

Predicting Annual Transit Fare Revenue from Midyear Results

SAMUEL SEWARD, RICHARD P. GUENTHNER, and HASSAN KH. NASSER

ABSTRACT

Uncertainty about the future and the possibility of major changes are problems that face all transit systems. As a tool in facing these uncertainties, contained in this paper is an exploration of the possibility of predicting total annual revenues for transit systems based on the revenue collected part of the way through the year. Six Wisconsin cities representing three city sizes (large, medium, and small) were used as a case study. A single prediction along with the limits of the 95 percent confidence interval were computed. Also, the percentage errors (as compared to the actual annual revenues) were calculated and the distribution of these errors was examined. The findings revealed that the proposed method is applicable to the Milwaukee County Transit System because it results in a prediction of annual revenue that will have less than 5 percent error 95 percent of the time. This method was also found to be applicable to small Wisconsin cities. However, in the case of the only medium Wisconsin city, the percentage error was higher because of anomalies in some of the data.

After World War II, the United States transit industry suffered a general decline marked by dwindling patronage and increasing costs to a point where fares would no longer cover costs for service levels desired by most urban communities. The termination of private transit services in many cities resulted, and the federal government responded in 1961 with an aid program for transit. This aid program had evolved into a full spectrum of transit assistance programs by the mid-1970s. By 1980, the U.S. Congress was providing several billion dollars annually for transit. Also, a number of states started to provide capital and operating aid directly to local transit agencies to supplement federal matching grant programs.

All these funds and grants from federal, state, and local government might appear sufficient to ensure that transit operators are able to cover their operating costs. Unfortunately, this has not been the case. Transit operating ratios still declined through the late 1970s. Because of these trends and recently decreasing federal funds, transit systems have had to be increasingly concerned with predicting annual revenues in order to balance budgets.

The purpose of this paper is to test a methodology for using partial-year transit revenue as a means for predicting total current-year revenue for a transit system. By applying such a method early in the year, transit managers will have an indication of whether to expect a shortage, surplus, or balance of funds at the year's end. The methodology will also give budgeting personnel better information for developing their annual budgets, which typically are prepared during the middle of the preceding year.

For this analysis, past revenue data were collected from six Wisconsin cities: Milwaukee, Madison, Racine, Kenosha, Janesville, and Green Bay. These

cities were chosen to represent Wisconsin's large, medium, and small cities. Some statistics for these cities are given in Table 1. This paper is presented in two sections. The first section describes the methodology that was used in making the predictions, and the second describes the results from applying the model.

UNCERTAINTY IN FORECASTING

One reason that transportation planning has limited influence in the policy process is the proven inaccuracy of the forecasts. Unfortunately, this limitation also applies to financial and budgeting forecasts for transit systems. The nature of forecasts is to be in error, and no one can ever eliminate all such errors. However, realizing this, concern should be focused on anticipating the errors and limiting both their size and their consequences.

Past research has focused on the sources of errors in forecasting and ways to reduce them (1). Although this is a worthwhile avenue to pursue, an insufficient amount of attention has been devoted to characterizing uncertainty and conveying useful information about it to decision makers. In characterizing uncertainty, researchers are primarily concerned with preparing realistic estimates of the likely range of key forecast values (2). Assessing the level of uncertainty in the final outputs is not easy, largely

TABLE 1 Selected Information Concerning the System Analyzed

System	Population	No. of Routes	1984 Vehicle-Hours	Fleet Size	1984 Operating Budget (\$)
Milwaukee	965,000	71	1,617,000	616	63,493,000
Madison	225,000	21	337,000	194	14,909,000
Green Bay	141,000	16	84,500	29	2,559,000
Kenosha	94,000	7	63,600	29	1,717,000
Janesville	52,000	7	30,300	22	929,000
Racine	85,000	12	105,800	39	2,700,000

Note: From operating reports of respective transit authorities.

