1. Report No. FHWA/TX-95/1235-6	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle TRIPCAL5 DOCUMENTATION MANUAL, REVISED EDITION		5. Report Date February 1992 Revised: January 1995
		6. Performing Organization Code
7. Author(s) David F. Pearson, Charles E. Bell, and George B. Dresser		8. Performing Organization Report No. Research Report 1235-6 (Revised)
9. Performing Organization Name and Address Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. Study No. 0-1235
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Transfer Office P. O. Box 5080 Austin, Texas 78763-5080		13. Type of Report and Period Covered Interim: September 1989 - January 1995
		14. Sponsoring Agency Code

15. Supplementary Notes
Research performed in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration
Research Study Title: Improving Transportation Planning Techniques

16. Abstract

This manual is designed to provide technical documentation for the trip generation program TRIPCAL5. It was originally published in February 1992 and is published here in revised form to include documentation on the default models; this documentation is presented as a technical appendix.

In an effort to update the Texas Department of Transportation's transportation planning process, the trip generation program, TRIPCAL5, was developed in 1990. TRIPCAL5 is a multi-functional, flexible program for estimating trip productions and attractions for multiple trip purposes via user-specified models. Trip productions and attractions may be estimated for up to 10 trip purposes and 9,999 zones. The program includes such features as user-specified trip production and attraction models, input of user-developed disaggregate data at the zone level, and/or the disaggregation of the zonal data using default models within the program. The program's flexibility allows the trip generation process to be designed to maximize the use of local data and provides a quantum improvement in the trip generation process.

Included in this report are program options; a brief discussion of the function and purpose of each subroutine; cross-reference of the subroutines and functions; description of each of the variables by labeled common statements; description of the sorts and sort keys; data set formats; how the data flow through the program; discussion of the results of the program tests which were done; and a summary.

The appendix provides technical documentation on the default models which provide flexibility to the user in the application of TRIPCAL5. The use of a disaggregate cross-classification trip generation model is assisted by the provision of these default models in TRIPCAL5. These models will disaggregate households at the zone level and produce reasonable estimates of the percentages of households by household size, household income, and vehicle availability. A default model is also included which estimates the total truck and taxi trip productions within the urban area.

17. Key Words Household Size, Household Income, Auto Ownership, Vehicle Availability, Distributions, Trip Generation, Disaggregation, Cross- Classification, Trip Productions, Transportation Planning, Regression Model	18. Distribution Statement No Restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif.(of this report) Unclassified	20. Security Classif.(of this page) Unclassified	21. No. of Pages 152	22. Price

TRIPCAL5 PROGRAM DOCUMENTATION MANUAL
REVISED EDITION

by

David F. Pearson
Associate Research Engineer
Texas Transportation Institute

Charles E. Bell
Systems Analyst
Texas Transportation Institute

and

George B. Dresser
Research Scientist
Texas Transportation Institute

Research Report 1235-6
Research Study Number 0-1235
Research Study Title: Improving Transportation Planning Techniques

Sponsored by the
Texas Department of Transportation
In Cooperation with
U.S. Department of Transportation
Federal Highway Administration

February 1992
Revised: January 1995

TEXAS TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas 77843-3135

IMPLEMENTATION STATEMENT

This report provides technical documentation of the mathematical models provided in the trip generation program, TRIPCAL5. These models provide the program with the flexibility to disaggregate households at the zone level for input to a disaggregate cross-classification trip generation model. These models may be implemented by TRIPCAL5 users to develop estimates of trip productions and attractions for any urban area in Texas.

DISCLAIMER

The contents of this report reflect the views of the authors who are responsible for the opinions, findings, and conclusions presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration or the Texas Department of Transportation. This does not constitute a standard, specification, or regulation. Additionally, this report is not intended for construction, bidding, or permit purposes. George B. Dresser, Ph.D., was the Principal Investigator for this project.

TABLE OF CONTENTS

LIST OF TABLES

SUMMARY

This manual is designed to provide technical documentation for the trip generation program TRIPCAL5. It was originally published in February 1992 and is published here in revised form to include documentation on the default models; this documentation is presented as a technical appendix.

Transportation planning typically involves a four-step process consisting of trip generation, trip distribution, mode split, and traffic assignment. These steps systematically produce estimates of travel demand which are used for short-term and long-term transportation planning. The first step in this process is trip generation in which estimates of the number of trips being produced and attracted are developed. These estimates are generated for different trip purposes for sub-areas called zones. The basis for the estimates are the socioeconomic characteristics of the households or activities within the zones.

In an effort to update the transportation planning process employed by the Texas Department of Transportation (TxDOT), a new trip generation program, TRIPCAL5, was developed in 1990 (1). TRIPCAL5 is a multi-functional, flexible program for estimating trip productions and attractions for multiple trip purposes via user-specified models. Trip productions and attractions may be estimated for up to 10 trip purposes and 9,999 zones. The program includes such features as user-specified trip production and attraction models, input of user-developed disaggregate data at the zone level, and/or the disaggregation of the zonal data using default models within the program. The program's flexibility allows the trip generation process to be designed to maximize the use of local data and provides a quantum improvement for the TxDOT in the trip generation process.

This manual is supplementary and complementary to two prior reports which detail the program specifications for TRIPCAL5 (1) and the instructions for the setup and operation of TRIPCAL5 (2). Following this section, the program options are discussed in more detail to provide the reader with some understanding of the overall capability of TRIPCAL5. The subroutines in TRIPCAL5 are presented in the third section with a brief discussion of the function and purpose of each. The fourth section presents a cross-reference of the subroutines and functions which indicates the routines that are called from each program. The fifth section

presents a description of each of the variables in the program by labeled common statements and the sixth section presents a cross-reference of the labeled common statements and the subroutines containing these statements. The seventh section presents a description of the sorts and sort keys that are used in the program. The eighth section presents the data set formats used in the program, and the ninth section discusses how the data flow through the program. The tenth section presents a discussion of the results of the program tests which were done and, the last section presents a summary. The documentation of the default models contained in the program is provided as an appendix to this report.

PROGRAM OPTIONS

TRIPCAL5 is designed to be flexible and allow different options for the user in the development of estimates of travel demand. This allows the trip generation process to be designed to use available data and improve the overall travel demand estimates. The different options available include trip production models, trip attraction models, disaggregation models, multiple trip purposes, and user-selected data inputs.

Up to 10 trip purposes may be used with specific trip rates or models for each. The only limitation is that the same type of cross-classification model must be used for each run where those trip purposes are being estimated using a cross-classification model.

Four trip production models may be used to estimate trip productions. A two-way cross-classification model may be selected and trip rates stratified for up to six categories for each independent variable. A three-way cross-classification model may be selected and trip rates stratified for up to six categories for two of the independent variables and up to four categories for the third independent variable. A simple linear regression may also be used with up to six independent variables. The fourth optional production model is a cross- classification regression model which requires a different program setup under a two-way cross-classification option. This setup is discussed following this section.

Five trip attraction models may be used to estimate trip attractions. A two-way cross-classification model may be selected and attraction trip rates stratified for up to six categories for each independent variable. A three-way cross-classification model may be selected and attraction trip rates stratified for up to six categories for two of the independent variables and up to four categories for the third independent variable. A cross-classification regression model may also be specified with trip rates stratified for up to 24 generation areas and by households and employment type. A simple linear regression model may also be used with up to six independent variables. The fifth option available for attractions is the use of a two-tier regression model. Each regression model may use up to six independent variables.

The disaggregation models provided in TRIPCAL5 are for three variables: households by household size, households by household income, and households by auto ownership. For any one of these three variables, the user may chose to input the marginal distribution for each zone, input a disaggregation curve for the urban area which is used to develop a marginal distribution for each zone, let the default model in the program compute the marginal distribution for each zone, or use a combination of these three methods. While TRIPCAL5 is oriented to those three variables, the user may select and use any other variable desired as long as the marginal distribution is input for each zone.

TRIPCAL5 is designed to use socioeconomic data normally used in trip generation. Provisions are included to allow the user to input and use non-typical variables and/or combinations of typical variables. User selected data may be input and used in either cross-classification models or regression models.

TRIPCAL5 also provides the option of selecting the information to be output from the program. Depending upon the model selected and size of the area, the output can exceed 100,000 lines of print. Options are provided for to select the output and, in certain reports, select the zones for which the information is desired.

DESCRIPTION OF SUBROUTINES

TRIPCAL5 has 87 subroutines including the main program. The names of each subroutine and a brief description of what it does and/or its purpose is presented in the following discussions:

AUDFLT	This subroutine produces the default distribution of auto ownership. The output is the number of households with 0 autos, 1 autos, 2 autos and 3 or more autos. The input variables are income and number of households.
AUDSUG	This subroutine disaggregates auto ownership data from the auto ownership disaggregation curve data (AU) cards. The output is the number of households with 0 autos, 1 autos, 2 autos, and 3 or more autos. The input variables are income and number of households.
BLDHDG	This subroutine builds headings for most reports. The trip purpose names are built into headings of up to 11 characters and 4 lines. The subroutine separates words from the heading and places some words on separate lines to build a heading for each trip purpose. A separate subheading of "P" and "A" is built under the trip purpose heading.
BLOCK	This subroutine initializes data in labeled common areas. Some of the data are for missing values. Other data are for character descriptions.
B10HD	This subroutine builds headings for some reports. The trip purpose names are built into headings of up to 10 characters and 5 lines. The subroutine separates words from the heading and places some words on separate lines to build a heading for each trip purpose.
B9HD	This subroutine builds headings for some reports. The trip purpose names are built into headings of up to 9 characters and 5 lines. The subroutine separates words from the heading and places some words on separate lines to build a heading for each trip purpose.
CHKCC	This subroutine checks cross-classification production and attraction rates against models specified on production trip rate (PT) data cards and attraction trip rate (AT) data cards. A check is made to see if the correct number of rows, columns, and planes are specified for each trip purpose.
CHKZN	This subroutine checks to see that the required zonal data are present. The data cards checked are the DA1, DA2, DA3, DA4, and DA5. DA1 data cards are always required. DA2 data cards are required only if regression models reference data in the DA2 data cards. DA3 data cards are required

if a production cross-classification is run and the default marginal code, COLOPT, on the production column information (PCI) data card specifies no default column marginal model. DA3 data cards are also required if a two-way or three-way attraction cross-classification model is run. DA4 data cards are required if the production row information (PRI) data card specifies no row marginal default model, ROWOPT. DA4 data cards are also required if a two-way or three-way attraction cross-classification model is run. DA5 data cards are required if the production depth information (PDI) data card specifies no depth marginal default model, DPTOPT, and a three-way production cross-classification model is run. DA5 data cards are also required if a three-way attraction cross-classification is run.

CONTRL	This subroutine reads control and data cards. All control and model data must be in a group prior to the zonal data. This subroutine calls other subroutines to check and process the actual data.
CPY6	This subroutine copies a temporary output data set to Unit 6. This subroutine allows the output reports to be produced in the same order on different runs.
DISAG2	This subroutine does a two-way marginal disaggregation for a two-way cross-classification model.
DISAG3	This subroutine does a three-way marginal disaggregation for a three-way cross-classification model.
DMOD	This subroutine reads zone data sorted by district, zone, and card type and calculates district model attractions.
DZAG	This subroutine disaggregates the two-tier attractions and then sorts by zone.
FILLMR	This subroutine fills marginals for two- and three-way cross-classification models by calling subroutines INCDSG, INDFLT, HHSDSG, HHSDFL, AUDSG, and/or AUDFLT. The marginals are either filled from the default models or from other models if they are available.
HHSDFL	This subroutine fills the household size marginal from the default model.
HHSDSG	This subroutine fills the household size marginal from disaggregation data input on the HH data cards.

Description of Subroutines

HSRNG	This subroutine is called by ZMOD to build interpolation coefficients for the Household Size Model if the HS cards are input.
INCDSG	This subroutine fills the income marginal with data input from the IC data.
INDFLT	This subroutine fills the income marginal from the default model. Data from the IR data cards are used.
LNEBLD	This subroutine is used to print ranges by subroutine CHKZN.
MAIN	This is the main program for TRIPCAL5. It calls various subroutines to control the reading data, produce reports, sort data, and run various requested models.
PRTOE	This subroutine prints tables of equals. The tables of equals are sector, district, and area type.
RDACI	This subroutine reads Attraction Column Information (ACI) data cards.
RDACR	This subroutine reads Regional Distribution for Attraction (ACR) Cross-Classification data cards.
RDACV	This subroutine reads Attraction Cross-Classification Variable Number Record (ACV) data cards. This data card has been superseded by the Cross-Classification Variable Number Record (CCV) data card.
RDADI	This subroutine reads Attraction Depth Information (ADI) data cards.
RDAD1	This subroutine reads Input Data Card 1 (DA1) zonal data cards. The data contained on this data card type are acres, population, households, average household size, household income, total employment, total basic employment, total retail employment, and total service employment. The zone number is checked for the range of 1 to the number of zones from the PS data card. The other numbers are checked to see that they are positive or zero. A check is made to see that the three classes of employment sum to the total employment. The number of errors found is summed in variable NRR. The record is copied to Unit 1 with table of equals added for district, sector, and area type. The added data are written as three fields of I5 added to the end of the record.
RDAD2	This subroutine reads DA2 zonal data cards. The data contained on this data card is user-defined data. The zone number is checked for the range

	of 1 to the number of zones from the PS data card. If an error is found, 1 is added to the NRR variable. The user-defined fields are checked to see if they are positive or zero. If any values are negative, NWARN is incremented by 1.
RDAD3	This subroutine reads DA3 and DA4 zonal data cards. The zone number is checked for the range of 1 to the number of zones from the PS data card. The DA3 data card contains production column marginal data and attraction column marginal data DA4 contains Production Row and Attraction Row Marginal Data. The attraction column or row marginal data is required for a two-way or three-way attraction cross-classification model. The marginal data are checked to see if they are positive; any negative values cause 1 to be added to the NRR variable. If all marginal data are blank then an error message is written; and 1 is added to NRR.
RDAD5	This subroutine reads DA5 zonal data cards. The zone number is checked for the range of 1 to the number of zones from the PS data card.
RDAMC	This subroutine reads and edits attraction cross-classification data cards.
RDAMR	This subroutine reads and edits attraction linear regression model data cards.
RDARI	This subroutine reads and edits attraction row information data cards.
RDAT	This subroutine reads and edits attraction trip rate data cards.
RDAU	This subroutine reads and edits auto ownership disaggregation curve data cards.
RDCCV	This subroutine reads and edits the cross-classification variable number data card.
RDCMT	This subroutine reads and edits special generator COMMENT information data cards.
RDDR	This subroutine reads and edits attraction district regression model data cards.
RDDZR	This subroutine reads and edits district to zone regression model data cards.
RDEA	This subroutine reads and edits area type table of equals data cards.

Description of Subroutines

RDED	This subroutine reads and edits district table of equals data cards.
RDES	This subroutine reads and edits sector table of equals data cards.
RDHH	This subroutine reads and edits household disaggregation curve data cards.
RDHS	This subroutine reads and edits household size ranges data cards.
RDIC	This subroutine reads and edits income disaggregation curve data cards.
RDIR	This subroutine reads and edits income ranges data cards.
RDNAM	This subroutine reads and edits independent variable name data cards.
RDPCI	This subroutine reads and edits production column information data cards.
RDPCR	This subroutine reads and edits regional distribution for production cross-classification data cards.
RDPCT	This subroutine reads and edits percent of trips by category index and Trip purpose data cards.
RDPDI	This subroutine reads and edits production depth information data cards.
RDPMR	This subroutine reads and edits production simple regression model data cards.
RDPRI	This subroutine reads and edits production row information data cards.
RDPS	This subroutine reads and edits the program control/specification record data card.
RDPT	This subroutine reads and edits production trip rate data cards.
RDSEL	This subroutine reads and edits the zones to be printed data cards.
RDTBL	This subroutine reads and edits reports to be printed data cards.
RDTP	This subroutine reads and edits the trip purpose record data cards.
RPT1	This subroutine prints Table 1: TRIP MODEL INPUT.

Description of Subroutines

RPT15	This subroutine prints Table 15: STUDY AREA CHARACTERISTICS SUMMARY BY SECTOR.
RPT16	This subroutine prints Table 16: STUDY AREA CHARACTERISTICS SUMMARY BY ZONE WITHIN SECTOR.
RPT17	This subroutine prints Table 17: STUDY AREA CHARACTERISTICS SUMMARY BY AREA TYPE.
RPT18	This subroutine prints Table 18: STUDY AREA CHARACTERISTICS SUMMARY BY ZONE WITHIN AREA TYPE.
RPT19	This subroutine prints Table 19: STUDY AREA CHARACTERISTICS SUMMARY BY ZONE.
RPT20	This subroutine prints Table 20: FINAL PRODUCTIONS AND ATTRACTIONS.
RPT21	This subroutine prints Table 21: ZONE COMMENTS.
RPT5	This subroutine prints Table 5: REGIONAL DISTRIBUTION INPUT. Part A is for two-way cross-classification data and part B is for three-way cross-classification data.
R12HD	This subroutine builds the variable part of headings for Reports 6, 7, and 8.
SCALE	This subroutine calculates scale factors to either the total attractions to the total productions or the total productions to the total attractions by trip purpose as specified. This subroutine also prints Table 10: SCALING FACTOR COMPUTATIONS.
SGAO	This subroutine read special generator and add on cards.
SORT	This subroutine calls subroutine INVOKE with one of seven different sort statements as indexed by the argument JTYPE.
STAMP	This subroutine calls subroutine DATER for the date. It then prints the date and the TRIPCAL5 capacities for data in terms of zones, districts, sectors, and area types.
SUMDAT	This subroutine counts the number of data cards by zone for data card

	types DA1, DA2, DA3, DA4, DA5, AOA, AOP, SGA, SGP, DA3, DA4, DA5, CMT, and SGZ.
T12INT	This subroutine initializes arrays for summation for Tables 6, 7, and 8.
XPT	This subroutine prints four tables whose headings follow. Table 11: AGGREGATE PRODUCTIONS AND ATTRACTIONS BY SECTOR, Table 12: AGGREGATE PRODUCTIONS AND ATTRACTIONS BY ZONE WITHIN SECTOR, Table 13: AGGREGATE PRODUCTIONS AND ATTRACTIONS BY AREA TYPE, and Table 14: AGGREGATE PRODUCTIONS AND ATTRACTIONS BY ZONE WITHIN AREA TYPE.
XPT6	This subroutine calls subroutines T12INT, R12HD, X6A, X6B, X6D, and X6X to print Table 6: DISAGGREGATE ZONE RESULTS. The subroutine XPT7 is called to print Table 7, and subroutine XPT8 is called to print Table 8.
XPT7	This subroutine prints Table 7: DISAGGREGATE SECTOR RESULTS. This subroutine calls subroutines X6A, X6B, X6D, and X6X to print TABLE 7.
XPT8	This subroutine prints Table 8: DISAGGREGATE AREA TYPE RESULTS.
XPT9	This subroutine prints TABLE 9: TOTAL UNCALLED PRODUCTION AND ATTRACTION RESULTS BY ZONE AND TRIP PURPOSE.
XPT9HD	This subroutine builds headings for subroutine XPT9.
X6A	This subroutine is used to print the main part of production and attraction cross-classification models disaggregate trips for Tables 6, 7, and 8. Table 6 is printed by zone, Table 7 is printed by sector, and Table 8 is printed by area type.
X6B	This subroutine is used to print the main part of production and attraction regression models disaggregate trips for Tables 6, 7, and 8. Table 6 is printed by zone, Table 7 is printed by sector, and Table 8 is printed by area type.
X6D	This subroutine is used to print the totals part of production and attraction regression models disaggregate trips for Tables 6, 7, and 8. Table 6 is

printed by zone, Table 7 is printed by sector, and Table 8 is printed by area type.

X6X	This subroutine is used to print the totals from production and attraction cross-classification models disaggregate trips for Tables 6, 7, and 8. Table 6 is printed by zone, Table 7 is printed by sector and Table 8 is printed by area type.
ZMOD	This subroutine does zonal modeling.
ZNDAT	This subroutine controls reading of zonal data cards. After editing for errors, the CMT records are written to Unit 4. All other records are copied to Unit 1.

CROSS-REFERENCE OF SUBROUTINES AND FUNCTIONS

The following listing presents a cross-reference of the subroutines and functions called from each routine. The calling program is listed first and is followed by a listing of the subroutines and functions called from that program.

Calling Program	**Subroutines Called**
CHKZN	LNEBLD, NUMDT
CONTRL	RDACI, RDACR, RDACV, RDADI, RDAMC, RDAMR, RDARI, RDAT, RDAU, RDCCV, RDDR, RDDZR, RDEA, RDED, RDES, RDHH, RDHS, RDIC, RDIR, RDNAM, RDPCI, RDPCR, RDPCT, RDPDI, RDPMR, RDPRI, RDPS, RDPT, RDSEL, RDTBL, RDTP
FILLMR	AUDFLT, AUDSG, HHSDFL, HHSDSG, INCDSG, INDFLT
MAIN	CHKCC, CHKZN, CONTRL, DMOD, DZAG, PRTOE, RPT1, RPT15, RPT16, RPT17, RPT18, RPT19, RPT20, RPT21, RPT5, SCALE, SORT, STAMP, XPT, ZMOD, ZNDAT
RDAD1	SUMDAT
RDAD2	SUMDAT
RDAD3	SUMDAT
RDAD5	SUMDAT
RDAMC	ITPCHK
RDAMR	ITPCHK
RDAT	ITPCHK
RDCMT	SUMDAT
RDDR	ITPCHK

Calling Program	Subroutines Called
RDDZR	ITPCHK
RDPMR	ITPCHK
RDPT	ITPCHK
RDTP	ITPCHK
RPT15	BLDHDG
RPT16	BLDHDG
RPT17	BLDHDG
RPT18	BLDHDG
RPT19	BLDHDG
RPT20	BLDHDG
RPT21	NUMDT, SORT
R12HD	B10HD, B9HD
SGAO	SUMDAT
SORT	INVOKE
STAMP	DATER
XPT	BLDHDG
XPT6	CPY6, R12HD, T12INT, X6A, X6B, X6D, X6X, XPT7, XPT8
XPT7	CPY6, R12HD, X6A, X6B, X6D, X6X
XPT8	CPY6, R12HD, X6A, X6B, X6D, X6X
XPT9	CPY6, XPT9HD

Calling Program	**Subroutines Called**
ZMOD	BLDHDG, DISAG2, DISAG3, FILLMR, HSRNG, XPT6, XPT9
ZNDAT	RDAD1, RDAD2, RDAD3, RDAD5, RDCMT, SGAO

DESCRIPTION OF VARIABLES BY LABELED COMMON

Included in TRIPCAL5 are 12 labeled common statements. These provide an efficient method of utilizing variables between routines. Each labeled common name follow with a listing of the variables included in the common statement and what each contains.

SUBSCRIPT DICTIONARY FOR LABELED COMMON DESCRIPTIONS

The following subscripts have a common meaning for all labeled common blocks. All names described below are either 2 or 3 characters long. Subscripts with a length of 1 character generally have a different meaning for each occurrence.

Name	Range	Description
ITP	1-10	Trip purpose number.
IAT	1-24	Area type number.
JR	1-6	Row index for cross-classification models.
JROW	1-6	Row index for cross-classification models.
JC	1-6	Column index for cross-classification models.
JCOL	1-6	Column index for cross-classification models.
JD	1-4	Plane index for cross-classification models.
JDPTH	1-4	Plane index for cross-classification models.
JSECT	1-99	Sector number.
MTP	1-3	Production model type index.
MTA	1-5	Attraction model type index.
IZ	1-9999	Zone number.
JZONE	1-9999	Zone number.
JTBL	1-21	Table number (printed report number).
JRINC	1-25	10 times (the ratio of zonal median income to regional median income).
ICH	1-6	Household income category.
IHS	1-6	Household size category.

JRHS	1-25	1 to 10 times (the ratio of zonal average household size to regional average size index).
JREG	1-10	Index into array of production or attraction regression trip purposes.
JCC	1-10	Index into array of production or attraction cross-classification trip purposes.

LABELED COMMON ATTR

This labeled common groups variables and arrays used in attraction calculations.

Variable	**Description**
MDLTA	Not used.
MDLOPA(ITP)	Attraction model type code by trip purpose.
MDLSMA(ITP)	Not used.
NIVA(ITP,IAT)	Number of regression variables for the trip purpose (indicated by variable ITP) and the area type (indicated by variable IAT).
IVA(ITP,IAT,L)	Regression variable index for each trip purpose (ITP), area type (IAT), and regression variable L. This variable indexes into the array RGV in subroutine ZMOD. The RGV array contains zonal data from the DA1 data card and user zonal data from the DA2 data card.
ACOEF(ITP,IAT,L)	This array contains attraction regression model coefficients. If the regression model is not by area type, then ACOEF(ITP,1,1) contains the constant coefficient and ACOEF(ITP,1,L+1) contains the coefficient for regression variable L. If the regression is by area type then ACOEF(ITP,IAT,L+1) contains the coefficient for area type IAT and regression variable L; and a constant coefficient is not used.
NMACOL	Number of attraction cross-classification column classes.
NMAROW	Number of attraction cross-classification row classes.

NMAPLN	Number of attraction cross-classification depth classes.
NIVD(ITP)	Number of independent variables for attraction district regression models.
IVD(ITP,J)	Independent variable numbers for attraction district regression models for trip purpose ITP and variable J.
DCOEF(ITP,L)	Coefficients for attraction district regression model for trip purpose ITP and variable L-1. The constant coefficient for the attraction district regression model is in DCOEF(ITP,1).
NIVDZ(ITP)	Number of independent variables for attraction district to zone allocation regression model for trip purpose ITP.
IVDZ(ITP,L)	IVDZ(ITP,L) contains the regression variable index into array RGV for variable L of the district to zone allocation model.
DZCOEF(ITP,L)	Coefficients for district to zone allocation model. This allocation model is a regression which allocates the attractions from the two-tier attraction model for a particular trip purpose back to a zone. DZCOEF(ITP,1) contains the constant term of the regression equation. DZCOEF(ITP,L+1) contains the coefficient for variable L in the regression equation.
ATRREG(JR,JC,JD)	Regional cross-classification for either two-way or three-way attraction cross-classification. JR is the row index, JC is the column index, and JD is the depth index.
ATRATE(JR,JC,JD,ITP)	Attraction trip rates for two-way and three-way cross-classification models.
JACV	Attraction plane variable index in the array RGV. This variable is used only for attraction three-way cross-classification models.

LABELED COMMON ATTRC

This labeled common groups character variables and arrays used in attraction calculations.

Variable	Description
COLNMA	The name of the column variable for two-way or three-way attraction cross-classification models.
COLDSA(J)	The description of the J'th attraction column cross-classification.
ROWNMA	The name of the row variable for two-way or three-way attraction cross-classification models.
ROWDSA(K)	The description of the K'th attraction row cross-classification.
PLNNMA	The name of the depth variable for three-way attraction cross-classification models.
PLNDSA(M)	The description of the M'th attraction depth cross-classification.

LABELED COMMON CHR

This labeled common groups the header and the regression and/or cross-classification character variable names.

Variable	Description
HDR	Header for TRIPCAL5 run.
NAMRV(L)	Name of regression variables and cross-classification variables.

LABELED COMMON CTL

This labeled common groups various control variables.

Variable	Description
NZONE	Number of zones.
NSECT	Number of sectors.

NGEN	Number of generation areas.
JYEAR	Data year for TRIPCAL5 run.
JPRS	Vehicle or person trips flag.
JDISTM	One purpose model flag.
JAON	Add-on records code.
JDSPEC	Read from columns 45-46 of the PS card but not used.
CPI	1967 consumer price index.
EOD	End of data flag for Unit 5.
FATAL	Fatal error flag.
NRR	Number of data errors.
NWARN	Number of data warnings.
TPPCT(ITP)	Control total for trip purpose ITP.
TPACT(ITP)	Not used.
TPBAL(ITP)	Trip purpose balance option: 0 = balance to attractions, 1 = balance to productions.
PRCT(L,ITP)	Production plane information.
REGINC	Regional median income.
REGHHS	Regional household size.
INCFLG	Flag for presence of income disaggregation data and income ranges.
HHFLG	Household size disaggregation curve flag. This flag is set to true if exactly 25 HH records have been read.
AOFLG	Auto ownership disaggregation curve data flag. This flag is set to true if any AU records have been read.

