Ferry Pricing Strategies Analysis

WA-RD 193.1

Final Report
November 1989

Washington State Department of Transportation
Planning, Research and Public Transportation Division

Final Report

Research Project GC8286, Task 10
Ferry Pricing Strategies Analysis

FERRY PRICING STRATEGIES ANALYSIS

by

Cy Ulberg
Research Associate

Washington State Transportation Center (TRAC)
University of Washington, JE-10
The Corbet Building, Suite 204
4507 University Way N.E.
Seattle, Washington 98105

and

Institute for Public Policy and Management (IPPM)
University of Washington
Graduate School of Public Administration
Parrington Hall, DC-13
Seattle, Washington 98195

Washington State Department of Transportation
Technical Monitor
D.M. Nutter
Director of Planning

Prepared for

Washington State Transportation Commission
Department of Transportation

November 6, 1989

WASHINGTON STATE DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
WA-RD 193.1		

4. TITLE AND SUBTITLE	5. REPORT DATE
FERRY PRICING STRATEGIES ANALYSIS	November 1989
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Cy Ulberg	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
Washington State Transportation Center (TRAC) University of Washington, JE-10 The Corbet Building, Suite 204; 4507 University Way N.E. Seattle, Washington 98105	11. CONTRACT OR GRANT NO. GC8286, Task 10
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Washington State Department of Transportation Transportation Building, KF-10 Olympia, Washington 98504	Final Report
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This report describes the results of research on ridership response to various fare pricing strategies. This research builds on the analysis of fare elasticity conducted by the Washington State Transportation Center (TRAC) under contract with the Washington State Department of Transportation (WSDOT). The basic objective of this study was to provide information from which to predict changes in revenues with changes in the fare structure.

Specific recommendations for an efficient and equitable fare structure require knowledge about more than the fare elasticities. The fare structure needs to reflect policies concerning the provision of mobility to island residents, achievement of a desirable ratio of revenue to operating costs, an optimal vehicle and walk-on passenger mix, and the like. However, fare elasticities can be used to explore the implications of different policies designed to address these issues.

On the basis of the fare elasticity research, these observations can be made: (1) Care should be taken when the fares in categories with elasticities less than -1 are increased. The probable result will be a net loss in revenue. (2) The loss in commuter ridership on Cross-Sound and Vashon routes will probably be greater than any increase in fares. A reduction in fare may actually increase ridership enough to offset the loss. A properly priced monthly pass may be a very good way to attract more of these riders. (3) Riders on the Vashon and Cross-Sound routes have a very strong tendency to shift from vehicles to walking onto the ferries when the fare increases. The fare structure can be used as a way to control the mix of vehicles and walk-on passengers. (4) For all three categories of oversized vehicles included in this study, ridership was elastic with respect to fares. Increasing those fares apparently causes people not to make those trips or to divert them around the Sound. This finding lends support to the idea of providing an off-peak discount for those vehicles.

To achieve a higher degree of certainty in these findings and to make quantitative estimates of revenue impacts of fare changes, a different approach is required than the one used in this study. Data need to be collected on a disaggregate basis. The report outlines how this research should be carried out.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
pricing, fare elasticity, ridership, data collection, ferry	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22616

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None	75	

DOT FORM 310-022 (11/86) -1216-

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Washington State Transportation Commission, Department of Transportation, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The purpose of this report is to describe the results of research on ridership response to various fare pricing strategies. This research builds on the analysis of fare elasticity conducted by the Washington State Transportation Center (TRAC) under contract with the Washington State Department of Transportation (WSDOT). The basic objective of this study was to provide information from which to predict changes in revenues with changes in the fare structure.

To accomplish this objective, four courses of action were pursued:

1) ferry fare elasticities were estimated for three route groups and nine fare categories

2) an outside consultant (Cambridge Systematics, Inc.) reviewed the previous work on fare elasticity,

3) an experiment with a fare change was proposed,

4) further analysis of the detailed data files was conducted,

5) a recommendation for a different approach to evaluation and prediction of ridership response to fare changes was developed.

Following the executive summary and recommendations, this report describes previous and current research on fare elasticity, and discusses the implications of the this research, on the basis of the consultant's findings. Following that is a presentation of results from other project activities.

EXECUTIVE SUMMARY AND RECOMMENDATIONS

Specific recommendations for an efficient and equitable fare structure require knowledge about more than the fare elasticities. The fare structure needs to reflect policies concerning the provision of mobility to island residents, achievement of a desirable ratio of revenue to operating costs, an optimal vehicle and walk-on passenger mix, and the like. However, because of data limitations, the estimated elasticities are not extremely stable, in that fairly small changes in model specifications result in different elasticities. The elasticity findings in this report should therefore be considered indications of the order of magnitude of the true underlying elasticities, rather than precise measures.

On the basis of the fare elasticity research, some observations can be made concerning the revenue impacts of changing fares. However, limitations inherent in the use of aggregate data should be considered.

1) The fact that ridership is elastic with respect to fares in several categories means that care should be taken when the fares are increased in those categories. The probable result will be a net loss in revenue. For instance, all ridership categories for San Juan routes have elastic relationships. Any increase in fares on those routes will probably mean a loss in total revenue.

2) The loss in commuter ridership on Cross-Sound and Vashon routes will probably be greater than any increase in fares. These riders constitute the bulk of ferry system ridership, so it is important not to alienate them. A properly-priced monthly pass may be a very good way to attract more of these riders. A reduction in fare may actually increase ridership enough to offset the loss. Theoretically, there should be an increase in net revenue. For both categories, however, a possible lag in the effect of up to one year is important to note.

3) Riders on the Vashon and Cross-Sound routes have a very strong tendency to shift from vehicles to walking onto the ferries when the fare increases. The fare structure can be used as a way to control the mix of vehicles and walk-on passengers. For instance, a fare increase applying to vehicle fares only will probably increase net revenue from vehicle passengers **and** increase revenue from walk-on passengers as a result in the shift from vehicles. The important point is that the total loss in ridership from a vehicle-only fare increase would be minimal. The result would be a shift of passengers to the walk-on category and a net increase in revenue.

4) For all three categories of oversized vehicles included in this study, ridership was elastic with respect to fares. Increasing those fares apparently causes people not to make those trips or divert them around the Sound. The result of a fare increase in these categories will probably be a net loss in revenue. Conversely, a decrease in those fares, will probably result in a net increase in revenue. This finding lends support to the idea of providing an off-peak discount for those vehicles. On the basis of this finding, an experiment with decreasing fares for commercial vehicles during off-peak travel times and directions was recommended. However, the recommendation was rejected for reasons detailed later in the report.

To achieve a higher degree of certainty in these findings and to make quantitative estimates of revenue impacts of fare changes, a different approach is required than the one used in this study. Data need to be collected on a disaggregate basis.

As detailed in the section entitled "Fare Change Response Evaluation Plan," interviews with a large number of travelers who **could** use the ferry should be conducted to determine individual traveler fare elasticities. The cost for such a study (largely, the cost for data collection) would be on the order of $300,000 to $500,000. This approach to evaluation of traveler response to fare changes could be expected to provide far greater accuracy in estimates of fare elasticity than the analysis of aggregate data used in this project. Furthermore, because the information would be at the individual level, responses to different kinds of fare changes (e.g., for different times of day or for different trip purposes) could be determined. In addition, a by-product of the research would be detailed information on customers and potential customers of the ferry system that could be used for other purposes such as marketing and service planning.

Because of the complications inherent in analyzing fare elasticities using traditional time-series regression analysis and aggregate data on Washington State Ferries' patronage, this traditional type of research should not be pursued further.

FARE RESPONSE RESEARCH

Since the Washington State Ferries (WSF) was formed with the purchase of the privately-owned Black Ball lines in 1951, the fare structure has not changed significantly, except for periodic, across-the-board increases to cover rising operating costs. Relatively minor changes in the structure have included various kinds of discounts for frequent users and surcharges applied during the summer months.

In 1982, a study of fare elasticity was published by the TRANSPO group for the WSF.(1) Data from 1970 to 1981 were used in the analysis. The report contained analysis of fare elasticity for four routes and for the total system. The analysis had two important shortcomings for analyzing fare elasticity. The first is that the analysis was confined primarily to the time period before 1979, when gasoline prices and area employment were rising constantly. The results represented relationships in an economic environment that was quite different from today's.

The second fault was that lags in the effects of variables were not considered. The study considered only concurrent effects of variables on ridership. However, changes such as fare levels may have effects that are not realized for several months, or even years. For instance, decisions to reside or work in places that require travel by ferry are influenced by ferry fare levels. However, such decisions cannot be made or implemented in a very short time. As a result, the impact of fare changes on ridership may lag. By not considering these lagged effects, the long-term impacts of fare changes were probably underestimated in the TRANSPO study.

---

1 Bullock, Kari and Leonard, Elena. "Fare Elasticity Study," The TRANSPO Group, February, 1982.

Another recent study of ferry patronage was an analysis of service elasticity conducted by Ritchie in 1985.(2) While fare levels were not explicitly considered in the analysis, the methodology employed was similar to the one TRAC employed.

Numerous studies of fare elasticity for public transportation agencies have been conducted in the past several years. Among them were studies by Mayworm, Lago and McEnroe(3) and Cervero.(4) A recent study conducted by Kyte, Stoner, and Cryer reported an analysis of ridership at Tri-Met in Portland, Oregon, that closely followed the Box-Jenkins methodology.(5) This study employed a similar approach.

ELASTICITY (DEFINITION)

Elasticity is a concept used in economics to describe the relationship between two variables. It is the ratio of the change in one variable to the change in another variable. For instance, the elasticity of ridership with respect to fare is -0.3 if a 10 percent increase in fare is accompanied by a 3 percent decrease in ridership.

Elasticity can be positive or negative depending on the relationship between two variables. If the elasticity is less than -1 or greater than +1, the relationship is said to be "elastic." The relationship is "inelastic" if it is between -1 and +1. In the analysis of fare elasticity, this distinction is important. If fare elasticity is elastic (that is, less than -1), this fact implies that a fare increase will lead to a loss in

---

[2] Ritchie, Stephen G. "Washington State Ferries Service Elasticity Study," October, 1985.

[3] Mayworm, Patrick; Lago, Armando M.; and McEnroe, J, Matthew. Patronage Impacts of Changes in Transit Fares and Services, U. S. Department of Transportation, Urban Mass Transportation Administration, Washington, D. C., 1980.

[4] Cervero, Robert. "Examining Likely Consequences of a New Transit Fare Policy, Transportation Research Record 877, 1982, pp. 79-84.

[5] Kyte, Michael; Stoner, James; and Cryer, Jonathon. "A Time-Series Analysis of Public Transit Ridership in Portland, Oregon, 1971-1982," submitted for publication in the Transportation Research Record, October 1986.

revenue, since the percentage loss of riders will be greater than the percentage gain in average fare. If the relationship is inelastic and negative, a fare increase will lead to a loss of riders, but an increase in total revenue.

There is no theoretical reason to believe that elasticity will be constant over the wide range of values that the variables may take on. In general, ridership will become more elastic if the fare increases substantially. Elasticities are usually estimated over a narrow range of values and can be used to predict changes within that range. Care should be exercised in predicting the impacts of major changes in fare levels. Historical elasticities may underestimate the impact of a major fare increase.

USE OF RESULTS

Fare elasticities can be used to investigate the ridership and revenue impacts of various fare structures. By holding constant all other variables that might influence ridership, the marginal differences in ridership resulting from different fare structure scenarios may be computed. From these marginal differences, the impact on revenues may be estimated.

The elasticity data for other variables, such as employment and gasoline price, may be used to predict actual ridership in the future. Of course, the use of these data requires that employment and gasoline price also be forecast. While it is difficult to predict exactly what will happen to economic factors such as these, ridership can be projected under various economic scenarios, and ranges of ridership forecasts can be established.

DATA SOURCES AND GENERAL DESCRIPTIONS

The data for the fare elasticity study came from a variety of sources. All methods for forecasting transportation demand generally draw on a similar set of variables, for two reasons: 1) all kinds of transportation demand decisions are based on the same set of criteria, and 2) a limited number of reliable and complete time series data sets are available for the researcher's use.

To develop a model for fare elasticity, time series for ridership (in various disaggregations) and fare levels were required. In addition, an effort was made to find other factors that could explain variance in ridership and that could not be explained by fare level alone. These factors did not have a strong effect on the relationships between fare and ridership, but they did account for additional variance and provided more precise measurements of fare elasticities.

Factors that were considered included those that 1) indicate the level of demand, such as population, employment, weather, and measures of economic activity such as retail sales, and 2) measure the cost of ferry transportation and of competing modes, such as fares, service levels and gasoline price. Population data were not included because information on a monthly, quarterly, or annual basis was not very accurate. Retail sales data were not included because they were very highly correlated with employment and did not add new information. Using historical data on number of vehicle spaces provided on a run-by-run basis, the impact of service levels on ferry patronage was evaluated. No signficant relationships were found.

In addition to time series data that were available, special events such as strikes, dock outages, and the Hood Canal bridge destruction in 1979 were found to be important. This winnowing process left six major categories of data to deal with: patronage, fares, special events, employment, gasoline price, and weather.

Patronage

Summary data file. All patronage data were derived from the summary data file (File 82) compiled by the WSF and recorded in machine-readable form since 1977. This data file included 38 categories of ridership. Some of these categories were used for only a short time and were not considered in this analysis. Other categories involved such small numbers that reliable relationships could not be found. Total ridership was taken to be the sum of total passengers (PASS-TOTAL) and total vehicles (VEH-TOTAL). Total ridership was divided into two classifications: type of ticket and travel mode. The first class distinguished between commuter and non-commuter patronage. Commuters were composed of the sum of commuter passengers (PASS-COMM) and commuter vehicles (AUTO-COMM). Non-commuters were the difference between total riders (defined above) and commuter riders.

