U.S. Department of
Transportation

Bus Route Demand Models

Portland Prototype Study

June 1983

Bus Route Demand Models:
Portland Prototype Study

Final Report
June 1983

Prepared by
Tri-County Metropolitan Transportation District
Portland, Oregon 97202

Prepared for
Office of Planning Assistance
Urban Mass Transportation Administration
U.S. Department of Transportation
Washington, D.C. 20590

In Cooperation with
Technology Sharing Program
Office of the Secretary of Transportation

DOT-I-83-37

PREFACE

This report summarizes the results of a study conducted by Tri-County Metropolitan Transit District (Tri-Met), Portland, Oregon, the purpose of which was to develop a route-level transit patronage forecasting procedure. The study was funded by the Urban Mass Transporation Administration (UMTA) under the Section 8 Planning Assistance program. Major portions of the model development work were performed by Cambridge Systematics, Inc. under contract to Tri-Met. The original report draft was prepared by Cambridge Systematics and subsequently reviewed and edited by Tri-Met and COMSIS Corporation under special contract to UMTA.

FOREWORD

Most transit operators are faced with the need to maximize the performance of the transit services they operate. The prospect of improving performance is enhanced if operators have the ability to directly estimate the short-range patronage, cost and revenue impacts of different transit service improvement options. Route-level transit patronage forecasting models can provide this capability, but they are in limited use in the transit industry.

To assist these operators, UMTA's Office of Planning Assistance, through its Special Studies Program, has supported route-level patronage model development efforts at transit properties in four locations: Portland, Oregon; Cleveland, Ohio; Albuquerque, New Mexico; and Los Angeles, California. This report summarizes the efforts of the Tri-County Metropolitan District (Tri-Met) in Portland. We believe that these develpment efforts will be of value to operators who are interested in developing a route-level transit planning capability.

Additional copies of this report are available from the National Technical Information Service (NTIS), Springfield, Virginia, 22161 at cost.

Charles H. Graves, Director
Office of Planning Assistance (UGM-20)
Urban Mass Transportation Administration
U.S. Department of Transportation
Washington, D.C. 20590

Alfonso B. Linhares, Director
Office of Technology and Planning Assistance (I-30)
Office of the Secretary
U.S. Department of Transportation
Washington, D.C. 20590

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

APPENDIX B: MODEL APPLICATION

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

In the face of declining resources for the support of public transit programs, transit operators must increasingly attempt to maximize the performance of transit services. In order to plan and operate efficient services, it is essential that transit operators have the capability to estimate the cost, ridership and revenue implications of different service improvement options. These same economic pressures are also causing a shift in planning emphasis, away from long-range, system-wide development programs to route-level service improvements that have near-term impacts on patronage and productivity. Because the need for accurate short-range, route-level planning capability is relatively recent, there has been little progress in the development and use of route-level models. Transit operators have primarily relied on operating experience and rules of thumb for making short-range service adjustments.

This report summarizes efforts undertaken by the Tri-County Metropolitan Transportation District (Tri-Met) in Portland, Oregon to develop a short-range, route-level patronage model. The study was funded by a special Section 8 grant from the Planning and Research Division of the Urban Mass Transportation Administration. The actual model development work was performed by Cambridge Systematics, Inc., under contract to Tri-Met.

The objective of this study was to develop a method for directly estimating short-range transit patronage impacts, on an individual route level, resulting from typical short-range transit improvements, such as changes in service frequency, route alignment, travel time, and accessibility. A supporting objective was that the method should be easy to apply and use, requiring data and staff capabilities commonly available to a transit operator, while at the same time yielding accurate predictions.

The resultant method, the "Route Patronage Forecasting Model", entails a four-step process, which incorporates both trip generation and distribution. The method's distinguishing characteristic is its use of simplified disaggregate logit-type equations to simulate selection between alternative transit paths and routes based on relative "ease of travel." The method was developed in two forms, "automated" and "manual." The first is designed to be operated in a full-fledged network simulation context. This version is data-intensive, and gives full consideration to the complexities of service changes and consumer choice in a large, interconnected transit system. The manual version modifies the automated approach to allow use of the Route Patronage Forecasting Model with manually-prepared data from a sample of affected origins and destinations.

This report describes the development and application of the Tri-Met Route Patronage Forecasting Model. The report is organized as follows:

- Chapter 2 provides a description of Tri-Met and the study region.
- Chapter 3 reviews the methods used by Tri-Met for transit planning before the current study, and describes the goals and objectives of the new model development.
- Chapter 4 describes the development and characteristics of the new Route Patronage Forecasting Model.

- Chapter 5 provides instruction in application of the model, and gives the result of test trails for both the automated and manual form.

- Chapter 6 summarizes the development effort and the strengths and weaknesses of the Route Patronage Forecasting Model.

- Appendix A provides supplemental detail on model derivation and development, and special issues with regard to estimation biases and application limitations.

- Appendix B summarizes the procedures for model application, in particular special data handling software developed to operate the Route Patronage Forecasting Model.

CHAPTER 2: TRI-MET STUDY REGION

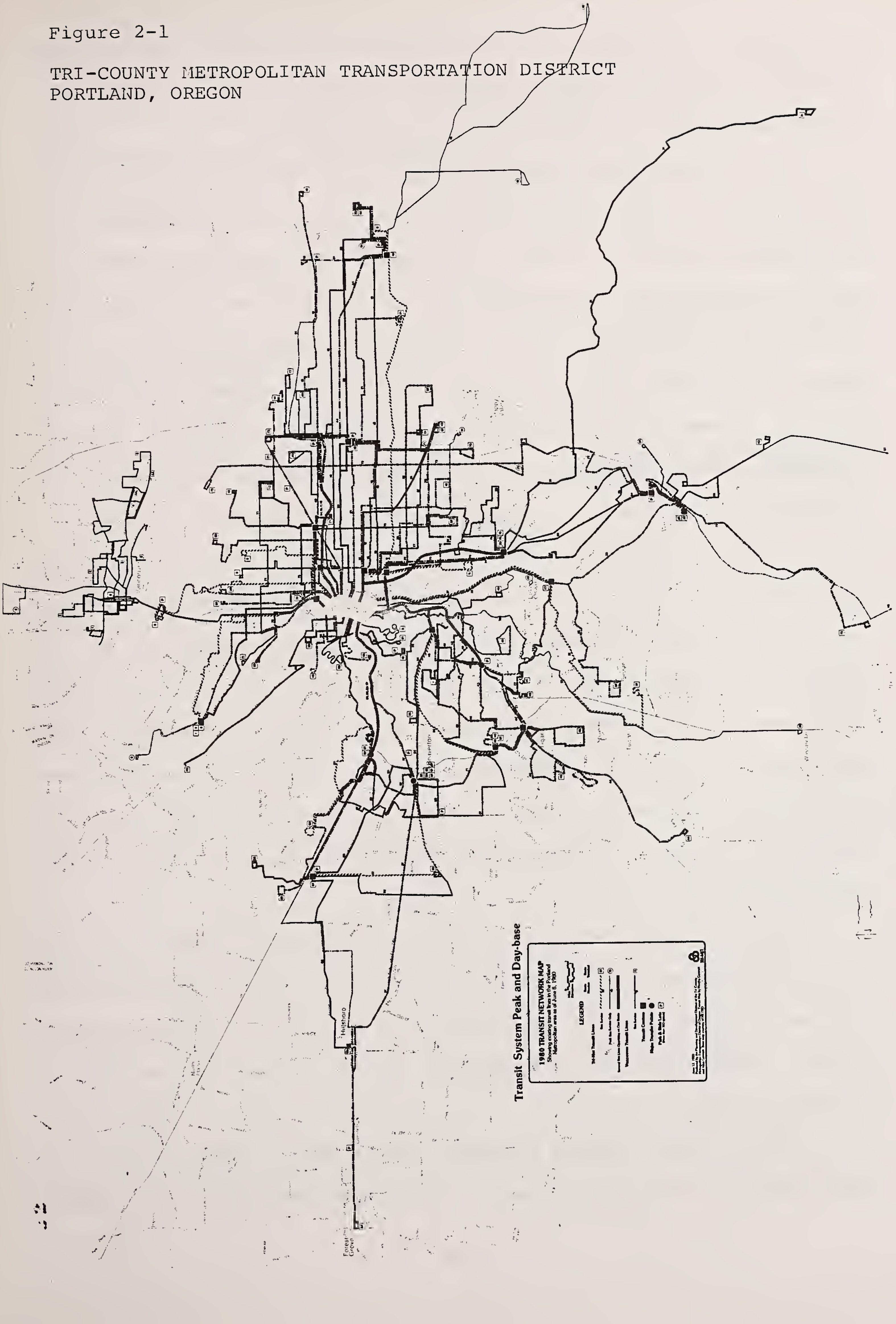
This project was undertaken by the Tri-County Metropolitan Transportation District of Oregon (Tri-Met), the mass transit operating agency for the Portland metropolitan area. The Tri-Met region includes three Counties (Multnomah, Washington, and Clackamas) and covers an area of approximately 1,000 square miles. This makes Tri-Met the largest transit district in Oregon and the fifth largest transit operator on the West Coast of the U.S. The total population of the Portland metropolitan area, according to the 1980 census, is 1,050,192. A map of the study region is provided as Figure 2-1.

Tri-Met is a non-profit, municipal corporation established under state law. The agency provides a number of public transportation services, including fixed-route transit, ridersharing, and contracted accessible services for the handicapped. The fixed-route bus system contains 71 routes with over 8,000 bus stops and 68 park-and-ride lots. Service is supplied by a fleet of 566 diesel coaches operating over 22 million annual miles. The system carries 48 million passengers per year. Operations are supported from three divisional bases.

The ridesharing program provides regional carpool matching services, with over 350 monthly inquiries. Tri-Met issues monthly Carpool Parking Permits for carpools of three or more persons, which allow unlimited parking at certain downtown parking meters. This program has more than 1,200 users. Tri-Met also disseminates technical guidance and evaluation assistance to employers

Figure 2-1

TRI-COUNTY METROPOLITAN TRANSPORTATION DISTRICT
PORTLAND, OREGON



who sponsor employees' vans for commuting purposes. This program has also produced eight buspools that provide service between major employers and local communities.

Tri-Met operates a special accessible service for physically or mentally disabled patrons who are unable to use the regular bus system. These services are coordinated by Tri-Met and operated by independent suppliers using over 40 specially equipped vehicles. Wheelchair lifts have also been introduced into the regular bus fleet, and the Light Rail line now under construction will be accessible for riders in wheelchairs. Finally, Tri-Met provides transit service in rural areas through contracts with independent suppliers.

In the Portland region at the present time, about four percent of all trips are made by transit, and 96 percent by car or other modes. Twenty-eight percent of all residents in the three-county area ride the bus at least twice per month. During peak hours, 35 percent of all trips to and from the downtown area are made by transit. Ridership in the Tri-Met system has been growing at an average rate of seven percent per year since its inception in 1969. Currently, ridership stands at about 135,000 originating riders per average weekday.

Recent counts indicate that 76 percent of all fares are paid at the adult rate, 12 percent at the student rate, seven percent at the honored citizen rate, and five percent at miscellaneous rates. Approximately 43 percent of all trips are made with monthly passes.

The Tri-met system contains three fare zones. Zone 1 is also known as "Fareless Square." It allows unlimited free travel within the downtown area of Portland, except between 3:00 p.m. and 7:00 p.m. on weekdays. As of October 1, 1980, fares have been set at the following levels:

Student	$ .45	
Senior Citizen & Disabled	.25	midday hours
Adult	.65 .90	up to two zones for three zones
Adult (Portland-Vancouver)	1.00	flat rate
Monthly Pass (Youth)	14.00	
Monthly Pass (Adult)	21.00 29.00	up to two zones for three zones
Monthly Pass (Portland-Vancouver)	35.00	

CHAPTER 3: STUDY DESIGN

This chapter reviews the methods used by Tri-Met before the current study and describes the development of the study approach.

3.1 Pre-Study Planning Methods

Prior to the development of the Route Patronage Forecasting Model, evaluation of service changes at Tri-Met was done using one of two methods: (1) regional travel demand forecasting models; or, (2) empirical estimation techniques.

3.1.1 Regional Demand Models

The Metropolitan Service District (MSD), the regional planning agency for the Portland area, has developed a regional travel demand forecasting process which incorporates a series of models developed under the Urban Transportation Planning System (UTPS). The models are of the disaggregate logit form and were calibrated from data gathered in an extensive household travel survey conducted in 1977. This forecasting package consists of: (1) auto ownership models; (2) trip generation models (for each of three trip purposes); (3) preliminary mode choice models; (4) destination choice models; (5) mode choice models; and (6) auto occupancy models.

Although these regional models have been successfully used in support of several large corridor studies, they have not been used for route-level analysis for two main reasons. First, the models require a considerable amount of data and computer resources to operate, including UTPS-coded highway and transit networks and detailed travel data by census tract. This significantly affects response time. Second, the models do not accurately reflect route-level travel decisions. The standard UTPS model system is unable to handle situations where the traveler has a choice of more than one bus route. It similarly has trouble in handling route interconnections or transfer points, which are important features in the Tri-Met system.

3.1.2 Empirical Estimation Techniques

The process used most frequently by Tri-Met to forecast patronage for service changes affecting only a few routes is the so-called RPSH (riders per service hour) method. Tri-Met maintains historical records on the number of peak and off-peak riders per service hour (RPSH) for each of the lines in the system. The RPSH factors are categorized by route type (regional, urban, radial, peak hour, local radial, grid feeder, owl) and/or area type (urban, suburban, etc.).

When service changes are to be made, the change is translated to number of (revenue) service hours, SH, for the route type in question. Average boarding rides per service hour, or BR/SH, are then calculated for similar route types from the historical data. The BR/SH average is then multiplied times the proposed change in service hours, SH, to estimate the new boarding rides expected.

3.1.3 Conclusions on Pre-Study Planning Methods

In summary, Tri-Met's pre-study capability for performing route-level transit planning was quite limited. The computer-based regional demand models suffered from being both cumbersome and inaccurate for route-level applications. Virtually all route-level changes were evaluated through the RPSH method. While this method is relatively easy to apply (typically operated by a $7/hr. technical analyst) and is directly applied at the route-level, it is a very naive method, and not terribly accurate. It depends greatly on the similarity between the routes and the service changes for which the RPSH factors are being shared. The more complex the service charge, and the more the characteristics of the route differ from the norm, the less accurate are the forecasts from the RPSH method.

3.2 Review of Existing Experience

Tri-Met's quest for a relatively simple technique to directly forecast patronage on individual routes was preceded by a review of previous experience with these techniques. The review concentrated on what were regarded as the most extensive efforts to date in subarea or route-level patronage forecasting, which included the following reports:

1. **Transit Corridor Analysis--A Manual Sketch Planning Technique,** COMSIS et al., April 1979, UMTA Report UMTA-MD-06-0046079-1.
2. **Quick Response Urban Travel Estimation Techniques and Transferable Parameters,** COMSIS, et al., 1978, TRB/NCHRP Report 187.
3. **Traveler Response to Transportation System Changes,** Barton-Ashman and R.H. Pratt & Co., July 1981, USDOT Report DOT-FN-11-9579.
4. **Characteristics of Urban Transportation Demand--A Handbook for Transportation Planners,** Wilbur Smith, Deleuw Cather & Co., April 1978, UMTA Report IT-OG-0049-78-1.

These reviews failed to uncover any promising alternatives to Tri-Met's existing methods. The first three references were found to employ various shortcuts of the traditional (network simulation-based) travel forecasting process. Neither adequately addressed the issue of traveler selection among multiple routes which is a particular concern in route patronage forecasting at Tri-Met. The other two references offered only empirical "ball-park" guidelines, similar to the RPSH method already in use.

3.3 Project Objectives

Tri-Met's objective was to develop a capability to produce accurate estimates of patronage impacts with quick turnaround response and with modest data and computational requirements. In the ideal, a method was desired that embodied the simplicity and ease of application of the RPSH approach, but with a higher degree of accuracy. By "accurate", it was hoped that the methods would more directly reflect the market and service characteristics of the given route than did the RPSH factors, and in particular be responsive to the need to forecast impacts in a highly interconnected route system. By "quick", it was hoped that the model could respond to virtually all problems within a day.

The developed methods were to be responsive to the following types of service changes:

- service frequency
- travel time
- route alignment

- accessibility (location of bus stops, stations, or park-and-ride lots)
- interconnections with other routes.

It was also Tri-Met's desire that the forecasting techniques be designed around existing data resources. The reasons for this were both the costs of additional data gathering and the desire to maintain simplicity in the approach. As summarized in Table 3-1, these data included: census tract level population and employment data; route level transit ridership data (passenger counts, on/off counts, and on-board passenger survey information), and data on transit and auto system performance as taken from Urban Transportation Planning System (UTPS) and Run Cutting and Scheduling (RUCUS) networks.

