Development and Application of Time-Series Transit Ridership Models for Portland, Oregon

MICHAEL KYTE, JAMES STONER, and JONATHAN CRYER

ABSTRACT

Described in this paper are the development and application of a methodology to identify and analyze the factors that influence changes in public transit ridership. The data used in the model development and testing are from Portland, Oregon and cover the period 1971 through 1982. Models were developed at the system, sector, and route levels, and were 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 here 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 here 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.

Analysis of past variation in transit ridership and forecasting of future ridership 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 exogenous factor such as a change in gasoline supply or price.

There are usually several processes that are occurring simultaneously, each affecting ridership in some way. 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 factors that are also exhibiting variation. Similarly, to satisfactorily forecast future transit ridership, a clear understanding of these factors is necessary.

Two basic classes of models (cross-sectional and time-series) have been developed by transportation analysts. Each class seeks to define the nature of travel demand and the factors that influence it. Cross-sectional models are developed using data collected at one point in time. Often, intensive travel surveys are undertaken and detailed characteristics of the transportation system are measured. The level of detail of the data allows the development of models that are able to relate microlevel characteristics of the system. For example, characteristics of individual trip-making patterns such as traveler demographics and travel costs and time by competing modes can easily be handled with cross-sectional models. However, using these models to assist in evaluating the impacts of a change over time involves some degree of risk. It is not clear that structural relationships estimated at one point will remain stable over time. In addition, data are expensive and time-consuming to collect and analyze. Time-series models are based on data collected over a period of time and thus allow for direct measurement of the nature of these dynamics. The trade-off is that the level of detail for time-series data is usually not nearly as great as for cross-sectional data. This reduces the precision with which time-series models can approximate true time-dependent structural relationships in the data. However, time-series data are typically collected regularly by the transit operator and are readily available to the analyst. Because the nature of these relationships may itself change over time, it seems clear that models based on time-series data are more likely to capture these dynamics than those based on 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 (1,2), Kemp (3,4), and Wang (5,6). The data in this paper are built on the work of these researchers, and extend 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 com-

puter software makes use of this approach practical and available to most analysts.

METHODOLOGY

The proposed 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 (Figure 1).

Model Development

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 (1,2), Kemp (3,4), and Wang (5,6).

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 can be written as

$$R_t = F(SL_t, TC_t, MS_t, S_t, I_t) + N_t \quad (1)$$

where

R_t = transit ridership,
SL_t = level of transit service,
TC_t = travel costs by automobile 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, and so forth, and
N_t = the noise model or error structure.

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 on whether the change in the input is large or small? It is assumed here that changes in transit ridership resulting from changes in service 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 system (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. 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 2.

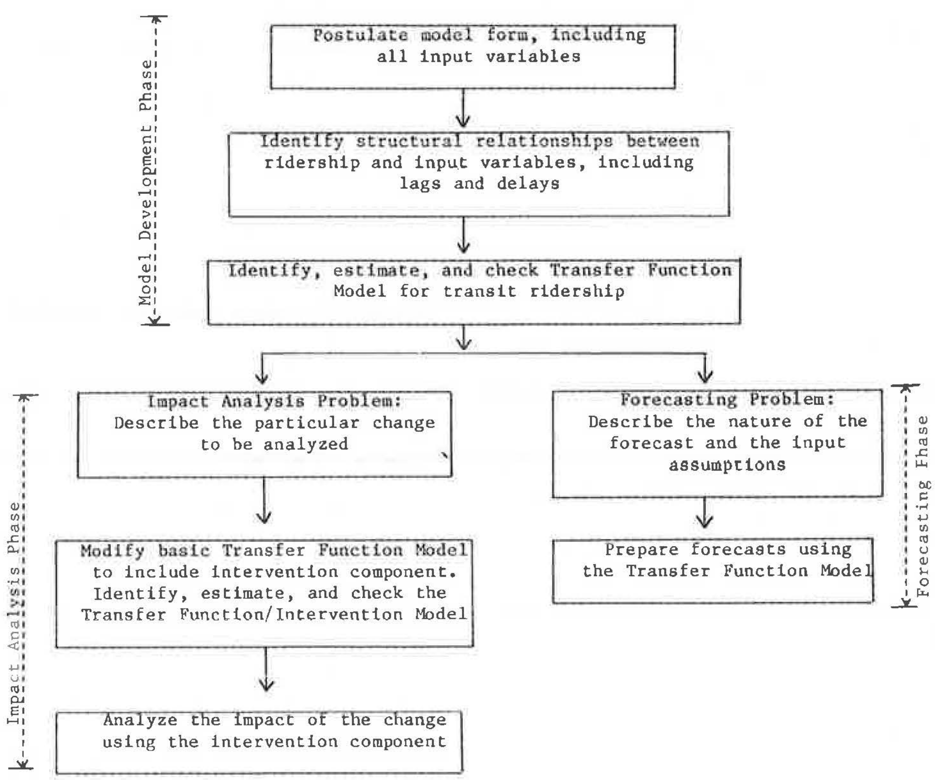
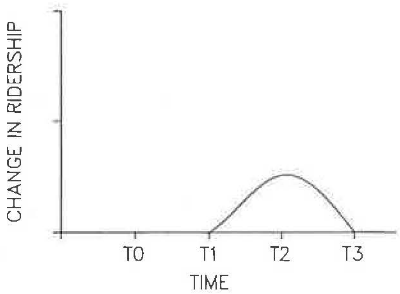