The second classification was by mode of travel and included vehicles (with driver), passengers in vehicles, and walk-on passengers. Walk-on passengers was a separate classification in the summary data file (WALK-ON). Vehicles (with driver) was simply the total vehicle (VEH-TOTAL) classification, and passengers on vehicles was total ridership minus the walk-ons and total vehicle counts.

A third type of analysis was performed on commercial and over-sized vehicles using the classifications for under (TRK-REG) and over 48-foot (TRK-EX) trucks, recreational vehicles (O-S-REG) and trailers (TRAILER). This analysis is reported separately from the main analysis of fare elasticities.

Data were also classified according to route group. The summary data file contained patronage data for each route in the system. However, because of the variability in patronage, consistent relationships on a route-by-route basis were difficult to find. By combining similar types of routes together, some of the wide fluctuation in data was reduced and the underlying relationships were more readily apparent.

Three route groups were employed: Vashon, Cross-Sound and San Juan routes. Vashon routes included the following:

- Fauntleroy-Vashon,
- Vashon-Southworth, and
- Tahlequah-Point Defiance.

These routes were thought to be similar since they all involved trips to and from an island accessible only by ferry.

Cross-Sound routes included

- Fauntleroy-Southworth,
- Seattle-Bremerton,
- Seattle-Winslow,
- Edmonds-Kingston, and
- Mukilteo-Clinton.

These routes included destinations that were accessible by land, so that competing modes of transportation were feasible. In addition, they tended to have a high degree of commuter and residence use, as opposed to tourist use.

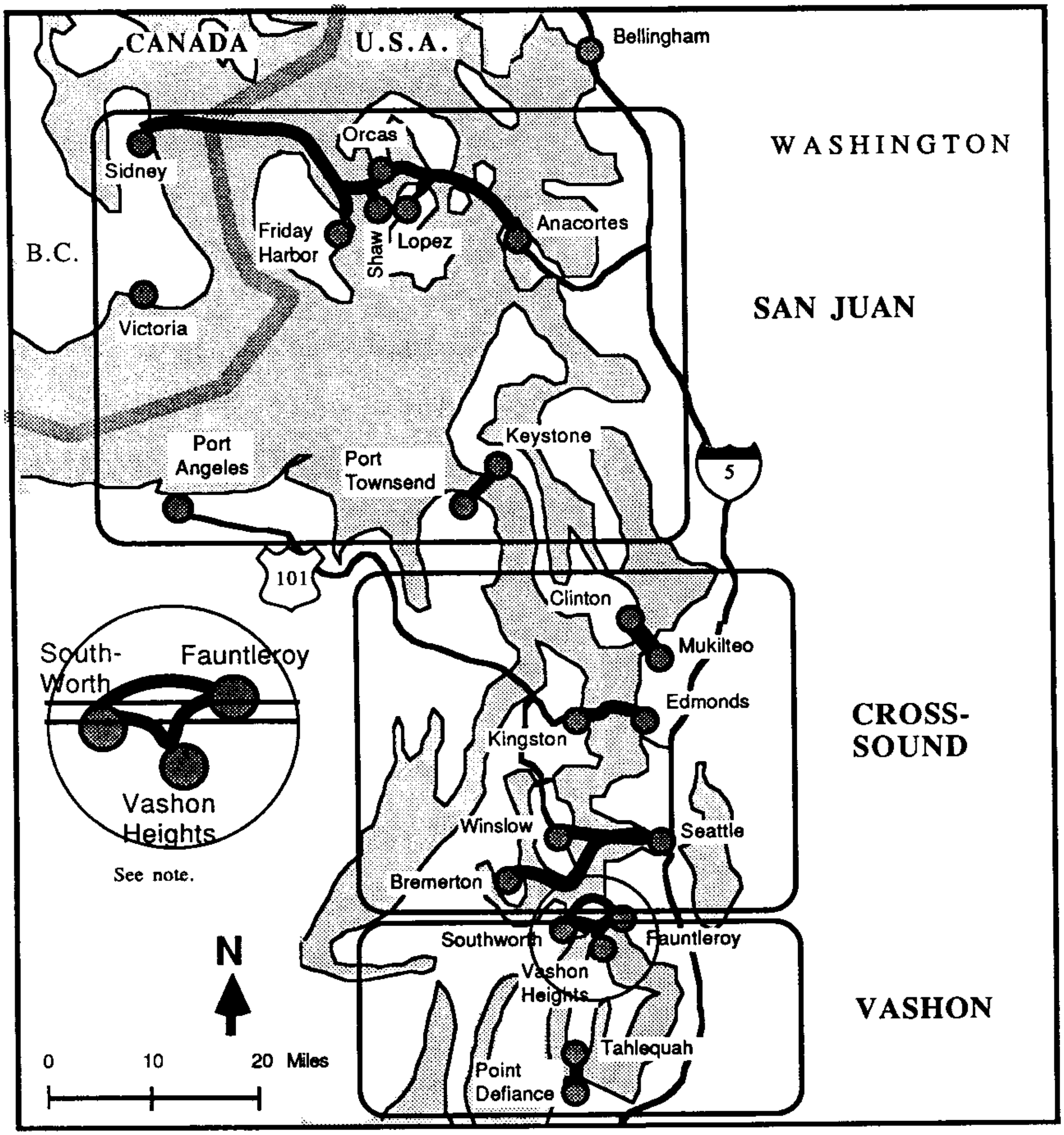
San Juan routes included all routes between and among Anacortes, the San Juan islands, and Sydney. In addition, the Keystone-Port Townsend route was included in this group. The latter route could be considered a Cross-Sound route, but it had a high percentage of tourist use, which made it more like the other San Juan routes. Figure 1 shows the route groups.

Downloading and data reduction. The summary data file was stored on the IBM 370 mainframe computer at Service Center 5 in files formatted for access by COBOL. The data were reduced and downloaded to floppy disks for analysis on a microcomputer. The daily data were collapsed into monthly records for each route. Four monthly records were produced:

- average data for Monday through Thursday,
- average data for Fridays,
- average data for Saturdays, and
- average data for Sundays.

The averages excluded holidays and days that might be affected by holidays. There were 16 routes, 115 months, four types of days, and 38 classifications of data, for a total of 280,000 data points.

These data points were further collapsed to get weekly numbers for the three route groups, reducing the total number of data points to about 13,000. Only ten



Note: The direct crossing between Fauntleroy and Southworth is in the Crosssound group. The crossing via Vashon is in the Vashon group.

Figure 1. Route Groups

types of ridership classification are used in the fare elasticity analysis, leaving a total of about 3,450 data points.

Two types of anomalies in the data for February through October 1980 were corrected before further analysis was performed. First, missing data in that period for some classifications were replaced with the averages for the same months in 1979 and 1981. Second, inn some cases, only one-way data were recorded during that period. The numbers were doubled before further analysis was performed.

Fares

Data on fares were collected directly from schedules provided by the WSF since 1977. Since all fare increases (with one exception) have been the same across the board since 1977, only the fare levels from the route with the highest patronage in each route group were used in the analysis. (Slight differences in fare changes among the routes were produced by the requirement to round off fares to the nearest 5 cents.)

The one exception to the across-the-board fare increases was in 1977, when the vehicle toll for the Anacortes to Sydney run was the only increase. In 1981, all fares increased, except for that route. Since the Anacortes- Sydney route was a relatively small part of the total patronage in the San Juan route group, these differences were ignored in the analysis.

The fare levels were adjusted for inflation by dividing by the Consumer Price Index (CPI) for the Seattle-Everett area and multiplying it by 100. This adjustment converted the fares to 1967 dollars. Adjustment for inflation is supported by economic theory, since the adjusted value reflects the cost relative to other expenditures that a potential ferry patron can make. Preliminary analysis confirmed that the deflated fare was a better predictor of ridership than the inflated fare.

Special events

Several events occurred during the ten years that were analyzed for this study that could potentially have had an effect on ridership. The primary ones were the destruction of the Hood Canal Bridge by a windstorm in February 1979 and the imposition of a 90-day period for commuter tickets to be valid in 1980. Other events included strikes and dock outages, but these were not very visible in the data since they were short or only affected one particular route. Major weather events such as the snow storm of November 1985 were also separated in the model separately from the weather variables that reflected longer term influences on ridership.

Employment

Employment data were derived from data published by the Washington State Employment Security Department. Estimates were made on a monthly basis for each county in the state. Data were analyzed for King, Snohomish, Pierce, and Kitsap counties from January 1977 through July 1986.

Gasoline price

Gasoline price represents one of the major costs for operating automobiles and usually has an effect on overall travel levels. This impact can affect ferry ridership. Monthly data were derived from the Lundberg letter for the time period of interest. The Lundberg association publishes data on the average price of gasoline at the pump for several regions in the United States and for each type of gasoline. For this study, the average price of unleaded gasoline for the Seattle area was used.

As with fares, economic theory justifies the adjustment of gasoline prices for inflation. The gasoline price data were deflated using the Consumer Price Index for the Seattle-Everett area.

Weather data

Variations in weather have not often been used in travel forecasting except for the impact of major storms. However, to analyze the patronage on the WSF, the study team hypothesized that longer-term weather trends might have significantly influenced ridership, since much of ferry travel is discretionary. Long periods of good or bad weather conceivably could have affected the level of out-of-state tourist traffic on the ferries and most likely did affect the discretionary use of the ferries by Puget Sound residents.

Data on average temperature, average percentage of sunshine, rainfall, and snowfall were taken from National Oceanographic and Atmospheric Administration (NOAA) information collected at SeaTac Airport. Time series data on rainfall and snowfall did not explain a significant amount of the variance in ridership, except for a few major storms. Those were considered special events and analyzed accordingly. On the other hand, temperature and sunshine did influence ridership and were included in the main analysis of the fare elasticity.

DATA ANALYSIS

The aim of the data analysis was to determine how changes in different categories of ridership were affected by changes in fares. Other variables were included in the analysis to explain changes in ridership that could not be explained by changes in fares alone. This section contains a short introduction to time series analysis, results and implications of the preliminary analysis, and the final results of the analysis.

Time series analysis

The analysis of time series is the same as the analysis of any other kind of data set, except that it must consider that data are measured in a continuous set of time periods. There are important differences between the analysis of a time series, such as monthly ferry ridership, and a cross-sectional set of data, such as the percentage of households owning a car in each county in the state. In a time series, successive measurements are probably (but not necessarily) related to each other. In a cross-sectional data set of county car ownership levels, arranged in alphabetical order, a relationship probably does not exist between adjacent values.

The basic statistical analysis method employed in this study was regression analysis. The reader should be familiar with a few terms to understand the discussion in this section of the report. The **dependent variable** is the one that is to be predicted (ferry patronage, in this case). The **independent variables** are the ones used to make the predictions (in this case, fares, gasoline price, employment and weather variables). **Residuals** are the differences between the actual values of the dependent variables and the predicted values. **Regression coefficients** are the values that relate the independent variables to the dependent variable. The prediction of the dependent variable is a constant plus a linear combination of the independent variables, as follows:

$$Y = C_0 + C_1 * X_1 + C_2 * X_2 \ldots$$

where Y is the dependent variable, the X's are the independent variables and the C's are the regression coefficients.

Pitfalls in time series analysis. In time series analysis, one characteristic of the data that the researcher must deal with is that successive values usually depend on previous ones. Unless this characteristic of time series is accounted for, successive residuals will be statistically related to each other (serially correlated), and one of the assumptions of least squares regression will be violated. On the other hand, methods exist to take advantage of this characteristic of time series data, and these are described in the next section.

A second issue to be dealt with is that the independent variables may not immediately affect the dependent variable. In standard regression techniques, only concurrent effects of independent variables on the dependent variables are considered. In a time series, however, the impact of a change in an independent variable can be detected after a lag in time and may continue to be effective for some period after the initial change.

Box-Jenkins approach. The Box-Jenkins, or ARIMA, approach to time series analysis takes advantage of some of the characteristics of time series to improve forecasts and to avoid methodological faults in the use of least squares regression analysis. That approach has been followed in this study. Three basic elements, or tools, are used in this approach: autoregressive terms (AR), integrated terms (I), and moving average terms (MA). Hence, the approach is called ARIMA.

The autoregressive term uses a lagged value of the dependent variable to help predict future values. The lags can be of any order, but three or four lags tend to be the limit of useful autoregressive terms. This tool minimizes the serial correlation of residuals and takes advantage of the structure of time series data to explain and forecast the dependent variable.

The integrated term is used to handle a time series that tends to drift over time. Its use entails replacing the values in the time series with the differences between adjacent terms (for a first-order integrated term). A second-order term is composed of the differences of the first-order terms. In other words, X_0 is replaced with $X_0 - X_{-1}$ for a first order difference. The first order difference is replaced with $(X_0 - X_{-1}) - (X_{-1} - X_{-2})$ or $X_0 - 2X_{-1} + X_{-2}$ for the second order difference. Often, the use of an integrated term eliminates the serial correlation of residuals. If the original data are replaced by their natural logarithms and a first order difference is applied, the resulting regression coefficients may be interpreted as elasticities. The analysis in this study employed this technique.

The moving average term uses the lagged values of the residual (rather than the dependent variable) to improve the explanatory power of the regression. It is useful when the autocorrelation of the dependent variable dies out very quickly.

The Box-Jenkins approach addresses seasonal effects in the data by employing the three tools to use appropriate seasonal lags. However, this study dealt only with seasonally adjusted data to simplify the presentation of the results (see the section on seasonal adjustments, below).

MicroTSP. MicroTSP is a time series analysis package designed for use on microcomputers. The program was written by David M. Lillien and is distributed by Quantitative Micro Software in Irvine, California. It was a very useful tool for the present analysis, since it can easily transform and manipulate time series data sets and perform regression analysis using autoregressive terms and moving average terms.

