.S. Department of Energy; and the ideas, encouragement and assistance of Philip Patterson. The basic survey data were provided through the generosity of J. K. Pollard of the Transportation Systems Center and Anton Morton of Arthur D. Little, Inc. Data on destination charges were provided by the various auto companies.

A DISAGGREGATE MODEL OF AUTO-TYPE CHOICE

Charles A. Lave and Kenneth Train*

Abstract

Previous models of auto-type choice have not been able to disentangle very much of the structure of the household's auto-choice decision: the models assumed that very few auto characteristics affect choice, and often these few parameters were estimated with low precision. Hence the models had only limited use in forecasting the effects of government policies to influence transportation energy consumption. The present paper introduces a multinomial logit model for the type of car that households will choose to buy. The model includes a large variety of auto characteristics as explanatory variables, as well as a large number of characteristics of the household and the driving environment. The model fits the data quite well, and all of the variables enter with the correct signs and plausible magnitudes.

A DISAGGREGATE MODEL OF AUTO-TYPE CHOICE

Charles A. Lave and Kenneth Train

Introduction

This paper presents a disaggregate model of the household's car buying decision: what type of car will be bought, choosing among the great number of new cars available. Such a model has very direct relevance for U.S. energy planning: automobiles consume about a quarter of our energy; there are many cars available which would reduce this energy consumption by more than 50%; and hence it is of interest to know the sensitivity of car choices to the various possible government policy instruments. Although there have been many aggregate models in this area, the model presented here is the first one with a rich enough specification to be able to explore the consequences of alternative policies.

The model is estimated on a stratified random sample of 1976 new car buyers in seven cities. The universe of new car types is classified into ten size/price categories. Multinomial logit analysis is used to estimate the probability of choice among these ten categories, as a function of a variety of car characteristics, household characteristics, and characteristics of the driving environment.

The resultant model fits the data quite well, and all of the variables enter with the correct signs and reasonable significance levels.

Preliminary sensitivity checks indicate that changes in the policy variables influence market-shares in the expected direction, and with plausible magnitudes.

Previous Research

All previous econometric models of auto-type choice have used aggregate data. To our knowledge, a total of six such models have been estimated: Chamberlain (1974), Chase Econometrics (1974), Energy and Environmental Analysis, Inc. (1975), Difiglio and Kulash (1976), Ayres *et al*. (1976), and Wharton (1977). These models are comprehensively reviewed in Train (1978).

The primary limitation of these aggregate econometric models is that they do not include as explanatory variables the whole array of automobile characteristics which affect consumers' choices of type and number of automobiles to own. The only automobile characteristics they use as explanatory variables are price and fuel economy. Other characteristics, such as weight, external dimensions, passenger space, horsepower, and so on, are not included in the models.

The omission of these variables was not an oversight on the part of the researchers using the aggregate econometric approach: inherent problems with the approach make inclusion of these variables difficult, if not impossible.

The first problem is that the automobile characteristics do not vary substantially over time or across regions. The weight, horsepower, and so on, of new automobiles in each class are the same for all regions of the United States for a given year. Furthermore, these characteristics do not change much from one year to the next. However, in order to estimate the effect of automobile characteristics on consumers' buying patterns, some variation in the characteristics must exist. When little variation exists in explanatory variables, they cannot be expected to explain much variation in the dependent variables.

This problem of insufficient variation in automobile characteristics was encountered by the aggregate model-builders even though they included only the characteristics price and fuel economy. If insufficient variation exists in an independent variable, then the standard error for the estimated coefficient will be very large. The aggregate models evidence this problem.

Chamberlain obtained an extremely large standard error for the variable she used as a proxy for auto price (the standard error was more than 17 times the size of the coefficient in one equation). Difiglio and Kulash also obtained fairly large standard errors for their price variable, stating that "the estimated own-price coefficient for medium car shares and large car shares are only significant at a 25% and 40% level of error respectively." Other models leave some important variables out of some of the equations, indicating that their standard errors must have been too large to allow inclusion. For example, the model of Chase Econometrics does not include the price of compacts in the equation for the market share of compacts; the equation for the market share of subcompacts includes neither the price of subcompacts nor the fuel economy. Similarly, the EEAI model does not include fuel economy of compacts in the compact auto share equation or the price of subcompacts in the subcompact auto share equation.

