REPORT NO. IA-11-0005

DEVELOPMENT OF TIME-SERIES BASED TRANSIT PATRONAGE MODELS

VOLUME 1

MODEL DEVELOPMENT FOR EVALUATING SERVICE LEVEL AND FARE STRATEGIES

MICHAEL KYTE

JAMES STONER

JONATHAN CRYER

The University of Iowa
College of Engineering
Iowa City, Iowa 52240

MARCH 1985

FINAL REPORT

Prepared for

U.S. DEPARTMENT OF TRANSPORTATION
URBAN MASS TRANSPORTATION ADMINISTRATION
OFFICE OF TECHNICAL ASSISTANCE
UNIVERSITY RESEARCH AND TRAINING
WASHINGTON, D.C. 20590

DEVELOPMENT OF TIME-SERIES BASED TRANSIT PATRONAGE MODELS

Volume I - Model Development for Evaluating Service Levels and Fare Strategies

Prepared By:

Michael Kyte
James Stoner
Jonathan Cryer
Rachel McQuillen
Lama Farsakh

The University of Iowa

March 1985

Technical Report Documentation Page

1. Report No. UMTA-IA-11-0005-85-1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Development of Time Series Based Transit Patronage Models. Volume I - Model development for evaluating service level and fare strategies		5. Report Date March 1985
		6. Performing Organization Code
7. Author(s) M. Kyte, J. Stoner and J. Cryer		8. Performing Organization Report No.
9. Performing Organization Name and Address The University of Iowa College of Engineering Iowa City, Iowa 52242		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. IA-11-0005
12. Sponsoring Agency Name and Address U.S. Department of Transportation Urban Mass Transportation Administration Office of Technical Assistance Univ. Research & Training Program, Washington, D.C. 20590		13. Type of Report and Period Covered Final Report July 1983 - March 1985
		14. Sponsoring Agency Code URT-33
15. Supplementary Notes UMTA Project Director - Judy Z. Meade Technical Advisor - Thomas J. Hillegas		

16. Abstract

This report describes the development and application of a methodology to identify and analyze the factors that influence changes in public transit ridership. The data used in model development and testing is from Portland, Oregon and covers the period 1971 through 1982. Models are developed at the system, sector, and route levels, and are used to assess the impacts of past changes in service level and fare, as well as to forecast future transit patronage. The statistical approach used was developed by Box and Jenkins for time-series data and is therefore more appropriate and powerful than the more traditional regression analysis. Of particular interest is the identification of the lag structures and functional forms that constitute the relationships between transit ridership, level of service, travel costs, and market size.

17. Key Words Transit, Time Series, Modeling, Forecasting	18. Distribution Statement		
19. Security Classif. (of this report) unclassified	20. Security Classif. (of this page) unclassified	21. No. of Pages	22. Price

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

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

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

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.

Approximate Conversions from Metric Measures

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

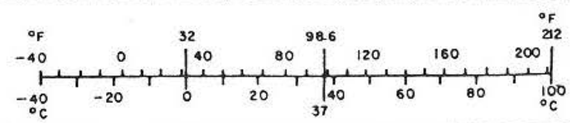


FIGURE 3. METRIC CONVERSION FACTORS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. EXECUTIVE SUMMARY

OVERVIEW

This report is the first of two technical reports that describe a methodology for analyzing and forecasting public transit ridership. The methodology is applied to data from the public transit system in Portland, Oregon. The second report, A Handbook for Developing Time-Series Transit Ridership Models, focuses on the statistical methodology for developing time-series models.

Analyzing past variation in transit ridership and forecasting future ridership variation are two important concerns for the public transit analyst. Before a service or fare change is instituted, its potential impact on ridership must be assessed. After implementation, and equilibrium conditions have been reached, the impact of the change must be analyzed. Has ridership increased or decreased, and has this been the result of the service or fare change? Often it is difficult to isolate the variation in ridership that can be attributed to a fare or service level change from the effects of some exogeneous factor such as a change in gasoline supply or price.

There are usually several processes that are occurring simultaneously, each in some way affecting ridership. A change in transit ridership in 1979, for example, might have been strongly related to rapidly increasing gasoline prices and supply constraints. But changes in the size of the travel market or in the level of transit service would also have had a direct impact on ridership levels if these variables were also changing during this time. Thus, any study of the variation in transit ridership must consider all of the relevant influencing factors that are themselves changing. Similarly, to satisfactorily forecast future variation in transit ridership, a clear understanding of these factors is necessary. Because the nature of these relationships may themselves change over time, it seems clear that models based upon time-series data are more likely to capture these dynamics than those based upon cross-sectional data.

There have been several important efforts in recent years in the development of time-series based transit ridership models. Of particular importance is the work of Gaudry (1975, 1978), Kemp (1981a, 1981b) and Wang (1981, 1982). This report describes the results of a project which builds upon the work of these researchers and extends it into several important areas:

1. A methodology is proposed that provides a logical framework for the analysis and forecasting of transit ridership. The essence of the methodology is that in order to assess past impacts or to forecast future variation, a model must be developed that is time-series in nature and explicitly considers all of the relevant factors that influence transit ridership.

2. Consideration is given to the functional relationship between the input variables and transit ridership, particularly the nature of the delay that exists between a change in an input variable and when its effects in ridership can be measured. Also of importance is the method of specifying transit service level when using time-series data.

3. Extensive use is made of a statistical methodology that has not had wide application in transportation, the Box-Jenkins time-series models. This technique resolves several problems that occur when standard regression models are used with time-series data, including multicollinearity and serial correlation. Recent availability of the appropriate computer software makes use of this approach practical and available to most analysts.

THE METHODOLOGY

The basic methodology used in this project is shown in Figure 1. It includes three phases:

1. Model development phase
2. Impact analysis phase
3. Forecasting phase

In the Model Development Phase, a model form is postulated that includes a description of the variables that are assumed to affect transit ridership. The structural relationships between transit ridership and the input variables are then identified and estimated. This includes identification of the lag structure that exists. Finally, the complete model is estimated and checked to insure consistency with the appropriate statistical assumptions. The model proposed here is known as a Transfer Function Model, as developed by Box and

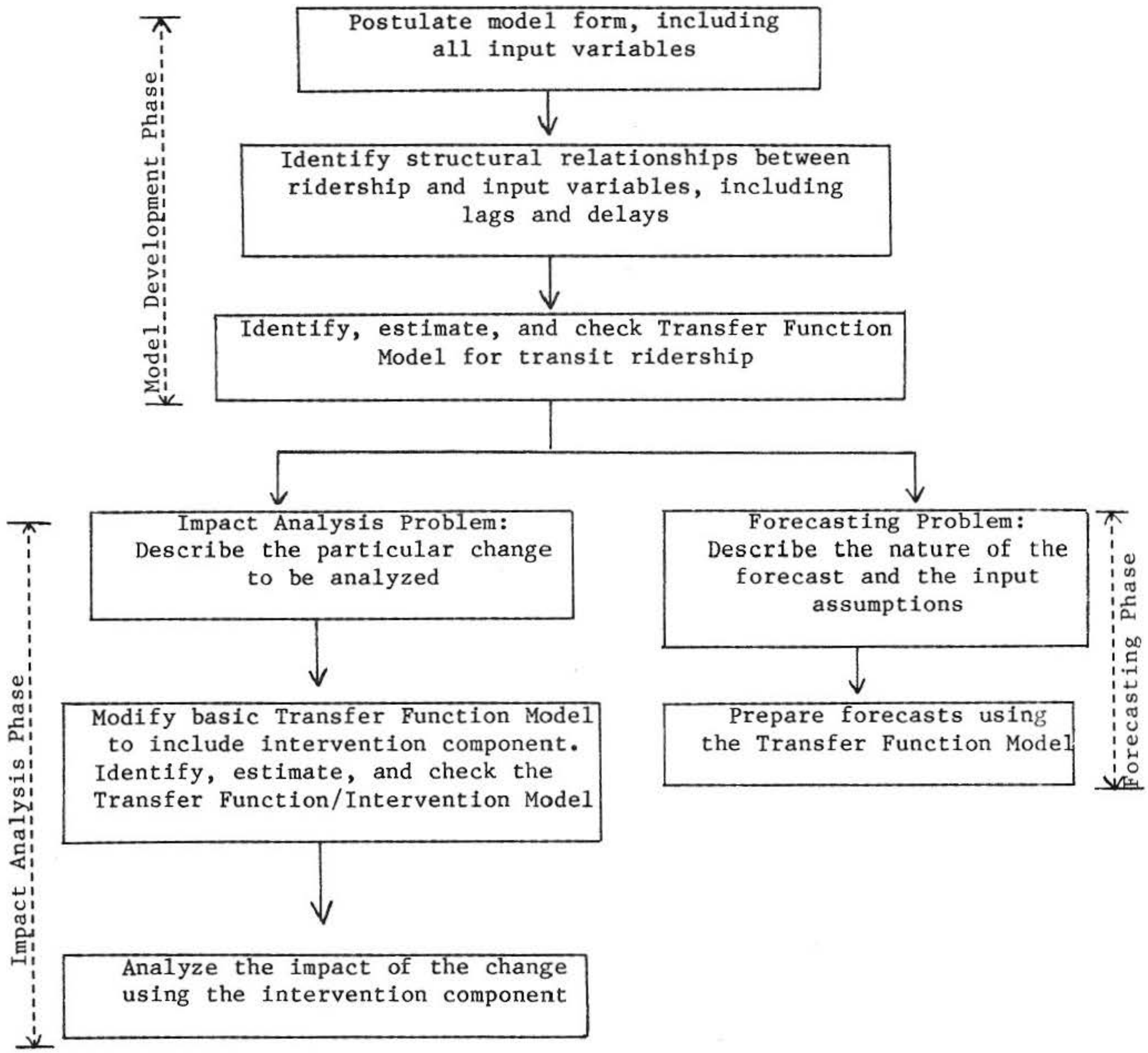
Jenkins (1976).

The Transfer Function Model can then be used in the Impact Analysis Phase to analyze the impact on transit ridership of past changes in service level, fare or other factors. The model coefficients provide an estimate of the average response to all previous changes in each of the input variables. To analyze the impact of a specific change, an intervention variable is introduced to the Transfer Function Model. The intervention variable is a binary variable which assumes a value of zero when the change is not in effect, and a value of one when the change is in effect. The Transfer Function Model can also be used in the Forecasting Phase when an assessment of a proposed future change is desired.

FIGURE 1

METHODOLOGY FOR ANALYSIS AND FORECASTING

OF PUBLIC TRANSIT RIDERSHIP



FINDINGS AND RESULTS

Model Development Phase

Data for Portland, Oregon covering the period 1971 through 1982 were used to develop a total of sixteen transit ridership models: one for the system as a whole, six representing distinct geographic sectors of the Portland region, and nine for individual routes in the Portland transit system. Figure 2 illustrates the three different data sets and the models that have been developed.

Four input variables were used for each of the models: transit service level, transit fare, gasoline price as a surrogate for auto operating costs, and employment as a measure of the travel market size (see Figure 3). Natural logarithms of the data were used, so that model coefficients give the elasticities directly for each variable. The nature of the market response was included in the model by introducing lagged variables. This allowed a direct assessment of the time delay between the introduction of a service level or fare change and when a change in ridership could be measured. Service level delays ranged from one to ten months for the system model and zero to three quarters for the sector and route models. Fare delays ranged up to two quarters. A summary of the elasticities and lags are given in Table 1.

Examination of Table 1 shows that there are some important consistencies in the results obtained by the three model categories. For example, the response delay to service level changes tends to be about two to three times longer for urban routes than for suburban routes. Another comparison is the consistency of the elasticities for the four input variables between the system model and the sector models, as given in Figure 4. Note that the elasticities estimated for the six sector models tend to vary around the system mean for each variable.