S. Seward, Milwaukee County Transit System, 1942 N. 17th Street, Milwaukee, Wis. 53205. R.P. Guenther, Department of Civil Engineering, Marquette University, Milwaukee, Wis. 53233. H. KH. Nasser, Community Design Center, Inc., 340 West Brown Street, Milwaukee, Wis. 53212.

TABLE 2 Milwaukee Monthly Revenue (\$)

Year	January	February	March	April	May	June	July	August	September	October	November	December
1976	1,424,372	1,413,677	1,580,086	1,460,876	1,415,669	1,354,835	1,308,211	1,306,352	1,513,938	1,556,177	1,439,284	1,429,220
1977	1,434,763	1,435,700	1,618,745	1,413,143	1,436,619	1,380,091	1,286,282	1,368,485	1,569,966	1,624,888	1,597,294	1,571,302
1979	1,676,124	1,599,742	1,750,953	1,604,397	1,596,898	1,520,803	1,494,836	1,565,873	1,620,860	1,781,055	1,662,921	1,548,043
1980	1,696,710	1,713,554	1,809,995	1,747,461	1,696,813	1,605,917	1,618,049	1,607,891	1,768,568	1,869,083	1,648,512	1,670,711
1981	2,084,294	1,998,171	2,238,437	2,052,699	1,940,304	1,915,990	1,908,101	1,839,716	2,010,404	2,103,085	1,943,993	1,880,787
1982	2,120,047	2,259,030	2,496,680	2,400,401	2,089,531	2,021,844	1,919,912	1,957,116	2,164,924	2,260,056	2,113,576	1,884,130
1983	2,139,726	2,259,448	2,493,455	2,183,944	2,177,069	2,065,629	1,958,554	2,005,853	2,183,183	2,187,442	2,188,989	1,804,155
1984	2,196,698	2,275,843	2,393,248	1,105,655	2,220,354	2,023,739	1,984,635	2,137,704	1,994,967	2,272,296	2,069,856	1,946,431

because of the multiplicity of inputs, estimates, assumptions, and the uncertainty associated with each. At least as important is finding constructive ways to convey the nature and significance of this uncertainty to the users of the forecasts.

Yet the difficulties that researchers have in dealing with uncertainty suggest that there are many risks involved in revealing this information. The risks come from the possibility of frightening the decision makers when the reality of uncertainty is spoken about openly.

One way to deal with these issues will be discussed in the next section. The attempt here is to understand the characteristics of errors to convey useful information about them. This has been done through studying the seasonal trends of transit revenues and, consequently, developing a methodology to predict these revenues, with a reasonable, understandable error. Also, although much of planning addresses the long range, the method presented here is applicable to the short range. This method is also simple and easy to understand.

METHODOLOGY

This study was primarily done for the Milwaukee County Transit System (MCTS), which is a mass transit organization owned by Milwaukee County, but managed and operated by Milwaukee Transport Services, Inc., a private, nonprofit organization. For the first 115 years of its history, the Milwaukee Transit System was able to provide reasonable service at acceptable rates at no cost to nonusers. The transit system was a taxpayer during this period and the system was supported solely by the fares. The Milwaukee Transit System continued as a private system many years longer than did most transit systems in other cities.

But, with a less dense community, more scattered riding requirements, and subsidized highway programs, the time was reached in Milwaukee when quality transit service could no longer be supported by the farebox alone. Beginning in 1975, the system has been subsidized by federal, state, and county funds. In fiscal year 1984, the total system operating cost was \$63.5 million. Operating revenue (primarily fares) covered \$30.5 million or 48 percent of that cost. The remaining cost was covered by \$7.4 million

in federal funding, \$21.8 million in state funding, and \$3.7 million in county funding. As can be seen, even though the farebox does not cover all the costs, its contribution is large enough that short-range predictions would be useful.

