usually depend on the reliability of the connection, which would be enhanced by increasing schedule reliability. Hence, the increases in user satisfaction caused by implementing timed transfers and increasing schedule reliability may exceed the sum of the benefits derived from using those two components individually. Furthermore, some options have more widespread applicability than others; through-routing, for instance, can probably be implemented on a wider range of property types than pulse scheduling, although pulse scheduling has more far-reaching effects. Each operator must evaluate the service, cost, and demand conditions on the property and the consequences of alternative policies to determine which actions would be the most productive.

ACKNOWLEDGMENT

The material presented in this paper is drawn from the results of a recently completed study funded by the Transportation Systems Center (TSC) under the Service and Methods Demonstration program, UMTA. We would like to thank Robert Casey of TSC and Thomas Parody of Charles River Associates for their many helpful comments and observations. Also, the contributions of the transit professionals who participated in this study are gratefully acknowledged. However, the opinions and conclusions expressed or implied in this paper are ours, and we are solely responsible for any errors of content or ommission that remain.

REFERENCES

1. Charles River Associates. State of the Art of Current Practices for Transit Transfers. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, June 1980.
2. Charles River Associates. Transit Operator Guidelines for Transfer Policy Design. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, June 1980.
3. M. Nelson, D. Brand, and M. Mandel. Use and Consequences of Timed Transfers on U.S. Transit Properties. TRB, Transportation Research Record 798, 1981, pp. 50-55.

Publication of this paper sponsored by Committee on Bus Transit Systems.

Short-Term Ridership-Projection Model

CY ULBERG

A statistical model has been developed for use by a transit agency in making short-term forecasts of transit ridership. These factors have been used successfully to plan service changes and to forecast revenues. A by-product of the use of the model is an increasing understanding by staff members of determinants of ridership changes and a corresponding reduction in the emphasis on ridership as a performance indicator by the agency. The model uses a combination of multiple-regression and time-series analyses to produce monthly projections of ridership. The variables included in the model were chosen for simplicity, ease of collection, and explanatory power. The validity and reliability of the model are quite strong, given its simplicity. During a two-year validation period, the average monthly error was 2 percent. Errors in annual totals were 0.9 and 1.7 percent, respectively. One objective in the development of the model was to make it a useful tool for planners and managers within the agency. A monthly report has been developed that has become a part of the decisionmaking process in the agency. Even though experience with a model has been limited, it has been demonstrated that a transit agency can make use of a relatively sophisticated (although simple) statistical technique to develop ridership forecasts.

In the past several decades, transit ridership has varied dramatically. Long-term trends have been influenced by phenomena such as the rising popularity of the automobile, world wars, and population shifts from farms into cities and suburbs. In contrast to these long-term trends, short-term ridership gains and losses occur due to more rapidly varying factors such as seasonal effects, service levels and quality, fares, gasoline prices and supply, parking rates, employment, and population. This paper describes one transit agency's experience with producing useful short-term forecasts.

Transit agencies use a variety of nonstatistical and quasi-statistical methods to produce forecasts of ridership. Generally, these methods use interpretations of past trends modified by management objectives for increasing ridership. Most agencies try to predict the impact of fare changes and service changes on ridership. In the Seattle metropolitan area, Metro Transit traditionally has projected ridership by using a modified Delphi technique. Objectives for productivity (passengers per hour) were set by using qualitative assessments of the environment, particularly the impact of fare and service changes. Service hours were projected by using budget constraints and perceived ridership demand. Total ridership projections were determined by multiplying productivity and service hours.

