1970

A socio-economic traffic demand prediction model based on a lumped system approach

John Edward Thompson

A SOCIO-ECONOMIC TRAFFIC DEMAND PREDICTION MODEL BASED ON A LUMPED SYSTEM APPROACH

BY

JOHN EDWARD THOMPSON, 1946-

A

THESIS

submitted to the faculty of

UNIVERSITY OF MISSOURI - ROLLA

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN CIVIL ENGINEERING

Rolla, Missouri

1970

Approved by

Frank A. Dering Jr. (advisor) John L. Best

Charles H. Adams

ABSTRACT

The objective of this investigation was to devise a method for predicting transportation system demand based on socio-economic variables of certain physical areas. This pilot study was performed utilizing traffic flow, but minor mathematical alterations could change the transportation system to urban mass transit usage demand.

An area of uniform socio-economic characteristics can be delineated and the socio-economic characteristics isolated. Each area will place a distinct demand on some existing transportation facility based on those characteristics. The physical demand can be approximated by a theoretical probability distribution. Once this internal flow generation has been approximated by the frequency distribution, it must pass through the system to the point of exit. Since flow does not occur instantaneously, it must be routed utilizing a traffic routing procedure. This procedure moves the flow through the system, altering the flow characteristics in both time and space, until the exit point is reached. The altered flow leaving the exit point represents the impact through time on the transportation facility in question.

ACKNOWLEDGMENT

The author wishes to express sincere appreciation to his advisor, Dr. Frank A. Gerig, Jr., for his dilligent assistance and time devoted toward the completion of this thesis.

Appreciation is also extended to the members of his faculty committee, Dr. John Best and Dr. Curtis Adams, for the time they so generously devoted to this project.

Especially to be noted is the contribution of my wife, Meredith. Her encouragement and patience made the completion of this thesis possible.

Finally, sincere thanks is extended toward my parents, Mr. and Mrs. Douglas R. Allen, for their guidance in my earlier years.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

I. INTRODUCTION

A. Orientation

Before Abraham made his trek from Ur to Palestine, man had carved out paths over which he could satisfy his need to move. Since that time, in ever increasing numbers and over ever increasing distances, man has developed means of travel for persons and for things. The forces which induced man to make these movements have been widely varying but have perhaps stemmed from man's restless nature. Abraham [redacted] made his move because of divine inspiration. Two thousand years later, the Roman Senate initiated the construction of the great road network from Rome to the far reaches of the empire in order to maintain control over these vast reaches. The camel trails across the Middle East provided spices and silks for Europeans not accustomed to such fineries. The oceanic trade routes served as the catalyst for the advent of Venice as the trading center of Europe. Completion of the intercontinental railroad at Promontory Point, Utah, in 1869 signaled the end of a project designed to link the eastern and western parts of the United States.

Today man travels for an increased variety of reasons. He goes to work, to recreation, to shop for food and clothing, etc. He uses means of transportation which have continually improved in comfort and in speed during the last few decades. As the means of obtaining these improved methods of travel become more available, the desire for travel increases correspondingly. This increased desire for travel is a reflection of man's changing socio-economic status in his environment.

The provision of facilities to meet these demands has long been a serious problem. Not only is the construction of the facilities a time-consuming and wealth-consuming problem, but also the prediction of the demand and the location of the proposed transportation facilities is a process often plagued by inaccurate results. This prediction process involves the interaction of transportation with the spacial distribution of activities, consumer preference for transportation in terms of willingness of pay, and a knowledge of the impact of transportation on land use as well as the socio-economic characteristics of the population. Historically, the system of predicting demand involves the prediction of numbers of trips generated, then distribution between arbitrarily selected zones or locations, the assignment of these trips to available and planned routes, and the selection of a division of the demand between the various modes provided or projected (18). Often the interzonal transfers are calculated using a gravity model (17). The gravity model is formulated as

$$T_{ij}^{P} = T_i^{P} \frac{K_{ij}^{P} A_j^{P}/t_{ij}^{p}}{\sum_{k=1}^{n} K_{ik}^{P} A_k^{P}/t_{ik}^{p}} \qquad \text{for all } k \neq i \qquad 1.1$$

where

T_{ij}^{P} = interzonal trips from zone i to zone j for purpose P,

T_i^{P} = interzonal trips originating in zone i for purpose P

$= \sum_{k=1}^{n} T_{ik}^{P}$ for all $k \neq i$,

A_k^{P} = interzonal trips ending in zone j for purpose P, railroad had been predicted based upon the number of persons traveling by land between the east and west coast in 1865, there would have been no probable requirement for the intercontinental railroad. We can see from history that when the facility was provided, the demand increased enormously. Ideally, estimation of demand would consider each aspect in terms of a supply and a demand. Supply defines the relationship between travel costs or impedance and flow. Demand describes the relationship between the volume demands and travel costs. The resulting supply-demand relationships can then be used to determine equilibrium flows on each individual system. It should be clear that socio-economic factors of several types affect the supply-demand relationship and these predictions (20). A need, then, exists to develop a method which would be descriptive of the supply-demand relationship and would incorporate a full range of the necessary socio-economic factors.

Upon development of this method, a technique will have been presented that relates transportation to a supply-demand relationship. Although this initial pilot study will be primarily concerned with vehicular movement, it must be remembered that only a few minor mathematical alterations would make this method applicable to urban mass transportation usage demand.

The application of the methods developed by this project will be very useful. Since transportation demand will be heavily dependent on socio-economic and physical factors, prediction of this demand could also be applied to areas still in the planning stage. All needed data could be obtained from the area developers' plans as well as past records of the class of people moving into certain types of neighborhoods. This would be invaluable in predicting, say, a planned subdivision's effects on existing or planned transportation facilities.

The University of Missouri - Rolla is conducting studies in cooperation with the Urban Mass Transportation Administration, U.S. Department of Transportation to develop an improved method of predicting the demand for movement which will consider socio-economic factors and the dynamic influences of external stimuli. The author completed this study under the direction of the University of Missouri - Rolla Transportation Institute. This Institute embraces members from Civil Engineering, Electrical Engineering, Computer Science, Mathematics, Social Sciences, Humanities, and Urban Planning Faculties. Invaluable aid was also provided by faculty and graduate students of the Department of Sociology and Anthropology, University of Missouri - St. Louis.

B. Operation Techniques

Assume a series of essentially homogeneous areas. By homogeneous areas, it is implied that the homogeneity will be based on several socio-economic factors. These factors would include demographic, social, and functional variables. These variables must be selected with the primary concern being their possible effect upon traffic generation for their particular area.

There will also be a need for the collection of a series of physical variables for each area. These physical variables must describe conditions peculiar to each area that are causes of traffic slowdown and stoppage.

The peak demand time for each area must be determined and isolated.

During this peak time, data must be taken over small time increments in all areas to determine the frequencies of individuals leaving their homes to start a journey contributing to the peak volume period.

This departure frequency can be estimated with a theoretical probability distribution. Examination of similar departure frequencies by Wattleworth (2) show a pronounced peak, but not always a symmetrical distribution. Three probability distributions appear to offer the possibility of describing the observed data distributions. These are the beta distribution (Pearson Type I), the gamma distribution (Pearson Type III), and the normal distribution. Of these distributions, the beta and gamma allow the greatest flexibility. The peaks of these two distributions may be altered quite readily in the horizontal and/or vertical directions with manipulation of the distribution parameters. Figures 1.1 and 1.2 illustrate the beta distribution and gamma distribution, respectively, with changing constants. These individual parameters could be related to the socio-economic variables of the regions by utilizing a multiple regression technique. This would allow the traffic inflow prediction based only upon the area's socio-economic factors. A chi-square goodness of fit test will determine the most applicable distribution.

The traffic outflow prediction can be obtained through the use of the traffic routing procedure. This procedure takes into account two variables, x and k, which describe volume and time distortions of the traffic flow as it passes through the area. These variables could be determined by equating them to the physical characteristics of the area by again utilizing a multiple regression technique.