Because of the focus on short-range, route-level projections, it was determined that the model need not be sensitive to such long-range or system-wide factors as:

- major land use changes
- major socioeconomic changes (population or employment)
- major changes in the highway network (such as a new freeway or bridge)
- exogenous conditions such as fuel rationing, shifts toward multi-family housing, or changes in general economic conditions
- large, capital-intensive transit improvements such as a fixed-guideway system

3.4 <u>Preliminary Study Approach</u>

Tri-Met's initial study design recommended a two-step forecasting procedure as the most logical approach to route forecasting. This process would include an initial transit trip

Table 3-1

EXISTING TRI-MET AND MSD DATA RESOURCES

DATA VARIABLES	SOURCE	DATES	LEVEL OF DETAIL CENSUS TRACT	 TRANSIT LINE
Socioeconomic				
Population	*MSD	1977	Yes	
Employment (by 2-digit SIC)	MSD	1976	Yes	
Ridership				
Passenger Counts				
- Weekday by quarter	Tri-Met	Present	---	Yes
- Saturday by quarter	Tri-Met	Present	---	Yes
- Sunday by quarter	Tri-Met	Present	---	Yes
- By time of day	Tri-Met	Present	---	Yes
- By fare category	Tri-Met	Present	---	Yes
- On/Off by stop (Section 15 reporting)	Tri-Met	Present	---	Yes
On-Board O&D Survey	Tri-Met	May '80	Yes	Yes
Transit Operations				
UTPS (INET network simulation)		Present	---	Yes
RUCUS (Description of Service)		Present	---	Yes
Miles			---	
Hours			---	
Layovers			---	
Deadheads			---	

*MSD (Metropolitan Service District) is the Regional Planning Agency for the Portland Region.

generation phase (number of trips made) followed by a separate trip distribution and assignment phase (where the trips are made). It was expected that each step would be built around one or more regression models, which would be developed from the existing data summarized in Table 3-1.

Tri-Met anticipated that the trip generation models would be developed at the census tract level of detail to reflect the nature of the data and the desire to produce simplistic, easily-applied formulations. It was expected that a set of equations or equivalent look-up tables would be developed to yield estimates of total transit travel from a small number of key independent variables. Tri-Met expected that these variables would include measures of accessibility (% of population or employment within specified distance of transit) and level of service. It was hoped that these methods would be simple and flexible enough to allow application under different response conditions, ranging from users with detailed data desiring fairly precise estimates to those with little or no primary data who would be satisfied with "rough-cut" estimates. Simplified expressions and development of "default factors" and procedures for adjusting the model coefficients were seen as the way to accomplish this objective.

The trip distribution model in the two-part planning system was then expected to focus the overall transit ridership predictions from step 1 into specific route-level impacts. This was not only for the purpose of examining ridership impacts on a particular route, but to address important design-related questions such as loadings by route segment, seat turnover, and direction of travel. It was anticipated that the distribution of transit trips to the individual route would be accomplished by first dividing the route into segments, then allocating the trips to individual segments using some type of weighting or proportioning scheme.

CHAPTER 4: PROJECT RESULTS

This chapter describes the results of the route-level model development effort. The Route Patronage Forecasting Model (RPFM) developed by the study reflects a viable route planning capability, though it also marks a departure from several of the project objectives. The approach incorporates a four-step process, rather than the two-step technique suggested in the initial study design. It is, however, structurally similar to the original approach. It still begins with a separate trip generation step, and follows this with trip distribution. The difference is in the detail afforded the trip distribution task; three individual steps are developed to handle the aspects of path choice, route choice, and segment loadings.

The present model is not the highly simplistic, manually-operable technique that was initially forseen. During the initial development phase, typical Tri-Met planning applications were inventoried and their impact on the forecasting methodology assessed. It was quickly realized that the issue of interconnectivity and multiple choice of routes would be a major factor in determining the accuracy and degree of complexity of the eventual model. Tri-Met is a highly interconnected system, with an average of over 10 intersecting routes per line (exclusive of connections in downtown Portland). The western portion of the network incorporates a "timed-transfer" (pulse point) system where nine lines are focused on two transit centers. The eastern portion incorporates a grid system with 16 radial lines and eight crosstown lines. The systemwide transfer rate is about 28 percent of all originating rides.

The members of the project team determined that the RPFM would be unrealistic and of limited use if it did not allow for interconnectivity and user choice among several transit alternatives. This feature greatly expanded the sophistication of the modelling approach, the data requirements, and the effort in applying the model to planning tasks. The basic model, in the so-called "automated" version, is rather data intensive and requires a main frame computer for operation. Various software aids (utility programs) and systematic procedures have been devised to expedite operation, but application of the automated RPFM is still a complex and lengthy operation. A "manual" adaptation of the RPFM has been developed, which is designed to reduce the methodology to a level of sophistication compatible with minicomputer or hand-calculator application. While initial results are described in this report, however, the manual version is still in the development phase.

4.1 Overview of Model

The four steps in the Route Patronage Forecasting Model procedure are:

Step 1:	transit trip generation
Step 2:	trip distribution among alternative transit paths
Step 3:	trip distribution among alternative transit routes within a given path
Step 4:	estimation of trip loadings by individual route segment

The purpose of the individual steps in the process is best visualized through a simple example, as offered by Figure 4-1. In this example it is desired to forecast transit ridership on a route "E" which connects two points, represented by traffic analysis zones* 1 and 2, in the diagram. An individual wishing to travel between zones 1 and 2 has several options. He may:

- take path 1, which is served by Route C directly;
- take path 2, which includes Route A with a transfer to Route B;
- take path 3, which includes Route D with a transfer to either Route E or Route F.

The four-step Route Patronage Forecasting Model is used to estimate the ridership impacts on Route E in the following manner:

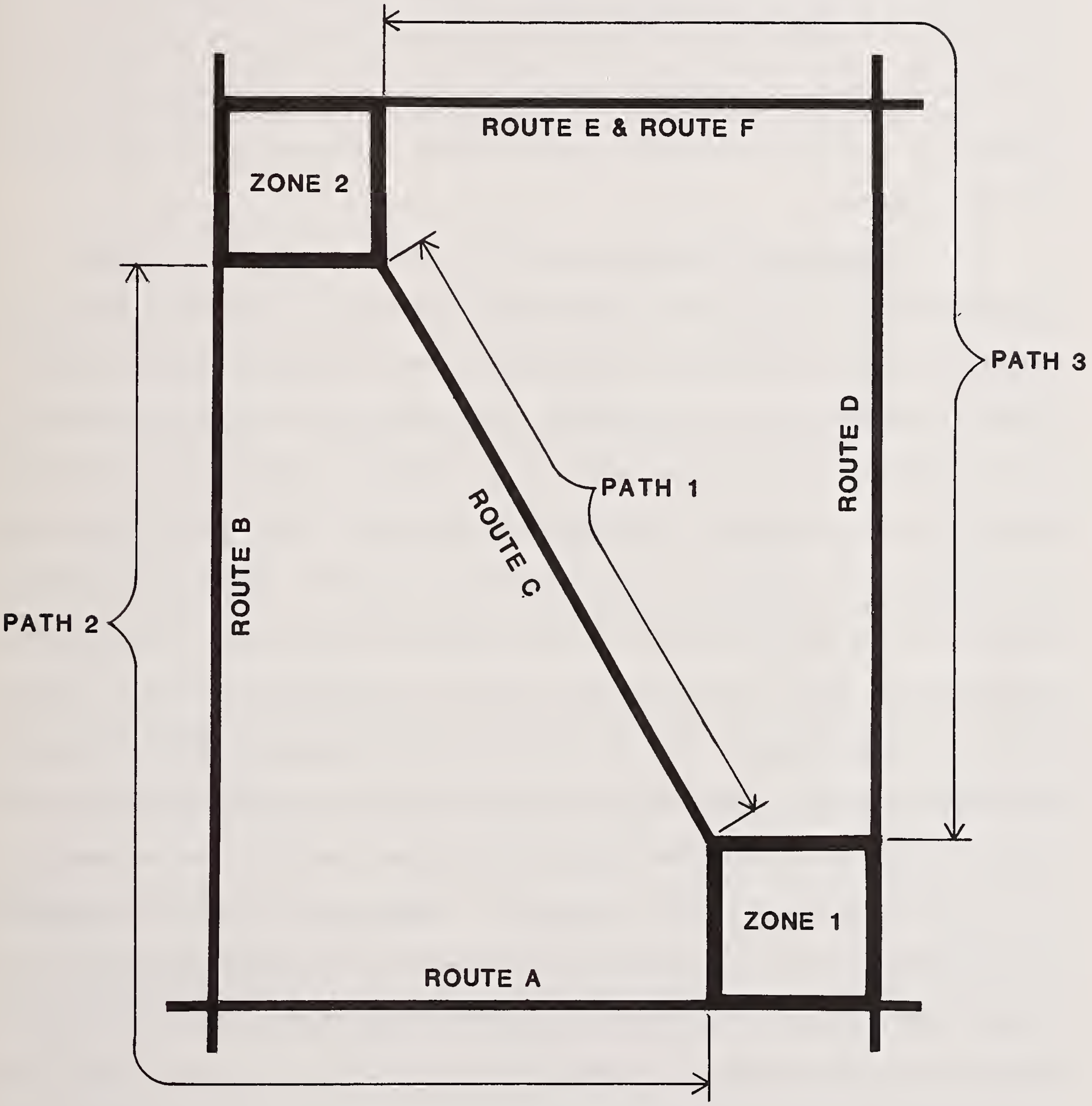
1. Predict the total number of transit trips (N) taken from zone 1 to zone 2.
2. Predict the proportion of trips using path 3 (P3), and calculate the number of trips on path 3 ($N \cdot P3$).
3. Determine the proportion of trips using Route E (RE), and calculate the number of trips on Route E ($N \cdot P3 \cdot RE$).
4. Determine the total number of trips on the given segment of Route E by performing the same calculations for all other zone pairs for which there is a possible transit path that utilizes Route E, and sum these trips over the segment.

---

*The traffic analysis zone is a convention of the Urban Transportation Planning System (UTPS), somewhat analagous to the census tract. It is a geographic/demographic sector of variable size and population, developed to segment and codify the regional transportation (chiefly highway) network and population into convenient units for analysis.

Figure 4-1

ILLUSTRATION OF MODEL APPLICATION



Because the Route Patronage Forecasting Model is characteristically thorough in its consideration of interdependent relationships, it was necessary to consider limiting the scope of the analysis in order to minimize the computational requirements and the actual time to perform the analysis. As a result, criteria were developed to control both the number of routes and the number of origin-destination zone pairs affected by a particular service change.

It is nearly always possible to use a particular route for travel between any given origin-destination pair, provided enough transfer and circuity are tolerated. To eliminate these theoretically possible but unrealistic alternatives, two criteria were established. First, at most three transit paths are allowed between each origin and destination. Second, only paths with two or fewer transfers are considered. The net effect of these restrictions is to significantly reduce the number of origin-destination zone pairs which must be considered for a given service adjustment, and to restrict the number of bus routes presumed to have been affected by a given service change. A third criterion was developed to further reduce the number of zone pairs entering the analysis. This rule limits eligible origin-destination pairs to those which currently generate at least nine transit trips on the route in question, or which have more than 50 automobile trips.

4.2 Model Development

This section describes in greater detail the development of the individual steps in the Route Patronage Forecasting Model. Discussion is limited to the essential details of conceptualization and estimation only. The more technical aspects related to derivation of model form, composition of variables, estimation difficulties, and discussions of statistical accuracy are contained in Appendix A.

Step 1: Forecast Transit Trips Between Zone Pairs

The number of transit trips taken between an origin and destination zone pair of zones is predicted with trip generation equations. These equations were developed using ordinary least squares regression techniques. The RPFM includes separate expressions for each major trip purpose: home-based work (HBW); home-based other (HBO); and non-home-based (NHB). In predicting the number of trips that will occur, the models consider individual rates of travel, travel population and employment, accessibility of transit service, and the comparative ease, or "impedance", of travel by auto and transit. "Impedance" is nothing more than an expression of the cost of travel, i.e., travel time and money cost.

The individual trip generation models are shown in Figure 4-2, with the coefficients as estimated by the regression. Goodness of fit may be evaluated using the t-statistics which appear in parentheses below each parameter estimate. Each estimate is significant at the 95 percent confidence level.

Figure 4-2

TRIP GENERATION MODELS

HBW Transit Trips

$$\ln N_{ij} = \underset{(10.30)}{-\ .098} + \underset{(39.95)}{.642\ \ln}(Q_{ij} \cdot R_i \cdot S_j) + \underset{(16.15)}{.333}\ YT_{ij}$$

$$\underset{(29.99)}{-\ .665}\ \ln(e^{YT_{ij}} + e^{YA_{ij}})$$

$R^2 = .986$
MSE = .2874

HBO Transit Trips

$$\ln N_{ij} = \underset{(4.36)}{-\ .733} + \underset{(**)}{1}\ \ln(Q_{ij} \cdot R_i) + \underset{(**)}{1}\ YT_{ij}$$

$$\underset{(22.4)}{+\ .599}\ \ln(e^{YT_{ij}} + e^{YA_{ij}})$$

$R^2 = .987$
MSE = 1.657

NHB Transit Trips

$$\ln N_{ij} = 0^{*} + \underset{(**)}{1}\ \ln(Q_{ij} \cdot R_i) + \underset{(43.14)}{.285}\ YT_{ij}$$

$$\underset{(20.01)}{-\ .293}\ \ln(e^{YT_{ij}} + e^{YA_{ij}})$$

$R^2 = .998$
MSE = .222

where:

N_{ij} = the number of transit trips of the stated purpose between zone i and zone j

Q_{ij} = the total number (auto and transit) of trips of the stated purpose taken from home in zone i to work in zone j

R_i = the proportion of population in zone i served by transit

S_j = the proportion of employment in zone j served by transit

YT_{ij} = transit impedance between zone i and zone j

YA_{ij} = auto impedance between zone i and zone j

*Coefficient constrained to theoretically derived value of zero
**Coefficient constrained to theoretically derived value of 1

The models are very similar in structure. Each forecasts transit zone-to-zone trips based on the total number of trips by all modes between the given origin and destination (Q_{ij}), the proportion of the population in the origin zone living within 1/4 mile of transit service (R_i), and the impedance associated with travel by auto and transit (YA_{ij} and YT_{ij}, respectively). The home-based work model is slightly different in that it includes transit access in the destination as well, or % of employment within 1/4 mile of transit (S_j). Data on the number of transit trips, N_{ij}, was developed from Tri-Met's 1980 on-board transit survey. Data on total trips, Q_{ij}, was obtained from a 1977 regional origin-destination survey, and transit and auto impedance, YT_{ij} and YA_{ij}, was taken directly from the Metropolitan Service District's regional network planning models. The impedance terms reflect a weighted sum of in-vehicle time, wait time, walk time, and cost for the respective mode. A complete description of the impedance functions is provided in Appendix A.

The trip generation equations in Figure 4-2 are modifications of the original models. The essential changes were to add accessibility measures to the specification of all models, and to eliminate a destination choice feature from the non-work (HBO and NHB) models.

The original HBW model predicted transit work trips between two zones as a function of total interzone travel, and transit and auto travel impedance. The study team reasoned that the model would be more accurate if it took into account the nearness of transit at both the origin (residence) and destination (employment) ends of the trip. This was accomplished by adding the terms R_i and S_j to the model, representing, respectively, the percent of residences and employment within 1/4 mile of transit service.

The HBO and NHB equations were also revised to include a transit origin accessibility measure (R_i). In addition, the equations were modified to eliminate the estimation of trips between zone pairs. The revised set of equations is designed to use existing trip estimates taken from UTPS (total person) trip tables. The details of the modifications to all models are explained in more detail in Appendix A.

To apply the trip generation models for the example in Figure 4-1, it is necessary only to indicate the change in transit service to the model, as represented through the transit impedance function, YT_{ij}. The values for all other inputs to the equation remain unchanged. The YT_{ij} which is input to the model is the weighted sum of transit impedance over all feasible paths (as defined in the previous section). This is calculated as follows:

$$YT_{ij} = \sum_k P_{ij}^k \cdot YT_{ij}^k$$

where P_{ij}^k is the percentage of transit trips between i and j that use path k and YT_{ij}^k is the transit impedance for path k between i and j. Thus, transit impedance between zones 1 and 2 is equal to:

$$YT_{1,2} = (P_{1,2}^1 \cdot YT_{1,2}^1) + (P_{1,2}^2 \cdot YT_{1,2}^2) + (P_{1,2}^3 \cdot YT_{1,2}^3)$$

In our example, paths 2 and 3 involve multiple routes. Therefore, impedance for path 2 is calculated as the sume of the impedances of the two routes which must be used to travel from zone 1 to zone 2, or:

$$Y^{2}_{1,2} = Y^{\text{route A}}_{1,2} + Y^{\text{route B}}_{1,2}$$

Similarly, the impedance for travel on path 3 is calculated as the sum of the impedance on the first leg, route D, plus the second leg, served by routes E and F. Because routes E and F represent alternative choices on the second leg, it is necessary to use the average impedance which is calculated as the sum of the impedances for each route weighted by the percent of riders using each route, or:

$$Y^{3}_{1,2} = Y^{\text{route D}}_{1,2} + (P^{\text{route E}}_{1,2} \cdot Y^{\text{route E}}_{1,2} + P^{\text{route F}}_{1,2} \cdot Y^{\text{route F}}_{1,2})$$

Of course, changing level of service along one or more transit paths affects traveller selection of path, and hence, P^{k}_{ij}. For this reason even though steps 2 and 3 of the RPFM, which involve choice of path and route, are presented as being performed **after** the trip generation step, in actuality they are performed first, in order to supply the necessary information about P^{k}_{ij} for the calculation of impedance.

Step 2: Distribution Among Alternative Paths

In this step the proportion of trips taken on each feasible path is estimated. This is done by comparing transit impedance along each path. This is the same measure of transit impedance as used in the trip generation models, which is a weighted sum of in-vehicle time, wait time, walk time, and cost. The impedance formula is taken directly from MSD's regional travel model, and is illustrated in Appendix A.

The equation for estimating the distribution of trips among alternative paths has the form of a logit model. This mathematical formulation is commonly used for predicting consumer choice among travel alternatives. The advantages of the logit approach are both demonstrated statistical accuracy in predicting choice, as well as the representation of the choice decision as a probability relationship. This latter characteristic is very convenient when the **proportion** of the applicable market which will use each alternative is desired.