SUMPOP	The total population read from the DA1 records.
SUMHH	The total households read from the DA1 records.
SUMINH	The sum of the households multiplied by the household income from all DA1 record.
NTPP(MTP)	Integer array which is a collection of three counts of production model types. NTPP(1) is the count of production two-way cross-classification models. NTPP(2) is the count of production three-way cross-classification models. NTPP(3) is the count of production regression models. NTPP indices 4-6 are not used but are included, because the similar attraction array NTPA is dimensioned 6.
NTPA(MTA)	Integer array which is a collection of five counts of attraction model types. NTPA(1) is the count of attraction two-way cross-classification models. NTPA(2) is the count of attraction three-way cross-classification models. NTPA(3) is the count of simple attraction regression models. NTPA(4) is the count of attraction cross-classification regression models. NTPA(5) is the count of attraction two-tier regression models.

LABELED COMMON CTLCHR

This labeled common groups the header, date, trip purpose names, urban area name, and the reread data.

Variable	**Description**
RRD	This character variable contains the last data card read. It is used to reread the data card by a format which is appropriate for the record type.
TPNAME(ITP)	Trip purpose name.
INAME(L)	Independent variable name.
URBARA	Name of urban area.

DATE	Date that TRIPCAL5 started execution of present run.

LABELED COMMON EQUALS

This labeled common groups the following variables:

Variable	Description
NUMAT	Number of zones equivalenced to area types.
MXATZN	Maximum zone number equivalenced to an area type
MXAT	Maximum area type.
NUMED	Number of zones equivalenced to districts.
MXEDZN	Maximum zone number equivalenced to a district.
MXED	Maximum district.
NUMES	Number of zones equivalenced to sectors.
MXESZN	Maximum zone number equivalenced to a sector.
MXES	Maximum sector.
JEQAT(IZ)	Area type table of equals.
JEQD(IZ)	District table of equals.
JEQES(IZ)	Sector table of equals.

LABELED COMMON PROD

This labeled common groups production model information.

Variable	Description
MDLTP(ITP)	Non-home based or Truck-taxi trip purpose code: 0 = other, 1 = Non-home based, or 2 = Truck-taxi trip purpose.

MDLOPT(ITP)	Production trip purpose code: 1 = two—way cross-classification model, 2 = three—way cross-classification model, 3 = simple regression model, and 6 = use attraction model for productions.
MDLSM	Not used.
MDLDG	Not used.
NUMCOL	The number of columns in a two-way production cross-classification model or a three-way production cross-classification model.
NUMROW	The number of rows in a two-way production cross-classification model or a three-way production cross-classification model.
NUMPLN	The number of planes in a three-way production cross-classification model.
NAU	The number of auto ownership disaggregation curve (AU) records.
AUX(K,JAU)	This array holds the data from the auto ownership disaggregation curve (AU) records. The index JAU is the record index. The index K is for the six data items on one record.
NIVP(ITP)	Number of independent variables by trip purpose for the production regression models.
IVP(ITP,JV)	Production regression model variable numbers for up to six variables indexed by JV for each trip purpose ITP.
PCOEF(ITP,JV+1)	Production regression model coefficients for trip purpose ITP and regression variable JV. PCOEF(ITP,1) is the constant coefficient for trip purpose ITP.
RDU(JR,JC,JD)	Regional cross-classification for either two-way or three-way production cross-classification. JR is the row index, JC is the column index, and JD is the depth index.
PT(JR,JC,JD,ITP)	Production trip rates for two-way and three-way cross-classification models.

DINC(ICH,JRINC)	Data from income disaggregation curves (IC). The index ICH is the household income category. Multiplying 0.1 by JRINC is the ratio of the zonal median income to regional median income(nondimensionalized income). The nondimensionalized income has a range of 0.1 to 2.5 in steps of 0.1.
RNGINC(ICH)	Data from the income ranges (IR) data card. These data are the ending value for each income range.
DHH(IHS,JRHS)	Data from household size disaggregation curves (HH). The IHS index is the household size category index. JRHS is the ratio of zonal average household size to regional average size index (nondimensionalized household at size index). The nondimensionalized household size index is converted to the value by multiplying by 0.1 and adding 1.
RNGHS(IHS)	This array contains data from the household size ranges read from the household size ranges (HS). The values are the ending value of the household size range by IHS index the household size category index.
JPCV	The index of the production variable distributed by the cross-classification. The default is 3 which is households. See Table 4 of the **TRIPCAL5 Users Manual** for the independent variable number.

LABELED COMMON PRODC

This labeled common groups production description information.

Variable	Description
COLNAM	Production column name for production cross-classification models.
COLDES(JC)	Column JC description.
ROWNAM	Production row name for production cross-classification models.
ROWDES(JR)	Row JR description.

PLNNAM	Production depth name for production three-way cross-classification model.
PLNDES(JD)	JD depth description.
COLOPT	Default column marginal model code.
ROWOPT	Default row marginal model code.
DPTOPT	Default depth marginal model code.

LABELED COMMON RPTS

This labeled common groups printing control variables.

Variable	Description
MXTBL	Maximum number of tables that the program controls printing (currently 50).
TBL(JTBL)	Logical array. If the value is true, the table JTBL is printed.
SEL(JZONE)	Controls printing zone JZONE in tables. If SEL(JZONE) is true, output for zone JZONE is printed in tables.

LABELED COMMON SC

This labeled common groups special generator data.

Variable	Description
PROD(ITP)	Sum of the unscaled productions by trip purpose.
AOP(ITP)	Sum of the add-on productions by trip purpose.
SGP(ITP)	Sum of the production special generators by trip purpose.
SGA(ITP)	Sum of the attraction special generators by trip purpose.
AOA(ITP)	Sum of the add-on attractions by trip purpose.

ATT(ITP)	Sum of the unscaled attractions by trip purpose.
SCP(ITP)	Production scale factor by trip purpose.
SCA(ITP)	Attraction scale factor by trip purpose.
RESIDP(ITP)	Production residual rounding factor by trip purpose.
RESIDA(ITP)	Attraction residual rounding factor by trip purpose.
THH	Sum of the study area households.
TRETL	Study area total retail employment.
TBASIC	Study area total basic employment.
TSERVC	Study area total service employment.
TRKOPT	Default truck-taxi control total model flag.

LABELED COMMON T12

This labeled common groups subtotals for reports and indices by specific model types for reports.

Variable	**Description**
NCRP	Count of production regression models.
JQCRP(JREG)	The trip purpose numbers which are production regression models. The JREG index varies from 1 to NCRP.
NCRA	Count of attraction regression models.
JQCRA(JREG)	The trip purpose numbers which are attraction regression models.
NCCP	Count of production cross-classification models.
JQCCP(JCC)	The trip purpose numbers which are production cross-classification models.

NCCA	Count of attraction cross-classification models.
JQCCA(JCC)	The trip purpose numbers which are attraction cross-classification models.
CCPS(JCOL,JROW,	Production total for cross-classification models. JCOL JDPTH,JPUR,JSECT)is the column index. JROW is the row index. JDPTH is the depth index if a three-way cross-classification is used; otherwise JDPTH is 1. JPUR is the purpose index; JPUR = 11 is a sum of all purposes. JSECT is the sector index.
CCPSAO(ITP,JSECT)	Production add-on total. ITP is the purpose index. JSECT is the sector index.
CCPSSG(ITP,JSECT)	Production special generator total. ITP is the purpose index. JSECT is the sector index.
CCAS(JCOL,JROW, JDPTH,JPUR,JSECT)	Attraction total for cross-classification models. JCOL is the column index. JROW is the row index. JDPTH is the depth index if a three-way cross-classification is used; otherwise JDPTH is 1. JPUR is the purpose index; JPUR = 11 is a sum of all purposes. JSECT is the sector index.
CCASAO(ITP,MXSECT)	Attraction add-on total. ITP is the purpose index. JSECT is the sector index.
CCASSG(ITP,JSECT)	Attraction special generator total. ITP is the purpose index. JSECT is the sector index.
CCPT(JCOL,JROW, JDPTH,JPUR,JAT)	Production total for cross-classification models. JCOL is the column index. JROW is the row index. JDPTH is the depth index if a three-way cross-classification is used; otherwise JDPTH is 1. JPUR is the purpose index; JPUR = 11 is a sum of all purposes. JAT is the area type index.

CCPTAO(ITP,JAT)	Production add-on total. ITP is the purpose index. JAT is the area type index.
CCPTSG(ITP,JAT)	Production special generator total. ITP is the purpose index. JAT is the area type index.
CCAT(JCOL,JROW, JDPTH,JPUR,JAT)	Attraction total for cross-classification models. JCOL is the column index. JROW is the row index. JDPTH is the depth index if a three-way cross-classification is used; otherwise JDPTH is 1. JPUR is the purpose index; JPUR = 11 is a sum of all purposes. JAT is the area type index.
CCATAO(ITP,JAT)	Attraction add-on total. ITP is the purpose index. JAT is the area type index.
CCATSG(ITP,JAT)	Attraction special generator total. ITP is the purpose index. JAT is the area type index.
CCP(JCOL,JROW, JDPTH,JPUR)	Cross-classification production subtotal by trip purpose index JPUR. Index 11 of JPUR is a subtotal for all trip purposes. JCOL is the column index. JROW is the row index. JDPTH is the depth index if a three-way cross-classification is used; otherwise JDPTH is 1.
CCA(JCOL,JROW, JDPTH,JPUR)	Cross-classification attraction subtotal by trip purpose index JPUR. Index 11 of JPUR is a subtotal for all trip purposes. JCOL is the column index. JROW is the row index. JDPTH is the depth index if a three-way cross-classification is used; otherwise JDPTH is 1.

LABELED COMMON XFAC

This labeled common groups the following variables:

Variable	**Description**
HHSMDL	This flag indicating that a default household size model can be used because an HS data record has been read.
INCMDL	Not used.
AUTOMD	Not used.

XF(J,L)	Interpolation factor produced by subroutine HSRNG and used by subroutine HHSDFL. Index J is the user range described by beginning of RNGS(J) and ending of RNGE(J) in subroutine HSRNG. The index L is the integer household size data plus 1. The value in XF(J,L) is the proportion of the integer household range that lies in the user range.

CROSS-REFERENCE OF LABELED COMMON NAMES AND SUBROUTINES

The following is a listing of each labeled common name and the subroutines which contain the labeled common statement.

Labeled Common	Subroutines Using Labeled Common
ATTR	BLOCK, CHKCC, CHKZN, DMOD, DZAG, MAIN, RDACI, RDACR, RDACV, RDADI, RDAMC, RDAMR, RDARI, RDAT, RDCCV, RDDR, RDDZR, RDTP, RPT1, RPT5, R12HD, T12INT, XPT6, XPT7, XPT8, XPT9HD, ZMOD
ATTRC	BLOCK, CHKCC, CHKZN, DMOD, DZAG, MAIN, RDACI, RDACR, RDACV, RDADI, RDAMC, RDAMR, RDARI, RDAT, RDCCV, RDDR, RDDZR, RDTP, RPT1, RPT5, R12HD, T12INT, XPT6, XPT7, XPT8, XPT9HD, ZMOD
CHR	BLOCK, MAIN, PRTOE, RPT1, RPT15, RPT16, RPT17, RPT18, RPT19, RPT20, RPT21, RPT5, XPT, XPT6, XPT7, XPT8, XPT9HD, X6A, X6B, X6D, X6X, ZMOD
CTL	BLDHDG, BLOCK, CHKCC, CHKZN, CONTRL, DMOD, FILLMR, ITPCHK, MAIN, PRTOE, RDACI, RDACR, RDACV, RDADI, RDAD1, RDAD2, RDAD3, RDAD5, RDAMC, RDAMR, RDARI, RDAT, RDAU, RDCCV, RDCMT, RDDR, RDDZR, RDEA, RDED, RDES, RDHH, RDHS, RDIC, RDIR, RDNAM, RDPCI, RDPCR, RDPCT, RDPDI, RDPMR, RDPRI, RDPS, RDPT, RDSEL, RDTBL, RDTP, RPT1, RPT15, RPT16, RPT17, RPT18, RPT19, RPT20, RPT21, RPT5, R12HD, SCALE, SGAO, STAMP, T12INT, XPT, XPT6, XPT7, XPT8, XPT9, XPT9HD, X6A, X6B, X6D, X6X, ZMOD, ZNDAT
CTLCHR	BLDHDG, BLOCK, CHKCC, CHKZN, CONTRL, DMOD, FILLMR, ITPCHK, MAIN, PRTOE, RDACI, RDACR, RDACV, RDADI, RDAD1, RDAD2, RDAD3, RDAD5, RDAMC, RDAMR, RDARI, RDAT, RDAU, RDCCV, RDCMT, RDDR, RDDZR, RDEA, RDED, RDES, RDHH, RDHS, RDIC, RDIR, RDNAM, RDPCI, RDPCR, RDPCT, RDPDI, RDPMR, RDPRI, RDPS, RDPT, RDSEL, RDTBL, RDTP, RPT1, RPT15, RPT16, RPT17, RPT18, RPT19, RPT20, RPT21, RPT5, R12HD, SCALE, SGAO, STAMP, T12INT, XPT, XPT6, XPT7, XPT8, XPT9, XPT9HD, X6A, X6B, X6D, X6X, ZMOD, ZNDAT
EQUALS	BLOCK, CHKZN, DMOD, DZAG, MAIN, RDAD1, RDAD2, RDAD3, RDAD5, RDCMT, RDEA, RDED, RDES, RPT1, RPT15, RPT16, RPT17,

Labeled Common	Subroutines Using Labeled Common
	RPT18, RPT19, RPT5, SGAO, XPT, XPT6, XPT7, XPT8, XPT9HD, ZMOD
NUMDAT	CHKZN, RPT21
PROD	AUDSG, BLOCK, CHKCC, CHKZN, DISAG2, DISAG3, FILLMR, HHSDFL, HHSDSG, HSRNG, INCDSG, MAIN, RDAU, RDCCV, RDHH, RDHS, RDIC, RDIR, RDPCI, RDPCR, RDPDI, RDPMR, RDPRI, RDPT, RDTP, RPT1, RPT5, R12HD, SCALE, T12INT, XPT6, XPT7, XPT8, XPT9HD, ZMOD
PRODC	AUDSG, BLOCK, CHKCC, CHKZN, DISAG2, DISAG3, FILLMR, HHSDFL, HHSDSG, HSRNG, INCDSG, MAIN, RDAU, RDCCV, RDHH, RDHS, RDIC, RDIR, RDPCI, RDPCR, RDPDI, RDPMR, RDPRI, RDPT, RDTP, RPT1, RPT5, R12HD, SCALE, T12INT, XPT6, XPT7, XPT8, XPT9HD, ZMOD
RPTS	BLOCK, MAIN, RDSEL, RDTBL, RPT15, RPT16, RPT17, RPT18, RPT19, RPT20, RPT21, SCALE, XPT, XPT6, XPT7, XPT8, ZMOD
SC	RDPS, SCALE, ZMOD
T12	R12HD, T12INT, XPT6, XPT7, XPT8, ZMOD
XFAC	HHSDFL, HSRNG

DESCRIPTION OF SORTS

The following describes the sorts and sort keys that are executed in TRIPCAL5.

Sort 1: Sort data are ZONE DATA after the table of equals have been added. The sort key is DISTRICT, ZONE number, and DATA CARD TYPE.

```
SORT FIELDS=(81,5,CH,A,4,5,CH,A,1,3,CH,A)
```

Sort 2: Sort data are district to zonal regression allocation values. The sort key is District number, ZONE number, and DATA CARD TYPE.

```
SORT FIELDS=(91,5,CH,A,4,5,CH,A,1,3,CH,A)
```

Sort 3: Sort data are productions and attractions disaggregated from district data. The sort key is ZONE number.

```
SORT FIELDS=(1,10,CH,A)
```

Sort 4: Sort data are ZONE DATA. The sort key is ZONE number and DATA CARD TYPE (columns 1-3 of the data records).

```
SORT FIELDS=(4,5,CH,A,1,3,CH,A)
```

Sort 5: TABLES 11-14 data are productions and attractions by sector and zone. The sort key is RECORD type, SECTOR number, and ZONE number.

```
SORT FIELDS=(1,20,CH,A)
```

Sort 6: Sort data are comment records. The sort key is ZONE number and INPUT LINE number.

```
SORT FIELDS=(4,5,CH,A,81,5,CH,A)
```

DATA SET FORMATS

This section presents a brief description and discussion of the data set formats used in TRIPCAL5.

ZONAL DATA SORT DATA SET[1]

This data set has two record types which are a data record and a trailer record.

Zonal Data Sort Format

Columns	**Format**	**Description**
1-80	A80	Zonal data as read from input data, see TRIPCAL5 User's Guide.
81-85	I5	District equated to zone.
86-90	I5	Sector equated to zone.
91-95	I5	Area type equated to zone.

SCALED TRIPS SORT DATA SET

This data set has six record types. The first character of each record specifies the major record type. Column 9 specifies the minor record type of "A" or "B" which is necessary if more than five trip purposes are run.

---

[1] A trailer record for this data set is written. This trailer record contains a character string of 85 nines.

SA Sector Sort Record

Columns	Format	Description
1	A1	"S".
2-4	I3	Sector number.
5-8	I4	Zone number.
9	A1	"A".
10-19	I10	Production for trip purpose 1.
20-29	I10	Attraction for trip purpose 1.
30-39	I10	Production for trip purpose 2.
40-49	I10	Attraction for trip purpose 2.
50-59	I10	Production for trip purpose 3.
60-69	I10	Attraction for trip purpose 3.
70-79	I10	Production for trip purpose 4.
80-89	I10	Attraction for trip purpose 4.
90-99	I10	Production for trip purpose 5.
100-109	I10	Attraction for trip purpose 5.

SB Sector Sort Record

Columns	Format	Description
1	A1	"S".
2-4	I3	Sector number
5-8	I4	Zone number
9	A1	"B".
10-19	I10	Production for trip purpose 6
20-29	I10	Attraction for trip purpose 6
30-39	I10	Production for trip purpose 7
40-49	I10	Attraction for trip purpose 7
50-59	I10	Production for trip purpose 8
60-69	I10	Attraction for trip purpose 8
70-79	I10	Production for trip purpose 9
80-89	I10	Attraction for trip purpose 9
90-99	I10	Production for trip purpose 10
100-109	I10	Attraction for trip purpose 10

SD Sector Sort Record

Columns	Format	Description
1	A1	"S".
2-4	I3	Sector number.
5-8	I4	Zone number.
9	A1	"D".
10-12	A3	"DA1".
13-17	I5	Zone number.
18-24	F7.0	Special generator acreage.
25-31	F7.0	Special generator population.
32-38	F7.0	Special generator households.
39-45	F7.0	Special generator persons/household.
46-52	F7.0	Special generator zonal income.
53-59	F7.0	Special generator total employment.
60-66	F7.0	Special generator basic employment.
67-73	F7.0	Special generator service employment.
74-80	F7.0	Special generator retail employment.
81-89	19X	Blank.
90-99	F10.5	Zonal autos per household.

TA Area Type Sort Record

Columns	Format	Description
1	A1	"T".
2-4	I3	Area type number.
5-8	I4	Zone number.
9	A1	"A".
10-19	I10	Production for trip purpose 1.
20-29	I10	Attraction for trip purpose 1.
30-39	I10	Production for trip purpose 2.
40-49	I10	Attraction for trip purpose 2.
50-59	I10	Production for trip purpose 3.
60-69	I10	Attraction for trip purpose 3.
70-79	I10	Production for trip purpose 4.
80-89	I10	Attraction for trip purpose 4.
90-99	I10	Production for trip purpose 5.
100-109	I10	Attraction for trip purpose 5.

TB Area Type Sort Record

Columns	Format	Description
1	A1	"T".
2-4	I3	Area type number.
5-8	I4	Zone number.
9	A1	"B".
10-19	I10	Production for trip purpose 6.
20-29	I10	Attraction for trip purpose 6.
30-39	I10	Production for trip purpose 7.
40-49	I10	Attraction for trip purpose 7.
50-59	I10	Production for trip purpose 8.
60-69	I10	Attraction for trip purpose 8.
70-79	I10	Production for trip purpose 9.
80-89	I10	Attraction for trip purpose 9.
90-99	I10	Production for trip purpose 10.
100-109	I10	Attraction for trip purpose 10.

TD Area Type Sort Record

Columns	Format	Description
1	A1	"T".
2-4	I3	Area type number.
5-8	I4	Zone number.
9	A1	"D".
10-12	A3	"DA1".
13-17	I5	Zone number.
18-24	F7.0	Special generator acreage.
25-31	F7.0	Special generator population.
32-38	F7.0	Special generator households.
39-45	F7.0	Special generator persons/household.
46-52	F7.0	Special generator zonal income.
53-59	F7.0	Special generator total employment.
60-66	F7.0	Special generator basic employment.
67-73	F7.0	Special generator service employment.
74-80	F7.0	Special generator retail employment.
81-89	19X	Blank.
90-99	F10.5	Zonal autos per household.

DATA FLOW

Data flows through several character data sets and six sorts. Following this description of data flow there are four cross-references of data set reads and data set writes. Next there is a table of data set usage by data set type.

ZONAL DATA INPUT AND EDITING

Basic zonal data input is controlled by subroutine ZNDAT. Subroutine ZNDAT calls subroutines RDAD1, RDAD2, RDAD3, RDAD5, and SGAO to read and edit zonal data. The data are written to Unit 1 by these subroutines. Subroutine RDAD1 reads and edits AD1 data. Subroutine RDAD2 reads and edits AD2 data. Subroutine RDAD3 reads and edits AD3 and AD4 data. Subroutine RDAD5 reads and edits AD5 data. Subroutine SGAO edits special generator data. The data card types processed are AOA, AOP, SGP, and SGA. All of these subroutines write records to Unit 1. The output records have the district number, the sector number, and the area type number added to the end of the data card record in a format of 3I5. Subroutine ZNDAT writes a trailer record of all nines to the end of Unit 1.

DATA FLOW IN SUBROUTINE SORT

Subroutine SORT calls subroutine INVOKE which links to the system sort package. The system sort package uses Unit 9 as input data and writes its output to Unit 10. The sorting order is specified by a sort statement which specifies primary and secondary sort fields. For each sort field, the following data are specified: sort field beginning column, size of the field, data type, and ascending or descending sort. All sorts by TRIPCAL5 are specified as character data and ascending sorts.

DATA FLOW IN THE MAIN PROGRAM

The MAIN program calls subroutines which do data transfers. Also the MAIN program copies data sets to the sort input which is Unit 9. The MAIN program also saves data from one sort by copying the data to Unit 13.

- Subroutine ZNDAT is called to read zonal data cards and write this data to Unit 1 with district number, sector number, and area type number added on.
- DO loop 20 in the MAIN program copies the Unit 1 data to Unit 9, the sort input file.
- A call to SORT(1) sorts the zonal data (Unit 9) by district number. The sort records are written to Unit 10.
- Subroutine DMOD is called if district modeling is done and writes district attraction totals to Unit 12 as binary records. District to zone disaggregation data are written to Unit 11 as binary records.
- A call to SORT(2) {SORT FIELDS=(91,5,CH,A,4,5,CH,A,1,3,CH,A)} sorts data from Unit 9 by area type and writes the sorted records to Unit 10.
- Subroutine DZAG is called to disaggregate district data to zone data. This subroutine reads binary Units 11 and 12 which were written by DMOD.
- A call to SORT(3) {SORT FIELDS=(1,10,CH,A)} sorts the district attraction data by zone. The sort records are written to Unit 10.
- DO loop 25 in the MAIN program copies sort out (Unit 10) to Unit 13.
- DO loop 30 in the MAIN program copies Unit 1 (zonal data) to Unit 9 (SORTIN).
- A call to SORT(4) {SORT FIELDS=(4,5,CH,A,1,3,CH,A)} sorts the zonal data by zone and card type. The sort records are written to Unit 10.
- Subroutine ZMOD is called by the MAIN program to produce the unscaled trips. Subroutine ZMOD reads sorted zonal data from Unit 10.
- Subroutine ZMOD is called by the MAIN program to produce the scaled trips. Subroutine ZMOD reads sorted zonal data from Unit 10.
- DO loop 35 in the MAIN program copies Unit 2 (temporary final productions and attractions) to Unit 9 (sortin).
- Subroutine SORT(5) {SORT FIELDS=(1,20,CH,A)} is called. This sorts by zone.
- Subroutine RPT21 is called which calls subroutine SORT(6) {SORT FIELDS=(4,5,CH,A,81,5,CH,A)}. This sorts the temporary data by district.

DATA SET VERSUS SUBROUTINE CROSS-REFERENCES

These cross-references are for all FORTRAN input output data sets except Units 5 and 6. The subroutine INVOKE, which is shown in these tables, does not do input and output; but it calls the system sort routine and passes the data set name of FT09F001 (Unit 9) to the system sort as sort input. It also passes the data set name of FT10F001 (Unit 10) to the system sort as sort output. The output written to Unit 8 is dependent on the variable DEBUG which is coded in several different subroutines to FALSE. When debugging is done on these subroutines this variable is changed to TRUE.

CROSS-REFERENCE OF FORMATTED WRITES

Unit	Subroutine Names
1	RDAD1, RDAD2, RDAD3, RDAD5, SGAO, ZNDAT
2	ZMOD
8	CHKCC, CHKZN, DISAG3, RPT1, T12INT, XPT6, ZMOD
9	DZAG, MAIN, RPT21
10	INVOKE
13	MAIN
14	X6A, X6X
15	X6B, X6D
16	X6B, X6D
17	X6A, X6X
18	XPT9
20	ZMOD
21	ZMOD
22	ZMOD

CROSS-REFERENCE OF FORMATTED READS

Unit	Subroutine Names
1	MAIN
2	MAIN, RPT20
9	RPT19, INVOKE
10	DMOD, MAIN, RPT15, RPT16, RPT17, RPT18, RPT21, XPT, ZMOD
13	ZMOD

14	CPY6
15	CPY6
16	CPY6
17	CPY6
18	CPY6

CROSS-REFERENCE OF UNFORMATTED WRITES

Unit	Subroutine Names
11	DMOD
12	DMOD

CROSS-REFERENCE OF UNFORMATTED READS

Unit	Subroutine Names
11	DZAG
12	DZAG

DATA SET USAGE BY UNIT NUMBER

Unit	Data Set Use
1	Zonal data with equivalences added.
2	Temporary data for Tables 11-20.
8	Debugging output.
9	Sort input.
10	Sort output.
11	District to zone allocation values.
12	District attractions.
13	Zonal attractions from district attraction model.
14	Temporary report output for reports 6, 7, and 8.
15	Temporary report output for reports 6, 7, and 8.
16	Temporary report output for reports 6, 7, and 8.
17	Temporary report output for reports 6, 7, and 8.
18	Temporary report output for report 9.
20	Generation card output for trip purposes 1 to 4.
21	Generation card output for trip purposes 5 to 8.
22	Generation card output for trip purposes 9 and 10.

PROGRAM TESTING

As previously discussed, TRIPCAL5 was designed to be flexible and adaptable to meet the individual needs of an urban area. As such, different options were built into the program which allow the user to estimate trip productions and attractions in a variety of ways using different models and/or variables. As a part of the software development, every effort was made to insure that each option and/or possible combination of options would execute properly. This section presents a description of the major tests that were done and a comparison of the results of some of those tests with other programs which use different models.

The primary objective of testing TRIPCAL5 was to insure the program would execute properly and produce the results expected. A secondary objective was to perform tests which would allow comparison with other trip generation programs in terms of overall results. The purpose of a comparison was to determine if the results from TRIPCAL5 were within a reasonable range of those being produced and used from other trip generation programs. While an effort was made to insure that the same basic input data were used, there were some instances where certain assumptions were made in order to be able to run the program and test different options.

One primary data set was used in the testing. It was obtained from the North Central Texas Council of Governments (NCTCOG) for the Dallas-Fort Worth area and consisted of the zonal demographic data for nearly 6,000 traffic analysis zones used in their trip generation program. The tests performed with TRIPCAL5 used the data aggregated to a 605-zone level (i.e., the NCTCOG Regional Analysis Area level). A second data set from the Austin urban area was also used in a comparison test to illustrate the difference in the results from TRIPCAL5 from those from TRIPCAL3 and TRIPCAL4. It consisted of the demographic data for 635 serial zones used in travel demand forecasting in Austin.

The trip generation program used in the Dallas-Fort Worth area was developed by the NCTCOG. The model used in the program is a cross-classification model for both productions and attractions which are estimated for seven trip purposes. Four of those trip purposes are home based work trips which are estimated for different income groups. Trip productions and attractions were also estimated for home based non-work, non-home based and other trip purposes. Home based work, home based non-work, and non-home based trip productions were estimated using a two-way cross-classification model where trip rates (i.e., trips per household) were stratified by four income groups (income quartiles) and six household size groups. Home based work attractions were estimated using a three-way cross-classification model where trip rates were stratified by four income groups, five area types, and three employment categories. Home based non-work and non-home based attractions were estimated using a cross-classification regression type model where trip rates were stratified by five area types for three categories of employment and households. Other trip productions and attractions were estimated using a cross-classification regression type model where trip rates were stratified by five area types for three categories of employment and households. The models developed for the Dallas-Fort Worth area were based on a 1984 travel survey and were developed specifically for that area. TRIPCAL5, with its flexibility, is capable of using the same models for estimating trip productions and attractions.