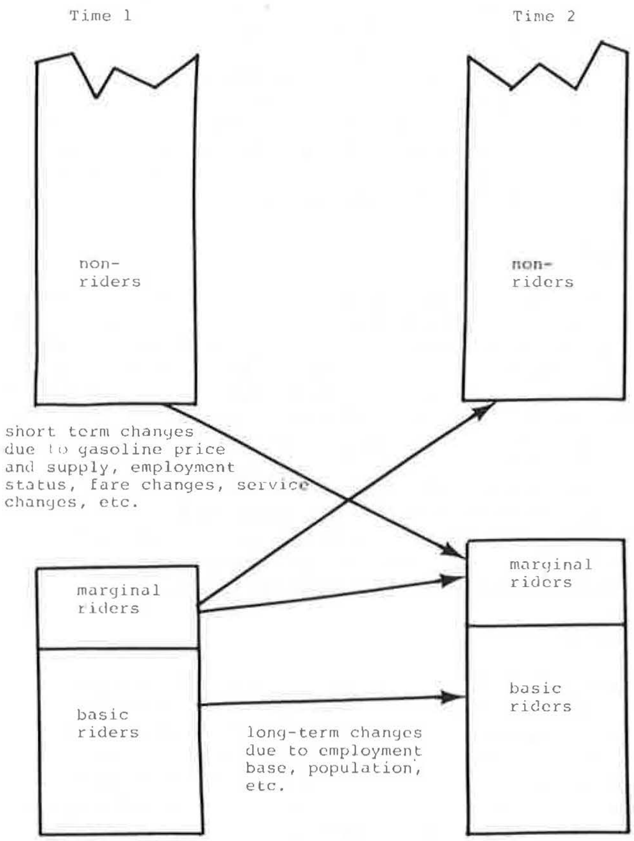
When ridership changes were relatively stable (such as between 1975 and 1979), these methods worked fairly well. However, in 1980 ridership trends changed abruptly. A gasoline crisis and rising employment were followed by a drop in gasoline price and declining employment trends. A major fare increase was implemented. Rapid increases in ridership changed to a leveling-off period. The extent of the change was unanticipated and resulted in major adjustments in service planning and budgeting.

In order to anticipate similar short-term changes in the future, Seattle's Metro Transit has developed a short-term ridership-projection model. It has been used during the past year to assist in the preparation of revenue projections and in planning service changes. It has also been used to anticipate the impact of a fare increase implemented in February 1982. Because the model uses variables extraneous to Metro's control, such as gasoline price and supply and employment, it has helped develop a new perspective on the use of ridership data for evaluating the effectiveness of the transit agency and its components.

BASIC STRUCTURE OF MODEL

A major objective in the development of the model

Figure 1. Model of behavioral assumptions that explain ridership.



was simplicity. Only five variables were used in projecting ridership:

1. Gasoline price,
2. Gasoline supply,
3. Service changes,
4. Fare changes, and
5. Employment.

Several other variables were considered for use but were rejected either because the data concerning them are inadequate or because they do not add significantly to the predictive ability of the model. Discarded variables include quality of service, population, employment in the central business district (CBD), parking prices, fuel efficiency of cars, and data disaggregated by time of day, route, or region served.

The model basically uses a multiple-regression approach. However, the dependent variable is the monthly change in ridership rather than the nominal value for ridership. Thus, the model is a time-series analysis. In addition, model development included investigation of the lagged effects of the independent variables. The assumption is that people do not respond immediately to changes in the economic environment.

The development of the model was guided by the constraint that data collection and analysis should be simple and straightforward. Statistical sophistication and rigor were sacrificed sometimes for simplicity of data collection and explanation of the model. A balance between rigor and simplicity was the goal. The development of the model was based on assumptions about aggregate responses to changes in the environment. The underlying behavioral model assumes that there are two types of transit riders: basic and marginal. Basic riders consist of transit dependents and those who ride for other reasons that do not fluctuate in the short run. This group changes with population and employment base. Marginal riders, or more aptly rides, are influenced by short-term phenomena such as gasoline price, gasoline supply, fares, employment status, and other economic factors. Figure 1 illustrates this model.

By using this model for underlying aggregate behavior, one would expect that the form of a model varies from agency to agency, depending to a great extent on the ratio of basic to marginal riders. Where basic riders compose the bulk of transit patronage, ridership would vary slowly except in response to changes in geographic coverage of the routes. Where marginal riders predominate, large fluctuations would be expected in response to changes in economic variables.