The Tri-Met study team conceptualized* that the proportion of trips utilizing a given alternative path is predicted by the following equation:

$$P_{ij}^{k} = \frac{e^{YT_{ij}^{k}}}{\sum_{l} e^{YT_{ij}^{l}}}$$

---

*The path choice relationship was conceptualized during the study as a logical representation of the probable basis on which such a travel decision is made. Consistent with most travel demand forecasting approaches, the path choice model assumes that travellers will choose the path of least resistance in terms of relative travel time and cost penalties among alternatives.

where:

P^k_{ij} = the proportion of transit trips from zone i to zone j that utilize path k

YT^k_{ij} = the travel impedance encountered in traveling by transit from zone i to zone j along path k

$\sum_l e^{YT^l_{ij}}$ = the logrithmic sum of travel impedance over all feasible alternative transit paths l between zone i and zone j

For the example from Figure 4-1, the proportion ($P^3_{1,2}$) of trips between zone 1 and zone 2 that would take path 3 (which contains Route E) is calculated as:

$$P^3_{1,2} = \frac{e^{(\text{travel impedance, path 3})}}{e^{(\text{tr imp,path 1})} + e^{(\text{tr imp,path 2})} + e^{(\text{tr imp,path 3})}}$$

When applying the model it is necessary only to furnish information to recalculate transit impedance for the service improvement. All other inputs to the model remain unchanged.

Step 3: Distribution of Trips Among Individual Routes

Because it is possible to have more than one route within each of the paths in step 2, a separate step is necessary to distribute path trips among alternative routes.

To allocate trips among routes, the Tri-Met study team conceptualized a relationship which assumes that riders will take the first available carrier they encounter along a path leg. The proportion of trips using a given route is calculated as:

$$P_{ij}^{r} = \frac{C_r}{\sum_k C_k}$$

where:

P_{ij}^{r} = the proportion of transit trips between zone i and zone j taking path k which use route r

C_r = Number of carriers per hour along path leg k for route r

$\sum_k C_k$ = Number of carriers per hour along path leg k for all routes

To illustrate using the example from Figure 4-1, if route E on path 3 has 20-minute headways (three vehicles per hour) and route F on path 3 has 15-minute headways (four vehicles per hour), step 3 would predict that three-seventh's of the trips along path 3 would utilize route E.

Step 4: Calculating the Total Number of Trips on Route

If steps 1 through 3 are performed on all zonal pairs which generate transit trips that use route r, total ridership on route r may be calculated as:

$$T_r = \sum_{ij} N_{ij}^{k} \cdot P_{ij} \cdot P_{ij}^{r}$$

where:

T_r = total trips using route r over segment s

N_{ij} = number transit trips between zone i and zone j (step 1)

P_{ij}^{k} = proportion of transit trips between zone i and zone j taking path k which contains route r (step 2)

P_{ij}^{r} = proportion of transit trips on path k that would be taken on route r (step 3)

If it is desired to forecast the number of trips taking place on a specific segment s of route r (for illustration, the leg of path 3 that carries route E in the example in Figure 4-1), the equation above is restricted to those zone pairs ij which utilize segment s. In other words, this program module simply allows for the orderly summation of effects over the route segment of interest.

CHAPTER 5: MODEL APPLICATION AND VALIDATION

The Route Patronage Forecasting Model (RPFM) presented in the previous chapter is a reflection of the complexities inherent in evaluating service changes in a large, interconnected regional transit system. The trade-off in of building a model to accurately simulate multiple route alternatives and choice behavior is that it becomes computationally cumbersome. In its unaltered form, the RPFM requires a mainframe computer installation to process the vast amounts of data which it acquires from the Urban Transportation Planning System (UTPS) data files.

Two strategies were employed to make the RPFM more efficient and usable. The first entailed the development of special software to manage the vast amounts of data necessary to perform a normal full run. These software packages, which are described in Appendix B, significantly reduced the amount of effort and computation time to set up and run the model. The second strategy was the development of a procedure for applying the model in a manual mode. This modification responds to situations where the computer resources to produce the data file are prohibitive or where the user does not have access to a coded UTPS network. The manual method relies instead on a sampling system to produce the necessary data for the model.

This chapter presents separate descriptions of both the fully "automated" and the "manual" versions of the RPFM model. The discussion for each begins with a summary of the data requirements and the special data handling procedures which are used. A subsequent section discusses important issues in application, including calibration requirements and rules for using the model for typical application tasks. The discussion then covers testing and validation of the model using data on test cases drawn from existing Tri-Met services. A final section describes the time and effort involved in operating RPFM for typical planning tasks in the respective automated or manual mode.

5.1 Automated Method

5.1.1 Dataset Development and Data Processing Procedures

For the automated version of RPFM, application involves three processes: (1) developing the data input file; (2) calibrating the model; and (3) running the simulation using a specially developed SAS (Statistical Analysis System~) utility program. These steps are discussed briefly below.

The data for the Route Patronage Forecasting Model comes from a variety of sources. The basis for the information on transit paths, transit travel, access and transfer times and transit fares was the information collected and stored in INET (Integrated Transit Network), a module of the Urban Transportation Planning System (UTPS) package. Transit population and employment access (percent served) figures were derived from transit route maps and land use maps, while auto travel times and costs, and auto and transit trip tables were derived from the UTPS network. Auto ownership factors were developed from origin-destination survey data and the census.

These data are both varied and voluminous, which necessitated development of special data-handling procedures. While the effort associated with initial development of the raw data elements listed above cannot be avoided, a special data handling procedure was developed by Tri-Met to ease the problem of transforming this data into appropriate format for the operation of the automated RPFM for each planning application. This program, "ROOTINFO", extracts the necessary information from the INET data files and places it in the correct format. Four types of information are managed by ROOTINFO:

1. Accessibility--percent of zone population/employment served by transit;
2. General zonal characteristics--proportions of households with zero, one, or two-plus autos;
3. Impedance--travel times and costs along each path, transit line, and leg of the transit system operating between zones; and
4. Total person trips by purpose--HBW, HBO, and NHB (also broken down by auto and transit for calibration purposes).

A more complete discussion of ROOTINFO may be found in Appendix B-1.

Another special data handling issue results from the multiple path determination feature of the RPFM. RPFM, of course, accounts for transfer movements between the transit route line examined and all other lines in the system. Determination of paths among zonal interchanges is, however, a complex process, requiring computer assistance when more than a few zones are involved. While RPFM allocates up to three paths between zone pairs, the model itself is unable to identify paths, so this information must be determined externally. The forecasting model uses UTPS to determine paths and interchange information, but since the path assignment procedure (UPATH) in UTPS uses an all-or-nothing assignment logic, the UTPS reference must be run three times, with manipulation to force selection of three reasonable paths. This procedure has been formalized and is discussed in greater detail in Appendix B-2.

Finally, a series of special utility programs has been developed to handle special problems related to formatting and transfer of data into and out of ROOTINFO, and construction of interim workfiles. Information on these procedures may be obtained on request from Tri-Met.

5.1.2 Model Application

Once the input dataset ROOTINFO is in place, the RPFM model may be applied to a variety of route-level planning and analysis tasks. This section describes, in very general terms, the procedures to be followed when applying the model to the different types of planning problems. A more detailed discussion of the application process is presented in Appendix B-3.

Before applying the RPFM model to the evaluation of service changes, it may be necessary to calibrate the model to the new situation. Calibration, which involves adjustment in the model parameters, is necessitated by differences between the conditions reflected in the base data from which the model is estimated and the new application. Generally, the model should be calibrated when it is being used for entirely new proposed services, or when being transferred to new locations. New routes, route extensions, changes in alignment, and major changes in land use in the service area are all applications where calibration of the model before use is advisable.

For applications where calibration is undertaken, it is recommended that a service area be identified with sociodemographic characteristics and levels of transit service similar to the proposed new service. The model is calibrated using an interactive process, which is described in detail in Appendix B-4. The model is first run on the sample route, and predicted vs. actual ridership levels are compared, resulting in an adjustment factor. If a good fit is not observed, the intercept term in the model is modified using the adjustment factor. The process is repeated until a satisfactory fit is realized.

For planning applications other than those listed above, i.e., to evaluate operational changes which affect only travel time or service frequency, the RPFM models may be used "as is." However, for best results, it is recommended that the model be used to predict the **percentage change** in transit trips resulting from a service modification, and apply this percentage as a factor on the observed existing ridership level to obtain the estimated new ridership, as follows:

$$P^* + P^* \left[\frac{\hat{P}_c - \hat{P}_o}{\hat{P}_o} \right]$$

where:

P^* is pre-change actual ridership

$\hat{P}_o$ is pre-change predicted ridership (from RPFM)

$\hat{P}_c$ is post-change predicted ridership (from RPFM)

This means running the model twice: once to predict ridership under existing service conditions (to the extent the prediction differs from actual), and again under the revised service conditions.

5.1.3 Testing and Validation of Automated Method

The Tri-Met study team hoped to validate the RPFM model by testing it on some typical local service changes. However, very few service changes were found to have occurred after June 1979, and ridership data by line prior to that time was relatively poor. The few post-June 1979 service changes that did occur were either extremely minor, or were such radical realignments that considerable recoding of the input dataset, as well as model recalibration, would have been required.

As a result, the RPFM model was tested on its ability to replicate current ridership on several existing lines. Six lines were selected, which represent a cross-section of Tri-Met's service area and line types. They included:

- Line 2--a long line serving residential areas, industrial employment, retail areas and the downtown
- Line 12--one end of this double-ended line serves the downtown, while the other end serves a shopping mall with residential areas between
- Line 34--connects downtown Portland with two major suburban centers and passes through residential areas
- Line 46--connects the downtown with a major shopping mall and serves residential areas
- Line 53--a relatively short line which connects a high-density residential area with the downtown
- Line 71--a crosstown line connecting a major industrial employment area with residential areas

The RPFM was calibrated to replicate ridership patterns on Line 53, and then tested to see how well it could replicate ridership on the other lines in the test group. Predicted trips from the model were compared with quarterly passenger counts made by bus operators on the respective routes. Results of the tests are shown in Tables 5-1, 5-2, 5-3, and 5-4, for HBW, HBO, NHB and total trips, respectively. The path and route selection subcomponent modules of the RPFM were not individually tested or validated. These relationships were developed from hypotheses about travel choice and not empirical data; therefore there was no test data against which the models could be run.

In the tests, the number of trips predicted by the RPFM was always greater than the actual number of trips on the given line. There are two factors partially responsible for this. First, the predictions represent the total number of transit trips by purpose, i.e., they have not been split between the line of interest and competing lines such as occurs with steps 2 and 3 of the RPFM. Second, the route selected for calibration (Line 53), was the most active (heavily used) line in the Tri-Met system. This line is a radial route serving a densely populated area near the downtown, with a high proportion of medium-low income people; in other words, this route serves the ideal transit market. A more rigorous and extensive test would have recalibrated the model for similar line types or area types, but the current study ran short of time and resources.

Apart from the identifiable biases, however, the model does not predict existing ridership with great accuracy. The model

Table 5-1

ROUTE PATRONAGE FORECASTING MODEL (Automated Version)
COMPARISON OF PREDICTED VS. ACTUAL RIDERSHIP

Home-Based Work Trips

Route	Actual Trips	Predicted Trips	Difference (Pred - Act)	% Difference
53	8,663	8,939	276	63%
2	12,110	14,797	2,687	22%
12, #1	3,351	5,204	1,853	35%
12, #2	3,802	5,870	2,068	54%
34	8,005	8,971	966	12%
46	810	1,951	1,141	141%
71	2,063	4,068	2,005	97%
			1,787	64%

Table 5-2

ROUTE PATRONAGE FORECASTING MODEL (Automated Version)
COMPARISON OF PREDICTED VS. ACTUAL RIDERSHIP

Home-Based Other Trips

Route	Actual Trips	Predicted Trips	Difference (Pre - Act)	% Difference
53	4,829	5,061	232	5%
2	6,605	6,582	-23	-.3%
12, #1	2,426	2,820	394	16%
12, #2	2,590	3,029	439	17%
34	3,813	5,188	1,375	36%
46	682	1,146	464	68%
71	2,188	2,548	360	16%
			509	26%

Table 5-3

ROUTE PATRONAGE FORECASTING MODEL (Automated Version) COMPARISON OF PREDICTED VS. ACTUAL RIDERSHIP

Non-Home-Based Trips

Route	Actual Trips	Predicted Trips	Difference (Pred - Act)	% Difference
53	1,768	1,860	92	5%
2	1,749	2,062	313	18%
12, #1	832	1,229	397	48%
12, #2	855	1,293	438	51%
34	1,569	1,644	75	5%
46	150	551	401	267%
71	409	847	438	107%
			344	83%

Table 5-4

ROUTE PATRONAGE FORECASTING MODEL (Automated Version)
COMPARISON OF PREDICTED VS. ACTUAL RIDERSHIP

Total Trips

Route	Actual Trips	Predicted Trips	Difference (Pred - Act)	% Difference
53	15,260	15,860	600	4%
2	20,464	23,441	2,997	15%
12, #1	6,609	9,253	2,664	40%
12, #2	7,247	10,192	2,945	41%
34	13,387	15,803	2,416	18%
46	1,642	3,648	2,006	122%
71	4,660	7,463	2,803	60%
			2,632	49%

predicted ridership on Line 53 within 4 percent, which is to be expected. However, it overpredicted ridership by as much as 141 percent on Line 46, a residential-to-downtown service, and 97 percent on Line 71, the crosstown route. Overall, the model overpredicted total transit trips by 49 percent. For home-based work trips the model overpredicted by 64 percent, for home-based other trips it overpredicted by 26 percent, and for non-home-based trips, it overpredicted by 83 percent. There is a pattern to the overpredictions, as the model overpredicts total trips by about 2,000 for each of the lines. It is believed that an overprediction with this consistency may reflect a bias deriving from the calibration on Line 53. A good follow-on test would be to recalibrate the model on a less active line.

While a replication type test may appear to be a fundamental way of validating a model, it may not be entirely indicative of the model's true capability. The primary application of the model will be to predict **changes in patronage** due to **changes in service.** Predicting **absolute patronage** is a far more demanding requirement than computing percentage changes, for any forecasting model. In general, more testing is required before the full capability and behavior of the model will be known.

5.1.4 Resources Required to Operate Automated Method

It is difficult to accurately project the time and effort required to operate the automated version of the RPFM. Time estimates vary greatly depending on the particular application and the availability and condition of the data. There is also the problem that the experience with RPFM at Tri-Met has been largely developmental, and hence not as efficient as if the process had been formalized.

The automated RPFM is fairly data intensive, and hence considerable effort is required for its initial set-up. However, once the basic data file has been established (see ROOTINFO discussion in Appendix B-1), the additional effort to run analyses with RPFM is modest. For small changes in service, the analysis can be localized (study area narrowed), thus minimizing the computational requirements of analyzing systemwide effects. Also, once the baseline conditions in ROOTINFO have been specified, parameters related to the transit changes must be revised before another run. Clearly, the automated RPFM must be viewed as an investment in a process to be used many times.

The automated data base is large. For a 242-zone system like Portland, 50 cylinders of storage are required on an IBM 3330 Mod 2 disk pack (the number of records required is approximated by $n(n-1)/2$). Most of the requirements for the data base are met through the regional UTPS highway and transit networks and summary (HBW, HBO, NHB) trip tables. Based on the Tri-Met experience, it appears that with the network and trip tables on hand, the basic database can be assembled in about 2 weeks. The most time-consuming activity is the development of the alternative transit paths between each O-D, as discussed in Section 5.1.1.

The time and effort to run the automated RPFM once setup has been accomplished varies predictably with the application. For simple service changes, results can be seen in 1-3 hours. For the most complex changes, such as estimating the impacts of land use changes or major realignments, it can be necessary to reestimate the total person trip tables, which can require several days of effort. A listing of the range of application tasks and the relative effort involved is summarized below:

RPFM Operating Requirement	Applicable Analysis	Approximate Effort
Re-estimate Total Person Trip Table, or Alter Zone Characteristics	Change in land use Major Realignments	Several days
Alter Impedance Data in ROOTINFO	Fare changes Certain special service changes	1-2 days
Alter Accessibility Data in ROOTINFO	Realignments Land use changes	1 day
Recalibrate Model	Realignments Land use changes	2-3 hours
Change Service Parameters	Change frequency Change travel time Minor alignment changes	1-3 hours

5.2 Manual Method

As described above, the Route Patronage Forecasting Model in its fully automated form is very data intensive. A significant amount of the time and effort involved in applying RPFM to planning tasks is simply in development of the data file, which is performed on the computer. Recognizing that this feature would limit the use of RPFM under conditions of limited time and resources, the Tri-Met study team developed a variation of the technique which is referred to as the "manual version". Some caution should be exercised in this description, however, since this does not mean that RPFM itself is manually operated, but rather that the data preparation tasks are made manual.

The manual method relies on sampling methods to reduce the set of origin-destination zones that are used in the analysis. Typically five or six zone pairs are adequate to perform an analysis. This significantly reduces the effort in the subsequent delineation of alternate travel paths and processing of travel data necessary to set up and run RPFM. Thus, while the RPFM model itself is still operated on the mainframe computer, the manually-developed data file reduces both the set-up time as well as the computer operating time in responding to a planning problem.

There are three basic steps in the manual approach: (1) selection of the zone pair sample; (2) creation of the data file; and (3) calibration and application of the model. Steps (1) and (2) are described below in Section 5.2.1, and are illustrated in the process chart in Figure 5-1. Application and testing of the RPFM manual version on local data is discussed in Section 5.2.2, and a summary of the time and effort to apply the manual RPFM is presented in Section 5.2.3.