A second problem confronting the builders of aggregate models is that, over time, automobile characteristics tend to vary together. Weight and external dimensions tend to move in tandem, with both increasing or decreasing at the same time. Similarly horsepower and price tend to be correlated. To obtain reliable estimates of the relationships between variables, however, each explanatory variable must vary fairly independently of the other explanatory variables. Otherwise, it is not possible to determine which variable's variation is explaining the variation in the dependent variables.

The only automobile characteristics which Difiglio and Kulash entered were price and fuel economy. Yet, with only these two variables, the problem of collinearity was encountered. Kulash (1975) stated the difficulty explicitly:

> Car prices, gasoline prices, and fuel economy all have a bearing on the overall cost of owning and operating a vehicle. But over the last 15 years, these three measures have all changed in highly interrelated ways, as reflected by simple correlation coefficients between them of .9 or greater. These interrelations make it difficult to use statistical techniques to separate the impact on sales of one price element relative to others. On the other hand, automotive fuel economy price proposals would influence various combinations of each of these elements, making it essential that the effect of each be isolated.

Difiglio and Kulash were forced to specify a "generalized" price combining of auto price, gas price, and fuel economy; and then entered this single variable into their model rather than estimating the separate effects of auto price, gas price, and fuel economy.

Because of these two problems (lack of sufficient variation over time and across regions in automobile characteristics, and large correlations among automobile characteristics) the possibility seems small of any aggregate econometric model being able to include a full array of automobile characteristics. Consequently, it seems that a disaggregate econometric approach should be attempted.

Description of the Data

Household Data: we utilized a stratified random sample of new car buyers, collected during the Summer of 1976 by Arthur D. Little, Inc., in seven cities--Atlanta, Buffalo, Chicago, Denver, Indianapolis, Los Angeles, and New Orleans. The stratification produced an equal number of households who had purchased small, medium, and large cars (classes 1-4/5-6/7-10; as explained below). Each home interview recorded the type of auto purchased and a variety of household characteristics. The means of some of the socioeconomic variables are shown in Table 1.

Automobile Data: information on various physical characteristics for each type of car was taken from the 1976 Automotive News Market Data Book. Price data for each car were constructed as the sum of A) sticker price, B) destination charges specific to each city, and C) taxes specific to each city. Destination charge information was obtained from the auto manufacturers. Tax information (registration fees, personal property taxes, local sales tax, and state sales tax) was obtained from various federal publications, and from individual cities.

Definition of Choice Categories: All of the makes and models of cars were classified into ten categories, with cars in each category

TABLE 1

Socioeconomic Characteristics of Sample

	Mean	Standard Deviation
Number of people in household who are 18 years and over	2.36	.92
Number of people in household who are 0-17 years old	.94	1.13
Number of licensed drivers in household	2.28	.96
Number of autos owned by household members	1.93	.72
Number of employed persons in household	1.68	.97
Location of home (1 = urban (2 = suburban (3 = rural	1.78	.48
Age*	3.55	1.35
Education*	2.78	1.00
Income**	4.06	1.22

* Coding shown in footnotes 4,5 of Table 2.

**1 = less than $7500; 2 = 7,500-9,999; 3 = 10,000-14,999; 4 = 15,000-24,999; 5 = 25,000-34,999; 6 = 35,000 and over.

chosen to be relatively homogeneous in size and price. The household is then choosing between these ten categories. We constructed a "representative" car for each category by taking a sales-weighted average of the characteristics of the cars in that category. The categories are summarized at the top of Table 2.

A Model of Auto-Type Choice

The model estimates the probability that a given household will choose to buy a new auto within a particular class of auto types (where the classes are defined as above), conditional upon the household's choice of a new auto over a used one. The model is thereby restrictive in two ways: (1) only the choice of new auto types is considered and (2) classes of autos rather than makes and models are the choice alternatives. The former restriction was imposed by the data: only a sample of new auto purchasers was available for analysis. The latter restriction was imposed by computer capacity: estimation of a model with a separate alternative for each make and model would entail more computer space than is available.