FIGURE 1 Methodology for analysis and forecasting of public transit ridership.



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

FIGURE 2 Lagged response.

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 automobile 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 (1) and Kemp (3,4), who each attempted to construct waiting time and in-vehicle time time-series for 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 nonservice 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.

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 automobile 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 (3,4) for a discussion of this approach]. However, it was found here that nondeflated prices are more directly correlated with transit ridership. The size of the travel market (market size) is described by employment.

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 nonsummer months (particularly during the winter) often make transit more attractive than walking or using an automobile. 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 previously 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. The models themselves are known as autoregressive-integrated moving average (ARIMA) models. This approach is based on 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 that may need to be repeated until a satisfactory model results. These steps include identification of a tentative model based on 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 on 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. For a discussion of the ACF and PACF, the reader is referred to Box and Jenkins (7). The class of ARIMA models of particular interest here is the transfer function model, which can be written as

$$Y_t = \sum_i [\omega_i(B)/\delta_i(B)] X_{it} [\theta(B)a_t/\phi(B)] \quad (2)$$

where 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 (month) after a fare change. Then, the general model (Equation 2) can be written as

$$R_t = (\omega_0/1-\delta B)SL_t + \omega_1 F_{t-1} + [\theta(B)a_t/\phi(B)] \quad (3)$$

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. Differentiate between 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 so that

$$Y_t = \theta(B)a_t/\phi(B) \quad (4)$$

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 so that

$$Y_t = v_0 X_{1t} + v_1 X_{1t-1} + v_2 X_{1t-2} + \ldots + [\partial(B)\ a_t/P(B)] = v(B)x_{1t} + [\theta(B)a_t/\phi(B)] \quad (5)$$

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 $v'(B)$ so that

$$Y_t = \nu'(B)x_{1t} + [\theta(B)a_t/\phi(B)] \quad (6)$$

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

$$Y_t = \nu'(B)x_{1t} + \nu'_2(B)x_{2t} + [\theta(B)a_t/\phi(B)] \quad (7)$$

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 of

$$Y_t = \sum_i [\omega_i(B)/\delta_i(B)]\ X_{it} + [\theta(B)a_t/\phi(B)] \quad (8)$$

where

Y_t = the output series,
$\omega_i(B)/\delta_i$ = the transfer function polynomial ratio,
X_{it} = the input series,
$\theta(B)/\phi(B)a_t$ = the ARIMA noise model, and
B = 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. Although 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 (8).

Impact Analysis

The transfer function model developed in the first phase (model development) provides an indication of the average response of transit ridership to changes in service level or transit fare. The model is estimated based on all of the service level or transit fare changes that occur during the period for which the data are 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 (9), is based on 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 gasoline shortage, is represented by a binary variable ζ_{jt}, which assumes a value of 0 before or after the event and a value of 1 during the time that the event or intervention is taking place.

