OST - 77-25

REPORT NO. DOT-TSC-OST-77-25

STUDY DESIGN FOR A METHOD OF PROJECTING VEHICLE MILES OF TRAVEL

F.T. Rabe

ENVIRONMENTAL IMPACT CENTER, INC.
55 Chapel Street
Newton MA 02158

August 1977

FINAL REPORT

Prepared for:

U.S. DEPARTMENT OF TRANSPORTATION
OFFICE OF THE SECRETARY
Office of the Assistant Secretary for Systems
Development and Technology
Office of Systems Engineering
Washington DC 20590

Technical Report Documentation Page

1. Report No.	DOT-TSC-OST-77-25
2. Government Accession No.	
3. Recipient's Catalog No.	
4. Title and Subtitle	STUDY DESIGN FOR A METHOD OF PROJECTING VEHICLE MILES OF TRAVEL
5. Report Date	August 1977
6. Performing Organization Code	
7. Author(s)	F.T. Rabe
8. Performing Organization Report No.	
9. Performing Organization Name and Address	Environmental Impact Center, Inc.* 55 Chapel Street Newton MA 02158
10. Work Unit No. (TRAIS)	OS714/R7508
11. Contract or Grant No.	DOT-TSC-10596
12. Sponsoring Agency Name and Address	U.S. Department of Transportation Office of the Secretary Office of the Assistant Secretary for Systems Development and Technology Office of Systems Engineering Washington DC 20590
13. Type of Report and Period Covered	Final Report July 1975 - Sept. 1975
14. Sponsoring Agency Code	
15. Supplementary Notes	*Under contract to: U.S. Department of Transportation Transportation Systems Center Kendall Square Cambridge MA 02142

16. Abstract

Vehicle miles of travel (VMT) by passenger automobiles is an important determinant of gasoline consumption, ambient air quality, highway safety, and personal and corporate financial conditions in the United States. Changing patterns and trends in VMT, therefore, have profound implications for energy conservation, environmental quality, and economic stability. Forecasts of likely future levels of VMT have become a central input to transportation policy analysis.

This report is an effort to assess the state of the art of VMT forecasting and map out strategies for extending it. The work included an inventory of data sources and a review of existing VMT models. Recommendations for long-, intermediate-, and short-range future research are included, along with estimates of effort and costs required for carrying out each possible research strategy.

17. Key Words	18. Distribution Statement
Travel forecasts, Short-, Intermediate- and Long-Term Strategies	DOCUMENT IS AVAILABLE TO THE U.S. PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	56	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

PREFACE

The transportation Energy Efficiency Program (TEEP) of the U.S. Department of Transportation has made development of an improved vehicle miles of travel (VMT) model one of its major objectives. The work described in this report, sponsored by the TEEP, is an effort to assess the state of the art of VMT forecasting and map out strategies for extending it. Included in the work is an inventory of data sources and a review of existing VMT models, along with recommendations for future research and cost estimates.

This report was prepared by the Environmental Impact Center, Newton, Massachusetts for the U.S. Department of Transportation (DOT), Transportation Systems Center (TSC).

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F

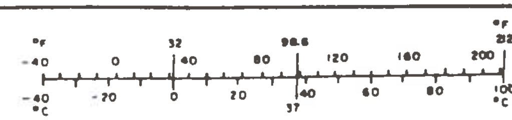

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION

1.1 GENERAL

Vehicle miles of travel (VMT) by passenger automobiles is an important determinant of gasoline consumption, ambient air quality, highway safety, and personal and corporate financial conditions in the United States. Changing patterns and trends in VMT, therefore, have profound implications for energy conservation, environmental quality, and economic stability. Forecasts of likely future levels of VMT have become a central input to transportation policy analysis.

At present, however, VMT forecasts are made with outmoded, simplistic analytic techniques. Most projections rely on historical trend analysis, an approach that can be useful when the data exhibit a stable pattern over time and only short-term extrapolation is required. Unfortunately, the historical stable growth in VMT was eroded by the Arab oil embargo and associated increases in gasoline price. Further, effective policy analysis requires long-term forecasts to evaluate long-range energy conservation and environmental protection strategies. While a few simple models of VMT, embodying some behavioral economic relationships, have been developed in the past several years, they contain numerous theoretical and empirical shortcomings that render their forecasts little better than standard trend extrapolation.

1.2 OBJECTIVES FOR VMT FORECASTS

Because of these limitations in the state of the art, together with the central significance of auto travel for public planning, the Transportation Energy Efficiency Program (TEEP) of the U.S. Department of Transportation, has made development of an improved VMT model one of its major objectives. The work reported here, sponsored by the TEEP, was an effort to assess the state of the art and map out strategies for extending it. In further research, new models of VMT will be specified, empirically tested, and added to the TEEP forecasting methodology.

The principal objective of the TSC VMT model is to provide estimates of national travel subject to variations in policy. However, because of the role of VMT in a broader systems analysis, several more detailed objectives can be identified. It would be desirable, for example, to estimate travel by class and age of automobile, since emissions, safety, user costs, and fuel consumption depend in part on these two variables. Disaggregation by purpose or location of trip -- journey to work versus leisure, urban versus intercity -- would also be appropriate since these travel characteristics influence the other elements of analysis and since travel sensitivities to other factors probably vary with the type of trip.

The extent to which these phenomena can be modeled depends on the quantity and quality of available VMT data. Without adequate measures of auto travel and the factors that influence it, no modeling or forecasting would be possible. More important, the character of the available data determines what hypotheses about VMT may be tested.

1.3 STUDY METHOD

An improved approach to forecasting VMT should therefore be based in equal measure on firm empirical evidence and sound theoretical specification. This work has included both an inventory of data sources and problems, presented in Section 2, and a review of existing VMT models and their theoretical adequacy, documented in Section 3. Ultimately, of course, theory and data must be merged coherently. Section 4 presents recommendations for further research, including short- and long-range strategies, in which both optimal theoretical structures and empirical support are identified. Finally, Section 5 provides estimates of effort, costs, and schedules for each proposed strategy.

2. DATA SOURCES AND PROBLEMS

2.1 GENERAL

The absence of data has been a major obstacle to a thorough model of auto travel. No direct measurement of VMT has ever been made; the available information consists of estimates from various sources. In addition, data on several important explanatory factors -- gasoline prices, fuel economy, and mass transit, for example -- are scanty at best. The problem is further complicated by the need for data on several levels of aggregation, building up to a nationally representative sample. Compilation of a comprehensive, internally consistent data base, therefore, represents a sizable task.

This inventory of data sources was intended to evaluate the feasibility of compiling the necessary data base. However, since there are several potential theoretical approaches to VMT modeling, no single source of information would necessarily satisfy all requirements. In this section, various sources of data, their scope and quality are described without attempting to tie theoretical and empirical elements of a modeling effort together. The latter is accomplished in Section 4.

Of necessity, most efforts were focused on data for VMT, rather than explanatory factors. Since VMT "data" are actually estimates, it was necessary to explore and understand the estimation procedures, some of which are rather convoluted, as well as to make an assessment of the accuracy of the data. Data on explanatory factors could then be matched to data on VMT, the dependent variable.

2.2 VMT DATA

Only three general sources of information on auto travel currently exist.* The Federal Highway Administration (FHWA) has

*NHTSA has initiated a program of auto travel measure through odometer readings for a selected sample of households which, when completed, should provide a limited amount of new information on VMT.

collected and published travel data in its annual report, Highway Statistics. In conjunction with the 1970 Census, the FHWA also sponsored the Nationwide Personal Transportation Survey, in which a sample of households was taken on travel behavior. Finally, a very large number of metropolitan transportation studies has collected information on travel patterns in the region. Each of these sources has strengths and weaknesses as a base for VMT modeling.