The model is multinomial logit (MNL) with the probability of choosing a new auto within class i defined as:

$$P_i = \frac{e^{\beta' z(x_i, s)}}{\sum_{j=1}^{10} e^{\beta' z(x_j, s)}} \qquad (1)$$

where P_i is the probability of choosing a new auto in class i;

x_i is a vector of attributes of autos of class i (e.g., cost, weight, horsepower, etc.);

s is a vector of attributes of the household;

z is a vector-valued function of x_i and s; and

β is a vector of coefficients.

The term $\beta'z(x_i,s)$ is called the "representative" utility of auto class i. The MNL model and a method for estimating the model parameters are discussed in McFadden (1973).[1]

Table 2 presents the model of new auto class choice. The independent variables are the elements of $z(x_i,s)$ in (1).

The second and third columns record the estimates and t-statistics, respectively, of the elements of β. The independent variables are fairly complex and require explanation.

Specification of the Model

In broad termsthe probability of choosing some given car type can be thought of as depending on three kinds of explanatory variables

P_i = F(household characteristics; car characteristics; environment)

Breaking down each of these main categories still further, we get

P_i = F(income, education, age, family size, # cars, # miles driven; car price, MPG, horsepower, interior size, weight; local price of gasoline)

And these are, in fact, the variables which are included in the final model. But the actual specification of the model is a good deal more complex for two reasons: first, the relation of many of the variables to the choice-probabilities is quite non-linear, and hence a number of "squared" terms are included in the equation; second, since the <u>objective</u> attributes of cars do not vary across households (hence there is no variance in these explanatory variables across the observations), we must use <u>subjectively-</u>perceived car characteristics in the equation. That is, since two households looking at, say, a Class 7 car are both faced with identical objective car characteristics, we must look at the subjective car characteristics in order to obtain the variance necessary to estimate our equation.

We assume that subjectively-perceived attributes result from the interaction of household characteristics and car characteristics. For example, a low income household will perceive an $8000 price tag in quite different manner than would a high income household. Thus, the subjective cost can be approximated by using the variable "cost/income." Likewise we have reason to suspect the following other interactions between household characteristics and car characteristics: family size and car size, age and performance (acceleration), age and weight, income and weight. Thus the equation to be estimated then becomes

$$P_i = F(\text{income, family size, \# cars owned, \# miles driven;}\ \frac{\text{cost}}{\text{income}}, \frac{\text{\# seats}}{\text{\# people}}, \frac{\text{gas price}}{\text{MPG}}, \text{weight} \times \text{age, weight} \times \text{education, performance} \times \text{age})$$

<u>Results of the Estimation</u>. Table 2 shows the estimated coefficients. We discuss each variable in turn.

TABLE 2

A Model of Auto-Type Choice

Alternatives:

Class 1: Subsubcompact (e.g., VW Beetle)
Class 2: Sports Cars (e.g., Porsche)
Class 3: Subcompact-A (e.g., Vega)
Class 4: Subcompact-B (e.g., Mustang II)
Class 5: Compact-A (e.g., Valiant)
Class 6: Compact-B (e.g., Granada)
Class 7: Intermediate (e.g., Cutlass)
Class 8: Standard-A (e.g., Impala)
Class 9: Standard-B (e.g., Riviera)
Class 10: Luxury (e.g., Cadillac)

Model: Multinomial Logit, Fitted by the Maximum Likelihood Method

Independent Variable (The variable takes the described value in the alternatives listed in parentheses and zero in non-listed alternatives).	Estimated Coefficient	T-Statistic
Initial Auto cost divided by household income[1] (1-10)	-9.35	3.23
Initial Auto cost divided by household income, quantity squared[1] (1-10)	2.61	2.81
Auto operating cost per mile[2] (1-10)	- .350	1.58
Weighted seats[3] (1-10)	.558	1.79
Dummy variable for whether the household owns more than two autos (1-4)	.815	3.53
Dummy variable for whether the household's income exceeds $25,000/year (9-10)	.474	1.52
Number of persons in household (1-3)	.590	2.45
Number of persons in household (4-10)	.378	1.44
Vehicle miles traveled per month by the household (8-10)	.000348	1.55
Auto weight times age of respondent[4] (1-10)	.0527	2.58

