REPORT NO. IA-11-0005

DEVELOPMENT OF TIME-SERIES BASED TRANSIT PATRONAGE MODELS

EXECUTIVE SUMMARY

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

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UMTA-IA-11-0005-85-1		

4. Title and Subtitle Development of Time Series Based Transit Patronage Models. Executive Summary	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. Dpeartment of Transportation Urban Mass Transportation Administration Office of Technical Assistance Univ. Research & Training Prog, Washington, D.C. 20590	13. Type of Report and Period Covered Final Report July 1983 - March 1985 14. Sponsoring Agency Code

15. Supplementary Notes

UMTA Project Director - Judy Z. Meade
Technical Advisor - Thomas J. Hillegas

16. Abstract

This document is a summary of research addressing the use of time series models to both evaluate service and fare policy changes by public transit operators, and forecast transit system ridership at the route, service, and system level. 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	18. Distribution Statement
Transit, Time Series, Modeling, Forecasting	

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

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

I. EXECUTIVE SUMMARY

OVERVIEW

This research develops a methodology for analyzing and forecasting public transit ridership. The methodology is applied to data from the public transit system in Portland, Oregon, and the statistical methodology for developing time-series models is described.

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. 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. This research builds upon the work of past 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 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.

3. Extensive use is made of the Box-Jenkins time-series models, resolving several problems that occur when standard regression models are used with time-series data, including multicollinearity and serial correlation.

THE METHODOLOGY

The basic methodology used in this project 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.

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.

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.

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. 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.

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. 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. 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.

A forecast of system ridership was made for twelve periods (months) ahead based on models developed from the previous eleven year's data. 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.

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: 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. 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.