2.2.1 FHWA-Highway Statistics

The FHWA's Highway Statistics reports[1*] are, in effect, the official source of information on several kinds of travel in the United States, including auto travel. They are the most widely cited of VMT data, and they form the empirical basis of every statistical model of VMT developed to date. Unfortunately, the method of compilation of the FHWA data lends itself to estimation errors and bias. These errors are particularly detrimental to modeling VMT because of the forms of estimation procedures used. It is, therefore, worthwhile to examine the procedures in some detail.

The FHWA VMT estimates generally are derived from fuel consumption data for each state, relying on the identity:

$$VMT = FE * C$$

where FE is average fuel economy (miles per gallon), and C is total gasoline consumption (gallons). While gasoline consumption is known with good accuracy from tax data, estimates of average fuel consumption are a major source of error. In principal, the average fuel economy in a state depends on:

1) The mix of cars by age and weight driven in the state;

2) The spatial distribution of travel in the state (most importantly, among urban roads, rural roads, and highways); and

3) Physical features of the state (climate, topography, etc.).

It is theoretically possible, then, to start with independent information on fuel economy by weight class and age of vehicle, and

*Superscripts refer to references listed in Appendix A.

develop a composite weighted average fuel economy that accounts for the type of vehicle driven, driving cycles, and physical factors. Unfortunately, the precise influence of driving cycle, climate, topography, etc., on fuel economy is not known, so that even the most careful estimates would introduce some error.

In practice, the situation is much worse. Each state develops its own VMT estimate using any procedure it wishes. The procedures fall into three broad categories:

a) Simple trend extrapolation based on socioeconomic changes;

b) Extrapolation of traffic counts, based on number of vehicles per mile of roadway; and

c) Some variant of the procedure outlined above, usually based on a very rough estimate of average fuel economy.

In addition, there are motivations for the states to inflate their VMT estimates. Some forms of Federal aid are tied to miles driven in the state, and high travel estimates reduce the traffic accident and fatality rates per mile. As a result, VMT estimates are biased as well as error-prone. Claffey[2] found discrepancies of up to 14 percent between published state estimates and his own careful calculations based on fuel consumption and average fuel economies. On the national level, the Highway Statistics reported VMT for 1970 was some 20 percent higher than that reported in the Nationwide Personal Transportation Survey for the same year.

In spite of their questionable accuracy, the FHWA data have supported all VMT modeling efforts to date. They have been used because no other time series is available, and the national VMT estimates can be readily matched to aggregate socioeconomic data necessary for calibrating statistical models. Given the scarcity of travel data, this might be defended as the most practicable approach if it were not for a statistical problem apparently overlooked in previous modeling efforts. A goal of VMT modeling is to relate the amount of travel occurring to its cost, usually expressed as a function of gasoline price and fuel economy. Unfortunately, the fuel economy data on the right-hand side of the equation are derived from VMT data (FE = VMT/C) or used to derive VMT estimates (VMT = FE * C), depending on state practices. Therefore, any statistical findings of high correlation between fuel economy and VMT are suspect since the relationship is implicit in the data.

To summarize, the FHWA data represent the only time series currently available. Their accuracy is questionable because of non-uniform, overly simplistic estimation procedures. They probably contain upwardly biased errors of up to 20 percent. In addition, the typical estimation procedure prevents the use of FHWA fuel economy data as an explanatory factor in a VMT model. Nevertheless, they have been frequently cited and used in travel studies because they are readily available and easy to match with other aggregate statistics.

2.2.2 Nationwide Personal Transportation Survey

The only other nationwide data set on VMT is the Nationwide Personal Transportation Survey (NPTS)[3] conducted by the Bureau of the Census of the U.S. Department of Commerce for the Federal Highway Administration. This survey provides cross-sectional data on the number of vehicle miles traveled per household and per vehicle based on a multi-state probability sample of 6,000 households taken in 1969-1970. The questionnaire used also recorded information on:

1) general socioeconomic characteristics of the surveyed households such as income, family size and race;

2) number of autos and type of model and year;

3) frequency and distance of trips to work, school and shopping; and

4) availability of public transit.

Survey data were published in aggregate form focusing on six variables in relationship to VMT per household.

a) number of cars per household;

b) age of auto by year;

c) car purchase - new or used;

d) annual income of household;

e) occupation of principal automobile operator; and

f) place of residence classified as incorporated or unincorporated.

While the survey data were published only in national aggregate form, the original household data are available on tape from the FHWA.

As a one-time-only survey, the NPTS could not make direct odometer measurements of VMT. Instead, each household respondent was asked to estimate the total amount of auto travel in the previous year. Such an approach, while far from precise, probably represents an improvement over the Highway Statistics data, and should be unbiased in any event.* Equally important, the survey includes data on form of auto travel -- work, shopping, and leisure, for example -- so that each form can be analyzed individually.

The greatest value of the NPTS data, however, lies in the microscale level of disaggregation. While national aggregate data are much easier to work with, they may obscure behavioral relationships and omit much of the statistical variation found at the microlevel. In addition, a VMT model representing behavior at the level of the individual decisionmaker is theoretically preferable to one based solely on aggregate relationships.

There are three major drawbacks to using the NPTS data as a major source of information for VMT modeling. Most important is the absence of gasoline price data in conjunction with travel estimates. No such complimentary price data were collected during the survey itself, and there appears to be no way of tracing the

---

*FHWA reports that VMT estimates are lower than might be expected, and are in reference to the national Highway Statistics which appear to be biased upward.

observations in the data to specific geographic locales to determine gas prices.* Hence, only rough measures of the cost of travel could be used in a modeling effort.

The second principal drawback is the time at which the survey was made, nearly four years before gas prices increased to their current high levels. If the relationship of VMT to costs of travel is nonlinear, or has changed since 1970, statistical results could be rendered invalid.

A final drawback is the purely cross-sectional character of the data, which raises problems for developing a forecasting model. The data represent, in essence, a snapshot of conditions for 1970. It is unclear whether relationships within the data represent stable long-term conditions, unstable short-term conditions, or even non-characteristic mid-response relationships. This means that elasticities derived from the data could be long-term, short-term, or some average of the two. Forecasts based on such elasticities might show too rapid responses to altered conditions.

In summary, the NPTS household response data represent a nearly complete, internally consistent base of VMT and socioeconomic information. Unfortunately, the major gap in the data -- gasoline price -- is a crucial variable, and cannot be easily added to the data base. The age of the data and absence of time series are further drawbacks. On the other hand, there appears to be a large amount of information on auto travel in the NPTS data -- particularly on travel by trip purpose -- which has not been fully explored.

2.2.3 Metropolitan Travel Studies

Virtually every metropolitan area in the country has, at one time or another, conducted a survey of regional travel patterns for

*Information acquired from George Gray of the Bureau of Census. A substantial fraction of observations may be traced to specific geographic locations, but the resulting subsample will not be fully representative of national behavior.

transportation planning. Such studies, however, are geared toward daily trips rather than total miles driven. While the data collected are similar to the NPTS, the detailed information varies from area to area. In any given metropolitan area, such studies are conducted only at long intervals -- usually a decade or more.

It therefore appears impracticable to compile a nationally representative, internally consistent data base from these studies. They reflect conditions at different points in time, without a coherent time series, and with incomplete consistency from region to region. Their utility is limited to auto travel models for individual metropolitan areas, and comparison of travel behavior from region to region.

2.3 SOURCES OF SUPPLEMENTAL DATA

In addition to VMT data, consistent information on the related social and economic factors is required. A basic listing of such factors includes:

1) income (preferably both permanent and current levels);

2) gasoline price;

3) average fuel economy; and

4) stock of automobiles.