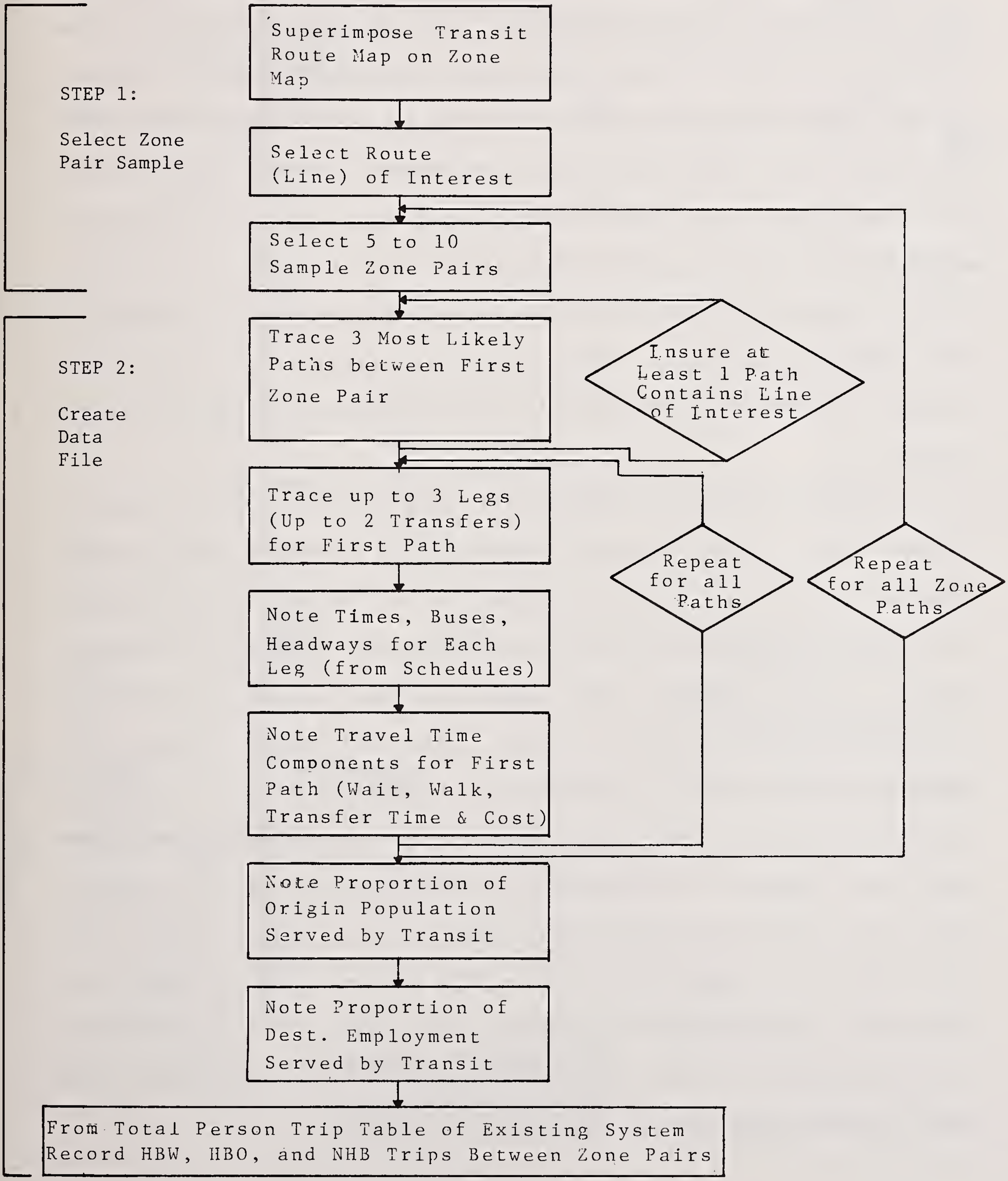
5.2.1 Dataset Development and Data Processing Procedures

(1) Selection of Zone-Pair Sample

The objective in the manual method is to select a small, but representative, sample of origin-destination zone pairs to reflect the range of behavioral reactions to a proposed route

Figure 5-1

PROCEDURE FOR DATA BASE DEVELOPMENT
MANUAL METHOD



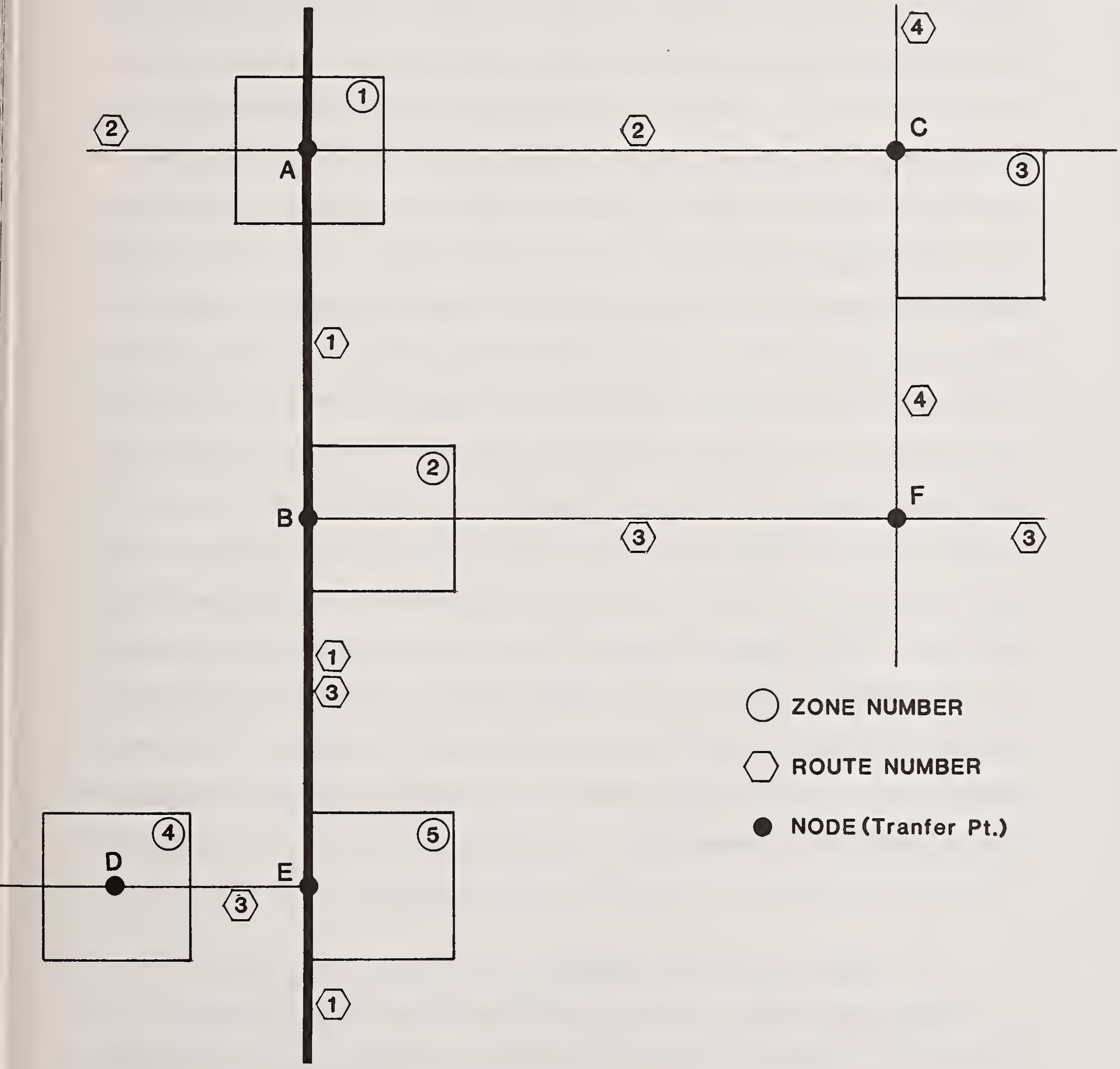
change. Clearly, the concern and skill is in identifying those zone pairs that best represent the travel environment.

A system of rules and procedures has been developed to help guide the zone selection process. The procedure begins by categorizing origin-destination zone pairs which contribute ridership to the route in question into three groups: (1) those where both the origin and destination zones are directly served by the line, (2) those where only the origin or destination zone is directly served by the line, implicitly requiring at least one transfer, and (3) those where neither zone is served directly, implying at least two transfers. These characteristics are determined through inspection of existing transit origin-destination trip tables.

This designation of zone-pair categories is illustrated by Figure 5-2. In the figure, route [1] is the line of interest, represented by the heavy vertical line. Boxes [1] through [5] are travel zones, which may be either trip origins or destinations. In the example, zone pairs [1]-[2], [1]-[5], and [2]-[5] are all directly served and fall into the first group. Zone pairs [1]-[4], [2]-[4], [2]-[3] and [3]-[5] all fall into the second group, i.e. only one trip end is directly served and at least one transfer is necessary to complete the trip by transit. Zone pair [3]-[4] is the only pair in the third group, i.e., where neither trip end is directly served, and at least two transfers are necessary to complete the trip. Pairs [5]-[4] and [1]-[3] have no trips served, and hence, fall into none of the above categories.

FIGURE 5-2

SAMPLE SELECTION DIAGRAM



ROUTE OF INTEREST 1

ZONES SERVED DIRECTLY
BY ROUTE OF INTEREST ① ② ⑤

The procedure for selecting the sample of zone pairs is as follows: After first superimposing a transit route map on the network of travel zones, the user lists all individual zones which receive direct service from the line of interest. One of these zones that is either served exclusively or predominately by the line of interest (such as a highly residential area or strictly industrial area) is then chosen to develop a profile of the major zone pairs that could use the line. Using the selected zone as a destination and an existing zone-to-zone transit trip table, a list is developed of zone pairs with high trip volumes. From this listing and a route map, several zone pairs are then selected in such a manner as to represent both the largest proportion of trips as well as the geographic range of trips. One or two other zones that are high trip generators are also selected, and the entire process is repeated. From the overall listing, a final set of approximately five zone pairs is selected as the sample. The actual number of pairs selected depends on various factors, including the length of the line, variations in conditions along the line, number of activity centers served, and the judgment of the analyst. However, the sample should be kept as small as possible to expedite the analysis.

(2) Creating the Data File

Once the sample of zone pairs is identified, the data file necessary to operate RPFM is developed according to the procedure outlined in Figure 5-1. The first step is the determination of transit paths and legs for each zone pair. A "path" is defined as one or more segments of a route or routes which, when linked together, comprise a line of travel between two zones. A "leg" is one of the above-mentioned segments of a route. Using the previous Figure 5-1, c-a-b-e and c-f-b-c are examples of paths connecting zone 3 and 5, whereas c-a and a-b-e are examples of legs which comprise path c-a-b-e.

First, using a route map, all feasible paths between the zone pairs in the sample are traced. Up to three paths are identified, one of which must be the line of interest. Once the paths and component legs are determined, the number of routes traversing each leg is established, along with the corresponding in-vehicle time and headway. All transit performance measures are developed manually: travel time and headway information are taken from schedules, and determination of average travel time and headway in the event of service redundancies are all done by hand. Finally, all information on the zone pairs, transit paths and legs, and associated travel times are placed into the format required by ROOTINFO, the utility program which runs RPFM on the mainframe computer. All other data necessary to run RPFM, including auto performance data, socioeconomic data, and accessibility data, are already present within ROOTINFO and need not be redeveloped from scratch.

5.2.2 Model Application and Testing

Once the sample selection and data file development are complete, operation of RPFM is virtually identical to the automated method. The same guidelines regarding model calibration and application relative to the type of service change under consideration as described in Section 5.1.2 for the automated method still apply. The only difference is that the model is run against the sample zone pair dataset instead of the complete ROOTINFO dataset. What this means is that the model predictions will reflect only trips between the sample zones. As with the automated method, the manual method is most effective if used to predict percentage change in ridership, which is then applied as a factor to base ridership to forecast the change. If the user desires to predict total ridership directly from the model with the sample data, it is necessary to develop expansion factors for the sample. This process is described in Appendix B-5.

To test the manual method, three test lines were used: Lines 2, 53, and 71. These lines were also test lines for the automated method so that the accuracy of the two methods of generating data sets and applying RPFM may be compared.

As described earlier, Line 2 is a radial line from North Portland to the downtown core, Line 53 is a relatively short line that originates in northwest Portland, a densely populated area close to the CBD, and ends in the CBD, and Line 71 is a relatively long crosstown line which provides service between a mostly residential area and an industrial center, and never goes to the CBD. Because of some special characteristics of the crosstown line--that it is a comparatively long line, not highly focused, and with a more heterogenous population served than usual--six zone pairs were used to represent the 71 line, somewhat more than the usual sample. Lines 2 and 53 have more typical characteristics--moderate coverage, highly focused on one activity center

(CBD), and a fairly homogenous population of users. Thus five zone pairs were considered an adequate sample. The sampling rates for each of the cases are summarized below.

LINE NO.	NO. OF ZONES SERVED	NO. OF SAMPLE PAIRS
2	19	5
53	7	5
71	17	6

The model results summarized in Table 5-5 show that the model is much better at predicting total ridership than at splitting ridership by trip purpose. The model did best on Line 2 where the service characteristics and travel patterns are relatively simple. Line 71 was the furthest off, both in total ridership and in forecasting trips by purpose. This can be partly attributed to the difficulty of determining travel patterns on a crosstown line and the geography of the industrial center the line serves. Because of its isolated location (on a peninsula with one access road) it is perceived as inaccessible and thus may have a lower mode split than the rest of the region. Selecting more zone pairs for the sample for this line may have diminished the problem with over-prediction. In any case, when using a sample of zone pairs to estimate ridership, the model does better at forecasting typical routes with easily defined travel patterns.

Table 5-5

ROUTE PATRONAGE FORECASTING MODEL TEST RESULTS FOR MANUAL METHOD

Line No.	Predicted Riders	Actual Riders	Difference (Pred-Act)	Percent Difference
2	3,255	3,550	- 295	- 0.8%
53	4,265	5,730	-1,465	- 26.0%
71	3,299	1,433	-1,866	-130.0%

5.2.3 Resources Required to Operate Manual Method

The time needed to build a data file manually depends on the number of zones served by the line and the number of zone pairs selected. The average time by activity for the routes tested here are:

Sample selection	1 hour
Path determination	2 1/2 hours
Specification of ROOTINFO	2 1/2 hours
Determine expansion factors	1 1/2 hours
	7 1/2 hours

These are liberal estimates of the time actually taken. Line 71 took the longest, a total of 8 hours, while Line 53 took only about 5 hours. If computer support for an agency is minimal, this is a fairly quick means of supplying the necessary data file to run the Route Patronage Forecasting Model.

CHAPTER 6: CONCLUSIONS

The efforts of the Tri-County Metropolitan Transportation District in developing a route-level transit patronage forecasting methodology have been a qualified success. As a result of the study Tri-Met now has access to a set of computer-based models that will allow evaluation of a wide range of transit service improvements, where previously no such capability existed. Before the study, virtually all service planning analyses were done using a simplistic system which applied ridership factors from similar routes to the service hours of change on the study route.

As stated, the results represent a qualified success. First, the test results from the initial model applications indicate that some additional refinement and testing of the models is necessary before they can be used with confidence. Second, the models are based on a mainframe computer and require some effort to apply, a product of making the methodology suit local conditions and accuracy needs.

Initial tests of the (automated) RPFM on local data were rather discouraging, with the model overpredicting by about 64 percent on HBW trips, 26 percent on HBO trips, 83 percent on NHB trips, and 49 percent overall. The reason for this is believed to be a matter of having calibrated the model on a route with above-average ridership characteristics. Further testing is necessary before the full capability of the model is known.

With regard to the form of the final patronage forecasting model, it is quite clear that the eventual model is much more elaborate and complex than that which was first anticipated. Some simple econometric "direct-demand" type models were originally planned, which could be used manually with data on-hand from the census, transit schedules, and travel surveys. However, it was judged that such simplistic techniques could not adequately address the phenomenon of route interconnectivity and traveler choice among routes, a problem that was seen to be especially prevalent in Portland. The resultant RPFM method, with its individual modules to forecast choice of path and route, was seen as the only way to effectively address this planning issue.

In its fully developed or automated version, the RPFM operates on a large mainframe computer and extracts data from existing Urban Transportation Planning System data files. Extensive initial data processing is required to develop the necessary data file, although special utility programs have been developed to reduce the human effort in this task. These programs also have the capacity to retain baseline information from earlier tests, so that subsequent trails require only modest amounts of new data. Analysis may also be restricted to a reduced area of influence within the system. The system is viewed as an investment with relatively high initial set up costs, which are offset by modest variable costs on subsequent operational runs.

To further reduce costs and improve freedom and spontaneity in applying RPFM, a short cut version was developed which uses a manually prepared dataset based on only a subset of the origin-destination zone pairs in the study area. This approach uses the same RPFM, with some as yet unresolved bias issues, but with considerably reduced set-up and operating time.

Additional testing and development work is planned for both automated and manual methods.

APPENDIX A

MODEL DERIVATION

A-1. Specification, Estimation and Modification of Zone to Zone Trip Generation Equations

The Route Patronage Forecasting Model's (RPFM) first step is to predict the number of transit trips taken between origin-destination zone pairs in the travel simulation network. The equations used to make these predictions were derived from travel behavior models developed previously by the Metropolitan Service District (MSD) for the Portland region. These models are currently used for long-range regional travel predictions.[1]

Development of the trip generation models was a two-step process: first the development of basic models designed to predict transit trip volumes from raw data inputs; second, modification of these models to eliminate destination choice and run directly off trip table estimates, thereby improving transferability. Separate equations were specified and estimated for each of three purposes: HBW, HBO, and NHB. The three equations have similar forms and interpretations, the derivation of which is first described in **general** form, and then for each trip purpose as a **specific** type of the general form.