FIGURE 2

SUMMARY OF MODELS DEVELOPED

AND THEIR INTERRELATIONSHIP

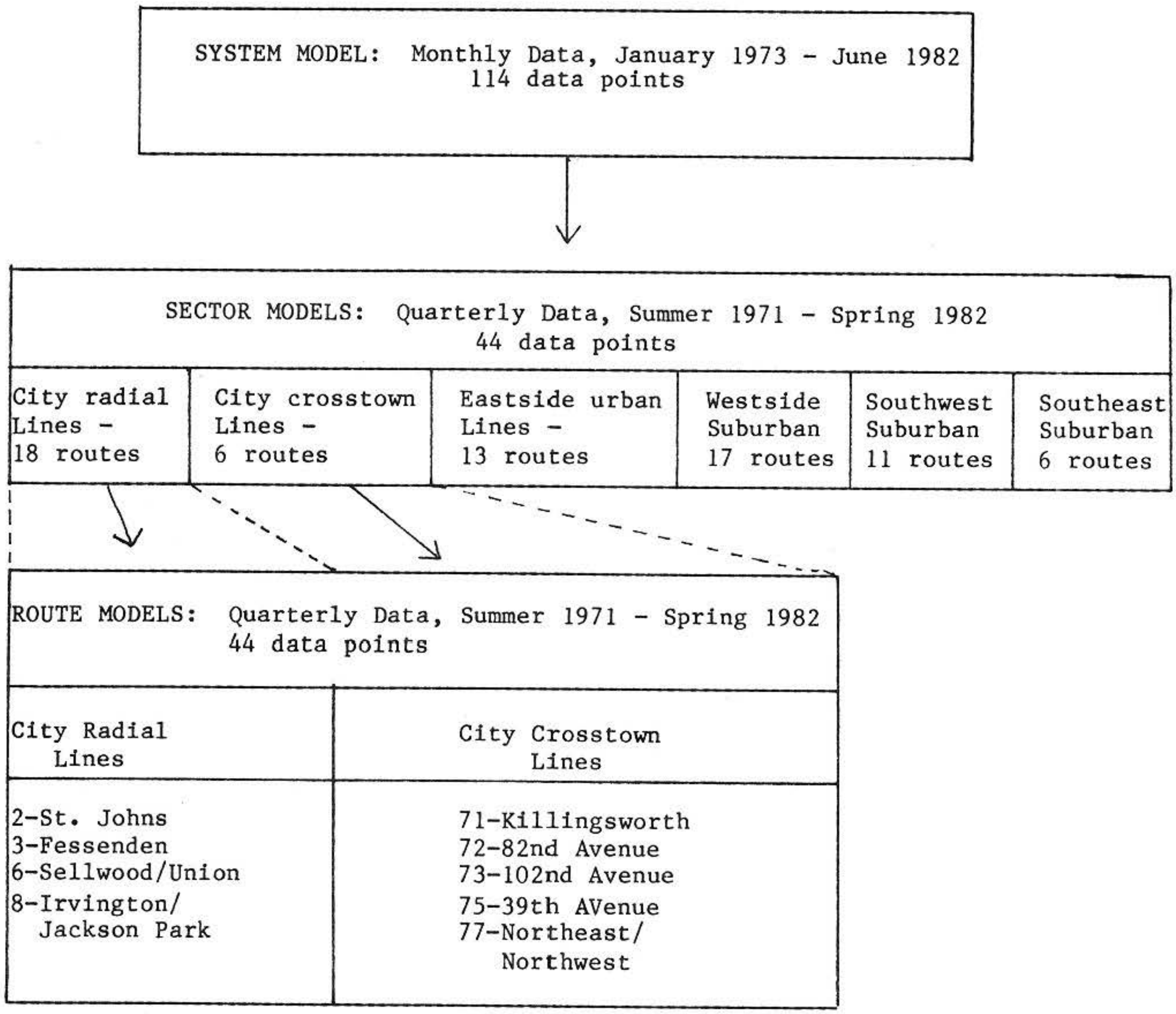


FIGURE 3

BASIC MODEL FORM

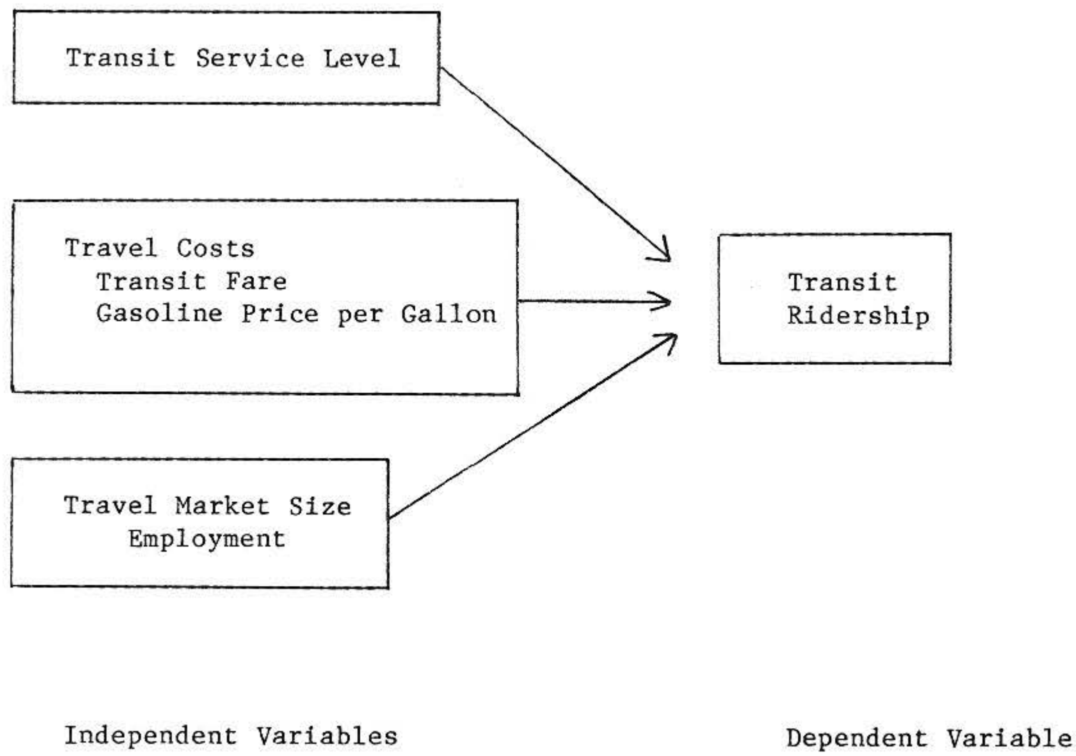


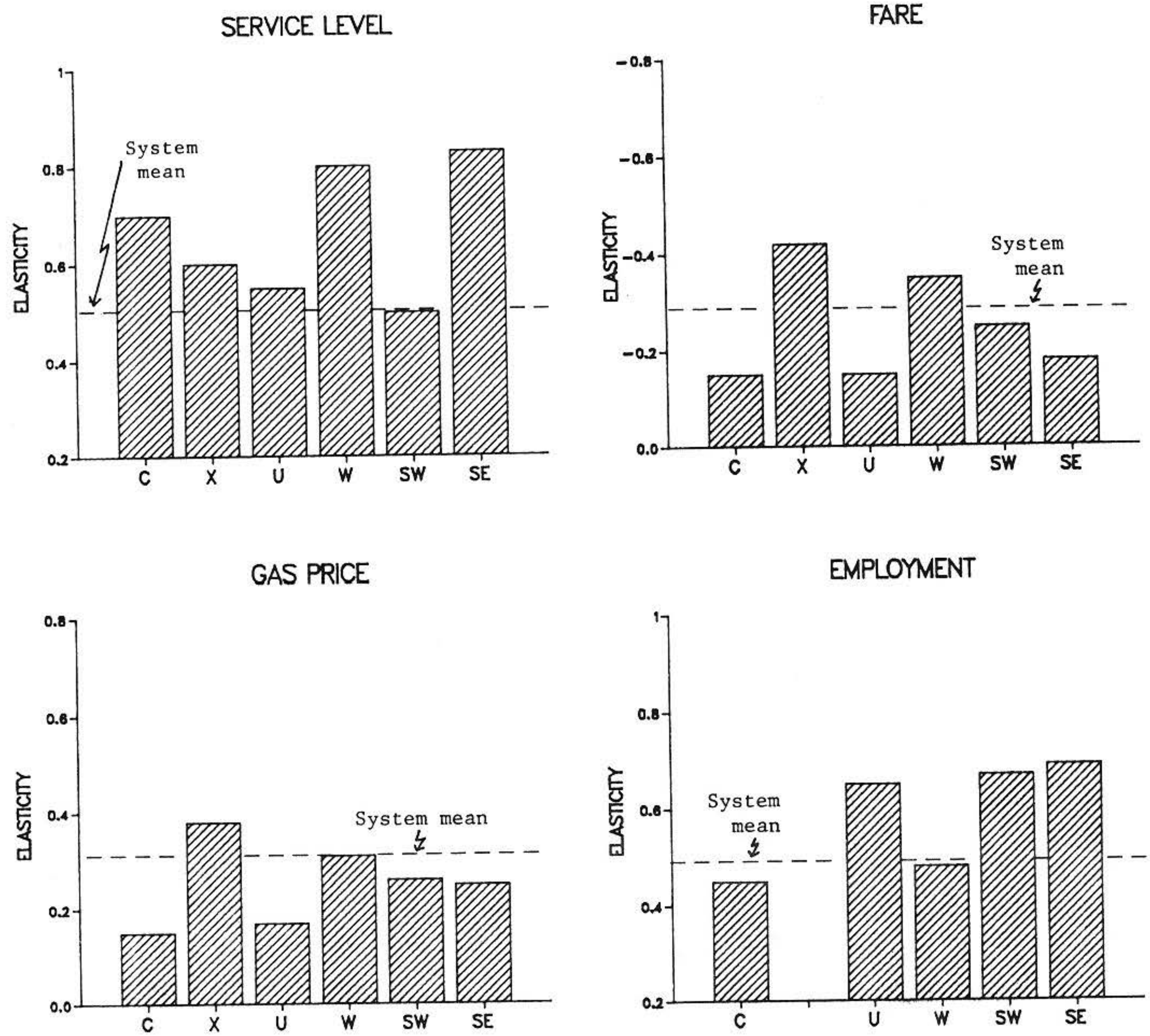
TABLE 1

SUMMARY OF MODELS

MODEL DESCRIPTION: Data Aggregation	Data Period	Model Description	SERVICE LEVEL Elasticity	Lag	FARE Elasticity	Lag	GAS PRICE Elasticity	Lag	EMPLOYMENT Elasticity	Lag
System	Monthly	System	.51	1,10	-.29	0	.32	0	.49	0
Sector	Quarterly	City radial lines	.71	2	-.13	0	.14	0	.43	0
		City crosstown lines	.60	0-3	-.42	0	.39	0	-	
		Urban Eastside lines	.55	2	-.15	0	.18	0	.65	0
		Westside sub. lines	.80	0	-.32	0	.31	0	.47	0
		SW Suburban lines	.49	0	-.22	1	.28	0	.67	0
		SE Suburban lines	.88	0,2	-.16	0	.27	0	.69	1
Route	Quarterly	City radial line-								
		Route 2	1.81	0,2	-.39	0	.72	0	1.14	2
		Route 3	1.73	0,2,3	-.90	0,1	1.39	0-3	-	
		Route 6	.23	0	-.80	0	.62	0	.95	0
		Route 8	.25	3	-.35	2	1.23	0,1	-	
		City Crosstown line-								
		Route 71	.72	0	-		3.24	2	-	
		Route 72	.55	0	-		.68	3	-	
		Route 73	-		-		.60	0	-	
		Route 75	-		-		1.72	3	-	
		Route 77	.35	0	-		.24	2	-	

Elasticity = total elasticity for given variable
Lag = lag or delay for which change in ridership was measured. A lag of 2 using quarterly data for example, indicates that a change in ridership was measured two quarters after the input variable was changed.

FIGURE 4
CONSISTENCY OF MODEL COEFFICIENTS BETWEEN SYSTEM AND SECTOR MODELS



LEGEND:
- C City Lines Sector
- X Crosstown Lines Sector
- U Eastside Urban Sector
- W Westside Suburban Sector
- SW Southwest Suburban Sector
- SE Southeast Suburban Sector

Impact Analysis Phase

The elasticities computed in the model development phase represent an average elasticity for a given variable over the entire study period. If four service changes were implemented during a given period, for example, the service level elasticity would be an average of the impact of each service level change. However, to study the impact of a specific service level change, an intervention variable, which represents that change alone, must be added to the model. The model is then re-estimated with the intervention variable and the coefficient yields the elasticity of the specific change under study. If the variable coefficient is not statistically significant, it can be concluded that the change had no measurable impact on ridership.

Eleven service changes instituted between 1973 and 1979 were analyzed using the intervention analysis technique. The results are summarized in Table 2. Seven of the eleven changes were found to have had a significant impact on ridership.

Forecasting Phase

The models developed in the initial phase of this project can be used to forecast future transit ridership variation. For example, the impact of a future fare change can be estimated using the appropriate model. But because the model depends upon future variation in gasoline price and employment as well, these variables must also be forecasted or assumptions must be made about their future values.

Table 3 shows the results of a forecast of system ridership for twelve periods (months) ahead. It was assumed that service level and fare were set by policy and that gas price and employment had to be forecasted using time-series models. These results, with a mean absolute percent error of 2.1%, show the high quality of forecast that can be achieved using this approach. Actual observed values and the forecast are compared graphically in Figure 5.

TABLE 2

IMPACT ANALYSIS OF PAST SERVICE CHANGES

AT THE ROUTE LEVEL

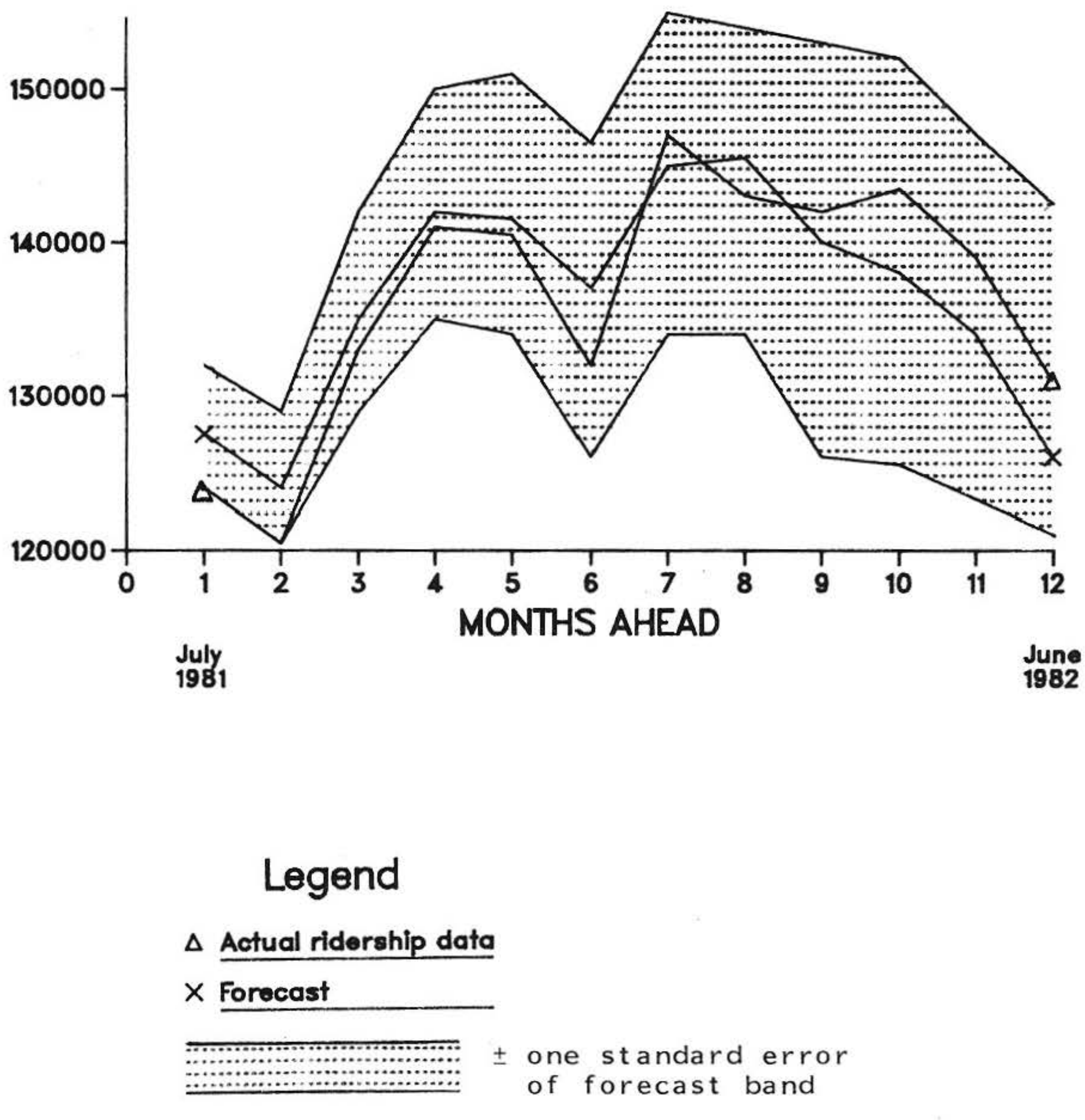
Route	Date	Type of Change	Significant Impact?	Coefficient of the Intervention Variable
2	1975	Frequency improvement	Yes	.13
	1978	Route extension	No impact	-
3	1973	Frequency improvement	Yes	.11
	1974	Frequency improvement, Route extension	Yes	.13
	1978	Service reduction	No impact	-
6	1974	Route extension	No impact	-
	1975	Frequency improvement	Yes	.23
71	1979	Frequency improvement, Route extension	Yes	.72
72	1976	Route extension	Yes	.81
75	1979	Route extension	No impact	-
77	1979	Frequency improvement	Yes	.35

TABLE 3
FORECAST OF TRANSIT RIDERSHIP SYSTEM DATA

Month	Actual	Forecast	% Error
July 1981	125,800	128,300	+2.0%
August 1981	121,400	124,700	+2.7%
September 1981	132,600	135,200	+2.0%
October 1981	141,700	142,600	+0.6%
November 1981	141,700	141,600	-0.1%
December 1981	132,900	136,200	+2.5%
January 1982	146,100	144,800	-0.9%
February 1982	142,500	145,800	+2.3%
March 1982	141,900	140,000	-1.3%
April 1982	143,200	138,400	-3.4%
May 1982	139,400	133,700	-4.1%
June 1982	132,700	127,600	-3.8%
Annual Total	1,641,900	1,638,900	

Monthly Total, Mean Absolute Percent Error 2.1%
Annual Total, Percent Error 0.2%

FIGURE 5
COMPARISON OF FORECASTS, SYSTEM MODEL



COMPARISON WITH STANDARD REGRESSION MODELS

It has been traditional to use multiple regression models when developing models relating transit ridership to explanatory variables. Using time-series data with regression models, however, invariably leads to a variety of statistical problems. Table 4 highlights the major areas in which problems are likely to arise by contrasting standard regression with transfer function models: multicollinearity, autocorrelated errors, lag structures, and coefficent estimates and standard errors. To determine whether these problems would, in fact, result, both standard regression and transfer function models were developed using the Portland system data.

Using the non-differenced data, a high degree of correlation was found among the input variables. Seven of the ten input variable combinations were highly correlated, with correlation coefficients of 0.60 or greater (see Table 5). Second, the residuals were highly correlated and not independent as required for regression models. Third, the delay in the response to service level changes would have been missed if only contemporaneous correlations were included in the model. Finally, the biased standard errors from the regression model would have erroneously lead to the conclusion that one of the variables (service level-suburban lines) was statistically significant when in reality, it wasn't. These results argue for the wider application of the appropriate statistical methodology when time-series data is used.

TABLE 4

COMPARISON OF STANDARD REGRESSION AND TRANSFER FUNCTION MODELS

	Comparison	Standard Regression	Transfer Function
1.	Correlated input variables	Yes, the input variables are highly correlated. Multicollinearity is present	No, data is differenced
2.	Autocorrelated errors	Yes, the error structure is highly autocorrelated, violating basic model assumptions.	Yes, but model structure allows for correlated errors
3.	Lag structure for input variables	No, only contemporaneous correlation assumed	Yes, methodology directly investigates the nature of dynamic relationships
4.	Coefficient estimates and standard errors	Estimates are in-efficient and the standard errors (and thus the significance tests) are biased.	Estimates are efficient and the standard errors are unbiased

TABLE 5

MULTICOLLINEARITY OF NON-DIFFERENCED DATA

Correlation Matrix - Input Variables	Service Level City Lines	Service Level Suburban Lines	Fare	Gasoline Price	Employment
Service Level City Lines	1.00	.96	.45	.85	.89
Service Level Suburban Lines	.96	1.00	.48	.88	.84
Fare	.45	.48	1.00	.80	.60
Gasoline Price	.85	.88	.80	1.00	.89
Employment	.89	.84	.60	.89	1.00

TABLE 6

COMPARISON OF COEFFICIENT ESTIMATES

STANDARD REGRESSION VS. TRANSFER FUNCTION MODELS

Variable	Coefficient Estimate and Standard Error	
	Regression	Transfer Function
Service level - city lines	.39 ± .21	.28 ± .17
Service level - suburban lines	.31 ± .12	.08 ± .06
Fare	-.30 ± .08	-.28 ± .07
Gas Price	.27 ± .07	.25 ± .11
Employment	.48 ± .09	.57 ± .26

II. PREVIOUS RESEARCH IN TIME-SERIES TRANSIT DEMAND MODELING

The methodologies used in the analysis of transportation changes over time can be divided into two broad categories: those that use before-and-after data only and those that attempt to develop structural relationships or models for the processes under study.

With respect to public transportation, before-and-after studies have focused primarily on the effects of fare changes implemented by various transit systems. Examples of these include efforts by Rainville (1948), Curtin (1968), and Kemp (1974). A number of such studies have been summarized by Barton Aschmann (1981) and Ecosometrics (1981).

Studies that modeled the relevant structural relationships have relied upon several different approaches. Three groups will be discussed here. The first group used traditional multiple regression techniques to relate changes in transit ridership over time to changes in service level, transit fare, and other factors. Agrawal (1978) used annual time series data for Philadelphia for the period 1964 to 1974. The independent variables used included transit fare, annual miles of bus service, jobs in Philadelphia, and a dummy variable for miscellaneous events or interventions. Uhlborg (1982) used monthly time series data for Seattle and his model included transit fare, gasoline price and supply, and employment. Agrawal did not specifically account for serial correlation in his data, while Uhlborg used first-differences to account for first-order serial correlation. Since the major purpose of both studies was to develop forecasting tools, no detailed assessment of past service or fare changes was considered. In the Uhlborg study, service level was not even found to be a statistically significant variable. It can be argued that using a total system measure of service level with no consideration of the lag structure that may be involved would tend to mask any effects in such a model. Use of annual data as by Agrawal would also preclude such an analysis. Cherwony and Polin (1977) used daily ridership data to develop models for new transit routes. They found that patronage on a new route generally builds up rapidly during the first few days after service has begun, and then stabilizes after a period of time. They used a logistic curve to approximate this relationship. Finally, Bates (1981) analyzed transit ridership for Atlanta using multiple regression models that included service level, transit fare, and gasoline price inputs.

The second group of studies used econometric methods to study the interrelated effects of ridership, service level, and travel costs in the context of supply, demand, and cost functions. These studies also were cognizant of the more complex error structure inherent in time-series analysis by using first-order and sometimes twelfth-order autocorrelation terms. Gaudry (1975) analyzed fifteen years of system level transit data for Montreal (1956-1971) using linear regression techniques in conjunction with Box-Jenkins procedures for specification of the rth-order autoregressive process of the error terms. He developed separate functions for demand, supply, and cost and used generalized least square procedures to estimate these models. He has expanded this work (see Gaudry and Wills (1978)) to include a more thorough analysis of the functional forms relating ridership to service level, and travel costs. Kemp (1981) used a simultaneous equations model using pooled time-series and cross-sectional data to estimate transit ridership for specific bus routes for the San Diego transit system. Separate equations were developed for passenger volumes, level of service, and service operation. Kemp used two-stage least squares methods with correction for first-order autocorrelation only. Moody (1976) and Schmenner (1976) have also considered econometric methods using multiple equations for demand, supply, and cost.

The third group used the techniques of Box and Jenkins to model changes in travel demand over time. Der (1977), Elder (1977), Holmesland (1979), Ahmed and Cook (1979), and Nihan and Holmesland (1980) developed autoregressive integrated moving average (ARIMA) models to study changes in traffic volumes over time. Harmatuck (1975) and Wang (1981) used intervention ARIMA models to study the effects of intervening events such as transit strikes and fare changes on transit ridership levels. McLeod, Everest, and Paully (1980) developed both univariate ARIMA models and transfer-function models to study air and rail passenger traffic between London and Glasglow. Polhemus (1976, 1979) studied air traffic volumes using both univariate and transfer function models.

This previous work has provided the basis for the research undertaken here. It is apparent that a number of factors influence changes in transit ridership over time, including service characteristics of the competing modes (e.g. auto and transit), transportation costs, and size of the travel market. While a complete specification of these input variables is necessary, Gaudry (1975) found that the interdependence of several key input variables affected his model estimation process adversely. Specification of transit level of service can also be a problem if this data is not further refined or categorized in any way. Note, for example, the difficulty that Uhlborg had in using bus hours as a surrogate for service level. In general, these studies did not attempt to review the functional forms that related transit ridership to the various explanatory variables, particularly with respect to any lag structure that might exist. (Exceptions are Gaudry (1978) and Uhlborg (1982) who included some delays in their models). This is particularly distressing since most theories of market response assume that there is usually a delay between the introduction of a change and the response to that change. The fact that transit demand and supply are interrelated functions must also be explicitly accounted for. Finally, the autocorrelation of the time series data (usually both month-to-month as well as year-to-year) must be more firmly introduced into accepted analytical methodologies. Some of the studies cited above do directly account for serial correlation; most, however, ignore it completely.

III. METHODOLOGY

The methodology that has been used in this research includes three phases. The first phase, Model Development, consists of postulating the form of the model, identifying the structural relationships between transit ridership and the input variables, estimating the model parameters, and checking the validity of the model. Impact analysis is the second phase. Here, the model that has been developed is used to determine the impact on ridership of a previous change in transit service level or fare. The final phase is Forecasting, in which the model is used to forecast future transit ridership levels.

PHASE I: MODEL DEVELOPMENT

Model Form

It is hypothesized that transit demand can be described as a function of level of service, cost, and market size. This approach has been variously used by Gaudry (1975, 1978), Kemp (1981a, 1981b), and Wang (1981, 1982), whose general model structures are given in Table 7.

A model structure suggested by theory must be tempered with the reality of the data that is actually available. The model considered here has been developed with this balance in mind. The model has the form:

(1) $$R_c = F(SL_t, TC_t, MS_t, S_t, I_t) + N_t$$

where R_t = transit ridership

SL_t = level of transit service

TC_t = travel costs by auto and by transit

MS_t = size of the travel market

S_t = seasonal factors such as weather

I_t = interventions such as gasoline shortages, marketing plans, etc.

N_t = the noise model or error structure

TABLE 7

GENERAL MODEL FORMS
PROPOSED BY OTHER RESEARCHERS

Gaudry (1975, 1978)

$$R_t = \rho_1 R_{t-1} + \rho_{12} R_{t-12} + \sum_{i=0}^{n} \beta_i X_{it} - \sum_{i=0}^{n} \sum_{\ell=1,12} \beta_i \rho_\ell X_{i,t-\ell} - e_t$$

where R_t = transit ridership
X_{it} = the independent variables (travel time, travel cost, comfort, activity levels, etc)
ρ = the autoregressive structure
e_t = the error term

Kemp (1981a, 1981b):

(1) Demand function:
Transit ridership is a function of bus fare, gasoline price, bus speed, passenger waiting time, number of school days.

(2) Performance function:
Bus Speed is a function of bus stop density and traffic congestion

(3) Supply function:
Seat miles operated is a function of patronage, variable cost per seat mile, subsidy level, available seat capacity

Equations were estimated using Two-Stage Least Squares with correction for autocorrelation.

Wang (1981, 1982)

$$R_t = \sum_{i=0}^{n} \beta_i X_{it} + U_t$$

where R_t = transit ridership
X_{it} = the independent variables (service level, bus fare, gasoline price, seasonal dummy variables, working day variables)
U_t = stationary ARIMA error terms

Equations were estimated using generalized least-squares.