TABLE 2 (continued)

Independent Variable	Estimated Coefficient	T-Statistic
Auto weight times age of respondent, quantity squared[4] (1-10)	- .0000545	1.39
Auto weight times education of respondent[5] (1-10)	.0282	0.971
Auto weight times education of respondent, quantity squared[5] (1-.0)	- .000152	1.94
Auto "performance" times age of respondent[6] (1-10)	1.12	1.22
Auto "performance" times age of respondent, quantity squared[6] (1-10)	- .0230	1.38
Dummy variable for sports/specialty (2)	1.80	.621
Dummy variable for subcompact A (3)	-3.46	.340
Dummy variable for subcompact B (4)	.273	.148
Dummy variable for compact A (5)	1.62	.718
Dummy variable for compact B (6)	1.75	.758
Dummy variable for intermediate (7)	3.10	.838
Dummy variable for standard A (8)	1.86	.418
Dummy variable for standard B (9)	2.82	.584
Dummy variable for luxury (10)	4.67	.773
Likelihood ratio index	.1257	
Log likelihood at zero	-1246.	
Log likelihood at convergence	-1089.	
Percent correctly predicted	24.77	
Sample size	541	

[1]Initial auto cost is sticker price, taxes, and destination charges, in dollars. Income is in dollars per year.

TABLE 2 (continued)

[2]Auto operating cost per mile is in cents per mile and is defined as the price of a gallon of gasoline divided by the auto's miles per gallon. Price for each city from the Oil and Gas Journal.

[3]Weighted seats is a variable which gives a weight of one for each seat in an auto up to the number of persons in the household and a weight of one-half for each seat in an auto more than the number of persons in the household. Thus, a household with three members will have a value of 2 for auto classes with 2 seats, a value of 3 for auto classes with 3 seats, and a value of 3.5 for auto classes with 4 seats.

[4]Weight is in hundreds of pounds. Age is coded as follows:

1 = teens
2 = twenties
3 = thirties
4 = forties
5 = fifties
6 = sixties and above.

[5]Weight is in hundreds of pounds. Education is coded as follows:

1 = high school not complete
2 = high school complete
3 = one to three years of college
4 = four or more years of college.

[6]"Performance" is defined as horsepower divided by weight, which is an excellent predictor of relative acceleration ability. See footnote 4 for definition of weight and age.

Dividing auto cost by income allows the importance of auto cost to vary across households according to their income; and we include two terms (one squared and one not) to allow for a non-linear relation between initial cost and "representative" utility. The estimated coefficients indicate that, over all the combinations of initial cost and income observed in the data, an increase in initial cost makes the "representative" utility of an auto class lower, and an increase in income makes the "representative" utility higher. That is, a rise in initial cost for an auto class decreases the probability that an auto in that class will be chosen; and a rise in income increases the probability of choosing an expensive auto class.

Auto operating cost per mile has a negative sign, indicating that an increase in the MPG of a given class will increase the "representative" utility of that class, and hence increase the probability of choosing that class. An increase in the price of gasoline decreases the "representative" utility of all auto classes, yet it increases the probability of choosing fuel-efficient autos <u>relative</u> to inefficient autos.

The positive sign for the coefficient of the weighted seats variable indicates that an increase in the number of seats, in autos of a particular class, results in an increase in the probability of that class being chosen. (See note 3 of Table 2 for the definition of this variable.)

It was thought that households which own many autos are more likely to choose a small auto when they buy an extra one (they can afford to have special purpose cars). The dummy variables for whether the household owns more than two autos enters the "representative" utility of the first four auto classes but not that of other auto classes. The positive estimated coefficient confirms this hypothesis.

A dummy variable for incomes greater than $25,000 per year enters the "representative" utility of auto classes 9 and 10 only. The positive sign for this variable indicates that high income people are more likely to choose large, expensive autos independent of the fact (expressed in the first two variables of the model) that high income people care less about cost. The effect being captured in this dummy variable is perhaps a prestige effect, and the positive coefficient would be expected.