The expected number of daily trips that a person (labeled r) living in zone i will take by transit to zone j is:

$$(1) \qquad N_{ij}^{r} = TR^{r} \cdot R_{ij}^{r} \cdot P_{ijt}^{r} ,$$

---

[1] Metropolitan Service District, **Regional Transportation Plan** (forthcoming), Appendix 2, "Assumptions and Models: Travel Demand Models."

where

TR^r = person's daily trip rate (i.e., number of trips taken per day);

R^r_{ij} = probability that the person will travel from i to j, given that he/she takes a trip;

P^r_{ijt} = probability that the person will take transit, given that he/she travels from i to j.

The individual travel models estimated by MSD specified the probabilities R^r_{ij} and P^r_{ij}. In particular, the probability of choosing destination j was specified to be logit:

$$(2) \qquad R^r_{ij} = \frac{e^{W^r_{ij}}}{\sum_{\ell} e^{W^r_{i\ell}}}$$

where W^r_{ij} is the utility (or satisfaction) the person derived from traveling to zone j, given that he/she lives in zone i. This W^r_{ij} depends on the attractions in zone j as well as the difficulty of traveling to zone j.[1] The summation in the denominator is taken over all zones that the person could travel to.

Similarly, the probability that the person chooses transit, given that he is traveling from i to j, is specified as logit:

$$(3) \qquad P^r_{ijt} = \frac{YT^r_{ij}}{YT^r_{ij} + Y^r_{ij}}$$

---

[1] In Metro's models of destination choice, upon which W_{ij} is based, the difficulty of traveling to a zone is measured by the auto in-vehicle time required to drive to the zone.

where YT_{ij}^{r} is the utility to the person of traveling by transit between i and j, and YA_{ij}^{r} is the utility to the person of traveling by auto between i and j. These utilities are better known as "travel impedance", calculated as weighted sums of cost, in-vehicle time, wait time, and walk time.

Substituting (2) and (3) into (1), the expected number of trips taken by person r from i to j by transit may be expressed as:

$$(4)\qquad N_{ij}^{r} = TR \cdot \left(\frac{e^{W_{ij}^{r}}}{\sum_{\ell} e^{W_{i\ell}^{r}}} \right) \cdot \left(\frac{e^{YT_{ij}^{r}}}{YT_{ij}^{r} + YA_{ij}^{r}} \right)$$

The above model applies to individual travelers. To obtain the **total** number of trips from i to j, the sum is obtained over all travelers who are available to take trips from i to j. That is,

$$(5)\quad N_{ij} = \sum_{r \varepsilon S_{ij}} N_{ij}^{r} = \sum_{r \varepsilon S_{ij}} TR^{r} \cdot \left(\frac{e^{W_{ij}^{r}}}{\sum_{\ell} e^{W_{i\ell}^{r}}} \right) \cdot \left(\frac{e^{YT_{ij}^{r}}}{e^{YT_{ij}^{r}} + e^{YA_{ij}^{r}}} \right),$$

where S_{ij} is the set of people who might take trips from i to j.

If it is assumed that TR^{r}, W_{ij}^{r}, YT_{ij}^{r}, and YA_{ij}^{r} are the same for all people who might travel from i to j, and equal to TR, W_{ij}, YT_{ij}, and YA_{ij}, then the sum expressed in (5) becomes:

(6) $$N_{ij} = M_{ij} \cdot TR \cdot \left[\frac{e^{W_{ij}}}{\sum\limits_{\ell} e^{W_{i\ell}}}\right] \cdot \left[\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}}\right],$$

where M_{ij} is the number of people who are available to take trips from i to j. Consolidating terms, this becomes:

(7) $$N_{ij} = M_{ij} \cdot TR \cdot \frac{e^{W_{ij}+YT_{ij}}}{\left(\sum\limits_{\ell} e^{W_{ij}}\right)\left(e^{YT_{ij}} + e^{YA_{ij}}\right)}$$

Taking logs, the expression becomes:

(8) $$\ln N_{ij} = \ln M_{ij} + \ln TR + (W_{ij} + YT_{ij}) - \ln\left[\left(\sum\limits_{\ell} e^{W_{ij}}\right)\left(e^{YT_{ij}} + e^{YA_{ij}}\right)\right]$$

To account for individual differences (that is, W_{ij}, YT_{ij}, etc. are not the same for all people), and since the individual choice models are not perfectly accurate, parameters are added to this equation so as to allow for behavioral "shifts" among individuals:

(9) $$\ln N_{ij} = \alpha_1 + \beta \ln M_{ij} + \alpha_2 \ln TR + \theta(W_{ij} + YT_{ij}) - \phi \ln\left[\left(\sum\limits_{\ell} e^{W_{ij}}\right)\left(e^{YT_{ij}} + e^{YA_{ij}}\right)\right] .$$

This is the general form of the equation that is used for prediction of zone-to-zone transit trips. For each trip purpose, this equation was specified to represent that particular purpose, and then the parameters of the equation were estimated. The following subsections discuss the equations for each of the three purposes.

A-1.1 Home-Based Work Trips

In the short-run (for which the forecasting method is intended), the location of residences and work places is fixed. From an individual worker's point of view, the choice of destination is fixed: he/she will necessarily travel from home to the workplace. Therefore, the term R^r_{ij} in (1) is necessarily 1 for a person living in zone i and working in zone j, and is necessarily zero otherwise. Therefore, for HBW trips, (6) becomes:

$$\text{(6HBW)} \qquad N_{ij} = M_{ij} \cdot TR \cdot \left[\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}} \right],$$

where M_{ij} is now the number of people who live in zone i and work in zone j. Since $M_{ij} \cdot TR$ is, by definition, the number of work trips (both auto and transit) taken by people living in i and working in j, the above equation becomes:

$$\text{(6HBW')} \qquad N_{ij} = Q_{ij} \cdot \left[\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}} \right],$$

where Q_{ij} is the number of work trips taken by people living in i and working in j. Taking logs gives the HBW version of (8):

$$\text{(8HBW)} \qquad \ln N_{ij} = \ln Q_{ij} + YT_{ij} - \ln \left[e^{YT_{ij}} + e^{YT_{ij}} \right] .$$

Adding shift parameters gives the HBW version of (9):

$$\text{(8HBW)} \qquad \ln N_{ij} = \alpha_1 + \beta \ln Q_{ij} + \theta\, YT_{ij} - \phi \ln \left[e^{YT_{ij}} + e^{YA_{ij}} \right]$$

The parameters of this equation were estimated from data on-hand at Tri-Met. The number of HBW trips from zone i to zone j was derived from an on-board survey conducted by Tri-Met in 1980. The raw counts were "factored-up" by Tri-Met to reflect the proportion of trips sampled on each route and during each time of day. The number N_{ij} that was used as the dependent variable for the HBW equation included the number of one-way trips from home in zone i to work in zone j **and** the number of one-way trips from work in zone j to home in zone i. (Note that zone i is the home zone and zone j is the work zone for both trips.)

The total number of HBW trips by auto and transit between home in i and work in j (that is, Q_{ij}) was calculated as the sum of N_{ij} and the number of HBW auto trips from i to j, with the auto trip counts obtained from an origin-destination survey in 1977.

Time and cost data for auto and transit and socioeconomic data for each zone were obtained from the Metropolitan Service District's regional travel models. These data were used to calculate YT_{ij} and YA_{ij}, which represent the utility or impedance of travel from i to j by transit and auto, respectively. The formulae for the terms YT_{ij} and YA_{ij} are given in Figure A-1.

There are 246 zones in the Portland area, giving rise to 60,516 zone pairs. However, only some of these zone pairs were used as observations in estimation. Specifically, the study team felt that if the "factored-up" on-board survey indicated that eight or fewer HBW trips were taken from one zone to another, then the figure was (a) unreliable and (b) too small to represent systematic behavior. Consequently, a zone pair was not included in the estimation if the number of HBW transit trips between the home zone and the work zone were eight or less. This restriction, along with estimations due to missing data, left 1805 observations for estimation of the parameters in equation (9HBW).

Using ordinary least squares regression, the following parameter estimates were obtained for the HBW model:

Parameter	Estimate	t-Statistic
$\alpha 1$	.885	10.38
β	.642	39.95
θ	.333	16.15
ϕ	.665	29.99

(R^2 = .986)
(MSE = .2874)

Figure A-1

CALCULATION OF TRANSIT AND AUTO TRAVEL IMPEDANCES HOME-BASED WORK TRAVEL

$$YT_{ij} = -\ .1619 \times Walk_{ijt} - .0073 \times CST_{ijt} - .0311 \times lV_{ijt}$$
$$- .0528W1_{ij} - .078W2_{ij} + 1.4 \times CBD_i$$

$$YA_{ij} = -\ .1619 \times Walk_{ija} - .0073 \times CST_{ija} - .0813 \times lV_{ija}$$
$$+ 1.1 \times A1_i + .975 \times A2_i$$

where:

YT_{ij} = impedance of traveling by transit from home in zone i to work destination in zone j;

YA_{ij} = impedance of traveling by auto from home in zone i to work destination in zone j;

$Walk_{ijm}$ = walk time from zone i to zone j, using mode m (m = a for auto; m = t for transit);

CST_{ijkm} = cost of travel from zone i to zone j, using mode m;

lV_{ijm} = in-vehicle time from zone i to zone j, using mode m;

$W1_{ij}$ = time waiting for first carrier for transit travel from zone i to zone j;

$W2_{ij}$ = time waiting at a transfer for transit travel from zone i to zone j;

CBD_i = a dummy variable indicating zone i is in Portland's CBD;

$A1_i$ = proportion of population in zone i with one auto available;

$A2_i$ = proportion of population in zone i with two or more autos available.

Source: Metropolitan Service District HBW Mode Choice Model. These measures were obtained from the HBW mode choice model estimated by Metro.

Modification

Following the original estimation, several modifications were made to the HBW model in an attempt to make it more accurate. The modification consisted of adding "accessibility" measures--percent served by transit in the origin and destination zones--to the model structure. Using the same notation as in the original derivation, the revised equation for predicting HBW transit trips is specified as follows:

$$(6HBW'^*) \quad N_{ij} = Q_{ij} \cdot R_i \cdot S_j \cdot \left[\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}} \right]$$

where:

N_{ij} = the number of HBW transit trips from home in zone i to work in zone j;

Q_{ij} = the number of HBW trips (both auto and transit) taken from home in i to work in j;

R_i = the proportion of population in i served by transit;

S_j = the proportion of employment in j served by transit; and

YT_{ij} and YA_{ij} are transit and auto impedance, respectively.

Specifically, $Q_{ij}R_iS_j$ is the number of HBW trips that have the option of using transit.

Taking logs:

$$(8HBW^*) \quad \ln N_{ij} = \ln (Q_{ij} \cdot R_i \cdot S_j) + YT_{ij} - \ln (e^{YT_{ij}} + e^{YA_{ij}}),$$

And adding shift parameters yields the following expression:

$$(9HBW^*) \quad \ln N_{ij} = \alpha' + \beta \ln (Q_{ij} \cdot R_i \cdot S_j) + \theta \, TY_{ij} - \phi \ln(e^{YT_{ij}} + e^{YA_{ij}})$$

This is the equation to be used for predicting the number of HBW transit trips from zone i to zone j. Note that this equation is the same as the previous (9HBW), but with $\ln Q_{ij}$ replaced by $\ln (Q_{ij}R_iS_j)$. The shift parameters were not re-estimated. Instead, the model uses the previously estimated values of β, θ, and ϕ, with the intercept term, α', adjusted to account for the replacement of $\ln Q_{ij}$ by $\ln (Q_{ij}R_iS_j)$, or:

$$\alpha + \beta \ln(Q_{ij}R_iS_j) = \alpha + \beta \ln Q_{ij} + \beta \overline{\ln R_iS_j} = (\alpha + \beta \overline{\ln R_iS_j}) + \beta \ln Q_{ij}$$

where the bar denotes the mean. Therefore, the new intercept, α', is equal to the previously estimated intercept α minus $\beta \overline{\ln R_iS_j}$, or -.098.

The study team relied upon an econometric identity in electing to not re-estimate the HBW equation following the addition of R_i and S_j. The econometric identity states that if a variable whose coefficient in a true equation is not zero and ordinary least squares (OLS) is applied to an equation that omits this variable, then the coefficients of the included variables are estimated without bias if the omitted variable is uncorrelated with the included variables. In this case, it was argued that the revised equation (9HBW*) is the "true" expression for HBW travel, while the original HBW equation (*8HBW) was the true equation with an omitted variable.

A-1.2 Home-Based Other Trips

In developing the original HBO model, it was assumed that only people who were served by transit would make non-work trips by transit. A person was considered to be served by transit if he/she lived within a quarter-mile of a transit line.

Using this assumption, the summation in equation (5) is taken, for HBO trips, over all people served by transit in the origin zone. Unlike work trips, the destinations of non-work trips are not fixed, even in the short-run. Therefore, the HBO version of equation (6) is:

$$(6\text{HBO}) \qquad N_{ij} = M_i \cdot TR \cdot \left(\frac{e^{W_{ij}}}{\sum_{\ell} e^{W_{i\ell}}} \right) \cdot \left(\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}} \right) ,$$

where:

N_{ij} = the number of HBO trips taken between home in zone i and a non-work location in zone j , and

M_i = the number of people in zone i who are served by transit (that is, live within a quarter-mile of a transit line).

Taking logs and adding shift parameters gives the HBO version of equation (9):

$$(9\text{HBO}) \qquad \ln N_{ij} = \alpha_1 + \beta \ln M_i + \alpha_2 \ln TR + \theta (W_{ij} + YT_{ij})$$

$$- \phi \ln \left[\left(\sum_{\ell} e^{W_{ij}} \right) \left(e^{YT_{ij}} + e^{YA_{ij}} \right) \right] .$$

The parameters of this equation were estimated by ordinary least squares regression, with observations eliminated if eight or fewer HBO trips were taken between the zones, or if data were missing for any of the variables. Note that, analogous to the HBW regression, N_{ij} includes the number of one-way trips from home in zone i to non-work locations in zone j and the number of one-way trips from non-work locations in zone j to home in zone i.

The population served by transit in each zone, i.e., M_i, and the time and cost data for auto and transit, attraction data for each zone, and socioeconomic data for each zone were obtained in a manner equivalent to the HBW model. These data were used to calculate:

- W_{ij}, which represents the utility of choosing to travel from home in zone i to a non-work location in zone j. This measure is obtained from MSD's HBO model of destination choice, and depends on highway travel time from i to j and the attractions in j. The function used to calculate W_{ij} is given in Figure A-2.
- YT_{ij}, which represents the impedance for travel from i to j by transit. This measure is obtained from MSD's HBO model of mode choice and is a weighted sum of transit times and costs. The function used to calculate YT_{ij} is given in Figure A-3.

Figure A-2

CALCULATION OF DESTINATION CHOICE UTILITY HOME-BASED OTHER TRAVEL

$$
\begin{aligned}
W_{ij} = {} & \ln(ATT_j) - 2.580\sqrt{lV_{ija}} + .1422 \times lV_{ija} \\
& - .6523 \times WE_{ij} - .6843 \times NE_{ij} + .644 \times NW_{ij} \\
& - 1.002WN_{ij}
\end{aligned}
$$

where:

W_{ij} = utility of travelling from home in zone i to a non-work destination in zone j;

lV_{ija} = auto in-vehicle time from zone i to zone j;

WE_{ij} = a dummy indicating that travel from zone i to zone j involves crossing the Williamette River from west to east;

NE_{ij} = a dummy indicating that travel from zone i to zone j involves crossing the Columbia River only, in a southbound direction;

NW_{ij} = a dummy indicating that travel from zone i to zone j involves crossing the Columbia and Williamette Rivers, in a southbound direction;

ATT_j = attraction of zone j, measured as .228 x (population of zone j) + 6.23 x (retail employment in zone j) + .49 x (non-retail employment in zone j).

Figure A-3

CALCULATION OF TRAVEL IMPEDANCE
HOME-BASED TRAVEL

$$YT_{ij} = - .1544 \times Walk_{ijt} - .0156 \times CST_{ijt} - .0297 \times 1V_{ijt}$$
$$- .0504 \times W1_{ijt} - .0744 \times W2_{ijt} + 1.4CBD_{i}$$

$$YA_{ij} = - .1544 \times Walk_{ija} - .0156 \times CST_{ija} - .0755 \times 1V_{ija}$$
$$+ 1.9 \times A1_{i} + .975 \times A2_{i}$$

where:

YT_{ij} = impedance of traveling by transit from home in zone i to work destination in zone j;

YA_{ij} = impedance of traveling by auto from home in zone i to work destination in zone j;

$Walk_{ijm}$ = walk time from zone i to zone j, using mode m (m = a for auto; m = t for transit);

CST_{ijkm} = cost of travel from zone i to zone j, using mode m;

$1V_{ijm}$ = in-vehicle time from zone i to zone j, using mode m;

$W1_{ij}$ = time waiting for first carrier for transit travel from zone i to zone j;

$W2_{ij}$ = time waiting at a transfer for transit travel from zone i to zone j;

CBD_{i} = a dummy variable indicating zone i is in Portland's CBD;

$A1_{i}$ = proportion of population in zone i with one auto available;

$A2_{i}$ = proportion of population in zone i with two or more autos available.

- YA_{ij} , which represents auto travel impedance from zone i to zone j by auto. This measure is also obtained from MSD's HBO model of mode choice, and is illustrated in Figure A-3.