The first issue to be considered with respect to model form is the level of change that can be expected for a given change in the input. In other words, does that relative change in transit ridership depend upon whether the change in the input is large or small? It is assumed here that changes in transit ridership resulting from changes in sevice level or travel costs are subject to the law of diminishing returns. That is, for a fixed market size, there is a maximum number of transit riders that can be expected to use the transit sytem (assuming no capacity constraint) even if service level is raised to an extremely high level and if the transit fare is zero. For a variety of reasons, some travelers must or will always use their automobile no matter how attractive public transit becomes. Thus, for each additional increment of service level that is added, for example, there will be a smaller increase in the number of new riders that result. Figure 6 illustrates the use of a log model that approximates a diminishing return function. While a more generalized functional form can be used, log transformations, which have other useful properties as well, have been used here.

The second issue with respect to model form is that of lagged response. Changes in service level, travel costs, or market size do not always result in instantaneous changes in transit ridership. It takes time for potential riders or current riders to hear about or perceive a change in the level of service, for example, and then make decisions about whether to change their pattern of usage. For this reason, the function relating transit ridership to changes in the independent variables must allow for these lag effects. While the form of the lag is unknown, it may have the form as shown in Figure 7.

Previously, the variables of the model were listed in general form as service level, travel cost, market size, and seasonal variation. The final issue with respect to model form is the specific form of the variables.

Service Level. One of the major determinants of transit ridership is the level of service available on the transit system. Most cross-sectional travel demand models use such measures as in-vehicle time, waiting time, and access time by transit and by auto for each origin-destination pair to describe level of service. In time series models, however, the data is simply not available at this level of disaggregation. Typically, time-series demand models use such measures as platform hours or miles of service as a surrogate for transit service level. (Exceptions are Gaudry (1975) and Kemp (1981), who each attempted to construct waiting time and in-vehicle time time-series for

FIGURE 6
EFFECT OF DIMINISHING RETURNS

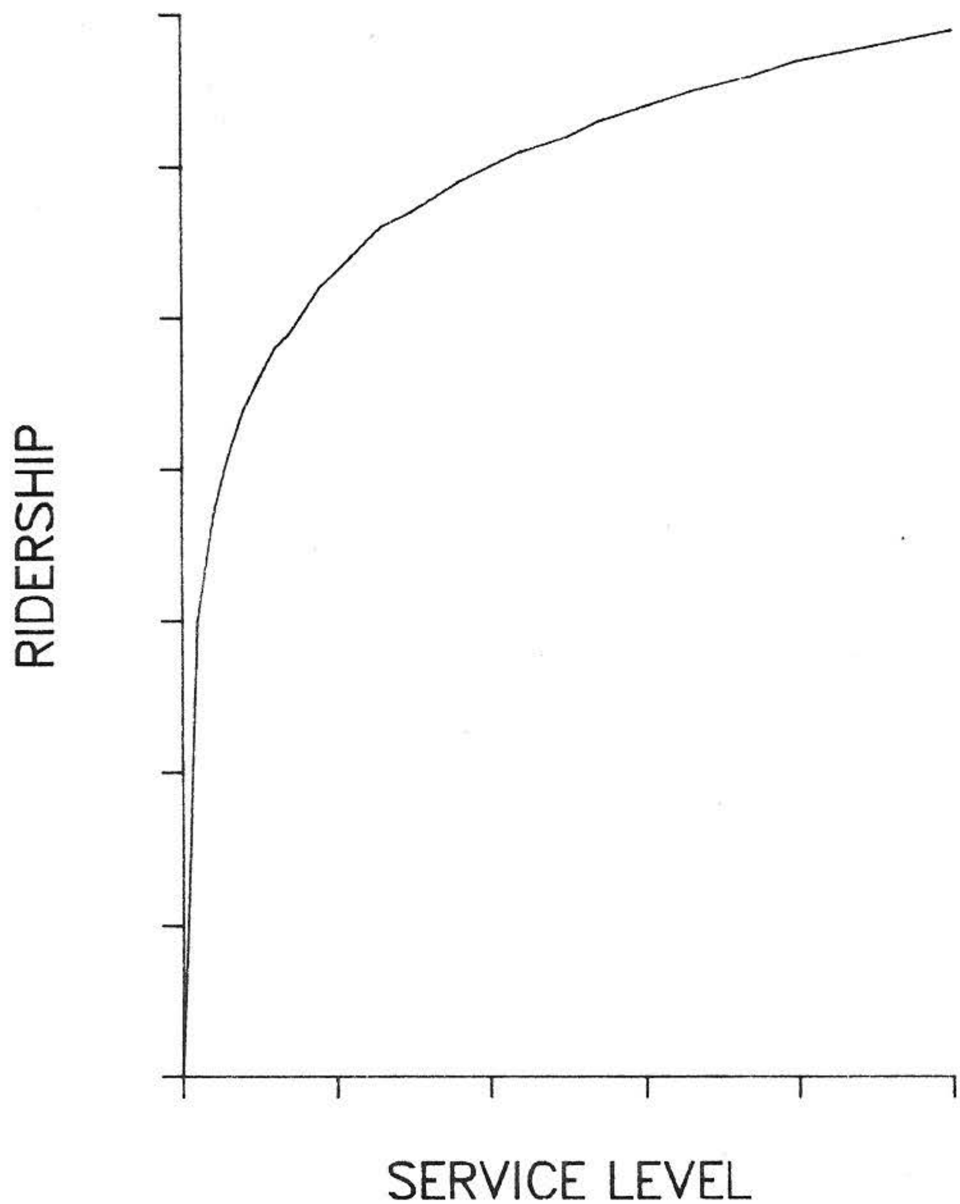
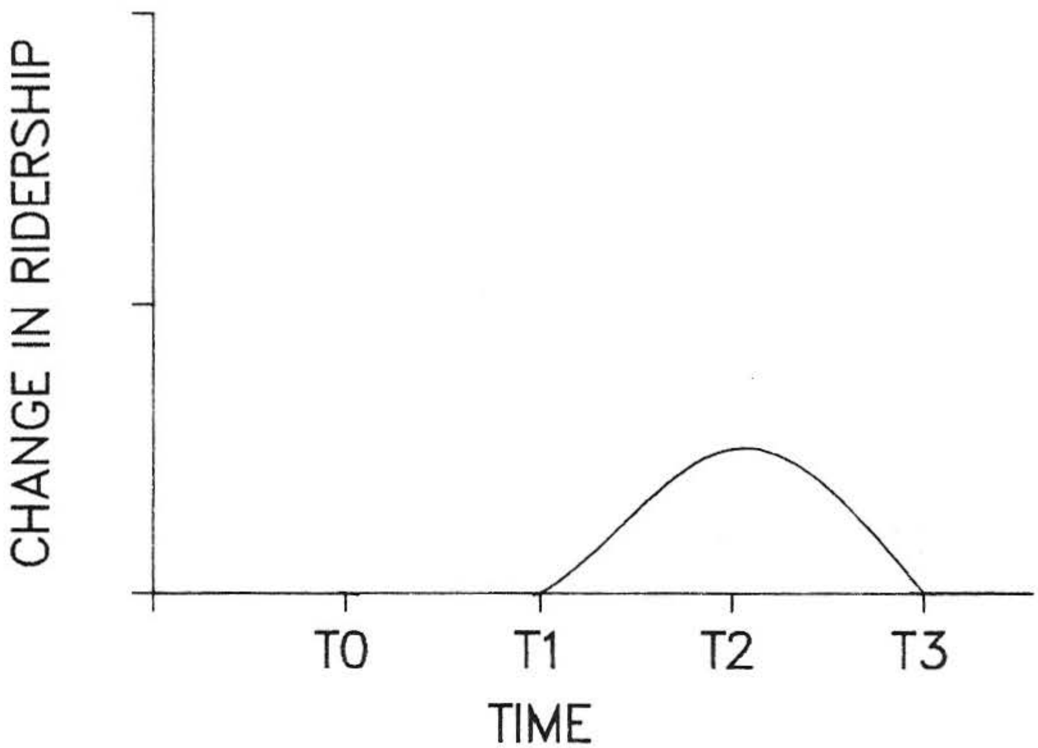


FIGURE 7
LAGGED RESPONSE



T0 = Time at which change in service level or fare is implemented
T1 = First response in ridership to the change is measured
T2 = Maximum response
T3 = Time at which response/effect disappears

Montreal and San Diego. Gaudry was working at the system level, while Kemp was working at the route level). Here, platform hours, platform miles, and route miles are used.

Platform hours and platform miles are gross measures of the amount of service provided each day, but each also includes non-service layover and deadhead time. Route miles describes the extent of the coverage of the system. Classification of the data by service change category (frequency of service, times of operation, network modification, new route, service reduction, and route elimination) provides a further useful refinement.

Combinations of these three variables are also of interest. Platform miles per platform hour yields a crude measure of system speed, while platform miles per route mile describes the intensity of service over a given network.

Service Level Descriptor	Variable
Gross Service Available	Platform hour or miles
Extent or coverage of network	Route miles
System speed	Platform miles/platform hours
Intensity of service	Platform miles/route miles

For the Portland data, at the route and sector levels, these variables are reasonable estimates for level of service. At the system level, the aggregation of service level into one variable such as platform hours results in a variable that is insensitive to the variation of ridership productivity by geographic sector of the service area. For this reason, the service level variable has been disaggregated by sector, even when using the system data.

Travel Cost. Two variables are used to describe travel cost: transit fare and gasoline price. Transit fare is the actual (average) cost for a transit trip, while gasoline price is a surrogate for the cost of an automobile trip. Assuming that trip lengths have remained fairly stable between 1973 and 1982, gasoline price is a reasonable estimate of auto travel costs. It can be argued on economic grounds that both transit fare and gasoline price should be deflated using the consumer price index (see Kemp for a discussion of this approach). However, it was found here that non-deflated prices are more directly correlated with transit ridership.

Market Size. Employment is used to describe the size of the travel market.

Seasonal Variation. Transit ridership varies in a seasonal manner for two major reasons. First, ridership declines in the summer are directly related to vacations from school and work. Second, adverse weather conditions during the non-summer months (particularly during the winter) often make transit more attractive than walking or using the auto. In regression analysis, seasonal variation must be specifically accounted for by dummy variables. Seasonal variation can be considered in the transfer function models more simply by adding a seasonal difference and/or a seasonal multiplicative component to the error structure of the model.

Other Variables. The variables listed above are the primary ones considered here. Others that could be tested include the effects of gasoline supply constraints (1973-1974 and 1979), marketing and promotional programs, and construction of capital facilities.

Identifying, Estimating and Checking the Model

The statistical methodology that has been used to develop these models has come to be known as the Box-Jenkins approach. This approach is based upon the philosophy that models should be parsimonious (or represented with the smallest possible number of parameters) and that model building should be iterative. That is, there is a logical sequence of steps and checks that should be followed when constructing a model and which may need to be repeated until a satisfactory model results. These steps include identification of a tentative model based upon various statistics constructed from the data itself, estimation of parameters for the tentatively identified model, and diagnostic checking for model adequacy. One of the most important aspects of this approach is that the form of the model is not assumed in advance but is inferred based directly upon the data. While theory may provide some guidance regarding which variables to include and the signs of the model coefficients, the analyst must look to the data for clues regarding the lag structure of the independent variables and the error structure of the model.

Tentative models are identified by analysis of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of a given series z_t. The ACF between z_t and $z_t + k$ (that is, at lag k) is defined as follows:

$$(2) \qquad ACF(k) = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\{E[z_t - \mu)]^2 E((z_{t+k} - \mu)^2]\}^{1/2}}$$

$$= \gamma_k / \gamma_0$$

where γ_k is the autocovariance of z at lag k

γ_0 is the variance $\sigma_z 2$ of the process z

The PACF can be described qualitatively as the autocorrelation that exists between z_t and z_{t+k} after all of the intervening correlation has been accounted for. The characteristics of the ACF and the PACF provide the guidelines for selecting an autoregressive (AR), a moving average (MA) or a mixed ARMA process for a given set of data. For example, a pure moving average process of order one has a non-zero ACF at lag one and is zero for all other lags, and a PACF that dies out exponentially for increasing lags. Such a process, known as an MA(1) process, is written as:

$$(3) \qquad z_t = a_t - \theta a_{t-1}$$

Conversely, a pure autoregressive process of order one has a non-zero PACF at a lag one and is zero for all other lags, and an ACF that dies out exponentially for increasing lags. An AR(1) process is written as:

$$(4) \qquad z_t = \phi z_{t-1} + a_t$$

The general autoregressive moving-average model of order (p,q) is written as:

$$(5) \qquad \phi_p(B)z_t = \theta_q(B)a_t, \text{ or}$$

$$(6) \qquad z_t-\phi_1 z_{t-1}-\phi_2 z_{t-2}-\ldots\phi_p z_{t-p} = a_t-\theta_1 a_{t-1}-\theta_2 a_{t-2}-\ldots-\theta_q a_{t-q}$$

where z_t is the time-series under consideration

$\phi_p(B)$ is the autoregressive polynomial of order p in B

$\theta_q(B)$ is the moving average polynomial of order q in B

a_t is a random shock that is $N(0, \sigma^2)$

B is the backshift operator, where $B(z_t)=z_{t-1}$

It is assumed that the series z_t is stationary, that is, it varies about some mean value. In addition, stationarity assumes that the covariance structure between z_t and z_{t+k} depends only upon the lag k, and not on t. Stationarity can frequently be induced in a series by taking differences; usually first or second differences are sufficient. When a series is non-stationary, and differencing is necessary, the more general autoregressive integrated moving average (ARIMA) model results:

$$(7) \qquad \phi_p(B)(1-B)^d z_t = \theta_q(B)a_t$$

where z_t is the time-series under consideration

$\phi_p(B)$ is the autoregressive polynomial of order p in B

$\theta_q(B)$ is the moving average polynomial of order q in B

d is the order of differencing necessary to induce stationarity in the series z_t

a_t is a random shock that is $N(0,\sigma^2)$

B is the backshift operator, such that $Bz_t=z_{t-1}$

Also of interest in the univariate case is the seasonal multiplicative model, which has the form:

$$(8) \qquad \phi_p(B)\Phi_P(B)(1-B)^d(1-B^s)^D z_T = \theta_q(B)\Theta_Q(B)a_t$$

where ϕ, d, θ are the non-seasonal components of the model

Φ, D, Θ are the seasonal components of the model

s is the seasonal period

The class of ARIMA models of particular interest here is the transfer function model, which can be written:

$$(9) \qquad Y_t = \sum_i \frac{\omega_i(B)}{\delta_i(B)} X_{it} + \frac{\theta(B)a_t}{\phi(B)}$$

Y_t is the dependent variable, or the transit ridership series in this case. The X_{it} terms are the independent variables or those factors that explain or effect the variation in Y_t. The polynomial ratio $\omega_i(B)/\delta_i(B)$ represents the lag structure associated with the variable X_{it}. The error structure is represented by the ARIMA model $\theta(B)a_t/\phi(B)$.

An example may help to illustrate this general form. Suppose that two factors, service level (SL) and transit fare (F) are found to affect transit ridership. Further, the effects of a service level change begin immediately and decay over the next several time periods, while transit fare has an impact one period after a fare change. Then the general model (9) can be written:

$$(10) \qquad R_t = (\omega_0/(1-\delta B)SL_t + \omega_1 F_{t-1} + \theta(B)a_t/\phi(B)$$

Note that the coefficient of SL_t is simply an exponentive decay term. Several methodologies exist for identifying the form of the transfer function model. The one used here is not unlike stepwise regression in which one variable is added to the model at a time. The following steps are included in this process.

Step 1. Difference each series of interest so that each is stationary.

Step 2. Analyze the ACF and PACF for the dependent variable (or output series) Y_t. The ARIMA model suggested for this series should then be used as the first approximation for the noise model of the transfer function model.

$$(11) \qquad Y_t = \theta(B)a_t/\phi(B)$$

Step 3. Add the first variable X_{1t} to the model with a lag structure sufficient to cover all lags possibly suggested by theory. Estimate the parameters of this model using generalized least squares methods.

$$(12) \qquad Y_t = \nu_o X_{1t} + \nu_1 X_{1t-1} + \nu_2 X_{1t-2} + \dots + \theta(B)\, a_t/\phi(B)$$
$$= \nu(B)x_{1t} + \theta(B)a_t/\phi(B)$$

Step 4. Analyze the coefficients $\nu(B)$ representing the lag structure for the variable X_{1t} and keep only those that are statistically significant ($\nu'(B)$) and of the correct sign. Re-estimate the model parameters using only those coefficients $\nu'(B)$.

$$(13) \qquad Y_t = \nu'(B)x_{1t} + \theta(B)a_t/\phi(B)$$

Step 5. Add the second variable X_{2t} and follow the procedure of steps 3 and 4. After analysis and re-estimation, the model will be of the form:

$$(14) \qquad Y_t = \nu'(B)x_{1t} + \nu'_2(B)x_{2t} + \theta(B)a_t/\phi(B)$$

Step 6. After all of the input variables have been added in this manner, and the significant ones identified and estimated, the model can be estimated in its more parsimonious form:

$$(15) \qquad Y_t = \sum_i \frac{\omega_i(B)}{\delta_i(B)} \quad X_{it} + \frac{\theta(B)a_t}{\phi(B)}$$

where Y_t is the output series

$\omega_i(B)/\delta_i(B)$ is the transfer function polynomial ratio

X_{it} are the input series

$\theta(B)/\phi(B)a_t$ defines the ARIMA noise model

B is the backshift operator

Step 7. Finally, the independence of the residuals a_t, the adequacy of the noise model $\theta(B)a_t/\phi(B)$ and the independence of the a_t series with each X_{it} series can be checked.

If all conditions are satisfied, the model is assumed to be in its final form. It should also be noted that a one-way relationship is assumed between X_{it} and Y_t; that is, X_{it} may cause changes in Y_t, but not vice-versa. While this assumption is a reasonable approximation for this case, it should be pointed out that, in fact, a two-way relationship does exist. For example, continued growth in transit ridership will eventually require an increase in capacity and thus in level of transit service provided. This case can be handled by the general multiple time-series model, but will not be covered in this report. For a discussion of the multiple time-series methodology, see Tiao and Box (1981).

PHASE 2: IMPACT ANALYSIS

The transfer function model developed in Phase 1 provides an indication of the "average" response of transit ridership to changes in service level or transit fare. The model is estimated based upon all of the service level or transit fare changes that occur during the period for which the data is available and thus the elasticities represented by the model coefficients represent the combined effect of all of these changes. If, however, the analyst desires to study the impact of one particular change, that change must somehow be isolated from the other changes that occurred during the study period. This can be achieved using intervention analysis.

Intervention analysis, developed by Box and Tiao (1975), is based upon the transfer function model but with the addition of a variable that represents one specific change or event. The event, which could be a strike, the implementation of a marketing program, or a period of gasoline shortage, is represented by a binary variable ζ_{jt} which assumes a value of zero before or after the event and a value of one during the time that the event or intervention is taking place.

The basic form of the transfer function model with intervention is:

$$(16) \quad Y_t = \sum_i \frac{w_i(B)}{\delta_i(B)} B^{b(i)} X_{it} + \sum_j \frac{w_j(B)}{\delta_j(B)} B^{b(j)} \zeta_{jt} + \frac{\theta(B)}{\phi(B)} a_t$$

The variables of equation (16) are the same as previously defined for equation (15), with the addition of the j intervention variables ζ_{jt}.

The following steps are included in the Impact Analysis:

Step 1. Identify, estimate, and check the transfer function model. This represents the Model Development Phase.

Step 2. Describe the past change whose impact is to be analyzed. Formulate an intervention variable to represent this change.

Step 3. Modify the data base to eliminate the effects of this change from the other data representing this variable. For example, if the impact of a previous five cent fare increase is to be analyzed, this increase should be "subtracted out" of the fare data. This is illustrated in Table 8.

Step 4. Re-estimate the model with the intervention variable included, as in equation (16). If the coefficient of the intervention variable is statistically significant, the coefficient represents the effect of the specific change under analysis. If the coefficient is not statistically significant (that is, not significantly different than zero), then the intervention had no measurable impact on transit ridership.