The number of persons in the household enters the "representative" utility of classes 1 and 3 with a coefficient of .58 and classes 4 through 10 with a coefficient of .37. This indicates that increasing household size decreases the probability of choosing a sports/speciality auto (class 2) independent of the issue of how many seats are in the auto (which is already represented by the weighted seats variable.) This effect could perhaps be called a "family man" effect.

Vehicle miles traveled per month enters the "representative" utility of classes 8-10. The positive estimated coefficient indicates that, as expected, households which drive a lot tend to choose large autos.

To examine the effect of auto weight on choice probabilities we used two interactive variables--a weight-age interaction term, and a weight-education interaction term; both in non-linear form.

The weight-education interaction has a complex, though quite plausible pattern; and it is shown in Table 3. The numbers in the table can be regarded as relative utilities. Reading across the rows we see that, all else held equal, people with little education (row 1) place a high value on big cars: the greater the car weight, the greater the "representative" utility. Whereas for the most educated group, row 4, the highest utility is associated with cars in the 2000-3000 pound range; and the value of larger cars drops off quite rapidly. At intermediate levels of education, the pattern of size preferences lies between these two opposite taste patterns.

We can examine the weight-age interaction by calculating a similar table. It shows a simpler pattern. For any given age class, all else held equal, there is a positive effect between the choice probability and the weight of the car. But the degree of positive slope increases greatly with age: an increase in car weight has a much stronger positive influence on old people than it does on young people.

The "performance" variable approximates the relative acceleration of different cars; and again we have allowed for a possible non-linear relationship. The results seem quite reasonable: for any given age class of people, performance has a positive slope; that is, an increase in the performance of autos in a particular class increases the probability that the class will be selected. But the degree of positive slope goes down with increasing age: young people are much more influenced by an increase in performance than older people.

In summary, the flexible form allowed by our various non-linear interaction terms has enabled us to discover a number of complex, but highly plausible interactions between household characteristics and car choices.

TABLE 3

The Interaction of Auto Weight and Respondent

Education in "Representative" Utility

	Auto Weight (in hundreds of pounds)			
Education Level (coded)	20	30	40	50
1. (High school not complete)	.503	.709	.884	1.03
2. (High school complete)	.885	1.45	1.28	1.30
3. (One to three years of college)	1.14	1.31	1.91	.810
4. (Four or more years of college)	1.28	1.20	.621	- .440

Each element in the table is

$$[.0282 * \text{WEIGHT} * \text{EDUCATION}] - [.000152 * (\text{WEIGHT} * \text{EDUCATION})^2]$$

where WEIGHT and EDUCATION take the values given in the row and column heads.

A dummy variable was included for each auto class[2] to capture the common effects, on all consumers, of variables which are not included in the model. In particular, these variables capture the effects of the relative prestige and comfort of the autos in each class. Since these variables cannot be measured, they cannot enter the model directly.

The "Auto Operating Cost" Variable. Auto purchase price was divided by income to allow the importance of cost to vary with income, but auto operating costs were not divided by income. The reason we did not treat the two kinds of costs in a parallel fashion was that we could not do so. A variety of alternative specifications to accomplish this were tried, but they simply did not yield significant results.

In particular, entering "auto cost per mile divided by income" rather than "auto cost per mile" decreased the log likelihood of the model; the t-statistic for this new term was only 0.49. This indicates that the importance of auto cost per mile does not vary with the inverse of income.[3]

It was thought that, perhaps, households consider auto cost per month when choosing an auto rather than the cost per mile. That is, the importance of auto cost per mile varies with the vehicle miles traveled by a household. To test the proposition, the model of Table 2 was estimated with "auto cost per month" replacing "auto cost per mile," the former being defined as the product of the latter and the vehicle miles traveled by the household. This model again attained a lower log likelihood than the model of Table 2, and the t-statistic for the coefficient of "auto cost per month" was only 1.25.

This ratio is constant for any C which contains i and k (including, of course, the set containing only i and k) and any attributes of alternatives (except i and k) in C.

The model of Table 2 seems particularly likely to violate the IIA assumption. For example, it seems doubtful that the probability of choosing a class 9 auto over a class 3 auto remains the same whether or not the possibility of owning a class 8 auto exists. If the probability is not constant, then the property of independence from irrelevant alternatives does not hold and the MNL model is inappropriate.