Using these data, initial parameter estimates were obtained for equation (9HBO). Some of these initial estimates had the wrong sign and consequently were constrained to the values that would obtain if the models of individual travel behavior were completely accurate and all people within a zone were the same. The parameters estimated under these constraints are given below. Note that, under the assumption that TR does not vary over zones, the y-intercept of the regression equation is an estimate of the term $\alpha_1 + \alpha_2 \ln TR$.

Parameter	Estimate	t-Statistic
$\alpha_1 + \alpha_2 \ln TR$	-.733	4.36
β	1	(by constraint)
θ	1	(by constraint)
ϕ	.599	22.4

(R^2 = .987)
(MSE = 1.657)

Modification

As with the Home-Based Work model, various modifications were made to the Home Based Other trip model after the original estimation. The purpose of these modifications was twofold: (1) an attempt, as with the HBW model, to improve the realism and accuracy of the model, and (2) to eliminate a site-specific feature of the model, which Tri-Met felt would limit the range of its use, both in Portland and potentially other locations. The site-specific attribute of the original model was encountered as a result of including the MSD destination choice utility (W_{ij}) within the model structure. As seen in Figure A-2, W_{ij} incorporates Portland-specific dummy variables (river crossing measures) to allow calibration. The revised HBO model relies instead on zone-to-zone trip tables, which implicitly incorporate traveller's destination choice. The revised equation for predicting HBO transit trips is derived as follows. The number of HBO transit trips from i to j (N_{ij}) is defined as:

$$(6HBO^*) \qquad N_{ij} = P_i \cdot TR \cdot H_{ij} \cdot R_i \left[\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}} \right]$$

where P_i is the population of zone i;

TR is the trip rate;

H_{ij} is the number of trips taken to zone j, as a proportion of trips taken to all zones;

R_i is the proportion of population served by transit in zone i; and

YT_{ij} and YA_{ij} are transit and auto impedances, respectively.

Note that by definition:

$$H_{ij} = Q_{ij}/\sum_{\ell} Q_{i\ell}$$

where $Q_{i\ell}$ is the number of HBO trips from i to ℓ. Note also that $P_i \cdot TR = \sum_{\ell} Q_{i\ell}$; that is, the total number of HBO trips originating in i . Therefore:

$$P_i \cdot TR \cdot H_{ij} = \sum_{\ell} Q_{i\ell} \cdot \frac{Q_{ij}}{\sum_{\ell} Q_{i\ell}} = Q_{ij} \quad ,$$

and (6HBO*) becomes:

$$N_{ij} = Q_{ij} \cdot R_i \cdot \left(\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}} \right)$$

Taking logs:

$$\ln N_{ij} = \ln(Q_{ij}R_i) + YT_{ij} - \ln\,(e^{YT_{ij}} + e^{YA_{ij}})$$

And adding shift parameters yields:

$$(9HBO^*) \quad \ln N_{ij} = \alpha + \beta \ln\,(Q_{ij}R_i) + \theta\, YT_{ij} - \phi \ln\,(e^{YT_{ij}} + e^{YA_{ij}})$$

which is the equation to be used for predicting HBO transit trips. Note that it is the same as the old equation (9HBO) except:

(a) $\ln M_i$ (where M_i is population served by transit) is replaced with $\ln Q_{ij}R_i$ (where $Q_{ij}R_i$ is the trips that can be taken by transit).

(b) W_{ij}, which denotes the destination choice utility and includes the river crossing dummies, does not enter the equation.

The values of α, β, θ and ϕ are as originally estimated. The study team felt that no adjustment of α was required since including $\ln Q_{ij}R_i$ just compensates for eliminating $\ln M_i$ and the terms including W_{ij}.

A-1.3 Non-Home Based Trips

Equation (1), describing individual behavior, is appropriate for NHB trips if TR is interpreted to be the probability of taking a NHB trip, times the probability of starting the trip in zone i. Then R^r_{ij} is, as before, the probability of choosing destination j and P^r_{ijt} is the probability of choosing transit.

In summing to obtain the total number of NHB trips from zone i to zone j, the summation is over all people in the Portland area, since each person **can** take an NHB trip from any particular zone. The NHB version of (6) therefore becomes:

$$(6\text{NHB}) \qquad N_{ij} = M \cdot TR \cdot \left(\frac{e^{W_{ij}}}{\sum_{\ell} e^{W_{i\ell}}} \right) \cdot \left(\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}} \right)$$

where M is the population of Portland.

Taking logs and adding shift parameters results in:

$$(9\text{NHB}) \qquad \ln N_{ij} = \alpha_1 + \beta \ln M + \alpha_2 \ln TR + \theta (W_{ij} + YT_{ij}) - \phi \ln \left[\left(\sum_{\ell} e^{W_{i\ell}} \right) \left(e^{YT_{ij}} + e^{YA_{ij}} \right) \right] .$$

Data on the number of NHB trips from each zone to each other zone (i.e., N_{ij}), cost and times for travel by transit and auto, attractions in each zone, and socioeconomic variables for each zone were again compiled in a manner analagous to the HBW model. These data were used to calculate W_{ij}, YT_{ij}, and YA_{ij}. The formulae for these terms are given in Figures A-4 and A-5.

Observations with missing data or with eight or fewer NHB trips were eliminated from the estimation sample. Using ordinary least squares, the parameter estimates given below were obtained. Note that since M is the same for all observations (i.e., the population of Portland), and assuming TR is constant over all zones, the y-intercept of the equation is an estimate of $_1$ + $\beta \ln M + \alpha_2 \ln TR$.

Parameter	Estimates	t-Statistic
$\alpha_1 + \beta \ln M + \alpha_2 \ln TR$	4.58	28.99
θ	.285	43.14
ϕ	.293	20.01

(R^2 = .998)
(MSE = .222)

Modifications

The NHB model was modified in the same manner as the HBO model, to operate directly off existing trip tables. The purpose of the change was to improve accuracy and eliminate site-specific characteristics.

Figure A-4

CALCULATION OF DESTINATION CHOICE UTILITY
NON-HOME BASED TRAVEL

$$W_{ij} = \ln (ATT_j) - 2.0228 \times \sqrt{lV_{ija}} + .089 \times lV_{ija}$$
$$- .5422 \times WE_{ij} - 1.1305 \times NE_{ij} + .2598 \times NW_{ij} - .6333 \times WN_{ij}$$

where:

W_{ij} = utility of traveling from home in zone i to a non-work destination in zone j;

ATT_j = .228x(population of zone j) + 6.23x(retail employment in zone j) + .49x(non-retail employment in zone j).

lV_{ija} = auto in-vehicle time from zone i to zone j;

WE_{ij} = a dummy indicating that travel from zone i to zone j involves crossing the Williamette River from west to east;

NE_{ij} = a dummy indicating that travel form zone i to zone j involves crossing the Columbia River only, in a southbound direction;

NW_{ij} = a dummy indicating that travel from zone i to zone j involves crossing the Columbia and Williamette Rivers, in a southbound direction;

Figure A-5

CALCULATION OF TRAVEL IMPEDANCE NON-HOME BASED TRAVEL

$$YT_{ij} = -\ .0436\ Walk_{ijt} - .0317xCST_{ijt} - .0084x1V_{ijt} - .04142xW1_{ij} - .021xW2_{ij}$$

$$YA_{ij} = -\ .0436\ Walk_{ija} - .0091xCST_{ija} - .0218x1V_{ija} + 3.502$$

where:

YT_{ij} = impedance of traveling by transit from home in zone i to work destination in zone j;

YA_{ij} = impedance of traveling by auto from home in zone i to work destination in zone j;

$Walk_{ijm}$ = walk time from zone i to zone j, using mode m (m = a for auto; m = t for transit);

CST_{ijkm} = cost of travel from zone i to zone j, using mode m;

$1V_{ijm}$ = in-vehicle time from zone i to zone j, using mode m;

$W1_{ij}$ = time waiting for first carrier for transit travel from zone i to zone j;

$W2_{ij}$ = time waiting at a transfer for transit travel from zone i to zone j;

CBD_{i} = a dummy variable indicating zone i is in Portland's CBD;

$A1_{i}$ = proportion of population in zone i with one auto available;

$A2_{i}$ = proportion of population in zone i with two or more autos available.

The new equation for predicting NHB transit trips is derived as follows. The number of NHB transit trips from i to j is:

$$(6NHB^*) \qquad N_{ij} = M \cdot TR \cdot H_{ij} . R_i \left[\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}} \right]$$

where all terms are defined analogously to those in (6HBO*) except M, which is the population of Portland. Using the fact that $M \cdot TR \cdot H_{ij} = Q_{ij}$, this becomes

$$N_{ij} = Q_{ij} \cdot R_i \cdot \left[\frac{e^{YT_{ij}}}{e^{YT_{ij}} + e^{YA_{ij}}} \right]$$

Taking logs and adding shift parameters yields:

$$(9NHB^*) \qquad \ln N_{ij} = \alpha + \beta \ln (Q_{ij} R_i) + \theta \, YT_{ij} - \phi \ln (e^{YT_{ij}} + e^{YA_{ij}})$$

This is the equation that will be used for predicting NHB transit trips. Note that it is the same as the old equation (9NHB) except:

(a) ln M is replaced with $\ln Q_{ij} R_i$;

(b) all terms including W_{ij} are eliminated.

The previously estimated values of θ and ϕ are also used in this version of the model. The value of β, which was not estimated in previous equations (since ln M was subsumed into the intercept) is set to its theoretically derived value of one. The intercept of the new equation is set to its theoretically derived value of zero, since the previously estimated value cannot be adjusted without knowing the value of α_2 ln TR in (9NHB).

A-1.4 Special Issues in Estimation and Application of Trip Generation Models

The equations described in Section A-1.1 are used to predict, for each trip purpose, the number of transit trips between each pair of zones. The models contain several important features which should be noted by the user to assure that they are properly applied.

A-1.4.1 Bias Considerations with Log-Linear Models

The trip generation equations in RPFM share the characteristic that the dependent variable is expressed as the logarithm of the number of trips by purpose. This means that the equations, developed through regression, give predictions of N_{ij} rather than N_{ij} directly. In such a "log linear" model estimated by ordinary least squares, with ln x as the dependent variable, an unbiased prediction of x is not obtained by simply taking the exponential of ln x. If x is estimated by the following model:

$$\ln x = \ln \alpha + \beta \ln z + \varepsilon$$

where:

z is an explanatory variable

α and β are estimated parameters, and

ε is the residual error due to regression

the expected value of ln x is equal to $\ln \alpha + \beta \ln z$. However, the expected value of x is not $\exp(\ln \alpha + \beta \ln z)$, since such a transformation introduces a bias in the amount of the error term, ε. Since ε is assumed to be distributed normally, with zero mean and variance of σ^2 (where σ is the standard error of the regression), the expected value of ε can be expressed as $\exp(\frac{1}{2}\sigma^2)$. This means that the expected value of x (E(x)) is more correctly:

$$E(x) = \exp(\ln \alpha + \beta \ln z) \cdot \exp(\tfrac{1}{2}\sigma^2)$$

Therefore, an unbiased estimate of x from ln x can be obtained through application of the adjustment factor $\exp(\frac{1}{2}\sigma^2)$ to the value of x calculated from the regression equation. The value of $\exp(\frac{1}{2}\sigma^2)$ to be used with the trip generation models are:

HBW model	1.1545
HBO model	2.2905
NHB model	1.1171

A-1.4.2 Use of Average Transit Impedance in Trip Generation Models

When estimating the trip generation equations in Section A-1.1, zone-to zone transit costs and times were calculated using the "minimum path", which is defined as the path with the lowest impedance (travel time and cost) connecting the zones. That is, YT_{ij} was calculated with costs and times for the minimum transit path between zones i and j.

If the equations were implemented exactly as they were estimated, using only the minimum path, then changes in transit service along paths other than the minimum path would not affect the predicted number of trips between zones. The effect of the service change is picked up by running the operations using a measure of transit impedance which is averaged over the feasible paths. That is, YT_{ij} is calculated as:

$$YT_{ij} = \sum_{k=1}^{3} P_{ij}^{k} \cdot YT_{ij}^{k}$$

where Y_{ij}^{k} is transit impedance along path k from zone i to zone j, k is the set of feasible paths, and P_{ij}^{k} is the proportion of trips between zones i and j that are predicted to utilize path k. Because the proportions, P_{ij}^{k} are not known in advance, it is necessary to first operate Step 2, the path selection module of RPFM. While Step 2 follows Step 1 logically in the structure of the RPFM model, in process Step 2 is actually performed prior to Step 1 and the results of Step 2 are used in calculating Step 1.

A.1.4.3 Limiting Range of Influence of Transit Route or Service Change

When estimating the trip generation models, zonal pairs were eliminated if the "factored-up" on-board survey data indicated that eight or fewer transit trips of the appropriate purpose were taken between the zones. It is necessary to also reflect this truncation in the application of the equations. The most obvious way to account for this would be to **predict** the number of transit trips **only** for zone pairs between which there were more than eight transit trips in the "factored-up" on-board survey, i.e., for those zone pairs that were used in the estimation. This approach requires the assumption that transit service changes have negligible effect for zone pairs between which there were initially few transit trips. While this assumption is probably accurate for changes in **existing** service, such as probably not accurate for new lines. For example, there may be currently few transit trips between two zones because the transit service between those zones is very circuitous; putting in a direct line between the zones could result in a large number of transit trips.

To allow for this possibility, the equations are applied for all zone pairs between which there are more than eight transit trips **or** more than 50 auto trips. This approach assumes that the effect of changes in transit service would, in fact, be negligible for zone pairs between which there are few transit **and** auto trips, since it is unlikely that many transit trips will be taken, no matter how good the service, if there are few auto trips. For zone pairs between which there are 8 or fewer transit trips **and** 50 or fewer auto trips,* the numbers from the "factored-up" on-board survey are used rather than the predicted number of trips.

---

*Note that the number 50 was chosen as the cut-off for auto trips under the following reasoning. Suppose there were 50 auto trips between two zones and no transit trips because of no transit service. If a transit route between the zones were initiated, and the mode share captured by transit was 15%, then the number of transit trips taken between the zones would be approximately 8, which is the cut-off for bus trips.

APPENDIX B

MODEL APPLICATION

APPENDIX B-1: ROUTE PATRONAGE FORECASTING MODEL INPUT DATASET DEVELOPMENT (ROOTINFO)

B-1.1 Introduction

The key to successful application of the Route Patronage Forecasting Model is the input data file "ROOTINFO." This file contains all of the data elements that relate to factors relevant in making a particular transit trip--given that a certain number of total trips are made between each zone (i-j) pair. Basically, these factors are: (1) **accessibility** (percent of zone population/employment served by transit); (2) **general zonal characteristics** (proportion of households with zero, one, two plus autos); (3) **impedance** (travel times and costs along each path, transit line, and leg of the transit system operating between zones); and (4) **total person trips** by each of three purposes--home-based work, home-based other, and non-home based. Although the total person trip tables are actually entered separately into ROOTINFO as transit and auto trips, the transit trip table is only really necessary if calibration of the model is desired. These data elements are quantified for each origin-destination zone pair to make up the ROOTINFO file.

This section describes the automated procedure and format for setting up the ROOTINFO data file for the entire transit system--all paths, lines and legs serving each zone pair in the region. The process is somewhat involved as it requires a fair amount of data processing using UTPS programs and some special utilities. The resulting complete data set would be worth the effort required if a considerable number of route changes are to be tested. If, however, only one or a few routes are to be tested, there is a more direct and faster way to set up ROOTINFO. This is described in the section, "Manual Method of Application."

Once established, the data file described herein can be used directly to test patronage changes in service frequency (headways) or travel time on existing routes. However, if the Route Patronage Forecasting Model is to be used to test effects of service on a new or altered route, then ROOTINFO should be altered to include the new accessibility and impedance values. For testing those new conditions, the model should also be recalibrated as described in the calibration discussion (Section 5.2.2) of this report.