Table 1 presents a listing of the tests which were done using the demographic data from the Dallas-Fort Worth area. The results of the runs are not presented because they could be misleading and misinterpreted. In several cases, input data were created for the purpose of testing the program to determine if the program would properly execute that particular option. For example, the simple and two-tier regression models were assumed and not based on any type of data or analysis. It must be understood that the examples shown in Table 1 are not all inclusive and many others are both possible and likely. These were done to insure that the program was operating correctly. The assumption is that since these executed properly, the other options available would also execute properly.

COMPARISON TESTS

Two types of comparison tests were done with TRIPCAL5. The first was to determine if the program was operating correctly in terms of producing about the same results as another program when both were using the same models. The Dallas-Fort Worth data were used for this comparison since results were available from the NCTCOG trip generation program which was using state-of-the-art models and provided a unique opportunity to test different setups in TRIPCAL5. The second was to compare the results of TRIPCAL5 with results from the old trip generation programs used by the TxDOT, TRIPCAL3 and TRIPCAL4. Data from the Austin Transportation Study were used for that test.

Table 1
TRIPCAL5 Test Runs Using Dallas-Fort Worth Input Data

Run No.	Trip Purpose	Production Model	Attraction Model	Income Model	H.H. Size Model	Auto Model
1	HBNW[1]	2-Way Cross-Classification	Regression Type Cross-Class.	Yes	Yes	No
	NHB[2]	2-Way Cross-Classification	Regression Type Cross-Class.	Yes	Yes	No
2	HBNW	2-Way Cross-Classification	2-Tier Regression	Yes	Yes	No
	NHB	Simple Regression	Simple Regression	Yes	Yes	No
3	HBW[3]	2-Way Cross-Classification	Regression Type Cross-Class.	Yes	Yes	No
	Other[4]	Simple Regression	Regression Type Cross-Class.	Yes	Yes	No
4	Dummy	Simple Regression	Regression Type Cross-Class.	No	No	No
	Dummy	Simple Regression	3-Way Cross-Classification	No	No	No
5	Other	2-Way Cross-Classification	Regression Type Cross-Class.	No	No	No
6	HBW	3-Way Cross-Classification	3-Way Cross-Classification	Yes	Yes	No
7	Other	2-Way Cross-Classification	Regression Type Cross-Class.	No	No	Yes

1 Home Based Non-Work
2 Non-Home Based
3 Home Based Work
4 Includes all trip purposes other than those specifically defined

The trip generation program used in the Dallas-Fort Worth area was unique in that different variables were used in the cross-classification models depending on the trip purpose, and the number of cross-classification categories were not the same for all trip purposes. Tables 2 and 3 present the models used. TRIPCAL5 was designed to be flexible enough to allow most models to be used in trip generation. This flexibility has certain limitations which require multiple runs for certain situations. The use of TRIPCAL5 to estimate trip productions and attractions for the Dallas-Fort Worth area using the same models as used by NCTCOG required more than one run. This was necessary due to the use of different variables and the number of categories in the cross-classification models as shown in Tables 2 and 3. For example, TRIPCAL5 does not estimate trip productions for both a two-way cross-classification model and a three-way cross-classification model in the same run. In that situation, it is necessary to run the program for each type of production model.

Table 2
Dallas-Fort Worth
Trip Production Models

Trip Purpose	Model Type	No. Rows[1]	Row Variable	No. Column[2]	Column Variable	No. Depth[3]	Depth Variable
Home Based Work Income Quartile 1	2-Way Cross-Classification	1	Income Quartile 1	6	Household Size	NA	NA
Home Based Work Income Quartile 2	2-Way Cross-Classification	1	Income Quartile 2	6	Household Size	NA	NA
Home Based Work Income Quartile 3	2-Way Cross-Classification	1	Income Quartile 3	6	Household Size	NA	NA
Home Based Work Income Quartile 4	2-Way Cross-Classification	1	Income Quartile 4	6	Household Size	NA	NA
Home Based Non-Work	2-Way Cross-Classification	4	Income Quartiles	6	Household Size	NA	NA
Non-Home Based	2-Way Cross-Classification	4	Income Quartiles	6	Household Size	NA	NA
Other	2-Way Cross-Classification	4	Employees & Households[4]	5	Area Type	NA	NA

[1] Number of Row Categories
[2] Number of Column Categories
[3] Number of Depth Categories
[4] Three Categories of Employment

Table 3
Dallas-Fort Worth
Trip Attraction Models

Trip Purpose	Model Type	No. Rows	Row Variable	No. Columns	Column Variable	No. Depth	Depth Variable
Home Based Work Income Quartile 1	3-Way Cross-Classification	1	Income Quartile 1	5	Area Type	3	Emp.[1]
Home Based Work Income Quartile 2	3-Way Cross-Classification	1	Income Quartile 2	5	Area Type	3	Emp.
Home Based Work Income Quartile 3	3-Way Cross-Classification	1	Income Quartile 3	5	Area Type	3	Emp.
Home Based Work Income Quartile 4	3-Way Cross-Classification	1	Income Quartile 4	5	Area Type	3	Emp.
Home Based Non-Work	2-Way Cross-Classification	4	Employees & Households	5	Area Type	NA	NA
Non-Home Based	2-Way Cross-Classification	4	Employees & Households	5	Area Type	NA	NA
Other	2-Way Cross-Classification	4	Employees & Households	5	Area Type	NA	NA

[1] Employment, i.e., Basic, Retail, and Service

The second comparison test done was different from the first. The first test (using Dallas-Fort Worth data) replicated the trip generation models used in the NCTCOG area. The second test was set up to illustrate the difference between the results from the old programs, TRIPCAL3 and TRIPCAL4, and TRIPCAL5 when using the default models contained within TRIPCAL5. These models allowed a comparison between the two methods for estimating trip productions. The data used in the test were from the Austin transportation study. The models used in Austin were unique in that a three-way cross-classification model was employed for estimating trip productions. This was accomplished by using of different zonal incomes to represent zones with different average household size. For example, zones falling in the medium income range (usually indicated by inputting a household income of $17,000) were given a household income of $17,000, $17,002, or $17,004 depending upon whether the zone's average household size was between 1 and 2.5, between 2.5 and 3.0, or 3.0 and more. This enabled the TRIPCAL4 program to utilize additional trip rates for zones depending upon their average household size. While this provided an improvement in the estimation of trips, the models were still applied in an aggregate manner, even though the trip production model was a three-way cross-classification. The test which was done used the same three-way cross-classification model and trip rates, but the default models in TRIPCAL5 were used to disaggregate the data at the zone level and then apply the trip rates. The difference in the results illustrates the difference in the model applications for each method.

Tables 4 and 5 present the results of both comparisons. As shown in Table 4, TRIPCAL5 produced basically the same estimates as those produced by the trip generation program used in the Dallas-Fort Worth area by the NCTCOG. The differences between the two are attributable to different methods of rounding used in the trip generation programs, the different levels used in computing trip productions and attractions, and the slightly different methods of estimating the marginal distributions. The NCTCOG program generates productions and attractions at the traffic analysis zone level and aggregates the zonal estimates to the regional analysis area level. For purposes of testing, the productions and attractions were estimated in TRIPCAL5 at the regional analysis area level. It is of interest to note that the difference between the total trip productions estimated by the NCTCOG program and the TRIPCAL5 program was less than 1 percent. Overall, the results indicated in Table 4 are comparable and indicate that the TRIPCAL5 program can be set up for generating estimates of trip productions and attractions for the Dallas-Fort Worth area using the same type of models as used by the NCTCOG.

The results shown in Table 5 are quite different. In that test using data from the Austin transportation study, TRIPCAL5 was set up with the same type model as used in the TRIPCAL3/4 program. The differences in the trip productions are due to the use of disaggregate data at the zone level for estimating the trip productions. This is a primary difference between the model theory applied in TRIPCAL5 and that used in TRIPCAL3/4. It should be noted that the trip attractions were the same for both programs. This was expected since both use the same models and same data. The difference in total trips being produced was only 5.1 percent. The lower estimate was produced by the TRIPCAL5 program. This appears to be insignificant until further examination of the differences

Table 4
Comparison Results for Dallas-Fort Worth

Trip Purpose		Dallas - Fort Worth Region		
		NCTCOG[1] Program	TRIPCAL5 Program	Percent Difference
Home Based Work Income Quartile 1	P[2]	469476	499800	1.065
	A[3]	455387	454038	-0.30
Home Based Work Income Quartile 2	P	652847	645808	-1.08
	A	669717	686678	2.53
Home Based Work Income Quartile 3	P	818143	805009	-1.61
	A	858756	866724	0.93
Home Based Work Income Quartile 4	P	928936	902219	-2.88
	A	918522	897858	-2.25
Home Based Non-Work	P	5063069	5047754	-0.30
	A	4876147	4698677	-3.64
Non-Home Based	P	2855511	2750383	-3.68
	A	3204212	3187949	-0.51
Other	P	1621484	1636122	0.90
	A	1586413	1610085	1.49

[1] North Central Texas Council of Governments
[2] Productions
[3] Attractions

between the estimates of trip productions by trip purpose and estimates at the zone level. The estimate of home based work trips from TRIPCAL5 was 14 percent less than the estimate from TRIPCAL4. The estimate of home based non-work trips from TRIPCAL5 was 8.4 percent less than the estimate from TRIPCAL4. The estimate of non-home based trips

Table 5
Comparison Results Between TRIPCAL3/4 and TRIPCAL5 for Austin

Trip Purpose		Austin Person Trips (Unscaled)		
		TRIPCAL3/4	TRIPCAL5	Percent Difference
Home Based Work	P[1]	1881848	1617317	-14.06
	A[2]	1880494	1880493	0.00
Home Based Non-Work	P	3474594	3184596	- 8.35
	A	3860049	3860048	0.00
Non-Home Based	P	1832108	2021574	10.34
	A	1635149	1635149	0.00

[1] Productions

[2] Attractions

from TRIPCAL5 was 10.3 percent higher than the estimate from TRIPCAL4. It was expected that overall, TRIPCAL5 would generally estimate fewer trips than TRIPCAL4 due to the disaggregation of the data at the zone level. It was initially surprising that TRIPCAL5 produced higher estimates of non-home based trips. This was a result of the disaggregation and the application of the trip rates at the disaggregate level. To illustrate the differences between the two models at the zone level, 12 zones were randomly selected. A comparison of relevant statistics and number of trips produced by trip purpose for each zone is presented in Tables 7 and 8. For all but the low income zones, TRIPCAL5 estimated fewer trips. This is consistent with earlier findings (1) where a comparison was done between TRIPCAL4 and TRIPCAL5. That comparison found that TRIPCAL4 typically underestimated the number of trips for low income zones. It is of interest to note that while the difference in the overall total number of trips was about 5 percent, the

Table 6
Comparison of Various Statistics from TRIPCAL4 and TRIPCAL5
Austin Data

Zone	Household Size	Household Income	Autos Per Household		Autos Per Person		Trips Per Household		Trips Per Person	
			TRIPCAL4	TRIPCAL5	TRIPCAL4	TRIPCAL5	TRIPCAL4	TRIPCAL5	TRIPCAL4	TRIPCAL5
44	2.26	$ 25,000	1.92	1.64	0.85	0.72	10.19	9.64	4.51	4.27
53	2.26	$ 25,000	1.92	1.64	0.85	0.72	10.89	10.47	4.82	4.63
84	2.04	$ 25,000	1.92	1.64	0.94	0.80	11.32	10.87	5.55	5.33
85	2.20	$ 25,000	1.92	1.64	0.87	0.74	12.15	11.75	5.52	5.33
189	2.52	$ 35,002	2.19	1.77	0.87	0.70	9.25	8.10	3.67	3.21
306	2.22	$ 35,000	2.22	1.76	1.00	0.79	10.89	10.23	4.90	4.60
331	1.21	$ 8,000	1.13	1.11	0.93	0.92	5.44	5.81	4.49	4.80
422	3.22	$ 12,504	1.37	1.34	0.43	0.42	10.69	9.69	3.32	3.01
424	2.40	$ 8,000	1.13	1.06	0.47	0.44	6.14	6.85	2.56	2.85
431	2.75	$ 35,002	2.19	1.79	0.79	0.65	16.70	15.70	6.06	5.71
598	2.06	$ 25,000	1.92	1.64	0.93	0.79	8.79	8.06	4.27	3.92
606	2.77	$ 17,502	1.59	1.49	0.57	0.54	8.73	7.66	3.15	2.77

Table 7
Comparison of Trip Productions from TRIPCAL4 and TRIPCAL5
Austin Data

Zone	Household Size	Household Income	Home Based Work Trips		Home Based Non-Work Trips		Non-Home Based Trips		Total Trips	
			TRIPCAL4	TRIPCAL5	TRIPCAL4	TRIPCAL5	TRIPCAL4	TRIPCAL5	TRIPCAL4	TRIPCAL5
44	2.26	$ 25,000	652	584	1225	1118	476	524	2353	2294
53	2.26	$ 25,000	1230	1103	2312	2110	1208	1352	4750	4565
84	2.04	$ 25,000	1535	1360	2886	2606	1739	1948	6160	5914
85	2.20	$ 25,000	1123	1003	2111	1922	1600	1751	4834	4677
189	2.52	$ 35,002	1705	1485	3263	2825	158	177	5126	4487
306	2.22	$ 35,000	26	23	50	45	22	24	98	92
331	1.21	$ 8,000	3321	4339	9805	9646	1089	1202	14215	15187
422	3.22	$ 12,504	855	662	1563	1366	757	849	3175	2877
424	2.40	$ 8,000	264	360	778	801	236	264	1278	1425
431	2.75	$ 35,002	163	144	312	274	410	414	885	832
598	2.06	$ 25,000	4167	3695	7834	7094	997	1123	12998	11913
606	2.77	$ 17,502	6348	5074	11108	9999	1005	2600	18461	17673

differences by zone and trip purpose showed more variability and ranged in values from -12.5 percent to 11.5 percent. This implies that the differences in the number of trips at the zone level could have an impact on the results from later stages of the travel demand modeling process.

The tests done using TRIPCAL5 indicate that the program is correctly computing the estimates of trip productions and attractions for the different models as designed into the program. Since the program has a great deal of flexibility and allows the user to be creative in modeling efforts, every test conceivable was not done. It is anticipated that with increased usage of the program, corrections and changes to the program will be made as needed.

REFERENCES

1. D. Pearson and G.B. Dresser. *TRIPCAL5 - Program Specifications Informational Report #6*. Research Report 1235-2. Texas State Department of Highways and Public Transportation Research Study 2-10-90-1235. Texas Transportation Institute, The Texas A&M University System, College Station, Texas, August 1991.

2. Unpublished report on the development and calibration of the trip generation models for Atlanta, Georgia, based on a 1980 auto use survey and 1972 home interview survey.

3. G.B. Dresser and J.L. Gattis. *Household Travel Surveys*. Texas State Department of Highways and Public Transportation Research Study 2-10-88-947. Texas Transportation Institute, The Texas A&M University System, College Station, Texas, February 1989.

4. D. Pearson, C.E. Bell, and G.B. Dresser. *TRIPCAL5 Users Manual.* Research Report 1235-3. Texas State Department of Highways and Public Transportation Research Study 2-10-90-1235. Texas Transportation Institute, The Texas A&M University System, College Station, Texas, November 1990.

5. Alan M. Voorhees & Associates. *Factors and Trends in Trip Lengths.* National Cooperative Highway Research Program Report 48. Highway Research Board, National Research Council, 1968.

TECHNICAL APPENDIX

AUTO OWNERSHIP/AVAILABILITY DISAGGREGATION MODEL

The number of trips made by household members has been found to be highly correlated with income, household size, and autos owned or available to the household. In general, the more autos owned or available to a household, the greater the propensity to travel. For this reason, autos owned or available is used in many disaggregate cross-classification models for estimating trip productions. It has been used in estimating trip productions in Texas for many years.

In developing TRIPCAL5 a new trip generation model for the Texas Department of Transportation (TxDOT), the flexibility to use variable trip production and attraction models was included. To ensure that TxDOT would be able to continue using past models it was necessary that TRIPCAL5 have the capability to use autos owned or available as one of the independent variables in the cross-classification model(s). Thus, TRIPCAL5 must include a means to either input data, by which households could be disaggregated into auto ownership categories, or it must have a default model which would accomplish this at the zone level. In the following sections, autos owned and autos available are considered to be the same, but the terms autos owned or auto ownership will be used. The following sections describe the methods by which the number of households within a zone may be disaggregated in categories of auto ownership.

PROBLEM STATEMENT

Using cross-classification trip generation models, it is necessary to be able to disaggregate households at the zone level into the appropriate categories being used in the model. For example, if household trip rates are cross-classified by household size and auto ownership, the number of households within each category (e.g., one person household with one auto) must be estimated in order to apply the trip rate and develop estimates of trip productions for the zone. If autos owned is one of the independent variables in the cross-classification model, a method by which households could be disaggregated into the auto ownership categories at the zone level is necessary.

PROPOSED ALTERNATIVES

Three alternatives were identified for obtaining estimates of the number of households by auto ownership category at the zone level. The first was to let the distribution of households by auto ownership be input directly for each zone. This is a feature of TRIPCAL5 whereby the distribution of households (in percentages) may be input for each independent variable being used in the cross-classification model for each zone. If autos owned is one of those variables, zonal distributions can be input for individual zones or for all zones.

The second alternative was to input a disaggregation curve for the urban area under study. This curve(s) would consist of a table of values which, given a mean income for a zone, would provide a percentage of households estimated to have 0, 1, 2, or 3+ autos. This alternative assumes that the distribution of households by auto ownership at the zone level will be similar to that of the urban area, depending upon the mean income for the zone.

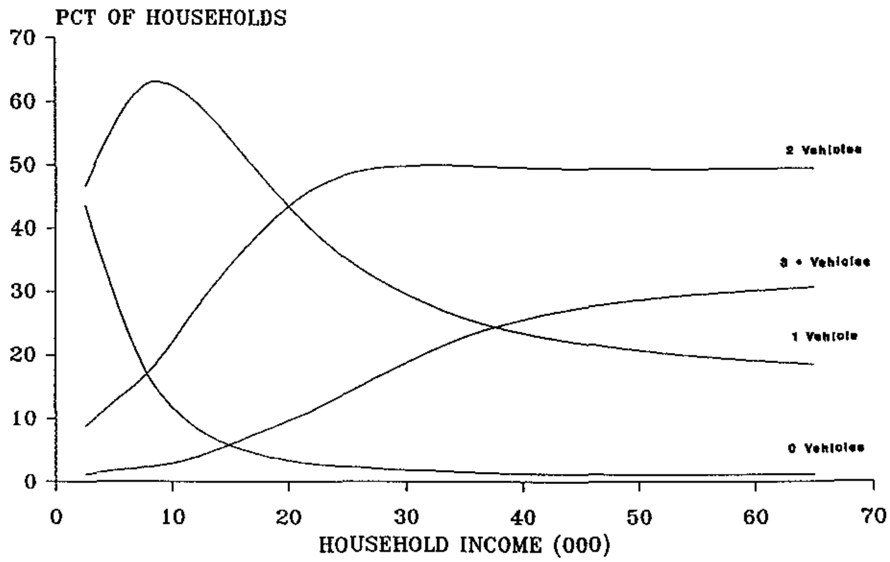
The third alternative was to develop a theoretical method for estimating a distribution of households by auto ownership at the zone level.

The following sections describe the methodology for developing the second and third alternatives. The methodology for the first alternative must be developed by the local area planning agency.

DISAGGREGATION CURVE METHODOLOGY

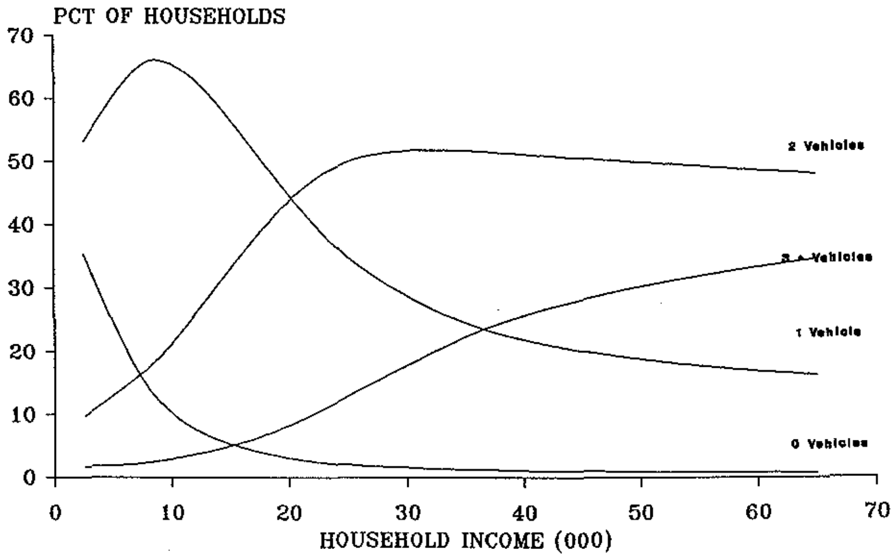
The development of disaggregation curve(s) for an urban area should be accomplished using census data. Using data from the 1980 Census Urban Transportation Planning Package for four Metropolitan Statistical Areas (MSA) within Texas, curves reflecting the distribution of households by auto ownership and income were plotted and are shown in Figures 1 through 4. The general shape of the curves is very similar; and when plotted on the same graph (shown in Figures 5 through 8) the shapes are almost identical. This similarity implies that one curve could be developed for each level of auto ownership and could be used to estimate the percentage of households at each income level for any one of the four MSAs. This was done by averaging the observations and hand fitting curves to the data for the four MSAs combined (i.e., Figure 5).

Figure 1
Distribution of Autos Available
By Household Income – San Antonio



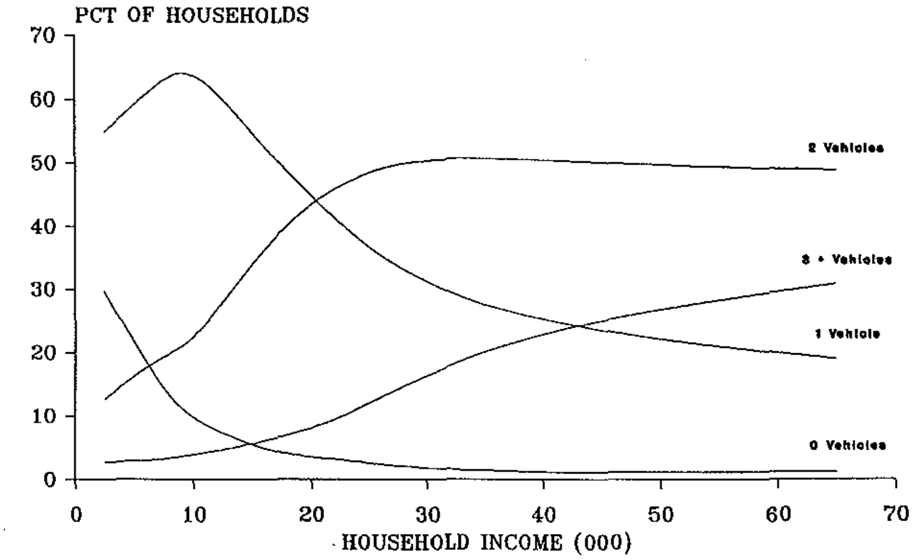
1980 UTPP DATA
San Antonio MSA

Figure 2
Distribution of Autos Available
By Household Income – Dallas–Ft Worth



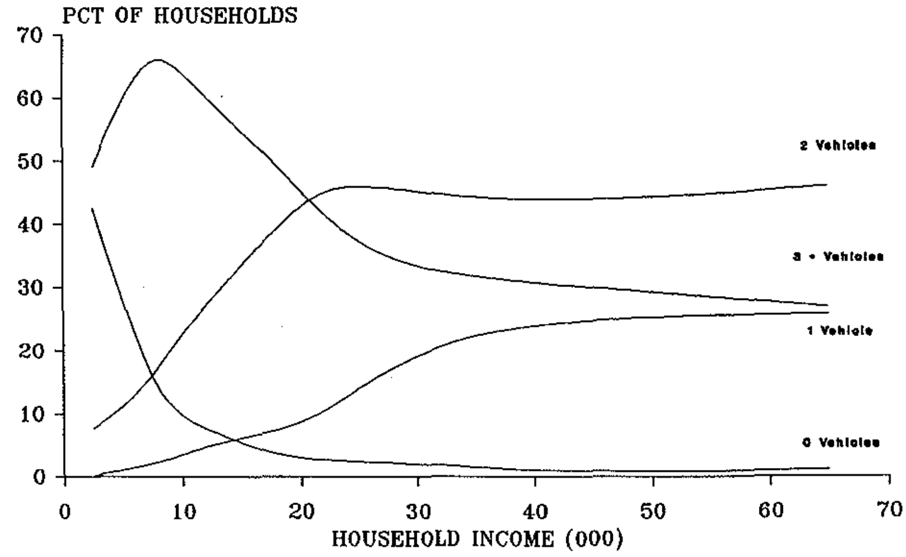
1980 UTPP DATA
Dallas – Ft Worth MSA

Figure 3
Distribution of Autos Available
By Household Income – Austin



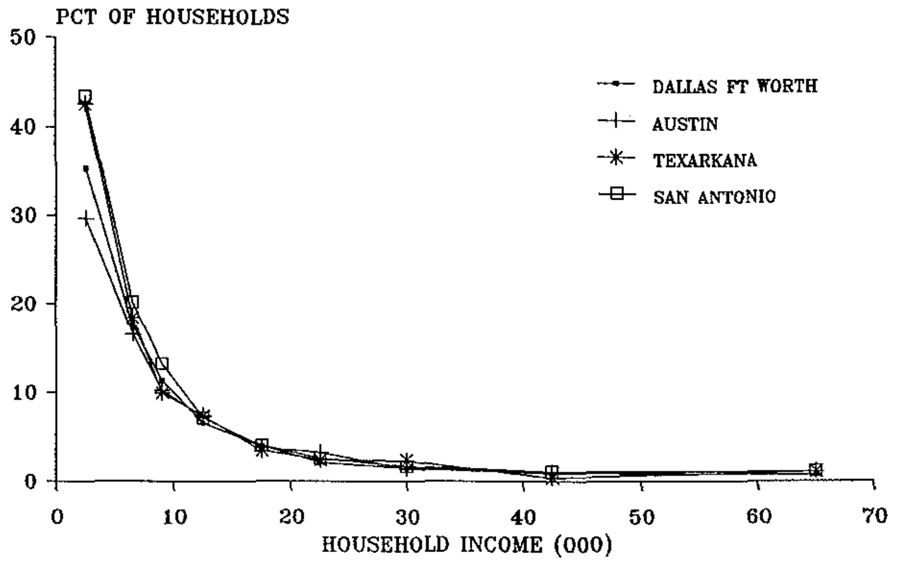
1980 UTPP DATA
Austin MSA

Figure 4
Distribution of Autos Available
By Household Income – Texarkana



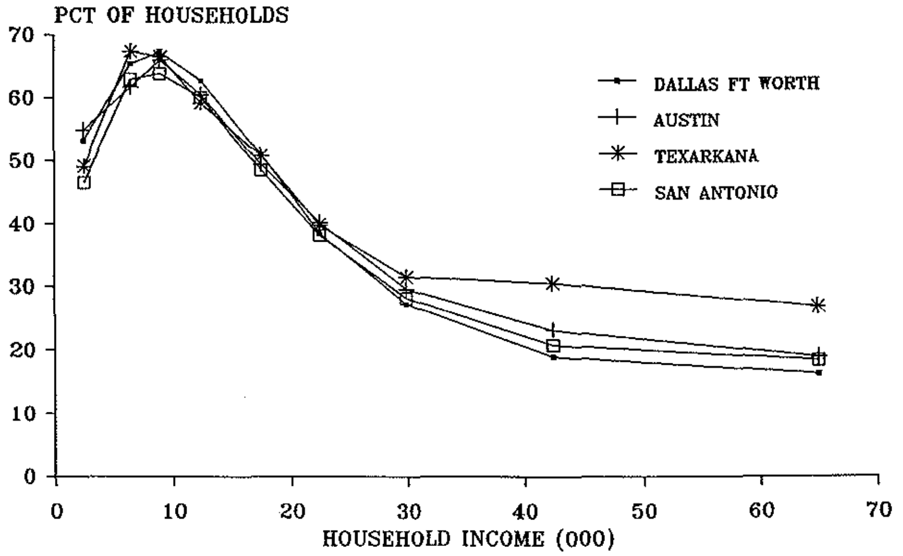
1980 UTPP DATA
Texarkana MSA

Figure 5
Distribution of Autos Available
By Household Income – 0 Vehicles



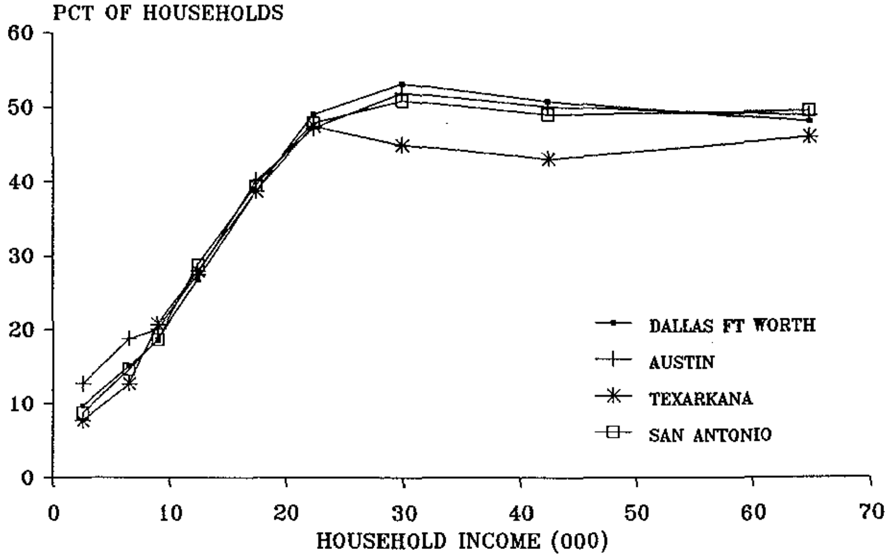
1980 UTPP DATA
Zero Vehicle Households

Figure 6
Distribution of Autos Available
By Household Income – 1 Vehicle



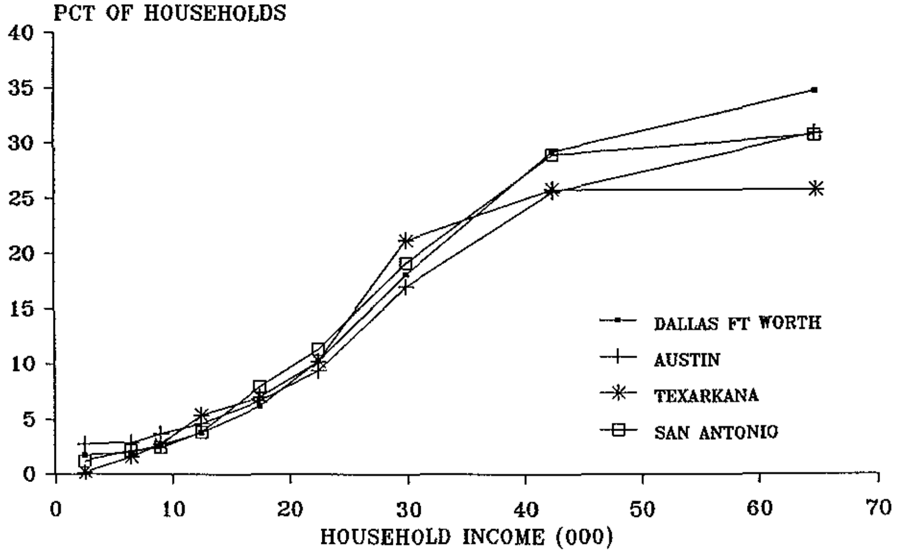
1980 UTPP DATA
One Vehicle Households

Figure 7
Distribution of Autos Available
By Household Income – 2 Vehicles



1980 UTPP DATA
Two Vehicle Households

Figure 8
Distribution of Autos Available
By Household Income – 3+ Vehicles



1980 UTPP DATA
Three Vehicle Households

Data points were measured from the hand fitted curves to create a table of values which reflect the estimated distribution of households by auto ownership at thousand dollar increments. The results are shown in Table 1.