REGRESSION VARIABLES

In the process of choosing variables for the regression, several options were considered, including seasonal adjustments, lag times, and sources of data. This section details these choices and the development of data for the regression.

Ridership

Ridership is estimated monthly at Metro by using revenue data, including farebox collections, pass sales, and other estimates for special types of services. Periodic surveys are used to determine the proportion of farebox passengers paying special fares, transferring, or taking more than one zone trip. Recently, the introduction of automatic passenger counters on some of the buses has afforded the possibility to conduct reliability checks on ridership estimates. These tests show that estimates based on revenue data are fairly close to estimates based on actual counts.

Metro historically has published ridership data unadjusted for seasonal or calendar effects. Variations in average weekly ridership occur for different parts of the year (on the order of 10 percent). The number of working days, compared with holidays and weekends, can have a great effect on monthly totals (March generally has 15 percent more riders than February simply due to more weekdays and fewer holidays). In order to eliminate confounding variables in the regression analysis, ridership data are converted to seasonally adjusted average weekly ridership for each month.

Before the seasonal adjustment is made, an adjustment must be made for school services. Since fall of 1979, Metro has provided special service for school children, who account for about 3.5 percent of the total ridership. Because this service was not provided over the entire period of the data base, it is excluded from the historical data, and forecasts are adjusted with a separate prediction of this school service.

The first step in calculating the seasonally adjusted average weekly ridership is to add daily ridership figures together to produce weekly figures for each week of the month. The effect of holidays is eliminated by normalizing. For instance, if a holiday occurs on a Friday, the average Friday ridership for the rest of the month is computed and substituted for that day's data. Each week is standardized (i.e., the first week ends on January 7, the second on January 14, etc.).

In the second step, these weekly figures are converted to monthly averages. Each month is standardized (January has four weeks, February has four weeks, March has five weeks, etc.) so that there is an integral number of weeks in each month. The month's average is simply the total for that month divided by the number of weeks in the month.

The third step is to apply the seasonal variations to the months. Multiplying each month's average weekly ridership by a monthly factor gives the seasonally adjusted average weekly ridership used as the basis for the independent variable in the regression. Monthly factors are computed by averaging the deviation of actual ridership figures each month from the ridership figures predicted by the regression equation.

The fourth step is to compute the percentage change in seasonally adjusted average weekly ridership for each month. Ridership tends to have a wide variation from month to month. Therefore, a smoothing technique was added to produce stability in the data. The monthly change is taken to be the change over the average of the previous three months' figures. This change is converted to the equivalent of one month's change.

Gasoline Price

The variable that represents the price of gasoline is based on the price of no-lead regular gasoline. The no-lead price is used because it is the largest volume type of gasoline consumed. The data come from the Lundberg Survey, Inc., Recap of Wholesale Prices--Seattle. The survey shows the average price from each major oil company and the independents. The model uses the average of those figures.

The gasoline price is divided by the consumer price index for all urban residents of the Seattle-Everett standard metropolitan statistical area (SMSA). Using the real gasoline price was found to improve the multiple correlation coefficient by 10 percent compared with using the nominal gasoline price. Because it was expected that changes in the gasoline price would not affect ridership behavior immediately, a study was made to determine the lag time that best predicted ridership changes. The lag time that resulted in the lowest average residual was two months. The two-month change was converted to the equivalent monthly change by taking compounding into account.

Gasoline Supply

Gasoline price alone does not explain all changes in ridership, particularly during the gasoline crises of 1974 and 1979. The supply of gasoline was shown to be another separate important independent variable. The 1974 crisis can be characterized as one that had a moderately high growth in real gasoline price coupled with an extremely short supply of gasoline, while the 1979 crisis had an extreme increase in the price of gasoline and a moderate shortage in supply.