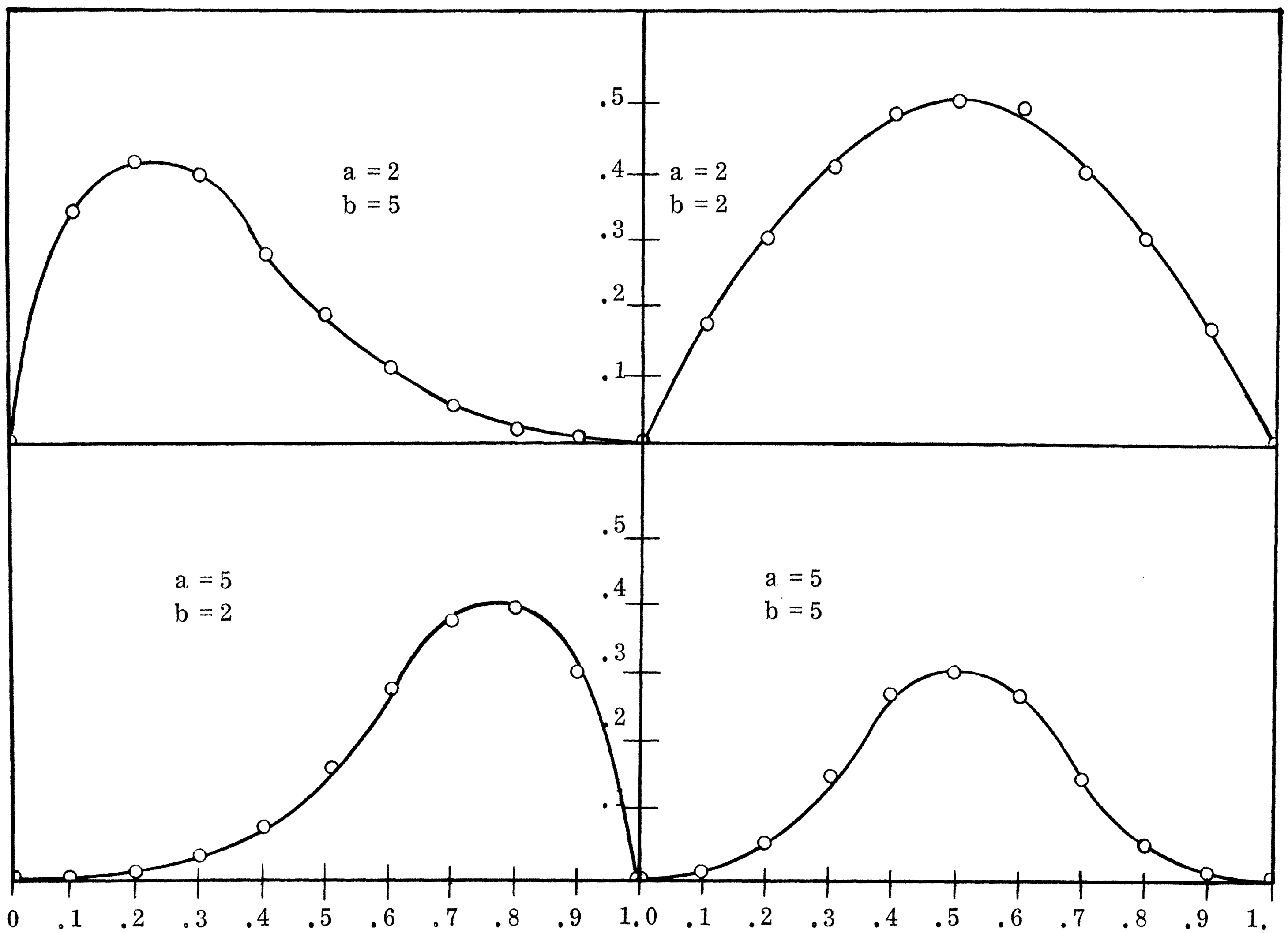

Figure 1.1. Beta Distribution with Changing Constants.

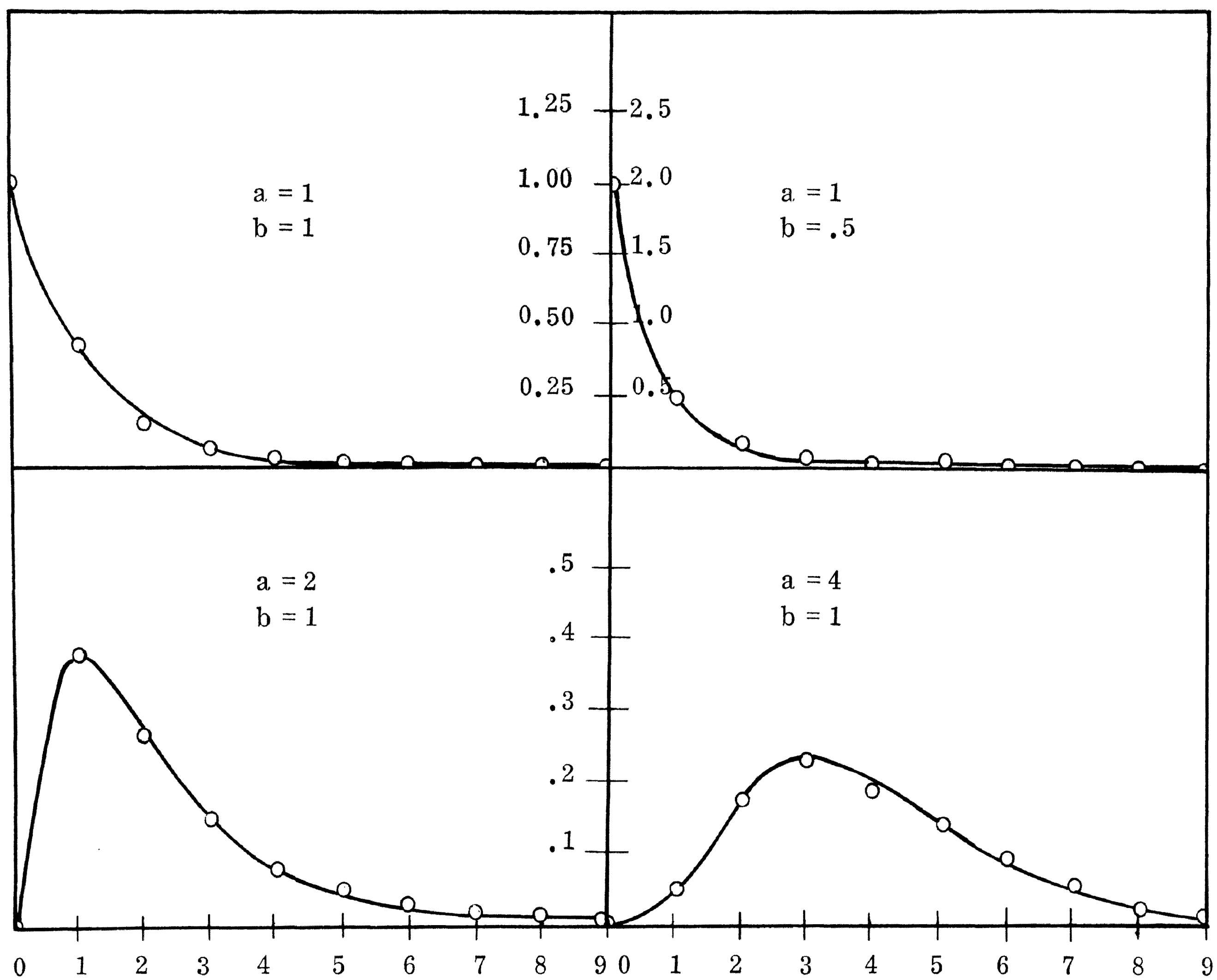

Figure 1.2. Gamma Distribution with Changing Constants.

It is now possible to find the inflow function from the area's socio-economic variables and the delay variables from the area's physical characteristics. The determination of the outflow characteristic is now a straightforward method utilizing the routing technique.

C. Objectives

As a part of the interdisciplinary studies of the Transportation Institute, this research has as its objectives:

1. Developing computer programs which will (a) perform the calculations involved in determining the most applicable of the theoretical probability distributions tested, (b) perform multiple regression analysis, and (c) model the actual traffic routing procedure and allied functions.
2. Testing the feasibility of the proposed method using data from census tracts and other studies.

II. REVIEW OF LITERATURE

A. Transportation System Demand Based on Socio-Economic Variables

The inability of current traffic prediction models to accurately predict demand is obvious. Owens (20) states:

> "The ... weakness of the current approach to urban transportation problems is that most of the steps taken thus far have been concentrated on the supply side of the problem. Little effort has been made to analyze the factors underlying transportation demand, with a view to bringing supply and demand into better balance. Yet the fact that demand continues to outrun supply regardless of efforts to increase transport capacity indicates the need for directing more attention to this neglected side of the problem."

In speaking on urban planning and development, Riley (21) states that economic, social, and legal factors are to be considered. Included among these factors are type of community, economy of area, population statistics, land use, and community facilities.

Levinson, et. al. (22) described a method of relating generated trips to ten socio-economic variables by utilizing a multiple regression technique. Cited correlation coefficients vary from a low of .30 to a high of .87. The average value of these cited correlation coefficients was .68.

There is no evidence of studies relating socio-economic factors to trip generation or frequency distribution of demand on a comprehensive basis.

B. The Application of Curve Fitting Procedures

A proposed model can be verified only by comparing its predictions with observed data. In the case of probability distributions models, the comparison is usually between the shape of the distribution and the shape of the data.

Varied applications of several distributions exist in transportation systems. Gerig (4) described encroachment onto a freeway from an acceleration lane with a beta distribution. Drew (11) applied a Poisson distribution to observed traffic flow exceeding an arbitrary speed. Benjamin and Cornell (6) illustrate the use of extreme value distributions in merging traffic problems. Additional examples may be readily located in traffic engineering textbooks.

C. Hydrologic Simulation of Traffic Events

One of the crucial parts of this investigation was the development of a traffic routing procedure. This routing procedure will be patterned after the Muskingum routing procedure. No literature was found which directly related to such a modelling process in a systems manner.

The Muskingum routing method (8, 9) routes a flood wave through a section of a stream to the section's exit point. This incoming flood wave, through time, forms the inflow hydrograph. In natural streams, the channel resistance and storage capacity are high; consequently, the flood-wave characteristics of the inflow hydrograph will be considerably changed as it passes along the stream. By utilizing the fact that a changing rate of storage exists, the inflow hydrograph is changed as it passes along the stream and forms the outflow hydrograph at the point of exit.