A determination of how well the values in Table 1 replicated the observed values for the four MSAs was considered next. This was tested by comparing the appropriate values from Table 1 with the observed values for each MSA and computing the correlation coefficient. The resulting values are shown in Table 2 and indicate a relatively good estimation using the values from Table 1.

The data presented in Table 1 could be used to estimate the distribution of households by auto ownership for any MSA in Texas. It is realized that many urban areas have access to more detailed local census data that will allow them to develop the same relationship for the area. The procedure for developing a similar table for use in a specific area is the same as used for Table 1. The following steps are required:

1. Obtain data from the census which contain the number of households by income and auto ownership. This was available in the 1980 Urban Transportation Planning Package and is anticipated to be available in the 1990 Package.

2. Plot the percentage of households versus income for each level of auto ownership. Usually this will result in four separate plots: 0-auto households, 1-auto households, 2-auto households, and 3+-auto households.

3. Hand fit curves to the data points using the above plots. Measurements may be taken directly from these hand fitted curves at desired intervals (suggest thousand dollar increments) to build a table of values similar to Table 1.

4. Adjust the percentages at each income interval to ensure that the total is 100 percent. Thus, for a given value of income, the table should give an estimate of the percentage of households with 0, 1, 2, and 3+ autos.

The resulting table can then be input to TRIPCAL5 and used to estimate the distribution of households by auto ownership at the zone level based on the zone's mean income. Note that the assumption in this procedure is that zonal distributions will be similar to the distribution for the region.

Table 1
Percentage of Households by Auto Ownership and Income

Income 1980 $	0-Autos	1-Auto	2-Autos	3+-Autos
1000	45.18	45.90	7.65	1.27
2000	40.92	48.60	9.00	1.48
3000	35.25	52.58	10.32	1.85
4000	30.60	55.48	12.01	1.91
5000	26.26	58.03	13.71	2.00
6000	21.31	61.43	15.05	2.21
7000	18.53	62.99	16.06	2.42
8000	14.91	64.59	18.02	2.48
9000	12.50	64.90	20.00	2.60
10000	11.18	64.35	21.22	3.25
11000	9.34	62.64	24.29	3.73
12000	7.74	61.56	26.38	4.32
13000	7.29	59.35	28.66	4.70
14000	6.28	57.16	31.39	5.17
15000	5.63	55.05	33.68	5.64
16000	4.98	53.54	35.26	6.22
17000	4.51	50.77	37.86	6.86
18000	3.89	48.68	39.94	7.49
19000	3.74	46.83	41.19	8.24
20000	3.53	44.22	43.54	8.71
21000	3.10	42.50	45.00	9.40
22000	2.81	41.34	45.73	10.12
23000	2.61	38.93	47.15	11.31
24000	2.50	37.54	47.55	12.41
25000	2.45	36.28	48.14	13.13
26000	2.32	34.83	48.52	14.33
27000	2.20	33.77	48.82	15.21
28000	2.10	32.57	49.05	16.28
29000	1.96	31.37	49.30	17.37
30000	1.85	30.67	49.36	18.12
31000	1.77	30.02	49.45	18.76
32000	1.66	29.54	49.33	19.47
33000	1.49	28.94	49.23	20.34
34000	1.34	28.25	49.07	21.34
35000 +	1.26	27.67	49.09	21.98

Table 2
Comparison of Observed Versus Estimated Percentage Of Households by Auto Ownership

Dallas-Fort Worth MSA

1980 Income Range	0- Autos Obs	0- Autos Est	1- Auto Obs	1- Auto Est	2- Autos Obs	2- Autos Est	3+- Autos Obs	3+ Autos Est
0 - < 5K	35.30	38.09	53.14	50.59	9.77	9.66	1.79	1.67
5K - < 8K	17.52	19.28	65.38	62.21	15.11	15.55	1.98	2.31
8K -< 10K	11.36	12.50	67.28	64.90	18.64	20.00	2.71	2.60
10K-< 15K	6.61	7.51	62.76	60.46	26.81	27.52	3.82	4.51
15K-< 20K	3.97	4.20	51.07	49.73	38.75	38.90	6.22	7.17
20K-< 25K	2.19	2.71	38.47	40.14	49.06	46.44	10.28	10.71
25K-< 35K	1.48	1.85	27.22	30.67	53.16	49.36	18.14	18.12
35K & Over	0.87	1.26	18.04	27.67	49.84	49.09	31.25	21.98

R-squared = 0.982

San Antonio MSA

1980 Income Range	0- Autos Obs	0- Autos Est	1- Auto Obs	1- Auto Est	2- Autos Obs	2- AutosE st	3+- Autos Obs	3+- Autos Est
0 - < 5K	43.41	38.09	46.52	50.59	8.83	9.66	1.23	1.67
5K - < 8K	20.18	19.28	62.99	62.21	14.68	15.55	2.15	2.31
8K -< 10K	13.27	12.50	65.99	64.90	18.64	20.00	2.44	2.60
10K-< 15K	7.19	7.51	60.09	60.46	28.84	27.52	3.88	4.51
15K-< 20K	4.04	4.20	48.48	49.73	39.53	38.90	7.95	7.17
20K-< 25K	2.57	2.71	38.19	40.14	47.91	46.44	11.33	10.71
25K-< 35K	1.74	1.85	28.15	30.67	50.91	49.36	19.15	18.12
35K & Over	1.15	1.26	20.07	27.67	49.22	49.09	29.57	21.98

R-squared = 0.986

Table 2 (continued)
Comparison of Observed Versus Estimated Percentages of Households by Auto Ownership

Austin MSA

1980 Income Range	0- Autos Obs	0- Autos Est	1- Auto Obs	1- Auto Est	2- Autos Obs	2- Autos Est	3+- Autos Obs	3+- Autos Est
0 - < 5K	29.63	38.09	54.88	50.59	12.70	9.66	2.80	1.67
5K - < 8K	16.62	19.28	61.71	62.21	18.80	15.55	2.87	2.31
8K - < 10K	10.28	12.50	65.99	64.90	20.06	20.00	3.67	2.60
10K - < 15K	7.26	7.51	60.57	60.46	27.51	27.52	4.65	4.51
15K - < 20K	3.80	4.20	49.37	49.73	40.14	38.90	6.69	7.17
20K - < 25K	3.26	2.71	40.18	40.14	47.18	46.44	9.38	10.71
25K - < 35K	1.45	1.85	29.61	30.67	51.91	49.36	17.03	18.12
35K & Over	1.09	1.26	21.83	27.67	49.72	49.09	27.36	21.98

R-squared = 0.985

Texarkana MSA

1980 Income Range	0- Autos Obs	0- Autos Est	1- Auto Obs	1- Auto Est	2- Autos Obs	2- Autos Est	3+- Autos Obs	3+- Autos Est
0 - < 5K	42.55	38.09	49.14	50.59	7.75	9.66	0.55	1.67
5K - < 8K	18.39	19.28	67.35	62.21	12.68	15.55	1.58	2.31
8K - < 10K	9.96	12.50	66.55	64.90	20.74	20.00	2.75	2.60
10K - < 15K	7.39	7.51	59.22	60.46	27.98	27.52	5.40	4.51
15K - < 20K	3.45	4.20	50.84	49.73	38.70	38.90	7.01	7.17
20K - < 25K	2.42	2.71	39.83	40.14	47.50	46.44	10.25	10.71
25K - < 35K	2.34	1.85	31.54	30.67	44.90	49.36	21.22	18.12
35K & Over	0.66	1.26	29.42	27.67	44.08	49.09	25.84	21.98

R-squared = 0.989

THEORETICAL METHODOLOGY

Developing a theoretical methodology has to address the problem of how to estimate households by auto ownership percentages at a small geographical level such as a census tract or traffic analysis zone. Though the use of a set of curves or relationships developed for a region might produce good results for the region, it would not necessarily produce good results at a zone level. Any methodology developed must also be predictable. A large number of variables might produce good results in terms of replicating observed data; the error introduced with the projection of each independent variable could result in less than reasonable projections.

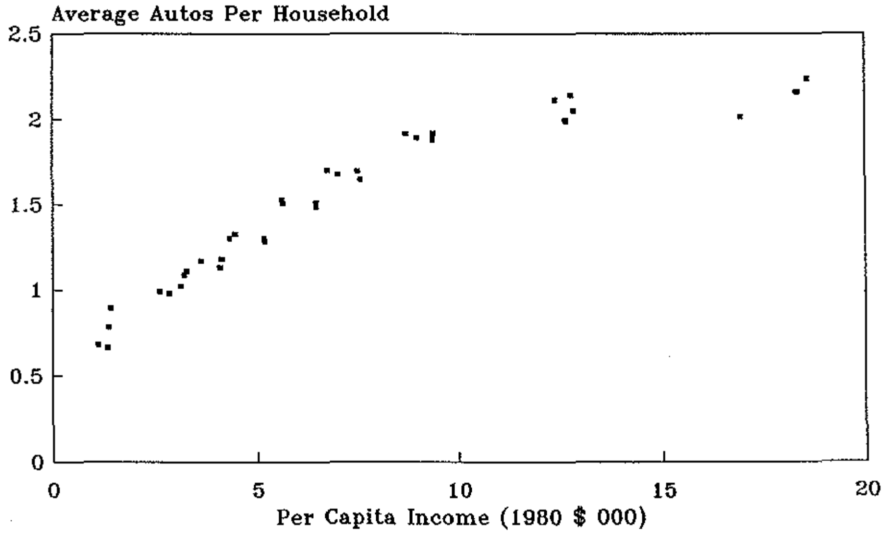
In reviewing the auto ownership relationships shown in Figures 1 through 4, there was no obvious mathematical formulation that would allow the percentage or number of households by auto ownership category to be calculated. A review of several research reports found that auto ownership models have been developed in other areas. A project done by the Harvard University Graduate School of Design, "Forecasting Auto Ownership and Mode Choice for U.S. Metropolitan Areas," developed a multi-variate model of auto ownership and mode choice using a sample of 346,000 households and 407,000 workers in the largest MSAs in the nation. That study found that auto ownership decisions depended on family type and composition, household income, residential location, workplace location, highway and transit service levels, and measures of overall urban spatial structure. The model actually predicted the probability of a household owning 0, 1, or 2+ autos given those variables. Use of these models was not considered practical due to the high number of input variables and the difficulty in forecasting them. Another research project done for the New York State Department of Transportation, "Predictive Accuracy of Aggregate and Disaggregate Auto Ownership Models," also developed several models for predicting autos per household. The primary variables used in those models were income, number employed in household, and residential density (population per residential land use in square feet). While these models appear to give fairly good results, they are not considered applicable due to the use of the number employed in the households as a variable. Both research projects indicate that income was strongly correlated with auto ownership, which led to the development and testing of alternative models using the 1980 census data for four MSAs in Texas.

First, a dependent variable was identified. The objective was a methodology by which the number of households that owned 0, 1, 2, and 3+ autos could be estimated for zones. Based on the review above and intuitive logic, it was felt that auto ownership could be related to income and/or household size. Since the data from the 1980 census included households by auto ownership by income and by household size, this was relatively easy to test. It was, however, still necessary to decide on what variable to estimate. The logical approach was to estimate the number of households in each auto ownership category based on income. This would result in four separate equations. Using the data for the four MSAs, a linear regression with total households as the dependent variable and income as the independent variable was performed for each auto ownership category, i.e., 0, 1, 2 and 3+ autos. This was accomplished with households disaggregated by income range as used in the census. The results were less than desirable. Correlation coefficient squared ranged from a low of 0.02 percent to a high of 0.21 percent. These were expected, since data plotted in Figures 1 through 4 had virtually no linear relationships.

Instead of attempting to estimate the number of households in each category independently (a procedure that would require four different models), the relationship between autos per household and independent variables such as household size and income were analyzed. The income ranges were the same for each urban area even though the areas were different. To account for the differences in income between the areas, a per capita income was computed for each income range. This was more realistic in terms of acting as a descriptive variable. Provided that a relationship could be established that would estimate autos per household, a subsequent model or method would be needed to use that estimate (with possibly other information) to estimate the number of households in each income category. This could be done by computing the total number of autos (e.g., in a zone) and then developing a routine to distribute households within the four categories of auto ownership, ensuring that the total number of autos and average auto ownership were accurately reflected by the distribution.

Using data for the four MSAs, the average autos per household were computed for each income range for which households were distributed (nine ranges with varying interval sizes). The per capita income and average household size were also computed for each income range. These values are shown in Table 3. Initially, a linear regression was performed with autos per household as the dependent variable and per capita income as the independent variable. The resulting fit was good with nearly 85 percent of the variation in autos per household being explained (as measured by the correlation coefficient squared). A plot of this fit, shown in Figure 9, reveals that the observed data points are more exponential than linear in form. A linear regression was performed again using the natural log of per capita income as the independent variable; a significant improvement was achieved with nearly 95 percent of the variation being explained (as measured by the correlation coefficient squared). A third linear regression was done, again using autos per household as the dependent variable, with average household size and per capital income as the independent variables. It was expected that household size would be significant as an explanatory variable. The resulting value of the correlation coefficient squared of 0.946 was a significant improvement over the use of just per capita income. A linear regression was next performed with both the average household size and the natural log of per capita income as independent variables. The combination of these two variables explained an estimated 96.5 percent of the variation in autos per household (as measured by the correlation

Figure 9
Autos Per Household Versus
Per Capita Income



Source: 1980 Urban Transportation Planning Package – Four MSA's in Texas coefficient squared). The use of that linear relationship is shown in Figure 10. The following equations were the result of the above regression analyses:

$$\text{APH} = 0.59614*(\text{Natural Log of Per Capita Income}) - 3.65027 \qquad \text{Eq. 1}$$

$$\text{R-square} = 0.9494$$

$$\text{APH} = 0.43596*(\text{Avg HH Size}) + 0.00005*(\text{Per Capita Income}) - 0.07659 \qquad \text{Eq. 2}$$

$$\text{R-square} = 0.9459$$

$$\text{APH} = 0.22447*(\text{Avg HH Size}) + 0.45939*(\text{Nat. Log of Per Capita Income}) - 3.08778 \qquad \text{Eq. 3}$$

$$\text{R-square} = 0.9645$$

where: APH = autos per household
HH = households

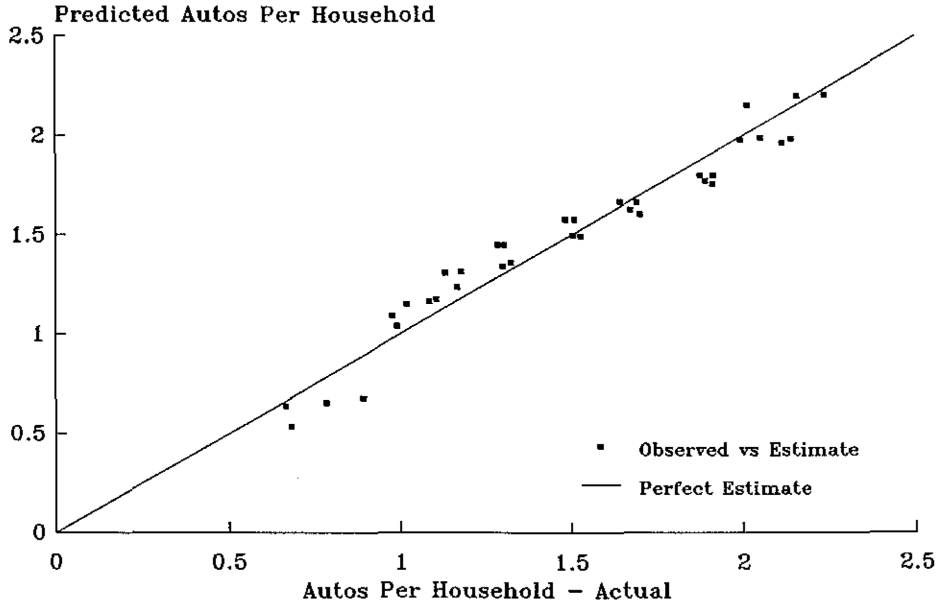
Any of the three relationships developed would produce reasonable results based on the correlations achieved. To select one to use in subsequent analyses, it was necessary to determine which one of the three replicated the actual number of autos for each of the four MSAs. This was somewhat of an aggregate measure, but it was reasonable since this would be estimated at the zone level in the application of the model. The result of this calculation is shown in Table 4. As will be noted, while the errors are all within acceptable ranges, the second and third equations were superior to the first. The second equation was attractive because the errors were similar for each area and were relatively small. The third was attractive simply because the accuracy was exceptionally good for three of the four areas, and the error for the fourth was small. The third equation was used because of its accuracy and its inclusion of household size with per capita income. The use of household size with per capita income is more explanative in predicting autos per household.

The next step was to develop a means by which the average autos per household could be used to estimate the number of households within each auto ownership category.

Table 3
1980 Auto Ownership Data

MSA	1980 Income Range	No. Households	Avg HH Size	Per Cap Income	Autos per Household
Texarkana	0 - < 5K	9015	1.873	1334.8	0.664
Texarkana	5K - < 8K	5078	2.289	2840.0	0.978
Texarkana	8K - < 10K	3202	2.489	3616.0	1.168
Texarkana	10K - < 15K	7587	2.808	4451.7	1.325
Texarkana	15K - < 20K	6520	3.103	5640.5	1.507
Texarkana	20K - < 25K	5250	3.208	7013.4	1.676
Texarkana	25K - < 35K	5523	3.339	8985.9	1.893
Texarkana	35K - < 50K	2363	3.354	12672.2	1.996
Texarkana	50K & Above	1124	3.539	16953.2	2.017
Dallas-Ft. Worth	0 - < 5K	110971	1.83	1366.5	0.784
Dallas-Ft. Worth	5K - < 8K	82605	2.079	3126.1	1.02
Dallas-Ft. Worth	8K - < 10K	62877	2.201	4089.5	1.132
Dallas-Ft. Worth	10K - < 15K	161225	2.405	5198.3	1.286
Dallas-Ft. Worth	15K - < 20K	154438	2.710	6456.8	1.485
Dallas-Ft. Worth	20K - < 25K	140365	2.997	7506.7	1.695
Dallas-Ft. Worth	25K - < 35K	189267	3.195	9389.0	1.916
Dallas-Ft. Worth	35K - < 50K	110837	3.321	12798.9	2.143
Dallas-Ft. Worth	50K & Above	63920	3.231	18570.6	2.237
San Antonio	0 - < 5K	52019	2.234	1119.3	0.681
San Antonio	5K - < 8K	36273	2.495	2605.6	0.992
San Antonio	8K - < 10K	25633	2.752	3270.2	1.107
San Antonio	10K - < 15K	59141	2.896	4316.3	1.302
San Antonio	15K - < 20K	51969	3.119	5610.8	1.530
San Antonio	20K - < 25K	41190	3.327	6763.7	1.703
San Antonio	25K - < 35K	47742	3.446	8706.2	1.914
San Antonio	35K - < 50K	23762	3.425	12408.1	2.116
San Antonio	50K & Above	11973	3.278	18304.5	2.160
Austin	0 - < 5K	27769	1.757	1423.2	0.892
Austin	5K - < 8K	19557	2.030	3202.3	1.085
Austin	8K - < 10K	13894	2.183	4122.3	1.179
Austin	10K - < 15K	31249	2.412	5181.9	1.305
Austin	15K - < 20K	26425	2.708	6462.9	1.511
Austin	20K - < 25K	22928	2.977	7557.8	1.646
Austin	25K - < 35K	29284	3.199	9376.7	1.879
Austin	35K - < 50K	16782	3.303	12867.1	2.054
Austin	50K & Above	8336	3.270	18348.1	2.158

Figure 10
Estimated Average Autos/Household
Vs. Actual Average Autos/Household



Using the average autos per household (calculated based on average household size and per capita income), an estimate of the total autos for a zone could be computed. Both the average autos per household and the estimation would serve as controls for the subsequent distribution of households by auto ownership. Using the mean income for a zone, an initial distribution of households by auto ownership could be obtained using Table 1. This distribution could then be adjusted to give the computational correct average autos per household (i.e., the estimate derived based on household size and per capita income). The initial distribution would be considered a best estimate (initially) and should define the relative relationship between the ownership categories. The assumption was made that the relationship in the initial distribution of the number of households in the categories on each side of the average would be held constant. This can best be illustrated by an example. If the initial distribution resulted in 20 0-auto households, 41 1-auto households, 40 2-auto households, and 15 3+-auto households and the desired average autos per household was 1.5, then to achieve the desired average it would be necessary to move households from the 0-auto and 1-auto categories to the other two categories. The assumption that relationships between categories in the initial distribution will remain constant means that in the final distribution, 30.76 percent of the households in categories 0-auto and 1-auto will fall in the 0-auto category; and 69.24 percent will fall in the 1-auto category (i.e., the categories to each side of the mean will hold the same relative percentages in the final distribution as in the initial distribution). The adjustments necessary to result in the correct average and to satisfy the above assumption are relatively straightforward and are not included here.

To test the above methodology, 1984 data for 546 zones in the Dallas-Fort Worth study were used. While the results are estimates for 1984, they were compared to the MSA data for 1980 to determine their "reasonableness." The overall average autos per household from the model was 1.67, an 8.3 percent increase from the 1980 observed average of 1.542 (based on the 1980 census). Total households for the same time period increased by 19.3 percent. The overall average of 1.67 was considered to be high, and the distribution of households by auto ownership was not considered realistic. For example, the 0-auto households had declined from 7.9 percent of the total to 3.25 percent. Upon further analysis, it was decided that the use of one value of the mean income for a zone resulted in a biased initial distribution of households by auto ownership. To alleviate this, the mean was used to estimate the distribution of households by income, to distribute the households within each income range (using the values from Table 1), and to accumulate the households within each category to develop a more reasonable estimate of the initial distribution of households by auto ownership. The result was a more realistic distribution in terms of 0-auto ownership, but the percentages in the 1- and 2-auto categories seemed out of proportion. For example, in 1980 the percentage of 1-auto households was 43.9. The estimate (using the model) for 1984 was 36.2 percent. Such a dramatic shift in four years was not considered realistic.

The projection of 1.67 autos per household in 1984 was too high relative to the observed value of 1.54 in 1980. It was reasoned that the average autos per household being projected for each zone used the average income for the zone, whereas the model had been calibrated using disaggregated data; i.e., autos per household for each MSA were related to household size and per capita income for each income range used in the census (there were nine ranges). In effect, the use of the zonal mean income for projecting the average autos per household was not consistent with the method used for calibrating the model. This was tested by using the model to compute the average autos per household for the Dallas-Fort Worth MSA with the MSA per capita income and average household size. The resulting estimate was 1.67 autos per household, while the actual observed value was 1.54 percent. It was then concluded that to achieve a realistic estimate, the households must be distributed by income at the zone level and an estimate of the autos per household computed for each income interval. This could be used to estimate the total autos for the households in that income interval which could be accumulated and used to compute an average for the zone (i.e., divide the total autos by the total households). This would result in a more accurate estimate of the average autos per household which, in the adjustment of the initial distribution of households by auto ownership, ultimately achieve a better estimate of the households by auto ownership.

The results are shown in Table 5. As will be noted, the estimate of the regional average autos per household using the revised methodology was 1.52, which is actually less than the observed value in 1980 of 1.54 percent. Table 5 also shows the resulting estimates of 1984 regional distribution of households by auto ownership as compared to 1980. These are considered realistic and reasonable. Subsequently, the model was selected for use as a default model for estimating the number of households by auto ownership at the zone level. Note that it is assumed that the categories being used are 0, 1, 2, and 3+ autos.

Table 4
Estimates of Total Autos Owned Versus Observed Values from 1980 Census

MSA	Total Autos Obs	Eq 1 Estimate	% Error	Eq 2 Estimate	% Error	Eq 3 Estimate	% Error
Texarkana	60,809	63,077	3.73	62,027	2.00	60,797	-0.02
Dallas/Ft. Worth	1,660,178	1,580,434	-4.80	1,632,669	-1.66	1,664,330	0.25
San Antonio	493,966	521,350	5.54	499,456	1.11	475,806	-3.68
Austin	288,301	271,797	-5.72	282,169	-2.13	289,695	0.48

Table 5
Comparison of 1980 Observed Data to 1984 Model Predictions for Dallas-Ft. Worth

Auto Ownership Category	1980 Households	%	Modeled 1984 Households	%
0-Autos	84969	7.9	101613	7.9
1-Auto	472407	43.9	575932	44.8
2-Autos	394666	36.7	468127	36.4
3+-Autos	124463	11.6	139077	10.8

HOUSEHOLD SIZE DISAGGREGATION MODEL

INTRODUCTION

The number of trips made by household members is highly correlated with the household size. With more members in a household, more trips are made on a daily basis. This was the primary reason it was recommended that the trip production model for TRIPCAL5 include household size as one of the independent variables in the cross classification model. The use of this variable necessitates the ability to project the number of households by household size and the ability to disaggregate the number of households within a zone into the number by household size. In the development of TRIPCAL5, work was undertaken to develop both a method for projecting households by household size and a method for disaggregating the households within a zone into the number of households by household size. The following sections describe the development of the methodology and the necessary assumptions which are a part of the methodology.