Preliminary analysis

A preliminary analysis was performed on the data to answer several early questions. Although most of the time series used for this analysis were available on a monthly basis, it was not clear whether one month was the appropriate time period to employ. Determining the seasonality of the data and the best way to deal with that factor was also important. Some variables were eliminated from the analysis, since they showed little, if any, relationship to ridership. An investigation of the effects of lags confirmed that lags had to be accounted for in the analysis.

Monthly versus quarterly analysis. In the first series of analyses, monthly time periods were employed. The pattern of relationships between fares and ridership was very unclear. Significant but small elasticities with lags distributed throughout a 12-month period tended to preceed the period of interest. These were interspersed with insignificant negative and positive elasticities and occasionally with significant positive elasticities.

These results implied that the effects of fare changes were felt for some months after they occurred but that unexplainable noise in the data was making the pattern difficult to decipher. To reduce this noise, monthly data were converted to quarterly data by being averaged over the three months in each quarter. The use of quarterly data resulted in more consistent fare elasticities, since the extraneous variation evened out when three months of data were combined into one data point. After this discovery, all subsequent analyses were performed on quarterly data.

Seasonal adjustments. Very clear seasonal patterns appeared in ferry ridership, with the greatest ridership occurring in the summer months. Since most of the fare increases occurred just before the summer months, detecting the impact of the fare changes on ridership by visually inspecting graphs of data that were not seasonally adjusted was very difficult. Although it would have been possible to account for seasonal effects in ridership using autoregressive or moving average terms in the regression, the study team believed that this method would have unnecessarily complicated the presentation of results.

As an alternative, the data were seasonally adjusted with the seasonal adjustment option available in MicroTSP **before** the logarithmic transformations, differencing, autoregressive, and moving average terms were applied. MicroTSP employs a traditional method to perform seasonal adjustment. For each period in the time series, the program computes the ratio between the value for that period and the average of the surrounding periods that comprise one year's worth of data. An average factor for each corresponding time period in a year is computed, and the time series is adjusted with these average factors.

The seasonally adjusted time series showed clearly the relative changes in data over the time period. For instance, a dip in summer ridership indicated that, in comparison with other year's summer quarters, that summer had experienced lower ridership, even though ridership had increased in comparison to the spring and fall quarters immediately surrounding it.

Significant variables. Some variables that were initially thought able to explain some of the variance in ridership were eliminated from further consideration after the preliminary analysis. These included employment data other than King County, service levels, rainfall, snowfall, and indicators of tourism levels.

Employment data were collected for King, Kitsap, Pierce, and Snohomish counties. Different combinations of these data series were tried for each of the route groups. King County employment was found to be more strongly related to the ridership data. The vast majority of the commuter traffic was generated by employment in King County, so it was not surprising that commuter traffic was related to that time series. In addition, King County employment was a good indicator of regional economic activity in general. Changes in the smaller counties reflected economic activity in those counties, but the bases were small enough that they did not have as strong a relationship with ridership as King County employment did.

Service levels were obtained for all route groups being analyzed. Preliminary analysis showed no relationships between service levels and ridership except for unusual circumstances such as strikes, dock outages, and the destruction of the Hood Canal bridge. These occurrences were represented separately in the analysis.

Rainfall and snowfall data were collected from NOAA, as well as average sunshine and temperature. No significant relationships were found between the time series representing rainfall and snowfall with the ridership data. In a quarterly time series, a winter quarter with a two-foot snowstorm appeared the same as one with six four-inch snowstorms. However, the impact of the snowfall in these two quarters on ridership was quite different. Major storms were better represented as special events rather than as a time series.

Since much of the ferry ridership was tourist oriented, an attempt was made to obtain an independent measure of tourism. One commonly used time series for this purpose is employment in the hotel and motel sector. These data were obtained from the Washington State Employment Security Department. However, all attempts to relate the data to ridership failed. This failure was probably due to the inclusion of gasoline price and employment data that were already related to tourism levels. Another effort was made to use data on hotel and motel occupancy. However, the available data sets were found to be inadequate for this study.

Analysis of lags. People are not always able to immediately change their traveling behavior in response to environmental changes. This is especially true for ferry patronage. When fares increase, commuter trips are not likely to change for quite a while. People may decide to relocate, change jobs, or buy an extra car so they can leave one at either end of the ferry route. New potential riders may not develop when the economic conditions are unfavorable. In any case, the full response to a fare change may not occur for several months. Preliminary analysis of the data revealed that responses to fare changes occurred for up to one year afterward. Responses to changes in gasoline price or employment usually occurred within two quarters. Responses to weather conditions (represented by average temperatures and percentage of sunshine), as well as special events, occurred immediately. The final analysis examined all possible lags that were uncovered in the preliminary analysis.

Final analysis

The preliminary analysis showed that five independent variables (plus intervention variables to represent special events) should be used in the final analysis to explain the dependent variable, ferry ridership. Ridership was disaggregated in three ways. First, it was separated into three route groups: Vashon, Cross-Sound, and San Juan routes. Secondly, it was separated into commuter and non-commuter use, according to the use of the frequent traveler coupon books. Thirdly, the ridership was separated by mode: walk-on, passenger on a vehicle, and driver of a vehicle.

None of the variables were stationary. To induce stationarity, first differences were employed in all the analyses. To be able to interpret regression coefficients as fare elasticities, a logarithmic transformation was applied before differences were computed.

The preliminary analysis also identified the range of lags that should be considered. The discovery that fare changes could affect ridership up to one year after had they occurred meant that lags of up to four quarters would have to be investigated for the fare variables. Changes in gasoline price and employment had little effect after three months, so no lag greater than one quarter was investigated for those variables. The weather variables, average sunshine and temperature, would only have concurrent effects. The same applied to intervention variables representing the effects of special events. Where a moving average or autoregressive term could improve the explanatory power of the regression equation, terms of order one were used. Higher order terms were not significant in any of the equations.

__Interpretation of confidence intervals__. Results for all of the elasticities are reported using confidence intervals. Since the data used in this analysis represent only a sample of all possible time periods, the resulting fare elasticities are **estimates** of the actual fare elasticities. By using the standard error of the coefficients, a confidence interval for each elasticity can be computed. In results for this study, 90 percent confidence limits are reported. The proper interpretation of these limits is that the reader can be 90 percent sure that the actual fare elasticity falls within the reported intervals.

In cases where a fare elasticity was significantly different from zero for successive lags, elasticities were summed and the standard error for the sum was computed from the individual standard errors. While the total impact of a fare change may not have occurred immediately, this sum of fare elasticities represents the effect that would have been felt after __n__ lags, where __n__ is the number of successive significant fare elasticities.

__Fare elasticity results__. Figures 2, 3, and 4 show the confidence intervals for fare elasticities for each of the route groups and for each of the methods of disaggregation.

The highest (most negative, that is) fare elasticities occurred for the San Juan routes, followed by the Cross-Sound routes and the Vashon routes. The confidence interval midpoint for the San Juan routes' total ridership was -1.02; the midpoint for the Cross-Sound route was -0.83, and the midpoint for the Vashon route was -0.62. These differences were expected, since a much higher percentage of passengers ride ferries on a discretionary basis on the San Juan routes than the other two route groups. These data showed that there is a better than 50 percent chance that San Juan ridership is elastic with respect to fares. In other words, fare increases on those routes would likely reduce total revenue. The other two route groups are