III. THEORETICAL DEVELOPMENT

In order to fully describe the distribution of the internal traffic generated, a usable probability distribution must be defined and shown to describe the phenomena. The probability distributions under consideration are the beta, gamma and normal distributions. As will be shown on the following pages, the beta and gamma distributions have numerical limitations placed on their random variables. Specifically, the beta distribution is a set of $[0, 1]$ numbers, and the gamma distribution a set of positively skewed $[0, +\infty]$ numbers. This entails shifting both distributions to accomodate these restrictions.

Once the inflow phenomena has been approximated by an appropriate probability distribution, a method must be developed to transport the flow to its exit point. This method will be known as the traffic routing procedure. The implementation of this procedure is crucial to the development of this investigation since this routing technique will determine the time and spacial distortions of the flow as it passes through an area to the outflow point.

A. The Beta Distribution (4, 10, 11, 12)

A random variable z is said to have a beta distribution if its density is given by

$$f(z) = \frac{z^{a-1}(1-z)^{b-1}}{\beta(a,b)} \qquad 0 \le z \le 1 \qquad 3.1$$

where $\beta(a, b)$ is called the beta function of a and b and is defined in terms of the gamma function by

$$\beta(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)} \qquad 3.2$$

By theoretical definition, the gamma function can be expressed as

$$\Gamma(n) = \int_0^{\infty} z^{n-1} e^{-z} \, dz \qquad 3.3$$

From the stated limits of z in Equation 3.1, it can be seen that traffic data must be translated to a set of [0, 1] numbers. Let c = variable range and let c_1 = minimum variable value.

The area between f(z) and the z axis is unity, and letting y = cz, we can translate the variable z to a unit length variable y.

$$\int_0^1 f(z)\,dz = \int_0^z [\beta(a,b)]^{-1} \left(\frac{y}{c}\right)^{a-1} \left(1 - \frac{y}{c}\right)^{b-1} c^{-1}\,dy \qquad 0 \le y \le c \qquad 3.4$$

Now to shift the distribution, a second variable, x, is introduced. Let $x = y + c_1$. Therefore

$$\int_0^c f(y)\,dy = \int_{c_1}^{c_1+c} [\beta(a,b)]^{-1} \left(\frac{x-c_1}{c}\right)^{a-1} \left(\frac{c-x+c_1}{c}\right)^{b-1} c^{-1}\,dx$$

$$c_1 \le x \le (c_1 + c) \qquad 3.5$$

Let $c_2 = c_1 + c$ = the maximum variable value. Rearranging Equation 3.5, we obtain the generalized form of the beta distribution

$$f(x) = \frac{(x - c_1)^{a-1} (c_2 - x)^{b-1}}{\beta(a,b)\,(c_2 - c_1)^{a+b-1}} \qquad c_1 \le x \le c_2 \qquad 3.6$$

In order to fit observed data to the theoretical distribution of Equation 3.6, the value of the parameters a and b must first be known. The mean and variance of the observed data may be found in terms of a and b, and the solution of the resulting two equations in two unknowns is relatively straightforward.

The mean and variance of the observed data can be found by taking moments about c_1. The generalized k^{th} moment about c_1 is

$$\mu_k^{c_1} = \int_{c_1}^{c_2} (x - c_1)^k f(x)\, dx \tag{3.7}$$

$$= \frac{1}{\beta(a,b)} \int_{c_1}^{c_2} \frac{(x - c_1)^k (x - c_1)^{a-1} (c_2 - x)^{b-1}}{(c_2 - c_1)^{a+b-1}}\, dx \tag{3.8}$$

$$= \frac{1}{\beta(a,b)} \int_{c_1}^{c_2} \frac{(x - c_1)^{a+k-1} (c_2 - x)^{b-1}}{(c_2 - c_1)^{a+b-1}}\, dx \tag{3.9}$$

Converting the expression under the integral to a form of the beta distribution by multiplying this expression by appropriate constants, we have

$$\mu_k^{c_1} = \frac{\dfrac{(c_2 - c_1)^k}{\beta(a,b)} \displaystyle\int_{c_1}^{c_2} [\beta(a + k, b)]^{-1} \frac{(x - c_1)^{a+k-1} (c_2 - x)^{b-1}}{(c_2 - c_1)^{a+k+b-1}}\, x\,dx}{\dfrac{1}{\beta(a + k, b)}} \tag{3.10}$$

Since the expression under the integral is now a generalized form of the beta distribution, the evaluation of this integral between limits c_1 and c_2 must be equal to 1. Therefore

$$\mu_k^{c_1} = \frac{\beta(a + k, b)}{\beta(a,b)} (c_2 - c_1)^k \tag{3.11}$$

Substituting the gamma functions in place of the beta functions, we have

$$\mu_k^{c_1} = (c_2 - c_1)^k \frac{\Gamma(a + k)\,\Gamma(a + b)}{\Gamma(a + k + b)\,\Gamma(a)} \tag{3.12}$$

The mean, by definition, is the first moment. By letting k = 1 in Equation 3.12, we obtain

$$\mu_1^{c_1} = \frac{\Gamma(a+1)\,\Gamma(a+b)}{\Gamma(a+b+1)\,\Gamma(a)} (c_2 - c_1)^1 = \frac{a}{a+b} (c_2 - c_1) \qquad 3.13$$

The mean about c, therefore, is

$$\mu = \mu_1^{c_1} + c_1 = \frac{ac_2 + bc_1}{a+b} \qquad 3.14$$

Similarly, by definition, the variance is the second moment. By letting k = 2 in Equation 3.12, we obtain

$$\mu_2^{c_2} = (c_2 - c_1)^2 \frac{\Gamma(a+2)\,\Gamma(a+b)}{\Gamma(a+b+2)\,\Gamma(a)} = (c_2 - c_1)^2 \frac{a(a+1)}{(a+b)(a+b+1)} \qquad 3.15$$

By statistical definition

$$\sigma^2 = \mu_2^{c_1} - (\mu_1^{c_1})^2 \qquad 3.16$$

Therefore

$$\sigma^2 = (c_2 - c_1)^2 \frac{a(a+1)}{(a+b)(a+b+1)} - (c_2 - c_1)^2 \frac{a^2}{(a+b)^2} \qquad 3.17$$

The variance about c_1, therefore, is

$$\sigma^2 = \frac{(c_2 - c_1)^2\, ab}{(a+b)^2 (a+b+1)} \qquad 3.18$$

B. The Gamma Distribution (4, 10, 11, 12)

A random variable y is said to have a gamma distribution if its density is given by

$$f(y) = \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by} \qquad 0 \leq y \leq \infty \qquad 3.19$$

where Γ is the gamma function described by Equation 3.3. The limits for the random variable are that it is equal to or greater than 0. Again the observed data must be shifted in order to obey these limits.

Let c_1 = the initial variable value. We can now translate variable y with the aid of variable t where $t = y + c_1$. Substituting t into Equation 3.19,

$$\int_0^\infty f(y)\,dy = \int_{c_1}^\infty \frac{b^a}{\Gamma(a)} (t - c_1)^{a-1} e^{-b(t-c_1)}\,dt \qquad 3.20$$

giving

$$f(t) = \frac{b^a}{\Gamma(a)} (t - c_1)^{a-1} e^{-b(t-c_1)} \qquad c_1 \le t \le \infty \qquad 3.21$$

The distribution parameters a and b must be calculated before observed data can be compared to Equation 3.21. The mean and variance of the observed data may be found and expressed in terms of a and b. Once this is completed, the solution of two equations in two unknowns can be solved readily.

The mean and variance can be found by taking moments. The k^{th} moment about c_1 is

$$\mu_k^{c_1} = \int_{c_1}^\infty \frac{b^a}{\Gamma(a)} (t - c_1)^k (t - c_1)^{a-1} e^{-b(t-c_1)}\,dt \qquad 3.22$$

$$= \frac{b^a}{\Gamma(a)} \int_{c_1}^\infty (t - c_1)^{k+a-1} e^{-b(t-c_1)}\,dt \qquad 3.23$$

Converting the expression under the integral to a form of the gamma distribution by multiplying this expression by appropriate constants, we have

$$\mu_k^{c_1} = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a+k)}{b^{(a+k)}} \int_{c_1}^\infty \frac{b^{(a+k)}}{\Gamma(a+k)} (t - c_1)^{k+a-1} e^{-b(t-c_1)}\,dt \qquad 3.24$$

Since the expression under the integral is now a generalized form of the gamma distribution, the evaluation of this integral between limits c_1 and infinity must be equal to 1. Therefore,

$$\mu_k^{c_1} = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a+k)}{b^{a+k}} \qquad 3.25$$

The mean, by definition, is the first moment. Letting k = 1 in Equation 3.25, we obtain

$$\mu_1^{c_1} = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a+1)}{b^{a+1}} = \frac{a}{b} \qquad 3.26$$

The mean about the origin, therefore, is

$$\mu = \mu_1^{c_1} + c_1 = \frac{a}{b} + c_1 \qquad 3.27$$

The variance, by definition, is the second moment. Letting k = 2 in Equation 3.25, we obtain

$$\mu_2^{c_1} = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a+2)}{b^{a+2}} = \frac{a^2+a}{b^2} \qquad 3.28$$

Using the relationship expressed in Equation 3.16, the variance is

$$\sigma^2 = \frac{a^2+a}{b^2} - \left(\frac{a}{b}\right)^2 = \frac{a}{b^2} \qquad 3.29$$

C. The Normal Distribution (10, 13)

A random variable x is said to have a normal distribution if its density is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty \le x \le \infty \qquad 3.30$$

From Equation 3.30, it can be seen that the variable is bounded only by $\pm\infty$. In this case, no variable transformations are necessary since the area under the density curve is 1 in all cases. The mean and variance of the normal distribution of Equation 3.30 can be found directly from observed data.