The basic form of the transfer function model with intervention is

$$Y_t = \sum_i [w_i(B)/\delta_i(B)]\ X_{it} + \sum_j [w_j(B)/\delta_j(B)]\ \zeta_{jt} + [\theta(B)/\phi(B)]a_t \quad (9)$$

The variables of Equation 9 are the same as previously defined for Equation 8, 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 \$0.05-fare increase is to be analyzed, this increase should be subtracted out of the fare data.

Step 4. Re-estimate the model with the intervention variable included, as in Equation 9. 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.

Forecasting

The transfer function model developed in the first phase (model development) can also be used to forecast future levels of transit ridership. But because the model depends on several inputs, these variables must also be assumed or forecast. 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 exogenous and these must be forecast 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.

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 exogenous input variables, such as employment and gasoline price.

Step 4. Using either the forecast 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 (10) and SCA (11) for further information.

CASE STUDY: PORTLAND, OREGON

The Portland, Oregon metropolitan area includes 1.2 million people and covers over 900 mi^2. 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 9-percent annual growth rate. The 3-year period 1973-1976 saw nearly a 20-percent annual increase. Platform hours and miles increased at an annual rate of nearly 7 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 10-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, automobile travel costs increased significantly during this period, while transit travel costs declined. Gasoline price increased at a 15.6-percent annual rate during the 10-year period, with the largest increase occurring 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 shown in Figures 3 through 7.

Model Development Phase

Data for Portland, Oregon covering 1971 through 1982 were used to develop a total of 16 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. The three different data sets used here and their interrelationships are given in Table 1.

Four input variables were used for each of the models: transit service level, transit fare, gasoline price as a surrogate for automobile operating costs, and employment as a measure of the travel market size. 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 1 to 10 months for the system model and 0 to 3 quarters for the sector and route models. Fare delays ranged up to 2 quarters. A summary of the elasticities and lags are given in Table 2.

Examination of Table 2 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 shown in Figures 8 and 9. Note that the elasticities estimated for the six sector models tend to vary around the system mean for each variable.

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 given in Table 3. Seven of the eleven changes were found to have had a significant impact on ridership.

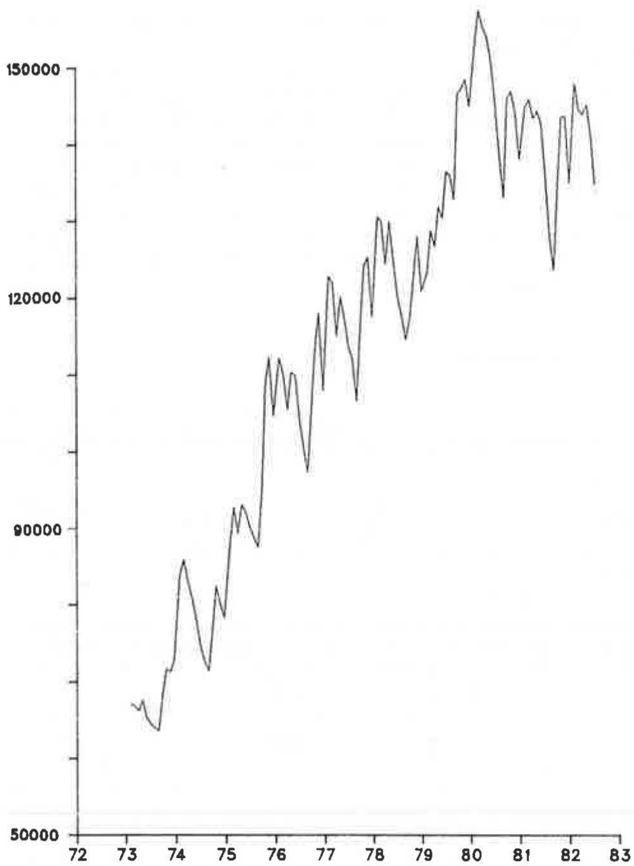


FIGURE 3 Transit ridership, system level, Portland data.