In addition to these items, which should be included on the basis of theoretical grounds, there are a large number of more marginal factors that might be tested in alternative specifications. These include:

a) cost of mass transit;

b) costs of alternative travel modes (e.g., air, bus, rail);

c) unemployment rate;

d) employment category;

e) availability of highways; and

f) index of urbanization.

There is at least some theoretical justification for testing these factors in a VMT model. Both the availability of data and the theoretical suitability of these factors depend on the level of aggregation of the model.

2.3.1 National Data

As might be expected, data on social and economic conditions are most readily available at the national level. This is a principal reason why previous modeling efforts have employed aggregate data. In addition to the FHWA, the Bureau of Labor Statistics (BLS)[4] and the Bureau of Census[5] publish national data annually.

The BLS provides all the necessary fiscal data. Their statistics are in real terms (corrected for inflation) and are available in a continuous time series. In addition to income and gasoline price, they offer information on unemployment rate, employment by category, and prices for other transportation modes.

The Bureau of Census provides annual demographic estimates, including number of households and population. These items are important for normalizing other variables on a per household or per capita basis. The Census Bureau also prepares a long-term projection of demographic conditions that are useful as inputs to VMT forecasts.

Automobile stock and fuel economy data are available from the FHWA. Independent estimates of stock may be obtained from the R. L. Polk Company.* As mentioned earlier, the FHWA fuel economy estimates are not independent of their VMT estimates; independent data on fuel economy by model year are available from EPA **

*R. L. Polk Company, Detroit MI, offers a range of information on automobile registrations, sales, and scrappage, disaggregated to the county level in most instances.

**EPA has not published the complete data on historical fleetwide fuel economy; for a partial compilation see Reference 6.

starting with 1957 models. The FHWA also publishes information on the total miles of highway (by several classifications) in existence in each state.

2.3.2 Microscale Data

The only relevant, nationally representative microscale data are a part of the Nationwide Personal Transportation Survey. The NPTS data set includes information on each household's estimate of its annual gross income. Information on disposable income, permanent income, or changes in income is not available. The data do, however, provide information on the occupational status of the head of the household which may be used as a proxy variable for permanent income.

The significance of the availability of public transit in demand for auto travel has not yet been tested because of the difficulty of measuring transit availability on the national level. One of the advantages of the NPTS data is that they provide information on the distance to public transit on the household level.

If a micromodel is used, the price of gas by geographical area will be needed. Platt's Oil Price Service[7] provides data on the average price of gas since 1963 at major brand stations for 55 cities in 48 states. Since independent dealers often charge lower prices, the cost of travel may be overstated. Fuel economy and cost of travel can be determined from the information on automobile make and year available in the NPTS data.

Degree of urbanization, a measure of physical distances separating people, could be a factor in a microscale VMT model. Indices can be constructed from census data for the regions in which surveyed households reside. This would be possible, of course, only if information on the geographic location of individual respondents can be recovered from either the FHWA or the Bureau of Census.

2.4 SUMMARY

This survey clearly indicates that the scarcity of good data on vehicle miles of travel is a major constraint to future modeling efforts. Currently available VMT estimates are of unknown but questionable accuracy. They can be easily correlated with explanatory socioeconomic and policy-related factors only at the national level of aggregation, where many behavioral relationships may be obscured. At the household level, where travel decisions actually occur, the only representative source of information is somewhat out of date and omits variables that are potentially important in influencing VMT.

In view of this situation, the cost effectiveness of further modeling efforts within the framework of existing data must be questioned. Certainly, the empirical support for such efforts would be suboptimal. At issue, however, is not simply the absolute integrity of the data, but also the extent to which useful information in the data has been extracted. In other words, have the VMT models developed to date made optimal use of the data in their theoretical specifications and tests of hypotheses? This question is addressed in the following section.

An alternative, of course, is to go outside of the context of currently available data and collect the necessary information for a rigorous VMT model. Such an approach, while time-consuming and expensive, could prove the most efficacious long-term strategy, and is discussed in Section 4 of this report.

3. VMT MODELS: THE STATE OF THE ART

3.1 GENERAL

In spite of its primary role in many transportation policy issues, national automobile travel was, until recently, the subject of only cursory analysis. The Federal Highway Administration made some projections of future VMT based on straightforward trend extrapolation,[8] but no causal models, much less robust forecasting techniques, were available.

With the Arab oil embargo and gasoline shortage, however, the critical nature of VMT as an energy policy variable became clear. Several more or less simultaneous efforts were launched to construct VMT models representing economic and social interrelationships, so that the impact of both the energy problem itself and energy-related problems might be anticipated. As a result, the first generation of VMT models evolved. These models represent today's state of the art.

Research on VMT was performed independently by the Transportation Systems Center, the RAND Corporation, and Chase Econometrics Associates, Inc. All three models are based on aggregate VMT data from the FHWA, but there are significant differences in the theoretical underpinnings and specifications.

In the following pages, each model is reviewed and critiqued with the objective of extracting or synthesizing an improved approach. Since all the models achieve high levels of statistical significance, the assessment is in mostly theoretical terms. The implications of the empirical results, along with a comparative discussion, are contained in the concluding portion of this section.

3.2 TRANSPORTATION SYSTEMS CENTER MODEL

The TSC model,[9] developed under the Automotive Energy Efficiency Project (AEEP), estimates vehicle miles of travel per household as a linear function of the preceding year's VMT, the number of drivers, disposable income (all per household), and the cost per mile of auto travel. The actual regression equation is as follows:

$$VMT/H(t) = \underset{(1.06)}{1590} + \underset{(6.21)}{0.6344}\ VMT/H(t-1) + \underset{(2.07)}{2153}\ D/H(t)$$

$$+ \underset{(2.57)}{0.3936}\ R/H(t) - \underset{(5.75)}{140{,}580}\ P/E(t) \qquad (R^2 = 0.995)$$

where the numbers in parentheses are t statistics and:

- VMT/H(t) = vehicle miles traveled per household, year t;
- VMT/H(t-1) = vehicle miles traveled per household, year t-1;
- D/H(t) = number of drivers per household, year t;
- R/H(t) = real disposable income per household, year t;
- P/E(t) = real cost per mile of driving, year t (gasoline price, P, divided by average fuel economy, E).

The regression produced a coefficient of determination (R^2) of 0.995 and one-year elasticities (measured at the mean) of 0.27 for the number of drivers, 0.24 for disposable income, and -0.22 for cost of travel. Because of the lagged VMT term in the quation, it is also possible to calculate long-term (equilibrium) elasticities. They are 0.71, 0.63, and -0.57, respectively.

This equation has several positive aspects. It explicitly includes both income and price variables, a prerequisite for demand models. The specifications of these variables as disposable income and cost per mile are theoretically appropriate. The inclusion of the demographic variable, number of drivers, helps to account for non-economic influences.

The major drawback of the model is the inclusion of lagged VMT per household as an explanatory variable. There is no sound theoretical reason why this term should be a behavioral influence on VMT. It represents, in essence, a trend variable, and its high level of statistical significance results from the stable growth in VMT over the estimation period of 1951-1974. This significance should be taken with a grain of salt; it is far more preferable to explain trends through underlying behavioral causes than to represent them as the inevitable result of the passage of time.

While its presence cannot be justified on theoretical grounds, the lagged VMT variable does permit calculation of long-term elasticities. It would, therefore, be desirable if its presence did not bias the other parameter estimates. Unfortunately, such a bias does occur because of the high correlation between lagged VMT and the other explanatory variables. In a regression analysis without lagged VMT, the following equation resulted:

$$\underset{}{VMT/H} = \underset{(1.4)}{3826} + \underset{(3.1)}{7034\ D/H} + \underset{(4.7)}{1.131\ R/H} - \underset{(2.5)}{174{,}000\ P/E}$$

$$(R^2 = 0.95)$$

Here, both the parameters and elasticities have changed dramatically, as shown in Table 3-1.