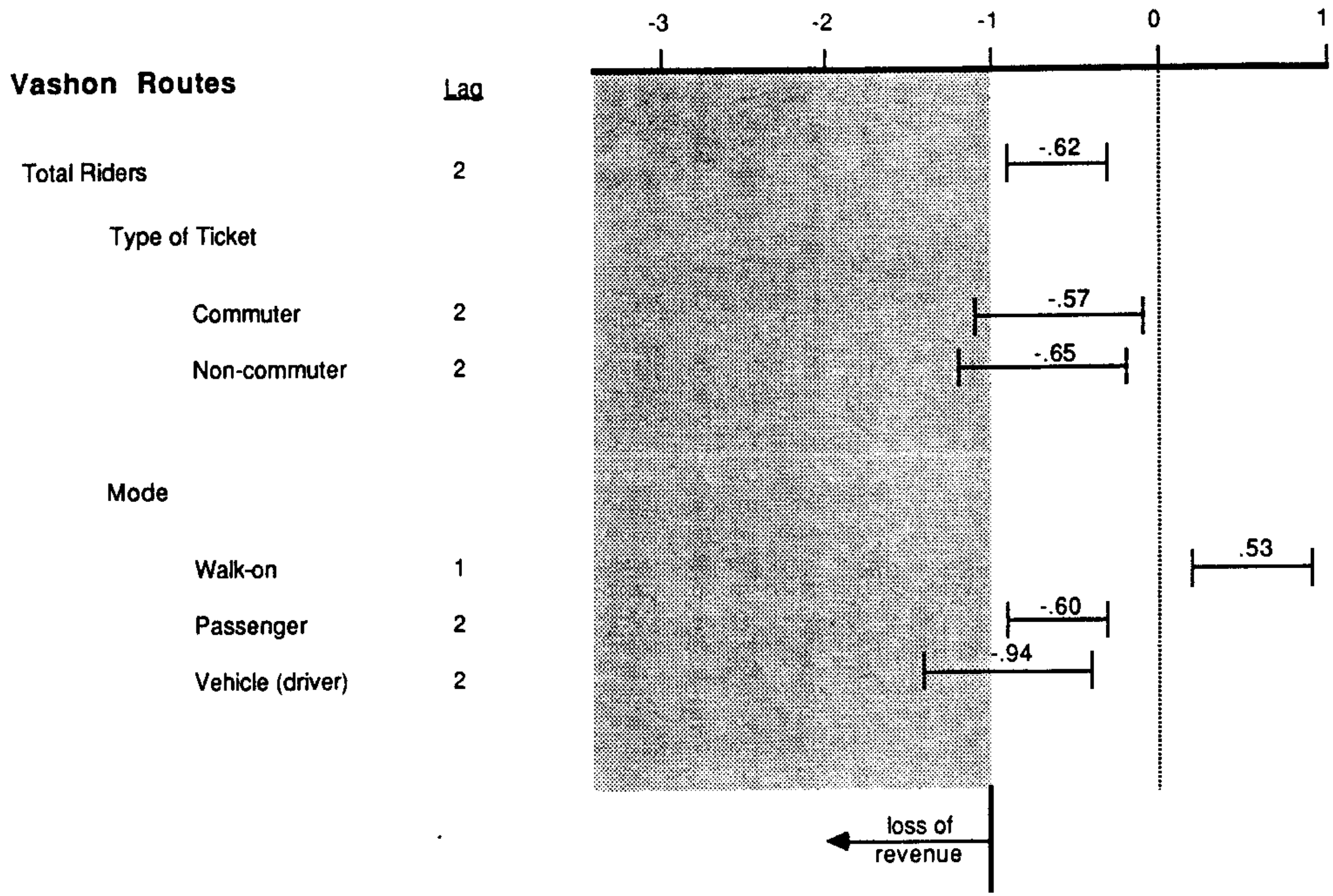


Figure 2. Fare Elasticity on Vashon Routes

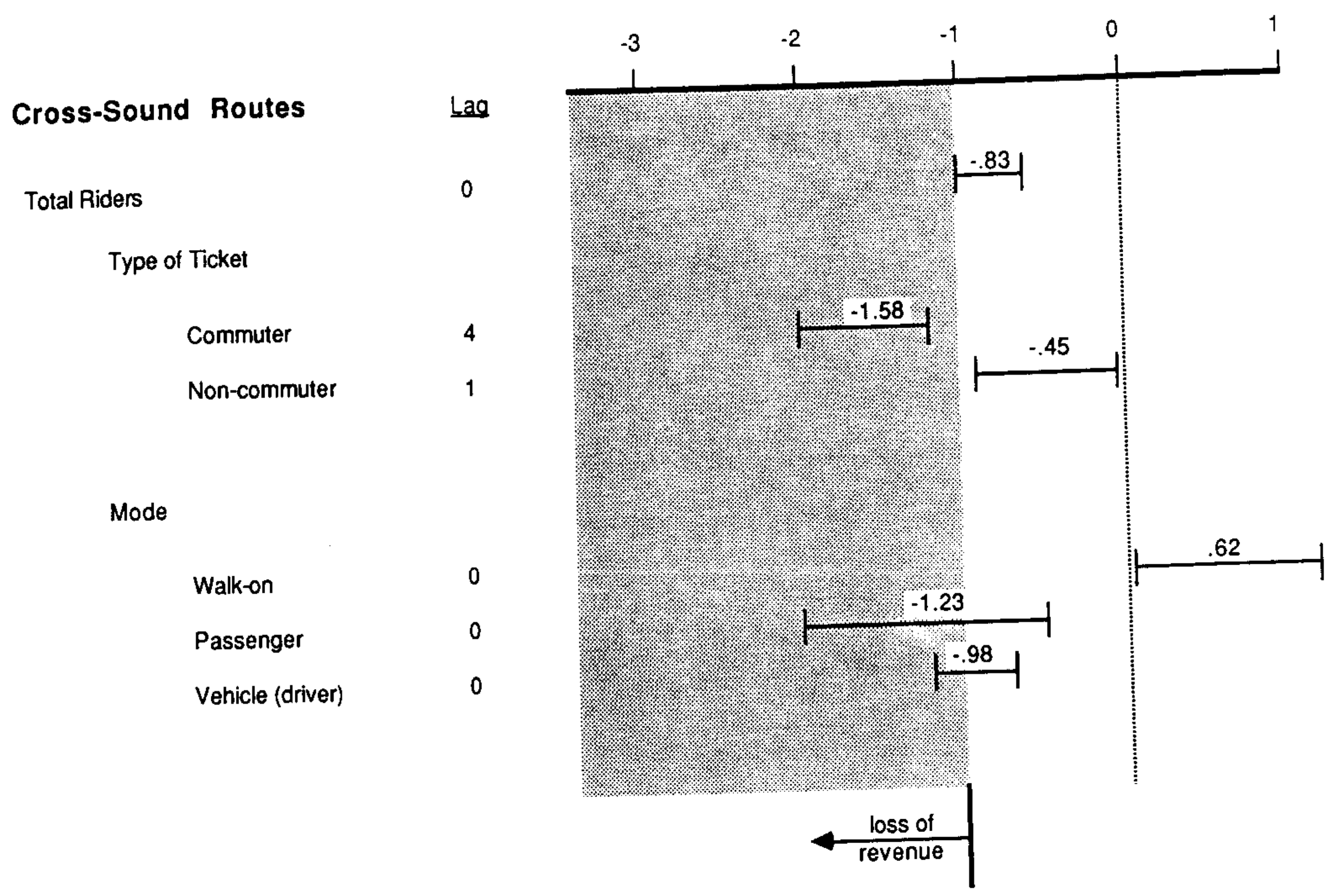


Figure 3. Fare Elasticity on Cross-Sound Routes

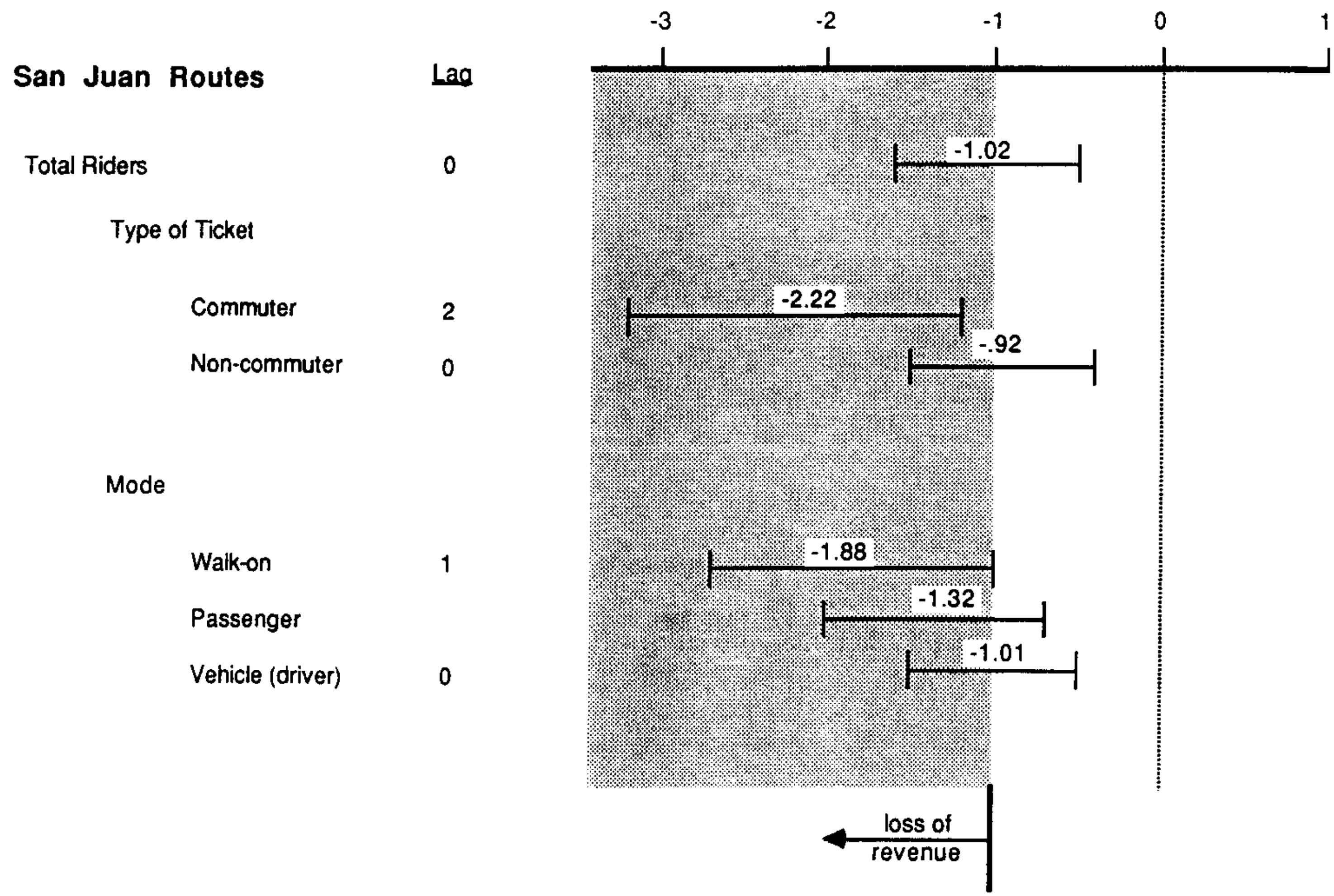


Figure 4. Fare Elasticity on San Juan Routes

probably inelastic, but individual categories of ridership must be examined to determine the impact on total revenue.

For the Cross-Sound and San Juan routes, commuter ridership is highly elastic with respect to fares. For Vashon routes, it is negative, but inelastic. For all three route groups, significant fare elasticities exist for commuter ridership for lags of up to four quarters. Losses in ridership after a fare increase tend to come from commuters to a greater extent than non-commuters, but the effect is not felt immediately.

These results should not be interpreted as a shift from commuter to non-commuters as a result of a fare increase, for two reasons. First, all fare elasticities for non-commuters are negative, meaning that a fare increase results in ridership loss in that category, as well as for commuters. Secondly, commuters probably do not buy an even more expensive ticket when the cost goes up. However, higher prices possibly lead to less use of the ferries in general and thus make the frequent user discount unavailable to some patrons.

Patterns by type of ticket vary among the three route groups. For the Vashon and Cross-Sound routes the fare elasticity for walk-on passengers is positive, while both passenger and drivers in vehicles show a strongly negative fare elasticity. The best interpretation of this finding is that, for these routes, a fare increase causes riders to shift out of their vehicles. They become walk-on passengers.

The pattern is very different for San Juan routes. While all ridership categories experience a loss as a result of a fare increase, the loss in walk-on passengers appears to be the greatest (the confidence intervals are overlapping, however). This is probably due to the fact that walk-on passengers on San Juan routes tend to be making discretionary trips to a greater extent than those on Vashon or Cross-Sound routes.

For the San Juan routes, any across-the-board fare increase results in a loss in total revenue. However, for the other two route groups the results are not as clear. Since the fare elasticity for vehicle drivers is about -1 for both route groups, total revenue for this category should be expected to remain constant. For the Vashon routes there would be a total revenue gain, both from passengers in vehicles and from walk-on passengers. For the Cross-Sound routes, the gain in revenue from walk-on passengers would likely make up for the loss in total revenue for vehicle passengers.

Influence of other variables. The primary objective of this study was to determine fare elasticities. However, some consistent relationships with other factors should be taken into account, even though little can be done to affect those variables in the future.

Special events can have a significant impact on changes in ridership during one quarter. To reduce errors in regressions and to develop more accurate estimates of fare elasticities, variables were introduced to represent these special events. In the language of ARIMA analysis, the factors are represented as **intervention variables**. They are quite similar to the dummy variables used in standard regression analysis.

For the purposes of this study, the special events that were explored included

- the 1978 strike (third quarter)
- the destruction of the Hood Canal bridge in February 1979,
- the 1980 strike (April 5-17),
- the 1981 strike (May 20-22),
- the construction of the Mukilteo dock in 1982 (first quarter), and
- a record snowfall in November 1985.

Only three of these special events had a significant impact on changes in quarterly ridership. These were the destruction of the Hood Canal bridge, the construction of the Mukilteo dock, and the snowstorm in 1985. They were not used in all of the regressions, but they did serve to reduce the error in several of the estimates of fare elasticity.

Employment had an uneven relationship with ridership. When a significant relationship existed between employment and ridership, it was positive. That is, higher employment was associated with higher ridership. It was consistently related with walk-on riders and commuters on the Vashon and Cross-Sound route groups. This implies that when employment increases in King County, the ferry system can expect to carry more commuters, especially as walk-on passengers.

For almost all categories of ridership, there was a negative relationship between gasoline price and ridership. As gasoline price increase, people generally travel less. This apparently leads to less travel on the ferries as well. Since there were no positive relationships between ridership and gasoline price for any category, the contention was not supported that riders switch to the ferry when the competing mode of driving becomes more expensive.

In virtually every category of ridership, a strong positive relationship existed between average temperature and ferry ridership. In a few cases there was also a positive relationship between average sunshine and ridership. Although a positive relationship between good weather and ridership was not surprising, the researchers were surprised that such strong and consistent relationships existed even when the weather variables were averaged over a whole quarter.

INTERPRETATION OF RESULTS

The major results of the study (the fare elasticities) were reported from three different classification methods: type of route, type of ticket, and type of mode. In this section, results are interpreted using these three classifications.

Ridership prediction

Fare elasticities can be used to predict ferry ridership. For fare elasticities with no lag, the method is very simple. Since the analysis used real fares, the actual fare increase should be reduced by the quarterly inflation rate before applying the fare elasticity. The percentage loss in ridership is estimated by multiplying the elasticity by the percentage change in **real** fares. For instance, suppose that an 11 percent fare increase is to be implemented during a time when the Consumer Price Index is going up 1.5 percent per quarter. The real fare increase would be 9.5 percent. Suppose further that the elasticity for a particular ridership group is -1.58. The estimated ridership loss would be 15 percent. That is, the ridership would be 15 percent less than one would predict **without** the fare increase. However, no fare increase is equivalent, in this case, to a 1.5 percent decrease in **real** fare, due to inflation. The prediction of ridership without a fare increase should take this into account. In the remainder of this section, references to fare changes should be interpreted with inflation in mind.

Route group differences

The three route groups (Vashon, Cross-Sound, and San Juan) were defined to represent different types of service and patronage. They were also defined so that they would be relatively homogeneous within themselves.

The Vashon routes are unique because they provide transportation to places accessible only by ferry. No highway alternatives exist. They are different from the San Juan routes because of the proximity of Vashon Island to Seattle, meaning that a high percentage of the trips are work-oriented. The routes are also shorter (in general) than the San Juan routes. However, the ridership shows significant seasonality because of the attractiveness of the island as a destination for mainland-based tourism during the summer months.

The Cross-Sound routes also carry some summer tourists, but the bulk of the ridership is related to the work commute for the people who live in Kitsap County and commute to King or Snohomish county to work. Another distinguishing feature of the Cross-Sound routes is the fact that people can drive around Puget Sound and reach the same destinations served by these routes.

The San Juan routes serve some work commute trips, but the bulk of the trips are recreation- or shopping-related. This applies both to residents of the islands and visitors from the mainland. These routes show a high degree of seasonality.

The differences among the routes accounts for differences in responses to changes in fares and to other factors used in the analysis. A major difference is the overall response to fare changes. The San Juan routes, which serve the highest number of discretionary trips, has the strongest relationship between fares and ridership. The relationship is so strong that a change in fares results in an even larger change in ridership. This means that a fare increase will actually reduce the total revenue on these routes. Conversely, a fare reduction would, theoretically, increase revenue from these routes.

The Cross-Sound and Vashon routes have similar relationships between fare changes and ridership, but the Cross-Sound routes show a slightly stronger relationship between fares and ridership than do the Vashon routes. This variance is probably due to the fact that a greater percentage of riders on the Vashon routes are captive riders than on the Cross-Sound routes.

The relationships for employment and weather variables are related to the percentage of discretionary trips in the route groups. A greater percentage of discretionary trips leads to a smaller relationship to employment and a correspondingly stronger relationship to weather variables.

Commuter vs. Non-commuter

The patterns of responses to fare changes by commuters and non-commuters vary by route group. On the Vashon routes, the commuters respond slightly less strongly to fare changes than do non-commuters, probably because commuters are captive riders. They have no other way to get to work. If they do make fewer trips, they probably cut out the discretionary recreational or shopping trips that they had been taking with their commuter tickets.

On the other hand, Cross-Sound commuters show a much stronger reaction to fare changes than do non-commuters on the same routes. The response to fare changes is not immediate, but the end effect of a fare increase is a large reduction in commuter trips on these routes. Three explanations can probably account for this reduction. First, for some commuters, the alternative of driving around is available and no doubt happens when fares increase. Second, with the 90-day limit on the use of commuter tickets and a reduction in the use of commuter trips for trips other than work trips, many riders probably shift from the commuter to the non-commuter type of ticket. Third, since fare increases tend to be related to ridership reductions for as long as one year after the reductions have occurred, some people eventually make different choices in location of employment or residence as a result of the fare increases.

The pattern in commuter/non-commuter responses to fare changes for the San Juan routes is the same as the pattern for Cross-Sound routes. The explanations are probably also the same, except for the fact that San Juan residents have no highway alternatives available.

As would be expected, relationships between employment and commuter ridership is stronger than those between employment and non-commuter ridership, for all routes. Higher employment levels lead to higher ridership, especially for commuters.

Conversely, non-commuter ridership is more strongly related to weather variables than is commuter ridership. Discretionary, recreation trips occur more often when the weather is especially good.

Mode

Fare changes have an interesting effect on the mode of ridership on the ferries. For Vashon and Cross-Sound routes, the number of walk-on riders actually increases when fares go up. However, clearly the additional riders come from the other two modal categories, vehicle drivers and passengers. A major response to fare changes is for people to ride the ferries without a vehicle. This is especially true for routes with a high percentage of commuter use.