In this paper, the predictions have been based on monthly revenues from 1976 through 1984 excluding 1978 due to a strike that made that year's annual total inapplicable. These values are given in Table 2. Annual revenue predictions were made from each month's cumulative total during the 8-year period. The second step was to obtain a 95 percent confidence interval for these predictions to provide a clear range as to where the estimates will fall.

The prediction was computed by first finding the cumulative monthly revenue and cumulative percentage of the total revenue for each month for all 8 years. The resulting annual percentage values for each month were then averaged over the 8-year period, and used to predict the annual revenue. These averages, along with the standard deviations, are given in Table 3.

The predictions for each month during the 8-year period were determined from the average cumulative percent and the revenue through the month in question, by using the following formula:

$$PRDI = (QDATA/QAVE) * 100 \quad (1)$$

where

$$\begin{aligned} PRDI &= \text{the middle prediction,} \\ QDATA &= \text{the cumulative monthly revenue, and} \\ QAVE &= \text{the cumulative percent average.} \end{aligned}$$

In other words, these values represent the predictions that could have been made at that point in time. For example, at the end of May 1982, the annual revenue estimate would have been made from the January through May totals. To determine the accuracy of the method, each prediction was then compared with the actual annual revenue. The percentage errors between the two are given in Table 4.

To determine the precision of each prediction, a 95 percent confidence interval can be computed. The high and low limits of this interval represent the 95th percentile confidence interval of the bounds of the cumulative average. In other words, transit systems will be able to predict the total annual revenue

TABLE 3 Cumulative Percentages of Milwaukee Revenue by Year and Month (%)

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.
1976	8.28	16.50	25.68	34.17	42.40	50.28	57.88	65.48	74.28	83.33	91.69
1977	8.09	16.18	25.31	33.28	41.38	49.16	56.41	64.12	72.98	82.14	91.14
1979	8.63	16.87	25.88	34.14	42.36	50.19	57.89	65.95	74.30	83.47	92.03
1980	8.30	16.67	25.52	34.07	42.36	50.21	58.13	65.99	74.63	83.77	91.83
1981	8.72	17.07	26.43	35.01	43.13	51.14	59.12	66.81	75.21	84.01	92.14
1982	8.25	17.05	26.77	36.11	44.25	52.12	59.59	67.21	75.64	84.44	92.67
1983	8.34	17.15	26.87	35.39	43.88	51.93	59.57	67.39	75.90	84.43	92.97
1984	8.57	17.46	26.80	35.02	43.68	51.58	59.33	67.67	75.46	84.32	92.40
Average	8.40	16.87	26.17	34.65	42.93	50.82	58.49	66.32	74.80	83.74	92.11
Standard deviation	0.203	0.378	0.591	0.841	0.906	0.964	1.040	1.105	0.894	0.723	0.537

TABLE 4 Percentage Error by Year and Month

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.
1976	-1.40	-2.20	-1.82	-1.37	-1.22	-1.08	-1.03	-1.28	-0.70	-0.49	-0.45
1977	-3.67	-4.06	-3.24	-3.96	-3.62	-3.29	-3.56	-3.32	-2.44	-1.91	-1.05
1979	2.77	-0.01	-1.06	-1.46	-1.32	-1.24	-1.02	-0.57	-0.67	-0.32	-0.09
1980	-1.21	-1.16	-2.43	-1.68	-1.32	-1.20	-0.62	-0.51	-0.22	0.04	-0.30
1981	3.78	1.20	1.04	1.05	0.46	0.61	1.07	0.72	0.55	0.32	0.03
1982	-1.72	1.06	2.33	4.22	3.07	2.54	1.89	1.33	1.12	0.84	0.60
1983	-0.65	1.68	2.74	2.14	2.21	2.18	1.85	1.60	1.47	0.83	0.93
1984	2.10	3.48	2.44	1.06	1.75	1.48	1.43	2.02	0.88	0.70	0.32
Standard deviation	2.41	2.24	2.26	2.43	2.11	1.90	1.78	1.66	1.20	0.86	0.58
95 percent interval	5.70	5.30	5.34	5.74	4.99	4.49	4.21	3.93	2.84	2.03	1.37