PHASE 3: FORECASTING

The transfer function model developed in Phase I can also be used to forecast future levels of transit ridership. But since the model depends upon several inputs, these variables must also be assumed or forecasted. Some of the input variables are under the direct control of the transit manager (e.g., service level and fare), and thus a given policy option (e.g., reduced fares) can be assumed. Other variables such as employment and gasoline price, however, are exogeneous and these must be forecasted directly. Forecasts of the input variables are accomplished by using "univariate" models. A univariate model for gasoline price is simply a model of today's gasoline price as a function of past values of gasoline price. For a further discussion of univariate models see Box and Jenkins (1976).

The following steps are included in the Forecasting phase:

Step 1. Identify, estimate, and check the transfer function model. This represents the Model Development Phase.

Step 2. Describe the nature of the forecast problem including the input assumptions and the length of the forecast period.

Step 3. Forecast the future values of the exogeneous input variables, such as employment and gasoline price.

Step 4. Using either the forecasted and/or assumed values for the input variables, forecast the future values of transit ridership.

The actual computations involved in transfer function forecasting are complex and are not described here. Several computer programs include the forecasting process and, once a transfer function model has been developed, are straightforward and easy to use. See, for example, SAS (1982) and SCA (1983) for further information.

TABLE 8

ELIMINATING THE EFFECT OF AN INTERVENTION

FROM ITS INPUT DATA BASE

Problem: Analyze the impact of a five cent fare increase that occurred in April 1980.

Period	Transit Fare	Modified Transit Fare Data With Effect of Intervention Removed
Jan. 1980	30¢	30¢
Feb. 1980	30¢	30¢
Mar. 1980	30¢	30¢
Apr. 1980	35¢	30¢
May 1980	35¢	30¢
June 1980	35¢	30¢

IV. CASE STUDY: PORTLAND, OREGON

OVERVIEW OF PORTLAND, OREGON

The Portland, Oregon metropolitan area includes 1.2 million people and covers over 900 square miles. The transit operator in Portland is the Tri-County Metropolitan Transportation District of Oregon (Tri-Met). Tri-Met was formed in 1969 by the Oregon legislature to take over the private bus operations within the City of Portland and to expand services into the rapidly growing three county area.

Starting from 50,000 weekday riders in 1970, ridership had grown to over 140,000 by 1980, averaging a nine percent annual growth rate. The three year period 1973-1976 saw a nearly twenty percent annual increase. Platform hours and miles increased at an annual rate of nearly seven percent between 1972 and 1982. The major period of expansion was from 1973 to 1976 when the annual growth rate was 14.5 percent. Area coverage, as measured by route miles, increased by 4.3 percent annually during this ten year period. Service level intensity (platform miles per route mile) increased by an annual rate of 11.5 percent from 1973 to 1976, but remained constant between 1976 and 1982.

By nearly all measures, auto travel costs increased significantly during this period, while transit travel costs declined. Gasoline price increased at a 15.6 annual rate during the ten year period, with the largest increase occuring between June 1979 and June 1980 when a 30 percent annual rate was recorded. Employment increased at an annual rate of between 2 and 5 percent until 1980 when it began to decline. Some of these trends are summarized in Table 9 and Figures 8 through 15.

Three basic data sets have been compiled and reduced. The first data set consists of monthly data for the transit system as a whole, a total of 114 data points covering the period January 1973 through June 1982. The second set consists of quarterly data for each of six geographic sectors of the Tri-Met service area. It includes 44 data points and covers the period Summer 1971 through Spring 1982. The third data set includes quarterly data for each Tri-Met bus route (a total of 71 routes) during this same time period. A summary of the data is given in Table 10.

TABLE 9

PORTLAND TRENDS, 1972-1982

	Annual Growth Rates			
	1972-1982	1973-1976	1976-1979	1979-1982
Weekday Riders	+ 8.9%	+19.4%	+ 3.2%	+ 8.5%
Platform Hours	+ 6.6%	+14.5%	+ 1.9%	+ 4.4%
Platform Miles	+ 6.8%	+17.4%	+ 0.3%	+ 3.8%
Route Miles	+ 4.3%	+ 5.7%	+ 1.4%	+ 2.8%
Platform miles per route mile	+ 3.5%	+11.5%	- 1.4%	- 0.4%
Platform miles per hour	+ 0.2%	+ 2.0%	- 1.4%	- 0.6%
Gas Price, unadj.	+15.6%	+15.6%	+ 6.8%	+23.8%
Gas Price, adj.	+ 6.9%	+ 6.5%	+ 0.5%	+21.4%
Fare, unadj.	+ 3.4%	- 7.4%	+ 9.3%	+11.7%
Fare, adj.	- 3.2%	-14.7%	+ 2.9%	+ 7.0%
Income, unadj.	+12.5%	+16.0%	+14.2%	+10.3%
Income, adj.	+ 4.7%	+ 2.8%	+ 7.5%	+ 0.8%
Population	+ 1.8%	+ 1.0%	+ 1.8%	+ 2.9%
Employment	+ 5.0%	+ 5.4%	+ 6.4%	+ 1.7%
Housing Units	+ 3.4%	+ 2.3%	+ 3.7%	---
CPI	+ 7.2%	+ 8.5%	+ 6.3%	+ 9.5%

Note: Growth rates were calculated using the month of January for the year shown.