B-1.2 Data Sources

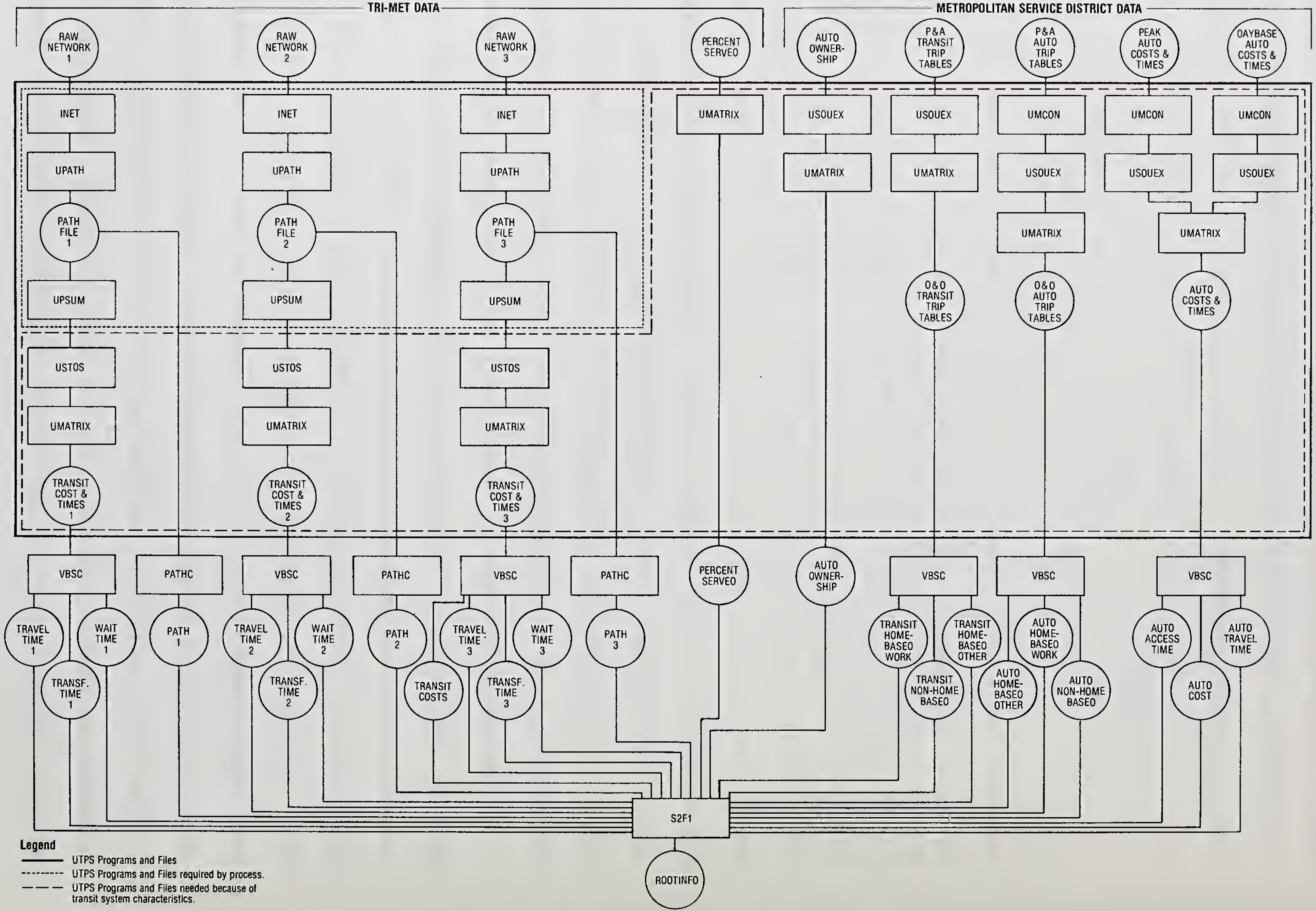
The input data file ("ROOTINFO") to the Route Patronage Forecasting Model is made up of information from a myriad of sources--both from within Tri-Met and from Metro, the local MPO. The transit trip table was factored from a 20,000-record on-board origin-destination survey taken by Tri-Met in May, 1980. The transit network used for developing paths, transit travel, transfer and access times and transit travel costs is a 1980 INET network. The "percent-served" factors were developed from 1981 land use and route maps showing the percentage of population and employment of each zone (tract) within one-quarter mile of the route. Auto ownership factors, auto costs and times and the auto trip tables were proved by Metro. The auto ownership factors were developed from a 1977 origin-destination survey. These survey results were used to factor 1970 network (travel times) and simulated projections from the regional trip generation, distribution modal split model. Auto costs and travel times come from the same network.

B-1.3 Process

This input data is processed through several computer programs to get them into the form needed for the Route Patronage Forecasting Model. The data flow is shown in Figure B-1.1. Many of the steps can be done using programs within the UTPS package. Those steps are enclosed in the solid box. Some of these steps are only necessary for the Tri-Met due to the unique features of the transit network and incompatibilities of the Tri-Met and MSD zone systems and are indicated by the dashed line. Those UTPS steps which are necessary for all applications are encircled by the dotted line. The remaining steps require two utility programs which were developed at Tri-Met. They are "VBCS" and "PATHC" (see section describing utility programs for details).

The UTPS-assisted process for multiple path finding shown on the left-hand side of Figure B-1 is documented in detail in the section on "Multiple Path Determination." Basically, three peak-period INET/UNET networks are developed (Raw Networks 1, 2 and 3). Network 1, in Tri-Met's approach, is the full network with the usual walk/wait time time coefficients used in UPATH. Networks 2 and 3 have different access links turned on and off. UPSUM would normally be the end of the path and cost development process since it produces time and cost files. However, Tri-Met uses transit centers with timed transfers which are not handled

Figure B-1.1
DATA PROCESS CHART FOR ROOTINFO



adequately in the basic process. Therefore, the lines which use these transit centers were identified using USTOS and the transfer times on these lines were reset to a maximum of two minutes using UMATRIX. See section on "Multiple Path Determination" for details.

The UTPS part of the process shown on the right side of Figure B-1 is specific to the transit system. The "percent-served" factors are stored as a UTPS Z File, and UMATRIX is used to convert it into matrix format. The auto trip tables developed by Metro utilize a zone system that is larger than that used by Tri-Met. UMCON is used to eliminate external "cordon zone." USQUEX is then used on both sets of trip tables and the auto ownership data to eliminate some small differences in the two zone systems.[1]

UMATRIX is used on the auto ownership data to change it into matrix format. UMATRIX is also used on the trip tables to change them from production and attraction tables to origin and destination tables. This is accomplished by dividing the matrices in half about the diagonal, transposing them and then adding the original halved matrix to the transposed halved matrix.

The zone system for auto costs must also be altered in a way similar to the zone system for auto trip tables. A UMATRIX run is done to combine peak and daybase times and costs. This is

---

[1] Although in Portland both the highway network and the transit network are based on zone systems corresponding to census tracts, over the years, small differences in the zone systems used for transit vs. highway analysis have "evolved." These differences needed to be reconciled in preparation of data for the ROOTINFO file.

done because the trip tables represent a full day's travel, and although transit travel times and costs do not change significantly between peak and daybase, auto times and costs do. The composite times and costs are factored using 68 percent of the daybase plus 32 percent fo the peak. These percentages are proportional to percentage of trips made by the time period.

The other method of handling this problem would have been to create peak trip tables and only estimate peak ridership. The results could then be factored up to full-day ridership.

This is the extent of the involvement of UTPS in the development process. Since UTPS files are not in a format easily utilized by other programs, two programs had to be developed to reformat the data. Tri-Met's PATHC (path conversion utility) converts the trees built by UPATH into a matrix format. This program is documented in the "Utilities" section. Another Tri-Met program, VBSC (variable block space conversion) takes a multi-tabled UTPS matrix, in any UTPS format, and outputs each table as a separate file in a standard format. The documentation of this program is also included in the "Utilities" section. The resulting tables can then be easily read into the "i-j" format for ROOTINFO.

The three sets of transit travel times and costs are separated into three transit travel time files, three transit wait time files, three transfer time files, and one transit cost file. Only one cost file is created because the data in this file reflect transit fares which do not change with the path used.

The auto cost and time file is separated into auto travel time, auto access time and auto cost. Both the transit and auto trip tables are separated into home-based work, home-based other and non-home based trips.

A final program in the sequence (S2F1--not specifically documented) reads in the 24 data files that have been generated elsewhere and combines them into a single data file.

Although transit access times were generated from the UTPS process, these numbers were discarded. This is because access time was determined to be a standard of 6 minutes, the walk time if the average access distance is one-quarter mile. These time factors were thus input to the processing.

B-1.4 Specification of Input File

The completed ROOTINFO file contains one record (observation) for each zone pair between which there is a feasible transit path. Since three paths are allowed between all zone pairs, fields must be allocated for information regarding each possible path--even if less than three paths exist. Our convention regarding cases where fields are not used is as follows: If less than three paths are available between two zones, put -8 in fields relevant to the missing path(s). For example, if only two paths exist, put -8 in fields 69 through 101 for that zone pair. If less than three legs constitute a certain path, then put -7 in fields relevant to the last leg(s). For example, if path 1 has two legs, put -7 in fields 22 through 30 for that zone pair. Finally, if a leg of some path has less than three lines traversing it, put -6 in all fields relevant to any missing line(s). For example, if path 2, leg 2 has two lines, put -6 in fields 53 and 54 for that zone pair. If any of the required data is unavailable, -9 will be put in the relevant field.

Each record will contain a field for each of the following pieces of information (see "ROOTINFO FILE STRUCTURE").

ROOTINFO FILE STRUCTURE

COLUMN	FORMAT	NAME OF VARIABLE IN PROGRAM	VARIABLE	DESCRIPTION
1- 3	I3	ZI	1	Origin zone
4- 6	I3	ZJ	2	Destination zone
7- 9	I3	Not used	3	First path (contains a 1)
10- 12	I3	NL11	4	Number of lines on leg 1 of path 1
13- 15	I3	IVT11	5	In-vehicle time for leg 1 of path 1
16- 18	I3	NOB11	6	Number of buses available on leg 1, path 1 after the three most available lines have been considered*
19- 21	I3	L111	7	Line number, first bus, leg 1, path 1
22- 24	I3	HW111	8	Headway for bus in item #7
25- 27	I3	L112	9	Line number, second bus, leg 1, path 1
28- 30	I3	HW112	10	Headway for bus in item #9
31- 33	I3	L113	11	Line number, third bus, leg 1, path 1
34- 36	I3	HW113	12	Headway for bus in item #11
37- 39	I3	NL12	13	Number of lines on leg 2, path 1
40- 42	I3	IVT12	14	In-vehicle time for leg 2,path 1
43- 45	I3	NOB12	15	Number of buses available on leg 2, path 1 after three most available lines have been considered*
46- 63	6I3	as in 7-12	16- 21	Correspond to items 7-12 for leg 2
64- 66	I3	NL13	22	Number of buses available on leg 3, path 1
67- 69	I3	IVT13	23	In-vehicle time for leg 3, path 1
70- 72	I3	NOB13	24	Number of buses on leg 3, path 1 after the three most available lines have been considered*
73- 90	6I3	as in 7-12	25- 30	Correspond to items 7-12 for leg 3
91- 93	I3	IVT1	31	Total in-vehicle time, path 1
94- 96	I3	W1T1	32	Wait time, path 1
97- 99	I3	W2T1	33	Transfer time, path 1

100-102	I3	WLK1	34	Walk time, path 1**
103-105	I3	CST1	35	Cost (in cents), path 1
106-108	I3	Not used	36	Second path (contains a 2)
109-204	32I3	as in 4-35	37- 68	Correspond to items 4-35, for path 2
205-207	I3	Not used	69	Third path (contains a 3)
208-303	32I3	as in 4-35	70-101	Correspond to items 4-35, for path 3
304-306	I3	IVAUTO	102	Number of minutes spent in-vehicle for travel by auto from origin zone to destination zone along "minimum" auto path"
307-309	I3	WLKC	103	Number of minutes of access (walk) time for travel by auto from origin zone to destination zone along "minimum auto path"**
310-312	I3	CSTC	104	Cost (in cents) for travel by auto from origin zone to destination zone along "minimum auto path"
313-315	I3	PCT1AUTO	105	Proportion of households living in origin zone with one auto
316-318	I3	PCT2AUTO	106	Proportion of households living in destination zone with two or more autos
319-321	I3	PCTSERV	107	Proportion of population in origin zone served by transit
322-324	I3	PCTEMP	108	Proportion of employment in destination zone served by transit
325-329	I5	NHBW	109	Number of transit trips taken from home in origin zone to work in destination zone (HBW)
330-334	I5	NHBO	110	Number of transit trips taken from home in origin zone to non-work places in destination zone (HBO)
335-339	I5	NNHB	111	Number of transit trips taken from origin zone to destination zone that are not home-based
340-354	3I5	NIJC1,NIJC2, NIJC3	112-114	Correspond to items 109-111, for auto trips

ROOTINFO FILE STRUCTURE

COLUMN	FORMAT	NAME OF VARIABLE IN PROGRAM	VARIABLE	DESCRIPTION
1- 3	I3	ZI	1	Origin zone
4- 6	I3	ZJ	2	Destination zone
7- 9	I3	Not used	3	First path (contains a 1)
10- 12	I3	NL11	4	Number of lines on leg 1 of path 1
13- 15	I3	IVT11	5	In-vehicle time for leg 1 of path 1
16- 18	I3	NOB11	6	Number of buses available on leg 1, path 1 after the three most available lines have been considered*
19- 21	I3	L111	7	Line number, first bus, leg 1, path 1
22- 24	I3	HW111	8	Headway for bus in item #7
25- 27	I3	L112	9	Line number, second bus, leg 1, path 1
28- 30	I3	HW112	10	Headway for bus in item #9
31- 33	I3	L113	11	Line number, third bus, leg 1, path 1
34- 36	I3	HW113	12	Headway for bus in item #11
37- 39	I3	NL12	13	Number of lines on leg 2, path 1
40- 42	I3	IVT12	14	In-vehicle time for leg 2,path 1
43- 45	I3	NOB12	15	Number of buses available on leg 2, path 1 after three most available lines have been considered*
46- 63	6I3	as in 7-12	16- 21	Correspond to items 7-12 for leg 2
64- 66	I3	NL13	22	Number of buses available on leg 3, path 1
67- 69	I3	IVT13	23	In-vehicle time for leg 3, path 1
70- 72	I3	NOB13	24	Number of buses on leg 3, path 1 after the three most available lines have been considered*
73- 90	6I3	as in 7-12	25- 30	Correspond to items 7-12 for leg 3
91- 93	I3	IVT1	31	Total in-vehicle time, path 1
94- 96	I3	W1T1	32	Wait time, path 1
97- 99	I3	W2T1	33	Transfer time, path 1

Notes: *The path of interest should always be included as one of the three detailed lines, even if it is not one of the three most available lines.

**In the application of our model, and in the estimation stage, this was considered constant at six minutes.

All times should be in terms of minutes.

All percents should be integers in the range 0 - 100.

This file requires 354 spaces for each record.

APPENDIX B-2: ROUTE PATRONAGE FORECASTING MODEL PATH DETERMINATION

Unlike regression or pivot point patronage estimation techniques, the Route Patronage Forecasting Model is able to account for transfer movement between the transit line being examined and all other lines in the system. This feature is particularly important in grid- or feeder-type configurations. The determination of paths among zonal interchanges is, however, a complex process, requiring computer assistance when more than a few zones are involved.

The Route Patronage Forecasting Model allocates trips to up to three paths between zone pairs. This allocation is based on relative travel impedance on those paths. The model is unable, however, to identify paths and this information must be predetermined and input to the model.

Interzonal path descriptions must be based on some representation of the transit system, usually in the form of links and node points. Like many transit operators, Tri-Met uses the UMTA UTPS package to build a representation of the system and determine paths among zones on the system. The programs INET or UNET are used to represent the system, the program UPATH determines all-or-nothing paths, and the program UPSUM summarizes trip interchange information. Tri-Met follows these three programs with programs USTOS and UMATRIX to make more travel time adjustments due to timed transfer activity at transit centers, but this is an optional feature and won't be discussed here. The coding and use of the UTPS programs is familiar to many and documentation is available from UMTA.

The Route Patronage Forecasting Model uses the UTPS program sequence to determine paths and interchange information, but since UPATH uses an all-or-nothing assignment and the Route Patronage Forecasting Model uses three paths in its assignment, the UTPS program sequence must be processed three times with some manipulation to force selection of three reasonable paths.

In Tri-Met's normal use of the path program, UPATH, the transit first wait time coefficients are set to 1.7 and the transfer coefficient is set to 2.0. Other maximum and minimum coefficients are used but need not be discussed here. The coefficients are factors which are applied to the wait time, calculated as one-half the boarding mode's headway. These coefficients are used only in the path selection process and are removed from the summary information. The coefficients represent the perceived inconveniences of waiting and boarding buses in addition to actual time doing so. Their use has a direct effect on the path selection process and the likelihood that transfer activity will take place.

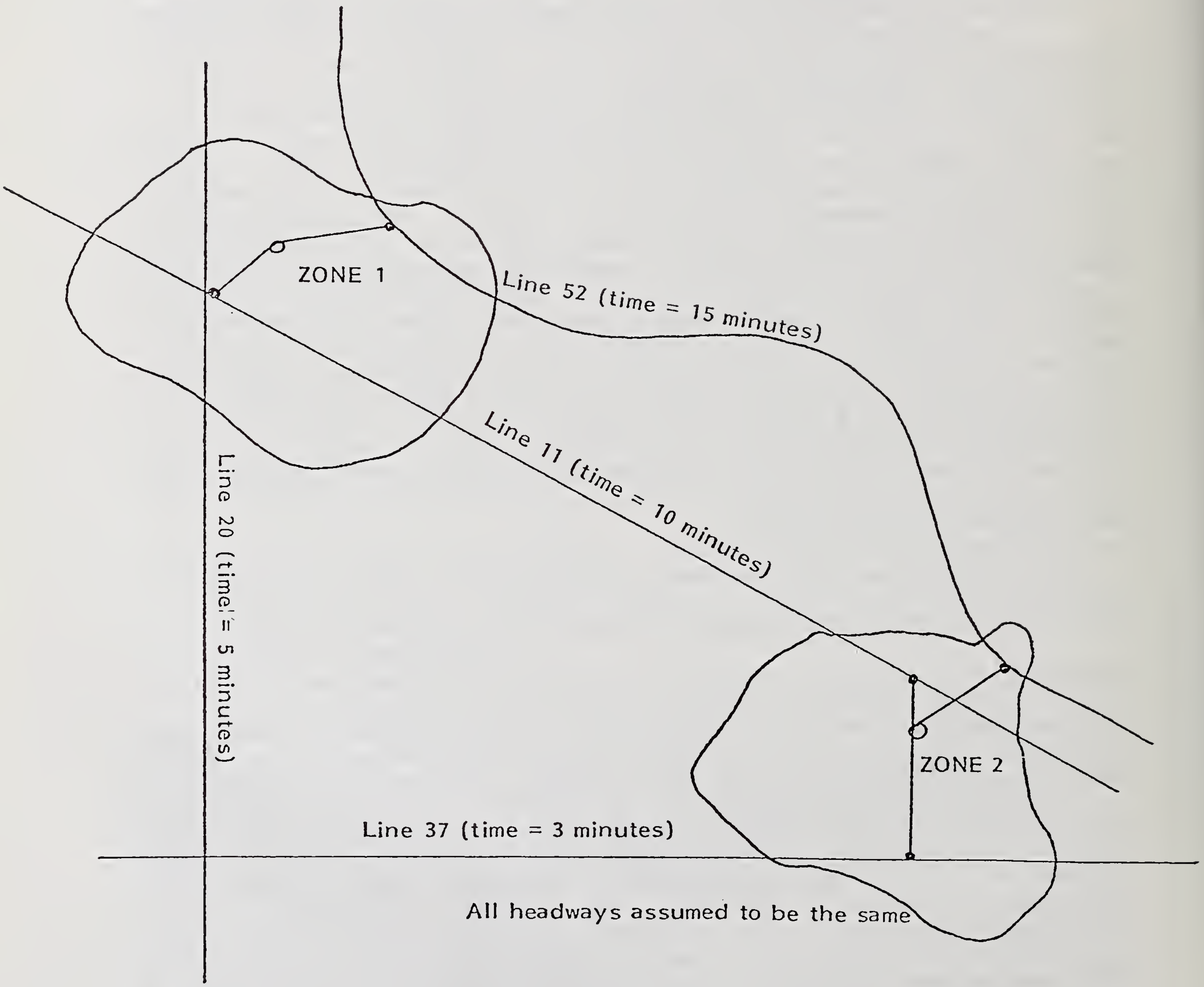
In normal use of the network program, INET or UNET, transit routes are accessed from a zone (represented by a zone centroid) with walk connectors. These are links from the trip producing centroid to one or more nodes (representing "bus stops") on the transit network. Tri-Met uses percent service or coverage factors based on the assumption that no one will walk more than one-quarter mile to catch a bus. Therefore, all walk access links are assigned a time of 3 minutes. (This percent served concept is also carried over the Route Patronage Forecasting Model.) The configuration of these access lines determines what routes through or near a zone will be accessible for trips in that zone. Most zones use two to four such access links. Some zones will have only one link and zones not served at all may be disconnected (no links).