and give a range in which the actual annual revenue will fall 95 percent of the time. Because the sample size is somewhat small (8 years of data), the normal distribution was inappropriate. Instead, the Student's T-distribution was used. The Student's T-distribution for 7 degrees of freedom (for 8 years of data) equals 2.365 (3). The resulting equations are

$$HI = QAVE + STUdT * QSSD \quad (2a)$$

and

$$XLOW = QAVE - STUdT * QSSD \quad (2b)$$

where

- STUdT = 2.365,
- HI = the upper limit of the mean standard error,
- XLOW = the lower limit of the mean standard error, and
- QSSD = the cumulative standard deviation.

The limits of each revenue prediction are

$$PRD2 = (DATA/HI) * 100 \quad (3a)$$

and

$$PRD3 = (DATA/XLOW) * 100 \quad (3b)$$

where DATA is the monthly revenue.

By using the mean and standard deviation for each column given in Table 4 and applying Equations 2 and 3, confidence intervals for the observed errors were computed. (These confidence intervals are given in Table 4.) These confidence intervals represent the expected range of the error of the predictions. For example, a prediction made using only information from January can be expected to be within 5.70 percent of the actual revenue 95 percent of the time.

Because the concept of the 95 percent confidence interval does not automatically have a great deal of meaning to a nonstatistician, the average error was also computed by separating the predictions from each month's data into positive and negative variations and then averaging those categories. Figures 1, 2, and 3 show these average errors as well as the largest negative and positive variations found in the predictions for each month. From this line, the range and direction of the variations can be seen.

It should be recognized that predicting revenues based on more limited data will reduce the accuracy of the results. Ideally, a prediction system such as this should be based on data from as many years as is feasible. By using the data in Table 3, the variation is shown to be much greater when a limited

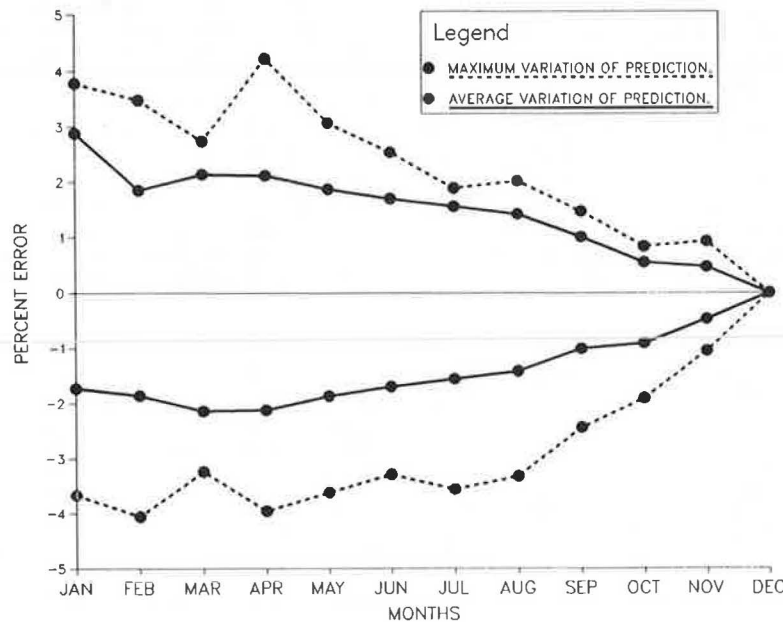


FIGURE 1 Largest errors in revenue predictions in Milwaukee.