Two factors mitigate the severity of this problem. First, it has recently been found that if alternative-specific constants are included as explanatory variables when an MNL model is being estimated, then these constants partially "correct" for violations of the IIA property. That is, if a MNL model is estimated in an application in which the IIA property does not hold, then the estimated values of the alternative-specific constants are automatically adjusted partially to correct for this problem. (See Train (1977) for a full discussion.) Since the model of Table 2 was estimated with an alternative-specific constant for each auto class, the problem of IIA is less severe than it would be if such constants were not included.

The second reason to discount the problem of the IIA property is based on empirical testing. McFadden, Tye, and Train (1976) developed methods for testing the IIA property of MNL models. These tests were applied to the model of Table 2, and in all cases the model passed. That is, the hypothesis that the IIA property holds was not rejected with any of the tests, indicating that perhaps the IIA property does not present

Finally test whether the importance of auto cost per mile varies with both vehicle miles traveled and income, the model of Table 2 was estimated with "auto cost per month divided by income" replacing "auto cost per mile." This model also attained a lower log likelihood than the model of Table 2, and the t-statistic for the coefficient of "auto cost per month divided by income" was only 0.18.

Problems with the Model

Several problems limit the plausibility and consequently the applicability of the model. First is the problem of the restrictions implied by multinomial logit (MNL). The MNL model, expressed in equation (1), assumes that the ratio of the probabilities of choosing any two alternatives is independent of the availability or attributes of other alternatives. This property is called the independence from irrelevant alternatives (IIA) property and can be demonstrated as follows. Consider the ratio of the probability of a person choosing alternative i to that of choosing alternative k, given that set C of alternatives is available:

$$\frac{P_i(C)}{P_k(C)} = \frac{e^{\beta' z(x^i,s)} \Big/ \sum_{j\in C} e^{\beta' z(x^j,s)}}{e^{\beta' z(x^k,s)} \Big/ \sum_{j\in C} e^{\beta' z(x^j,s)}}$$

$$= \frac{e^{\beta' z(x^i,s)}}{e^{\beta' z(x^k,s)}}$$

a problem in this application. (The details of these tests are available from the authors.) However, it must be mentioned that the power of these tests seems to be low. Consequently, it is quite possible that violations of the IIA property exist in this application but were not detected by the tests.

Another problem with the model concerns the method by which the model was estimated. As mentioned above, a stratified sampling procedure was used to obtain the households upon which the model was estimated. The stratification was based on auto size so as to obtain an equal number of households who had purchased small, medium, and large cars. As a result, households were selected on the basis of their chosen auto, rather than on the basis of some variable which is exogenous to the decision process being modeled.

Manski and Lerman (1976) show that the estimation method given in McFadden (1973), which was developed for use with exogenously chosen samples (and was used for estimating the model of Table 2) is not appropriate if the sample was chosen on the basis of the household's chosen alternatives. They demonstrate two differences between the maximum likelihood estimator which is appropriate when the sample is choice-based and the McFadden estimator. First, the estimated alternative-specific constants are different with the two estimators (though all the other estimated parameters are the same). Second, the estimated standard errors of all the parameters are different with McFadden's method than with the method which is appropriate for choice-based samples.

As a result of these findings, the estimated alternative-specific constants and the t-statistics in Table 2 should be viewed with caution.

Unfortunately, a software routine with the appropriate correction for chocie-based samples was not available at the time the model was estimated.

A last, and fundamental, problem with the model lies in the fact that most of the auto attributes do not vary over the population. That is, the weight, size, horsepower, number of seats, and so on, of a particular auto type is the same for all households in all parts of the country.[4] Because of this, auto attributes cannot enter directly into a model which has alternative-specific constants for each alternative (since the attributes are collinear with the constants). The only way an auto attribute which does not vary over the population can enter the model is by interacting with some characteristic of the household (which do vary over the population, of course) and/or by removing one or more of the alternative-specific constants. Both of these approaches have drawbacks.