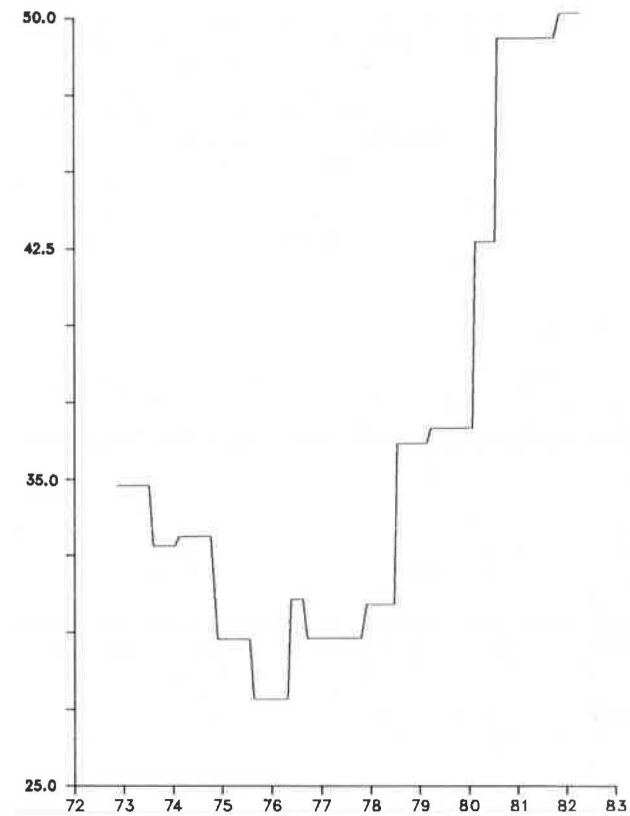


FIGURE 5 Average transit fare, Portland data.

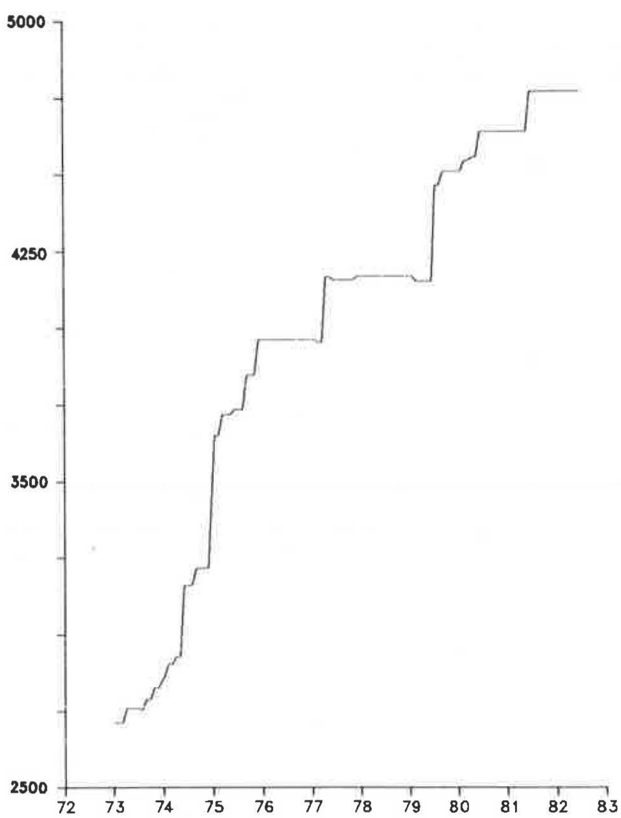


FIGURE 4 Platform hours, system level, Portland data.