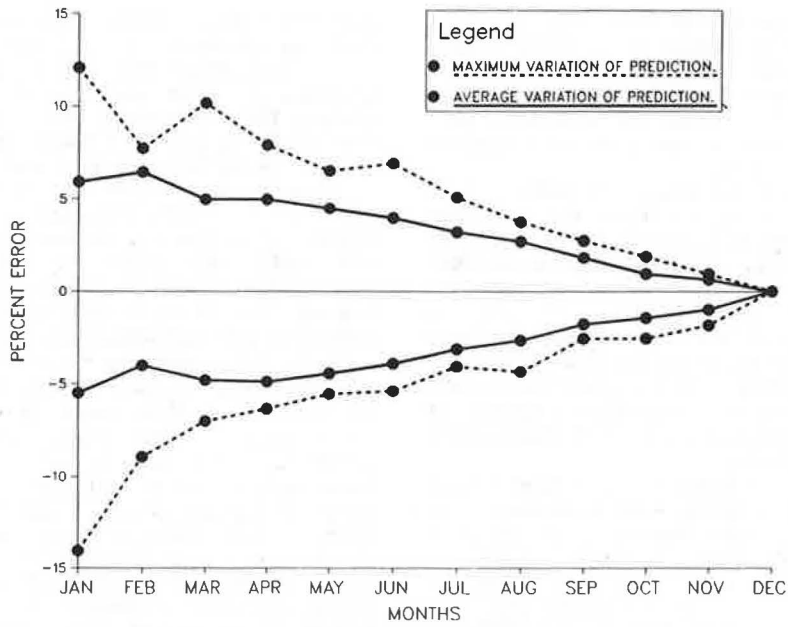


FIGURE 2 Largest errors in revenue predictions in Madison.

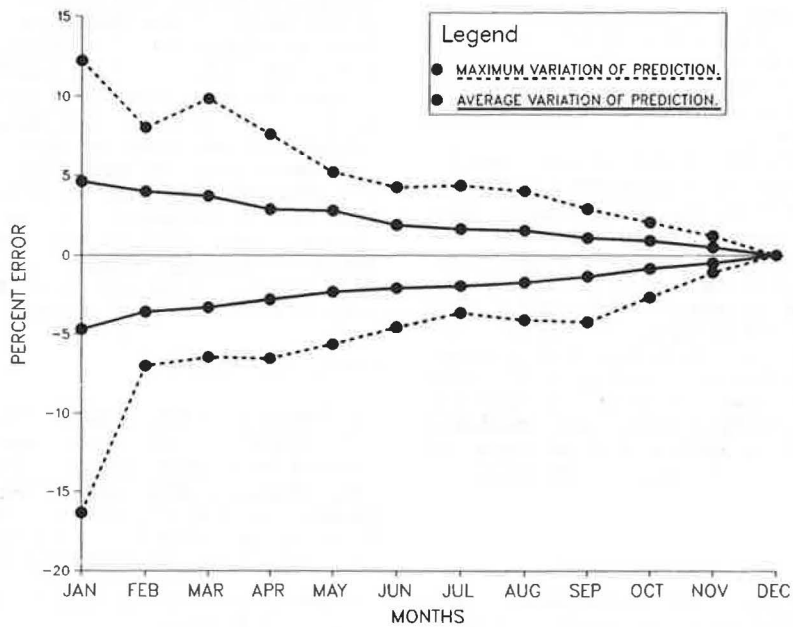


FIGURE 3 Largest errors in revenue predictions in small cities.

data base is used. For example, if a January prediction had been made in 1979 using only the 1976 and 1977 data, the prediction would have been 5.4 percent above the actual revenue. In comparison, the maximum error for a January prediction throughout the entire 8 years is only 3.78 percent.

DATA ANALYSIS AND RESULTS

Milwaukee

Predicting total annual revenue on the basis of the revenue collected in the first few months of that year is applicable to the MCTS. The average and maximum errors in predicting these revenues shown in

Figure 1 indicate an acceptable range of predictions. Table 4 indicates that the highest negative error is -4.06 percent in February 1977 and the highest positive error is 4.22 percent in April 1982. These figures are good indications that the prediction will stay in the range of -5 to +5 percent error because the highest and lowest errors in the actual data, which include 1984 predictions, do not exceed the 95th percentile confidence interval.