This instability in estimated coefficients is a classic result of multicollinearity, caused by the stable growth in most of the variables over the estimation interval. Since the results of both specifications are statistically significant, we have no way of knowing which set of coefficients comes closer to the true population parameters. Because there is no theoretical justification for the lagged VMT term, however, we may assume that it is distorting rather than improving the equation. Finally, the inclusion of the dependent variable in lagged form renders Ordinary Least Squares inappropriate as an estimation technique because of correlation between regressor and error terms.

Another, less crucial problem involves the drivers per household variable. If we view VMT as demand for auto travel, then the number of registered drivers is, in principle, a function of that demand rather than a determinant. That is, the causal link between VMT and number of drivers should be in the opposite direction. The problem is easily solved by substituting a more fundamental demographic determinant of demand for auto travel, such as the population aged 16 years and older. The implied causality would then be:

TABLE 3-1 RESULTS OF ALTERNATIVE VMT SPECIFICATIONS

Variable	With Lagged VMT	Without Lagged VMT	Percent Change
Coefficients			
Constant	1590	-3826	-341
Driver	2153	7034	+227
Income	0.3936	1.131	+187
Cost of Travel	140,580	174,100	+ 24
Short-term Elasticity			
Driver	.27	.88	+226
Income	.24	.69	+187
Cost of Travel	-.22	-.27	- 23

Driving-Age Population	→	VMT (Demand)	→	Number of Registered Drivers

Thus, the demand for driver's licenses would be derived from the demand for auto travel, a conceptual structure which we think more representative of true conditions.

Finally, the TSC model omits potential policy variables relating to mass transit and highway availability. In one sense, this is justified since the equation is already statistically adequate. On the other hand, the omission of such variables, if they are true causal influences, will bias parameter estimates.

3.3 RAND MODEL

The RAND Corporation's model[10] also estimates VMT on a per household basis, but in logarithmic rather than simple linear form. RAND derived two equations:

Estimated Form

$$\text{Log}(VMT/H) = \underset{(752)}{9.19} + \underset{(13.4)}{0.864}\ \text{Log}(A/H) - \underset{(3.52)}{0.369}\ \text{Log}(P) + \underset{(2.316)}{0.025}\ D \qquad (R^2 = 0.99)$$

Derived Form

$$\text{Log}(VMT/H) = 8.0 + 0.864\ \text{Log}\ (A/H) - 0.444\ \text{Log}(P) + 0.444\ \text{Log}\ (E) + 0.035\ D$$

where

A/H = auto stock per household;

D = dummy variable for Federal regulation;

and the other variables are as defined previously. The derived form was calculated with a further equation for fuel economy by assuming that elasticities for gasoline price, P, and fuel economy, E, must be equal and of opposite sign.[11] Since the equation is in logarithmic form, elasticities for auto stock, gas price, and fuel economy are equal to their coefficients.

RAND's VMT model is one of seven equations in a recursive system, the others being:

1) New-car sales;

2) Used-car price;

3) Used-car ownership;

4) Total car ownership;

5) Average fuel economy; and

6) Gasoline consumption.

Within this recursive system, it is assumed that VMT is derived from demand for automobiles; hence, no income variable is included in the VMT equation.

The auto stock variable explicitly represents the assumed influence of auto demand on travel demand. The elasticity of about 0.9 implies that each additional car in the stock brings a smaller percentage increase in VMT, representing the tendency for second and third cars in a household to be driven less than the first car.

The model represents the influences of fuel economy and gas price independently, rather than as a combined cost of travel variable. This is less desirable theoretically since it involves the assumption that the influences of the two variables are independent rather than joint. Further, the coefficient of fuel economy was externally derived rather than directly estimated because of collinearity with the stock variable. Hence, there is no way of knowing how reliable the coefficient is.

While the RAND VMT equation is statistically adequate, it suffers from some serious theoretical shortcomings. The entire concept of a recursive system in which auto demand determines VMT is questionable. It seems much more likely that demand for travel itself is fundamental, while demand for autos is derived. In such a case, the lack of income and travel price variables in the VMT equation are major deficiences. Further, the choice of a recursive rather than simultaneous equation system is theoretically unjustified although perhaps empirically necessary. Within the context of annual measurements, VMT, auto stock, used-car price, and new-car sales must certainly be viewed as simultaneous. The amount of aggregate data on these phenomena, however, does not offer enough degrees of freedom to estimate a simultaneous equation system. RAND's method of approach ignores rather than solves this problem. If the equations should be simultaneous, then to estimate them independently using Ordinary Least Squares will produce biased coefficients and forecasts.

3.4 CHASE ECONOMETRICS MODEL

The Chase Econometrics model[12] is also part of a system of equations. The model estimates VMT as a function of five variables:

$$VMT = \underset{}{3416.1} + \underset{(9.2)}{78.5\,A} - \underset{(2.5)}{3944.5\,P} + \underset{(2.3)}{5181\,\Delta PC1}$$

$$+ \underset{(2.1)}{7841\,V_{-2}} + 8.81\,\Delta YW \qquad (R^2 = 0.99)$$

where

A = automobile stock;

P = relative price of gas and oil;

$\Delta PC1$ = change in consumer price index;

V_{-2} = average price of new cars (2-year lag); and

ΔYW = change in wages and salaries.

Although more complex than the RAND model, Chase's equation is based on similar theoretical assumptions. Demand for auto travel is assumed to derive from demand for autos, so that no explicit income variable is included in the specification. As mentioned earlier, such an assumption is questionable, and the misspecification predisposes the model to biased parameter estimates.

The equation also lacks any reference to fuel economy and its influence on costs of driving. Chase justifies this with a statement that average fuel economy was relatively constant during the estimation period. Its omission from the equation prevents any confident interpretation of the coefficient of gas and oil prices.

The inclusion of variables such as lagged-unit new-car price and change in wages and salaries serves no clear theoretical purpose, and tends to obscure the conceptual structure of the model. They are used as general indicators in the context of Chase's overall macroeconomic model. They have little behavioral justification, however, and their statistical significance may well reflect aggregate trends in economic conditions without verifying underlying causal relationships.

3.5 SUMMARY AND CONCLUSIONS

Based on theoretical grounds, all of the models reviewed here suffer from a common set of drawbacks. From them, the following points about the current state of the art of VMT modeling might be made:

1) interactions between auto stock and VMT have not been adequately explored;

2) current models are likely to have biased parameter estimates because of omitted simultaneous relationships; and

3) cross elasticities of VMT with respect to the price and availability of competing modes of travel remain unknown.

In our opinion, these weaknesses result from a failure to evaluate thoroughly and address the complete system of interrelationships to which VMT belongs. The decisions leading to auto travel and ownership, new- and used-car purchases, scrappage, and the associated price structures and market clearing mechanisms cannot be considered independently.

The TSC equation seems theoretically superior to the other two models, primarily because it includes the disposable income and cost of travel variables omitted elsewhere. This judgment might seem unfounded in view of the general statistical adequacy of all three equations. Unfortunately, the nature of the data employed -- a stable time series with serially correlated, multi-collinear variables -- renders statistical results virtually useless for normative analysis. This is readily apparent in the *t* statistics and coefficients of determination for the three equations. Despite major changes in choices and definitions of variables, all parameters are highly significant and each R^2 is above 0.99. In such a situation, judgments must be made on other than empirical grounds.