A special case of the normal distribution occurs when μ equals 0 and σ^2 equals 1. This condition is called the unit normal distribution.

Since normal probability tables are derived from the unit normal distribution, it is advisable to transform the general normal distribution into the unit normal distribution. In order to accomplish this task, the x variables will be standardized to the unit normal z variables by

$$z = \frac{x - \mu}{\sigma} \qquad 3.31$$

This variable has 0 mean and unit variance. The unit normal density function may then be described as

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty \leq z \leq \infty \qquad 3.32$$

D. The Traffic Routing Procedure (3, 8, 9)

When a traffic phenomena passes through a system, or traffic reach, the inflow and outflow traffic graphs at the initial and final points, respectively, of the reach appear as in Figure 3.1. As shown in this figure, the peak volumes are attenuated and delayed. These differences are due to internal friction points in the reach, such as delays, traffic lights, accidents, etc. This can be referred to as the reach storage. In a system, the inflow minus the outflow

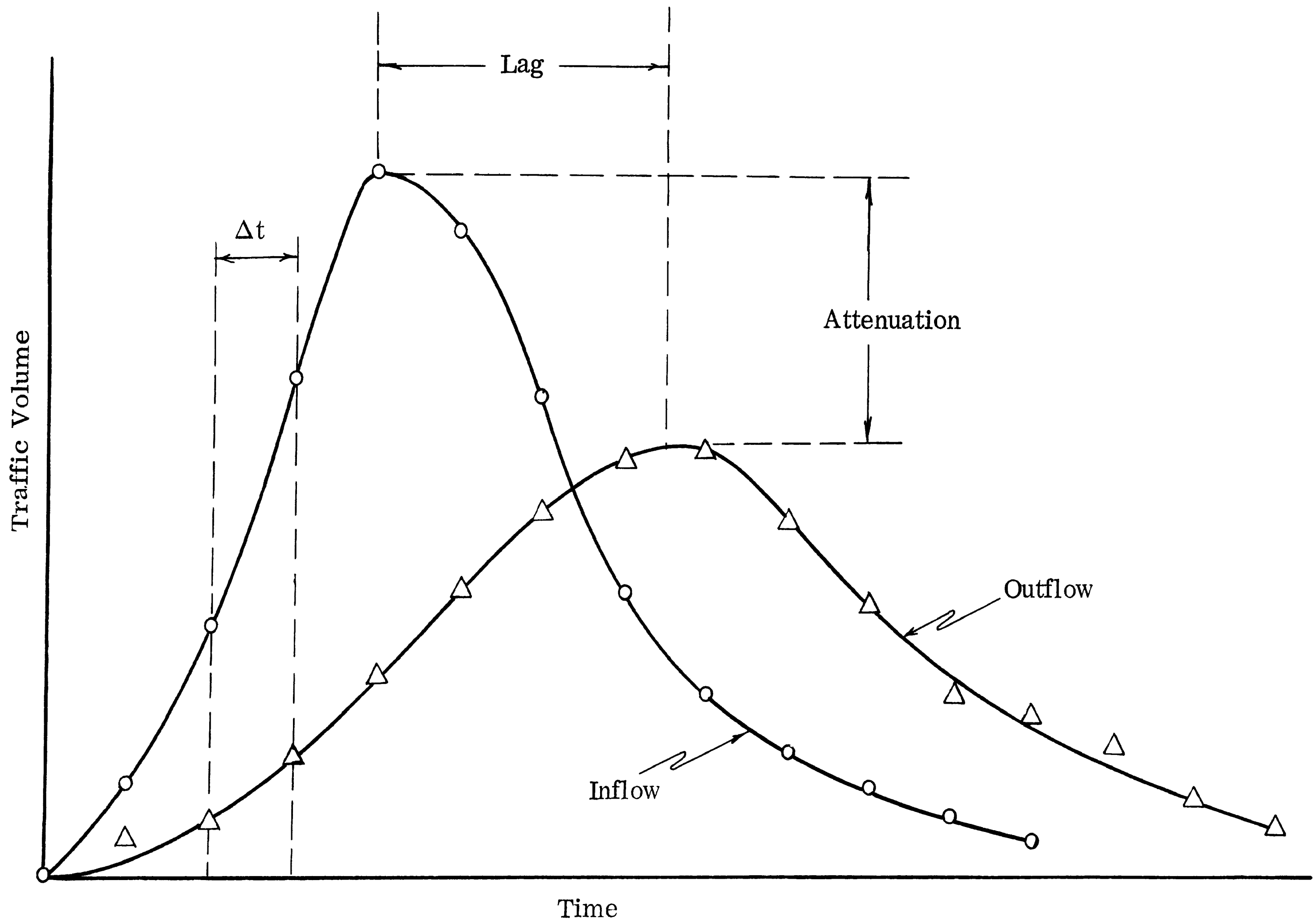

Figure 3.1. Traffic Inflow and Outflow Relationship.

is the change of storage with time, or

$$\frac{\Delta S}{\Delta t} = I - O \qquad 3.33$$

where

$\frac{\Delta S}{\Delta t}$ = the change of storage during the time period Δt,

I = the average inflow during Δt,

O = the average outflow during Δt.

$\Delta S/\Delta t$ is positive if storage is increasing and negative if storage is decreasing.

The total storage in the reach is a function of inbound volume and outbound volume. These two volumes will usually vary, so let x = the dimensionless factor that will assign relative weights to the two volumes.

The total storage can now be expressed as

$$S = xI + (1 - x)O \qquad 3.34$$

where S = total storage.

If storage were controlled only at the last point on the reach, x would be equal to 0. If control were equally divided by initial and final points, x would be equal to .5.

Storage does not occur instantaneously. The traffic must route itself through the area to the storage point. This time is called lag time.

Let k = the time factor describing the lag time. We can now express our total storage equation as

$$\frac{S}{k} = xI + (1 - x)O \qquad 3.35$$

or

$$S = k [xI + (1 - x) O] \tag{3.36}$$

Applying this equation to each incremental time slice Δt, we obtain a storage increment

$$\Delta S = k [x (\frac{I_1 + I_2}{2}) + (1 - x) (\frac{O_2 + O_1}{2})] \tag{3.37}$$

From Equation 3.33 we also know

$$\frac{I_1 + I_2}{2} + \frac{O_1 + O_2}{2} = \frac{\Delta S}{\Delta t} \tag{3.38}$$

Combining Equation 3.37 and Equation 3.38, we can express both relationships in terms of k.

$$k = \frac{\Delta t (\frac{I_2 + I_1}{2} - \frac{O_2 + O_1}{2})}{x(I_2 + I_1) + (1 - x) (O_2 - O_1)} \tag{3.39}$$

Separating Equation 3.39 to isolate O_2, we obtain

$$O_2 = I_2(\frac{.5\Delta t - kx}{k - kx + .5\Delta t}) + (\frac{kx + .5\Delta t}{k - kx + .5\Delta t}) I_1 + O_1(\frac{k - kx - .5\Delta t}{k - kx + .5\Delta t}) \tag{3.40}$$

Simplifying, we obtain

$$O_2 = c_0 I_2 + c_1 I_1 + c_2 O_1 \tag{3.41}$$

where

$$c_0 = \frac{-(kx - .5\Delta t)}{k - kx + .5\Delta t} \tag{3.42}$$

$$c_1 = \frac{kx + .5\Delta t}{k - kx + .5\Delta t} \tag{3.43}$$

$$c_2 = \frac{k - kx - .5\Delta t}{k - kx + .5\Delta t} \tag{3.44}$$

Equations 3.41, 3.42, 3.43, and 3.44 provide a convenient method of solution for each time slice Δt along the entire time period of observation. A solution must now be found to determine k and x. Examining Equation 3.36, we see that it resembles a linear equation

$$y = ma + b \tag{3.45}$$

with

$y = s,$

$m = k,$

$a = [\, xI + (1 - x)O\,],$

$b = 0.$

If flow data of previous traffic conditions is available, k and x can be determined by plotting x versus $[\, xI + (1 - x)O]$. An initial value of x is selected, and is then successively decreased until the resulting plotted configuration approximates a straight line as shown in Figure 3.2. The slope of this approximate straight line is therefore k.

With k, x, and t established, values of c_o, c_1, and c_2 can be computed. The traffic routing operation is then simply a solution of Equation 3.41 with the O_2 of one time slice becoming the O_1 of the succeeding period.