An attempt was made to use percentage shortfall in the state's allocation of gasoline as the variable to represent the supply problem. However, this variable was confounded so much by the cutback in use in response to the supply problem that it added little to the explanatory power of the model. The real problem that made people choose the bus was the inconvenience, as they saw it, in obtaining gasoline. A good quantifiable measure of this would be the average length of time waiting in line to get gasoline. Unfortunately, such data do not exist. As a surrogate for this information, newspaper articles written during the crisis were used as a basis for estimation. An energy specialist and I independently developed summaries of the problems on a month-to-month basis. These summaries were used to rate each month on a scale of 0-10 for severity of difficulty in obtaining gasoline. Adding these data to the regression served to reduce residuals during the 1974 and 1979 gasoline crises.

No gasoline shortfall is expected in the near future, so estimates of the variable that represent gasoline supply are not needed. If, however, one wished to assess the potential effect of a gasoline crisis, data from the 1974 crisis or the 1979 crisis could be introduced to represent the severity of a crisis at either of those two levels.

Fares

Four fare increases have occurred in the last eight years. In January 1977, fares were raised 10.5 percent on average. In January 1979, they went up 19 percent; in May 1980, 31 percent; and in February 1982, 10 percent. A trial regression was performed that included the effects of inflation on the fare price, but this added no explanatory power to the model; thus, in the interest of simplicity, it is not included in the data.

An investigation of the effect of lag in the fare variable revealed a relation similar to that with gasoline prices. The best predictor was the average monthly change over the previous two months.

Service

Each month the hours of service are computed by multiplying the number of hours of service on weekdays, Saturdays, and Sundays by the appropriate number of days in a standard month (i.e., 21.1 weekdays, 4.3 Saturdays, and 5.0 Sundays and holidays). The standard month is used because the dependent variable is weekly ridership adjusted for seasonal and calendar variations. Again, the effects of lag on the service-hours variable were investigated. No lag was found that resulted in a significant relation between service hours and ridership, except when service hours were lagged after ridership changes.

Employment

Raw employment estimates were computed monthly by the Research and Statistics Branch of the Washington State Employment Security Department. The model uses the change in employment for nonagricultural workers for the entire King County area, the service area for Metro Transit. Employment data were seasonally adjusted and the best lag was determined to be three months.

Calendar Variations

The level of ridership is highly influenced by the number of working days, Saturdays, Sundays, and holidays in a month. In transit agencies where ridership is recorded on a weekly or four-week basis, the only consideration is the number of holidays. However, at Metro, ridership data have traditionally been reported by calendar month, so an adjustment is necessary.

The dependent variable in the regression equation is a monthly percentage change in the seasonally adjusted average weekly ridership. By applying this percentage to the average of the previous three months' adjusted ridership, the average weekly ridership (adjusted for season) can be projected. The seasonal factor is applied to this figure to give average weekly ridership during a month. The next step in making the actual projection is to take into account the number of weekdays, Saturdays, Sundays, and holidays.

In order to compute the effect of the composition of a month, the total monthly raw ridership was divided by the average weekly ridership developed from revenue-based ridership estimates. This factor is generally slightly above 4. For each of the months during which ridership data exist, the number of weekdays, Saturdays, Sundays, and holidays was determined. By using multiple regression, a coefficient was determined for each type of day in a month where the dependent variable is the factor described above. Based on data through February 1982, the coefficients are as follows:

Type of Day	Coefficient
Weekday	0.1728
Saturday	0.0828
Sunday	0.0540
One-day holiday	0.0376
Two-day holiday	0.1773

Two-day holidays occur when a holiday falls on a Tuesday or Thursday, making either the Monday or Friday into a vacation day for many people. By applying these coefficients to the number of weekdays, Saturdays, and Sundays in future months, an estimate can be made of the calendar factor that must be applied to the weekly average ridership figure to compute the raw monthly ridership forecast.

REGRESSION ANALYSIS

The data were reduced to one dependent variable (the change in seasonally adjusted average weekly ridership) and five independent variables--change in real gasoline price, change in average fare, change in monthly service hours, change in perceived waiting time for gasoline, and change in employment. All of the variables used in the regression were converted to monthly percentage changes except the variable that represents gasoline supply. This allows regression coefficients to be interpreted as elasticities. It also allows a comparison of the strength of influence of each of the independent variables on changes in ridership. Each month, as new information is added to the data base, a new regression is performed.