The same difficulty applies in evaluating forecasting adequacy. It is all too frequently argued that statistical significance "proves" or "validates" a model's forecasting accuracy. If this were true, each of the models reviewed would suffice. In the present case, however, the statistical significance appears to arise from stable, joint trends in variables rather than direct causal relationships. Forecasting, therefore, amounts to little more than trend extrapolation; so long as the same trends occur, forecasts should be accurate. But high gas prices, the current economic situation, and continuing energy problems have already disturbed the basic pattern of growth on which trend extrapolation depends. Hence, the statistical significance of previous modeling efforts not only fails to establish forecasting adequacy, but actually obscures underlying behavioral relationships on which sound forecasts might be based.

Until alternative data reflecting more of the true variability in VMT are used for analysis, the appropriate theoretical questions and alternative specifications cannot be adequately tested. It might be argued with some justification that problems in the current state of the art result from the lack of adequate data for more thorough analysis, which, to a certain extent, is true. However, this report has determined that better use can be made of what information does exist, and that the collection of an adequate set of data is a viable long-term option. The following section explores these possibilities in detail.

4. RECOMMENDATIONS AND FURTHER RESEARCH

4.1 GENERAL

The preceding sections treated the nature and quality of currently available data and the state of the art of VMT forecasting. The resulting evaluation indicates that the quality of existing models is not commensurate with the significance of VMT as a policy variable. No well structured, policy-sensitive VMT model is now available.

The following section outlines strategies for improving the state of the art. The strategies involve both theoretical and empirical components. While many problems in VMT modeling can be traced to scarce or inaccurate data, the available information can be used more judiciously than it has been in previous efforts. At the same time, more complex and realistic hypotheses should be tested before accepting highly simplified models. It is here suggested that the available data can support some such tests and, further, that a Federally sponsored data-collection program could allow even greater advances in VMT forecasting.

These improvements will take substantial effort, however, while TSC has current and ongoing needs for VMT forecasts. For this purpose, separate strategies have been developed for short-term, intermediate-term, and long-range research programs.

Before presenting these recommendations in detail, however, it will be useful to review the principal issues raised thus far concerning VMT models. A description of a desirable general structure toward which future efforts might work can then be made.

4.2 ISSUES AND GOALS

Three major theoretical issues concerning VMT remain untested and unresolved. Each references a current weakness in the state of the art of VMT modeling, and offers an objective for further research.

Does VMT reflect a basic or derived demand? Two of the three models reviewed assume VMT to be subsidiary to automobile demand. It seems more likely, however, that demand for auto travel, based on day-to-day requirements for commuting, shopping, etc., dominates auto ownership decisions. Choice of specification for both auto ownership and VMT equations rests on the answer to this question.

Is VMT an independent demand or part of a simultaneous system of relationships? All three models assume that VMT is determined independently; their coefficients were all estimated using Ordinary Least Squares. Yet the concept of simultaneous, interdependent relationships applies very well to auto-ownership, travel-demand interactions. While it has been suggested that auto travel is the original quantity determined, it seems probable that one of the influencing factors may be current auto ownership. Households with more than one car, in this case, would desire to travel more by auto, simply because they have broader horizons. Concurrently, the desire to drive would more clearly affect ownership decisions. If these relationships are, in fact, simultaneous, then estimation of independent equations for VMT and auto stock or ownership will lead to biased, inconsistent parameters.

Is demand for auto travel affected by competing modes? Strict economic reasoning requires that demand for auto travel be evaluated in light of the availability and price of competing modes. They include not only mass transit, but also intercity transportation by bus, rail, and air. Unfortunately, economic analysis may prove inadequate since competing modes may tend to have irregular or understated prices. Nevertheless, some form of explicit recognition of cross elasticities is important, no matter how small they are, since modal shifts offer some of the best possibilities for energy saving.

Beyond these three basic questions, inquiry must be made as to the most appropriate theoretical specification for an improved VMT model. This step has been largely omitted from previous efforts, yet it should play a critical role in econometric model-building. Without careful specification of hypothetical relationships for testing, it is only possible to engage in invalid *a posteriori* rationalization of results. This procedure encourages statistical fishing to find significant parameters and an "acceptable" coefficient of determination even though such practices violate many basic rules of statistical analysis.

In principle, VMT can be represented as demand for auto travel in a specification equivalent to demand for other economic goods. The model should include at least three classical economic phenomena: income elasticity, price elasticity, and cross elasticities for competing goods. The precise fashion in which these phenomena are represented, however, depends on the overall modeling approach taken.

Variables used to reflect income effects on VMT should be different for microlevel and macrolevel models. Some forms of auto travel (e.g., vacations) are clearly dependent on current levels of disposable income. Other forms (e.g., commuting) should respond to changes in permanent (long-term) income levels. On a microlevel, some form of both variables might be tested, while on a macrolevel only aggregate or average national data may be used, and the distinction between permanent and current income is obscured. Further, the use of aggregate data dismisses income distribution effects and implies that an additional variable such as the unemployment rate should be included for complete specification.

The relevant price of auto travel is the variable cost per mile. In the short run, the only variable cost is that of fuel, which depends on gas price and fuel economy. Over a longer time period, vehicle maintenance would also be a variable cost, and in the long run, the annual cost of owning and operating a car would have to be included. Under any time horizon, the use of gas price alone, without adjustment for fuel economy, leads to bias in the estimation of the price coefficient and the resulting elasticity measure.

Viewing public transit as a substitute good, the choice of a transit variable should be its price. Data on price per mile for transit are seldom available, so that simple proxy measures for the availability of service often must suffice. The difficulty in obtaining data on transit service increases when a macroview is taken, for there are few meaningful aggregate measures of transit availability.

Other potential substitutes for intercity auto travel include air, rail, and bus transportation services. The average cost per mile for each of these services might also be included in the specification.

In addition to the above economic factors, several physical and social indices suggest themselves. A density or urbanization variable, for example, would reflect spatial distances over which travel must occur; in a cross-sectional or microscale approach such a variable would be appropriate, but definitional and statistical problems arise at the national level. The availability of good roads on which to drive should influence demand for auto travel, and represents a potential policy variable as well. On the microlevel, form of employment and household size should affect demand; on aggregate, these influences are largely averaged out.

Many of the above-mentioned factors have been omitted in previous modeling efforts. Such omissions, called "under-specification," cause bias in the coefficients of variables included in the model unless all explanatory factors are perfectly independent. This condition never occurs in practice, so that it may be assumed that existing models contain biased parameters.

Because of insufficient data, it is not possible to test a truly complete specification even if one were here designed. Some of the causal factors mentioned here cannot be measured with sufficient accuracy to allow tests of their influence on VMT. Others may prove to have insignificant effects when empirically tested. Nevertheless, parameters ultimately included in the model can be evaluated only after the most thorough possible specifications have been tested. Future modeling efforts should, therefore, begin with complex specifications to ensure that simplification which occurs in the normal course of hypothesis testing does not bias model coefficients.

4.3 SHORT-TERM STRATEGIES

By short-term is meant actions capable of completion in a period of one to three months; of necessity, this precludes any extensive new data collection and analysis. The most appropriate approach, therefore, is to modify and refine an existing model.

The TSC model is suggested for modification since it already includes the appropriate income and cost of travel variables. Recommended changes are: (1) deletion of the lagged VMT term; (2) inclusion of price or availability variables for competing modes; (3) inclusion of a highway availability variable; (4) replacement of the number of drivers with driving-age population; and (5) addition of unemployment rate.

1) Drawbacks associated with the lagged VMT variable have already been discussed. Its deletion, while preventing calculation of long-term equilibrium elasticities, will enhance the value of the remaining parameter estimates by removing a major source of multicollinearity and serial correlation. The problem of the applicability of Ordinary Least Squares to a specification containing a lagged dependent variable will also be avoided.