The determination of the three paths using the UTPS program sequence was done by (1) changing the UPATH transfer penalty coefficient, and (2) turning INET (UNET) zone access links "on" or "off". One run of the sequence uses a normally coded network with the UPATH transfer coefficient set to 0. This increases the likelihood of a path using a transfer being selected, since actual wait time is used rather than wait time factored by 1.7. The second and third run of the program sequence alternately turns access links on or off, forcing paths to use alternate routes. Other variations are possible using UPATH coefficients and network modifications.

To illustrate this process, Figure B-2.1 shows a simplified representation of two zones and available connecting routes. A traveler from zone 1 to zone 2 is faced with three possible paths and two possible ways to access the transit system. Normal Processing of UPATH would put the traveler on Line 11. However, if two network and path runs are made with alternative access links turned off, travel by way of Line 52 is identified. In these two runs, all coefficients are left in their normal settings. Selected access links are turned off by either removing them or giving them very high impedances. Note that if only one access

Figure B-2.1

EXAMPLE ZONE CONFIGURATION



link exist for a zone, it is always left "on" and the same path will be selected at least twice of the three runs. If three or more links exist, some links will have to be doubled up, i.e., two on, one off; one on, two off. This is a necessary arbitrary determination. A third processing would leave all access links turned "on" but sets the transfer penalty to 0. This places the traveler on a path using lines 20 and 37. The actual outcome of this process depends on route headways and route and access link configuration. It is very possible that the same path will appear each time. It is also possible, particularly when more than two access links are present, that some reasonable paths are overlooked. Redundant paths are handled by the UPATH conversion program. The limitation to three "reasonable" paths is a practical consideration and a necessary weakness of the process.

An alternative way to turn access links on or off would be to code access links with different mode numbers (i.e., 1, 2, or 3). The NOX parameter in program UPATH could then be used to turn links on or off.

The output of these three runs are read into the UPATH conversion program which reformats and combines the information with up to three path definitions for each one interchange. The entire process need not be repeated for both peak and daybase unless peak and daybase route and headway have significant relative differences. A daybase run of a normal network is desirable, however, to reflect changes in total travel time on the shortest path.

This process produces a reliable picture of interzonal travel options but it is somewhat cumbersome in that (1) the sequence of

UTPS programs must be run three times, and (2) the walk access link file must be modified to turn links on or off. If the manual method to estimate minor route changes is used, this process need only be carried out once for a base network. In any case, the process depends on the existence of a reliable base INET or UNET network, and an efficient procedure to run these programs.

APPENDIX B-3: MODEL APPLICATION GUIDE

Once the input dataset ROOTINFO is in place, and the RPFM model may be applied to a variety of route-level planning and analysis tasks. This section describes the general procedure to be followed, in terms of adjustments to the model or supporting software, to address typical applications.

Figure B-3.1 is a matrix which summarizes the operating actions necessary to apply the model to five types of service (or system) alterations, in addition to running the model to replicate baseline conditions (for calibration or validation). The operating actions include, in order of difficulty:

1. Translating the elements of the service change to the model, i.e., route number--inbound, outbound, headway, percent change in headway, and percent change in run time. This is done by altering the parameters in the SAS program. The SAS program uses these inputs to adjust the impedance terms in the RPFM model. Typically, this is the only action necessary in order to use RPFM to assess minor service changes like headways (frequency), stop spacing (travel time), etc.

2. Calibration of the model through adjustment of the "a" terms, as described in Appendix B-4. Model calibration would be advisable when investigating such major changes as land use or route alignment. The "?" notation in the figure related to fare change indicates that predictions for this type of alteration would

Figure B-3.1
APPLICATION OF ROUTE PATRONAGE FORECASTING MODEL

TYPE OF APPLICATION / ACTION(S) REQUIRED	Change Parameters in SAS Code (headway, % travel time)	Calibrate model by adjusting "a" terms	Alter Accessibility Data in ROOTINFO	Alter General Zone Characteristics in ROOTINFO	Alter Impedance Data in ROOTINFO	Re-estimate Total Person Trip Table
Replicate Existing Line Ridership	☆					
Change Service Frequency (Headway)	☆				?	
Change Transit Travel Time (Stop spacing, etc)	☆				?	?
Change Route Alignment	☆	☆	☆		?	?
Change Land Use		☆	☆	☆		☆
Change Fare (e.g. zone surcharge)		?			☆	

Code:

☆ means action probably required

? means model would be improved with action, but not strictly necessary probably be improved if a similar situation could be found to calibrate the model on.

3. Alteration of the accessibility data in the data base, ROOTINFO. Accessibility data includes the "percent (population/employment) served" factors, as well as walk and wait times. This must be done for the major alterations, i.e., land use and realignment. An important distinction here is that if the RPFM is being used to predict ridership changes in conjunction with route realignment, new paths and new path travel times must be derived. If the model is being used to predict ridership changes on an existing route experiencing land use changes, then new "percent served" factors are required, but not new paths.

4. Alter General Zone Characteristics in ROOTINFO. These factors include only the "autos owned per household" data, and are typically adjusted only in the event of a land use change. Even these may not need to be changed if the model is used applying a transit (rather than joint auto/ transit) trip table.

5. Alter impedance data in ROOTINFO. Impedance factors include travel time and travel cost. This clearly must occur in the analysis of a fare change. The "?" indicates that judgment must be applied in deciding the need to change impedance factors for these applications. In other words, if the proposed changes in

headway, travel time, route alignment, etc., are sufficiently minor so as not to radically change path/route assignment, the dataset can be left alone.

6. The maximum alteration to operate RPFM would include reestimation of the total person trip table. However, this action places the user into a decision as to whether the planning problem is in fact within the scope of the Route Patronage Forecasting process.

APPENDIX B-4: MODEL CALIBRATION

Step 1 of the Route Patronage Forecasting Model (RPFM) predicts the number of transit trips between zone pairs, which are subsequently divided among paths and routes by Steps 2 and 3. This section discusses circumstances under which calibration of the Step 1 equations is necessary before analyzing service changes, and the method by which the calibration is performed.

Econometric models must often be calibrated when applied to planning situations outside the range of the data from which the models were estimated. Such is the case when models are transferred to new sites or to new services where the markets are different and there is no operating experience.

If the Route Patronage Forecasting Model is to be used to analyze the effects of service change on an existing route, and patronage data are available for the route as it currently exists, then calibration of the model is not necessary. Instead, the model is used to predict the percentage change in ridership, and this percentage is used to modify the existing ridership. In other words, the RPFM is first run using the existing route service characteristics as input; the output of this run will be the predicted patronage under original service, $\hat{P}_o$. The model is then run a second time using the new (changed) characteristics as input, with the output being predicted route patronage under the changed service, $\hat{P}_c$. The best estimate of the change in patronage due to the service change is simply the known patronage of the route before the service change, P*, multiplied by the predicted **percent change** in patronage due to service changes (($\hat{P}_C$ - $\hat{P}_O$)/$\hat{P}_O$). Hence, total route patronage is simply before-change total patronage plus the predicted change, or:

$$P^* + P^* \left(\frac{\hat{P}_C - \hat{P}_O}{\hat{P}_O} \right)$$

It can be demonstrated that model calibration is not important when predicting the **percentage change** in patronage, and hence in predicting new route patronage for existing services.

Zone-to-zone transit trips are predicted by equations of the following general form:

$$\ln \hat{N}_{ij} = a + \text{other factors}$$

where: $\hat{N}_{ij}$ is the predicted number of transit trips from zone i to zone j (for a particular purpose);

a is an intercept term; and

"other factors" describe the transit and auto travel characteristics, or

$$\hat{N}_{ij} = e^{a} * e^{\text{other factors}}$$

The intercept term, a, in each of these equations has an important property: it can be set such that the predicted number of transit trips taken (for each purpose) from zone i to zone j. Calibration generally determines what value "a" should have such that this equality between predicted and actual number of transit trips holds.

The predicted number of zone-to-zone transit trips ($\hat{N}_{ij}$) is divided among different paths and routes in steps 2 and 3 of

RPFM. These steps develop a fractioning factor f_{ij} which is the fraction of transit trips taken from zone i to zone j on the route of interest, $\hat{P}_{ij}$ is then simply the total number of transit trips from zone i to zone j multiplied by the fractioning factor:

$$\hat{P}_{ij} = f_{ij} \cdot \hat{N}_{ij} = f_{ij} \cdot e^{a} * e^{\text{other factors}_{ij}}$$

and the total number of transit trips taken on the route of interest is simply the sum of predicted route-of-interest patronage in all zones:

$$\hat{P}_{\text{Total}} = \sum_{ixj} \hat{P}_{ij} \quad .$$

From this, the predicted **percentage change** in route patronage resulting from a service change can be written as:

$$\frac{\hat{P}_c - \hat{P}_o}{\hat{P}_o} = [\sum_{ixj} f^c_{ij} \hat{N}^c_{ij} - \sum_{ixj} f^o_{ij} \hat{N}^o_{ij}] / \sum_{ixj} f^o_{ij} \hat{N}^o_{ij}$$

$$= [\sum f^c_{ij} \; e^{a} * e^{\text{other factors under changed conditions}}$$

$$- \sum f^o_{ij} e^{a} * e^{\text{other factors under original conditions}}]$$

$$/ \sum f^o_{ij} \; e^{a} * e^{\text{other factors under original conditions}}$$

$$= \frac{e^{a} [\sum f^c_{ij} e^{\text{changed factors}} - \sum f^o_{ij} e^{\text{original factors}}]}{e^{a} \sum f^o_{ij} e^{\text{original factors}}}$$

$$= \frac{[\sum f^c_{ij} e^{\text{changed factors}} - \sum f^o_{ij} e^{\text{original factors}}]}{\sum f^o_{ij} e^{\text{original factors}}}$$

(where the superscript c indicates terms forecast under the changed service, and the superscript o indicates terms forecast under the **original** service conditions.)

This last equation does not include the intercept (calibration) terms "a", and from this it is concluded that having "a" set to an appropriate "calibrating value" is not important for the analysis of service changes on existing routes.

Although calibration is not necessary for analysis of service changes on existing routes, it may be desired for other applications, particularly for proposed new routes. Model calibration is important when there is no known base case to which percent changes can be applied. The suggested procedure is to calibrate the model on a route serving an area similar to that which the proposed new route will serve. By so doing, the particular under or over prediction patterns of the equations on this type of area will be identified, and will be corrected by the calibration term, a. Subsequently, when the model is used to forecast patronage on the new route, these under or over predictions will be automatically corrected for, and an accurate forecast derived.

In the case where a model has been calibrated, the procedure described above for analyzing service changes (on the route for which the model is calibrated) can be simplified: the model now need only be run once, using the new (changed) service conditions as input. Recall that under the procedure described above, the best forecast of new patronage was the sum of original patronage plus a percent change:

$$\text{new } \widehat{\text{patronage}} = P^* + P^* \cdot \left[\frac{\hat{P}_c - \hat{P}_o}{\hat{P}_o} \right]$$

When a model is calibrated, $\hat{P}_o = P^*$ (that is, patronage predicted by the mode under original service conditions should equal actual patronage), and so this equation reduces to:

$$\text{new patronage} = \hat{P}_c$$

This states that the best prediction of new patronage is simply the forecast of new patronage output by the model when it is run using the new (changed) service as input.

Currently, the forecasting model is calibrated for routes 163 and 164. (These routes represent inbound and outbound directions of actual Tri-Met Line 53.) If it is necessary to recalibrate the model to a different route (for example, for analysis of a proposed new route serving an area significantly different from that served by routes 163 and 164) the following three-step procedure is recommended:

(1) run the model using the existing service conditions on the route under analysis as input.

The model will provide as output (among other things) the following:

- actual (from survey) transit home-based work trips (NHBW) for the route of interest.
- forecast transit home-based work trips (NIJT1) for the route of interest.
- actual transit home-based other trips (NHBO) for the route of interest.
- forecast transit home-based other trips (NIJT2) for the route of interest

- o actual transit non-home based trips (NNHB) for the route of interest
- o forecast transit non-home based trips (NIJT3) for the route of interest.

If, in the judgment of the user, the forecast number of transit trips in each of the three purpose categories is sufficiently close to actual trips, no further calibration is required. Otherwise, the user should:

(2) calculate the three values:

ln (NHBW/NIJT1) = A
ln (NHBO/NIJT2) = B
ln (NNHB/NIJT3) = C

(where ln denotes the natural logarithm function).

Then:

(3) each of these terms should be added to the "a" terms in the corresponding zone-to-zone transit-trip prediction equation; that is, A should be added to the "a" term in the zone-to-zone home-based work transit trip prediction equation, and C added to the "a" term in the zone-to-zone non-home based trip prediction equation.

Because of non-linearities in the zone-to-zone transit trip forecasting equations, calibration of the model for a new route may require the user to go through several iterations of three steps: running the model, inspecting the output, and adjusting the calibration term a. When the forecast number of transit trips in each category becomes sufficiently accurate, calibration is completed.

APPENDIX B-5: DEVELOPMENT OF EXPANSION FACTORS FOR MANUAL METHOD

Estimating patronage using the sample zone pairs provides only a sample of ridership on the line of interest. Comparing "before and after" predictions of sample zone trips with the RPFM model will provide an estimate of the percentage change in ridership on the line of interest. However, to test whether the model is estimating ridership by line accurately, the predictions must be compared with actual ridership. Consequently, an **expansion factor** must be determined that expands the sample ridership up to total ridership. Since all trips could conceivably use transit if a trip end is served by the line of interest, the expansion factor becomes the number of total person trips between the sample zone pairs divided by the number of total person trips from all zones served by the line of interest. As was mentioned earlier, trips without a trip end in a zone served by the line of interest are excluded from the analysis. In some cases the analyst may know of situations where there are a substantial number of trips of this type. They should be included in the analysis.

A process is described below for developing the expansion factor. The process requires a combined purpose, total person trip table. It would be preferable to use a transit trip table (if available) to develop the expansion factor. Also, if either total person trip tables or transit trip tables by purpose (HBW, HBO, and NHB) are available, separate expansion factors can be computed by purpose. If only the total person trip table is available, it may be used bearing in mind the qualifications discussed later in this section. Finally, this procedure may be bypassed entirely if the user wants only to predict the percentage change in ridership.

To develop the expansion factor, the user first lists all zones served by the route of interest. A tally is then made of all trips that could use the route of interest, using a combined purpose total (auto and transit) person trip table. The tally includes trips in one direction only, with no double counting when both original and destination zones are served by the route of interest. This count, of course, includes all zone pairs which use the route under the 3-leg feasibility criteria, including those that would have been eliminated by the sample selection procedure. Once done, a separate tally is done for the sample zone pairs only, using the same trip table. The expansion factor is then calculated as the number of sample trips divided by the total number of potential trips. This factor is then applied to the predictions from the RPFM model to produce a corrected estimate of total trips.

Although the RPFM model estimates trips by individual purpose (HBW, HBO, and NHB), clearly the expansion factors above are not specific to purpose. The reason for this is that actual ridership counts by purpose are not available for comparison. Hence, the expansion factors are applied to the model estimates, **after** the estimates are summed by purpose, to give corrected **total ridership** on the route of interest.

Developing expansion factors using total person trips (i.e., both auto and transit trips) assumes that accessibility to transit service is as ubiquitously distributed as automobile accessibility. In the Portland region, this assumption is not perfect, but reasonable considering the extent and coverage of transit service--particularly in the suburban areas. However, even for regions with good transit service, it would be preferable to use a transit trip table, if one is available, to develop the expansion factor.

NOTICE

This report is being distributed through the U.S. Department of Transportation's Technology Sharing Program.

DOT-I-83-37

TECHNOLOGY SHARING

SPECIAL STUDIES IN TRANSPORTATION PLANNING (SSTP)

PROGRAMS OF THE U.S. DEPARTMENT OF TRANSPORTATION

1