The San Juan routes display a different reaction pattern. Walk-on passengers have a slight tendency to respond **more strongly** to a fare change than do vehicle passengers. This is probably because walk-on passengers on San Juan routes are more likely to be making discretionary trips that those on other routes.

Oversized Vehicles

The fare elasticity for oversized vehicles was investigated for this project. The method was exactly the same as for other ridership classifications. Numbers of oversized vehicles was used as the dependent variable in the time-series regressions analysis, and fare elasticities were derived from the coefficients for the fare variable. The following three types of vehicles were examined:

- commercial vehicles under 48 feet long,
- commercial vehicles over 48 feet long, and
- recreational vehicles.

Other classes of oversized vehicles, such as those with trailers and buses, had frequencies too low to detect significant relationships. For the three vehicle classes that were investigated, the frequencies were too low for significance on the Vashon and San Juan routes. Only results for the Cross-Sound routes are presented here.

Figure 5 shows the fare elasticity confidence intervals for three classes of oversized vehicles on the Cross-Sound routes. The fare elasticities are highly elastic. Clearly, the alternative of driving oversized vehicles around the Sound is an attractive one when fares are increased. In fact, fare increases appear to result in such a loss of use by these types of vehicles that total revenue from these classes is reduced when fares are increased.

The difference in fare elasticity between the under and over 48-feet commercial vehicles is insignificant. However, larger vehicles tend to opt for alternatives more when fares are increased.

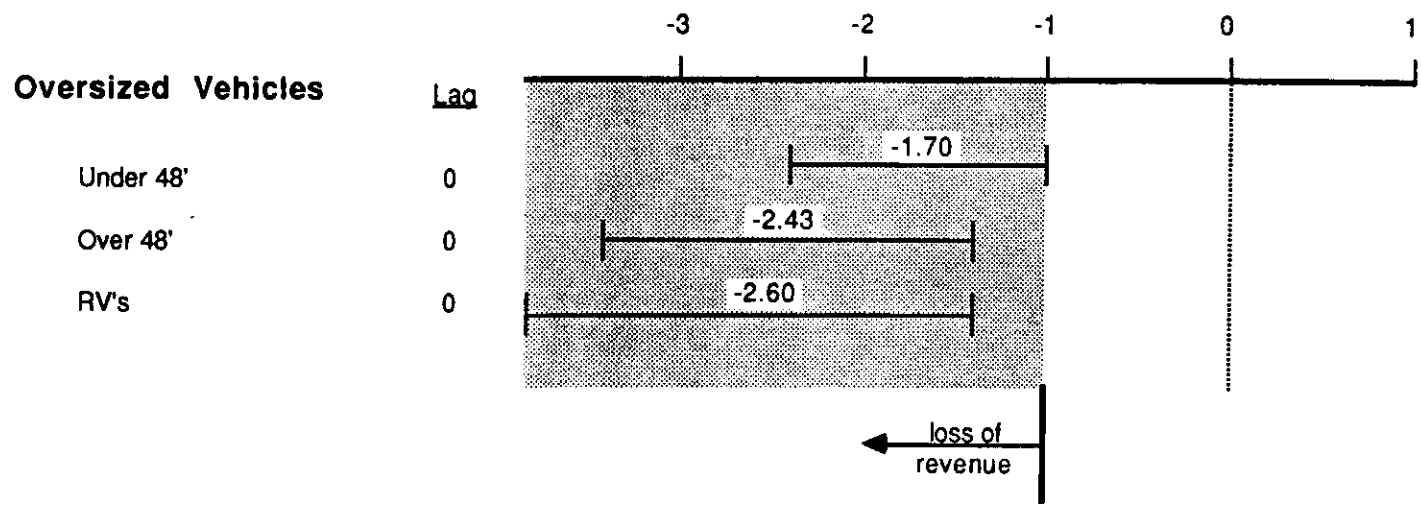


Figure 5. Fare Elasticity for Oversized Vehicles

EVALUATION OF RESULTS

The analysis of fare elasticity described in this section was reviewed by Cambridge Systematics, Inc., under subcontract with TRAC. A complete text of the review is contained in Appendix A. The first paragraph of the review states

> This study has been performed insightfully and competently. Its procedures are appropriate for the data that are available, are statistically rigorous, and are generally accepted and widely used in the demand analysis profession. The results are generally plausible, both theoretically and with respect to the particular data and situation being examined. Because of data limitations, the estimated elasticities are not extremely stable, in that fairly small changes in specification result in somewhat different elasticities. The estimates should therefore be considered indications of the order of magnitude of the underlying elasticities, rather than precise measures.

The report goes on to identify two major results of the study: 1) elasticities are fairly high, even exceeding 1 in some cases, and 2) elasticities are different for different routes and different types of travelers. Furthermore, it says that both of these results have policy implications that are justifiable on theoretical grounds.

The instability in the results is inherent in the data used in the analysis. The fare elasticities were based on aggregate data, namely, data on the total number of passengers and vehicles on each route during each time period. The estimates based on these data are imprecise because of fundamental limitations in the data. The two most important limitations are as follows:

1) **Insufficient variation in fare data.** Only a few general fare changes have occurred over the years for which comparable ridership data are available. Furthermore, the relations among fares on different routes have remained essentially the same over time. With little variation in fares, it is difficult to estimate meaningful fare elasticities.

2) **Aggregation over routes and days.** Patronage figures were aggregated over routes within a group and over days within a quarter. Given the lack of meaningful fare variation over routes with in a group and days within a quarter, this aggregation is necessary. However, it results in an average fare that does not correspond to the fare that any traveler actually faces. Since the fares used in estimation of elasticities are not the same as the fares that travelers actually face, the estimation of elasticities cannot be expected to be very precise.

These problems evidence themselves in large standard errors and in estimates that change considerably when the model specification is changed only slightly. It is important to note that these problems are endemic to the data, and that any analysis based on these aggregate data can be expected to encounter the same difficulties. To obtain better estimates of elasticities, another type of data, namely data containing greater price variation, is required. This is the motivation for collecting and analyzing customer-based data, which is the basis for the proposed evaluation method outlined in a later section.

OTHER PROJECT ACTIVITIES

COMMERCIAL VEHICLE DISCOUNT PROPOSAL

A proposal was developed in this project to experiment with a fare change to provide an empirical test of the fare elasticity results. Specifically, the proposal was to offer off-peak discount fares for commercial trucks on a demonstration basis on the Seattle-Winslow route. The discount was proposed as a special promotional tariff to be effective for three months starting in January 1989. The change was to receive a comprehensive evaluation by TRAC.

The justification for the proposal was to recapture some of the lost commercial traffic across the Sound. The arguments were stated as follows:

> The Washington State Ferries (WSF) is seeking to improve its relationship with the trucking industry. Some argue that high fares and reduced service has caused commercial truckers to seek alternatives to using WSF ferries. Highway facilities have been improved and, in some cases, provide an easier and faster route to and from the Olympic Peninsula than does WSF. Recent reductions in fuel prices have made driving to the peninsula even more attractive than it was in the past.
>
> A recent study of fares conducted by TRAC indicated that the fare elasticity for Cross-Sound commercial vehicles was relatively high. In fact, the research, based on experiences with fare changes over the last 10 years, indicated that increasing fares for commercial vehicles may have actually resulted in reduced revenues from this class of vehicle.

The advantages of this promotional demonstration are listed below. The proposal stated it would do the following:

- recapture part of the cross-Sound commercial vehicle market,
- use unsold space on off-peak sailings,
- reduce highway congestion and highway maintenance costs by taking some trucks off the road,
- reduce demand during the peak hours,

- possibly increase WSF revenues,
- reduce shipping costs, and
- develop a better understanding of user response to fare changes.

Specifically, the proposal included the following provisions:

- offer a 33 percent discount for off-peak sailings for commercial vehicles on the Seattle/Winslow and Winslow/Seattle routes
- off-peak sailings on the Seattle/Winslow route are all those except the 6:20 AM and the 3:45, 4:35, 5:25 and 6:10 PM sailings
- off-peak sailings on the Winslow/Seattle route are all those except the 5:35, 6:20, 7:10, 7:50 and 8:40 AM sailings

An analysis of patronage and revenues on the Seattle/Winslow route resulted in the following projected impacts of such a proposal:

- The off-peak sailings have the capacity to accommodate at least three times the current volumes of commercial vehicles (both in total space and in tall capacity).
- Assuming that 1) the level of shifts between other Cross-Sound routes or peak sailings on the target routes to the off-peak sailings on the target routes is between 10 and 25 percent and 2) the discount generates between 30 and 100 percent new commercial vehicle traffic on the off-peak sailings on the target route, the expected range of revenue impact is between a $32,000 loss and a $33,000 gain (on a total projected $300,000 revenue during the three months).
- Truck traffic on off-peak sailings on the target routes would approximately double and other cross-Sound commercial traffic would diminish by between 10 and 25 percent.

The Washington State Ferries (WSF) decided that conducting such a demonstration would not be prudent, for the following reasons:

- it was concerned about the public relations aspects of offering an off-peak discount on one route only and not on other routes;
- it was concerned about the possible operational impacts of too many tall vehicles if the demonstration exceeded expectations; and
- the proposed reconstruction of the Bremerton trestle would require a shutdown of vehicle service, resulting in a shift to Winslow, making the capacity constraints possibly problematic.

In addition, WSF was generally reluctant to increase commercial traffic at Colman Dock because of operational and other constraints.

As a result of the difficulties in proposing a feasible fare demonstration, there was no change in fare levels during the course of this project on which empirical tests of the fare elasticity results could be performed.

DATABASE EVALUATION

The original proposal for this project assumed that the detailed patronage file (File 47) could be used for an analysis of patronage responses to fare changes on a route-by-route and run-by-run basis. The proposers recognized that there would be problems with the data, but they realized that they could fix them by referring to the summary data file and by carefully searching for data points that were out of bounds.

Appendix B contains a memo that describes the state of the detailed data file and outlines some of the severe problems with using the data for the intended purposes. The primary problem was missing data, sometimes for months at a time. Another difficult problem was the frequent occurrence of duplicate records for the same sailing. Sometimes the duplication was legitimate because extra vessels were assigned to accommodate overloads. However, in other cases, the source of the duplicate record could not be determined.

The following suggestions, based on research experience, were offered to guide the improvements in data collection. They are compatible with the recommendations of the "Ferry Systems Data, Scheduling and Billing" study completed by TRAC in 1987. In addition, they are compatible with the Automated Revenue Control System (ARCS) under development by WSF.

1) Traffic data should be transferred electronically from the cash registers to the computer data file. This is not a new suggestion, but it is meant to reinforce the findings of many other people who have had to deal with the data. There are many instances of obvious mistakes

in recording data, including wrong dates, wrong run numbers, and clearly wrong ridership data.

2) Just one traffic data file should include all the detailed data. A summary data file can be produced by computer from the detailed data file. The format of the summary data file can be flexible to respond to the needs of the users. The advantages of one main file are that it will get more use and will be easier to keep up-to-date.

3) One person should be in charge of the data file. This person's responsibilities should include **monthly** entry of data into the file, with consistency checks immediately after data entry.

4) The computer should produce the totals included in the data file. The totals often do not agree with the sum of the individual ridership categories. There is no reason that totals should be moved along with the numbers in the individual categories.

5) Unique numbers should be run numbers for every sailing. If an extra vessel is added to a sailing, the sum of the ridership figures should be entered into the record and a vessel code should be created that will indicate that two specific vessels were used. Recording data this way will allow for consistency checks. No repeat run numbers should be used for the same day and route.

6) The data files do not have to be as large as they are. At a minimum, the size of the fields could be reduced from five to three characters for most of the ridership categories, and some unused categories could be eliminated. In addition, further reductions could be accomplished by packing the data (not storing it in character format). For our purposes, with minimal aggregation of categories, we were able to reduce the record length from 207 to 44 bytes. Further reductions would be possible with more efficient packing.

7) If the size of the records were reduced, data could be downloaded from the mainframe to floppy disks for use on PCs. The data would become more accessible and easier for most people to use. All of the detailed data for one year can be stored on five high density floppy disks.

As a result of the problems with the detailed data file, the project team recommended against using the data for a route-by-route and run-by-run analysis of fare elasticity. The incomplete time series and the number of errors in the existing data prevented the researchers from conducting the analyses that were envisioned. Instead, a different approach to analyzing responses to fare changes is proposed, as outlined in the section entitled "Fare Change Response Evaluation Plan."

COMPUTER PROGRAM DOCUMENTATION

One of the by-products of the analysis of the detailed data file was some PC-based software to analyze the data. The detailed data file was downloaded from the mainframe to floppy disks and converted to a format that can easily reside in moderately-sized hard disks. Planners can then readily access the information and also detect problems with the data.

Four separate programs were developed:

1) INDEX - prepares an index of the data to speed access to individual records,

2) MISSING - detects missing records,

3) DUPL - detects duplicate records for the same run, and

4) WSFTRAF - shows data in individual records or in averages by days, weeks, months, or parts of months.

The following sections describe the software that was developed.

Downloading the data

File 47 detailed ridership data are stored either on tape or as active ADABAS files. The length of each record (which gives statistics for a single run) is 207 bytes. Data for each year take up about 35 megabytes. For this investigation, the data were downloaded to floppy disks using SUPERNATURAL, software designed to produce reports from large databases.

The format chosen for the downloaded files resulted in 44 byte records. The reduction in record size was achieved partly by combining some categories of ridership statistics and partly by reducing the field size for each variable to the minimum size needed. The format for the downloaded data can be seen in Table 1. The data were also divided into separate files according to route. Each downloaded file contains a year's worth of data for one route. The size of the files depends on the number of daily runs on the route. The largest files contain about 250 kilobytes.