2) Variables representing availability and price of competing travel modes should be tested in new regression analyses. In theory, the influence of intra-urban mass transit and intercity air, rail, and bus travel should be distinguished. Since intercity auto travel is a relatively small fraction of total VMT,[13] however, one may expect the effects of competing intercity modes to be small. Mass transit is the major competing mode and the most important potential policy variable for transportation-related problems.

There are substantial difficulties in defining and measuring the true price of alternative travel modes at the national level. Data on total expenditures for fares and on total passenger miles might be used to derive a composite unit price, but the distinction between intra- and inter-urban bus and rail service would be obscured. This would prevent isolation of the effects of urban mass transit, a minor problem since all forms of auto travel are included in a single variable as well. More serious is the fact that expenditure and mileage data can provide only a crude measure of actual asking prices for each mode, the theoretically desirable measures of cross elasticities. Further inspection may indicate that proxy variables for availability or price would serve better.

The purpose of including variables for competing modes in spite of obvious measurement problems is to test both their statistical significance and their influence on the other model parameters. The recommended tests will indicate whether present coefficients are biased by under-specification. If the new variables prove to be insignificant and have little influence on presently included parameters, then they may be omitted with confidence.

3) Inclusion of a highway (or major road) availability variable will represent demand-inducing effects of an improved roadway system. Total lane-miles of interstate highways would be an appropriate measure, and this item is available from the FHWA.

4) The theoretical problem with using number of drivers as an explanatory variable was discussed in Section 3. By using the driving-age population instead, the appropriate causal sequence (driving-age population → demand for travel → demand for driver's licenses) will be represented.

5) As noted earlier, national data obscure the distinction between current and permanent income. The unemployment rate serves as a proxy by reflecting fluctuations in general economic conditions.

The basic specification for the new model is:

$$VMT/H = b_0 + b_1 (Y_D/H) + b_2(P_D/H) = b_3(C_A) + b_4(C_M) + b_5(C_B) + b_6(C_R) + b_7(C_P) + b_8(Hw) + b_9 (U_R)$$

where:

VMT/H = vehicle miles of travel per household;

Y_D/H = real disposable income per household;

P_D/H = driving aged members per household;

C_A = real cost of auto travel;

C_M = real cost of mass transit;

C_B = real cost of intercity bus travel;

C_R = real cost of intercity rail travel;

C_P = real cost of air travel;

Hw = highway availability; and

U_R = unemployment rate.

Using the FHWA national VMT data, such a regression equation would leave approximately 15 degrees of freedom for statistical analysis (assuming that post-World War II years are excluded). Hence, the expanded specification will not seriously detract from tests of significance or confidence intervals.

In spite of the low level of effort required, this short-term approach can accomplish several basic objectives. The theoretically preferred representation of VMT as a basic rather than derived demand will be explicitly incorporated. Cross elasticities for competing modes will be empirically tested rather than excluded a priori. Potential bias due to under-specification will also be empirically evaluated.

The resulting model, however, will continue to contain major weaknesses. It will be based on national data with the associated multicollinearity and serial-correlation problems. It will contain only crude mass transit measures. Most important, the underlying statistical analyses will not address the issue of interactions between travel demand and auto ownership in a simultaneous equation system. These problems can be addressed only in a more substantial research effort. The short-term model should be viewed as an interim forecasting technique.

4.4 INTERMEDIATE-TERM STRATEGIES

An intermediate-term research effort (approximately one year elapsed time) would permit substantially improved strategies for developing a VMT model. In particular, it would allow both microscale modeling with the Nationwide Personal Transportation Survey data and new analysis of the FHWA data on a cross-sectional as well as a time-series basis. It is probable that these approaches can be integrated for optimum model estimation.

The advantages of microscale analysis are several. Most important, the work focuses on the behavior of individual decision-making units as they respond to economic, social, and demographic influences. This is the meaning of the term "behavioral model." Statistically, this means that all of the variability in decision-making is reflected in the data, rather than averaged out through aggregation. In terms of specific modeling goals, a microscale effort allows:

1) A more precise measure of income elasticity, since income varies more among households than average household income varies over time.

2) Clarification of the role of income in household travel decisions, by eliminating the simultaneous equation bias inherent in an aggregate demand model. On the microlevel, it is reasonable to assume a one-way dependence of demand on income. On the macro-level, there is likely to be interdependence between aggregate travel demand and income.

3) Determination of the effect of the availability and/or the cost of public transit on VMT. These policy variables are difficult to define and measure at the national level.

4) An increase in the number of observations, which should improve the precision of parameter estimates as well as permitting evaluation of a simultaneous equation model of travel demand and auto ownership.

5) A reduction in multicollinearity and serial correlation relative to an aggregate time-series model, where income, price,

and population trends tend to move together.

While the NPTS data set offers these advantages, it also has deficiencies which make complete reliance on its information inadvisable. They include:

a) The age of the data. NPTS information is now five years old, and may not fully reflect current relationships. This is particularly true for gas prices, which rose dramatically three years after the survey.

b) Lack of time series. The NPTS data are for one point in time, so that a model based solely on this information would obscure dynamic properties for forecasting.

c) Missing information. The NPTS does not contain data on gasoline prices and the individual household observations cannot all be tied to specific geographic locations. Therefore, the original data will have to be supplemented and some observations may be lost, reducing the representative character of the data.

On balance, the NPTS appears to offer a large quantity of information on household VMT and the factors that influence travel decisions. This information has never been fully explored and deserves thorough scrutiny. Because of the drawbacks mentioned above, however, the NPTS cannot by itself support a national forecasting model. We, therefore, suggest that the survey data be analyzed in conjunction with additional information.

The only available source of supplemental data is the FHWA. Some of the incumbent problems in this data set can be avoided, however, by disaggregating at the state level to provide cross-sectional time-series information. Ordinarily, such disaggregation tends to enlarge measurement errors since random inaccuracies partially cancel out with aggregation. In the case of VMT, on the other hand, the data appear to be biased at the state level, so that no cancelling of errors would occur in aggregating to the national level. Hence, the state data should be no less accurate than the national data, while offering several clear advantages:

1) A cross-sectional time series would greatly increase the

number of the degrees of freedom available for parameter estimation. The small number of observations (roughly, 25) in the national time series has been a major deterrent to development of a simultaneous equation system. The suggested approach should permit such development.

2) The expanded data base should increase the range of variation in VMT measurements and independent variables, while at the same time reducing the degree of multicollinearity found in the data. This will provide a more severe and realistic test of hypothesized relationships, while minimizing the inefficiency in parameter estimates from multicollinearity.

3) State-level data permit inclusion of better mass transit measures since geographic as well as temporal differences may be represented. In addition, the appropriate explanatory variables at the state level are more consistent with microscale phenomena than are national variables. For example, degree of urbanization is a reasonable explanatory variable at the microlevel and state level, less appropriate in a national model.

Both the NPTS and the cross-sectional time series offer data superior to the FHWA aggregate information. Together they represent the best empirical support currently available for a VMT model. The microlevel but static information in the NPTS can be supplemented with dynamic state level data to help verify elasticity and parameter estimates.

For a theoretical approach, it is recommended that a simple two-equation system be explored. Possible specifications include fully interactive and simultaneous VMT and auto-ownership equations as well as a system in which auto demand is derived from demand for auto travel. It is not believed that the data will support more complex models making up the new- and used-car markets, price structures, and scrappage.