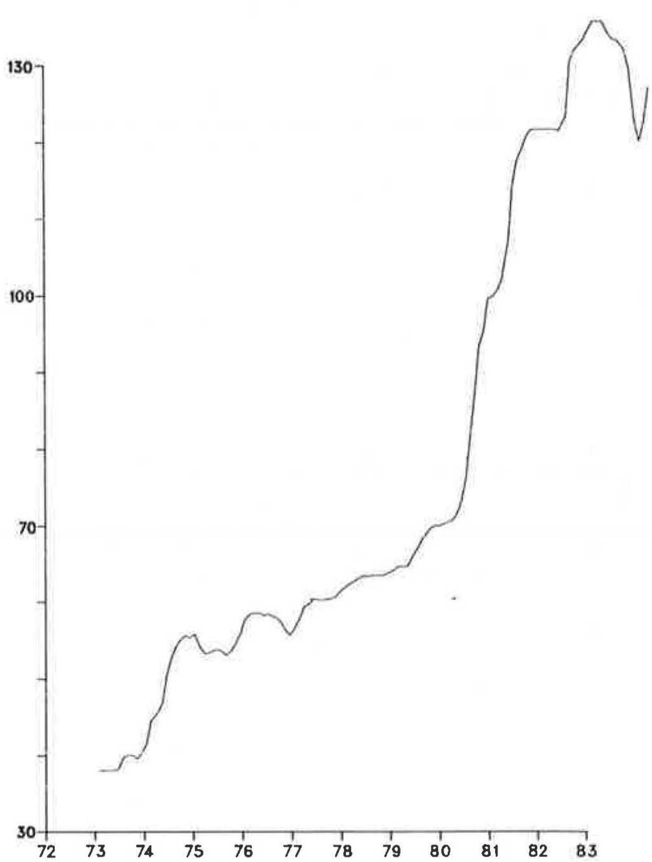


FIGURE 6 Gasoline price, Portland data.

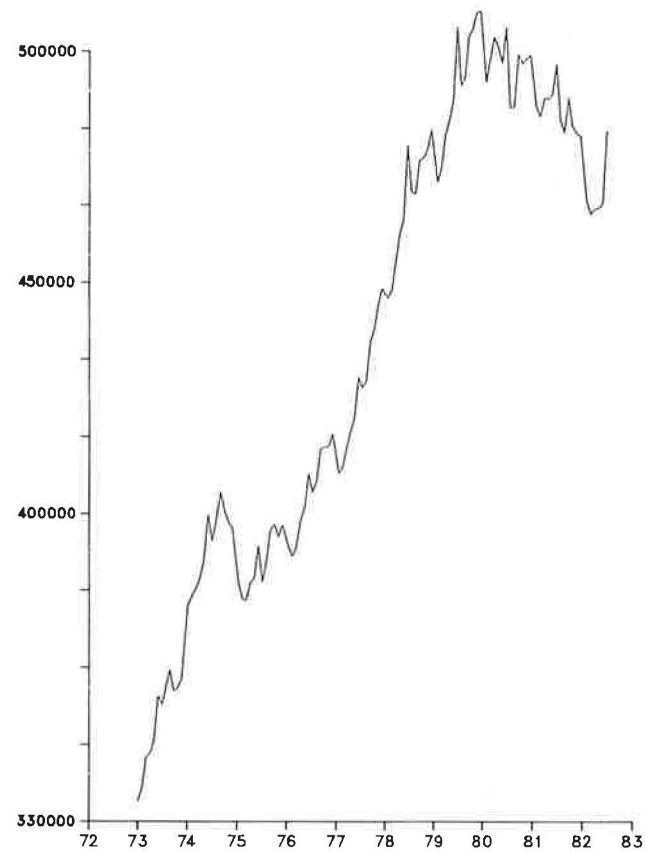


FIGURE 7 Tri-County employment, Portland data.

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 on future variation in gasoline price and employment as well, these variables must also be forecast or assumptions must be made about their future values.

Figure 10 shows the results of a forecast of system ridership for 12 periods (months) ahead. It was assumed that service level and fare were set by policy and that gas price and employment had to be forecast using time-series models. These results, with a mean absolute percent error of 2.1 percent, show the high quality of forecast that can be achieved by using this approach.

TABLE 1 Summary of Portland Data Base

Variable	Time-Series
System level data	
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	
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

COMPARISON WITH STANDARD REGRESSION MODELS

It has been traditional to use multiple regression models when developing models that relate transit ridership to explanatory variables. Using time-series data with regression models, however, invariably leads to a variety of statistical problems. Table 4 contains data that highlight the following major areas in which problems are likely to arise by contrasting standard regression with transfer function models: multicollinearity, autocorrelated errors, lag structures, and coefficient 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.