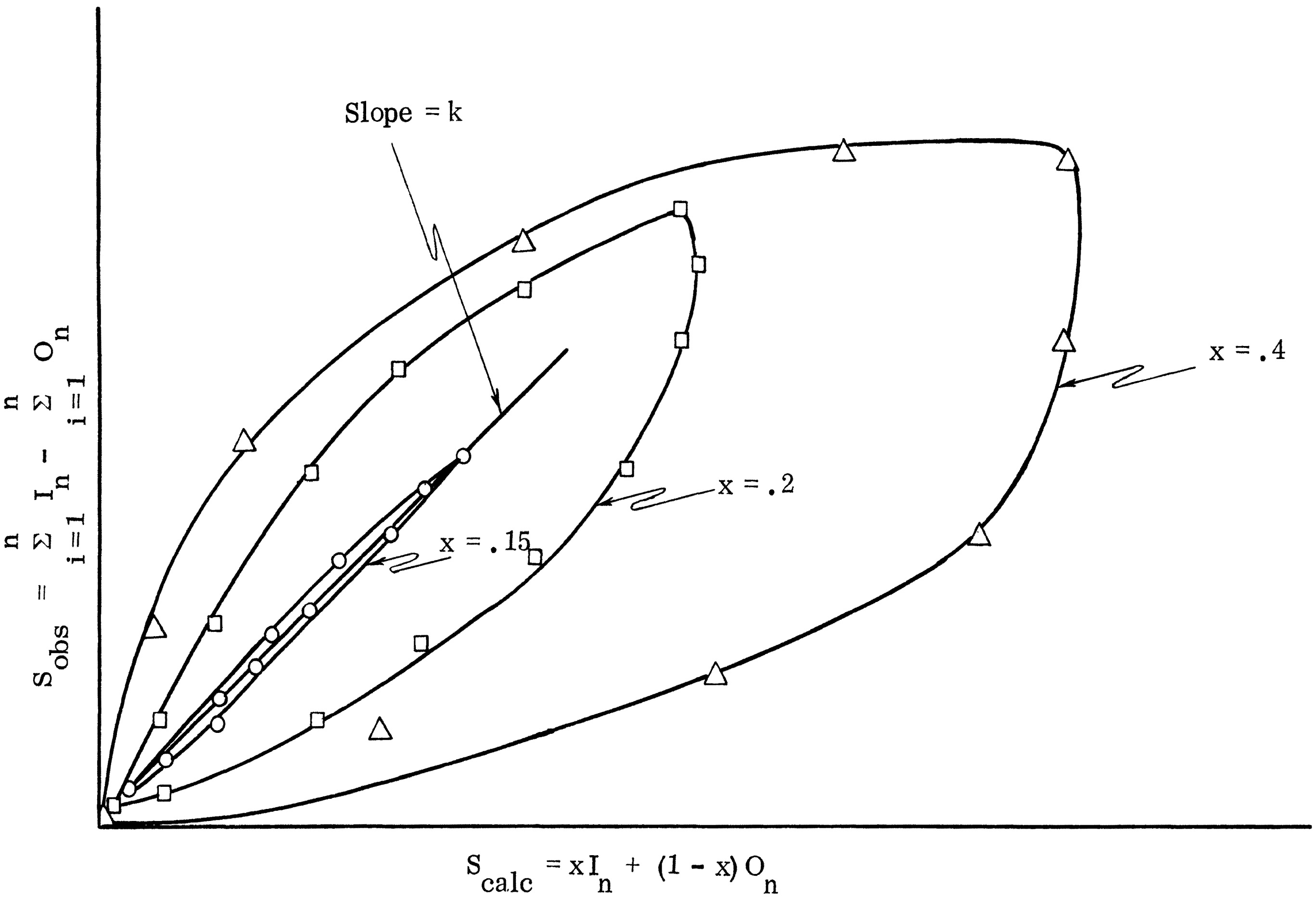

Figure 3.2. The Relationship between Observed and Calculated Storage for Varying Values of x.

IV. DATA COLLECTION

This chapter will define the type of data collected and where this collection occurred. All data, as well as the computer programs used to manipulate this data, are available at the Transportation Institute, University of Missouri - Rolla, Rolla, Missouri.

A. Selection of Areas

The individual areas selected for this study were in the city of St. Louis and St. Louis County, Missouri. The United States Department of Commerce, Bureau of the Census, divides St. Louis and St. Louis County into 225 individual census tracts as shown in Figure 4.1 (5). Tract boundaries were established co-operatively by a local committee and the Bureau of the Census, and were generally designed to be relatively uniform with respect to population characteristics, economic status, and living conditions. Of these 225 areas, 40 were selected for use in this study. Factors governing the selection were internal homogeneity, travel patterns of exit trips, available traffic flow data and number of major routes of exit.

The areas selected were areas 1C, 1D, 2A, 3B, 5C, 5D, 6A, 9A, 21C, 22B, 24A, 24B, 24C, 26B, 102, 103, 104, 111, 113, 115, 118, 138, 145, 149, 150, 154, 159, 162, 166, 182, 183, 184, 188, 192, 197, 199, 202, 206, 207, 211.

B. Selection of Socio-Economic Variables

The following variables, selected with the aid of Drs. Sarah and Solomon Sutker, Department of Sociology and Anthropology, University of Missouri - St. Louis, were felt to be the most significant in identitying with the traffic

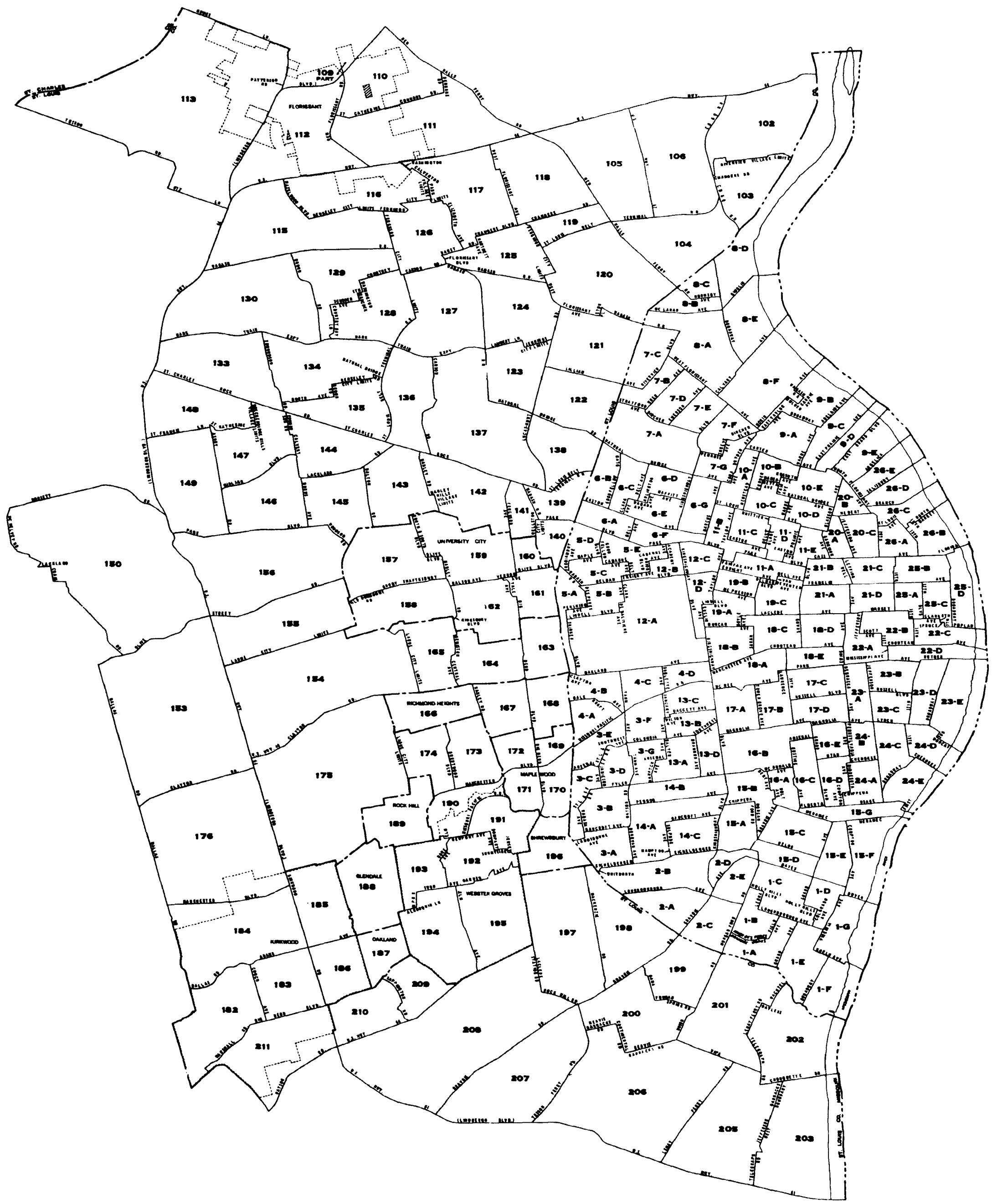

Figure 4.1. Census Tracts in St. Louis and St. Louis County, Missouri.

generation ability of the regions from which they were obtained. The actual data used in this pilot study was obtained from latest available data furnished by the Bureau of the Census. In this project, the more up-to-date data from the 1970 census was not available; consequently, the 1960 census data was utilized.