Generally, the VMT equation need not be dramatically changed from the specification listed in the preceding section. However, a more detailed level of disaggregation will require the inclusion of additional factors to account for inter-household or inter-state differences. Among them are:

a) degree of urbanization (state level);

b) occupation (household head);

c) racial or ethnic status (household);

d) size (household); and

e) age (household head).

Most of these influences on the household level can be represented through dummy variables, and since the analysis would no longer be constrained by degrees of freedom, their inclusion should present no statistical problem.

For an auto-ownership equation, the critical explanatory variable should be VMT. Other candidates are uncertain on theoretical grounds and should be empirically tested. Unit auto prices (or annual average costs of ownership), for example, may directly influence auto-ownership decisions; alternatively, they may exert their influence not on ownership but on the kind of car chosen after the ownership decision has been made. Income (or income class) may affect ownership only through its influence on VMT, or could have an independent influence. The same is true for the number of driving-age members of the household.

The preliminary suggestion for specifications, then, is:

<u>Interactive Version</u>

$$\begin{aligned} VMT = {} & b_0 + b_1(Y_D) + b_2(P_D) + b_3(C_A) + b_4(C_T) + b_5(I_U) \\ & + b_6(D_E) + b_7(D_R) + b_8(D_S) + b_9(D_A) + b_{10}(O_A) \\ & (+b_{11}(U_R)); \end{aligned}$$

$$O_A = c_0 + c_1(VMT) + c_2(I_C) + c_3(P_U) + c_4(P_D) \ (+ c_5(U_R));$$

<u>Derived Ownership Version</u>

$$\begin{aligned} VMT = {} & b_0 + b_1(Y_D) + b_2(P_D) + b_3(C_A) + b_4(C_T) + b_5(I_U) \\ & + b_6(D_E) + b_7(D_R) + b_8(D_S) + b_9(D_A) \ (+ b_{10}(U_R)); \end{aligned}$$

$$O_A = c_0 + c_1 (VMT) + c_2(I_C) + c_3(P_U) + c_4(P_D) \ (+ c_5(U_R));$$

where:

VMT = vehicle miles traveled per household;

Y_D = disposable income (real) per household;

P_D = people of driving age per household;

C_A = cost of auto travel;

C_T = cost (or availability) of mass transit;

I_U = index of urbanization (for example, population per square mile, state average);

D_E = dummy variable for employment category;

D_R = dummy variable for race;

D_S = dummy variable for household size;

D_A = dummy variable for age of household head;

O_A = number of autos owned per household;

I_C = income class of household;

P_U = unit price of car; and

U_R = unemployment rate (included in state-level data).

The first specification involves simultaneous equations that must be estimated with an advanced technique (e.g., two-state or three-state least squares) to avoid inconsistent and biased parameter values. The second equations are recursive rather than simultaneous and may be estimated with Ordinary Least Squares.

The only difference between the specifications is the inclusion of auto ownership (O_A) in the VMT equation for the interactive version. This implies two-way interaction between VMT and auto ownership, whereas in the second system the causal influence is one-way from VMT to ownership.

It should be noted that not all the variables specified above would necessarily be tested. The occurrence of intercorrelation between independent variables might make it more prudent to omit some terms. It is to be expected, also, that the models will be refined and variables deleted or added as the statistical analyses proceed. The equations are suggestive only.

The specifications should be tested using both the NPTS and cross-sectional time-series data The microlevel NPTS data should, in theory, provide better parameter estimates because of their greater detail. Tests using the cross-sectional time series, on the other hand, will reflect dynamic behavior obscured in the static NPTS data. The two sets of regression results should be compared to determine the relative size, stability, and variance of parameter estimates. It will then be possible to select coefficients judged more reliable from one data set and constrain their values in a regression with the other data set. This procedure is particularly valuable when the data, like the FHWA time series, contain multicollinear, serially correlated, or relatively unchanging variables that produce biased or inefficient parameter estimates. By constraining some coefficients to previously determined values, the influence of the remaining variables is more accurately estimated.

The process of integrating findings from both sets of data is a well established econometric technique and allows the maximum amount of information to be obtained from multiple sources of information that are individually incomplete. The resulting VMT model should represent the most accurate possible within the limitations of currently available data.

4.5 LONG-RANGE STRATEGIES

Both previous recommendations rely on published historical information. As a result, they are data-limited in two senses. Appropriate theoretical specifications must be constrained somewhat to match the available empirical support, and statistical results cannot be accepted with complete confidence because of probable inaccuracies in the data. In the long-term, these problems may be avoided only by collecting better information. There are several possible approaches, but the most cost-effective strategy is probably a repetitive survey based on multi-stage probability sampling.

Such a survey would involve in-the-field interviews or mailed questionnaires to acquire information on a nationally representative sample of households. It would be similar to the National Personal Transportation Survey, but with important exceptions. Auto travel should be measured directly through odometer readings. This implies at least one repetition to measure elapsed VMT; additional repetitions would provide a time series and increase the value of the data. To reduce costs, the number of households surveyed should be the minimum allowing for national representation.

Data to be collected for each household should include information on auto travel and ownership, demographic and economic characteristics, and geographic location of residence. While the ultimate content of the survey requires expert planning and design, a preliminary list of informational items is presented in Table 4-1.

TABLE 4-1 PRELIMINARY INFORMATIONAL ITEMS

<u>Travel Data</u>

VMT
auto travel to work
auto travel for shopping
auto travel for cultural and religious activities
auto travel for vacation and leisure
transit travel

<u>Demographic Data</u>

ages and sexes of household members
race and ethnic status

<u>Economic Data</u>

income (disposable and permanent)
employment status of household members
yearly auto-related expenditures
prices paid for autos
prices paid for gasoline
price (fare) for mass transit

<u>Locational Data</u>

type of location (urban, suburban, small town, rural)
distance to place of work

The expense of this undertaking is considerable (probably on the order of $100,000 for each repetition of the survey). The usefulness of the resulting information, however, should more than justify the expense. Such a data base would represent:

1) the first accurate measurement of personal auto travel;

2) the best means of estimating national aggregate VMT; and

3) a complete source of information for analyzing in detail the interactions of auto travel and ownership, new- and used-car sales, and price structures and elasticities.

The data would clearly support research not only on VMT, but on the entire system of interrelationships addressed by the AEEP in its Integrated Model.

If TSC is at a stage of planning for long-range fundamental research, this data base should receive a high-priority position in the program. Potential applications of the information should provide input in designing the survey, as has been here attempted, but detailed statistical analyses and modeling efforts can and should be planned in a more thorough fashion during the interval while the data are collected.

Clearly, support for such a program depends on Federal research priorities. Energy conservation and air quality maintenance, however, appear to be long-range national goals whose achievement will heavily rely on transportation policy. Collection of adequate data is a critical step in advancing the state of the art of policy analysis.

5. LABOR AND COST ESTIMATES

5.1 GENERAL

The research strategies presented in Section 4 were designed to provide both short-term forecasting capability for VMT and refinement of that capability over time. The combination of all three approaches should provide a balanced research program geared toward substantial improvements in the state of the art in the long term. Planning for such a program, however, requires more detailed information on the timing and level of funding needed for each task.

This section provides estimates of the level of effort labor costs, computer costs, and additional direct costs necessary to perform the recommended research. Where useful, the work is broken down into tasks for which separate estimates are provided. Also shown are schedules and elapsed time for each strategy.

Labor and labor-cost estimates implicitly reflect judgments concerning the experience or special expertise required for each task. Rather than imply a precision missing from the estimates, however, we have aggregated labor and costs to single values for each task. Computer costs are generally subdivided into central processing, input-output, and disk storage. Estimates reflect current typical commercial time-sharing rates.

It should be noted that the estimated expenses include only direct costs. No overhead or fees are included since they vary from establishment to establishment.