The table below gives the regression coefficients for the model as of February 1982 (note that R^2 = 0.694, F = 45.01 (df = 99), and Durbin-Watson statistic = 2.07):

Variable	Coefficient	t-Value
Gasoline price	0.29	5.44
Gasoline supply	1.13	11.13
Fare	-0.14	-4.45
Service hours	-0.01	-0.16
Employment	0.80	4.01
Constant	0.38	3.29

The coefficients for all of the variables are significantly different from zero except for service hours. With only 5 variables, the regression explains about 70 percent of the variance in monthly changes in ridership. If nominal values rather than changes were used in the regression, R^2 would be much higher but the variables would be serially correlated. By using change rates rather than nominal values, the Durbin-Watson statistic is well within allowable limits.

The lack of relation between service hours and ridership in the regression deserves special comment. At least three factors explain this phenomenon. First, total service hours, as an aggregate, is too gross a measure of quantity of service. Some service-hour additions immediately attract new ridership while others may take a couple of years. If the data were available, the model could be improved by disaggregating service hours. Second, in a service area like Metro's, marginal riders predominate. Thus, variations in economic factors obscure the effect of variations in service. Third, new service hours have been implemented in response to changes in demand rather than preceding demand. Historically, changes in service can be predicted by changes in ridership rather than the other way around.

This regression is based on eight years of historical data. The more data the regression is based on, the more confident one can be in the regression coefficients. However, there is no reason to rule out the possibility that the relations between the independent variables and ridership change over time.

In order to test for this possibility, regressions were performed on subsets of the data for the time intervals shown in Table 1. The regression coefficients are fairly constant except for employment. An explanation for this phenomenon is in order.

In the early years of Metro, the relation between employment levels and ridership was low. One interpretation of this finding is that there was a large untapped market of commuters in the early years, so that the level of employment had little effect on ridership. In the past four years, however, as the system has become more strongly oriented toward serving the commuter and as that market is approaching saturation, new riders must come from new employment rather than an increased share of the employed. Hence, the coefficient for employment growth in the regression is higher than it used to be.

VALIDATION OF MODEL

During 1980 and 1981, Metro ridership underwent dramatic changes. These changes were due to a sudden end to large increases in the price of gasoline, rapidly declining rates of employment growth, and a substantial fare increase in May 1980. The exact behavior of the independent variables could not have been known at the end of 1979. However, there were signs that gasoline prices would decline and that employment rates would go down. The fare increase was already planned. Prepared early in 1979, the budget document for FY 1980 predicted a ridership of 62.2 million. Large ridership increases continued during the first few months of 1980 and, as a result, in the spring the ridership estimate for the year was raised to 68.3 million. The projection for 1981 was increased to 77 million. However, by mid-year ridership growth began to decline. Actual 1980 ridership was 66.1 million; in 1981 it was 66.0 million.