First consider the approach of removing one of the constants. Say that auto weight is included as an explanatory variable and that the constant for auto class 10 is removed to allow the weight variable to enter. The estimated coefficient of the weight variable would be exactly equal to the constant which had previously been estimated for class 10 autos divided by the difference between the weight of class 10 and class 1 autos (since the constant for class 1 autos is normalized to zero). All of the constants for the other auto classes would be adjusted such that the sum of the new constant and the weight term would equal the previously-estimated constant. All the other estimated parameters would remain the same with this change in model specification.

Two points are important with regard to this approach. First, the coefficient to the weight variable can be calculated without actually entering it. Second, the coefficient of the weight variable would be different depending on which alternative-specific constant is removed.

This discussion shows that entering an auto attribute by removing one constant tells us nothing about the value of that attribute to consumers. It is simply a different, but equivalent, method for entering an alternative-specific constant.

If two or more alternative-specific constants are removed and one attribute is entered, it is possible that the estimated coefficient of this variable contains meaningful information. The more constants that are removed when one attribute is entered, the more information might be contained in the estimated coefficient. However, two problems occur in this regard. First, as mentioned above, the alternative-specific constants are useful in correcting for violations of the IIA property. If they are removed (and an equal number of attributes are not entered), then violations of the IIA property could cause important problems for the model. Second, the auto type choice model contains only nine alternative-specific constants (the tenth is normalized to zero). Consequently, even if all the constants were removed, few auto attributes could be entered with meaningful results.[5]

Because of the drawbacks involved with adding auto attributes by eliminating alternative-specific constants, the approach was adopted of interacting the attributes with the socioeconomic characteristics of the household. This approach allows the auto attributes to enter the model without eliminating the constants, but it has other drawbacks. The variation which occurs in an explanatory variable, that is defined as an interaction of an auto attribute with a socioeconomic variable, is entirely due to variation in the socioeconomic variable. Consequently, it is questionable whether the estimated coefficient of such an explanatory variable contains any information about the effect of the attribute on the choice of the decision-maker, rather than the effect of the decision-maker's tastes as captured by his socioeconomic characteristics.

The basic question is simply: how much can be learned about the effect of changes in auto characteristics from a sample in which such attributes do not vary? No simple, definitive answer to that question is available at the moment; and hence the estimated coefficients must be used with some degree of caution.

Implications of the Model

We have begun the lengthy process of exploring alternative policy scenarios with the aid of the model, and have two preliminary results: the effects of an increase in gasoline taxes, and the effects of an excise tax on larger cars. Utilizing the model of Table 2, we calculate each household's individual probabilities of buying a new car in each of the ten car classes. When these probabilities are summed across all households in the sample we obtain the first column of Table 5, the initial market shares of new car types.

To calculate the effects of a change in gasoline prices we make a 10% increase in the variable "auto operating cost per mile" for each separate household, and sum across households. These results are shown in the second column. The strong gainers are the subcompacts, classes 1 and 3; and the biggest losers are the intermediate and large cars, classes 7 to 10. The market share of compacts, classes 5 and 6, remains about the same, in apparent terms. But what is happening is that the increase in gasoline prices is shifting the entire profile of car choices downward. That is, the results are compatible with the notion that the tax causes people to move down one or two classes, rather than making a major shift: some luxury car buyers shift down to class 9; some intermediate buyers shift down to class 6; the net down-shift from the compacts is balanced by the net down-shift into the compacts; and more people are piling up at the bottom end, in the smallest cars.

The results of a 10% excise tax on intermediate, large and luxury cars is shown in column 3. The excise tax has the smallest effect on the shares of luxury cars, and expensive-large cars; presumably because people who buy such cars have little sensitivity to purchase price in the first place. The greatest change is in the sale of intermediate cars, with most of the "defectors" moving to class 2 sports cars, which seems plausible.

The gasoline tax increase has much broader effects across car classes. But neither the excise tax change, nor the gasoline tax change, causes large changes in market shares; and this is consistent with the observed change in market shares over the past four years. Much larger taxes than these will be necessary to cause substantial movements.