FIGURE 8
TRANSIT RIDERSHIP, SYSTEM LEVEL
PORTLAND DATA

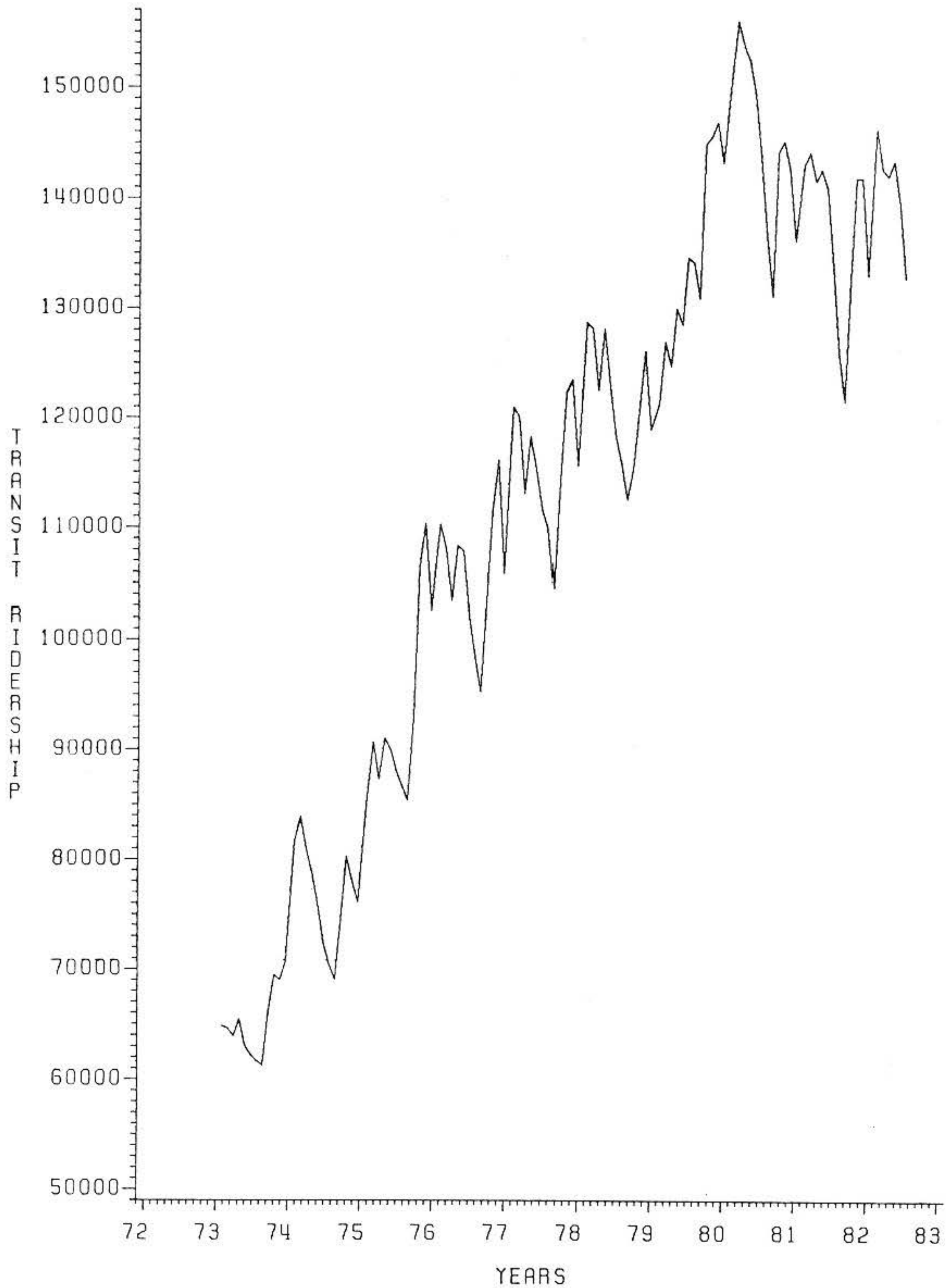


FIGURE 9
PLATFORM HOURS, SYSTEM LEVEL
PORTLAND DATA

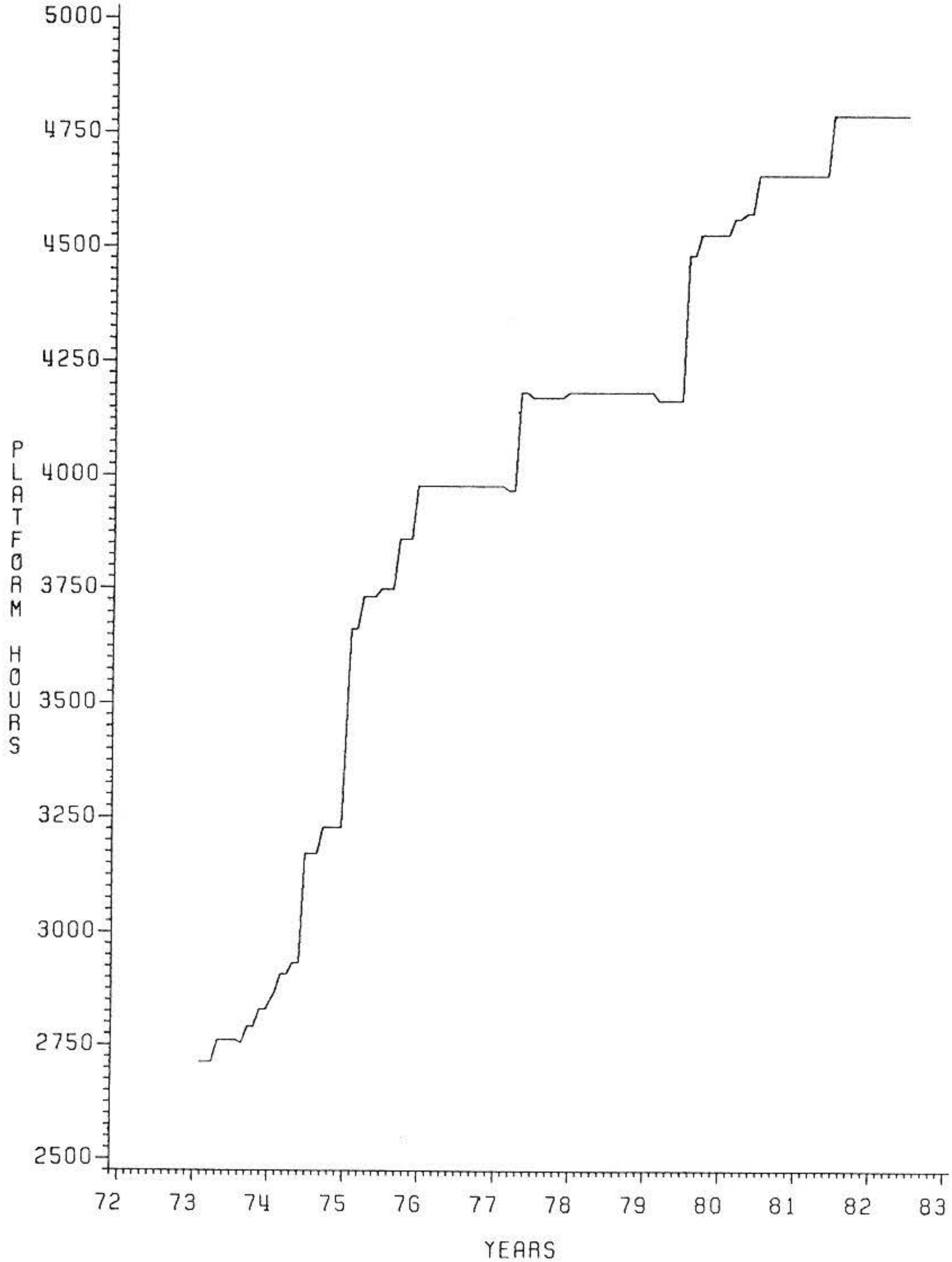


FIGURE 10
PLATFORM MILES, SYSTEM LEVEL
PORTLAND DATA

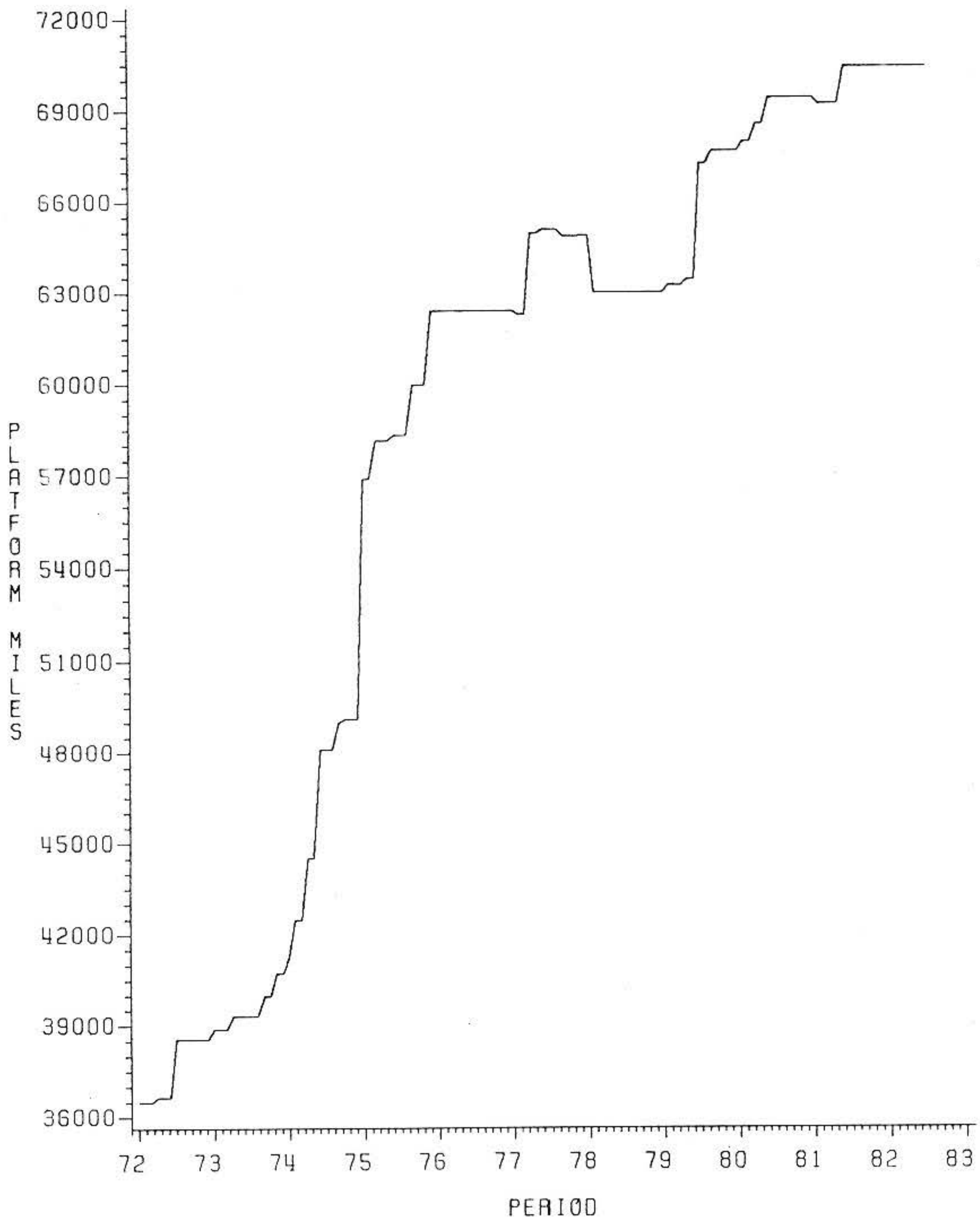


FIGURE 11
ROUTE MILES
PORTLAND DATA

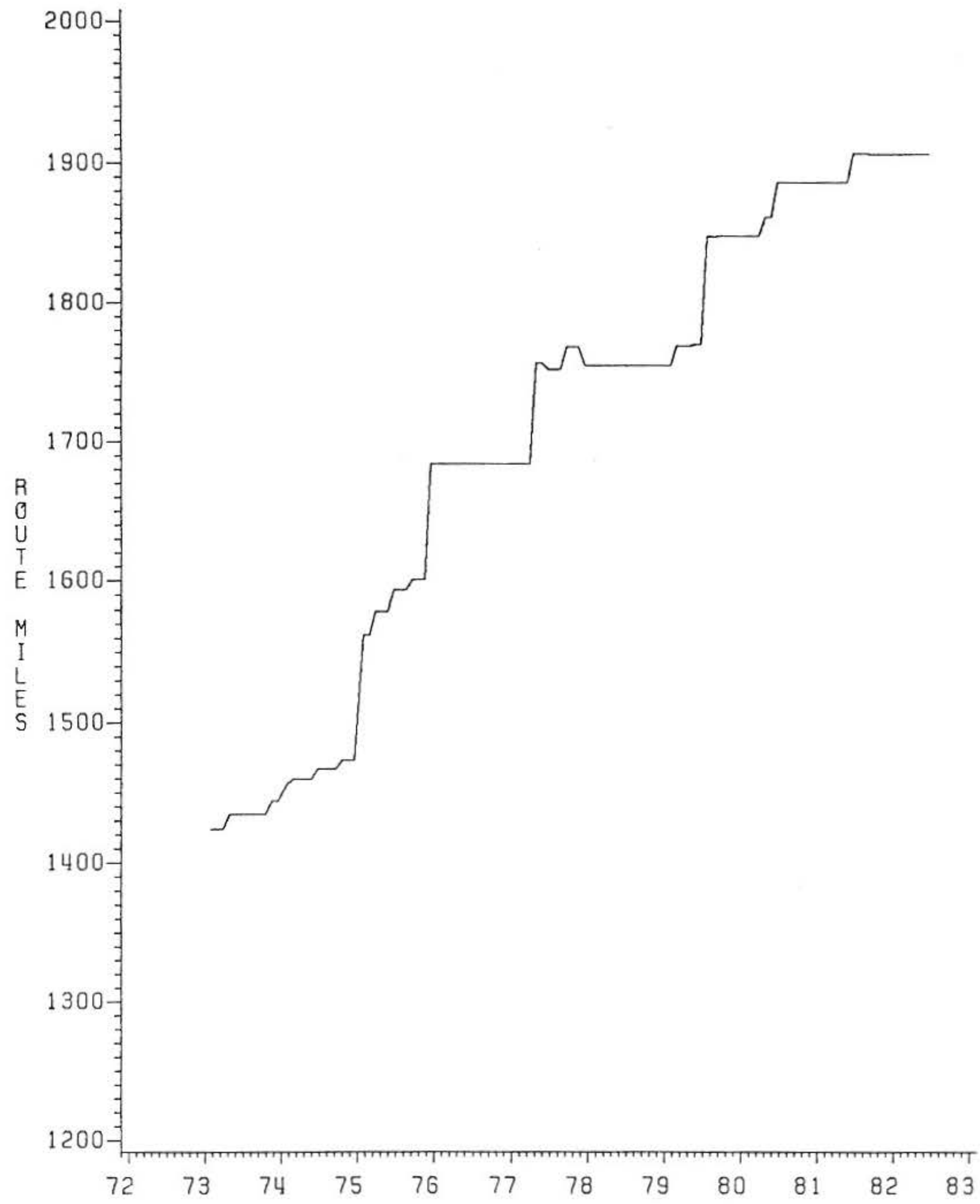


FIGURE 12
PLATFORM MILES PER ROUTE MILE, SYSTEM LEVEL
PORTLAND DATA

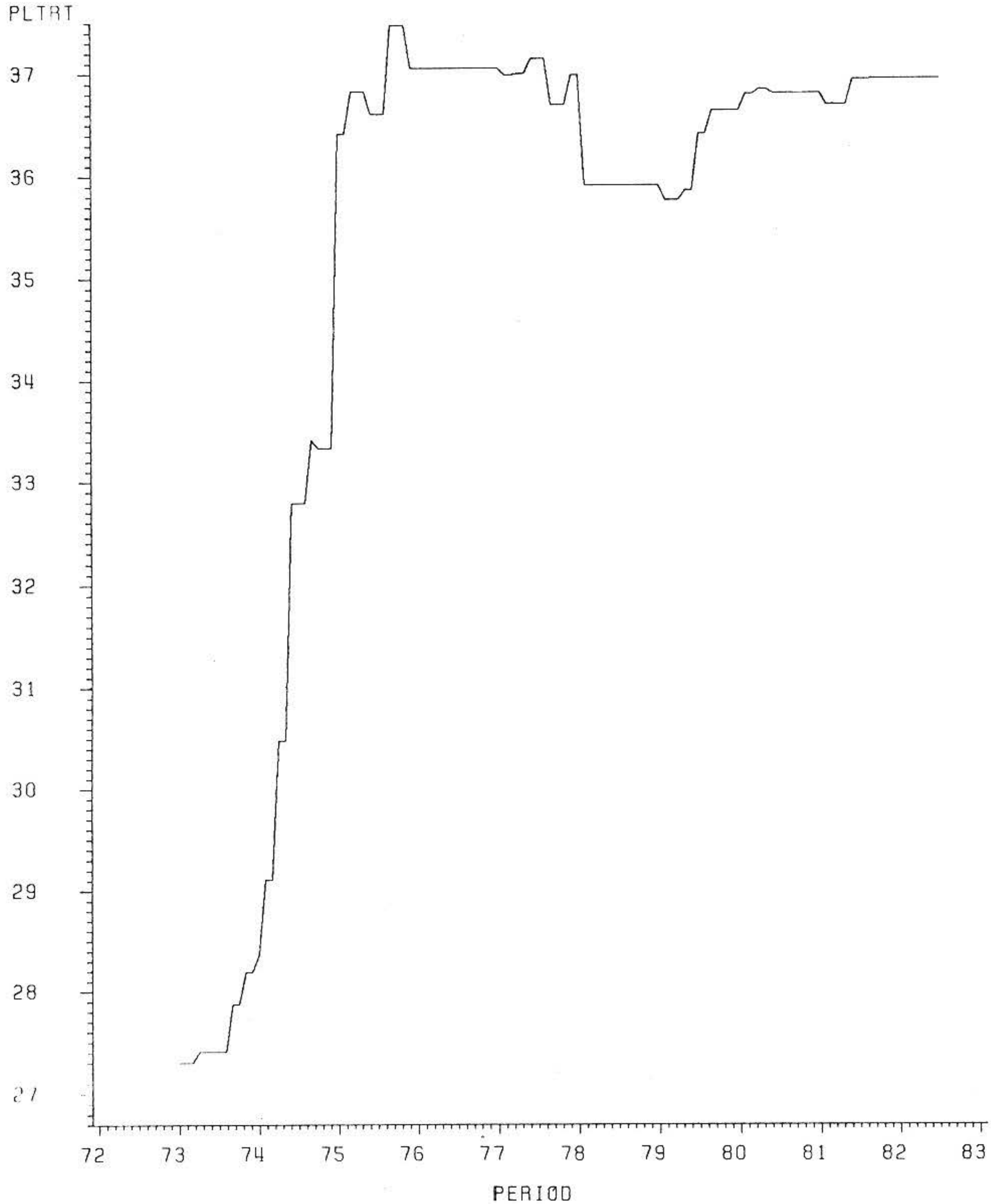


FIGURE 13
AVERAGE TRANSIT FARE
PORTLAND DATA

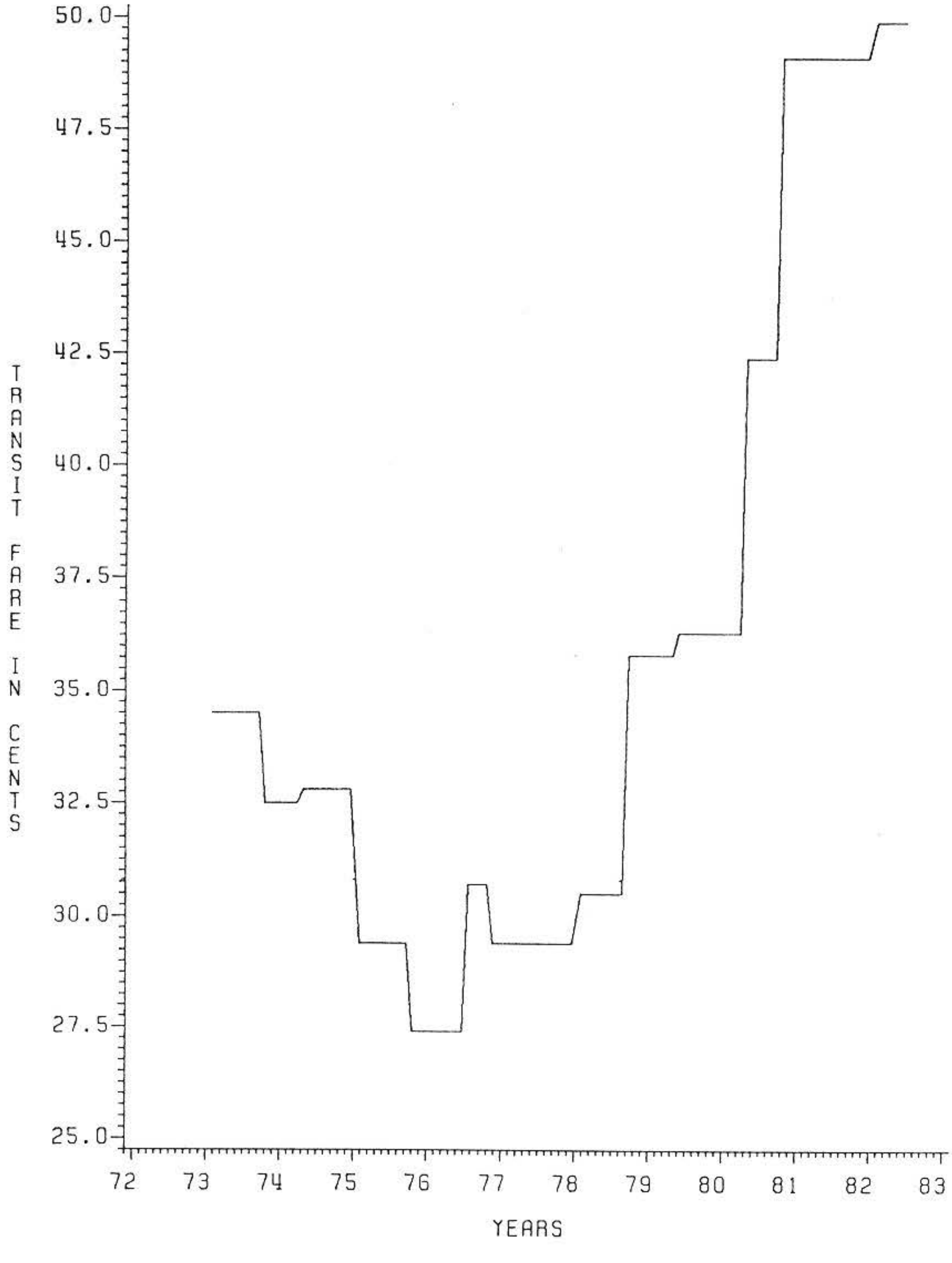


FIGURE 14
GASOLINE PRICE
PORTLAND DATA

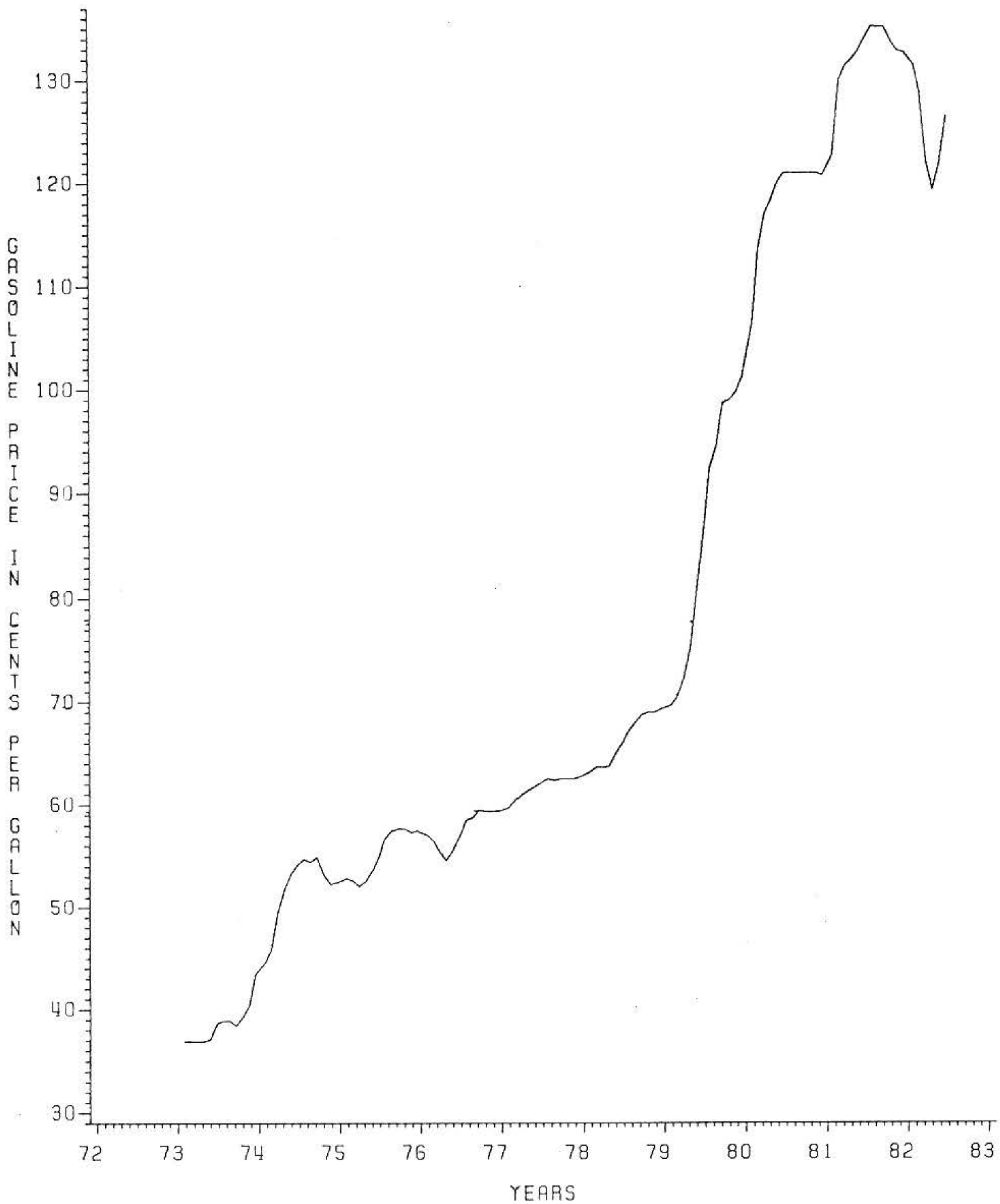


FIGURE 15
TRI-COUNTY EMPLOYMENT
PORTLAND DATA

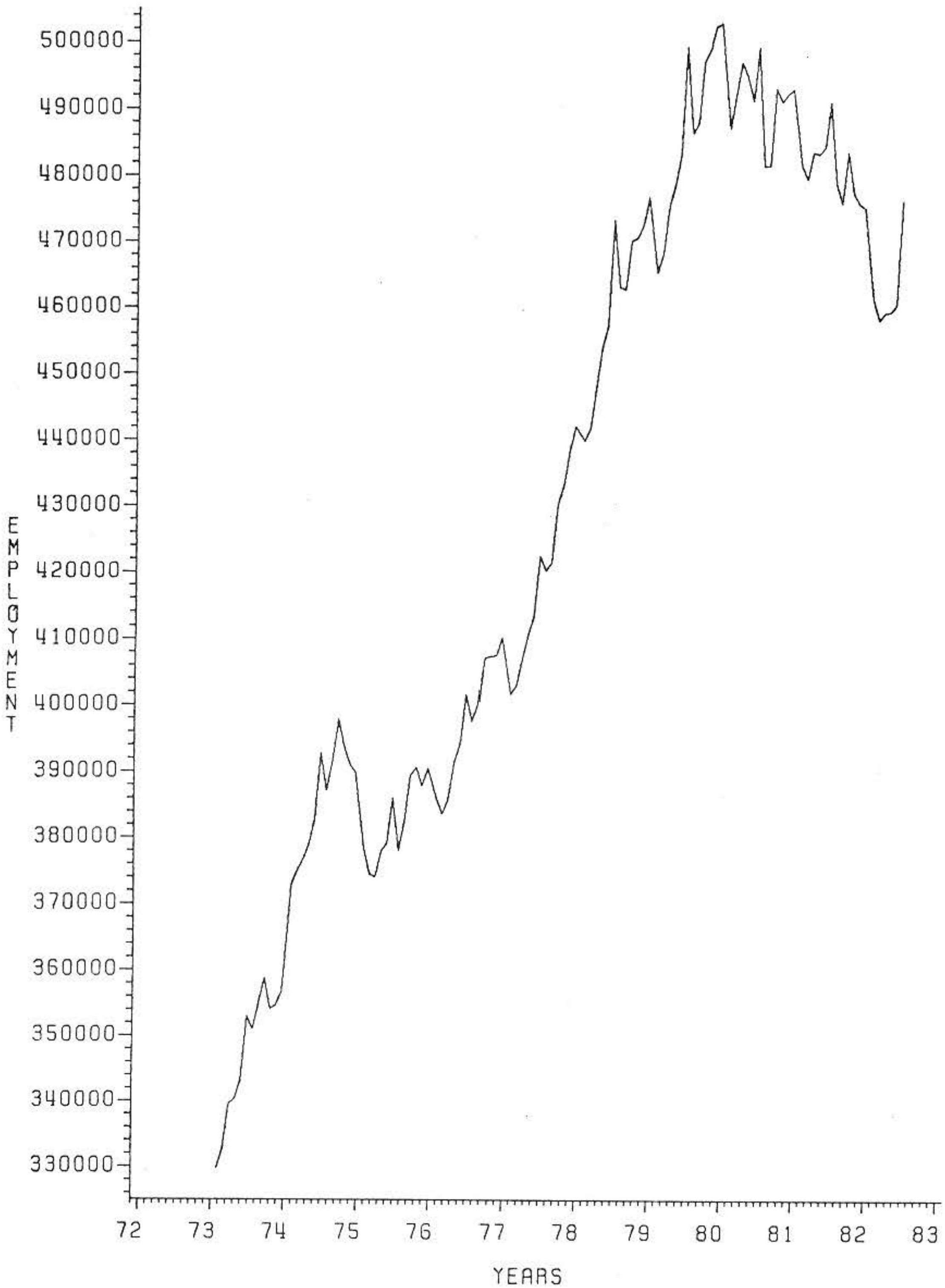


TABLE 10

SUMMARY OF PORTLAND DATA BASE

SYSTEM LEVEL DATA

Variable	Time-Series
Transit ridership	Average weekday originating transit riders
Service level	Daily platform bus hours Daily platform bus miles Daily route miles Daily platform miles per route mile
Travel costs	Average bus fare in cents Gasoline price per gallon in cents
Market size	Total employment by county

SECTOR AND ROUTE LEVEL DATA

Variable	Time-Series
Transit ridership	Total weekday boarding riders
Service level	Daily platform bus hours Daily platform bus miles Daily route miles Daily platform miles per route mile
Travel costs	Average bus fare in cents Base cash fare in cents Gasoline price per gallon in cents
Market size	Total employment by county

For the system level, a total of eight time-series are available. The transit ridership series is the total originating weekday riders, with transfer passengers excluded. Four series are available to describe transit service level: platform hours, platform miles, route miles, and platform miles per route mile. In addition, each series is categorized both according to geographic sector in which the service is provided, and by the category of service change. The service change categories include: change in the frequency of service or times or operation, modification of the service network, extension of a route, establishment of a new route, elimination of a route, or reduction in the service frequency or times of operation. Transit travel cost is measured by the average transit fare which includes consideration of the base cash fare plus any price discounts (i.e. monthly passes or senior citizen fares) that may be in effect. Auto operating cost is approximated by the current gasoline price per gallon. Market size is measured by the total employment by county.

The six sectors represented here include the aggregation of bus routes serving the following areas (see Figure 16).

1. Radial lines primarily serving the City of Portland only.
2. Crosstown lines primarily serving the City of Portland only.
3. Radial and local lines serving the City of Portland and East Multnomah County.
4. Radial and local lines primarily serving the Westside suburban area of Washington County.
5. Radial lines primarily serving the Southwest suburban areas of Washington and Clackamas Counties.
6. Radial and local lines primarily serving the Southeast suburban area of Clackamas County east of the Willamette River.

These areas are relatively homogenous with respect to transit service network, population density, and socio-economic characteristics of the population. Service level data for each sector is categorized into the same six groupings as described above for the system level data. Transit fare includes two different series: the average fare and the base cash fare. Gasoline price and employment are as described for the system level data.

The route level data is in the same form as for the sector level data.

MODEL DEVELOPMENT

As described earlier, it has been postulated that transit ridership is a function of service level, travel costs, and market size. But these relationships are often not contemporaneous. That is, a change in service level may not have an effect on ridership for several time periods. For example, it may take time for the public to hear about a change and to make decisions regarding travel patterns and habits. It is important, therefore, to determine the correlations that exist at different lags or the delay between a change in an input variable and when a change in the output variable can be measured.

Figure 17 illustrates the cross-correlation function for the system data relating transit ridership to service level, transit fare, and gasoline price. It indicates that ridership is affected by changes in service level at one month and at 8 to 9 months after a given change is implemented. Transit fare has an immediate effect on ridership (lag 0) as well as a smaller impact one month after a fare change. Gasoline price effects are completely contemporaneous. Once this lag structure is determined, the complete transfer function model can be estimated.

The system model was developed using service level, transit fare, gasoline price, and employment as input variables. Both platform hours and platform miles per route mile were used to approximate service level, with these variables further broken down by city and suburban service areas. Both models are summarized in Table 11. The error structures in both models included terms for lags 12 and 24 indicating a strong seasonal variation in transit ridership.

The sector models were developed using the same four variables. Platform hours, platform miles, and platform miles per route mile were variously used to approximate service level. The models are summarized in Tables 12, 13 and 14.

Route level models were developed for four city radial lines and five city crosstown lines. Transit service level, fare, gasoline price, and employment were used as independent variables. The models are summarized in Tables 15 and 16.