PROBLEM STATEMENT

In the majority of urban transportation studies, estimates of households and population are provided by the local planning agency (i.e., the Metropolitan Planning Organization) for both the entire urban area and each zone within the area. The estimation of the persons per household is a straightforward calculation. The difficult item to obtain is the number of households by household size (in many areas this is available only at the regional level). While difficult, it is not impossible because methods have been developed to compute these estimates. The recommendation for TRIPCAL5 was that a cross-classification model using income and household size as independent variables be used to estimate trip productions. It was, thus, necessary to develop and include a means by which reasonable estimates of households by household size could be obtained at the zone level.

PROPOSED ALTERNATIVES

There are three proposed alternatives to obtaining estimates of households by household size at the zone level. The first is to obtain the estimates for each zone from the local area planning agency. This is considered to be a viable option, and TRIPCAL5 has been developed to allow these estimates to be input with other zonal data.

The second alternative is to use the same technique as applied in other major urban areas. That technique consists of developing a set of curves which relate the distribution of households by household size to the average household size of a zone. These curves would be based on census data. The areas where this technique was used found that a strong correlation existed between the distribution of households by household size and the average household size using curves that were hand fitted to census data. Comparisons of estimated distributions (based on hand fitted curves) to observed values in Houston yielded correlation coefficients above 0.9.

The third alternative is to develop a theoretical method of estimating the distribution of households by household size. This would provide a default option which could be used in the event the others were not available.

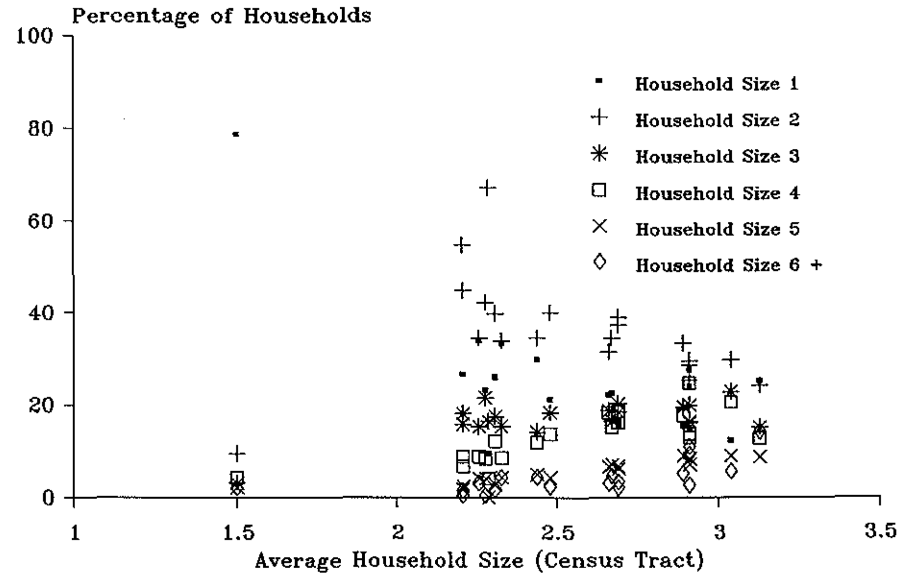
The following sections describe the methodology for the second and third proposed alternatives. The first alternative, while feasible, would be at the discretion of the local area planning agency which would develop the methodology for the area.

DISAGGREGATION CURVE METHODOLOGY

The disaggregation curve methodology is based on the assumption that a correlation exits between the distribution of households by household size and the average household size of a given area. The data analyzed in this methodology are at the census tract level. Provided a reasonable correlation can be found, it is assumed that the relationships will hold constant for the zones within the area and for projections into the future.

The first step in developing this methodology is to plot the relationship between average household size and the distribution of households by household size. Figure 11 shows this for the Bryan-College Station MSA. This is a small MSA with only 20 census tracts, resulting in fewer data points than would be available in the larger areas. This was considered to be a good case study in that if adequate results were possible for a small area, reasonable results could be expected for the larger areas. It was also known that this method was successfully developed and applied in both the Dallas-Fort Worth and Houston-Galveston areas. It was not known, however, whether it could be applied in the smaller areas.

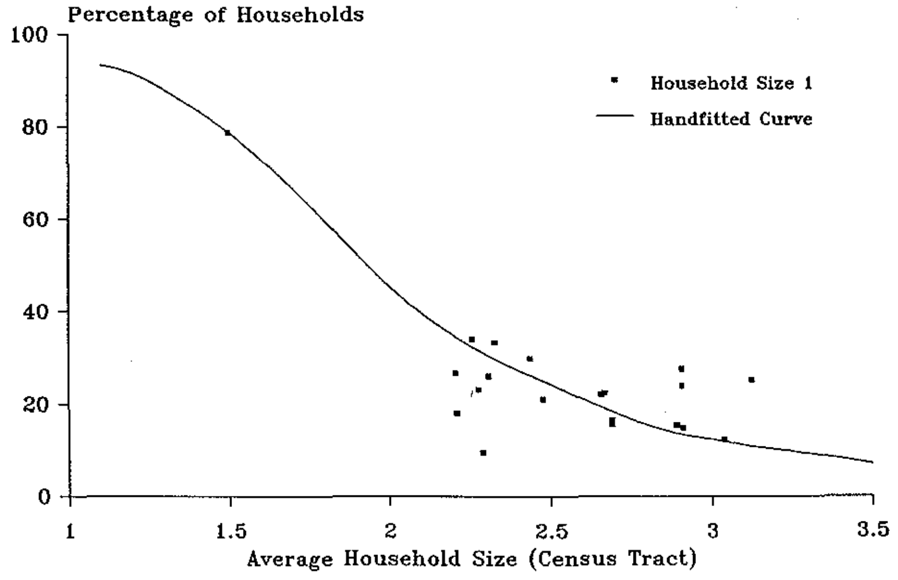
Figure 11
Distribution of Households by Household
Size vs. Average Household Size



Bryan–College Station 1980 Census Data

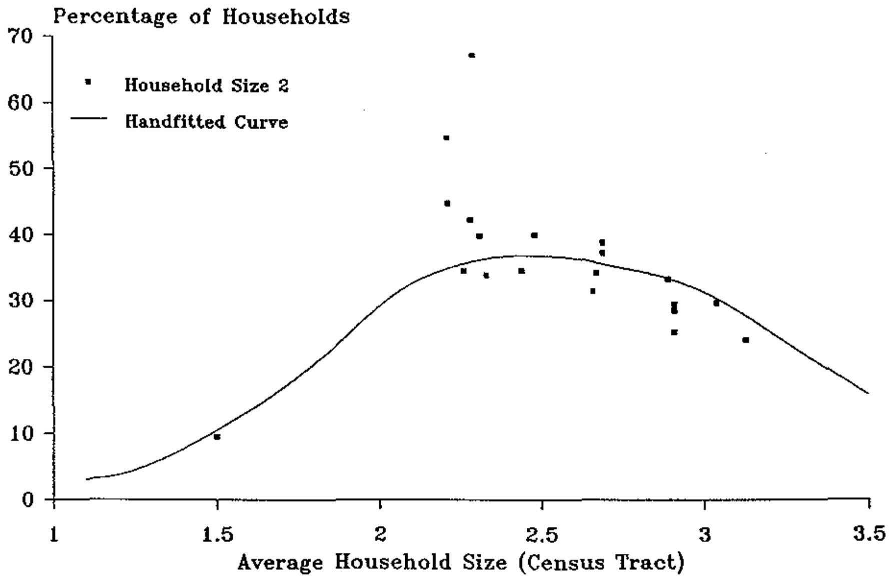
In reviewing Figure 11, it is difficult to ascertain any relationships in the format in which the data are plotted. The next step is to simplify the plot by developing individual plots for each household size. This is shown in Figures 12 through 17 for Bryan-College Station. These become the base data for developing curves for input to the trip generation model and for use in disaggregating households at the zone level by household size. These curves are generally hand fitted although, where appropriate, the analyst can use regression techniques or other mathematical relationships to develop the curves. In the case of Bryan-College Station, the curves were hand fitted and are also shown in Figures 12 to 17. The curve values were measured from the plots and are shown in Table 6 for household sizes ranging from 1.1 to 3.5. These values covered the range of possible values. To determine how well the curves would replicate the actual distributions, distributions of households by household size were computed for each census tract using the average household size of the tract and were compared with the actual census distribution by computing a correlation coefficient and "Z" statistic. The correlation coefficient value was 0.9117 with a "Z" statistic value of 16.6. This indicated that the disaggregation curves produced reasonably good overall results.

Figure 12
Distribution of One Person Households
vs. Average Household Size



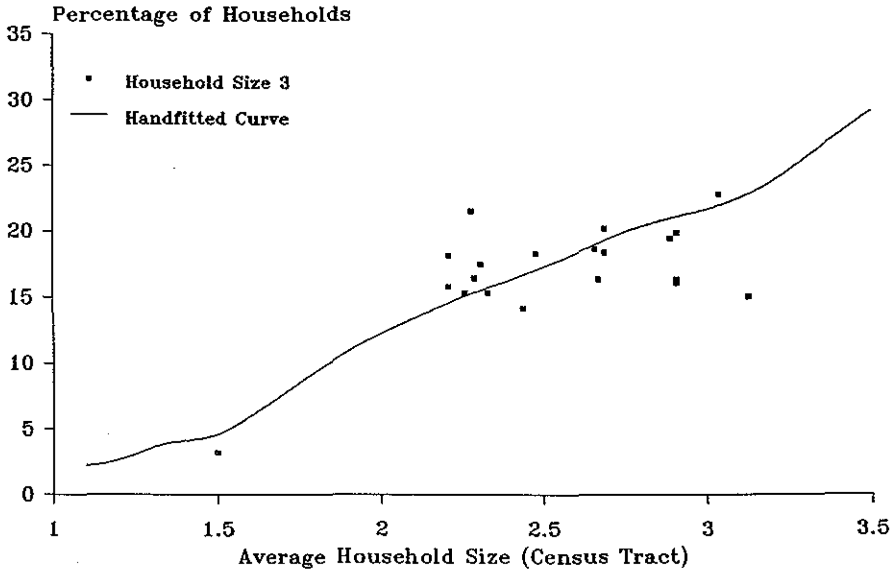
Bryan-College Station 1980 Census Data

Figure 13
Distribution of Two Person Households
vs. Average Household Size



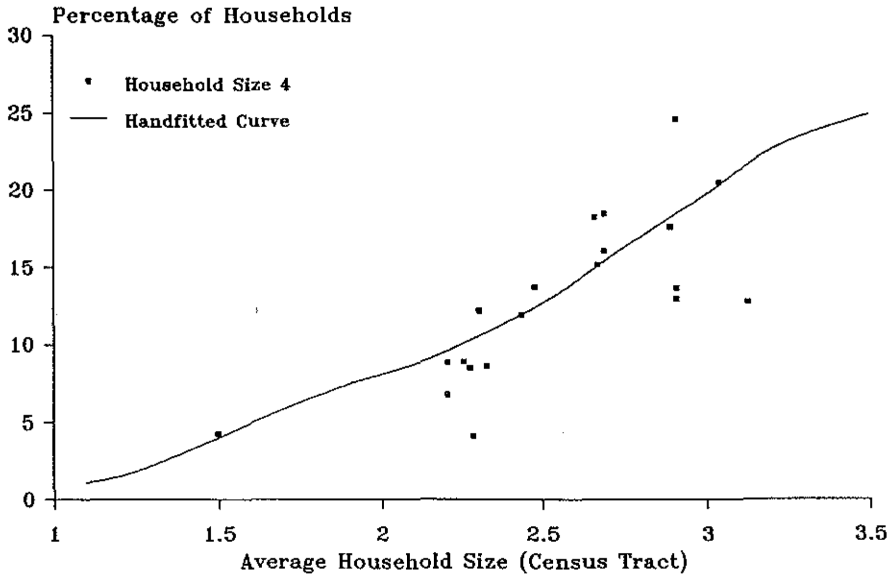
Bryan-College Station 1980 Census Data

Figure 14
Distribution of Three Person Households
vs. Average Household Size



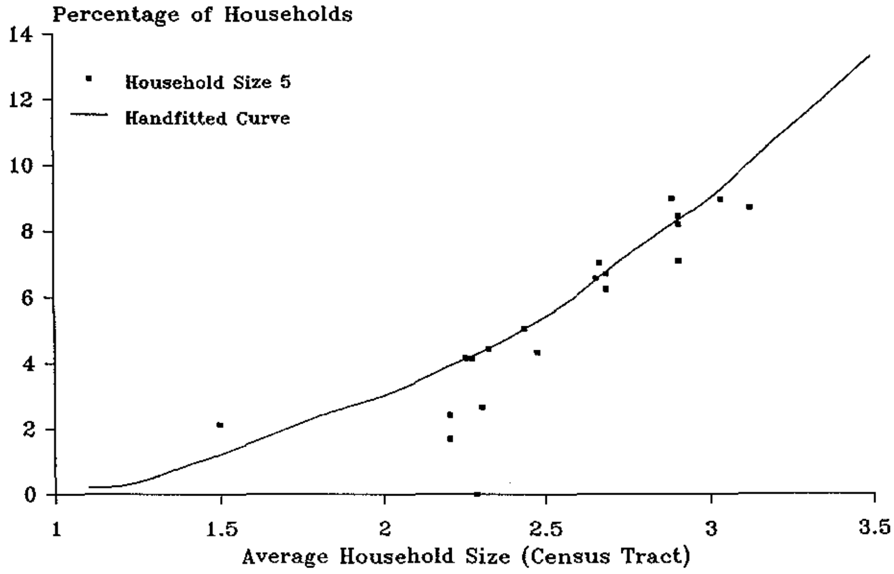
Bryan-College Station 1980 Census Data

Figure 15
Distribution of Four Person Households
vs. Average Household Size



Bryan-College Station 1980 Census Data

Figure 16
Distribution of Five Person Households
vs. Average Household Size



Bryan-College Station 1980 Census Data

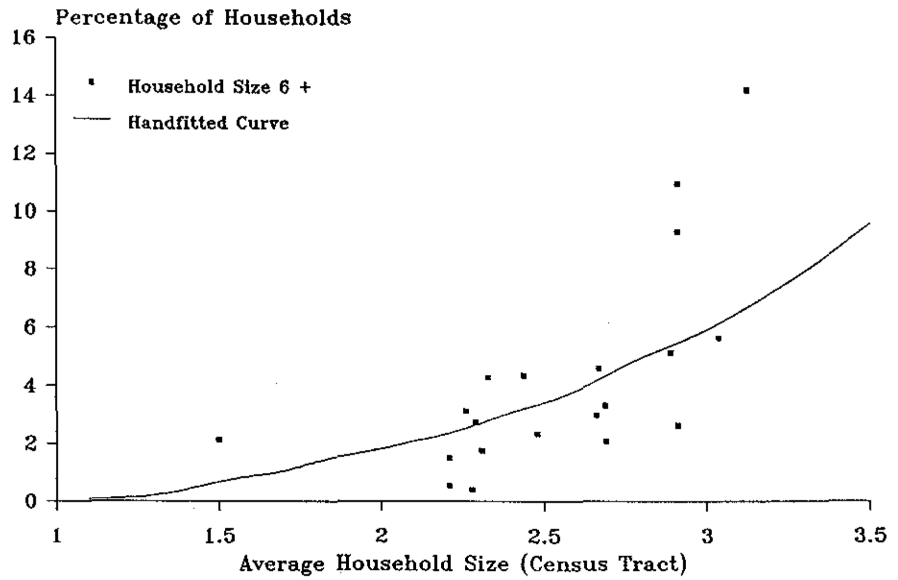


Figure 17
Distribution of Six + Person Households
vs. Average Household Size

Bryan-College Station 1980 Census Data

The operation of this methodology is fairly straightforward. Given the average household size of a zone, the estimated percentage of households by household size can be found by looking up the percentage shown in Table 6. For example, if a zone had an average household size of 2.4, use of the data in Table 6 would result in estimating that 27.3 of the zone's households were size 1, 37.0 percent were size 2, 16.3 percent were size 3, 11.5 percent were size 4, 4.8 percent were size 5, and 3.1 percent were size 6 or more.

THEORETICAL METHODOLOGY

The theoretical methodology is based on the distribution of households by household size as observed from census data. Observations from the 1980 census for MSAs in Texas indicate that these distributions are similar to those found for income and trip length frequency. The distributions were non-negative, had a minimum observed value, skewed to the left, and tailed off to the right. It was concluded that a similar methodology as that used in projecting income distributions was potentially applicable to predicting household size distributions. The first step in analyzing this possibility was to determine if the Gamma distribution would replicate the observed distributions from the 1980 census.

Households by household size data were available from the 1980 census for 26 MSAs in Texas. A program was written to input the distribution of households by household size and to fit a Gamma distribution to the data for each MSA using the method of moments. The results were good for most MSAs. The worst fits were for urban areas in the Rio Grande Valley, Brownsville, McAllen-Pharr, and Laredo. These areas were characterized as having the highest average household size in the state. With the exception of those three MSAs, the curve fits were reasonably good with correlation values ranging from 0.81 to 0.97. Twenty of the 26 MSAs had correlation coefficients of greater than 0.9 with most of those around 0.95. It should be noted that in the curve fit, the household size categories were converted to nondimensional values by dividing by the average household size for the MSA.

Table 6
Household Size Disaggregation Curve Data, Bryan-College Station

Percentage Households

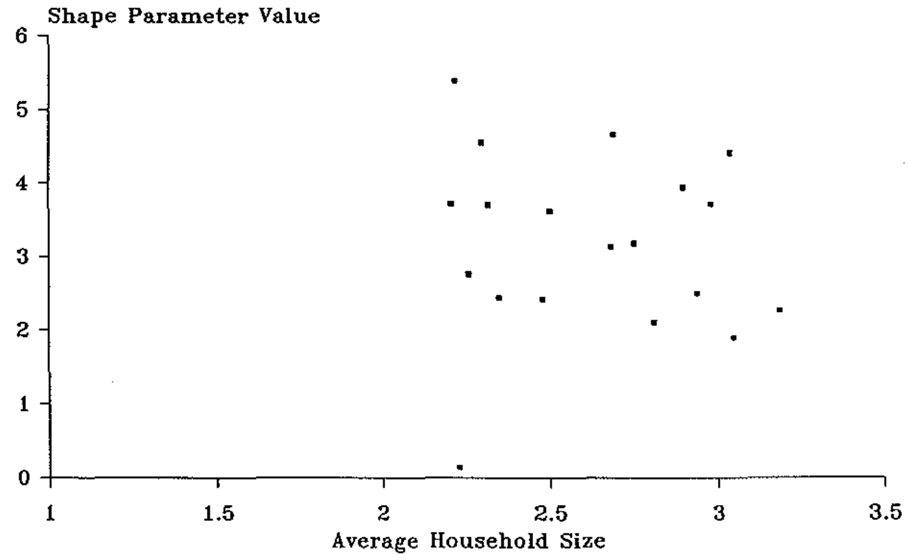
Avg HH Size	HH Size 1	HH Size 2	HH Size 3	HH Size 4	HH Size 5	HH Size 6+
1.1	93.3	3.1	2.2	1.1	0.2	0.1
1.2	92.2	3.5	2.4	1.4	0.2	0.1
1.3	87.9	5.4	3.8	2.2	0.5	0.2
1.4	83.9	7.6	4.1	3.1	0.9	0.4
1.5	79.0	10.7	4.4	4.0	1.2	0.7
1.6	73.1	13.4	6.0	5.0	1.6	0.9
1.7	66.7	16.7	7.7	5.9	2.0	1.0
1.8	59.8	20.3	9.4	6.7	2.4	1.4
1.9	52.6	24.6	11.0	7.5	2.7	1.6
2.0	45.4	29.4	12.3	8.1	3.0	1.8
2.1	39.6	32.8	13.4	8.7	3.4	2.1
2.2	34.9	34.9	14.5	9.5	3.9	2.3
2.3	30.8	36.2	15.5	10.5	4.3	2.7
2.4	27.3	37.0	16.3	11.5	4.8	3.1
2.5	24.5	36.8	17.3	12.6	5.4	3.4
2.6	21.2	36.7	18.3	14.0	6.0	3.8
2.7	18.1	35.5	19.5	15.6	6.9	4.4
2.8	15.4	34.6	20.4	17.0	7.6	5.0
2.9	13.4	33.4	21.1	18.4	8.3	5.4
3.0	12.5	31.5	21.6	19.7	8.9	5.9
3.1	11.1	28.8	22.5	21.3	9.8	6.5
3.2	10.3	25.4	23.7	22.7	10.7	7.2
3.3	9.3	22.1	25.6	23.6	11.5	7.9
3.4	8.2	19.0	27.4	24.3	12.4	8.7
3.5	7.1	15.9	29.2	24.9	13.3	9.6

To use the Gamma relationship for estimating distributions of households by household size, it is necessary to be able to estimate the shape parameter, alpha, for the following equation:

$$f(t) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$$

The household size categories were converted to nondimensional values by dividing by the average household size; therefore, beta, the value of the scale parameter, becomes equivalent to the shape parameter. Tests of the curve fit methodology found this to be true. A reasonable approach was to determine if a relationship existed between the value of the shape parameter and the average household size. Figure 18 presents a plot of this relationship which indicates no correlation. An unsuccessful attempt was also made to determine a relationship with population.

Figure 18
Average Household Size vs.
Gamma Distribution Shape Parameter



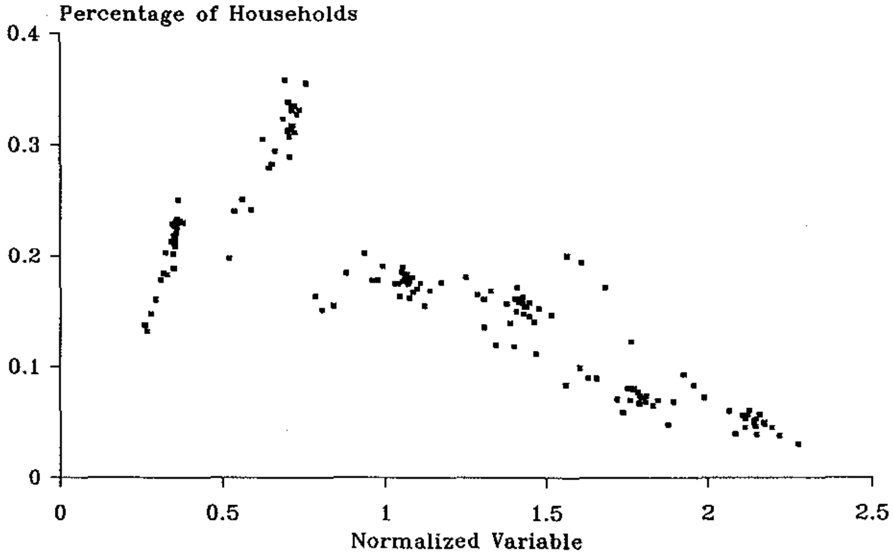
Bryan-College Station 1980
Census Data

The next attempt was to develop a generalized relationship for all the MSAs which potentially could produce reasonable results for individual MSAs. This type of generalized relationship is developed by combining all known data points into a nondimensional format, fitting a Gamma distribution to it and adjusting it to obtain the "best" fit.

Because each area is unique, the only distinguishing piece of information known now and in future applications is the average household size. Average household size was used to convert the household size categories to nondimensional values in the distributions by dividing each household size category by it and converting the number of households in each category to percentages. Figure 19 presents a plot of the resulting data for the 26 MSAs in Texas. While there is definite clustering of data, there also appears an overall shape similar to that of a Gamma distribution. A curve fit was performed and the resulting shape parameter was adjusted until the "best" fit was found, as indicated by the value of the correlation coefficient. The resulting overall correlation coefficient was 0.933. The curve fit estimates and observed data from the census are shown in Figure 20. It was interesting that the correlation coefficient for each individual MSA showed substantial improvement over the results from the individual curve fits. This implies that the curve fit methodology used was inappropriate, probably because of the low number of observations (i.e., only six for each MSA) input to the curve fit routine. Table 7 presents the results for both the individual curve fits and the overall aggregate fit routine. As indicated, the overall aggregate curve fit routine was superior in results; and the decision was made to use it for further analysis in disaggregating households at the subregional level.

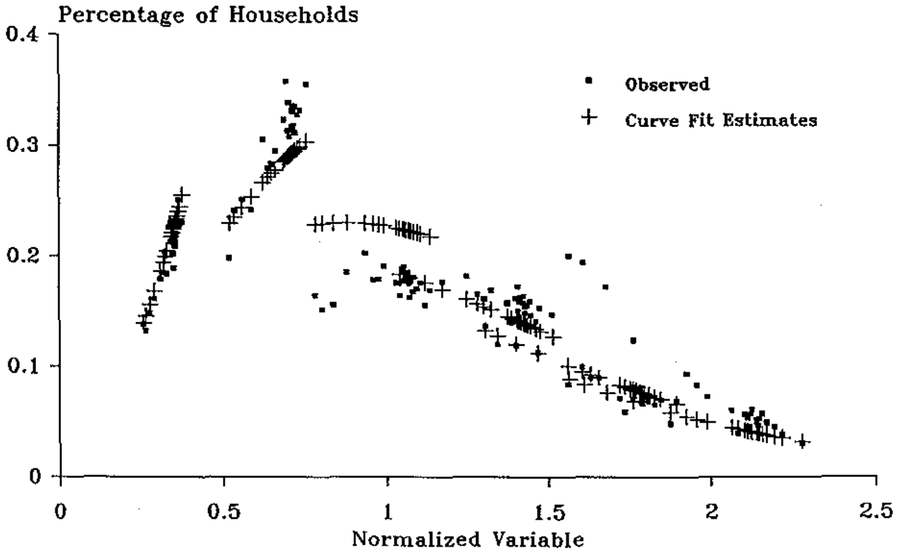
As before, the testing of a disaggregation model was accomplished using data for the Bryan-College Station MSA from the 1980 census. If the model could replicate the distribution of households by household size at the census tract level with a reasonable degree of accuracy, it would also be applicable to traffic serial zones. Since the results from estimating the MSA distributions were considered excellent, the model was applied directly to the census tracts for the Bryan-College Station MSA.

Figure 19
Observed Distributions of Households by
Normalized Household Size



Observed Values From 1980 Census
26 MSA's in Texas

Figure 20
Distributions of Households by Household
Size by Normalized Household Size



Observed Values From 1980 Census
26 MSA's in Texas

Table 7
Household Size Distributions
Curve Fitting Results

MSA	Avg HH Size	Individual Curve Fits		Combined Curve Fit	
		Alpha	Cor Coef	Alpha	Cor Coef
Abilene	2.84	2.022	0.914	2.76	0.989
Amarillo	2.71	2.654	0.961	2.76	0.992
Austin	2.74	2.097	0.951	2.76	0.991
Beaumont-Port Arthur	2.85	3.062	0.960	2.76	0.992
Brownsville-Harlingen	3.57	3.048	0.428	2.76	0.957
Bryan-College Station	2.88	1.266	0.834	2.76	0.983
Corpus Christi	3.12	2.954	0.915	2.76	0.989
Dallas-Ft. Worth	2.76	2.910	0.968	2.76	0.994
El Paso	3.40	2.937	0.870	2.76	0.986
Galveston	2.83	3.066	0.964	2.76	0.993
Houston	2.82	3.035	0.960	2.76	0.993
Killeen-Temple	3.20	1.473	0.814	2.76	0.993
Laredo	3.83	3.253	-0.007	2.76	0.948
Longview	2.81	2.807	0.959	2.76	0.992
Lubbock	2.91	2.268	0.923	2.76	0.989
McAllen-Pharr-Edinburg	3.72	3.158	0.206	2.76	0.946
Midland	2.80	3.076	0.963	2.76	0.992
Odessa	2.84	3.546	0.959	2.76	0.990
San Angelo	2.80	2.131	0.936	2.76	0.991
San Antonio	3.07	2.569	0.934	2.76	0.991
Sherman-Denison	2.64	2.765	0.951	2.76	0.987
Texarkana	2.78	2.967	0.957	2.76	0.992
Tyler	2.79	2.824	0.947	2.76	0.989
Victoria	3.02	2.965	0.959	2.76	0.993
Waco	2.77	2.353	0.939	2.76	0.989
Wichita Falls	2.79	2.007	0.925	2.76	0.990

Applying the Gamma distribution with a constant value for the shape parameter of 2.76 yielded fairly good results (the value of beta is set equal to the value of the shape parameter). The resulting distributions of households by household size were adjusted by varying the scale parameter, beta, to ensure the average household size for each census tract was the same as the input value. There are 20 census tracts in the Bryan-College Station area. The correlation coefficient computed from comparing the observed versus estimated households by household size was used as the basis for determining how well the model matched the actual census data. This was also compared with the correlation coefficients from the disaggregation curves described earlier. Table 8 shows that the disaggregation curves generally produced better results, but the results from the use of the Gamma distribution was felt to be adequate as a default model for use when other data were not available.