All of the data for one year could be contained in about 6 megabytes, or about 5 high density floppy disks.

The files are labeled with a name in the format RTxxCYyy.DTA, where xx is the route number and yy is the year. INDEX is a program that creates an index, making access to individual records identified by day, month, and run almost instantaneous. To run INDEX, the following steps are required:

1) make sure the file RTxxCYyy.DTA is in the active subdirectory along with INDEX.EXE;

2) type INDEX;

3) respond to the prompt [route =] with the two-digit route number xx; and

4) respond to the prompt [year = 19] with the two-digit year yy.

The program first has to blank out the index file. This process takes less than 1 minute, and the percentage of the process completed is displayed on the screen. Then the program indexes the records. The number of records completed and the number of duplicate records is displayed. When the indexing is complete, press any key to exit the program.

The index file corresponding to RTxxCYyy.DTA is called RTxxCYyy.IND. It will be roughly 60,000 bytes long, depending on the number of duplicate records. Once the index file has been created, the process need not be repeated unless the data file is changed.

<u>Checking for missing and duplicate records</u>

The process for checking for missing and duplicate records is the same. To run MISSING or DUPL, the following steps are required:

1) make sure the files RTxxCYyy.DTA and RTxxCYyy.IND are in the active subdirectory along with MISSING.EXE or DUPL.EXE;

2) type MISSING or DUPL;

3) respond to the prompt [route =] with the two-digit route number xx;

TABLE 1. DOWNLOADED DATA RECORD FORMAT AND DEFINITION

Starting Column	Ending Column	Variable Name	ADABAS Long Name
1	1	Route	ROUTE
2	2	Day	DAY
3	3	Month	MONTH
4	4	Year	YEAR
5	5	Run	RUN
6	6	Hour	HOUR
7	7	Minute	MINUTE
8	8	Vessel	VESSEL
9	10	Full Fare	FULLFARE
11	11	Half Fare	HALFFARE
12	13	Commuters	PASS-COMM
14	14	Bicyclists	BICYCLE-FF BICYCLE-HF BICYCLE-COMM BICYCLE-EXC-FF BICYCLE-EXC-FF
15	15	Bus Passengers	STAGE-FF STAGE-HF
16	16	Senior/Handicap	PASS-SEN-CITZ H-CAPP
17	18	Others	SCH-COMM EXC-FF EXC-HF FLAT-FEE PASS-MISC PASS-FREE
19	19	Auto Full Fare	AUTO-REG
20	20	Auto Commuters	AUTO-COM
21	21	Auto Seniors	VEH-SEN-CITZ
22	22	Oversize Full Fare	O-S-REG
23	23	Oversized Commuter	O-S-COMM
24	24	Motorcycles	M-CYCLE-REG M-CYCLE-COMM
25	25	Trailers	TRAILER
26	26	Buses and Vanpools	VAN STAGE
27	27	Trucks - Regular	TRK-REG
28	28	Trucks - Extended	TRK-EX
29	29	Miscellaneous Veh	VEH-MISC
30	30	Free Vehicles	VEH-FREE
31	32	Total Passengers	Computed
33	34	Total Vehicles	Computed
35	36	Total Full Fare	Computed
37	38	Total Half Fare	Computed
39	40	Total Commuters	Computed
41	42	Walk on Passengers	WALK-ON
43	44	Overload	O-L-AUTO O-L-TRUCKS

4) respond to the prompt [year = 19] with the two-digit year yy;

5) use the arrow keys to choose the appropriate run (the program allows three seconds to change the run, then is starts displaying the missing or duplicate records); and

6) press <esc> to exit the program.

The screen display is shown in Figure 6. For each month, the records missing for the current run are shown. If a day, such as Monday (indicated with a "Mon"), is shown on the left side of the screen, it means that records for the current run are missing or duplicate for all Mondays in the indicated month. If only some Mondays are missing, the specific dates are shown on the right side of the screen.

<u>Traffic Data Summaries</u>

A program called WSFTRAF was written to display traffic data for any month and any route. Five choices are available:

1) average for a particular day,

2) average for Monday through Friday,

3) average for Monday through Thursday,

4) average for Saturdays and Sundays, and

5) individual runs.

To use WSFTRAF, the following steps are required:

1) make sure the files RTxxCYyy.DTA and RTxxCYyy.IND are in the active subdirectory along with WSFTRAF.EXE;

2) type WSFTRAF;

3) respond to the prompt [route =] with the two-digit route number xx;

4) respond to the prompt [year = 19] with the two-digit year yy;

5) choose the desired data summary using the function keys; and

6) change the day, month or run using the instructions shown on the screen.

The summary data will appear as in Figure 7. This particular example shows Monday through Thursday averages for run number two on the Edmonds-Kingston

Run #2

Jan								Sun	1Th			
Feb								Sun	16Mo			
Mar								Sun				
Apr									5Su	19Su	26Su	
May								Sun	25Mo			
Jun									6Sa	7Su	14Su	21Su
Jul									19Su			
Aug												
Sep	Mon	Tue	Wed	Thu	Fri	Sat	Sun					
Oct	Mon	Tue	Wed	Thu	Fri	Sat	Sun					
Nov	Mon	Tue	Wed	Thu	Fri	Sat	Sun					
Dec	Mon	Tue	Wed	Thu	Fri	Sat	Sun					

Figure 6. Missing or Dupl Display

January, 1987 (ave. Mon-Thurs)
Edmonds/Kingston
Run #2 5:51 AM
obs. = 16
Missing or duplicate record(s)

	Next	Previous
Month	F1	Ctrl F1
Run	F2	Ctrl F2

Average Monday through Thursday

PASSENGERS	
Full Fare	16
Half Fare	1
Commuters	67
Bicyclists	0
Bus Passengers	0
Senior/Handicap	0
Others	0

VEHICLES	
Auto Full Fare	29
Auto Commuters	33
Auto Seniors	1
Oversize Full Fare	0
Oversized Commuter	0
Motorcycles	0
Trailers	0
Buses and Vanpools	3
Trucks-Regular	4
Trucks-Extended	3
Miscellaneous Veh	0
Free Vehicles	5

TOTALS	
Total Passengers	84
Total Vehicles	73
Total Full Fare	45
Total Half Fare	2
Total Commuters	10
Walk-on Passengers	45
Overload	0

Figure 7. WSFTRAF Display

route in January 1987. The average departure time was 5:51 AM. The traffic categories are the same as shown in Table 1. The upper middle part of the screen gives instructions on how to change months or runs. The program pauses for about two seconds before computing averages to allow the users to go through the months and runs quickly. Pressing <esc> takes the user back to the main menu. Pressing <esc> again takes the user out of the program.

FARE CHANGE RESPONSE EVALUATION PLAN

The uncertainties inherent in the analysis of aggregate data mean that the fare elasticities derived in previous research are not to be construed as exact measurements but only as indications of order of magnitude. To obtain better estimates of elasticities, another type of data, namely data containing greater price variation, is required. This is the motivation for collecting and analyzing customer-based data, which is the basis for the proposed evaluation method.

The procedure for analyzing probable customer response to fare changes starts with a sample of travelers drawn from the population of travelers who **could** use a ferry for their trips. This sample would include people who use the ferry as well as people who do not use the ferry but could. For each sampled traveler, information is obtained on the cost and time of travel by each possible mode. For example, for a person who could either take the ferry or drive without using a ferry, the data consist of the time and cost required to take the trip if the ferry is used, and the time and cost required to take the trip if the ferry is not used. These times and costs are point-to-point: that is, from the travelers' origin (usually the home) to the destination (say, place of work.) Hence, the cost of the trip by ferry includes the cost of driving from the origin to the ferry, any parking costs, and the fare for the ferry itself. If the vehicle is taken on the ferry, the cost of driving to the destination is also included.

Once these data are collected, models are estimated of traveler's choice of

1) whether to take a trip (e.g., whether to take a shopping trip to the city) and

2) which mode to use for the trip (e.g., whether to take the ferry).

These models give the probability that a traveler will take the ferry as a function of the cost and time of travel on each possible mode. Fare elasticity is determined for each traveler by calculating the impact of a fare change on that traveler's probability of taking the ferry. Average fare elasticities for a group of similar travelers, or for all travelers together, are determined by simply averaging the individual travelers' elasticities.

Traveler-based analysis of this type has several important features. First, the data contain considerable variation in cost even when there is little variation in ferry fares. For a trip using the ferry, the ferry fare is only one part of the total trip cost. Different travelers who use the same ferry will face different trip costs depending on the exact origin and destinations of their trips. This cost variation over individual travelers using the same ferry can be used to estimate travelers' responses to cost changes, including fare changes.

Second, the options of each traveler are explicitly incorporated. For each traveler, data are collected on the time and cost of travel by each possible mode. Travelers who can drive to their destinations without the ferry are thereby distinguished from those who must use the ferry if they are going to take their trip. Since fare elasticities can be expected to be quite different for travelers with other travel options than those without, this explicit accounting for options allows more precise estimates of fare elasticities. Furthermore, since the cost and time of each option is calculated, the ease by which a traveler can switch to another option is factored into the estimates of elasticities.

Third, factors other than cost, such as ferry headways and time spent driving to the ferry, can be incorporated into the analysis. These factors are important to include, even if only fare elasticities are needed, since these factors can determine a traveler's willingness and ability to respond to fare changes.

For all of these reasons, this approach to evaluation of traveler response to fare changes can be expected to provide far greater accuracy in estimates of fare elasticity. Furthermore, because the information is at the level of the individual, responses to different kinds of fare changes (e.g., for different times of day or for different trip purposes) could be determined.

The drawback to this approach is expense. The aggregate data analysis conducted in previous research relied on existing data sources. The cost for collecting those data has already been expended. This approach requires the collection of new data. Because of the complexity of the data required, it will also require that a personal interview be conducted, either by telephone or in person. The cost for data collection depends on the number of interviews desired, and the number of interviews depends on the level of information required. For instance, having good estimates of fare elasticity on a route-by-route basis would require more interviews than if the estimates applied to route groups. The total cost for such a study could easily be in the range of $300,000 to $500,000.

APPENDIX A

DATA INTEGRITY OF WSF DETAIL TRAFFIC STATISTICS FILE

4 May 1988

MEMO

TO: Cy Ulberg

FROM: Rob Fellows

SUBJECT: **Data Integrity of WSF Detail Traffic Statistics File**

I have spent a great deal of time during the last couple of months examining the Ferry System's detail traffic statistics file (file 47) between 1983 and August of 1987 in an attempt to determine what kind of shape it's in. While it's known that errors exist in this file, the extent of the errors has been unknown. This, and the long delay in keying the data result in the file being used less than it might be if the data were cleaner and more timely.

My purpose in writing this memo is to provide a sort of users guide to the data in file 47. Any study which requires time-of-day or directional information will need to rely on this data. File 82, which summarizes the same data contained in File 47, aggregates traffic data by route, day and fare type. While much of the information presented is known, I could find no one person who could provide a clear overview of what is in the file.

I have limited myself to examining file 47 in comparison with file 82. File 82 is used often, so erroneous data is more likely to be spotted. It is more carefully verified during data entry, and data sheets are not transported back and forth to Olympia for processing. For these reasons, there is significantly more confidence in the data contained in file 82.

Appendix B shows the overall results of comparing files 47 and 82 for each route graphically. Appendix C shows the result of a day-by-day comparison of total passengers and vehicles between the two files. Following is a summary of what I've found:

- There are several long gaps in the data, often for periods of a month or more at a time.

- During September of 1985, file 47 data was doubled on every route, in every data field.

- On route 41 (Seattle/Bremerton Express), the file 82 summary total is generated by doubling the count in one direction. Since this route operates during one peak only, and is complementary to route 9 (Seattle/Bremerton local), the summary file total is approximately 16% undercounted and the local is overcounted.

- In some cases, only the total fields contain values, fields for individual fare categories contain zeroes.
- While the traffic statistics data input form includes fields for transit passengers, special events full fare and special events half fare, these amounts are merged with the data in other fields and thus are not stored separately in either file.
- In a majority of cases, the sum of daily records by route in file 47 is equal to the value contained in the summary file. However, a significant number of errors exist due to incorrect totals (where the total field is not equal to the sum of the detail fields), miscoding the route, day, month or year fields, data missing or doubled for an individual day, and other errors.
- There are many cases where more than one record exists for a single ferry run, either by design (because there is more than one gate at a terminal, each tallying statistics separately), or due to miscoding, creating a possibility of error in calculating average volumes or counting trips.

Contents of File 47

File 47 contains data for each run in each direction on each route for each day for each type of fare. Data for each year takes up approximately 35 Megabytes. As of May 1988, years 1986 and 87 are stored as an active ADABAS file. Previous years are stored on tape in the following files:

HWY.XRA.FILE47.HIST.CY1983
HWY.XRA.FILE47.HIST.CY1984
HWY.XRA.FILE47.HIST.CY1985

Each file is in FB record format, record length is 207, block size is 6210. Previous years are also available, but because all data for 1982 is missing, this analysis examines only years 1983 on.

Appendix A shows the file definition for both files 47 and 82. As of November 1, 1985, the following fields were no longer used: School commuter, Full-fare excursion, Half-fare excursion, Full-fare bicycle excursion, and Half-fare bicycle excursion. Three new fields were added to the data entry form: Transit passengers, Special events full-fare, and Special events half-fare. When these are input, the value for Transit passengers is merged with the commuter passengers field, and Special events full-fare and half-fare are merged with the Passenger full-fare and half-fare fields respectively. Separate values for these three input fields are not stored in either file 47 or 82.

In all, seven fields are currently not in use, including the five discontinued in 1985, Reserved filed, and Half-fare stage passengers. (The size of the file could be cut by almost 1/3 if these fields were removed and data fields were shortened from 5 positions to 4.)