1. Total Population. These persons include any lodgers, foster children, wards and resident employees who share the living quarters of the household head and his or her family, if any.

2. The Ratio of White to Non-white Population. This ratio supplies an indication of the racial breakdown of an area since different races have varying effects on traffic generation. It is assumed that at least one white or one non-white person exists in each area.

3. Population per Household. This ratio gives an indication of the density of the population as well as family size.

4. Number of Married Couples. A married couple is defined as a husband and a wife living together.

5. Median Schooling of Population in Years. Elementary school includes grades 1 to 8, and high school includes grades 9 to 12. College includes junior or community colleges, and graduate or professional schools.

6. Median Family Income. This is the median of the sum of amounts reported separately for wage and salary income, self-employment income, and other income. It represents the amount received before deductions of any sort.

7. Cars Available. This represents the total number of cars available

in an area. This statistic is based solely on the theoretical cars available based on an area's median income.

8. Median Population Age. This represents the median age of all persons living within the area in question.

9. Ratio of Total Individuals Over 21 Years of Age to Total Individuals Under 21 Years of Age. This ratio gives an indication of the balance of minors to adults.

10. Total Labor Force. This consists of all civilians 14 years of age and over who receive wages for services rendered.

11. Total Unemployed. This consists of all civilians 14 years of age and over and not at work but looking for work. This group includes individuals laid off from their regular occupation.

12. Ratio of Male to Female Workers. This gives an indication of the total number of families with both husband and wife working.

13. Ratio of Workers Using Automobiles for the Trip to Work to Total Workers. This ratio indicates the relationship of private automobiles, as well as carpool use, for work trip purposes and total number of workers.

14. Median Rooms per Household. This gives an indication of the economic as well as density characteristics of the area.

15. Median Value of Owned Homes. This is the median of the monetary sale value of the homes on the open market.

16. Median Amount of Rent Paid. This is the median value of monetary contract rent paid by the renter.

17. Ratio of Renters to Owners. This is an indication of the relation-

ships between renters and owners in an area.

C. Selection of Physical Variables

These variables are the probable causes of time and volume dispersion as a traffic platoon moves through an area. Teams from the Transportation Institute, University of Missouri - Rolla, collected data on the roadway which served as the major collector for the traffic generated by the area in question. During the selection of areas to be used, one of the primary considerations involved was the presence of one main exit route.

1. The size of the area in square miles.

2. The length of the main road in miles.

3. The number of lanes available on the main road for all directions of traffic flow.

4. The mean lane width of the main road.

5. The number of stop signs and traffic lights along the main road.

6. The number of side streets and pedestrian crossings with right of way along the main street.

7. The speed limit along the main road.

8. A rating factor of the physical street condition where:

Rating Factor	Physical Street Condition
1	Very Poor
2	Poor
3	Average
4	Above Average
5	Very Good

9. A parking condition rating factor where:

Rating Factor	Parking Condition
1	One Side Parallel
2	One Side Slant
3	Both Sides Parallel
4	Both Sides Slant
5	Other (Including None)

10. A traffic flow rating factor where:

Rating Factor	Traffic Flow
1	Two Way
2	One Way

11. A driving comfort rating factor. This is an indication of the driving comfort felt by the driver while passing along the area's main road.

Rating Factor	Physical Comfort
1	Very Poor
2	Poor
3	Average
4	Above Average
5	Excellent

D. Traffic Flow Data

This data was obtained with the aid of the Missouri State Highway Department as well as studies performed by several independent consulting firms. Inflow data, or the number of vehicles leaving the home, was obtained between the hours of 7:00 AM and 8:30 AM at 5 minute intervals. Outflow data, or the number of people leaving the individual area at the major exit was obtained between the hours of 7:00 AM and 9:30 AM at 5 minute intervals. In each of the above cases, a baseflow was subtracted from all volumes. This baseflow represents the through traffic not generated by the area.

V. DISCUSSION OF RESULTS

A. Fitting the Beta, Gamma, and Normal Distributions

In order to select the most applicable probability distribution for the inflow data, computer programs were written which had as input the total number of observations as well as magnitude of those observations at five minute intervals through the entire data collection time period. In the programs, the observed data is grouped into intervals determined by sample size. The observed values falling into each time interval are determined. Knowing the mean and variance of the observed data, the distribution parameters can be found. Knowing these parameters, the theoretical number of observations falling into each time interval could be found by integrating the appropriate distribution. If less than five observed or expected events occurred in any one interval, the intervals were rearranged until none contained less than five observations. A chi-square test at a 99% confidence level then estimated if a fit had been achieved. It must be kept in mind that goodness of fit tests are only indicators, not determinants. As stated by Benjamin and Cornell (6), "... goodness of fit hypothesis tests are not designed to choose among contending models, but rather suggest that a given proposed model should or should not be retained." The adoption or rejection of various models could also be based on various other factors, such as cost, time, etc., as determined by sound engineering judgement.

Of the forty sets of data in these analyses, 15% were accepted at a 99% confidence level as fitting a gamma distribution, and only 2.5% were accepted as fitting either a beta or normal distribution.

Since the acceptance percentage was somewhat low, a closer examination of the gamma curve fit was undertaken. In most cases, the major problem in attempting the curve fit was the peak flow as well as the trailing side of the inflow traffic data. The observed data tended to peak very quickly for a short duration. The gamma distribution could not duplicate this peak but instead tended to level it over a greater time interval. The trailing end of the observed data did not close and left gaps often totalling 5% of the peak flow. The gamma distribution tended to close much more rapidly and left a much smaller opening. In relation to the problem of the open ended tail, Benjamin and Cornell (6) state, "One should place no great confidence in the predictions based on the tails of empirically adopted models."

If the engineer needs for some reason to make a final selection between the remaining contending models in an objective way, influenced only by the data, a closeness of fit statistic such as the chi-square statistic, can be used in a somewhat heuristic manner to accomplish the choice. It should be emphasized that goodness of fit hypothesis tests are not designed to choose among contending models but rather suggest that a given proposed model should or should not be retained.

In consultation with engineers experienced in the field (4), it was judged that the gamma distribution exhibited the characteristics which most likely will describe the phenomena and therefore the gamma distribution was used in routing of the flow through the system.

B. Prediction of the Parameters of the Gamma Distribution

In order to obtain an estimation of the gamma distribution parameters, a multiple regression computer program was employed. The type of regression procedure selected was a stepwise analysis. In speaking of the stepwise regression procedure, Draper and Smith (14) state, "We believe this to be the best of the various selection procedures ... and recommend its use."

In spite of its different name, this procedure is, in fact, an improved version of the forward selection procedure. The basic forward selection procedure, given a dependent variable Y and independent variables X, is as follows. The X_i variable most correlated with Y is selected, and a first-order, linear regression equation $\dot{Y} = f(X_i)$ results. A partial correlation between $\dot{Y}$ and all X_j $(j \neq i)$ is then determined. Based on that, the second most correlated X is incorporated into the $\dot{Y}$ fit. This continues until all independent variables, which meet a specified significance level for the F-test, are utilized. One of the drawbacks of the forward selection procedure is that it makes no effort to explore the effect that the introduction of a new variable may have on the role played by a variable which entered at an earlier stage. The improvement of the stepwise regression involves the re-examination at every stage of the regression of the variables incorporated into the model in previous stages. A variable which may have been the best single variable to enter at an early stage may, at a later stage, be superfluous because of the relationship between it and other variables now in the regression. To check on this, the F criterion for each variable in the regression at any stage of calculation is evaluated and compared with a preselected percentage point of the appropriate F distribution. This provides a judgment of the contribution made by each variable as though it had been the most recent variable entered, irrespective of its actual point of entry into the model. A 95% confidence F-level was specified for this model.

To determine the estimate of the gamma distribution parameters, known values of "a" and "b", obtained from the theoretical fit of the above section, were regressed individually against all socio-economic variables described in Chapter IV. The equation generated for the approximation of "a" is

$$a = \frac{25.7(X3)^{.277}(X6)^{.158}(X7)^{.280}(X9)^{.445}}{(X1)^{.198}(X2)^{.005}(X8)^{.148}(X10)^{.111}(X11)^{.015}(X12)^{.077}(X13)^{.174}}$$

$$\cdot \frac{1}{(X14)^{.552}(X15)^{.209}(X16)^{.106}(X17)^{.023}} \qquad 5.1$$

where

X1 = Total population,

X2 = The ratio of white to non-white population,

X3 = Population per household,

X6 = Median family income,

X7 = Cars available,

X8 = Median population age,

X9 = Ratio of individuals over 21 to total individuals under 21 years of age,

X10 = Total labor force,

X11 = Total unemployed,

X12 = Ratio of male to female workers,

X13 = Ratio of workers using automobiles for the trip to work to total workers,

X14 = Median rooms per household,

X15 = Median value of owned homes,

X16 = Median amount of rent paid,

X17 = Ratio of renters to owners.

This equation had a standard error of .0217 $\log_{10}$ units and a correlation coefficient of .875. Table I lists the known parameter "a", the estimated parameter "a" from Equation 5.1, and the accompanying residual for each individual area.