The application of this methodology is not elaborate. An average household size is computed using the estimated population and number of households for a zone. The household size categories are then converted to nondimensional values and input to the Gamma distribution equation with the shape and scale parameter values set equal to 2.76. The resulting values are converted to percentages, and these are used to compute an average household size based on the distribution from the Gamma equation. This average is compared with the originally input average (i.e., computed from the population and households for the zone). If not within ±1 percent, the scale parameter, beta, is adjusted by multiplying by the ratio of the distribution average to the desired average. A new distribution is computed using the shape parameter value of 2.76 and the adjusted scale parameter. This process is repeated until the distribution results in approximately the same average household size as computed from the zone's population and number of households.

This methodology has the advantage that it can be used by a local area planning agency to develop initial distributions of households by household size for each zone. These distributions can be adjusted and input directly to TRIPCAL5. If desired, local area planning agencies can use current information to estimate revised values for the shape and/or scale parameters to improve the accuracy of the zonal estimates. The conclusion is that the methodology gives reasonable results which can be modified to incorporate local knowledge and improve the final estimates.

Table 8
Results of Using Gamma Distribution to Estimate Distributions of Households by Household Size at Census Tract Level
Bryan-College Station

Census Tract	Gamma Distribution Correlation Coef.	Disaggregation Curve Correlation Coef.
1	0.929	0.998
2	0.956	0.990
3	0.942	0.994
4	0.721	0.593
5	0.478	0.333
6.01	0.933	0.999
6.02	0.864	0.787
7	0.977	0.991
8	0.898	0.980
9	0.987	0.999
10	0.924	0.938
11	0.960	0.971
13	0.954	0.978
14	0.803	0.813
15	0.689	0.690
16	0.929	0.969
17	0.927	0.895
18	0.900	0.962
19	0.877	0.936
20	0.943	0.994

FORECASTING INCOME FOR URBANIZED AREAS

INTRODUCTION

Estimating income for households, families, and/or individuals (i.e., per capita) has continued to pose a difficult problem for most urban areas. Income is widely used as a variable in models predicting travel demand. A recommendation was made that the cross-classification trip production model include household income as one of the independent variables. As part of upgrading the travel demand models for the Texas Department of Transportation (TxDOT) it was necessary to develop a method(s) by which reasonable income forecasts could be accomplished.

Prediction of income within an urban area is difficult because of the cyclical nature of urban economies, influences of national and international economies, inflation, and the lack of a comprehensive historical data base. While problematic, the task is not impossible when viewed in terms of developing a "reasonable" forecast, i.e., one that, while not exact, may be considered relatively accurate as a measure of the most likely change in terms of magnitude.

The number of trips made by household members has been found to be highly correlated with income. For this reason, it was recommended as one of the independent variables in the trip production model for the new trip generation program TRIPCAL5. A methodology for projecting and disaggregating households by income was needed, both for an urban area and for the zones being used in the development of the transportation plan for the area. To accomplish this, the first step was to review the historical pattern of income for several metropolitan areas within Texas. Census data were considered to be the best data available and since it was intuitively reasoned that estimates of households by income would be needed, the decision was made to use household distributions by income as the basis for analysis. While data were available for MSAs in Texas for the 1950, 1960, 1970, and 1980 census, the data were not reported in terms of constant dollars and were not directly comparable. The data for income were for the years 1949, 1959, 1969, and 1979. In addition, the intervals used in the distributions of families by income were also different. A distinction must be made at this point involving terminology. Census data up until 1979 reported only income distributions for families and unrelated individuals. These were not the same as households which were also reported in 1979. The assumption was made that the change in family income characteristics relative to the distribution by income would closely parallel that for households and therefore, could be used as a basis for developing the methodology for projecting household distributions.

The first difficulty to overcome was that of comparability in terms of dollar value. A dollar earned in 1949 is not comparable to a dollar earned in 1979. A method, therefore, was sought that could be used to put the dollar values for each census in terms of a common base which would allow them to be comparable. The best measure for this is the Consumer Price Index (CPI) compiled by the United States government since the mid-forties. It was realized that this measure would not be absolutely correct, but it represented the most accurate measure available for converting historical dollar values into comparable values. Since the CPI is measured in terms of 1967 dollars, it was decided to convert the dollar values from each census to 1967 dollars. The CPI for 1949 was 0.714, for 1959 was 0.873, for 1969 was 1.098, and for 1979 was 2.174.

The second difficulty was the inconsistent income ranges between census years. For example, in the 1949 census, family income was reported in terms of $500 increments up to $5,000, $1,000 increments from $5,000 to $7,000, $3,000 increments from $7,000 to $10,000, and the last interval was $10,000 or more. In the 1979 census, family income was reported as less than $5,000, $5,000 to $7,499, $7,500 to $9,999, $10,000 to $14,999, etc. In order to compare the distribution from one census year with that of another, consistent intervals of income in constant dollars were developed. The conversion of the dollar values for the intervals to constant dollars (i.e., 1967 $) was fairly straightforward, but it still left different numbers of intervals between the census years and different values. It was decided after some trial and error to use $1,000 increments for the intervals up to $35,000, with the last interval being $35,000 or more. Tables 9 and 10 give illustrative examples of the results of converting the family income distributions to constant dollars for each census year for the San Antonio MSA. The next step was to convert the census distributions into expanded distributions using constant dollars and consistent income intervals.

The conversion of the census distributions into expanded distributions with common income intervals required two major assumptions. The first was that families were equally distributed within each interval where the interval was split between two of the expanded intervals. For example, the 1949 distribution of families by income for San Antonio, using the 1967 dollar estimates for the interval, has 6.7 percent of the families in the first interval with less than $700 income and 6.9 percent in the second interval with $700 to $1,399 income. Using this assumption, the estimated number of families (percent) which would be in the interval $0 to $999 (1967 dollars) would be 6.7 plus 300 divided by 700 times 6.9 percent or 6.7 plus 3.0 which yields an estimate of 9.7 percent. This procedure works well when splitting one interval into two parts.

The next major assumption was that the actual distribution of families (and households) by income is similar to that of a Gamma distribution with the characteristics of being non-negative, i.e., having a minimum observed value, being skewed to the left, and having a long tail. The Gamma distribution has these characteristics and has been used in other studies to estimate and predict income distributions. Using this assumption allowed further assumptions to be made relative to the distribution of families within a single interval to multiple intervals. In most cases, a simple linear relation was assumed which, knowing the beginning point and the total families to be distributed, allowed the slope and intercept to be easily calculated. This was not the assumption used for distributing the families over the larger intervals which fell within the tail of the distribution. In those cases, the assumption was made that the distribution would follow a steadily declining curve which was calculated by knowing the beginning point and the total to be distributed and by assuming an ending point (i.e., a value for the last interval). An iterative procedure was used to calculate the parameter value of the declining function. These relationships and equations were first worked out by hand for each census. A computer program was then written to perform the computations for each of 15 MSAs in Texas. The program also performs several calculations to determine if the expanded distribution exhibits the same characteristics of the unexpanded distribution. Three comparisons were made: mean income, median income, and the percentage of income earned by each pentile (quintile) of families. A correlation coefficient was computed for each comparison to determine the statistical fit achieved between the actual data and the estimated variables. Of the three comparisons, none of the correlation coefficients were less than 0.99, indicating that the expanded distributions were reasonable estimates of the unexpanded census distributions.

Table 9
San Antonio MSA Family Income Distributions, 1949 - 1959

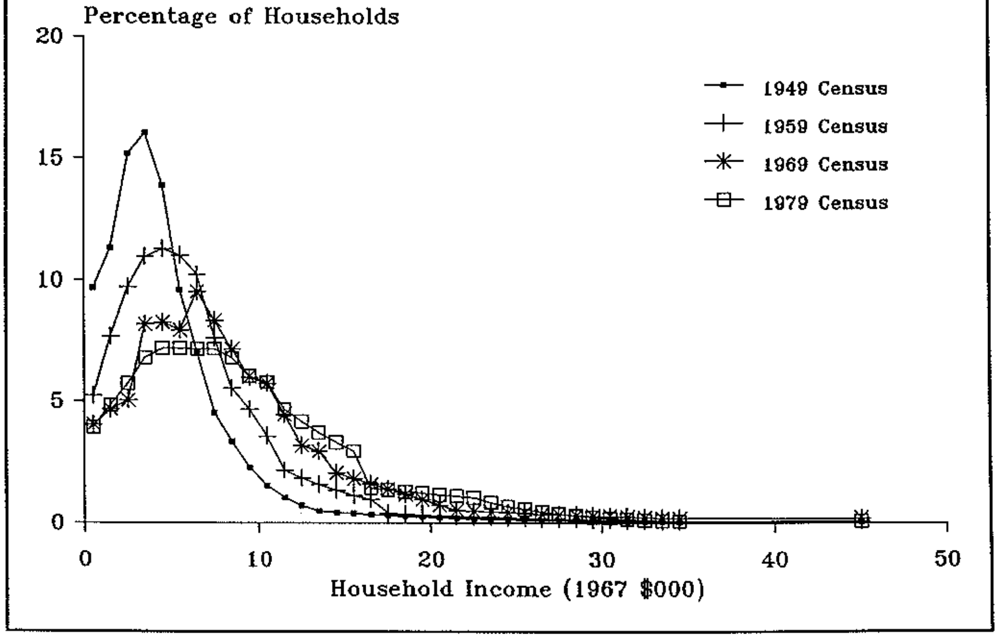
1949 Income	1967 Dollars	Pct Families	1959 Income	1967 Dollars	Pct Families
< $500	< $700	6.7	< $1000	< $1145	6.0
$500 to $999	$700 to $1399	6.9	$1000 To $1999	$1145 to $2290	9.3
$1000 to $1499	$1400 to $2099	9.1	$2000 To $2999	$2291 to $3435	11.9
$1500 to $1999	$2100 to $2800	10.3	$3000 To $3999	$3436 to $4581	13.0
$2000 to $2499	$2801 to $3500	12.6	$4000 To $4999	$4582 to $5726	12.8
$2500 to $2999	$3501 to $4200	9.9	$5000 To $5999	$5727 to $6872	12.2
$3000 to $3499	$4201 to $4901	10.0	$6000 To $6999	$6873 to $8017	8.7
$3500 to $3999	$4902 to $5601	7.3	$7000 To $7999	$8018 to $9163	6.3
$4000 to $4499	$5602 to $6301	5.8	$8000 To $8999	$9164 to $10308	5.2
$4500 to $4999	$6302 to $7001	4.5	$9000 To $9999	$10309 to $11454	3.6
$5000 to $5999	$7002 to $8402	6.3	$10000 To $14999	$11455 to $17181	7.7
$6000 to $6999	$8403 to $9803	3.6	$15000 To $24999	$17182 to $28636	2.1
$7000 to $9999	$9804 to $14004	3.9	$25000 and Over	$28637 and Over	1.2
$10000 and Over	$14005 and Over	3.1			

Table 10
San Antonio MSA Family Income Distributions, 1969 - 1979

1969 Income	1967 Dollars	Pct Families	1979 Income	1967 Dollars	Pct Families
< $1000	< $911	3.6	< $5000	< $2300	14.9
$1000 to $1999	$912 to $1821	4.1	$5000 to $7499	$2300 to $3449	8.6
$2000 to $2999	$1822 to $2731	5.1	$7500 to $9999	$3450 to $4599	9.1
$3000 to $3999	$2732 to $3642	6.2	$10000 to $14999	$4600 to $6899	16.9
$4000 to $4999	$3643 to $4553	7.3	$15000 to $19999	$6900 to $9199	14.9
$5000 to $5999	$4554 to $5464	7.8	$20000 to $24999	$9200 to $11499	11.8
$6000 to $6999	$5465 to $6374	8.1	$25000 to $34999	$11500 to $16099	13.6
$7000 to $7999	$6275 to $7285	7.9	$35000 to $49999	$16100 to $22999	6.8
$8000 to $8999	$7286 to $8196	7.5	$50000 and Over	$23000 and Over	3.4
$9000 to $9999	$8197 to $9107	6.3			
$10000 to $11999	$9108 to $10928	10.8			
$12000 to $14999	$10929 to $13660	10.5			
$15000 to $24999	$13661 to $22768	11.5			
$25000 to $49999	$22769 to $45536	2.7			
$50000 and Over	$45537 and Over	0.6			

Figure 21 presents a plot of the San Antonio expanded distributions for each of the census years. Figure 22 presents a plot of the Bryan-College Station expanded income distributions. Assuming that these distributions of families by income accurately reflect the actual distributions, the first observation made was that there have been significant shifts in real income growth and purchasing power between 1949 and 1979. The second observation was that while real family income has grown (as reflected by both the mean and median family income), the rate of growth has been steadily declining. For example, the mean family income in the San Antonio MSA in 1949 was $4,756 (1967 $). Between 1949 and 1959, the mean family income grew by 3.12 percent per year to $6,467 (1967 $). This rate of growth declined to 2.9 percent per year between 1959 and 1969 and to 0.66 percent per year between 1969 and 1979. Similar trends were noted in the other MSAs.

Figure 21
Expanded Household Income Distributions
San Antonio MSA



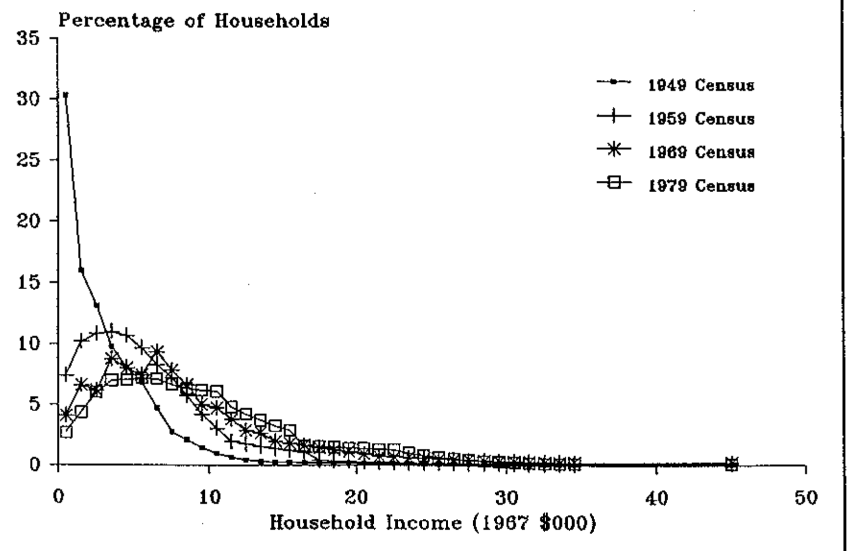


Figure 22
Expanded Household Income Distributions
Bryan–College Station MSA

The establishment of consistent comparable distributions of income provide a basis for developing a methodology for their estimation and subsequent projection. As mentioned previously, the distribution of families (and households) by income have been found to be similar to that of the Gamma distribution. The general formula for a two parameter Gamma distribution is as follows:

$$f(t) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$$

where:

$f(t)$ = *Relative percentage of households at income t*

α = *the shape parameter*

β = *the scale parameter*

$\Gamma(\alpha) = (\alpha - 1)!$

e = *base of natural logarithms* (2.71828)

t = *relative income*

To determine the applicability of using the Gamma distribution for estimating income distributions, the parameter values were calculated for each MSA income distribution, estimates of the percentage of families for each income interval calculated using the parameter values, and compared with the actual distribution values to determine the "fit". The comparison was done by calculating the correlation coefficient for each set of actual versus estimated values. Table 11 presents the results for each MSA and each census year. The results of the curve fit indicate that the Gamma distribution can be used to estimate distributions of families (and households) by income. The reason that the values of alpha and beta were the same was that the income values were nondimensionalized by dividing by the respective mean income prior to the calculation of the parameter values. This results in each distribution having a mean of 1.0 which negates the scale parameter (i.e., beta) and results in the need to only compute one parameter (i.e., the shape parameter, alpha).

It should be noted that the value of alpha changes over time. It was originally thought that one curve could be fitted to an urban area and then used to predict future distributions based on a projected mean income. The fact that alpha changes over time indicates that this procedure would not be applicable. Examining the relationship between the value of alpha and the mean income did appear to exhibit a strong correlation. A linear regression analysis was subsequently performed which yielded a correlation coefficient of 0.9156 with a "Z" statistic of 5.1193. This analysis was done using all 60 values from Table 11 and indicates a relatively good estimate of alpha may be obtained using the estimated mean income.

The data presented have been family income which while assumed to be relationally consistent with household income, is not the same as household income. For example, in 1979 the mean family income for San Antonio (in 1967 $) was $9,190; while the mean household income (in 1967 $) was $8,285. This relationship was consistent for all the MSAs being analyzed; the mean household income was significantly less than the mean family income. An analysis of the distribution of households by income also revealed that the distribution was more skewed to the left than that of the distribution of families by income. This was expected since both distributions have the same characteristics in terms of shape, and a lower mean would imply a distribution more skewed to the left. While the assumption that household distributions by income will follow the same historical pattern as family distributions is still considered valid, it is not valid to use the same values of alpha for estimating household distributions. Table 12 presents the results of fitting the Gamma distribution to the 1979 household distributions by income for the MSAs under study. Based on the results in Table 12, the Gamma distribution does a good job of estimating the distribution of households by income. A linear regression performed on the alpha values versus the mean income values resulted in a correlation coefficient value of 0.9728 with a "Z" statistic value of 3.2237. It was also observed in comparing the results of this linear regression with the one for family income distributions (i.e., alpha versus mean family income) that the slope of the linear relationships were nearly the same, .000214 versus .000242. This implies that the two distributions maintain a relatively consistent relationship with each other over time since the dependent variable in both equations was the shape parameter alpha. The conclusion of this analysis was that, given an estimate of the mean household income for an urban area, a value may be derived for alpha which when used in the Gamma distribution will provide a reasonable estimate of the distribution of households by income for the urban area. The next step was estimating the income for zones within the urban area.

Table 11
Income Distributions Gamma Curve Fit Results

MSA	Year	Mean Income	Alpha Value	Beta Value	Cor. Coef.	Z Stat.
Amarillo	1949	6,088	1.496	1.496	0.898	8.39
	1959	7,768	2.122	2.122	0.949	10.45
	1969	9,274	2.274	2.274	0.959	11.10
	1979	10,417	2.579	2.579	0.983	13.61
Austin	1949	5,098	1.278	1.278	0.944	10.19
	1959	6,994	1.604	1.604	0.965	11.59
	1969	9,894	2.052	2.052	0.977	12.75
	1979	10,566	2.486	2.486	0.981	13.31
Beaumont-Port Arthur	1949	5,516	2.002	2.002	0.917	9.03
	1959	7,296	2.172	2.172	0.927	9.40
	1969	9,102	2.300	2.300	0.950	10.51
	1979	10,676	2.728	2.728	0.934	9.72
Corpus Christi	1949	5,243	1.466	1.466	0.963	11.40
	1959	6,619	1.588	1.588	0.979	13.04
	1969	8,527	1.898	1.898	0.974	12.46
	1979	9,843	2.167	2.167	0.968	11.84
El Paso	1949	5,211	1.445	1.445	0.958	11.06
	1959	6,955	1.904	1.904	0.974	12.46
	1969	8,417	1.940	1.940	0.975	12.59
	1979	8,501	1.925	1.925	0.994	16.73
Galveston	1949	5,307	1.746	1.746	0.941	10.03
	1959	6,921	1.964	1.964	0.967	11.75
	1969	9,730	2.443	2.443	0.954	10.74
	1979	11,270	2.863	2.863	0.955	10.80
Laredo	1949	3,670	0.546	0.546	0.896	8.33
	1959	4,581	1.029	1.029	0.964	11.48
	1969	6,094	1.211	1.211	0.949	10.46
	1979	7,269	1.443	1.443	0.995	17.05
Lubbock	1949	5,757	1.414	1.414	0.937	9.83
	1959	7,469	1.686	1.686	0.947	10.36
	1969	9,016	1.856	1.856	0.962	11.36
	1979	9,947	2.284	2.284	0.988	14.57

Note: All mean incomes are in 1967 dollars as computed from expanded income distributions.

Table 11 (continued)
Income Distributions Gamma Curve Fit Results

MSA	Year	Mean Income	Alpha Value	Beta Value	Cor. Coef.	Z Stat.
San Angelo	1949	4,979	1.082	1.082	0.904	8.57
	1959	6,422	1.472	1.472	0.954	10.74
	1969	8,263	1.759	1.759	0.961	11.27
	1979	9,455	2.190	2.190	0.990	15.33
San Antonio	1949	4,757	1.294	1.294	0.941	10.03
	1959	6,467	1.683	1.683	0.973	12.37
	1969	8,607	1.894	1.894	0.967	11.75
	1979	9,190	2.111	2.111	0.982	13.55
Waco	1949	4,505	1.096	1.096	0.925	9.32
	1959	6,218	1.641	1.641	0.974	12.42
	1969	8,154	1.950	1.950	0.973	12.31
	1979	9,010	2.155	2.155	0.989	14.85
Wichita Falls	1949	5,379	1.241	1.241	0.882	7.96
	1959	6,992	1.935	1.935	0.955	10.81
	1969	8,594	1.874	1.874	0.948	10.40
	1979	9,582	2.278	2.278	0.989	14.98
Dallas-Ft. Worth	1949	5,636	1.571	1.571	0.934	9.69
	1959	7,756	1.892	1.892	0.963	11.37
	1969	10,619	2.413	2.413	0.963	11.42
	1979	11,304	2.637	2.637	0.979	13.01
Houston	1949	5,692	1.678	1.678	0.940	9.97
	1959	7,962	1.951	1.951	0.958	11.06
	1969	10,446	2.385	2.385	0.956	10.89
	1979	12,243	2.809	2.809	0.953	10.69
Bryan-College Station	1949	3,466	0.695	0.695	0.996	17.90
	1959	6,018	1.583	1.583	0.991	15.57
	1969	8,577	1.664	1.664	0.977	12.81
	1979	9,581	2.177	2.177	0.992	15.68

Note: All mean incomes are in 1967 dollars as computed from expanded income distributions.

Table 12
Household Income Distributions
Gamma Curve Fit Results

MSA	Year	Mean Income	Alpha Value	Beta Value	Cor. Coef.	Z Stat.
Amarillo	1979	9,071	1.935	1.935	0.977	12.75
Austin	1979	8,813	1.783	1.783	0.980	13.22
Beaumont - Port Arthur	1979	9,496	2.084	2.084	0.906	8.65
Corpus Christi	1979	8,953	1.834	1.834	0.966	11.66
El Paso	1979	7,918	1.713	1.713	0.993	16.11
Galveston	1979	9,965	2.181	2.181	0.934	9.69
Laredo	1979	6,786	1.282	1.282	0.998	19.34
Lubbock	1979	8,662	1.773	1.773	0.987	14.55
San Angelo	1979	8,146	1.657	1.657	0.991	15.43
San Antonio	1979	8,285	1.765	1.765	0.982	13.49
Waco	1979	7,648	1.567	1.567	0.987	14.52
Wichita Falls	1979	8,395	1.745	1.745	0.985	14.02
Dallas-Ft. Worth	1979	9,989	2.066	2.066	0.973	12.29
Houston	1979	10,906	2.246	2.246	0.950	10.51
Bryan-College Station	1979	7,467	1.418	1.418	0.995	16.99

Note: All mean incomes are in 1967 dollars as computed from expanded distributions.

ESTIMATING ZONAL INCOMES

The estimation of income for zones is a more difficult problem than that of estimating income and income distributions for an urban area. It is more difficult because of a lack of historical data at the zone level (zone boundaries may change over time), changes occur more frequently within zones, and some zones may be completely undeveloped in the base year being analyzed and projected for partial or full development in the forecast year.

The methodology as proposed in this report assumed the following:

1. A distribution of households by income has been developed for the area under

study, both for the base year and the forecast year.

2. An estimate of the mean household income has been developed for each zone for the base year under analysis. This would normally come from census data or historical information, if available.
3. Base year and forecast year estimates of population and number of households is available for each zone.

The first step is to identify those zones which are undeveloped in the base year and projected to be partially or fully developed in the forecast year. The methodology as proposed is predicated on each zone with households in the forecast year having an estimate of mean income in the base year. This is impossible for zones which are undeveloped in the base year and creates a situation where an assumption must be made relative to what the most likely income would be for those zones if they were developed in the base year. This estimate can be made based on the analyst's judgment or, as recommended here, be calculated based on the characteristics of the zones around the undeveloped zone. It is recommended that a weighted average be computed of the mean incomes of the zones (those developed) around the undeveloped zone and that this be input as the base year estimate of the undeveloped zone's mean income.

The second step is to develop initial growth factors for each income group being used in the trip production model. If preferred, these factors can be estimated for any set of income groupings; but it is suggested that no more than five or six be used. Five, for example, allows a logical break in terms of income pentiles (quintiles) and is used here for illustration.

Using the base year household income distribution and the forecast year household income distribution, the growth in income for each of the selected income groupings can be computed. This growth should be expressed as a factor for each income group. These factors then become the basis for the initial estimates of zonal mean income in the forecast year. This is done by determining the income group into which each zone falls based on its mean income in the base year. The growth factor for that income group is applied to the base year mean income for the zone to develop the initial projection of mean income for the zone.

The third step begins the iterative process involved in this methodology. With the initial estimate of mean income for each zone, a distribution of households by income for each zone is developed using the methodology as described in the previous section. These "zonal" income distributions are added together to create a distribution for the urban area which is then compared to that forecast for the urban area. Adjustment factors are computed for each income range being used in the distributions and applied to each zonal distribution. Using the adjusted distribution for each zone, a new estimate of mean income for each zone is calculated. This new estimate of the mean is used to compute a new distribution using the methodology described in the previous section, and the process is repeated. There are several constraints on the process simply because if left to iterate indefinitely, the end result would be that each zone would have the same mean and distribution as the urban area, which is not realistic or accurate. These constraints are both internal and external. Internally, the model will not allow the mean income for a zone to decline unless the population for a zone is projected to decline. This is based on the assumption that if an area is growing economically as reflected by growth in real income, this growth will occur across the entire area. Certain sectors of the area will grow at different rates while some sectors which are experiencing negative indicators of change may actually decline. For example, a negative indicator would be a declining population in a zone or simply a change in a neighborhood.

The external constraint indicates which zones should or should not be included in the iterations. This allows the model to incorporate local area conditions. Such conditions might include zones which are extremely stable and not likely to experience any significant change in income. These zones are included in the first iteration but excluded thereafter with regard to adjustments in their distribution or mean. The end results are zonal estimates of mean income which when distributed and combined reflect the projected distribution of households by income for the entire area. These zonal income distributions are also the basis for disaggregating households by income groups in the default disaggregation model described in a later section of this report.

The discussion up to this point has dealt with projecting income at the urban area level and at the zone level. This is the basis for the remaining sections which discuss the estimation of households within income ranges for use in the trip generation model TRIPCAL5. These are referred to as disaggregation models since the normal input to a trip generation model for a zone will be just the total number of households in the zone and the median income for the zone.

TRIPCAL5 Disaggregation Models

In order to use income in a cross-classification trip generation model, an estimate is needed of the number of households in each zone whose annual household income is estimated to fall within the ranges being used. TRIPCAL5 provides three ways for these estimates to be input. In the first, is where the user inputs the estimates for each zone into the program. This requires the user to develop the estimates of households within each income range for each zone. The second allows the user to input disaggregation curves which, given the median income of a zone, allows the percentage of households in the zone that fall within the ranges of income being used in the model to be estimated. The third, considered the default routine, uses the theoretical methods (as described in the previous sections) to estimate the distribution of households by income for each zone and computes from that the number of households within each income range in the model. The purpose of this section is to describe the methods by which the second and third ways are developed, including the assumptions and theoretical considerations involved in each.