It must be pointed out that our model is for the new car market only, and obviously there will be some feedback effects with the used car market. For example a significant excise tax on new large cars can be expected to increase the likelihood that consumers will decide to

TABLE 5

Market Share Sensitivities of New Cars

Class	Market Share (%)	Market Share After 10% Gas Tax (%)	Market Share After 10% Excise Tax on Class 7-10 Purchase Price
1	11.1	14.2	11.1
2	3.1	4.7	4.8
3	15.3	18.2	15.9
4	5.2	5.6	5.6
5	19.4	19.2	20.2
6	11.1	10.6	11.2
7	19.0	15.5	16.3
8	3.1	2.3	2.7
9	9.2	7.3	9.0
10	3.3	2.1	3.2

Note: Quota sampling was used to generate equal samples of classes 1-3/4-6/7-10.

keep their existing large cars for a longer period of time. And any general increase in new car prices is going to operate to somewhat reduce the total sales of new cars.

The total effects of the two kinds of taxes will also be somewhat different if we consider both the new and used car markets simultaneously. We would expect that the gas tax increase will operate to push people into smaller cars, in both markets. While the down-shift effects of an excise tax on large new cars will tend to be offset by the decision of existing large car owners to keep their cars for a longer period of time. Thus the gasoline tax approach would appear to have much broader effects.

Footnotes

*The authors, listed alphabetically, are in the economics department of the University of California, Irvine, and at Cambridge Systematics, Inc., Berkeley, California, respectively.

[1]The model was estimated as if the sample were drawn exogenously, whereas the sample was actually choice-based. This inconsistency is discussed below, in "Problems with the Model."

[2]A dummy variable was not included for auto class 1 since doing so would produce an identification problem. Not including the variable is equivalent to normalizing the representative utility function such that the coefficient of this variable is zero.

[3]The correct specification test is to include both variables in a more general model and test the hypothesis that the coefficient of the "auto cost divided by income" variable is zero. This procedure was not possible, however, because the capacity of the computer was insufficient to allow estimation of the model of Table 2 with an extra variable added.

[4]Initial cost varies because of differences in taxes and destination charges. Operating cost varies because of differences in the price of gasoline.

[5]MNL models have been estimated that include several attributes of the alternatives which do not vary over the population; destination choice models in urban travel demand analysis are an example. However, these models describe choice situations for which there are numerous alternatives (and alternative-specific constants are not included). For the model of auto type choice, there are only ten alternatives.

References

Ayres, R., et al. (1976), "Automobile Forecasting Models," prepared by International Research and Technology Corporation for the Office of Technology Assessment, U.S. Congress, Working Paper IRT-446-R.

Chamberlain, C. (1974), "A Preliminary Model of Auto Choice by Class of Car: Aggregate State Data," unpublished discussion paper, Transportation Systems Center, U.S. Department of Transportation.

Chase Econometric Associates (1974), The Effect of Tax and Regulatory Alternatives on Car Sales and Gasoline Consumption, NTIS Report No. PB-234622.

Energy and Environmental Analysis, Inc. (1975), Gasoline Consumption Model, unpublished report of EEAI, Arlington, Virginia, done for the Federal Energy Administration.

Difiglio, C. and D. Kulash (1976), Marketing and Mobility. Report of a panel of the Interagency Task Force on Motor Vehicle Goals beyond 1980; available through the Office of the Secretary of Transportation. Publishing Section, TAD-44.1, Washington, D.C.

Kulash, D. (1975), "Forecasting Long Run Automobile Demand," in Strategies for Reducing Gasoline Consumption through Improved Motor Vehicle Efficiency. Transportation Research Board Special Report #169.

Manski, C. and S. Lerman (1976), "The Estimation of Choice Probabilities from Choice-Based Samples," in press for Econometrica.

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in Frontiers in Econometrics, P. Zarembka, ed., Academic Press, New York.

McFadden, D., W. Tye, and K. Train (1976), "Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model," Working Paper 7616, Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley.

Train, K. (1977), "Auto Ownership and Mode Choice within Households," Sloan paper No. SL-7703, Chapter 2, Department of Economics, University of California, Berkeley.

Train, K. (1978), "Consumers' Reponse to Fuel Efficient Vehicles: A Critical Literature Review," unpublished manuscript.

Wharton Econometric Forecasting Associates (1977), An Analysis of the Automobile Market: Modeling the Long Run Determinants of the Demand for Automobiles. Volumes 1, 2, 3. Prepared for U.S. Department of Transportation/T.S.C.