The equation generated for the approximation of "b" is

$$b = \frac{.213(X1)^{.151}(X2)^{.002}(X4)^{.273}(X5)^{.053}(X8)^{.053}(X11)^{.011}(X12)^{.043}(X13)^{.182}}{(X3)^{.283}(X6)^{.144}(X7)^{.340}(X9)^{.342}}$$

$$\cdot\ (X14)^{.127}(X15)^{.180}(X16)^{.094}(X17)^{.029} \qquad 5.2$$

where independent variables are as described previously and

X4 = Number of married couples,

X5 = Median schooling of population.

This equation has a standard error of .016 $\log_{10}$ units and a correlation coefficient of .918. Table II lists the known parameter "b", the estimated parameter "b" from Equation 5.2, and the residual for each individual area. Further analysis was made in order to examine more closely the validity of both models. An examination of residuals, $Y - \dot{Y}$, did not give indication that they were not of mean 0, constant variance and members of normal populations.

Table I. The Relationship of Known and Estimated Values of the Parameter "a" in Terms of $\log_{10}$

Area Number	Known Value	Predicted Value	Residual
1	0.557	0.526	0.031
2	0.564	0.556	0.008
3	0.495	0.512	-0.016
4	0.529	0.546	-0.016
5	0.526	0.529	-0.003
6	0.525	0.521	0.003
7	0.507	0.523	-0.016
8	0.528	0.546	-0.017
9	0.564	0.562	0.002
10	0.519	0.504	0.014
11	0.526	0.526	0.000
12	0.539	0.511	0.028
13	0.537	0.524	0.013
14	0.503	0.517	-0.014
15	0.517	0.528	-0.011
16	0.506	0.529	-0.023
17	0.507	0.499	0.008
18	0.520	0.515	0.005
19	0.494	0.502	-0.008
20	0.560	0.557	0.003
21	0.485	0.500	-0.015
22	0.491	0.482	0.009
23	0.533	0.529	0.004
24	0.503	0.511	-0.008
25	0.527	0.522	0.005
26	0.533	0.524	0.009
27	0.558	0.553	0.006
28	0.509	0.504	0.005
29	0.528	0.512	0.016
30	0.549	0.542	0.007
31	0.519	0.526	-0.007
32	0.499	0.502	-0.003
33	0.509	0.521	-0.012
34	0.472	0.518	-0.047
35	0.530	0.544	-0.014
36	0.509	0.513	-0.004
37	0.552	0.529	0.023
38	0.555	0.531	0.024
39	0.551	0.532	0.019
40	0.497	0.508	-0.009

Table II. The Relationship of Known and Estimated

Values of the Parameter "b" in Terms of $\log_{10}$

Area Number	Known Value	Predicted Value	Residual
1	0.255	0.282	-0.027
2	0.254	0.258	-0.004
3	0.309	0.299	0.010
4	0.267	0.258	0.009
5	0.273	0.278	-0.005
6	0.280	0.287	-0.007
7	0.292	0.285	0.007
8	0.266	0.251	0.015
9	0.247	0.253	-0.006
10	0.286	0.296	-0.010
11	0.286	0.286	-0.000
12	0.272	0.289	-0.017
13	0.268	0.281	-0.012
14	0.288	0.281	0.007
15	0.297	0.281	0.015
16	0.295	0.276	0.019
17	0.294	0.304	-0.010
18	0.291	0.293	-0.002
19	0.310	0.305	0.005
20	0.251	0.254	-0.003
21	0.317	0.297	0.020
22	0.308	0.315	-0.007
23	0.270	0.271	-0.001
24	0.297	0.295	0.002
25	0.274	0.286	-0.012
26	0.271	0.279	-0.008
27	0.264	0.256	0.008
28	0.302	0.302	-0.000
29	0.284	0.288	-0.004
30	0.220	0.229	-0.009
31	0.283	0.281	0.002
32	0.290	0.287	0.003
33	0.297	0.289	0.008
34	0.311	0.283	0.028
35	0.278	0.266	0.012
36	0.298	0.294	0.004
37	0.266	0.276	-0.010
38	0.255	0.274	-0.019
39	0.268	0.280	-0.012
40	0.309	0.294	0.015

This observation justifies the use of the F-distribution in selecting critical values for variable significance and seems to indicate that no additional variables are needed to improve the $\dot{Y}$ estimates.

C. Determination of the Delay Parameters x and k

As described in Chapter III, Section D, a method exists for the solution of k and x. For each of the 40 stations, observed storage, $\Sigma I - \Sigma O$, was plotted against calculated storage $[xI + (1 - x)O]$. The range of variation of x was 0 to 1 in .05 increments. The x factor chosen for each area was the value which generated a plot most resembling a straight line. The slope of that line was then taken as k.

D. Prediction of the Delay Parameters x and k

In order to obtain estimates of the delay factors "x" and "k", the stepwise multiple regression computer program, described in Section B above, was again utilized. To determine the approximation to these delay variables, the known values of "x" and "k", obtained from the above section, were regressed individually against all physical variables described in Chapter IV. The equation generated for the approximation of "x" is

$$x = \frac{.389(Z1)^{.127}(Z4)^{.236}(Z5)^{.042}(Z11)^{.079}}{(Z2)^{.129}(Z6)^{.004}(Z7)^{.069}(Z9)^{.033}(Z10)^{.275}} \qquad 5.3$$

where

Z1 = Size of area,

Z2 = Length of main road,

Z4 = Mean lane width,

Z5 = Number of stop signs and traffic lights,

Z6 = Number of side streets and pedestrian crossings with right of way,

Z7 = Speed limit,

Z9 = Parking condition rating,

Z10 = Traffic flow rating,

Z11 = Driving comfort rating.

This equation had a standard error of .0684 $\log_{10}$ units and a correlation coefficient of .762. Table III lists the known factor "x", the estimated parameter "x" from Equation 5.3, and the accompanying residual for each individual area.

The equation generated for the approximation of "k" is

$$k = \frac{24.5(Z1)^{.090}(Z6)^{.008}(Z8)^{.056}(Z9)^{.043}(Z10)^{.469}}{(Z2)^{.105}(Z3)^{.111}(Z4)^{.210}(Z7)^{.121}(Z11)^{.078}} \qquad 5.4$$

where the independent variables are the same as previously designated and

Z3 = Number of lanes,

Z8 = Physical condition rating factor.

This equation had a standard error of .0941 $\log_{10}$ units and a correlation coefficient of .699. Table IV lists the known factor "k", the estimated factor "k" from Equation 5.4, and the accompanying residual for each individual area.

Table III. The Relationship of Known and Estimated Values of the Parameter "x" in Terms of $\log_{10}$

Area Number	Known Value	Predicted Value	Residual
1	-0.397	-0.424	0.027
2	-0.602	-0.540	-0.062
3	-0.397	-0.358	-0.039
4	-0.455	-0.474	0.019
5	-0.455	-0.392	-0.063
6	-0.455	-0.439	-0.017
7	-0.397	-0.420	0.022
8	-0.455	-0.517	0.062
9	-0.602	-0.430	-0.172
10	-0.346	-0.349	0.003
11	-0.397	-0.400	0.003
12	-0.455	-0.382	-0.073
13	-0.455	-0.463	0.008
14	-0.455	-0.407	-0.048
15	-0.455	-0.415	-0.040
16	-0.346	-0.454	0.107
17	-0.397	-0.453	0.056
18	-0.346	-0.398	0.052
19	-0.455	-0.432	-0.023
20	-0.455	-0.478	0.023
21	-0.455	-0.498	0.043
22	-0.455	-0.422	-0.033
23	-0.346	-0.415	0.069
24	-0.397	-0.448	0.051
25	-0.602	-0.480	-0.122
26	-0.522	-0.452	-0.070
27	-0.455	-0.450	-0.005
28	-0.346	-0.457	0.110
29	-0.346	-0.411	0.065
30	-0.522	-0.459	-0.064
31	-0.455	-0.440	-0.015
32	-0.602	-0.538	-0.064
33	-0.346	-0.408	0.062
34	-0.455	-0.474	0.018
35	-0.455	-0.458	0.003
36	-0.346	-0.424	0.078
37	-0.522	-0.504	-0.018
38	-0.455	-0.472	0.017
39	-0.455	-0.445	-0.010
40	-0.397	-0.445	0.048

Table IV. The Relationship of Known and Estimated Values of the Parameter "k" in Terms of $\log_{10}$

Area Number	Known Value	Predicted Value	Residual
1	0.939	0.986	-0.047
2	0.897	0.940	-0.042
3	0.982	0.958	0.023
4	0.973	1.013	-0.040
5	0.916	0.926	-0.010
6	0.903	0.814	0.088
7	0.931	0.936	-0.004
8	0.924	0.946	-0.021
9	0.944	0.941	0.003
10	0.900	0.954	-0.054
11	0.883	0.943	-0.060
12	0.929	0.967	-0.038
13	1.146	0.977	0.169
14	1.006	1.013	-0.007
15	0.980	1.014	-0.034
16	1.008	0.934	0.074
17	0.950	0.971	-0.021
18	0.975	0.950	0.025
19	0.946	0.955	-0.009
20	0.983	0.985	-0.002
21	1.049	1.049	-0.000
22	1.035	0.907	0.128
23	0.977	0.937	0.040
24	0.866	0.935	-0.069
25	1.010	0.960	0.050
26	0.982	0.980	0.002
27	0.532	0.890	-0.357
28	1.012	0.966	0.047
29	1.029	0.972	0.057
30	1.000	0.916	0.084
31	0.961	0.950	0.011
32	1.075	1.049	0.026
33	0.970	0.969	0.001
34	1.089	1.061	0.028
35	0.838	0.962	-0.123
36	0.921	0.891	0.030
37	1.045	0.978	0.066
38	0.950	1.033	-0.083
39	1.043	1.005	0.037
40	0.982	0.947	0.034

VI. CONCLUSIONS AND RECOMMENDATIONS

The conclusions reached from this research are based on the analysis of data from traffic flow patterns in St. Louis and St. Louis County, Missouri. It is believed that the conclusions are generally applicable to other urban areas throughout the United States.