FIGURE 16

GEOGRAPHIC SECTORS OF THE PORTLAND REGION

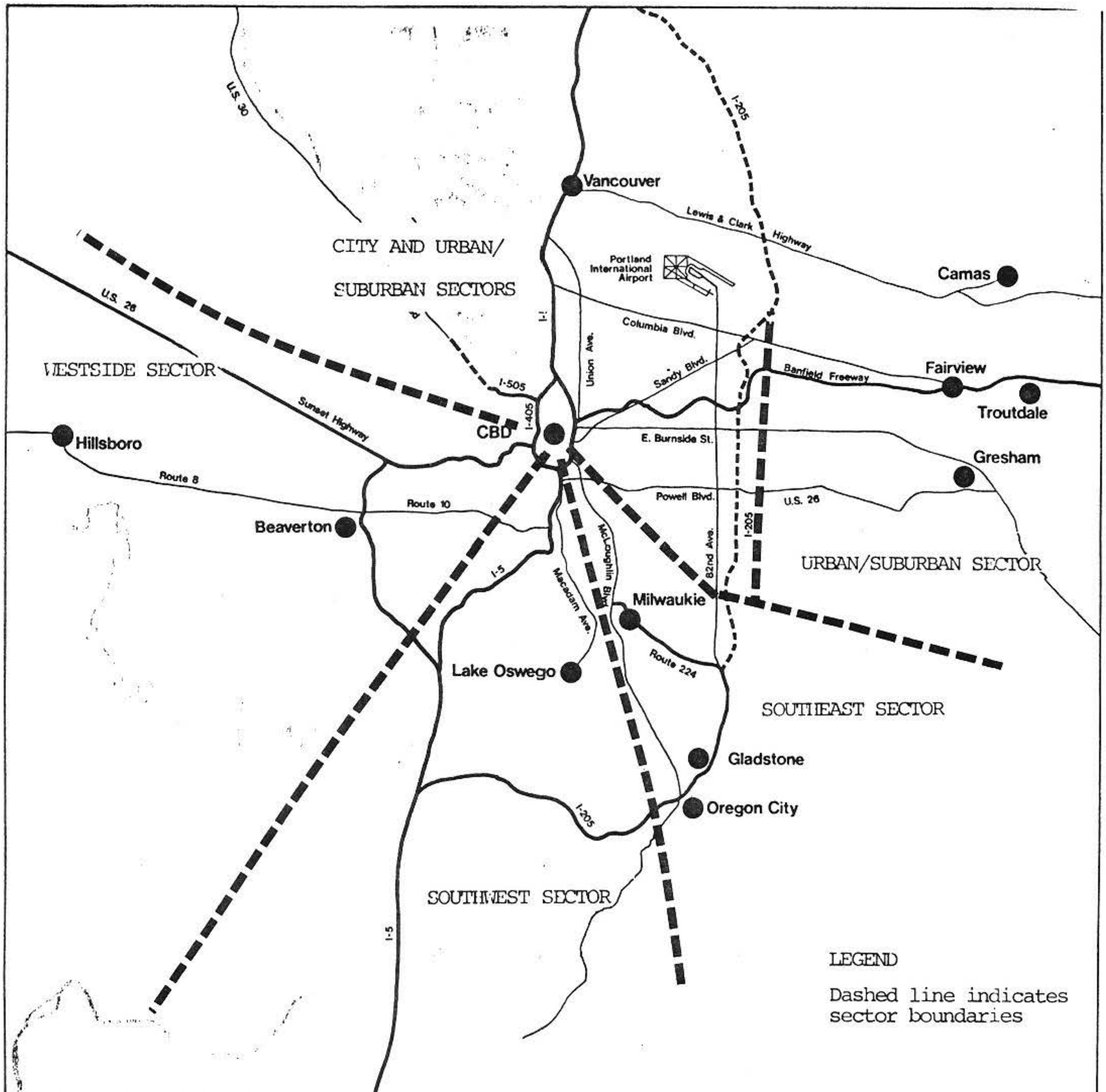


FIGURE 17
CROSS–CORRELATION FUNCTIONS RELATING
RIDERSHIP TO SERVICE LEVEL, FARE, AND GAS PRICE
SYSTEM DATA

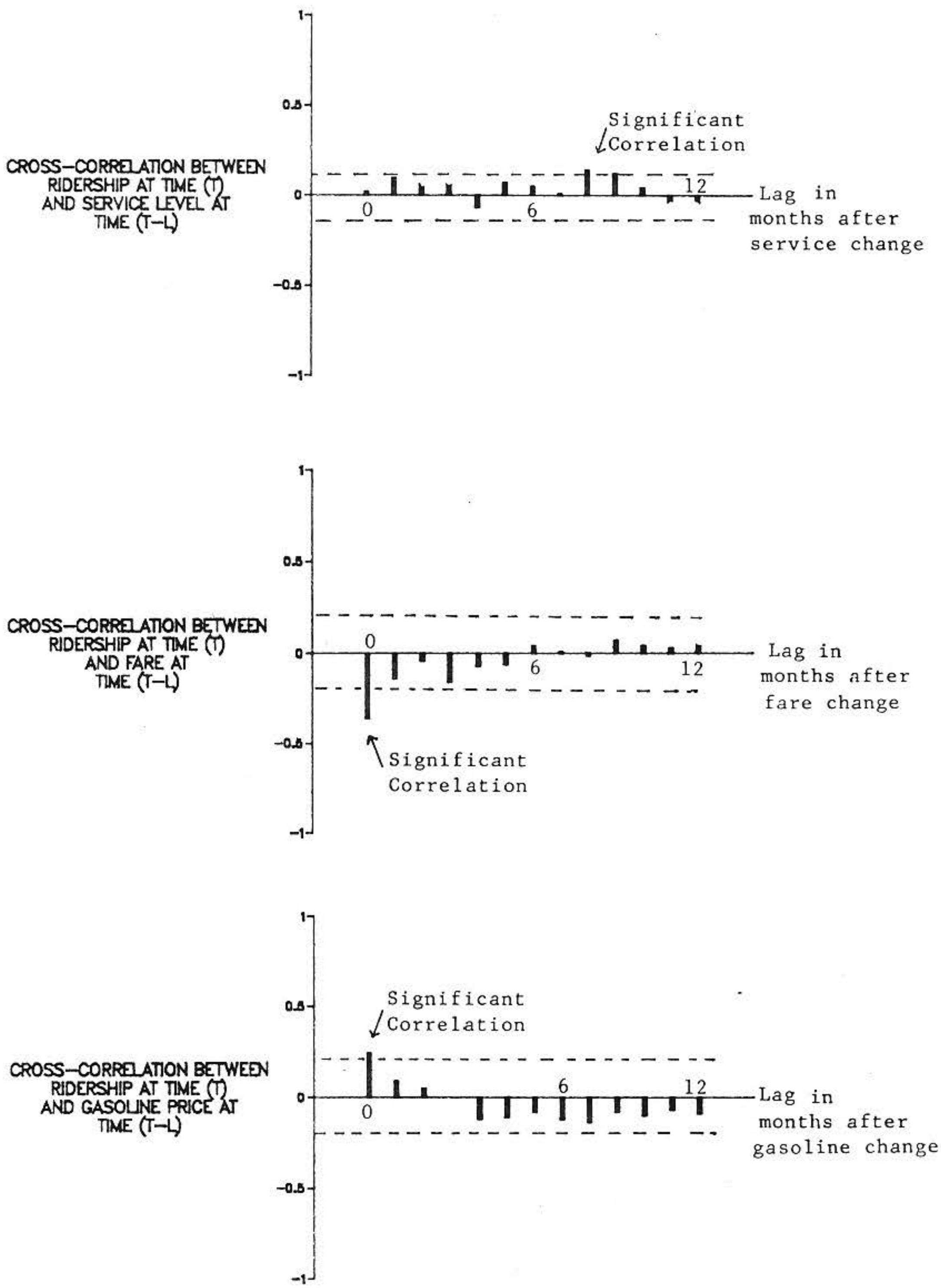


TABLE 11

SUMMARY OF SYSTEM MODELS

Model #1

INPUT VARIABLE	LAG (months)	COEFFICIENT	STANDARD ERROR
Service Level, City Lines (platform hours)	8	.31	.18
Service Level, Suburban Lines (platform hours)	1	.08	.07
Transit Fare	0	-.27	.07
Gasoline Price	0	.26	.12
Employment (Tri-county area)	0	.56	.28
Error Structure (a_{12})	12	.30	.11
(a_{24})	24	.31	.12

Model #2

INPUT VARIABLE	LAG (months)	COEFFICIENT	STANDARD ERROR
Service Level, City Lines (platform miles per route mile)	10	.29	.16
Service Level, Suburban Lines (platform miles per route mile)	1	.22	.09
Transit Fare	0	-.29	.07
Gasoline Price	0	.32	.12
Employment (Tri-county)	0	.49	.28
Error Structure (a_{12})	12	.29	.12
(a_{24})	24	.26	.12

TABLE 12

SUMMARY OF SECTOR MODELS

CITY RADIAL LINES SECTOR MODEL

INPUT VARIABLE	LAG (quarters)	COEFFICIENT	STANDARD ERROR
Service Level (platform hours)	2	.71	.28
Transit Fare	0	-.13	.10
Gasoline Price	0	.14	.07
Employment (multnomah county)	0	.43	.27
Error Structure (a_1)	1	.76	.13
(a_4)	4	.56	.16

CITY CROSSTOWN LINES SECTOR MODEL

INPUT VARIABLE	LAG (quarters)	COEFFICIENT	STANDARD ERROR
Service Level (platform miles)	0	.42	.21
	1	.18	.09
	2	.07	.03
	3	.03	.01
Transit Fare	0	-.42	.27
Gasoline Price	0	.39	.22
Employment (multnomah county)		N/S	
Error Structure (a_1)	1	.40	.17
(a_4)	4	.58	.16

TABLE 13

SUMMARY OF SECTOR MODELS

URBAN EASTSIDE LINES SECTOR MODEL

INPUT VARIABLE	LAG (quarters)	COEFFICIENT	STANDARD ERROR
Service Level (platform miles)	2	.55	.14
Transit Fare	0	-.15	.07
Gasoline Price	0	.18	.08
Employment (multnomah county)	0	.65	.30
Error Structure (a_1)	1	.75	.14
(a_4)	4	.67	.15

WESTSIDE SUBURBAN LINES SECTOR MODEL

INPUT VARIABLE	LAG (quarters)	COEFFICIENT	STANDARD ERROR
Service Level (platform miles)	0	.80	.07
Transit Fare	0	-.32	.08
Gasoline Price	0	.31	.08
Employment (Washington County)	0	.47	.12
Error Structure (a_3)	3	.49	.15

TABLE 14

SUMMARY OF SECTOR MODELS

SOUTHWEST SUBURBAN LINES SECTOR MODEL

INPUT VARIABLE	LAG (quarters)	COEFFICIENT	STANDARD ERROR
Service Level (platform miles per route mile)	0	.49	.19
Transit Fare	1	-.22	.09
Gasoline Price	0	.28	.17
Employment (tri-county)	0	.67	.29
Error Structure (a_4)	4	.99	.10

SOUTHEAST SUBURBAN LINES SECTOR MODEL

INPUT VARIABLE	LAG (quarters)	COEFFICIENT	STANDARD ERROR
Service Level (platform hours)	0	.32	.11
Service Level (platform hours)	2	.50	.11
Transit Fare	0	-.16	.13
Gasoline Price	0	.27	.16
Employment	1	.69	.30
Error Structure (a_1)	1	.30	.49
(a_4)	4	.19	.18

TABLE 15

SUMMARY OF CITY RADIAL LINES SECTOR AND ROUTE MODELS

	SERVICE LEVEL				FARE			GASOLINE PRICE			EMPLOYMENT			ERROR STRUCTURE		
MODEL	V	L	C	SE	L	C	SE	L	C	SE	L	C	SE	L	C	SE
City Radial Lines Sector Model	HR	2	.71	.28	0	-.13	.10	0	.14	.07	0	.43	.27	1	.76	.13
														4	.56	.16
Line 2 - St. Johns Route Model	HR	0	.91	.43	0	-.39	.20	0	.72	.15	2	1.14	.42	4	.64	.14
		2	.90	.43												
Line 3 - Fessenden Route Model	HR	0	.56	.25	0	-.46	.21	0	.69	.16	N/S			3	-.81	.13
		2	.51	.16	1	-.44	.19	1	.38	.09						
		3	.66	.25				2	.21	.05						
								3	.11	.03						
Line 6 - Sellwood/Union Route Model	HR	0	.23	.06	0	-.80	.20	0	.62	.27	0	.95	.64	1	-.47	.13
Line 8 - Irvington/ Jackson Park Route Model	HR	3	.25	.07	2	-.35	.23	0	.59	.29	N/S			1	-.43	.15
								1	.64	.28				5	-.49	.13
														8	.27	.14

Notes: V = the service level variable that is used for the particular model. HR indicates that platform hours was the variable used.
L = the lag, in quarters, for which the input variable affected transit ridership
C = model coefficient or elasticity
SE = standard error of the model coefficient
N/S = variable not statistically significant and was not included in the model

TABLE 16

SUMMARY OF CITY CROSSTOWN LINES SECTOR AND ROUTE MODELS

MODEL	SERVICE LEVEL V	SERVICE LEVEL L	SERVICE LEVEL C	SERVICE LEVEL SE	FARE L	FARE C	FARE SE	GASOLINE PRICE L	GASOLINE PRICE C	GASOLINE PRICE SE	EMPLOYMENT L	EMPLOYMENT C	EMPLOYMENT SE	ERROR STRUCTURE L	ERROR STRUCTURE C	ERROR STRUCTURE SE
City Crosstown Lines Sector Model	MI	0 1 2 3	.42 .18 .07 .03	.21 .09 .03 .01	0	-.42	.27	0	.39	.22	N/S			1 4	.40 .58	.17 .16
Line 71 - Killingsworth Route Model	I(31)	0	.72	.27	N/S			0	3.24	.63	N/S			4	.51	
Line 72 - 82nd Avenue Route Model	I(19) HR	0 0	.64 .30	.15 .23	N/S			3	.68	.30	N/S			1	.42	.15
Line 73 - 102nd Avenue Route Model	N/S				N/S			0	.60		N/S					
Line 75 - 39th Avenue Route Model	N/S				N/S			3	1.72		N/S			1 4	.40 .43	
Line 77 - Northeast/ Northwest Route Model	I(26)	0	.35	.11	N/S			2	.24	.18	N/S			1 2	.56 .42	.16 .18

Notes: V = the service level variable that is used for the particular model. HR = platform hours, MI = platform miles, and I(t) indicates an intervention variable at quarter t.
L = lag, in quarters, for which the input variable affected transit ridership
C = model coefficient or elasticity
SE = standard error of the model coefficient
N/S = variable not statistically significant and was not included in the model

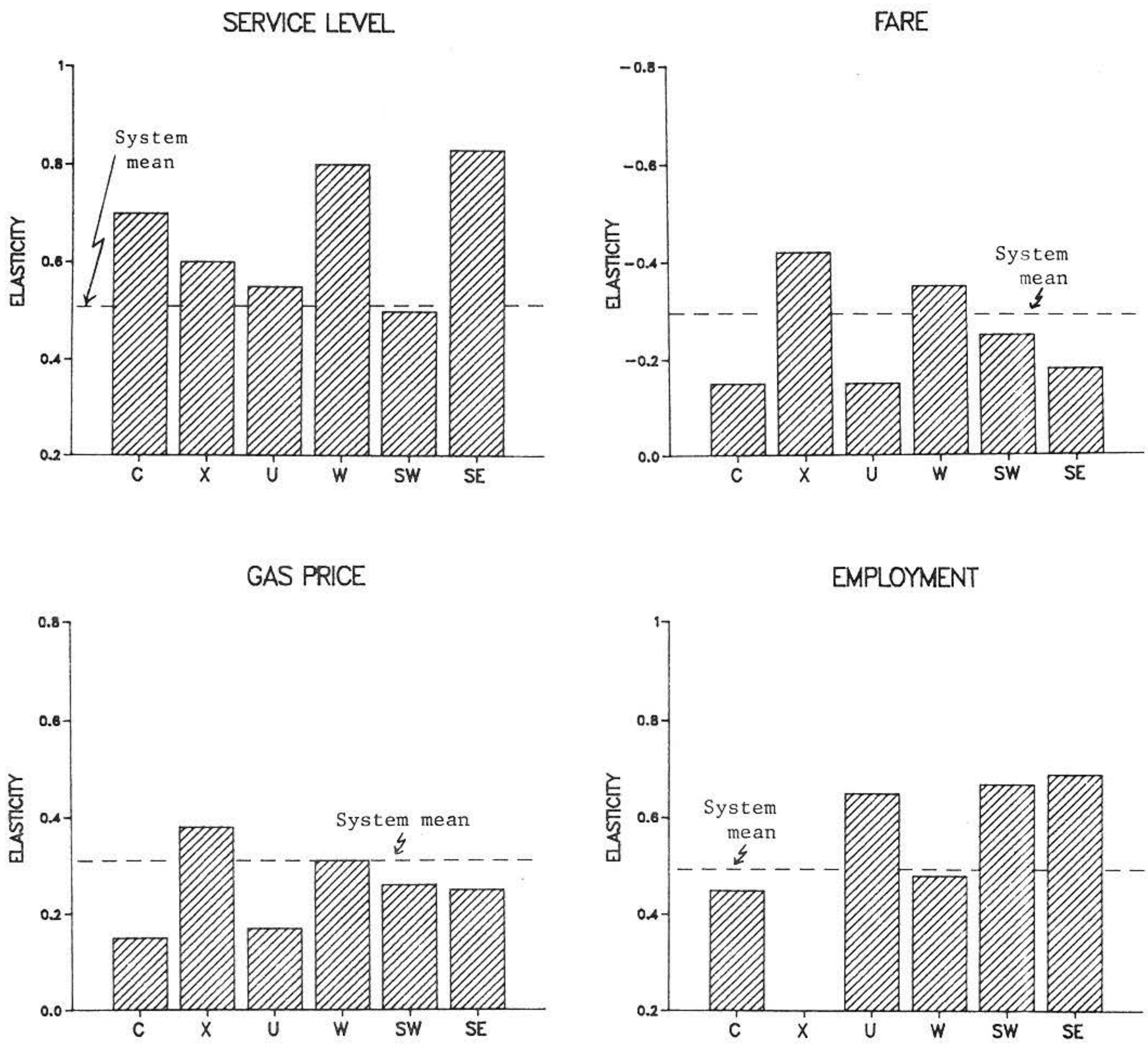
CONCLUSIONS REGARDING MODEL DEVELOPMENT

Several useful conclusions can be drawn from an analysis of the sixteen models that have been developed.

1. The Box-Jenkins approach for identifying, estimating, and checking transfer function models is an appropriate one for modeling the variation in public transit ridership over time. It has several useful attributes of particular interest here. The cross-correlation analysis clearly identifies the lag structure that exists between transit ridership and the input variables. The error structure is identified directly from the autocorrelation and partial autocorrelation functions of the output (ridership) series. This error structure is often complex because of both the month-to-month and year-to-year serial correlation in the output series. Misspecification of this noise structure, which often occurs when standard regression techniques are used with time-series data, yields spurious estimates of the model parameters and their statistical significance. See, for example,the work of Wang and others (1982) where a re-estimation of the Bates (1981) Atlanta models yielded elasticities that were half the original estimates.
2. For many of the models that have been developed, all four input variables were statistically significant contributors. This confirms the basic theoretical approach assumed here that these four variables in combination yield models for transit ridership that include sufficient explanatory power to allow for analysis of past changes as well as to forecast future changes.
3. Four different input variables were used to measure transit service level: platform hours, platform miles, route miles, and platform miles per route mile. Route miles was typically the least effective in approximating service level, leading to the conclusion that area coverage itself is not a good predictor of transit ridership. The other three variables were good measures of level of service, but their relative effectiveness varied from model to model. More analysis is needed before conclusions can be drawn regarding the situations most appropriate for each variable.
4. Typically, the impacts of service level changes in the urban sectors lagged about two quarters or eight to ten months. Suburban service changes usually showed less of a delay, often one quarter or one month.