DISAGGREGATION CURVES

Disaggregation curves are numerical tables which can be input to the trip generation program. Using a pre-selected measure (e.g., median household income), the information in the table allows an estimate of the percentage of households within each income range to be generated within the program. These curves are generally developed using census data for the urban area under study. This methodology has been used in several urban areas for disaggregating households into income groups and household size. The normal procedure is to plot the census data for the urban area and handfit curves to the plots. Two things have to be decided first in developing this procedure: the measure to be used in disaggregating the households for each zone and the income ranges used in the trip generation model.

Theoretically, it is logical to assume that the mean or median income for a zone will be correlated with the distribution of households by income. As will be demonstrated in the discussion concerning the third method, these values are used in estimating the distribution of households for the urban area. For purposes of TRIPCAL5, the variable used is the ratio of the zone's median income to the urban area's median income. The same result would be achieved using just the median income; but the use of the ratio allows more discrete values (i.e., 0.1 to 3.0 rather than 500.0 to 35000.0), which simplifies the data development and input requirements.

The next step is that of selecting the income ranges to be used in the trip generation model and the disaggregation model. For illustration purposes, income pentiles (also referred to as quintiles) are used in this report for determining the income ranges for use in the model. Income pentiles may be defined as those income ranges within which 20 percent of the households would fall in terms of their annual income. The basis for determining these values is to use the distribution of households by income as reported in the latest census. The percentage of households within each income range are simply added (starting with the lowest) to determine the income range that contains 20 percent of the households. For example, Table 13 shows the estimated percentage of households by income range for 1979 for the Bryan-College Station (B-CS) MSA. Using the information in Table 13, the first income pentile would fall in the income range $0 to less than $5,000. The second income pentile would fall in the range $5,000 to $9,999, the third in the range $10,000 to $19,999, the fourth in the range $20,000 to $24,999, and the fifth in the range $25,000 and more. It will be noted that the values were rounded up to be able to use the ranges as specified in the table. This round-up can create some problems depending upon the income intervals reported in the census. For example, the third pentile income range should begin somewhere between $7,500 and $9,999 and end somewhere between $15,000 and $19,999. For this reason, the distribution for B-CS expanded and put in terms of 1967 dollars for defining the income ranges for the income pentiles. The procedure used for this was described earlier and was applied to the census tracts as well. It should be noted that use of this procedure is optional. The resulting distribution for B-CS is shown in Table 14. Note that it also required some rounding but was not as significant as that required using the data as reported in the census. Either method should produce acceptable results.

Table 13
Bryan-College Station MSA 1979 Income Distribution

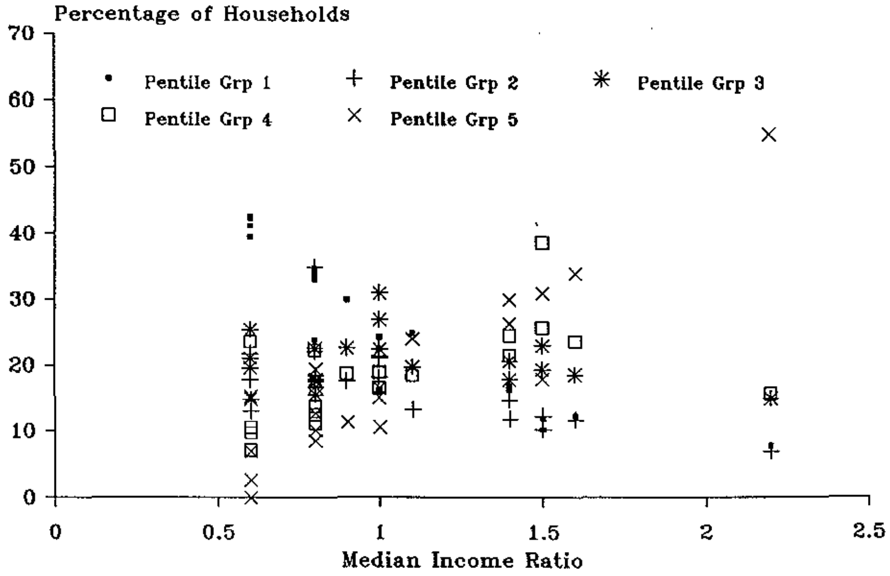
Income In 1979	No. Households	Percentage Of Households	Cumulative Percentage
Less Than $5000	6631	20.4	20.4
$5000 TO $7499	3621	11.1	31.5
$7500 TO $9999	3157	9.7	41.2
$10000 TO $14999	5245	16.1	57.3
$15000 TO $19999	3998	12.3	69.6
$20000 TO $24999	3178	9.8	79.4
$25000 TO $34999	3595	11.1	90.5
$35000 TO $49999	2037	6.3	96.8
$50000 or more	1031	3.2	100.0
TOTALS	32493	100.0	-

Source: 1980 Census

Once the income ranges have been identified, the next step is to tabulate the percentage of households in each income range for each census tract in the urban area. Census tracts are recommended for use because the data are readily available. The tabulated values for the census tracts in B-CS are shown in Tables 15 and 16. Table 15 shows the numbers as developed using the income ranges from Table 13, and Table 16 shows the numbers as developed using the income ranges from Table 14. These are for comparative purposes only. A comparison of the two tables does reveal a significant difference between the two relative to pentile four. The impact on estimates of travel could also be significant due to the difference in trip rates between income pentiles. For purposes of illustration and discussion relative to the presentation of this methodology, the analysis presented in this report will deal with the data presented in Tables 14 and 16.

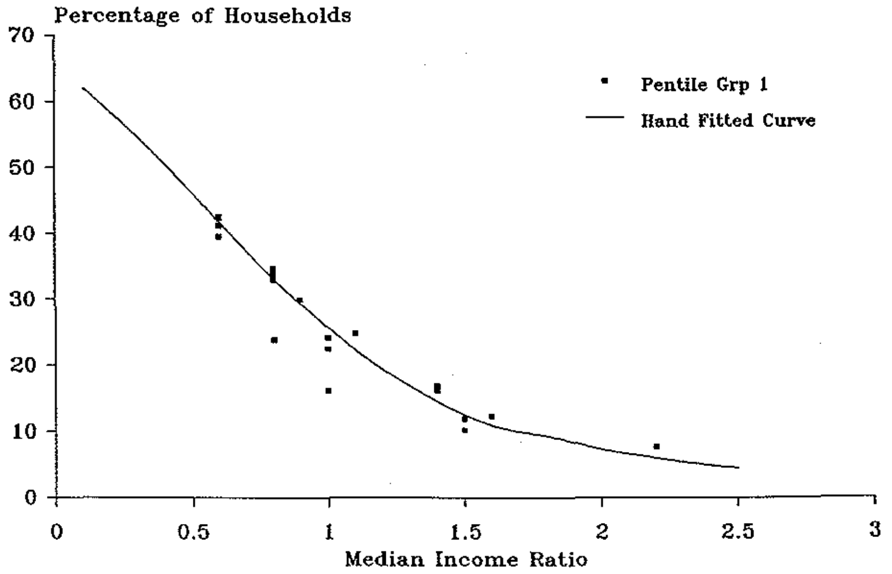
Figure 23 presents a plot of the data shown in Table 16, with the x-axis representing the ratio of a census tracts median income to the MSA median income. As will be noted, the data points are scattered and, when presented on the same graph, very difficult to ascertain any distinct relationship. Figures 24 through 28 present plots of the same data except the data for each pentile are shown separately, allowing the analyst to develop the relationships needed for input to the model. These are generally developed by handfitting curves to the data. Regression analysis could be used to fit both linear and nonlinear curves, but it will depend upon the analyst and the data being analyzed. For purpose of the example shown, curves were fitted by hand to develop the input data for input to the model. Figures 24 through 28 also show the handfitted curves. Using the curves, values were calculated for each point on the x-axis and tabulated. These values were forced to add to 100 percent for each value of x (i.e., the ratio of median income for a subarea to the median income for the entire area). The adjustments to ensure the sum was 100 percent were entirely based on the judgment of the analyst in terms of what would normally be expected. The results of those computations are shown in Table 17. Table 17 would then be the data actually input to the trip generation model and used to disaggregate the households in a zone into the number of households estimated to fall within each income pentile. To determine the relative accuracy of this method, the percentage of households in each income pentile was computed using Table 17 for each census tract in B-CS and compared with the actual percentages reported in the 1980 census. The comparison of values was done by computing the correlation coefficient for the two sets of values. The correlation coefficient was 0.85 which indicated a fairly good estimation of the actual values.

Figure 23
Distribution of Households by Pentile
Income Groups – Bryan–College Station



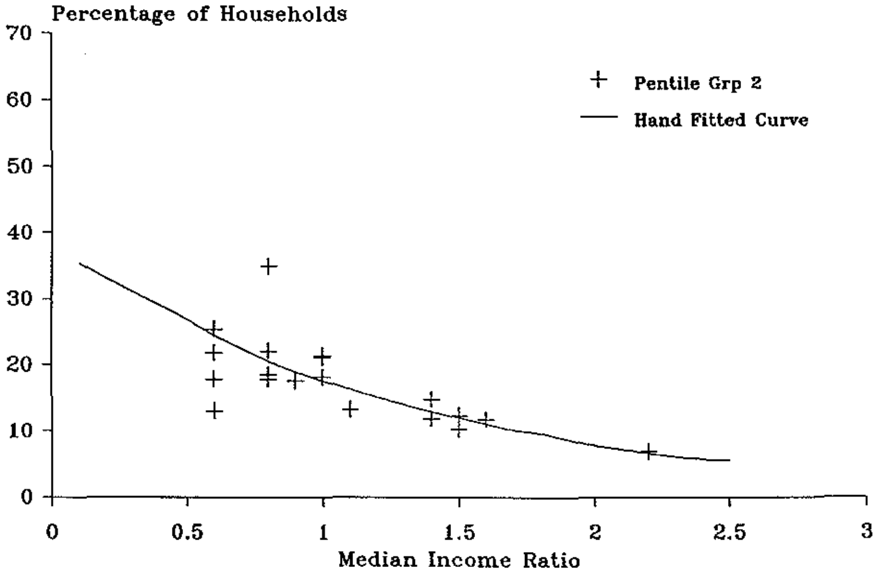
Computed from 1979 Census Data

Figure 24
Disaggregation Curve for Pentile
Group 1 – Bryan–College Station



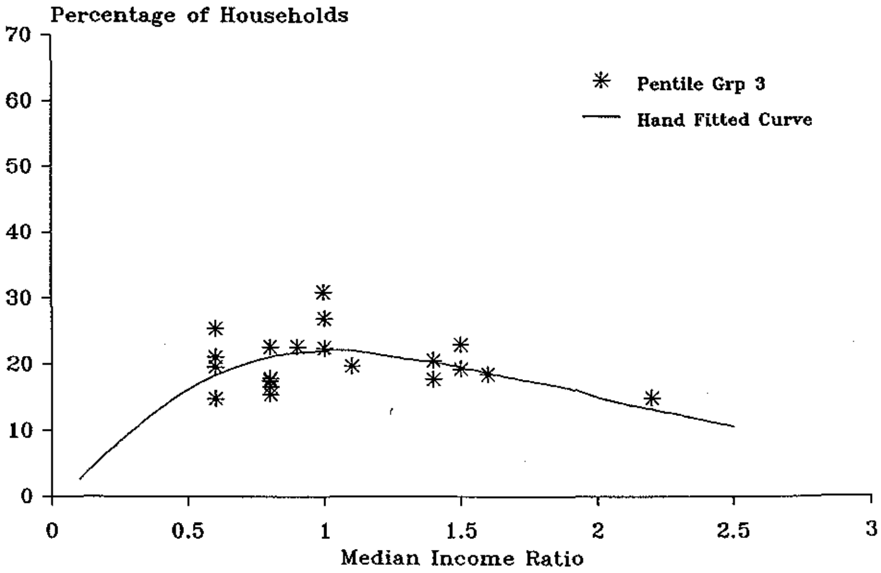
Computed from 1979 Census Data

Figure 25
Disaggregation Curve for Pentile
Group 2 – Bryan–College Station



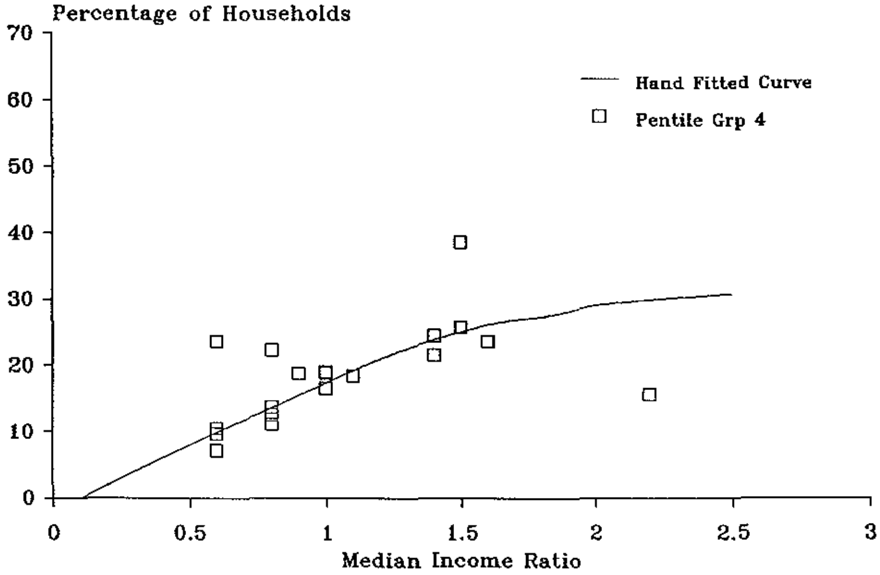
Computed from 1979 Census Data

Figure 26
Disaggregation Curve for Pentile
Group 3 – Bryan–College Station



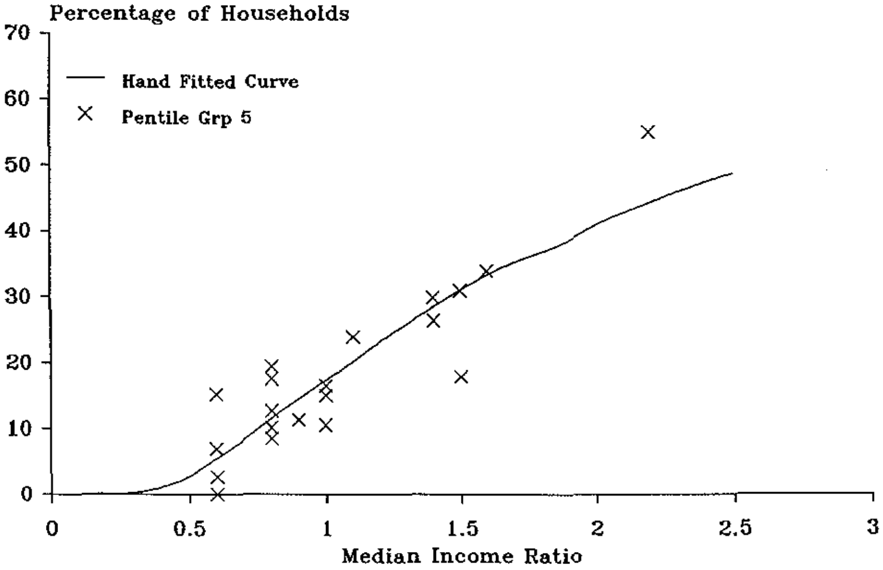
Computed from 1979 Census Data

Figure 27
Disaggregation Curve for Pentile
Group 4 – Bryan–College Station



Computed from 1979 Census Data

Figure 28
Disaggregation Curve for Pentile
Group 5 – Bryan–College Station



Computed from 1979 Census Data

Table 14
Bryan-College Station
MSA Expanded 1979 Income Distribution

1979 Income in 1967 Dollars	No. Households	Percentage Of Households	Cumulative Percentage
$0 TO $999	2821	8.68	8.68
$1000 TO $1999	2945	9.06	17.74
$2000 TO $2999	3069	9.45	27.19
$3000 TO $3999	2927	9.01	36.20
$4000 TO $4999	2559	7.88	44.08
$5000 TO $5999	2281	7.02	51.10
$6000 TO $6999	2226	6.85	57.95
$7000 TO $7999	1867	5.75	63.70
$8000 TO $8999	1609	4.95	68.65
$9000 TO $9999	1453	4.47	73.12
$10000 TO $10999	1382	4.25	77.37
$11000 TO $11999	1082	3.33	80.70
$12000 TO $12999	946	2.91	83.61
$13000 TO $13999	827	2.54	86.15
$14000 TO $14999	723	2.22	88.37
$15000 TO $15999	632	1.94	90.31
$16000 TO $16999	344	1.06	91.37
$17000 TO $17999	329	1.01	92.38
$18000 TO $18999	315	0.97	93.35
$19000 TO $19999	301	0.93	94.28
$20000 TO $20999	288	0.89	95.17
$21000 TO $21999	275	0.85	96.02
$22000 TO $22999	264	0.81	96.83
$23000 TO $23999	213	0.65	97.48
$24000 TO $24999	171	0.53	98.01
$25000 TO $25999	138	0.43	98.44
$26000 TO $26999	111	0.34	98.78
$27000 TO $27999	90	0.28	99.06
$28000 TO $28999	72	0.22	99.28
$29000 TO $29999	58	0.18	99.46
$30000 TO $30999	47	0.14	99.60
$31000 TO $31999	38	0.12	99.72
$32000 TO $32999	31	0.09	99.81
$33000 TO $33999	25	0.08	99.89
$34000 TO $34999	20	0.06	99.95
$35000 OR MORE	16	0.05	100.0
TOTALS	32493	100.00	-

Table 15
Bryan-College Station Census Tract Household Distributions by Income Pentile Based on Unexpanded Census Distributions

Percentage Households

Census Tract	Median Income Ratio	Pentile Group 1	Pentile Group 2	Pentile Group 3	Pentile Group 4	Pentile Group 5
1	1.4	10.5	19.0	29.5	13.0	28.0
2	1.0	18.1	20.8	31.9	11.1	18.1
3	1.5	9.1	13.0	33.0	25.9	19.0
4	0.8	27.1	25.5	29.9	6.9	10.6
5	0.6	32.3	27.9	28.0	4.0	7.8
6.01	0.6	24.2	29.0	35.5	11.3	0.0
6.02	0.9	23.6	20.5	32.4	11.3	12.2
7	1.0	9.2	23.7	42.6	8.4	16.1
8	1.6	8.9	12.5	27.8	14.9	35.9
9	1.0	13.9	26.1	39.8	8.7	11.5
10	0.8	25.1	24.3	24.8	7.5	18.3
11	1.5	6.5	10.8	35.0	15.2	32.5
13	0.6	33.9	24.1	20.3	5.8	15.9
14	0.6	28.5	32.5	34.0	2.3	2.7
15	0.8	17.5	39.7	22.2	7.9	12.7
16	0.8	24.9	24.5	22.8	7.1	20.7
17	0.8	25.2	24.7	24.9	16.3	8.9
18	1.4	11.5	13.0	28.7	14.7	32.1
19	2.2	6.4	6.2	22.6	7.3	57.5
20	1.1	18.8	16.2	28.3	11.0	25.7

Table 16
Bryan-College Station Census Tract Household Distributions by Income Pentile Based on Expanded Income Distributions

Percentage Households

Census Tract	Median Income Ratio	Pentile Group 1	Pentile Group 2	Pentile Group 3	Pentile Group 4	Pentile Group 5
1	1.4	16.8	14.7	17.8	24.4	26.3
2	1.0	24.2	18.1	22.4	18.8	16.5
3	1.5	11.9	12.3	19.3	38.6	17.9
4	0.8	34.1	22.1	22.6	11.1	10.1
5	0.6	41.2	21.8	19.6	10.5	6.9
6.01	0.6	42.3	13.0	21.1	23.6	0.0
6.02	0.9	29.9	17.5	22.6	18.7	11.3
7	1.0	16.2	21.4	30.9	16.5	15.0
8	1.6	12.3	11.7	18.5	23.6	33.9
9	1.0	22.5	21.1	26.9	19.0	10.5
10	0.8	33.6	18.5	17.9	12.5	17.5
11	1.5	10.2	10.2	23.0	25.7	30.9
13	0.6	42.5	17.8	14.8	9.7	15.2
14	0.6	39.5	25.4	25.4	7.1	2.6
15	0.8	23.8	34.9	17.5	11.1	12.7
16	0.8	33.0	18.4	15.5	13.7	19.4
17	0.8	34.7	17.8	16.7	22.3	8.5
18	1.4	16.2	11.8	20.6	21.5	29.9
19	2.2	7.7	6.9	14.9	15.6	54.9
20	1.1	24.8	13.2	19.7	18.4	23.9

Table 17
Bryan-College Station
Household Disaggregation Curve Data

Median Income Ratio	Pentile Group 1	Pentile Group 2	Pentile Group 3	Pentile Group 4	Pentile Group 5
0.1	62.0	35.3	2.7	0.0	0.0
0.2	58.4	33.1	6.5	2.0	0.0
0.3	54.7	31.0	10.3	4.0	0.0
0.4	50.5	29.0	13.5	6.0	1.0
0.5	46.2	27.0	16.3	8.0	2.5
0.6	41.7	24.3	18.5	10.0	5.5
0.7	37.4	22.5	20.0	11.8	8.3
0.8	33.0	20.4	21.3	13.6	11.7
0.9	29.3	18.9	21.8	15.5	14.5
1.0	25.7	17.5	22.1	17.3	17.4
1.1	22.2	16.3	22.2	19.3	20.0
1.2	19.2	15.0	21.4	21.0	23.4
1.3	16.8	13.9	20.8	22.6	25.9
1.4	14.4	12.8	20.4	23.9	28.5
1.5	12.4	12.0	19.6	25.0	31.0
1.6	10.8	11.0	18.7	26.2	33.3
1.7	9.9	10.0	17.9	26.9	35.3
1.8	9.4	9.7	17.1	27.2	36.6
1.9	8.4	8.7	16.6	28.0	38.3
2.0	7.2	7.7	14.9	29.2	41.0
2.1	6.6	7.2	14.0	29.5	42.7
2.2	6.0	6.6	13.2	29.9	44.3
2.3	5.3	6.1	12.4	30.2	46.0
2.4	4.9	5.7	11.5	30.5	47.4
2.5	4.4	5.5	10.6	30.8	48.7

The basic operation of the model is fairly straightforward. Given the median income for a zone and the median income for the urban area as a whole, a ratio is calculated by dividing the zone's median income by the area's median income. Using that value, estimates of the percentage of households in each income pentile may be obtained directly from Table 17 and multiplied by the number of households in the zone to estimate the number of households in each group. This then becomes the marginal distribution of households by income pentile for the zone.

DEFAULT DISAGGREGATION MODEL

The default disaggregation model in TRIPCAL5 is based on the methodology previously discussed concerning the projection of distributions of households by income for the urban area and for zones within the area. The basis for the default model is that, given an estimate of the mean income for a zone, a distribution of households by income for the zone may be obtained using the Gamma distribution and the methodologies presented in previous sections of this report. Because the normal input data for trip generation includes an estimate of the zone's median income, it is necessary to derive a method for estimating the mean income for the zone. A linear regression was done using mean income as the dependent variable and median income as the independent variable. The data were compiled from the 1980 census for 26 MSAs in Texas with the mean and median values converted to 1967 dollars. The correlation coefficient for the regression was 0.92. This indicated a relatively good "fit" and was subsequently made a part of the default model. The equation from the regression was as follows:

$$\text{Mean Income} = 1.0397 \times (\text{Median Income}) + 1355.02$$

Having an estimate of the mean income then allows an estimate of the value of alpha (shape parameter) for use in the Gamma distribution. This estimate is derived based on the analysis presented previously in this report and uses the following equation:

$$\text{Alpha} = 0.000242 \times (\text{Mean Income}) - 0.3006$$

The scale parameter, beta, is set equal to the shape parameter, alpha, initially. The income values are nondimensionalized by dividing each by the estimated mean income. The number of income intervals used is 36 with the range of each being \$1,000 (1967 \$'s). The mid-points of the intervals are \$500, \$1,500, \$2,500, etc. The zonal distribution is then calculated using the nondimensionalized values of income (mid-point values divided by mean) with the values of alpha and beta in the Gamma distribution equation. The mean of the calculated distribution is computed and compared with the estimated mean. If within ± 1 percent, the program continues to the final step. If not within ± 1 percent, the value for beta is adjusted by the ratio of the distribution mean to the estimated mean, and a new distribution is calculated. This process is repeated until the distribution mean is within ± 1 percent of the estimated mean.

The final step is the computation of the percentage of households within the income ranges being used in the trip generation model. This is done by adding the percentage of households as estimated by the distribution within each of the income ranges. The resulting values become the marginal distribution of households within each income range being used in the model.

A test of this default model was conducted using data for the Bryan-College Station census tracts to see how well its results compared with those from inputting handfitted disaggregation curves. The correlation coefficient which resulted from comparing the estimates of households by income pentile with the actual number of households by income pentile was 0.83. The correlation coefficient obtained from the use of the disaggregation curves was 0.85. The conclusion was that the results of both models were adequate and were considered reasonable.

TRUCK/TAXI DEFAULT MODEL

In developing TRIPCAL5, a need was identified for an optional model that would estimate the total number of truck and taxi trips. The trip generation program would then provide three methods for estimating truck and taxi trip productions. The first method would be to input a model for computing truck-taxi trip productions by zone. The second would be to input an estimate (generated by hand or other means exogenous to the trip generation program) of the total truck and taxi trip productions in the area. The third would be the default model for the program.

The following sections describe the research and analysis done in developing a recommended model for use as a default in TRIPCAL5. The first section presents a brief overview of the state of the practice relative to the estimation of truck trips in travel demand modeling. The second section discusses the process by which a model was developed for use in TRIPCAL5. The third section presents the results of the calibration of the model, while the fourth discusses the validation of the calibrated model. The final section presents the summary and recommendations of this study.

STATE OF THE PRACTICE

The first step in this task was to review models used in other areas for estimating truck and taxi trips. A majority of the urban areas reviewed do not use truck and taxi trips as a separate trip purpose. These trips were considered a part of the non-home based travel in the urban area. Those areas that did use truck and taxi trips as a trip purpose had developed models for predicting those trips. It should be clarified that estimates of truck and taxi trips are somewhat misleading in terminology. Generally, truck and taxi refers to commercial truck travel and seldom includes taxi trips. While the term "truck and taxi" is used in this discussion, the models presented were, in fact, used to estimate commercial truck trips. Typically, taxi trips are assumed to be such a small fraction of the total travel in an urban area that the effort to identify a model for their estimation would not result in any significant improvement in the overall travel estimates. For that reason, the term "taxi" is not used in the following discussions.

The model reviewed for Atlanta, Georgia, (1) was developed based on data from a 1972 home interview survey which was expanded to represent average daily travel for the entire Atlanta region. A linear regression model was developed which used six independent variables with truck trips being the dependent variable. The following was the final model selected:

$$\begin{aligned}\text{Truck Trips} \quad = \quad & \ 0.0872\ (\text{Population}) \\ & +\ 0.2253\ (\text{Retail Employment}) \\ & +\ 0.0857\ (\text{Commercial and Governmental Employment}) \\ & +\ 0.2481\ (\text{Industrial Employment})\end{aligned}$$

+ 1.4404 (Construction Employment)
+ 0.2481 (Other Employment)

This model predicted truck attractions at the zonal level, and the assumption was then made that the zonal truck productions would be the same as the attractions.

The second model reviewed was that for Denver, Colorado (2). As in Atlanta, truck trip attractions were estimated with productions being set equal to attractions at the zonal level. The model for Denver was calibrated using 1971 data. Mathematically, the model is a regression type. The model is as follows:

Truck Trips = 0.43 (Households) + 0.36 (Total Employment)

The third model reviewed was developed for the Minneapolis-St. Paul, Minnesota area (3). An analysis of the truck trip models developed in 1970 was done as part of an overall effort to update the trip generation models based on information from a travel survey done in 1982. The model which was subsequently selected was different from the 1970 models, but there was no reported significant difference in terms of accuracy. The new model was selected primarily because it used readily available forecast variables, whereas the old models would have required forecasting additional variables. For that reason, both models are presented. Both of the models are linear regression in form. The 1970 models were two-tiered; one model was used to predict truck trips at a district level, and another was used to predict truck trips at the zonal level within the district with the zonal totals constrained to match the district total. The 1970 models were as follows:

District Model:

Truck Trips = 25.3230 + 0.1598 (Manufacturing Employment)
+ 0.2382 (Transportation, Communications, & Utilities Emp.)
+ 0.0406 (Total Population)

Zonal Level Model:

Truck Trips = 76.5390 + 0.0744 (Manufacturing Employment)
+ 0.0262 (Transportation, Communications, & Utilities Emp.)
+ 0.6060 (Net Industrial Acreage)
+ 0.13 (Total Population)

The subsequent analysis resulted in two new truck trip models:

Truck Non-CBD Trips = 0.0708 (Total Employment)
+ 0.0317 (Population)
- 0.2338 (Retail Employment)

Truck CBD Trips = 0.0466 (Total Employment)
+ 0.0209 (Population)
- 0.1539 (Retail Employment)

In both equations there is a negative coefficient for retail employment. This is due to retail employment being weaker in terms of generating truck trips than other types of employment.