In using the nondifferenced 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

TABLE 2 Summary of Models

	Model Description		Service Level		Fare		Gas Price		Employment	
Data Aggregation	Data Period	Model Description	Elasticity	Lag	Elasticity	Lag	Elasticity	Lag	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 suburban 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	-	

Note: 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 2 quarters after the input variable was changed.

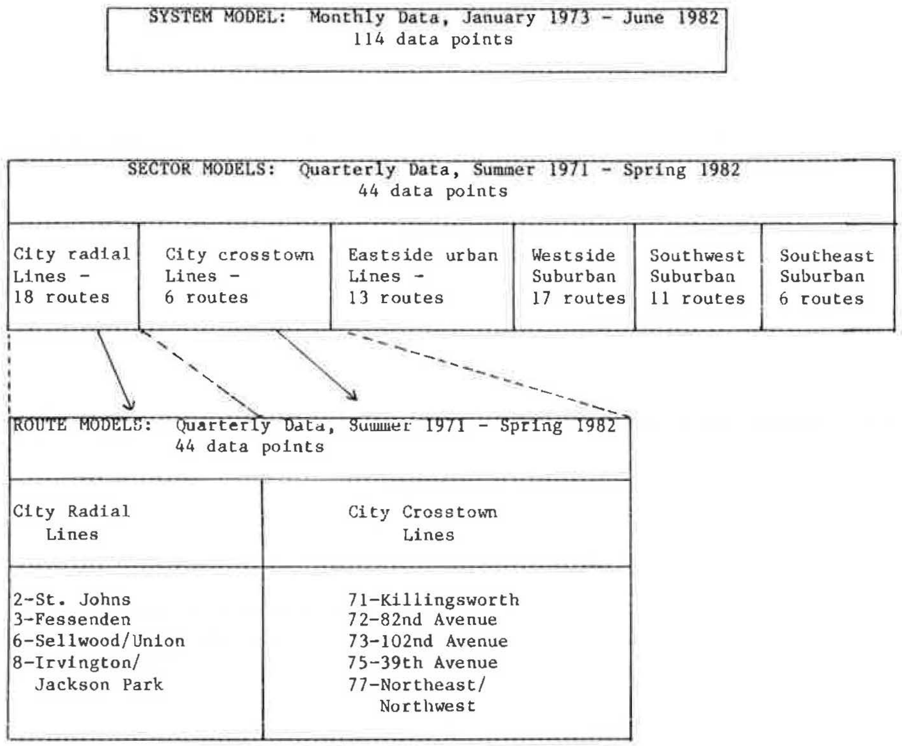


FIGURE 8 Summary of models developed and their interrelationship.