A. Conclusions

1. The gamma probability distribution was found to be more applicable to the traffic inflow pattern than the beta and normal distributions.
2. The parameters of the gamma distribution can be predicted from various socio-economic variables peculiar to the area which generated the distribution.
3. The socio-economic variables found to be most applicable to predict an area's internal traffic pattern are:
   a. Total population,
   b. Ratio of white to non-white population,
   c. Population per household,
   d. Number of married couples,
   e. Median schooling,
   f. Median family income,
   g. Cars available,
   h. Median population age,
   i. Age ratio of minors to adults,
   j. Total labor force,
   k. Total unemployed,

l. Ratio of male to female workers,

m. Ratio of workers using auto to other means,

n. Median rooms per household,

o. Median value of owned homes,

p. Median rent payment,

q. Ratio of renters to owners.

4. The time and space delay parameters, x and k, can be predicted from various physical characteristics of the area in question.

5. The physical variables found to be most applicable to predict time and space delays are:

   a. Area size,

   b. Length of main road,

   c. Number of lanes in main road,

   d. Mean lane width of main road,

   e. Number of stop signs and traffic lights on the main road,

   f. Number of side streets and pedestrian crossings with right of way along the main road,

   g. Speed limit along the main road,

   h. Physical rating factor of the main road,

   i. Parking conditions along main road,

   j. Type of traffic flow along main road,

   k. Driving comfort rating along main road,

6. The results of this investigation must be considered positive and further research in this area seems warranted.

B. Recommendations for Further Research

1. The probability distributions tested in this investigation were the beta, gamma, and normal. Since the acceptance percentage of the gamma distribution was somewhat low, several other distributions should be investigated for their possible use.
2. The goodness-of-fit criteria used in determining the most applicable distribution was the chi-square test. Because of the sensitivity and validity of the chi-square test, it is to some extent dependent on the selection by the investigator of the number of classes and the division point between the classes. In order to reduce this inherent bias, the test should be re-examined with the class selection not based on the total number of data points, but on the criteria that each class should contain an even number of data points.
3. A more simplified method for determining the "x" and "k" delay variables should be developed.
4. The need exists to define more physical variables which play an important role in predicting "x" and "k" in Equations 5.3 and 5.4.

C. Closure

This investigation was a pilot study to determine if the approach used was feasible. The results were encouraging, and, pending the improvements suggested above, the method can be made even more exact. Although automobile traffic was used in this study, the model can be converted to mass urban transit demand with minor mathematical changes. For such a conversion, the outflow can represent the arrival rate at a mass transit terminal.

The socio-economic variables would be programmed to reflect the changing conditions of the urban area. The closed system concept, which recognizes only one exit from the system, could be revised mathematically to allow for multiple exits. Once all desired changes and implementations have been completed, the entire system could be incorporated into a computer program which would require as input only the socio-economic and physical variables of the area in question.

VII. BIBLIOGRAPHY

1. Hoel, Lester A., "Analysis of Transportation Planning Education," Transportation Engineering Journal of ASCE, Vol. 96, No. TE2, May 1970.

2. Wattleworth, Joseph A., "System Demand - Capacity Analysis on the Inbound Gulf Freeway," Research Report Number 24-8, Texas Transportation Institute, Texas A & M University, College Station, Texas, October 1964.

3. McCarthy, G. T., "The Unit Hydrograph and Flood Routing," presented at Conference of North Atlantic Division, U. S. Corps of Engineers, June 1938.

4. Gerig, Frank A., "Development of a Model Describing the Use of Acceleration Lanes," Doctoral Dissertation, Texax A & M University, August 1967.

5. U. S. Bureau of the Census, "U. S. Censuses of Population and Housing: 1960," Final Report PHC(1)-131, U. S. Government Printing Office, Washington, D. C., 1962.

6. Benjamin, Jack and Cornell, C. Allen, Probability, Statistics, and Decision for Civil Engineers, New York: McGraw-Hill, 1970.

7. Wilbur Smith and Associates, "A Highway Planning Study for the St. Louis Metropolitan Area," Vol. I and Vol. II, Private Printing, 1959.

8. Harbaugh, T. E., "Flood Routing Techniques," Unpublished class notes, CE 338, University of Missouri - Rolla, Rolla, Missouri, 1969.

9. Chow, Ven Te, Open Channel Flow, New York: McGraw-Hill, 1968.

10. Wohl, Martin and Martin, Brian, Traffic System Analysis for Engineers and Planners, New York: McGraw-Hill, 1967.

11. Drew, Donald, Traffic Flow Theory and Control, New York: McGraw-Hill, 1968.

12. Ashton, Winifred, The Theory of Road Traffic Flow, London: Methuen & Co., Ltd., 1966.

13. Miller, Irwin and Friend, John, Probability and Statistics for Engineers, Englewood Cliffs, N. J.: Prentice Hall, 1965.

14. Draper, N. R. and Smith, H., Applied Regression Analysis, New York: John Wiley & Sons, 1963.

15. Cochran, William, Sampling Techniques, New York: John Wiley & Sons, 1963.

16. Carnahan, Brice, and others, Applied Numerical Analysis, New York: John Wiley & Sons, 1969.

17. Institute of Traffic Engineers, Traffic Engineering Handbook, ed. John Baerwald, Raleigh, N. C.: Edwards & Broughton, 1965.

18. U. S. Department of Commerce, Traffic Assignment Manual, U. S. Government Printing Office, Washington, D. C., 1964.

19. Hay, William, An Introduction to Transportation Engineering, New York: John Wiley & Sons, 1961.

20. Owen, Wilfred, The Metropolitan Transportation Problem, Washington, D. C.: The Brookings Institution, 1956.

21. American Society of Civil Engineers, "Urban Planning Guide," William Claire, ed., New York: Private Printing, 1969.

22. Levinson, Herbert and Wynn, Houston, "Effects of Density on Urban Transportation Requirements," Highway Research Board, Washington, D. C., Highway Research Record, No. 2, 1963.

VIII. VITA

John Edward Thompson was born April 20, 1946, in Mannheim, Germany. He received his Bachelor of Science Degree in Civil Engineering from the University of Missouri - Rolla, in Rolla, Missouri, in May 1969.

He has been enrolled in the graduate school of the University of Missouri - Rolla since June 1969 and will receive his Master of Science Degree in Civil Engineering in August 1970.

IX. APPENDICES

APPENDIX A

The Multiple Regression Technique (14)

Consider the first order linear model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \qquad \text{A.1}$$

The regression approximation of Equation A. 1 will be

$$Y = b_0 X_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n \qquad \text{A.2}$$

where

Y = the dependent variable,

X_0 = the independent dummy variable whose value always equals unity,

$X_1, X_2, \ldots, X_n$ = the independent variables,

$b_0, b_1, \ldots, b_n$ = the approximations to $\beta_0, \beta_1, \ldots, \beta_n$.

Let m = the number of data points, and n = the number of independent variables excluding the dummy variable X_0.

Utilizing matrix notation

$$\dot{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \cdot \\ \cdot \\ \cdot \\ Y_m \end{bmatrix} \qquad \dot{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdot & \cdot & \cdot & X_{n1} \\ 1 & X_{12} & X_{22} & \cdot & \cdot & \cdot & X_{n2} \\ 1 & X_{13} & X_{23} & \cdot & \cdot & \cdot & X_{n3} \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ 1 & X_{1m} & X_{2m} & \cdot & \cdot & \cdot & X_{nm} \end{bmatrix}$$

where

$\dot{Y}$ is an (m x 1) vector,

$\dot{X}$ is an (m x n) matrix.

As the solution

$$\dot{b} = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{Y} \qquad \text{A.3}$$

where

$\dot{b}$ is an (n + 1, 1) vector of estimates of the elements β.

APPENDIX B

Chi-Square Goodness of Fit Test (13, 16)

Consider a random variable falling into any one of k classes with respective probabilities $f_1, f_2, f_3, \ldots, f_k$. For a total of n observations, the expected number falling into each class is defined by nf_i. Let the actual observed values falling into each class be x_i. The x_i would obey the multinomial distribution

$$f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1!\,x_2!\ldots x_k!}(f_1)^{x_1}(f_2)^{x_2}\ldots(f_k)^{x_k} \qquad \text{B. 1}$$

and on this basis the variable

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - nf_i)^2}{nf_i} \qquad \text{B. 2}$$

obeys the χ^2 distribution approximately.

The expected frequencies are usually based on the hypothesis of a particular distribution. Therefore, this hypothesis can be tested by evaluating χ^2 and checking to see if its value is significant in view of the number of degrees of freedom involved and the confidence level required.