5. Average transit fare was used in the system model while both average fare and base cash fare were used in the sector models. Average fare was more effective in the city lines sector model and the southeast sector model, while cash fare was superior in the southwest sector model. Both series were nearly equally effective in the crosstown, urban eastside, and westside sector models. The systemwide fare elasticity was estimated to be 0.27. Sector fare elasticities showed two basic groups above and below this level: the crosstown and westside sectors at -0.40 and the other sectors at about -0.20. Effects of fare changes typically were instantaneous, usually occuring in the same period as the implementation of the fare change. However, some effects were measured for one month after a fare change for the system data thus suggesting an exponential decay function. Also, the effect of fare changes for the southwest sector was delayed by one quarter.
6. Gasoline price and employment series were effective inputs in nearly all models. The system gas price elasticity was estimated to be 0.26. Sector elasticities ranged from 0.19 to 0.40. Employment elasticities varied from 0.5 to 0.7. The effects of gasoline price changes were also usually instantaneous. However, strong correlations were identified at negative lags (up to two months or one quarter) thus showing some anticipatory effects of known future price changes. While transfer functions cannot handle this kind of two-way relationship, multivariate ARIMA models (the subject of future research efforts) can. The effects of employment changes were also instantaneous for all model groups except for the southeast sector which included an employment variable lagged one quarter.
7. The most price sensitive sector was found to be the crosstown lines sector which had high elasticities for both transit fare (-0.39) and gasoline price (0.40). The most price inelastic sectors were the city radial, eastside urban, and southwest sectors.
8. There was generally good consistency in the coefficients and lag structures when comparing models estimated at the system, sector, and route levels. This can be seen particularly in the system and sector model comparisons shown in Figure 18. Much more variation, however, shows up in the sector to route level comparisons (Figures 19 and 20). This larger variation at the route level is likely the result of using

FIGURE 18
CONSISTENCY OF MODEL COEFFICIENTS BETWEEN SYSTEM AND SECTOR MODELS



LEGEND:
C City Lines Sector
X Crosstown Lines Sector
U Eastside Urban Sector
W Westside Suburban Sector
SW Southwest Suburban Sector
SE Southeast Suburban Sector

FIGURE 19
CONSISTENCY OF MODEL COEFFICIENTS BETWEEN CITY RADIAL LINES SECTOR AND ROUTE MODELS

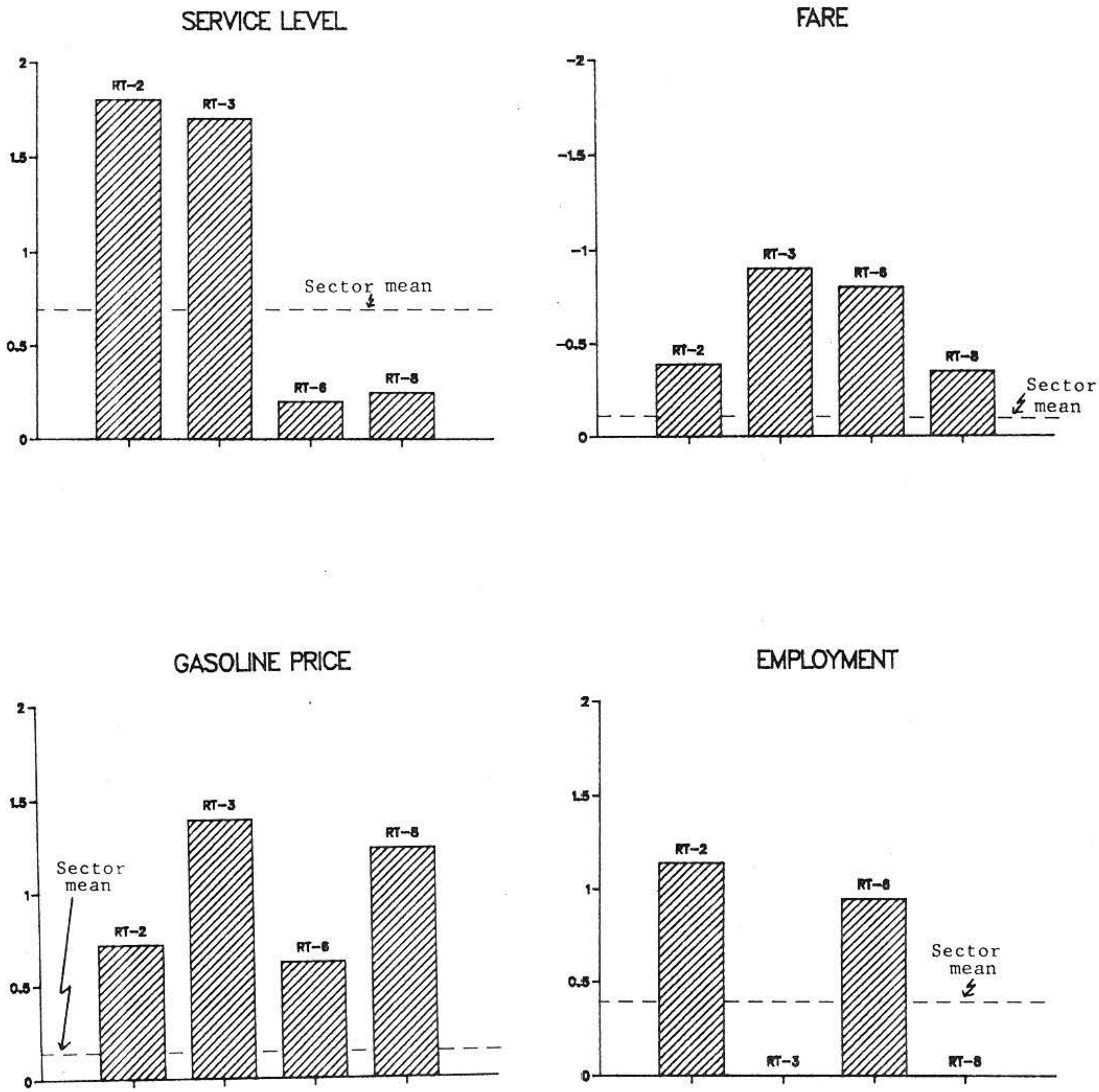
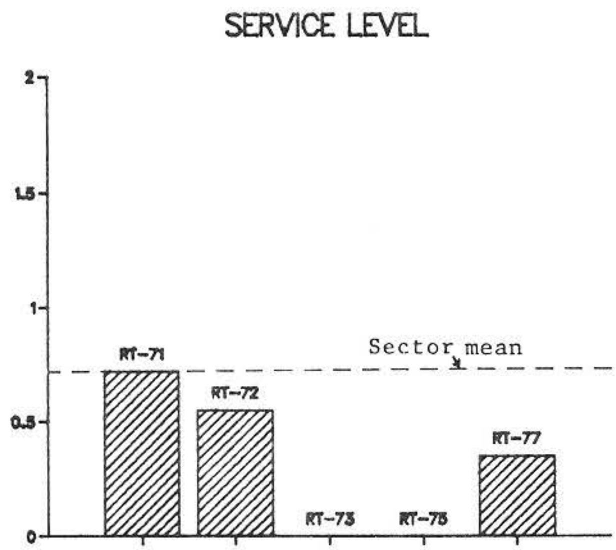
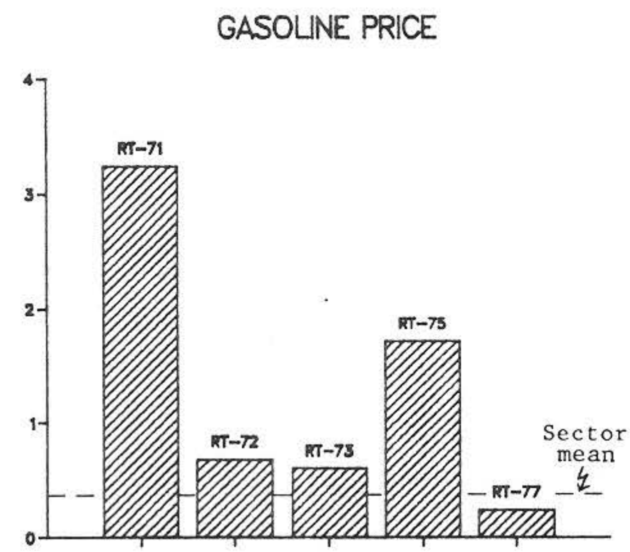


FIGURE 20
CONSISTENCY OF MODEL COEFFICIENTS BETWEEN CITY CROSSTOWN LINES SECTOR AND ROUTE MODELS





aggregated data to explain more disaggregate or micro-level changes. This is particularly true for the measurement of market size at the route level. The lowest level of disaggregation for employment is by county. Thus, for individual routes, it is obviously only an approximation of the market size. But while it can't account for the small-scale variation in trip-making activity along a given route, its statistical significance in some of the route models and most of the sector models indicates that it can account for some of the larger scale variation in market size that occurs over time.

9. Models for four individual routes from the city radial sector were developed. In total, the city radial lines sector includes eighteen radial routes that operate exclusively in the City of Portland. These routes serve the highest density and most transit dependent sections of the Portland metropolitan area and typically have the highest levels of service of any of the routes in the Tri-Met system. Service level elasticities tended to be higher on routes with lower levels of service (Routes 2 and 3) than for those routes with higher levels of services (Routes 6 and 8). The varying lag structures for service level, fare, and gasoline price indicates the different level of response for the rider populations served by each route. Finally, the lack of inclusion of employment for two of the route models shows the difficulty of trying to capture route level variation in market size using a county level variable.

10. The City Crosstown sector includes the eight crosstown routes that operate (or did operate) in the eastside section of the City of Portland during the study period. Service levels on these routes are typically poor (headways vary from 20 minutes to 60 minutes) and since the downtown is not served directly, the potential market from which these routes can draw is relatively small. The result is that most riders on the crosstown routes are captive riders. Models were developed for the five routes that operated through at least 37 of the 44 quarters of the study period. Employment was not a significant variable in any of these models, supporting the earlier statement that users of the crosstown routes were transit dependents (largely students), and illustrating the problem of using the county level data at the route level. The insignificance of fare reflects the transit dependent nature of the users as well. However,

gasoline price changes did have a significant effect on inducing new riders to the crosstown lines, particularly as service levels were increased in the late 1970's.

IMPACT ANALYSIS

The elasticities computed in the model development phase represent an average elasticity for a given variable over the entire study period. If four changes were implemented during a given period, for example, the service level elasticity would be an average of the impact of each service change. However, to study the impact of a specific service level change, an intervention variable which represents that change alone, must be added to the model. The model is then re-estimated with the intervention variable and the coefficient yields the effect of the specific change under study. If the variable coefficient is not statistically significant, it can be concluded that the change had no measurable impact on ridership.

Eleven service changes implemented between 1973 and 1979 on seven different routes were analyzed using the intervention analysis technique. The service changes included frequency improvements, route extensions, and service reductions. The results are given in Table 17.

Four of the service changes did not result in measurable impact transit ridership. While this sample is too small to yield any useful inferences, it can be noted that three of these changes were route extensions. This result may be consistent with that previously found in the use of route miles as a variable to represent service level: increased area coverage or extensions of service into new areas may not provide many new transit riders.

One of the route extensions that did have a significant impact on ridership was that implemented on Route 72 in 1976. The mean service level elasticity from the Route 72 transfer function model was estimated to be 0.45. When the intervention variable was added to the model, the mean service level elasticity dropped to 0.30. The route extension itself, when modeled with an intervention variable, was estimated to have an elasticity of 0.55. This example shows the value of modeling individual changes and how the elasticity for a specific change might vary from the mean value.

TABLE 17

IMPACT ANALYSIS OF PAST SERVICE CHANGES

AT THE ROUTE LEVEL

Route	Date	Type of Change	Significant Impact	Coefficient of the Intervention Variable	Mean Elasticity from Model
2	1975	Frequency improvement	Yes	.13	1.81
	1978	Route extension	No impact	-	
3	1973	Frequency improvement	Yes	.11	1.73
	1974	Frequency improvement, Route extension	Yes	.13	1.73
	1978	Service reduction	No impact	-	
6	1974	Route extension	No impact	-	
	1975	Frequency improvement	Yes	.23	.23
71	1979	Frequency improvement, Route extension	Yes	.72	.72
72	1976	Route extension	Yes	.81	.55
75	1979	Route extension	No impact	-	
77	1979	Frequency improvement	Yes	.35	.35

FORECASTING

The system model was used to make forecasts of future transit ridership. First, the model was re-estimated using only the first 102 data points. The last twelve data points were "held back" to enable a comparison with the forecasted values. The model coefficients in the forecasting model are approximately the same as for the "full" model (See Table 18). The difference in the employment coefficient probably results from the fact that the last twelve data represent a period of major decline in employment.

The next step is to determine the nature of the forecast to be made. The two variables over which transit management has control are service level and fare. It was assumed that these variables would take on their historical values for the forecast period, July 1981 to June 1982. Gasoline price and employment, the other two model inputs, are variables that are beyond the control of transit management. Thus, these variables must themselves be forecasted. To accomplish this task, univariate ARIMA models were identified and estimated. Essentially, the univariate models state that a present value of the variable (e.g., gasoline price) depends only on some combination of past values of the variable itself. The forecasted values for gasoline price and employment, as well as the assumed values for service level and fare, are given in Table 19.

Finally, transit ridership was forecasted. The results given in Table 20 and Figure 21, show a mean absolute percent error of only 2.1% for the twelve monthly ridership forecasts.

COMPARISON WITH STANDARD REGRESSION MODELS

It has been traditional to use multiple regression models when developing models relating transit ridership to explanatory variables. Using time-series data with regression models, however, invariably leads to a variety of statistical problems. Table 21 highlights the major areas in which problems are likely to arise by contrasting standard regression with transfer function models: multicollinearity, autocorrelated errors, lag structures, and coefficent estimates and standard errors. To determine whether these problems would, in fact, result, both standard regression and transfer function models were developed using the Portland System data.

Using the non-differenced data, a high degree of correlation was found among the input variables. Seven of the ten input variable combinations were highly correlated, with correlation coefficients of 0.60 or greater (see Table

22). Second, the residuals were highly correlated and not independent as required for regression models. Third, the delay in the response to service level changes would have been missed if only contemporaneous correlations were included in the model. Finally, the biased standard errors from the regression model would have erroneously lead to the conclusion that one the variables (service level-suburban lines) was statistically significant when in reality, it wasn't. These results argue for the wider application of the appropriate statistical methodology when time-series data is used.

TABLE 18

COMPARISON OF SYSTEM MODELS:

FULL-MODEL AND FORECSTING MODEL

MODEL DESCRIPTION	ELASTICITIES			
	SERVICE LEVEL	FARE	GASOLINE PRICE	EMPLOYMENT
Full-model	.51	-.29	.32	.49
Forecasting Model	.38	-.28	.33	.70

The Full-model was estimated using the complete data set of 114 points: January 1973 - June 1982
The Forecasting Model was estimated using 102 data points, with the last twelve data points (July 1981 - June 1982) withheld so that the forecasted values could be compared with the actual values for the period.

TABLE 19

INPUTS USED IN FORECASTING

SYSTEM RIDERSHIP

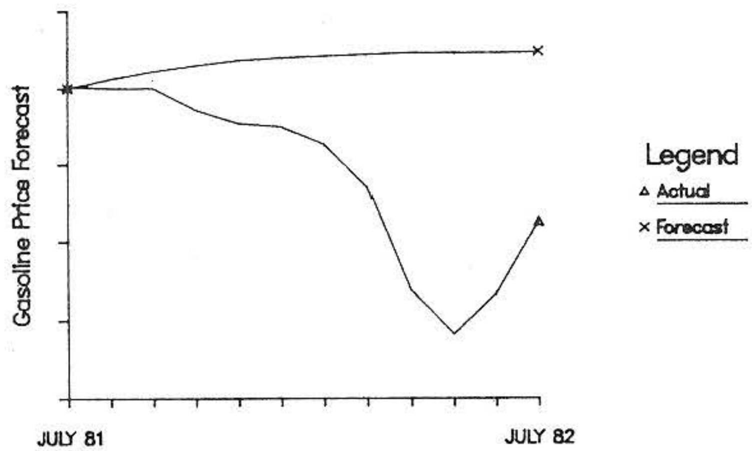
MONTH	Service Level Plat Hrs	Transit Fare Cents	Gasoline Price/ Gallon, Cents	Employment
	Actual	Actual	Forecast	Forecast
July '81	4,788	49.1	135.0	476,100
Aug '81	4,788	49.1	135.6	476,900
Sept '81	4,788	49.1	136.1	486,900
Oct '81	4,788	49.1	136.5	486,500
Nov '81	4,788	49.1	136.8	488,500
Dec '81	4,788	49.1	137.0	489,900
Jan '82	4,788	49.9	137.1	478,400
Feb '82	4,788	49.9	137.2	479,000
Mar '82	4,788	49.9	137.3	483,700
Apr '82	4,788	49.9	137.3	484,200
May '82	4,788	49.9	137.3	484,700
June '82	4,788	49.9	137.4	493,700

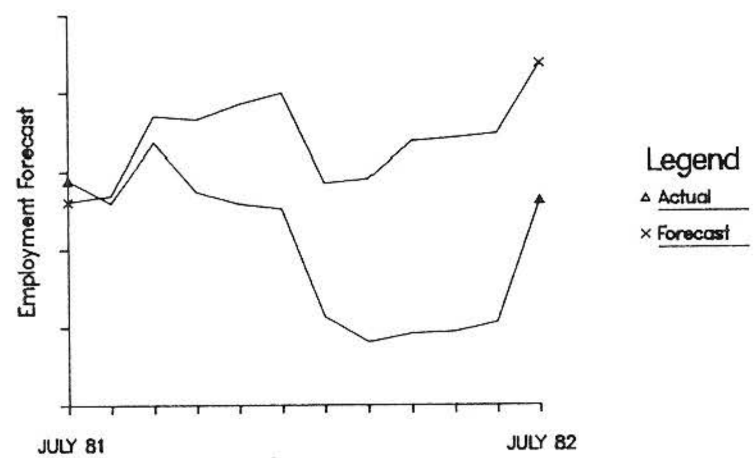
TABLE 20

SUMMARY OF FORECASTS

MONTH	FORECAST OF INPUTS: Gasoline Price Actual	Forecast	% error	Employment Actual	Forecast	% error	FORECAST OF OUTPUT: Ridership Actual	Forecast	% error
July '81	135.1	135.0	-0.1	478,900	476,100	-0.6	125,800	138,300	+2.0
Aug '81	135.0	135.6	+0.4	475,900	476,900	+0.2	121,400	124,700	+2.7
Sept '81	135.0	136.1	+0.8	483,600	486,900	+0.7	132,600	135,200	-2.0
Oct '81	133.6	136.5	+2.2	477,300	486,500	+1.9	141,700	142,600	+0.6
Nov '81	132.7	136.8	+3.1	475,800	488,500	+2.7	141,700	141,600	-0.1
Dec '81	132.5	137.0	+3.4	475,100	489,900	+3.1	132,900	136,200	+2.5
Jan '82	131.3	137.1	+4.4	461,300	478,400	+3.7	146,800	144,800	-0.9
Feb '82	128.5	137.2	+6.8	458,100	479,000	+4.6	142,500	145,800	+2.3
Mar '82	121.9	137.3	+12.6	459,200	483,700	+5.3	141,900	140,000	-1.3
Apr '82	119.1	137.3	+15.3	459,400	484,200	+5.4	143,200	138,400	-3.4
May '82	121.7	137.3	+12.8	460,600	484,700	+5.2	139,400	133,700	-4.1
June '82	126.3	137.4	+8.8	476,300	493,700	+3.7	132,700	127,600	-3.8
Mean Absolute Percent Error of Monthly Forecasts	5.9%			3.1%			2.1%		

FIGURE 21
SUMMARY OF FORECASTS





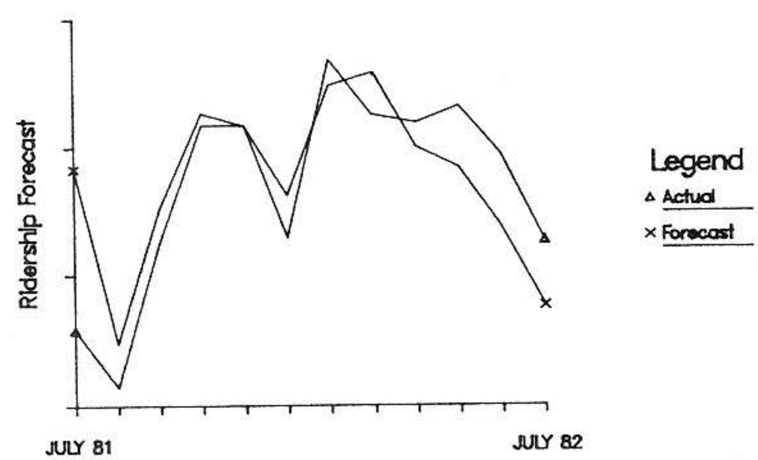


TABLE 21

COMPARISON OF STANDARD REGRESSION AND TRANSFER FUNCTION MODELS

	Comparison	Standard Regression	Transfer Function
1.	Correlated input variables	Yes, the input variables are highly correlated. Multicollinearity is present	No, data is differenced
2.	Autocorrelated errors	Yes, the error structure is highly autocorrelated, violating basic model assumptions.	Yes, but model structure allows for correlated errors
3.	Lag structure for input variables	No, only contemporaneous correlation assumed	Yes, methodology directly investigates the nature of dynamic relationships
4.	Coefficient estimates and standard errors	Estimates are inefficient and the standard errors (and thus the significance tests) are biased.	Estimates are efficient and the standard errors are unbiased

TABLE 22

MULTICOLLINEARITY OF NON-DIFFERENCED DATA

	Correlation Matrix - Input Variables				
	Service Level City Lines	Service Level Suburban Lines	Fare	Gasoline Price	Employment
Service Level City Lines	1.00	.96	.45	.85	.89
Service Level Suburban Lines	.96	1.00	.48	.88	.84
Fare	.45	.48	1.00	.80	.60
Gasoline Price	.85	.88	.80	1.00	.89
Employment	.89	.84	.60	.89	1.00

TABLE 23

COMPARISON OF COEFFICIENT ESTIMATES

STANDARD REGRESSION VS. TRANSFER FUNCTION MODELS

Variable	Coefficient Estimate and Standard Error	
	Regression	Transfer Function
Service level - city lines	.39 ± .21	.28 ± .17
Service level - suburban lines	.31 ± .12	.08 ± .06
Fare	-.30 ± .08	-.28 ± .07
Gas Price	.27 ± .07	.25 ± .11
Employment	.48 ± .09	.57 ± .26

V. ISSUES REQUIRING FURTHER RESEARCH

In every research project, not all issues that require attention can be adequately addressed. This project is no exception. There are several issues in particular that deserve further work.