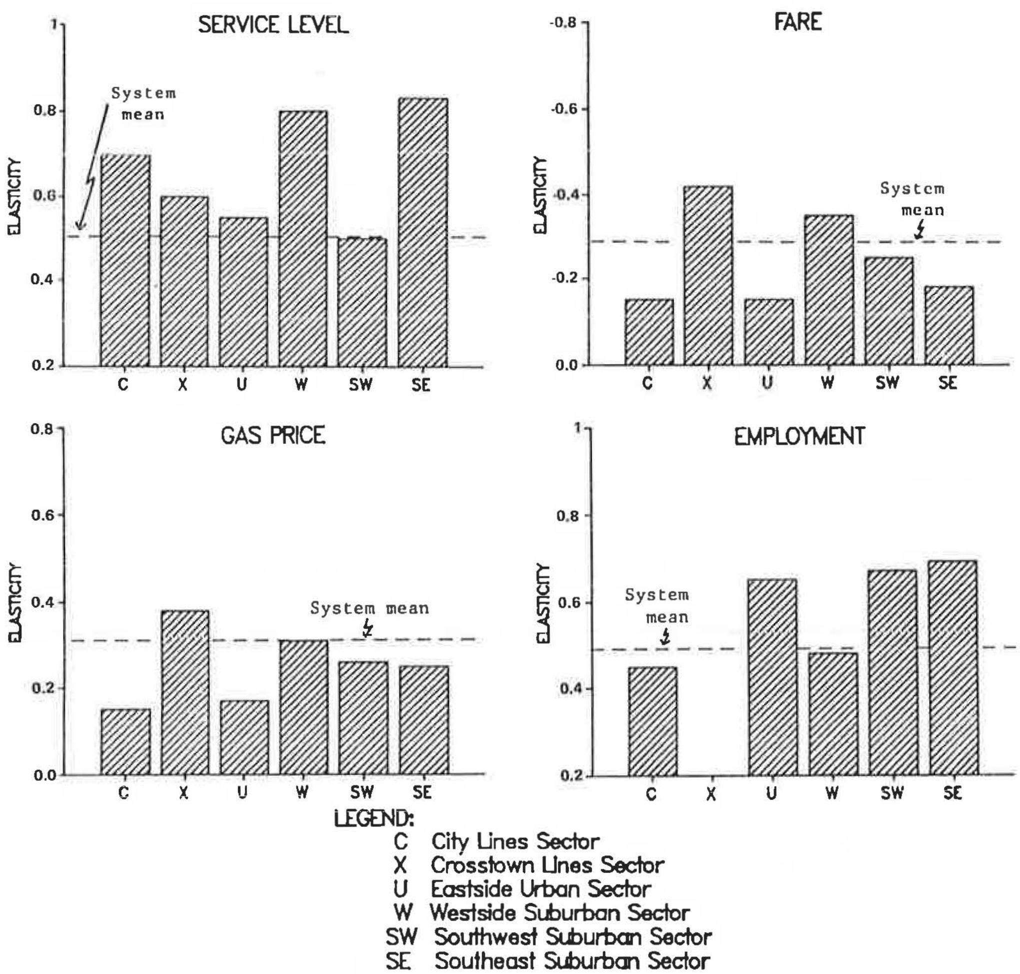


FIGURE 9 Consistency of model coefficients between system and sector models.

TABLE 3 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

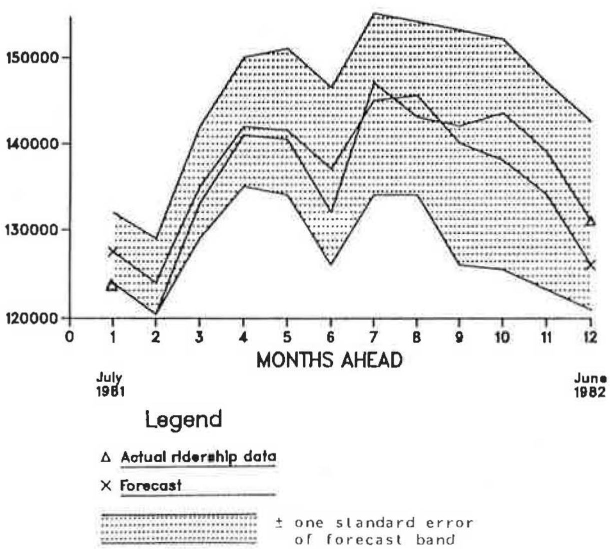


FIGURE 10 Comparison of forecasts, system model.

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 led to the conclusion that one of the variables (service level-suburban lines) was statistically significant when in reality it was not (see Table 6). These results argue for the wider application of the appropriate statistical methodology when time-series data are used.

CONCLUSIONS

This paper has presented the initial results from the development and application of Box-Jenkins transfer function and intervention models to time-series transit ridership and operations data for Portland, Oregon. The results indicate that this methodology is appropriate for evaluation and forecasting of transit ridership changes. Evidence is presented for the lag structure of the market response to the various factors that influence transit ridership. Service level changes, for example, may require up to 2 quarters or 10 months for their effects to be realized, while fare changes have lag effects of up to 1 quarter or 1 month. Response to gasoline prices and employment level changes are more rapid, though lag effects have been found at the route level for up to 3 quarters for gasoline price changes.

This work has also shown the consistency of results that may be obtained between system and route models when using different data bases. In addition, the effectiveness of intervention variables to model a specific change or event was demonstrated. Finally, some evidence was found on the variation of the structural relationships in the model over time.