The fourth model reviewed was used in Detroit, Michigan (4). As in the previous models, the trips being estimated were trip attractions with productions being set equal to attractions at the zonal level. The truck model for Detroit was also a linear regression model developed using data from a 1965 travel survey. The model was as follows:

Truck Trips = 187.03 + 0.30 (Basic Employment)
+ 0.74 (Wholesale Emp. + Retail Emp.)
+ 0.03 (Non-Basic Emp. - Wholesale Emp. + Retail Emp.)
+ 0.10 (Population)

where:

Basic employment included manufacturing, natural resources, transportation, communications, and utilities.

> Non-basic employment included wholesale, retail, service, public administration, finance, insurance, and real estate.

The statistical analysis of this model indicated (based on the t-statistic) that basic employment, wholesale plus retail, and population were the most significant variables.

Based on the review of those models, it was concluded that most models used for predicting truck trips utilize a linear regression. The independent variables typically include population and various categories of employment. Based on the data reviewed, a similar approach will be used to develop a general model for predicting truck trips.

MODEL DEVELOPMENT

Based on the review of the state of the practice, the most likely model for estimating truck and taxi trips is a linear regression type model. A general multiple regression model can be written as one of the following forms:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_{p-1} X_{p-1} \qquad \text{(intercept model)}$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_{p-1} X_{p-1} \qquad \text{(nonintercept model)}$$

In the above models, Y is the response or the dependent variable; and X_1 , X_2 ,......, X_p are the p independent variables. To estimate Y, regression coefficients α, β_1 , β_2 ,......, β_{p-1} must be calculated using the input data. Values of X_1 , X_2 ,......, X_p are normally given or known. Note that the nonintercept model is the one for which the regression line goes through the origin, i.e, $\alpha=0$.

Having selected the model's general form, the independent variables to be used in the model were selected. Since the model was to be incorporated into the trip generation program, TRIPCAL5, the most logical variables to be input to the trip generation models were households, basic employment, retail employment, and service employment. Selecting other variables would have required additional input and effort on the part of local agencies in projecting variables necessary for trip generation modeling.

The next step was to determine the appropriate coefficients for the model. Two approaches were considered feasible. One was to develop estimates of the coefficients based on the literature review and available research. The second was to use the origin-destination data from the surveys conducted in the 1960s and early 1970s to develop a model based on linear regression analysis.

Implementing the first approach meant assuming that each of the independent variables was significant and developing an estimate of the coefficient for use with each. The estimate of the coefficient for households was developed using data collected by the Transportation Planning and Programming Division of TxDOT from 1973 to 1975 (5). Trip generation statistics were compiled areas across the state. Part of the trip generation statistics gathered was the number of commercial truck trips to selected residential sites within the urban areas by dwelling unit type. These data were felt to be the best available and the most comprehensive, because they involved nearly 18,000 dwelling units. There were 3,856 commercial truck trips to 17,992 dwelling units which yielded an average truck trip rate of 0.214 per dwelling unit. This was selected as the coefficient for the household variable.

Developing coefficient estimates for employment categories was a more difficult task due to the lack of specific data on truck trips by employment categories. Using data from a study on transportation planning for goods and services (6) and from the Institute of Transportation Engineers' **Trip Generation**, 4th edition (7), coefficients were developed for each employment category. Table 18 shows the truck trip ends per acre for various land uses. The criterion for those selected land uses was that they were included in the definition of basic industries as used in travel demand modeling. Based on the data in Table 18, an average truck trip end rate of 10.25 trips per acre was calculated. Table 19 shows average land use densities in terms of employment for several basic type land uses. Based on the data in Table 19, an average employment density for basic land use was calculated as 15.62 employees per acre. The average truck trips per basic employee was then computed as 0.6565, and this was selected as the coefficient to use for basic employment.

Studies done in New York indicate that suburban shopping centers generated approximately 1.35 daily truck stops per 10,000 square feet of floor area (7). Data on commercial trip generation (6) were then used to compute an average number of employees per 10,000 square feet of floor area of 15.635. Assuming that these data are representative of all retail land use, an average truck trips per retail employee can be calculated as 0.086. This was selected as the coefficient for retail employment.

To develop a typical truck trip rate for service employment, it was first necessary to make a general assumption that service employment may be grouped into a single category, office building employment. While this is not strictly true in terms of definition, it was assumed for purposes of this effort that the truck trip generating characteristics of service employment would be similar to that of office building employment. Information from **Urban Transportation Planning for Goods and Services** indicated that truck generation rates for office buildings vary from 1.4 to 2.3 per day per 10,000 square feet. A rate of 2.0 is recommended. Information from **Trip Generation** indicated that typical employment densities for general office buildings ranged from 4.8 employees per 1,000 square feet of gross floor area for buildings with less than 100,000 square feet, to 4.4 employees per 1,000 square feet of gross floor area for buildings with 100,000 to 199,999 square feet, and to 3.5 per 1,000 square feet for buildings with over 200,000 square feet of gross floor area. For purposes of this analysis, a density rate of 4.4 employees per 1,000 square feet was selected. The resulting truck trip rate of 0.045 truck trips per employee was selected as the coefficient for service employment.

Table 18
Truck Trip Ends Per Acre for Selected Basic Industries

Land Use	Average
Industrial	5.90
Manufacturing	6.46
Non-Durable Manufacturing	4.10
Durable Manufacturing & Extractive	32.95
Wholesale	30.00
Commercial Wholesale & Storage	5.60
Transportation, Communications, & Utilities	1.98
Transportation & Utilities	0.45
Transportation-Warehouse	4.85

Source: Reference 2

Table 19
Basic Land Use Density

Land Use	Employees/Acre
General Light Industrial	17.4
General Heavy Industrial	7.6
Industrial Park	19.0
Manufacturing	20.1
Warehousing	14.0

Source: Reference 3

Combining the results of the individual analyses resulted in the following equation for estimating total truck trips in an urban area:

$$\begin{aligned} \text{Truck Trips} = {} & 0.214\ (\text{Total Households}) \\ & + 0.086\ (\text{Total Retail Employment}) \\ & + 0.650\ (\text{Total Basic Employment}) \\ & + 0.045\ (\text{Total Service Employment}) \end{aligned} \tag{1}$$

Unfortunately, the above relationship has no statistical measures by which to determine the level of accuracy which may be expected in applying the model.

The second approach was to develop a model using multiple regression analysis based on data from the early origin-destination surveys which included surveys of truck and taxi trips.

Table 20 shows information and data from selected urban areas where origin-destination and related transportation studies were conducted. The year in which the origin-destination or the transportation studies were conducted is also shown. The population and the employment data reported in each of the studies were not always consistent and secondary sources such as estimates from the U.S. Bureau of Census or the Texas Employment Commission were also used.

Table 20
Input Data

Urban Area	Year	Vehicle Trip	Truck Trip	Population	Dwelling Units	Basic Employment	Retail Employment	Service Employment
Abilene	1965	251882	37309	90368	36944	12194	6736	20761
Amarillo	1964	365913	49475	137969	50817	36108	29417	46610
Corpus Christi	1961	310098	48995	183055	55831	37923	16625	29490
El Paso	1970	609701	95191	322261	92704	34093	24160	50567
Harlingen-San Benito	1965	95175	15114	67653	20567	19242	11006	19951
Jefferson-Orange	1963	629926	72552	314714	100071	79893	32964	46604
Laredo	1964	101095	25580	64311	17686	2992	4545	11101
Lubbock	1964	344842	54695	152780	48441	25086	9989	33661
Midland-Odessa	1965	481032	30859	162337	54548	33257	13249	13968
San Angelo	1964	137404	23299	63438	22443	10729	4966	13054
Sherman-Denison	1968	171255	18227	62121	22698	12869	7292	11951
Tyler	1964	154885	26391	64512	21901	11996	7328	14247
Victoria	1970	131702	15793	45863	14717	5810	4514	8731
Waco	1964	255422	35119	132352	46740	16689	9284	24047
Wichita Falls	1964	259524	39032	107704	35495	13243	8325	19557

Source: Adapted from (8) to (9)

Three types of employment categories were considered based on the following Standard Industrial Classification (SIC) codes definition:

TYPE	SIC RANGE	INDUSTRY GROUP
Basic	1000-1499	Mining
	1500-1799	Construction
	2000-3999	Manufacturing
	4000-4999	Transportation, Communications, Public Utilities
	5000-5199	Wholesale Trade
Retail	5200-5999	Retail Trade
Service	6000-6799	Finance, Insurance, Real Estate
	7000-8199	Service
	8200-8299	Education Services
	8300-8999	Services
	9000-9799	Government

Because of limitations in gathering agricultural employment and the fact that agricultural employment was not used in forecasting urban transportation travel, data on agricultural employment were not included in the employment figures. Other types of employment figures such as self-employed employment were added to the service employment.

In some cases, the population and/or employment figures from secondary sources were obtained and interpolated because the information was not available for the desired year. To estimate the employment in those cases, for example, the percentage of basic, retail, and service employment were calculated based on the total employment for the year in which the data were available. These percentages were then applied to the data for the year in question.

In other cases, county and study area population and/or employment were both available for a particular urban area. In all cases, countywide figures were used to provide more consistent information.

Using the information from Table 20, 10 urban areas out of the 15 were selected to test various combinations of independent variables to determine their appropriateness for use in modeling truck and taxi trips. A multiple linear regression analysis was performed to test the combinations of variables. The results of each regression were measured using the correlation coefficient. The models producing the best results were then validated by applying them to the five urban areas not included in the calibration phase to determine how well the selected model(s) performed in estimating the number of truck and taxi trips recorded from the surveys.

Model Calibration

Several models were tested using the PROC REG (procedure regression) of the SAS statistical software package for PC. To use PROC REG,the general form (i.e., dependent and independent variables) of the model must be identified. In these preliminary tests, the following independent variables were included in the model:

- Number of dwelling units (DU)
- Total basic employment (BASIC EMP)
- Total retail employment (RETAIL EMP)
- Total service employment (SERVICE EMP)
- Total employment (TOTAL EMP)

The objective in testing different models was to select a model that produced coefficients with positive signs and provided a reasonable estimate of the number of truck and taxi trips. A regression model with no intercept term was desired although in all of the preliminary testing, regression models with and without the intercept term were considered.

Table 21 shows the partial listing of the models that was used in the testing procedure. Note that *N* in Table 21 refers to the number of observations that were used in different computer runs.

The preliminary tests produced negative coefficients for some independent variables. In some cases the resulting R^2 values for the models were low. Some additional tests were also done using transformed variables, e.g., the natural log of observed truck and taxi trips and ratios such as the observed truck and taxi trips divided by population. These generally had negative coefficients and low R^2 values.

Table 21
Model Results for Predicting Truck Trips

No	Intercept	DU	Basic EMP	Retail EMP	Service EMP	Total EMP	N	R^2
1		0.851	-0.436	-0.742	0.952		10	0.997
2	-2298	0.865	-0.427	-0.815	1.024		10	0.985
3		1.005				-0.088	10	0.961
4	6213	0.933				-0.104	10	0.831
5		0.861	-0.490	-0.408	0.782		15	0.990
6	-350	0.865	-0.491	-0.416	0.793		15	0.955
7		1.017				-0.094	15	0.961
8	5398	0.930				-0.099	15	0.851
9		0.878					10	0.960
10			1.275				10	0.975
11				2.296			10	0.858
12					1.553		10	0.952
13						0.575	10	0.901
14	5795	0.789					10	0.825
15	24481		0.711				10	0.387
16	18797			1.720			10	0.510
17	42340				1.434		10	0.776
18	13106					0.439	10	0.607
19		0.883					15	0.960
20			1.324				15	0.799
21				2.671			15	0.864
22					1.549		15	0.951
23						0.587	15	0.905
24	5316	0.790					15	0.846
25	20202		0.796				15	0.468
26	15177			1.865			15	0.576
27	4206				1.418		15	0.801
28	11168					0.456	15	0.667

Preliminary testing revealed that the number of dwelling units, basic employment, retail employment, and service employment were important independent variables based on individual t-tests. Two additional computer runs were performed after minor adjustments to the population and/or employment data in some urban areas. These adjustments (e.g., using countywide population) were necessary to reflect the growth in some areas. Following an analysis of the results shown in Table 21, additional runs were made.

In the first run, three independent variables, basic, retail, and service employment, were used. The following model was obtained:

$$TT = .443 \text{ (Basic)} - 1.784 \text{ (Retail)} + 2.085 \text{ (Service)} \qquad 2$$

For this model the coefficient estimate of the retail employment was negative, with the R^2 value of 0.96. The overall F-statistic of 61.04 was significant at the .05 level.

Two independent variables, number of dwelling units and service employment were used in the second run. The result was the following model:

$$TT = .497 \text{ (Dwelling Unit)} + .706 \text{ (Service)} \qquad 3$$

The R^2 value for the resulting model was 0.98. The coefficients for both independent variables were positive. For this model the overall F-statistic of 159.78 was significant at the .05 level.

Because of the uniqueness of the El Paso and Laredo urban areas, a third model was considered. The independent variables in this model were the same as the previous model (from the second run) except that the observations for El Paso and Laredo were deleted from the input data. The resulting model was similar to the case where all observations were used. Therefore, the two observations corresponding to El Paso and Laredo were retained in the input data, and the model with the number of dwelling units and service employment was selected as the final model. The overall performance of this model based on the F-statistic, R^2 value, and the individual t-statistics was superior compared to the other competing models.

In all of the above runs, the observed truck and taxi trips were used as the dependent variable. The degree of linear relationship between observed and estimated is determined by the simple correlation coefficient. The simple correlation coefficient is a scale between -1 to +1. The negative values indicate a negative correlation, whereas the positive values indicate positive correlation. The closer the absolute value of the simple correlation coefficient is to 1, the greater the degree of linear relation in the sample observation.

Numerical values of the observed and estimated truck and taxi trips are shown in Table 22 for the selected model e.g., Model 3. The correlation coefficient computed by comparing the observed and estimated values was 0.94 which indicated reasonably good estimates. Figure 29 shows the same values graphically.

Table 22
Observed and Estimated Values

Urban Area	Observed Value	Estimated Value by Model 3
Abilene	37309	33029
Amarillo	49475	58173
Corpus Christi	48995	48583
El Paso	95191	81800
Harlingen-San Benito	15114	24311
Jefferson-Orange Counties	72552	82667
Laredo	25580	16631
Lubbock	54695	47852
Midland-Odessa	30859	36989
San Angelo	23299	20376

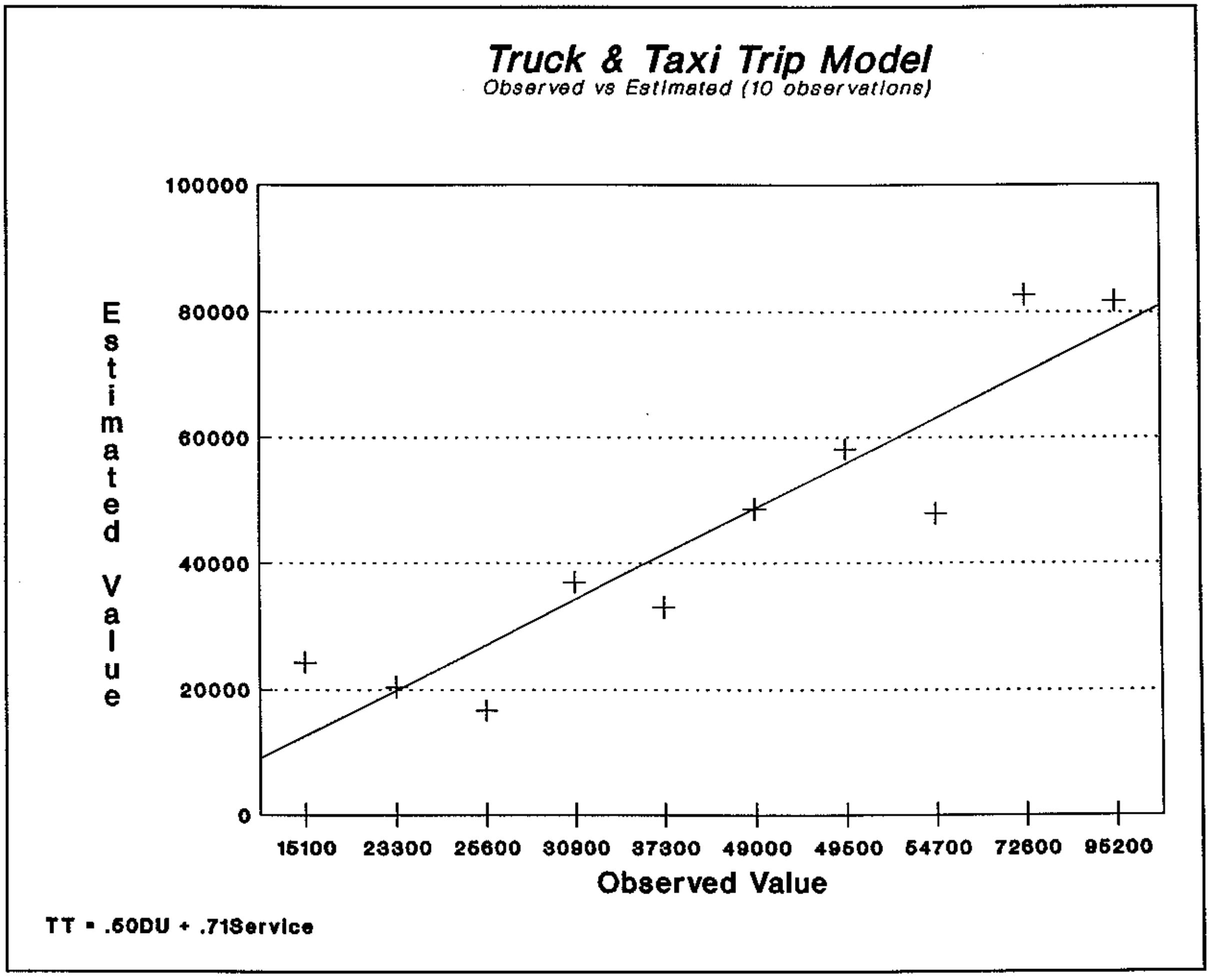


Figure 29. Calibration results using 10 observations.

Model Validation

In the validation stage, five urban areas were used to test the recommended model. Given the total number of dwelling units and the total service employment, truck and taxi trips were estimated for each urban area using Model 3. The truck and taxi trip estimates were also computed using the nonstatistically derived equation (i.e., Model 1). Both of these estimates and the actual truck and taxi trips are shown in Table 23.

Table 23
Observed and Estimated Values

Urban Area	Observed Value	Estimated by Model 3	Estimated by Model 1
Sherman-Denison	18227	19725	14387
Tyler	26391	20949	13756
Victoria	15793	13482	7707
Waco	35119	40220	22731
Wichita Falls	39032	31458	17800

Table 23 shows that Model 1 consistently underestimates the truck and taxi trips for the selected urban areas. For this model the correlation coefficient was 0.82. Estimated values by Model 3 were much closer to the observed values. For Model 3, the correlation coefficient was 0.88. This number indicates that the observed and the estimated values are highly correlated and that the existing relationship is positive. Figure 30 shows the observed and estimated truck and taxi values using Model 3.

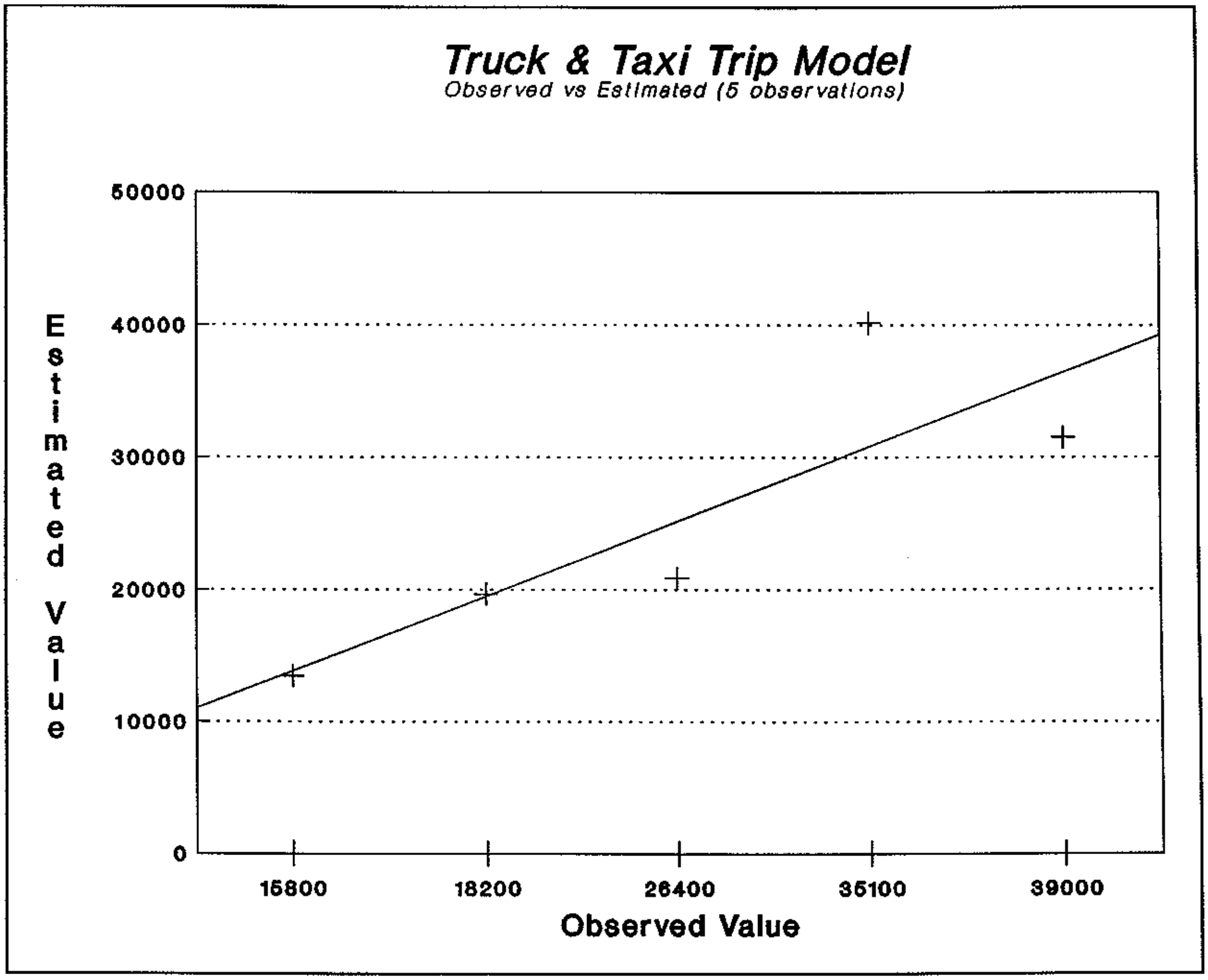


Figure 30. Validation results using 5 observations.

SUMMARY AND RECOMMENDATIONS

Based on the results in the calibration and the validation phases, the model with two independent variables, dwelling units and service employment, is recommended for use in estimating the total truck and taxi trips in an urban area. The final form of the model is recommended for use in TRIPCAL5 as a default model for estimating total truck and taxi in an urban area. The model is:

$$TT = .497 (DU) + .706 (SERVICE)$$

where:
TT = predicted truck and taxi trips
DU = total dwelling units
SERVICE = total service employment.

For this model the R^2 value was 0.98. This means that the variability in truck and taxi trips is reduced by 98 percent when total dwelling units and service employment are included in the model. This large proportionate reduction in the variability suggests that the multiple regression model is appropriate.

REFERENCES

1. Unpublished report on the development and calibration of the trip generation models for Atlanta, Georgia, based on a 1980 auto use study and 1972 home interview survey.

2. *Travel Models for Regional and Subarea Planning in the Denver Region.* Denver Regional Council of Governments, April 1987.

3. Barton-Aschman Associates, Inc. *Metropolitan Council Model Update Study.* The Metropolitan Council of the Twin Cities Region. Minneapolis-St. Paul, MN, December 1986.

4. Newhouser, R. and Hom, K.J. *Southeast Michigan Travel Forecasting Process.* Southeast Michigan Council of Governments, Detroit, MI, October 1984.

5. Transportation Planning Division. *Special Traffic Generator Study 1973-1975.* State Department of Highways and Public Transportation, 1975.

6. Christiansen, Dennis L. *Urban Transportation Planning for Goods and Services.* Texas Transportation Institute, The Texas A&M University System, prepared for the Federal Highway Administration, U.S. Department of Transportation, June 1979.

7. Technical Council Committee 6A-32. *Trip Generation, 4th Edition.* Institute of Transportation Engineers, 1987.

8. *Abilene Urban Transportation Plan, 1965-1985, Volumes 1 and 2.* Sponsored by the City of Abilene and the Texas Highway Department in cooperation with the U.S. Department of Commerce.

9. *Wichita Falls Urban Transportation Plan, 1964-1985, Volumes 1 and 2.* Sponsored by the City of Wichita Falls, Counties of Archer and Wichita, and the Texas Highway Department in cooperation with the U.S. Department of Commerce.

WORKS CITED

1. *SAS/STAT User's Guide, Release 6.03.* SAS Institute Inc., SAS Circle, Box 8000, Cary, NC, 1988.

2. *Amarillo Urban Transportation Plan, 1964-1985, Volumes 1 and 2.* Sponsored by the City of Amarillo, Counties of Potter, and Randall, and the Texas Highway Department in cooperation with the U.S. Department of Transportation.

3. *Corpus Christi Urban Transportation Plan, 1963-1964.* Sponsored by the City of Corpus Christi, County of Nueces, and the Texas Highway Department in cooperation with the U.S. Department of Commerce.

4. *El Paso Urban Transportation Plan, 1970, Volumes 5 and 6.* Sponsored by the City of El Paso, County of El Paso, and the Texas Highway Department in cooperation with the U.S. Department of Transportation.

5. *Harlingen-San Benito Urban Transportation Plan, 1965, Volumes 1 and 2.* Sponsored by the City of Harlingen, City of San Benito, County of Cameron and the Texas Highway Department in cooperation with the U.S. Department of Commerce.

6. *Jefferson-Orange Regional Transportation Study, 1963-1985, Volumes 1 and 2.* Conducted by the Texas Highway Department in cooperation with Jefferson-Orange Counties and the Bureau of Public Roads.

7. *Laredo Urban Transportation Study, 1964-1985, Volumes 1 and 2.* Sponsored by the City of Laredo, Webb County, and the Texas Highway Department in cooperation with the U.S. Department of Transportation.

8. *Lubbock Urban Transportation Plan, 1964-1985, Volumes 1 and 2.* Sponsored by the City of Lubbock, County of Lubbock and the Texas Highway Department in cooperation with the U.S. Department of Commerce.

9. *Midland-Odessa Regional Transportation Study, 1965-1985, Volumes 1 and 2.* Sponsored by the Cities of Midland and Odessa, Counties of Midland and Ector, and the Texas Highway Department in cooperation with the U.S. Department of Transportation.

10. *San Angelo Urban Transportation Study, 1964-1985, Volumes 1 and 2.* Sponsored by the City of San Angelo, Tom Green County, and the Texas Highway Department in cooperation with the U.S. Department of Commerce.

11. *Sherman-Denison Urban Transportation Study, 1968-1990, Volumes 1 and 2.* Sponsored by the Cities of Sherman and Denison, County of Grayson, and the Texas

Highway Department in cooperation with the U.S. Department of Transportation.

12. *Tyler Urban Transportation Study, 1964-1985, Volumes 1 and 2.* Sponsored by the City of Tyler, County of Smith, and the Texas Highway Department in cooperation with the U.S. Department of Commerce.

13. *Victoria Urban Transportation Study, 1970-1990, Volumes 1 and 2.* Sponsored by the City of Victoria, County of Victoria, and the Texas Highway Department in cooperation with the U.S. Department of Transportation.

14. *Waco Urban Transportation Plan, 1964-1985, Volumes 1 and 2.* Sponsored by the City of Waco and the Texas Highway Department in cooperation with the U.S. Department of Commerce.