5.2 SHORT-TERM STRATEGY

The recommended short-term strategy calls for re-estimation of the TSC model with modifications and additions to its current structure. This requires a minimum of new data collection and statistical as well as associated costs. Estimates of labor and cost follow in Table 5-1.

TABLE 5-1 SHORT-TERM LABOR AND COST ESTIMATES

Task	Effort (work-days)	Estimated Cost ($)
Data Collection and Preparation (costs of competing modes, mass transit and highway availability)	15	$ 900
Statistical Analysis and Evaluation	10	800
Total Direct Labor	25	1,700
Computer Costs* (400 CPU seconds plus input-output)		250
Total Direct Costs		$1,950

Elapsed Time: 1 Month

*Based on typical commercial time sharing costs.

5.3 INTERMEDIATE-TERM STRATEGY

The intermediate-term strategy calls for a substantial new data collection-and-preparation effort as well as more extended and complex statistical analyses. Tasks may be broken down into NPTS data preparation and supplementation, FHWA data collection and preparation, and statistical analyses. Cost and labor estimates are given in Table 5-2.

A representative schedule for the intermediate strategy would require approximately 10 work-months of elapsed time, as indicated in Table 5-3.

TABLE 5-2 INTERMEDIATE-TERM LABOR AND COST ESTIMATES

Task	Effort (work-months)	Estimated Cost ($)
A. NPTS data preparation. TSC currently owns a copy of the data tape, but appropriate variables must be prepared for analysis.	1	$ 1,500
B. Supplemental NPTS data collection. Observations must be matched to geographic areas and data on gasoline price, urbanization, etc., acquired.	2	3,000
C. FHWA data collection and preparation. The state VMT and highway data must be prepared. Concurrently, the requisite socioeconomic data must be acquired in time series at the state level.	4	6,000
D. Statistical analyses. Tests on the NPTS and cross-sectional time-series data may be carried out in parallel until consistent results are obtained. Then, the results must be integrated in constrained regressions. This task also includes evaluation and interpretation of statistical results.	6	12,000
E. Documentation. The complete set of statistical results as well as the final model should be described in detail.	2	4,000
Total Direct Labor	15	26,500
Computer: Central Processing	5 hours	4,000*
Input-Output		3,000*
Storage		1,000
Total Computer		8,000
Total Direct Expenses		34,500

*Typical commercial time-sharing rate.

TABLE 5-3 REPRESENTATIVE SCHEDULE FOR INTERMEDIATE STRATEGY

	Elapsed Time (Months)												
Task*	0	1	2	3	4	5	6	7	8	9	10	11	12
A	(1 work-month)												
B		(2 work-months)											
C	(4 work-months)												
D						(6 work-months)							
E									(2 work-months)				

*As defined in Table 5-2.

5.4 LONG-TERM STRATEGY

The recommended long-term strategy consists of a data collection effort in support of more thorough modeling activities. In this study, an attempt has not been made to define detailed modeling tasks since they should be designed in conjunction with the final data collection effort.

There are two potential approaches to the suggested effort. The first includes sample design identification of respondees in addition to the usual tasks of questionnaire preparation, conduction of the survey, and data processing. However, several survey firms have established so-called "household panels" that consist of a nationally representative group of households selected at the time of the last census of population who are reimbursed for responding to questionnaires. By taking advantage of the existing household panels, the expenses of sample design and identification of respondees are minimized. The arrangement is also helpful for repeat surveys in which the same households are questioned at various times. The use of household panels is, therefore, tentatively recommended with the caveat that the representativeness of the sample be evaluated prior to a final decision.

If this approach is taken, the major tasks would be questionnaire design, the actual carrying out of the survey, done by an experienced survey group, and data processing. It should be noted that questionnaire design is perhaps the most crucial step since all the information for intended modeling efforts must be obtained in the appropriate form. In cases such as mass transit variables, this will take careful and extended evaluation. Tables 5-4 and 5-5 give rough estimates for the major tasks.

TABLE 5-4 LONG-TERM ROUGH LABOR AND COST ESTIMATES

Task	Effort (work-months)	Estimated Cost ($)
A. Questionnaire design. Should be performed in conjunction with detailed modeling plans.	3	$ 6,000
B. Questionnaire preparation and carrying out survey. Mailed questionnaires.	(5,000 households)	17,000*
C. Data Processing. Management, keypunching, computer programming.	12	15,000

*Cost estimate based on 4-card questionnaire, 5,000-household sample; and typical charges of major survey firms.

TABLE 5-5 ESTIMATED COMPUTER COSTS

	Estimated Cost
Computer Expenses:	
Input-Output (1 million units)	1,000
Central Processing	5,000
Storage	1,000
Total Computer	7,000
Total Estimated Costs	$45,000

It is difficult to identify accurately a firm schedule for this work since survey companies make mailings only at selected times during the year. A minimum schedule might be the following in Table 5-6.

TABLE 5-6 MINIMUM SCHEDULE FOR LONG-TERM STRATEGY

Task	0	1	2	3	4	5	6	7
	Elapsed Time (months)							
Design Questionnaire	(3 work-months)							
Survey Preparation			*					
Survey					*			
Data Processing						(12 work-months)		

*Work by survey firm; no level of effort estimates.

As seen in Table 5-6, total elapsed time is 7 months, with an effort (not including the questionnaire printing, mailing and sorting) of 15 work-months. Estimated costs do not include statistical analysis of the data, which would depend on the extent of the modeling activities undertaken.

APPENDIX A - REFERENCES

1. Federal Highway Administration, Highway Statistics, published annually by the U. S. Government Printing Office, Washington DC.

2. P. J. Claffey, Travel Estimates from Fuel Consumption Information, prepared for the Federal Highway Administration, Report No. DOT-FH-11-7833, 1972.

3. Federal Highway Administration, Nationwide Personal Transportation Survey, Washington DC; national data published in 11 volumes, original household responses available on tape from the FHWA.

4. Bureau of Labor Statistics, Consumer Price Index; U.S. Department of Commerce, Survey of Current Business, U.S. Government Printing Office, Washington DC, various issues.

5. Bureau of Census Population of the United States, Trends and Prospects: 1950-1990, U.S. Government Printing Office, Washington DC, May 1974.

6. Austin and Hellman, "Fuel Economy of the 1975 Models," (SAE Paper No. 740970), October 1974.

7. Platt, Platt's Oil Price Handbook, McGraw Hill, New York NY, various issues.

8. Federal Highway Administration, Highway Travel Forecasts, U.S. Government Printing Office, Washington DC, November 1974.

9. D.E. Ward and L. Horan, "A Method for Projecting Aggregate Auto Miles Traveled," Working Paper on file at the Transportation Systems Center, U.S. Department of Transportation, Cambridge MA, May 1975.

10. S. Wildhorn, et al., How to Save Gasoline: Public Policy Alternatives for the Automobile, prepared for the National Science Foundation, RAND Corporation, Santa Monica CA, October 1974.

REFERENCES (CONTINUED)

11. Wildhorn, ibid., p. 151.

12. Chase Econometrics Associates, Inc., The Effect of Tax and Regulatory Alternatives on Car Sales and Gasoline Consumption, prepared for the Council on Environmental Quality, New York NY, May 1974.

13. Federal Highway Administration, Nationwide Personal Transportation Survey, "Purposes of Automobile Trips and Travel," Report No. 10, U.S. Government Printing Office, Washington DC, May 1974.

APPENDIX B - REPORT OF INVENTIONS

This report provides an inventory and evaluation of existing models for estimating vehicle miles of travel, and recommends areas for future research. After diligent investigations, no other innovations, discoveries, improvements, or inventions were revealed.

130 copies