While requiring a somewhat longer learning period than would more traditional multiple regression analysis, time-series ARIMA models offer a substantial advantage to the transportation analyst. With the recent availability of new computer software designed specifically to handle time-series problems, their use in transportation analysis will hopefully increase.

There are several areas in which further research is needed; some of this work is now underway by the authors of this paper, including:

1. Development of route level models for all 37 route pairs operated by the Portland transit system. This will enable a thorough statistical analysis of elasticity measures of service level and transit fare, and a better categorization of impact of these changes.

2. Development of multiple time-series models that will enable a study of two-way causality. Multiple time-series models are much more difficult to develop than transfer function and intervention models, but the results of the work provide more

TABLE 4 Comparison of Standard Regression and Transfer Function Models

Comparison	Standard Regression	Transfer Function
Correlated input variables	Yes, the input variables are highly correlated; multicollinearity is present	No, data are differenced
Automobile-correlated errors	Yes, the error structure is highly autocorrelated, violating basic model assumptions	Yes, but model structure allows for correlated errors
Lag structure for input variables	No, only contemporaneous correlation assumed	Yes, methodology directly investigates the nature of dynamic relationships
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 5 Correlation Matrix Showing Multicollinearity of Nondifferenced Data

Input Variables	Service Level City Lines	Service Level Suburban Lines	Fare	Gasoline Price	Employment
Service level					
City lines	1.00	.96	.45	.85	.89
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 versus Transfer Function Models

Input Variables	Coefficient Estimate and Standard Error: Regression	Transfer Function
Service level		
City lines	.39 ± .21	.28 ± .17
Suburban lines	.31 ± .12	.08 ± .06
Fare	-.30 ± .08	-.28 ± .07
Gas price	.27 ± .07	.25 ± .11
Employment	.48 ± .09	.57 ± .26

useful insights into the structure and dynamics of the factors that influence change in transit ridership.

ACKNOWLEDGMENTS

This research has been supported by the Urban Mass Transportation Administration, the Tri-County Metropolitan Transportation District of Oregon, and The University of Iowa.

REFERENCES

1. M. Gaudry. An Aggregate Time Series Analysis of Urban Transit Demand: The Montreal Case. Transportation Research, Vol. 9, 1975, pp. 249-258.
2. M. Gaudry, Seemingly Unrelated Static and Dynamic Urban Travel Demand Models. Transportation Research, Vol. 12, 1978, pp. 195-212.
3. M. Kemp. Bus Service In San Diego: A Study of Patronage Growth in the Mid-1970's. Working Paper 1470-1, The Urban Institute, Washington, D.C., 1981.
4. M. Kemp. A Simultaneous Equations Analysis of Route Demand and Supply, and Its Application to the San Diego Bus System. Working Paper 1470-2, The Urban Institute, Washington, D.C., 1981.
5. G.H. Wang. An Intervention Analysis of Interrupted Urban Transit Time Series Data: Two Case Studies. Proc., Business and Economic Statistics Section, American Statistical Association, Washington, D.C., 1981, pp. 424-429.
6. G.H. Wang, W. Maling, and D. Skinner. Modeling Monthly Ridership For Seven U.S. Transit Authorities. Transportation System Center, U.S. Department of Transportation, Cambridge, Mass., 1982.
7. G.E.P. Box and G.M. Jenkins. Time Series Analysis: Forecasting and Control, Revised Edition. Holden-Day, San Francisco, Calif., 1976.
8. G.C. Tiao and G.E.P. Box. Modeling Multiple Time Series with Applications. Journal of the American Statistical Association, Vol., 76, No. 376, 1981, pp. 802-816.
9. G.E.P. Box and G.E. Tiao. Intervention Analysis with Application to Economic and Environmental Problems. Journal of the American Statistical Association, Vol. 70, 1975, pp. 70-79.
10. SAS/ETS Users Guide--1982 Edition. SAS Institute, Cary, N.C., 1983.
11. The SCA System for Univariate-Multivariate Time-Series and General Statistical Analysis. Scientific Computing Associates, De Kalb, Ill., 1983.

Publication of this paper sponsored by Committee on Public Transportation Planning and Development.