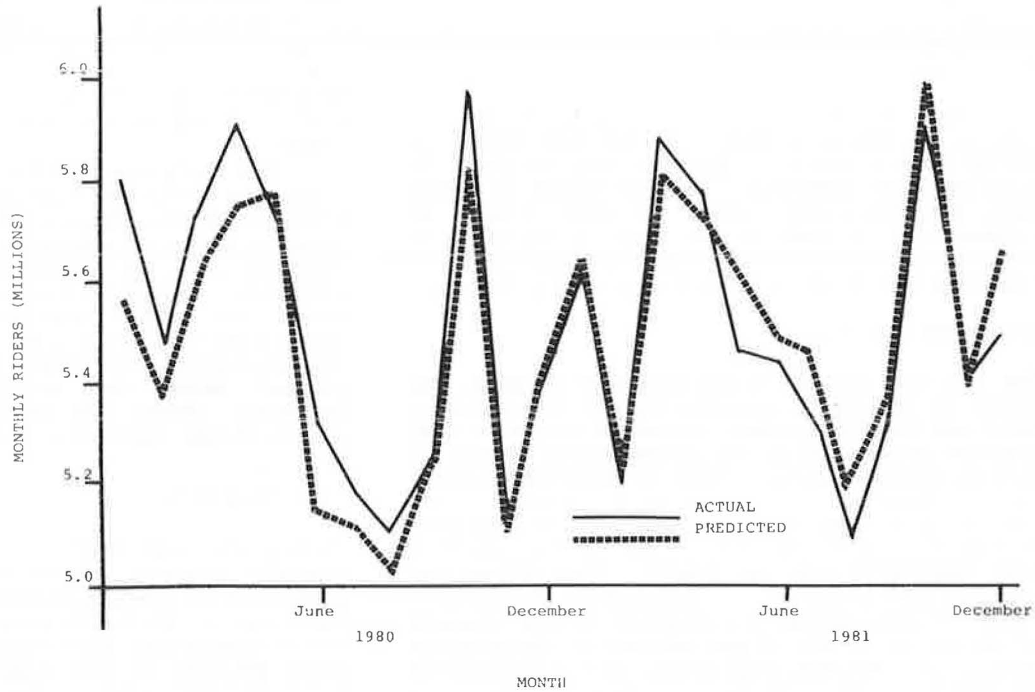
Figure 2 shows projections for ridership that the model would have produced at the end of 1979. These projections use data only up to that time as a basis for the regression. The projections for annual ridership in 1980 and 1981 would have been 65.0 and 66.6 million compared with the actual 66.1 and 66.0 million. The percentage of annual errors would have been 1.7 and 0.9 percent, respectively. The maximum monthly error would have been 4.1 percent, and the root mean square error over the 24 months, 2.0 percent. The correlation between actual and predicted ridership would have been 0.93. With seasonal and calendar adjustments taken out, the correlation would have been 0.75. All in all, use of the model would have allowed quite an accurate anticipation of the shifts in ridership patterns.

These years were very unusual, since there were wide variations in all the independent variables and the dependent variable. One should expect errors in

Table 1. Trends in regression coefficients.

Time Interval	Gasoline Price	Gasoline Supply	Fare	Service Hours	Employment	Constant Term
1/73-1/78	0.28	1.08	-0.14	-0.14	0.47	0.53
7/73-7/78	0.27	1.09	-0.14	-0.13	0.56	0.50
1/74-1/79	0.26	1.13	-0.15	0.09	0.61	0.48
7/74-7/79	0.34	1.15	-0.18	0.11	0.63	0.55
1/75-1/80	0.23	1.34	-0.19	0.19	0.42	0.61
7/75-7/80	0.23	1.34	-0.16	0.24	0.43	0.60
1/76-1/81	0.30	1.49	-0.13	0.09	1.17	0.09
7/76-7/81	0.32	1.50	-0.12	0.12	1.19	0.04
1/77-1/82	0.30	1.46	-0.14	0.08	0.97	0.23

Figure 2. Validation of 1980 and 1981 data.



monthly projections not to be so great during years when variables do not fluctuate so widely.

MAKING RIDERSHIP PROJECTIONS

One of the major objectives in the development of this short-term ridership-projection model was to produce information useful in service planning and budgeting. To this end, a monthly report was developed to present relevant information for wide distribution within the agency.

The first section of the report contains monthly projections for 24 months of ridership. Annual totals are included. These are monitored closely by several staff members and influence judgments about service increases and revenue estimates. Following these forecasts is a history of past projections. These are included primarily to give people a feeling for the accuracy of projections. It is more meaningful for people who have little statistical experience to see how the projections vary over time than it is to show the standard error or multiple correlation coefficient.

Most readers of the report are not interested in more than the first section. However, many readers have taken an interest in the data on the performance of the regression model. The regression coefficients are shown to give people a feeling for the importance of each variable. By showing the values of the coefficients for six months before, trends in the influence of the variables can be traced.

The last section of the report contains information about assumptions concerning independent variables. Historical information on real gasoline price and employment is included to give some basis for understanding the projections for these variables.

Publication of this paper sponsored by Committee on Bus Transit Systems.