Comparing File 47 with File 82

The method for compiling summary data from detail data differs from route to route and has changed over time. In all cases, the summary file is intended to contain the sum of all trips on a given route and day in both directions.

With the exception of routes 1, 4 and 5, each odd numbered route in file 47 contains data for westbound trips, and the even numbered route following contains data for eastbound trips. File 82 contains data for westbound routes only (excepting route 4), but the data is intended to represent 2 way volumes. Following is a summary of how the summary statistics have been compiled. Note that file 82 was not created from data entered in file 47, rather it was created from the same original data entry sheets as was file 47, and so should be comparable.

Routes 1, 4 and 5 (Vashon)

Summary data for routes 1, 4 and 5 in file 82 are obtained by doubling the sum of trips in file 47.

Routes 7 through 16 (Cross-sound)

Before 1/4/86, file 82 totals are obtained by adding the eastbound and westbound totals from file 47. As of 1/4/86, fares were no longer collected for passengers in the eastbound direction with the exception of bicyclists. Vehicle totals are still obtained by adding the eastbound and westbound totals, but passenger totals are obtained by summing the westbound non-bicycle trips and doubling them, then adding the bicycle trips for each direction.

Routes 17 and 18 (Keystone/Port Townsend)

This route pair is summarized in the same way as routes 7-16, except that passenger fares are collected for the even numbered (eastbound) direction, not the odd numbered direction. On 1/20/87, passenger fares were again collected in the westbound direction, so the file 82 summary for passengers is again obtained by adding the westbound and eastbound directions together.

Routes 19 and 20 (Anacortes/Interisland)

Before 6/17/84, summary data for this route was compiled by adding the eastbound and westbound trip totals. After 6/17/84, the summary total is equal to the westbound direction doubled.

File 47 contains no passenger entries in the eastbound direction, and only bicyclists have been counted in the westbound direction. For passengers, file 82 contains the number of westbound bicyclists only, doubled. Aside from bicyclists, no passenger volumes exists for these routes in either file 47 or 82.

Routes 21 through 28 (San Juans)

Before 6/17/84, summary data was compiled by adding the eastbound and westbound trip totals. After that date, File 82 summary data is obtained by doubling the westbound trip totals.

Routes 29 through 31 (Sidney)

On the Sidney routes, the file 82 summary totals are obtained by adding the eastbound and westbound trip totals.

Routes 41 and 42 (Bremerton/Seattle Express)

Summary data for route 41 is created by doubling the westbound passenger volumes. (As noted above, this results in undercounting by approximately 16% on the express, and overcounting on local route 9 after January 1987, because the westbound and eastbound volumes are not balanced after 1986.)

Missing Data

There are a number of days when all data is missing in file 47 for a particular route -- these are noted in Appendix C. Following is a list of dates where all file 47 data is missing for a period of several days:

- Mar. 1-15, 1985: Routes 1-14.
- Mar. 4, 1985: Route 15.
- October 1985: Routes 9 and 10.
- January 1986: Routes 11, 12, 17 and 18.
- February 1986: Routes 9-14.
- Feb. 10-28, 1986: Route 8.
- Feb 1-14 and 18-21, 1986: Routes 19-32.
- March 1986: Routes 4-13.
- Mar 1 - April 16, 1986: Routes 21, 23, 25, and 27.
- Mar. 1-14, 1986: Route 14.
- Mar. 20-31, 1986: Route 1.
- April 1986: Routes 11 and 12. (missing Jan-April.)
- June 1986: Routes 17 and 18.
- March-August 1987: Route 22.

File 82 contains volumes for these routes during the periods listed. For time series analysis, these months must be treated as missing data in file 47.

Doubled Data

During September, 1985, all data is doubled for all routes in file 47. This includes data for individual fare categories. For routes 19-32, all data is also doubled between Aug 16-31 of 1985. The simple solution is to halve all data during this time period.

More than One Data Record per Run

There are a number of cases where file 47 data is miscoded by route or date, leading to duplicate records for the same ferry run. When runs on a high volume route are miscoded as runs on a low volume route, as occurred on route 1 on 4/17, 5/24 and 6/24 of 1985, large errors result. In a few cases, records were simply entered twice, resulting in doubled data (this was not the case during September, 1985).

In some cases data is collected at more than one gate on separate coding sheets, and more than one record per run is appropriate. However, this creates problems for data users. In order to calculate statistics such as average volume by run, it is necessary to count runs, rather than simply counting records. While it is possible to do this for most runs by keeping track of which runs have been tallied and which have not, this is impossible for run 99 trips, because more than one run 99 trip can occur in a single day.

Ideally, file 47 should never contain more than one record per run -- when more than one input sheet exists for a run, they should be merged during the data input process. Duplicate records then would be an indication that a record has been miscoded or entered twice, and the error could be resolved. Averages and counts generated from the data would be more reliable.

Missing Detail

Between April 11 and October 2, 1985, and between January 19 and April 30, 1986, only totals were entered in the eastbound direction of the San Juans routes. Data fields for individual fare categories contain zeroes. While the eastbound runs are not used to compile the summary data in file 82, an analyst looking for directional data on these routes should be aware that this detail data is missing.

Incorrect Totals

In a relatively small number of cases (in 1983, about 400 records), total fields are not equal to the sum of the data fields for individual fare categories. This produced a relatively small number of relatively small errors, but they could be easily eliminated by checking to be sure the totals add up correctly during data entry.

Methodology, Potential Errors and Omissions, and Caveats

In order to examine file 47, I downloaded by file 47 and file 82 to a microcomputer, where I conducted analysis using Turbo Pascal. Errors could have occurred in the process of transmitting the data from the mainframe to the microcomputer, or in condensing it into a fewer number of data fields. A small number of records were rejected in the process because fields contained non-numeric characters or were the wrong record length (probably due to transmission errors).

The data in file 47 takes up about 30-35 megabytes of data per year, for a total of about 140 megabytes for the five year period. In order to analyze this amount of data, I had to significantly condense it (from 207 to 44 bytes per record), and I could only examine large trends. For example, I only identified when data was missing on a route for an entire day -- individual runs missing were not identified. I only compared data for the totals fields; I've made no attempt to determine whether files 47 and 82 are comparable for individual fare category data fields.

As I mentioned at the outset, I have restricted myself to examining the internal consistency of file 47 and comparing it with file 82. I have made no effort to determine if the original data collected is appropriate or correct.

APPENDIX B

TRAVELER-BASED ANALYSIS OF FERRY FARE ELASTICITIES

Traveler-Based Analysis of Ferry Fare Elasticities

by

Kenneth E. Train

At the meeting on June 1, 1989, staff from Washington State DOT, TRAC, and the Ferry System discussed and agreed to pursue a traveler-based approach to estimating fare elasticities.* The purpose of the present memo is to describe this approach briefly and delineate the reasons for pursuing it. As such, this memo serves as a recap of the discussion on June 1 and a focus for future discussions.

Previous estimates of fare elasticities have been based on aggregate data, namely, data on the total number of passengers and vehicles on each route during each time period. The estimates based on these data have been imprecise because of fundamental limitations in the data. The two most important limitations are:

(i) *Insufficient variation in fare data*. Only a few general fare changes have occurred over the years for which comparable ridership data are available. Furthermore, the relations among fares on different routes (that is, the fare structure) have remained essentially the same over time. With little variation in fares, it is not possible to estimate meaningful fare elasticities. This problem is exacerbated by the fact that nominal fares need to be expressed in real terms. When the time series of nominal fares in divided by the consumer price index (CPI), the variation in real fares reflects variations in the CPI more than in the fares themselves. Since it is doubtful that travelers accurately adjust fares each week for changes in the CPI, the conversion of nominal fares into real fares (which is necessary in order to meaningfully account for inflation) adds "noise" to the data, decreasing the accuracy of the elasticity estimates even further.

(ii) *Aggregation over Routes and Days*. Patronage figures are aggregated over routes within a group and over days within a quarter. Given the lack of meaningful fare variation over routes within a group and days within a quarter, this aggregation is necessary. However, it results in an average fare (that is, the average over all routes in the group and all days in the quarter) that does not correspond to the fare that any traveler actually faces. Since the fares used in estimation of elasticities is not the same as the fares that travelers actually face, the estimation of elasticities cannot be expected to be very precise.

*The consensus of the group was to consider using a traveler-based approach.

These problems evidence themselves in large standard errors and in estimates that change considerably when the model specification is changed only slightly. It is important to note that these problems are endemic to the data, in that any analysis based on these aggregate data can be expected to encounter the same difficulties. To obtain better estimates on elasticities, another type of data, namely data containing greater price variation, is required. This is the motivation for collecting and analyzing customer-based data.

Most modern studies of travel behavior are based on data collected at the level of the individual traveler. For the analysis of ferry system patronage, the procedure can be summarized as follows. A sample of travelers is drawn from the population of travelers that could use a ferry for their trip. (This sample includes people who use the ferry as well as people who do not use the ferry but could.) For each sampled traveler, information is obtained on the cost and time of travel by each possible mode. For example, for a person who could either take the ferry or drive without using a ferry, the data consist of the time and cost required to take the trip if the ferry is used, and the time and cost required to take the trip if the ferry is not used. These times and costs are point-to-point: that is, from the travelers' origin (usually his home) to the destination (say, place of work.) Hence, the cost of the trip by ferry includes the cost of driving from the origin to the ferry, any parking costs, and the fare for the ferry itself.[1]

Once these data are collected, models are estimated of traveler's choice of whether to take a trip (e.g., whether to take a shopping trip to the city) and which mode to use for the trip (e.g., whether or not to take the ferry.)[2] These models give the probability that a traveler will take the ferry as a function of the cost and time of travel on each possible mode. Fare elasticity is determined for each traveler by calculating the impact of a fare change on that traveler's probability of taking the ferry. Average fare elasticities for a group of similar

[1] This simple example assumes that the person does not take the car on the ferry. If the vehicle is taken, the cost of driving to the destination is also included.

[2] For work trips, it is most reasonable to assume that the worker will necessarily take the trip, such that the only choice is whether or not take the ferry (and perhaps whether to ride the ferry as a passenger or take his car on the ferry.)

travelers, or for all travelers together, or determined by simply averaging the individual travelers' elasticities.

Traveler-based analysis of this type has several important features. First, the data contain considerable variation in cost even when the there is little variation in ferry fares. For a trip using the the ferry, the ferry fare is only one part of the total trip cost. Different travelers who use the same ferry will face different trip costs depending on the exact origin and destinations of their trips. This cost variation over individual travelers using the same ferry can be used to estimate travelers' responses to cost changes, including fare changes.

Second, the options of each traveler are explicitly incorporated. For each traveler, data are collected on the time and cost of travel by each possible mode. Travelers that can drive to their destination without the ferry are thereby distinguished from those who must use the ferry if they are going to take their trip. Since fare elasticities can be expected to be quite different for travelers with other travel options that those without, this explicit accounting for options allows more precise estimates of fare elasticities. Furthermore, since the cost and time of each option is calculated, the ease by which a traveler can switch to another option is factored into the estimates of elasticities.

Third, factors other than cost, such as ferry headways and time spent driving to the ferry, can be incorporated into the analysis. These factors are important to include even if only fare elasticities are needed, since these factors can determine a traveler's willingness and ability to respond to fare changes.

Because of these features, analysis of traveler-level data can be expected to provide far greater precision in the estimation of ferry fare elasticities.

Review of *Ferry Fare Elasticity Study*

by

Kenneth E. Train

This study has been performed insightfully and competently. Its procedures are appropriate for the data that are available, are statistically rigorous, and are generally accepted and widely used in the demand analysis profession. The results are generally plausible, both theoretically and with respect to the particular data and situation being examined. Because of data limitations, the estimated elasticities are not extremely stable, in that fairly small changes in specification result in somewhat different elasticities. The estimates should therefore be considered indications of the order of magnitude of the underlying elasticities, rather than precise measures.

From a practical perspective, there are two main policy implications of the study. First, elasticities are fairly high, even exceeding one for some market segments, implying that raising fares will result in a substantial loss in patronage, and, for segments with elasticities over one, a loss in revenue. Second, elasticities are substantially different for different routes and different types of travelers, such that there are benefits to be gained from changing the fare structure to reflect these differing elasticities. These two implications are plausible and, for reasons I describe below, even reasonable. Furthermore, these implications are not very susceptible to the instability of the particular estimates, in that fairly large changes in the exact estimates can be obtained and these implications still be valid. Perhaps the most important function of the study is to focus attention on these two concepts. The estimated elasticities should probably be taken as first guesses only (as I understand they were intended to be, given the way the results are stated in the report), and more extensive investigation should be pursued to obtain a better idea of the exact elasticities.

Both of the policy implications are justifiable on theoretical grounds. The second of these implications, regarding differing elasticities, is undoubtedly correct. In fact, it could essentially be known to be correct without any data analysis. Different travelers have different purposes in traveling, different options regarding alternative modes, and so on. These differences necessarily result in different elasticities. The economic theory of pricing indicates that, for a supplier with high fixed costs, consumer surplus can be increased without reducing the revenues of the supplier by charging a price in each market segment that reflects the elasticity in that market.[1] Since the current fare structure was determined long ago, prior to the establishment of current travel and land use patterns, it is doubtful that it reflects the pattern of elasticities that exists today. The policy implication of the study that the fare structure should be reconsidered in light of differing elasticities for different segments is therefore important and valid, independent of whether the particular estimates are accurate for each market segment.