1. Only nine route level models have been developed. There is sufficient data to develop nearly forty more route level models for the Tri-Met system. When all models have been developed, a statistical comparison of service level and fare elasticities will be possible by route type and geographic sector served. Also, an impact analysis of all service changes, for example, would allow an analysis of elasticity by service change category.
2. It was determined here that route miles is not a useful variable in estimating level of service. But more analysis is needed to determine which data best represents level of service: platform hours, platform miles, or platform miles per route mile. It is likely that this will vary according to the route type and the kinds of service changes that are being modeled.
3. The transfer function models used here assume only a one-way dependence; that is, input variables affect the output variable, but not vice-versa. In actuality, there are feedback loops among several of the variables. For example, as capacity limits are approached as ridership increases, additional service might be required. Thus, ridership level will influence level of service. The general multiple-time series model developed by Tiao and Box (1981) has the ability to handle this two-way dependence.

VI. REFERENCES

Agrawal, D.C. (1978), "The Factors Affecting Mass Transportation Ridership: An Analysis," Transit Journal, (4)3, 55-66.

Alperovich, Gershon, MIchael A. Kemp, and Keith M. Goodman (1977), "An Econometric Model of Bus Transit Demand and Supply," Working Paper 5032-1-4, Washington, D.C.: The Urban Institute.

Ahmed, Mohamed S. and Allen B. Cook (1979), "An Analysis of Freeway Traffic Time Series Data using Box and Jenkins Techniques," Transportation Research Record 722, 1-8.

Barton Aschman Associates and R.H. Pratt and Co. (1981), Traveler Response to Transportation System Changes, Washington, D.C.: US Department of Transportation.

Bates, John (1981), "A Study of Demand For Transit Use," Ph.D. Dissertation, Atlanta, Georgia: Georgia State University, College of Business Administration.

Box, G.E.P. and D.R. Cox (1964), "An Analysis of Transformations," Journal of the Royal Statistical Society, B (26), 211-243.

Box, George E.P and G.M. Jenkins (1976), Time Series Analysis: Forecasting and Control, Revised Edition, San Francisco: Holden-Day.

Box, G.E.P. and G.C. Tiao (1975), "Intervention Analysis with Application to Economic and Environmental Problems," Journal of the American Statistical Association, (70), 70-79.

Charles River Associates (1972), Measurement of the Effects of Transportation Changes, Washington, D.C.: Urban Mass Transportation Administration.

Cherwony, Walter and Lewis Polin (1977), "Forecasting Patronage on New Transit Routes," Traffic Quarterly, 31, 287-295.

Curtin, John (1968), "Effects of Fares on Transit Riding," Highway Research Board, 213, 8-19.

Der, T.Y. (1977), "Some Investigations of Forecasting Freeway Occupancies," Illinois Institute of Technology, Chicago, M.S. Thesis.

Ecosometrics, Inc. (1981), "Patronage Impacts of Changes in Transit Fares and Services," Washington, D.C.: US Department of Transportation.

Elder, M. (1977), "Demand Predictors for Computerized Freeway Control Systems," Proceedings, 7th International Symposium on Transportation and Traffic Theory, (T. Sasaki and T. Yamaoka, eds.), Institute of Systems Science and Research, Kyoto, 341-370.

Gaudry, Marc (1975), "An Aggregate Time Series Analysis of Urban Transit Demand: The Montreal Case," Transportation Research, 9, 249-258.

Gaudry, Marc (1978a), "Seemingly Unrelated Static and Dynamic Urban Travel Demand Models," Transportation Research, (12), 195-212.

Gaudry, Marc and Michael Wills (1978b), "Estimating the Functional Forms of Travel Demand Models," Transportation Research, (12), 257-290.

Harmatuck, Donald (1975), "The Effects of a Service Interruption on Bus Ridership Levels in a Middle-sized Community," Transportation Research, 9, 43-54.

Hanssens, D.M. and Lon-Mu Liu (1983), "Lag Specification in Rational Distributed Lag Structural Models," Journal of Business and Economic Statistics, 4(1), 316-325.

Haugh, Larry and G.E.P Box (1977), "Identification of Dynamic Regression Models Connecting Two Time Series," Journal of the American Statistical Association, 72, 121-130.

Holmesland, Kjell O. (1979), "Use of the Box-Jenkins Time Series for Forecasting Commuter Traffic Volumes," Research Report No. 79-9, Seattle, Urban Transportation Program, Department of Civil Engineering and Urban Planning, University of Washington.

Kemp, Michael (1974), "Transit Improvements in Atlanta--The Effects of Fare and Service Changes," Paper 1212-2-1, Washington, D.C.: The Urban Institute.

Kemp, Michael (1981a), "Bus Service In San Diego: A Study of Patronage Growth in the mid-1970's," Working Paper 1470-1, Washington, D.C.: The Urban Institute.

Kemp, Michael (1981b), "A Simultaneous Equations Analysis of Route Demand and Supply, and its application to the San Diego Bus System," Working Paper 1470-2, Washington, D.C.: The Urban Institute.

Liu, Lon-Mu and D.M. Hanssens (1982), "Identification of Multiple-Input Transfer Function Models," Communications in Statistics A, 11(3), 297-314.

McLeod, G., J.T. Everest, and N.J. Paulley (1980), "An Analysis of Rail and Air Passenger Flows Between London and Glasgow Using Box-Jenkins Methods," Supplementary Report 524, Crowthorne (UK):
UK Transport and Road Research Laboratory.

Moody, James P. (1975), "Supply and Demand for Urban Bus Travel: Theoretical Revision and Empirical Estimation," unpublished mimeograph, Milwaukee (WI): University of Wisconsin-Milwaukee Department of Economics.

Multisystems, Inc. (1982), Route Level Demand Models: A Review, Washington, D.C.: Urban Mass Transportation Administration.

Nihan, Nancy L. and Kjell O. Holmesland (1980), "Use of Box and Jenkins Time Series Techniques in Traffic Forecasting," Transportation, 9, 125-143.

Polhemus, Neil W. (1976), "Time Series Analysis of Local Fluctuations in Traffic Parameters," Transportation Research, 10, 311-317.

Polhemus, Neil W. (1979), "A Multiple Tranfer Function Model For Air Traffic Control Systems," Transportation Research, 13, 229-236.

Rainville, Walter S. (1948), "Transit Riding, Revenues, and Fare Structures--Basic Approach and Computations by Statistical Laboratory, Division of Industrial Cooperation of Massachusetts Institute of Technology," Washington, D.C.: American Transit Association.

SAS Institute, Inc (1982) SAS/ETS Users Guide-1982 Edition, Cary, NC: Sas Institute.

SCA (1983), The SCA System for Univariate-Multivariate Time Series and General Statistical Analysis, DeKalb, IL: Scientific Computing Associates.

Schmenner, Roger W. (1976), "The Demand for Urban Bus Transit," Journal of Transport Economics and Policy, 10, 68-86.

Skinner, David, Robert Waksman, and George H. Wang (1983), "Empirical Modeling and Forecasting of Monthly Transit Revenue for Financial Planning: A Case Study of SCRTD in Los Angeles," Paper presented at the 1983 annual meeting of the Transportation Research Board.

Tiao, G.C. and G.E.P. Box (1981), "Modeling Multiple Time Series with Applications," Journal of the American Statistical Association, 76(376), 802-816.

Umashanker, Sushila and Johannes Ledolter (1983), Forecasting with Diagonal Multiple Time-Series Models: An Extension of Univariate Models," Journal of Marketing Research, XX(Feb 1983), 58-63.

Uhlborg, Cy (1982), "A Study of Transit Ridership in Seattle and Development of a Time Series Regression Model," Transportation Research Record, 854, 12-16.

Wang, George H. (1981), "An Intervention Analysis of Interrupted Urban Transit Time Series Data : Two Case Studies," 1981 Proceedings of the Business and Economic Statistics Section, American Statistical Association, 424-429.

Wang, George H., Walter Maling, and David Skinner (1982), "Modeling Monthly Ridership For Seven U.S. Transit Authorities," United States Department of Transportation, Transportation System Center: Cambridge, MA.

VII. APPENDIX - MODEL EQUATIONS

The models that have been described earlier in this report are presented in the equation form in the Appendix. Tables 24, 25, 26, and 27 show the standard time-series form. Tables 28 and 29 give the route level models in their expanded form thus showing the autoregressive nature of the models.

The notation used in the Appendix is defined below:

LR_t	= log of transit ridership at time t
$LHRC_t$	= log of city platform hours at time t
$LHRS_t$	= log of suburban platform hours at time t
LHR_t	= log of total platform hours at time t
$LM1_t$	= log of total platform miles at time t
$LPLTRTC_t$	= log of city platform miles per route mile at time t
$LPLTRTS_t$	= log of suburban platform miles per route mile at time t
LF_t	= log of average transit fare at time t
LAF_T	= log of average transit fare at time t
LCF_t	= log of basic cash transit fare at time t
LG_t	= log of gasoline price per gallon at time t
LE_t	= log of total tri-county employment at time t
$LMLE_t$	= log of Multnomah county employment at time t
$LWLE_t$	= log of Washington County employment at time
$LCLE_t$	= log of Clackamas County employment at time t
B	= backshift operator, which $BLHR_t = LHR_{t-1}$
a_t	= error term at time t
$(\pm.xx)$	= standard error of the coefficient estimate
ζ_t	= intervention variable, usually representing a specific service level change

TABLE 24

SYSTEM MODELS

$$LR_t = \underset{(\pm .18)}{.31LHRC_{t-8}} + \underset{(\pm .07)}{.08LHRS_{t-1}} - \underset{(\pm .07)}{.27LF_t} + \underset{(\pm .12)}{.26LG_t} + \underset{(\pm .28)}{.56LE_t} + \frac{(1 - \underset{(\pm .11)}{.30B^{12}} - \underset{(\pm .12)}{.31B^{24}})a_t}{(1 - B)(1 - B^{12})}$$

$$LR_t = \underset{(\pm .16)}{.29LPLTRC_{t-10}} + \underset{(\pm .09)}{.22LPLTRTS_{t-1}} - \underset{(\pm .07)}{.29LF_t} + \underset{(\pm .12)}{.32LG_t} + \underset{(\pm .28)}{.49LE_t} + \frac{(1 - \underset{(\pm .12)}{.29B^{12}} - \underset{(\pm .12)}{.26B^{24}})a_t}{(1 - B)(1 - B^{12})}$$

TABLE 25
SECTOR MODELS

City Radial Sector Model

$$R_t = \underset{(\pm.28)}{.71LHR_{t-2}} - \underset{(\pm.10)}{.13LAF_t} + \underset{(\pm.07)}{.14LG_t} + \underset{(\pm.27)}{.43LMCE_t} + \frac{(1 - \underset{(\pm.13)}{.76B})(1 - \underset{(\pm.16)}{.56B^4})a_t}{(1-B)(1-B^4)}$$

City Crosstown Sector Model

$$LR_t = \frac{.42LMI_t}{1-.42B} - \underset{(\pm.27)}{.42LCF_t} + \underset{(\pm.22)}{.39LG_t} + \frac{(1 - \underset{(\pm.17)}{.40B})(1 - \underset{(\pm.16)}{.58B^4})a_t}{(1 - B)(1 - B^4)}$$

Urban/Suburban Sector Model

$$LR_t = \underset{(\pm.14)}{.55LMI_{t-2}} - \underset{(\pm.07)}{.15LCF_t} + \underset{(\pm.08)}{.18LG_t} + \underset{(\pm.30)}{.65LMCE_t} + \frac{(1 - \underset{(\pm.14)}{.75B})(1 - \underset{(\pm.15)}{.67B^4})a_t}{(1-B)(1-B^4)}$$

Westside Suburban Sector Model

$$LR_t = \underset{(\pm.07)}{.80LMI_t} - \underset{(\pm.08)}{.32LAF_t} + \underset{(\pm.08)}{.31LG_t} + \underset{(\pm.12)}{.47LWCE_t} + \frac{(1 + \underset{(\pm.15)}{.49B^3})a_t}{(1-B^4)}$$

Southwest Suburban Sector Model

$$LR_t = \underset{(\pm.19)}{.49LPLTRT_t} - \underset{(\pm.09)}{.22LCF_{t-1}} + \underset{(\pm.17)}{.28LG_t} + \underset{(\pm.29)}{.67LE_t} + \frac{(1 - .99B^4)a_t}{(1-B)(1-B^4)}$$

Southeast Suburban Sector Model

$$LR_t = \underset{(\pm.11)}{.33LHR_t} + \underset{(\pm.11)}{.50LHR_{t-2}} - \underset{(\pm.13)}{.16LAF_t} + \underset{(\pm.16)}{.27LG_t} + \underset{(\pm.30)}{.69LCCE_{t-1}} + (1 - \underset{(\pm.19)}{.30B})(1 - \underset{(\pm.18)}{.49B^4})a_t$$

TABLE 26

CITY RADIAL LINE SECTOR
ROUTE MODELS

Route 2

$$(1 - B^4)LR_t = \underset{(\pm.43)}{(.91} + \underset{(\pm.43)}{.90B^2})(1 - B)LHR_t - \underset{(\pm.20)}{.39}(1 - B)LAF_t$$

$$+ \underset{(\pm.15)}{.72}(1 - B)LG_t + \underset{(\pm.42)}{1.14B^2}(1 - B)LMCEMP_t$$

$$+ (1 - \underset{(\pm.14)}{.64B^4})a_t$$

Route 3

$$(1 - B^4)LR_t = (\underset{(\pm.25)}{.56} + \underset{(\pm.16)}{.51B^2} + \underset{(\pm.25)}{.66B^3})(1 - B)LHR_t + (\underset{(\pm.21)}{-.46} - \underset{(\pm.19)}{.44B})(1 - B)LAF_t$$

$$+ \frac{\underset{(\pm.16)}{.69}(1 - B)LG_t}{(1 - \underset{(\pm.16)}{.55B})} + (1 + \underset{(\pm.13)}{.81B^3})a_t$$

Route 6

$$(1 - B^4)LR_t = \underset{(\pm.06)}{.23}(1 - B)\zeta_{17} - \underset{(\pm.20)}{.80B}(1 - B)LAF_t + \underset{(\pm.27)}{.62}(1 - B)LG_t$$

$$+ \underset{(\pm.64)}{.95B}(1 - B)E_t + (1 + \underset{(\pm.13)}{.47B})a_t$$

Route 8

$$(1 - B^4)LR_t = \underset{(\pm.04)}{.25B^3}(1 - B)\zeta_{13} - \underset{(\pm.23)}{.35B^2}(1 - B)LAF_t + (\underset{(\pm.29)}{.59} + \underset{(\pm.28)}{.64B})(1 - B)LG_t$$

$$+ (1 + \underset{(\pm.15)}{.43B} + \underset{(\pm.13)}{.49B^5} - \underset{(\pm.14)}{.27B^8})a_t$$

TABLE 27

CITY CROSSTOWN SECTOR ROUTE MODELS

Route 71

$$(1-B^4)LR_t = \underset{(\pm.27)}{.72}(1-B)\xi_{31} + \underset{(\pm\ .63)}{3.24}(1-B)LG_t + (1-.51B^4)a_t$$

Route 72

$$LR_t = \underset{(\pm.15)}{.64}\xi_{19} + \underset{(\pm.23)}{.30}LHR_t + \underset{(\pm.30)}{.68}LG_{t-3} + \frac{(1-\underset{(\pm.15)}{.42}B)a_t}{(1-B)}$$

Route 75

$$(1-B^4)LR_t = 1.72(1-B)LG_{t-3} + (1+.40B)(1-.43B^4)a_t$$

Route 77

$$LR_t = \underset{(\pm.11)}{.35}\xi_{26} + \underset{(\pm.18)}{.24}LG_{t-2} + \frac{(1-\underset{(\pm.16)}{.56}B)(1-\underset{(\pm.18)}{.42}B^2)a_t}{(1-B)}$$

TABLE 28

CITY RADIAL LINE SECTOR ROUTE MODELS
(EXPANDED FORM)

Route 2

$$LR_t = LR_{t-4} + .91LHR_t - .91LHR_{t-1} + .90LHR_{t-2} - .90LHR_{t-3} - .39\,LAF_t + .39LAF_{t-1} + .72LG_t - .72LG_{t-1} + 1.14LMCEMP_{t-2} - 1.14LMCEMP_{t-3} + a_t - .64a_{t-4}$$

Route 3

$$LR_t = LR_{t-4} + .56LHR_t - .56LHR_{t-1} + .51LHR_{t-2} + .15LHR_{t-3} - .66LHR_{t-4} - .46LAF_t + .02LAF_{t-1} + .44LAF_{t-2} + .69LG_t + .31LG_{t-1} + .10LG_{t-2} + a_t + .81a_{t-3}$$

Route 6

$$LR_t = LR_{t-4} + .23\zeta_{17} - .23\zeta_{16} - .80LAF_t + .80LAF_{t-1} \quad .62LG_t - .62LG_{t-1} + .95E_{t-1} - .95E_{t-2} + a_t + .47a_{t-1}$$

Route 8

$$LR_t = LR_{t-4} + .25\zeta_{10} - .25\zeta_{9} - .35LAF_{t-2} + .35LAF_{t-3} + .59LG_t + .05LG_{t-1} - .64LG_{t-2} + a_t + .43a_{t-1} + .49a_{t-5} - .27a_{t-8}$$

TABLE 29

CITY CROSSTOWN SECTOR ROUTE MODELS
(EXPANDED FORM)

Route 71

$$R_t = R_{t-4} + .72(\xi_{30}-\xi_{29}) + 3.24(G_{t-2}-G_{t-3}) + a_t - .51a_{t-4}$$

Route 72

$$R_t = R_{t-1}+.64(\xi_{19}-\xi_{18})+.30(HR_t-HR_{t-1})+.68(G_{t-3}-G_{t-4})+a_t-.42a_{t-1}$$

Route 73

$$R_t = R_{t-1}+.60(G_t-G_{t-1})+a_t-.86a_{t-1}+.32a_{t-4}-.28a_{t-5}+.26a_{t-8}-.22_{t-9}$$

Route 75

$$R_t = R_{t-4}+1.72(G_{t-3}-G_{t-4})+a_t+.40a_{t-1}-.43a_{t-4}-.17a_{t-5}$$

Route 77

$$R_t = R_{t-1}+.35(\xi_{26}-\xi_{25})+.24(G_{t-2}-G_{t-3})+a_t-.56a_{t-1}-.42a_{t-2}+.24a_{t-3}$$