The first policy implication of the study, that elasticities are fairly high and in some segments exceed one, is an empirical issue and depends on the particular estimates obtained in the study. However, there is a reason for believing this result independent of any limitation in the data and study. Elasticities are not fixed but change with price. An agency that is maximizing revenues will raise price until the elasticity of its demand is one. Consequently, if a researcher is estimating elasticity for a revenue-maximizing agency, then one would expect that the researcher would obtain an estimate of one. The question, therefore, is whether the Washington State Ferries are operated in a way that is similar to revenue maximization. Transportation agencies are generally required to meet a portion of their costs through farebox revenues, and many (if not most) find it difficult to do so. This difficulty means that the agency must maximize revenues (or approximately so) in order to meet its farebox requirements. If fares for the Washington State Ferries are set with revenue needs in mind, there will be a tendency to move to the point where elasticity is one.

1 Consumer surplus is maximized by setting prices higher in markets with lower elasticity, which is the well known inverse-elasticity, or Ramsey, rule. It is not necessarily the case that prices should be set in accordance with this rule, since goals such as equity and maintaining continuity with the past are important in addition to maximizing consumer surplus. The point is simply that the relation between fares and elasticities has an impact on consumer surplus, and this impact should enter in the determination of the fare structure.

I have some comments about the specific empirical analysis that was performed. These comments are not meant to detract from the value of the study, which, as I stated above, was well performed. Rather, my comments are of two kinds, namely, those that identify the types and direction of bias that might arise due to analytical and data limitations that are inherent in the study, and suggestions for directions that further work in this area might take to overcome some of the limitations of the current study.

1. As stated on page 32 of the report, most of the fare increases occurred shortly before the summer months, and the patronage data were seasonally adjusted to correct for the fact that ridership is generally highest in the summer such that the effect of fare increases is masked by the natural summer surge. It is clearly important to separate these two effects. However, the way in which the seasonal adjustment was performed tends to bias the estimated fare elasticities downward. This implies that the true elasticities could be expected to be even higher than those estimated.

The seasonal adjustment expresses ridership in each quarter as a deviation from the mean ridership for that quarter, with the mean taken over all years. This procedure incorporates part of the price response into the seasonal adjustment. The mean ridership in, say, the summer is calculated from the ridership figures for each of the summers; since these ridership figures reflect price responses, the mean also reflects price response. The deviations are therefore calculated from a base (mean) that already has price response incorporated, such that some of the price response has been removed from the deviations.

For illustration, consider the following simplified situation. Suppose there were only two years' worth of data and only two seasons, called summer and winter, and that the seasons are denoted such that winters precede summers. Without a fare increase ridership would be 6000 in each winter and 10,000 in each summer. However, suppose there was a fare increase immediately prior to the second summer (i.e., after the second winter) from $1.00 to $1.20, and that the response to this fare increase was a reduction in demand by 20%. The observed ridership figures are therefore 10,000 for the first summer and 8000 for the second, and 6000 for each winter. These figures are seasonally adjusted: the mean summer ridership is 9000 and the deviations are +1000 for the first summer and -1000 for the second; the deviations are zero for each of the winters. The data matrix is the following (I am not taking logs and differences to make the example easier to see; the same bias occurs with logs and differences.)

Observation	Ridership expressed as seasonal deviations	Fare
First winter	0	$1.00
First summer	+1000	$1.00
Second winter	0	$1.00
Second summer	-1000	$1.20

OLS applied to these data will compare the mean ridership before the price change (0+1000+0 divided by 3, or 333) with ridership after the price change (-1000) and infer that the difference in these amounts is the change in ridership attributable to the fare increase. Since the change from +333 to -1000 is -1333, OLS will estimate that the twenty cent fare increase resulted in a decrease in demand of 1333, which is less than the true decrease of 2000.

What is wanted is for each season's ridership to be deviated from the value that would occur due to seasonal effects only. For example, since we know in our example that summer ridership would be 10,000 without any fare changes, we would deviate each summers' ridership from this figure of 10,000. This would give ridership of 0 for the first summer and -2000 for the second summer. The data matrix would be:

First winter	0	$1.00
First summer	0	$1.00
Second winter	0	$1.00
Second summer	-2000	$1.20

With these data, OLS would compare average ridership before the fare increase (0) with that after (2000) and estimate the correct impact of the fare increase.

In this example, the correct seasonal adjustment is known. Unfortunately, in the real world, the level of demand that would occur without changes in fares is not known, such that the seasonal adjustment that we would like to apply is not possible. However, even though we cannot apply the correct adjustment, we can known the method applied in the study results in a smaller than actual estimated price response, giving greater credence to the conclusion that price elasticities are fairly high.

2. There is very little meaningful variation in the fare data. Nominal fares, which travelers observe, changed only a few times over the relevant period; real fares (i.e., deflated by CPI) are, except for 1977-9, are essential constant. This lack of variation poses a serious limitation to any attempt to use these time-series data for estimating price responses. The instability of the estimated elasticities is primarily due to this problem.

The problem of lack of fare variation is accentuated by the fact that travelers do respond to fare changes slowly over time instead of immediately. To account for delayed response, lagged fare variables are included in the model. However, lagged fares are correlated with current fares, such that there is very little independent variation with which to estimate coefficients. This problem is evident in the model results. For example, in the model for total ridership on the Vashon routes, the coefficients of the variables for current fares, one-period-lagged fares, and two-period-lagged fares are negative, positive, and negative, respectively. This implausible pattern of response is due to the fact that the model is attempting to obtain three pieces of information (namely, the time pattern of response) from a variable (fares) that does not even have sufficient meaningful variation to provide one reliable piece of information (total response.) The problem here is not with the model: since traveler's responses do occur over time, lagged fare variables are needed to represent this response. The problem is a lack of variation in the data. Future work should concentrate on obtaining data with greater variation, as described in point 5 below.

It is interesting to note that problems with the fare data might be inducing the estimation procedure to systematically *underestimate* the response to fare changes, further supporting the study's conclusion that elasticities are high. The fares that enter the analysis are fairly far removed from the fares that travelers observe. If travelers do not convert fares to real dollars by consistently knowing the CPI, then deflation by the CPI adds noise to the fare data in addition to its intended purpose of partially correcting for the effects of inflation. Furthermore, the fare for one route is used for all routes within a group, adding noise for travelers who do not take the designed route. It is a standard econometric result that when noise is added to a variable, the estimate of its impact is biased toward zero. Therefore, it is probably reasonable to expect that the elasticity estimates in the study have a tendency to be lower than their true values, supporting the conclusion that elasticities are high.

3. As stated above, the estimated elasticities are not particularly stable with respect to plausible changes in specification, due at least partially to the lack of variation in fare data. For example, the model for total ridership on Vashon routes includes two lagged fare variables that obtain implausible coefficients: the first lag enters positively while the second enters negatively. When these two variables are dropped from the model, the estimated elasticity changes by 30%. This comparison is not meant to imply that the model without the lags is better: there are theoretical reasons for including lags (i.e., people's responses are often delayed) and the estimated coefficients of the two lags, while implausible individually, might accurately reflect the total delayed response (that is, the sum of the coefficients might be accurate even if the individual coefficients are not.) The point is simply that different specifications, among which there is very little solid reason for choosing, result in fairly different estimated elasticities.

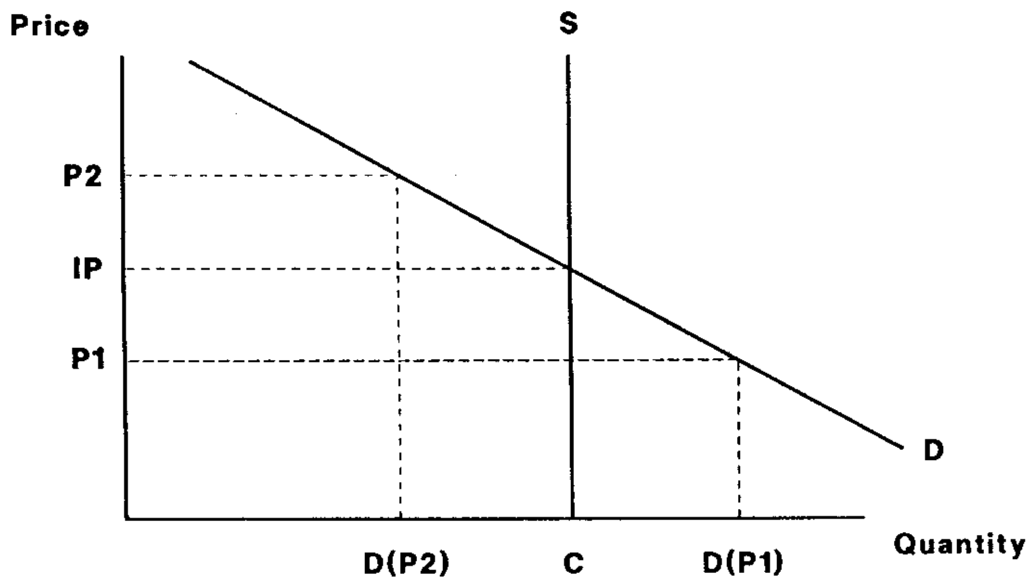
Other examples provide similar conclusions. If the moving average term is omitted from the model for total ridership on Vashon routes, then the estimated elasticity changes by over 10%. When the second-period lagged fare variable is omitted but the first-period lag included, estimated elasticity drops by 50%. For total ridership on cross-sound routes, the elasticity estimate drops by about 20% when a lagged fare variable is added. These are fairly large changes since there is little reason for choosing one specification over the other *a priori*. Furthermore, using goodness-of-fit criteria (such as R-squared and t-statistics) to choose among the specifications is hazardous in this situation since there is little meaningful variation in the relevant variables such that slight outliers or other anomalies can critically affect the results. Furthermore, performing specification tests on the same data set on which the parameters of the final specification are estimated violates the basic restrictions of classical statistics.

As stated above, the instability of the estimates is primarily due to the insufficient variation in fares. There is nothing the researcher can do statistically about this problem. The only solution is to collect data with greater variation. A method for doing so is described in point 5 below.

4. In estimating demand, it is important to identify what is happening on the supply side. If price is determined simultaneously by demand and supply, then attempts to estimate demand without explicitly accounting for supply will lead to bias. In the study, price is taken to be exogenous. This is appropriate and unbiased if supply is perfectly elastic at the price that is set. For the ferries, this assumption is met as long as everyone who wishes to take a ferry can do so (which is usually the case). However, if there is a constraint on supply such that travelers who have arrived in time for a ferry are not able to board due to excess demand, then the model will underestimate the demand response to price changes.

The accompanying figure illustrates the situation. Consider one ferry whose capacity is C. Suppose the fare is originally P1. At this price, more travelers arrive for the ferry than can be accommodated; while demand is actually D(P1), the observed number of riders is only C. Now

Ferry Demand and Supply with Supply Constraint



suppose the fare is raised to P2. At this price, demand is D(P2) which is less than the capacity such that all travelers who arrive for the ferry can be boarded. If the supply constraint is not considered, then the response to the price change is estimated to be the difference between C and D(P2), which is smaller than the true demand response of D(P1) minus D(P2).

In reality the situation is more complex than that depicted in the figure. Travelers who do not board one ferry can wait for another, and travelers who need to catch the ferry at a particular time can arrive earlier to be sure to be sufficiently close in line to be able to board. However, these adaptations impose a cost on the traveler, such that the cost of traveling by ferry is actually greater than the fare, in which case the fare is not the appropriate variable to use in analysis. The point is simply that the supply mechanism must be examined in order to correctly specify and interpret the demand model.

It is interesting to note that problems of supply constraints are not necessarily observable. For example, in the situation illustrated in the figure, the implicit cost of riding the ferry would rise to IP where demand equals to capacity of the ferry.[2] The ferry operators would be able to fit every traveler who arrives for the ferry, and would not turn away anyone. Consequently, it would *appear* as if there is no excess demand at the given fare when in fact there is.

5. Many of the limitations encountered in the current study can be overcome by using cross-sectional data on individual commuters, especially if used in conjunction with the time series data. The idea is to collect data on the trip-making behavior of a sample travelers. The data would contain information on the origin, destination and mode of trips made by the traveler, and the cost and time for taking those trips by each available mode. These data would be used to estimate models of choice of whether to take the ferry, separately for work and nonwork trips. Explanatory variables in these models contain the time

---

2 This price rise would take to form of the extra time spent by people arriving early to be able to board. People not willing to incur this time cost would choose not to ride the ferry, since they would know that if they arrived when they wanted to, they would be too far back in line to be able to board.

and cost of the trip by ferry and other alternative modes, if any, and can be used to estimate demand elasticities.

There are several advantages to this approach. (i) The data contain substantial variation in costs, such that demand elasticities can be estimated more precisely. The cost for any one traveler is the full cost of that traveler's trip by ferry, including the cost of driving to the ferry, riding the ferry, parking and any other expenses. These costs vary over travelers, depending on the locations of their homes and work, etc., even though the ferry fares are the same for all travelers on a ferry route. (ii) Factors other than cost, such a waiting time and ferry headways, can be included in the analysis. These factors are important to incorporate even if only price elasticities are needed, since these factors can determine a travelers willingness and ability to respond to price. (iii) Information about the alternatives available to each traveler (e.g., whether the traveler has the option to drive instead of taking the ferry, and if so what is the cost and time of driving) can be explicitly incorporated into the models. (iv) Models estimated on these data will allow separate elasticities for each traveler, such that, for example, travelers with different options and different travel costs will have different elasticities. By aggregating over travelers of a particular type, elasticities for market segments can be obtained.

For these and other reasons, most modern travel demand analysis is performed on cross-sectional data on individual travelers. Unfortunately, the cost, particularly for data collection, is substantial -- considerably more than that spent for the time series analysis in this study.

In summary, the study provides valuable information and draws two important implications that should definitely be considered in determining new fares. The study also identifies the need for further investigation, using more extensive data, so as to obtain more reliable estimates of fare elasticities.