In general, the revenue in January is about the same as that in February. Consequently, if the standard deviation of the cumulative percentage for January and February is less than twice that for January alone, more accurate predictions can be made. Examination of the standard deviation as given in Table 2 reveals that the cumulative percentage in

February is slightly more stable than that in January. A February prediction should be slightly better than one using information only from January. However, after April, the standard deviation is greater than four times that for January. The resulting predictions in April are not as accurate as those in January, February, and March.

Starting in May, the standard deviations have less-than-proportional increases from the previous months; consequently, the accuracy of the predictions starts to improve at that time. This indicates that no prediction should be made before April because the information during April is not sufficient to make the prediction. This can also be seen clearly by tracing the variation of error of the prediction shown in Figure 1. It appears that weather variations and the Easter holiday cause a significant amount of year-to-year variation (a high percent of Milwaukee's ridership is schoolchildren).

As can be seen from these numbers and from Figure 1, the best time (from a management standpoint) to make the prediction is during May or June. By that time, the 95th percentile error is within 5 percent. By then, the prediction is accurate and there is still time to react to it. This indicates that the method developed is highly applicable to the MCTS.

Madison

Unfortunately, Madison was an exception to the rule of consistent results because the error range was wide. As can be seen from Figure 2, the average errors do not enter the 5 percent range until March, and the limits of the maximum error found do not enter that range until July.

One important observation of the Madison results here is that the minimum worst possible errors occurred between 1974 and 1979; the maximum worst possible error occurred between 1981 and 1984. One reason for this is that changes in services must have occurred around 1980, such as adding new routes or increasing fares. Based on general information concerning the system, a large number of service and fare changes have occurred during the study period. Accordingly, no stable data were available for this analysis. One positive finding from the Madison results is that transit operators could have predicted their total annual revenue with a percent error of less than 6 percent during all years except 1982.

Small Cities

The results from all of the small cities were similar to each other. The dotted lines in Figure 3 repre-

sent the highest errors from the four cities for each month in question. The fluctuations of the graph are due to increasing services or a raise in fares. For example, a sudden increase of the percent error of January 1982 in Janesville was due to a fare increase from \$0.35 to \$0.50. Another increase had also been instituted in February 1981 from \$0.25 to \$0.35.

Another change that can cause variations is the addition of routes. This was the case for the Racine system, where two routes were added between June 1979 and April 1980. Also, in October 1982, peak-hour service was improved on four routes, with headways reduced from 30 to 20 min. These changes will affect ridership and, consequently, the monthly revenues.

With the exception of the first 4 months, the maximum errors stayed in the 5 percent range, which indicates that this model is applicable to Wisconsin's small cities. Also, with the exception of Janesville, the maximum and minimum worst errors for these cities varied between -6.73 in February 1979 in Green Bay and 10.47 in January 1984 in Green Bay. Unlike other small cities, Janesville's percentage error ranged from 12.23 percent in January 1984 to -16.35 percent in January 1981. Also note that these Janesville errors dropped to about one-half their magnitude in February to 6.32 and -6.00, respectively.

As was done for Milwaukee, the standard deviation and the 95th percentile confidence interval for the prediction errors were computed for Madison and each of the small cities. These are given in Table 5. Note that in Madison and Janesville, an error smaller than 5 percent could not be expected until the end of August. However, acceptable results could be obtained in the other small cities during the summer months between May and July. At this time, adjustments for fall could still be made. None of these results was as consistent as those in Milwaukee. However, wider ridership fluctuations can be expected in smaller cities. Also, because the small cities have much smaller operating budgets, an erroneous prediction of 5 percent would not be as much in terms of dollars as would one in Milwaukee.

USING THE FINDINGS

To demonstrate how future predictions can be made, the following predictions were computed using the revenue data from January, February, and March of 1985 obtained from the MCTS. The actual monthly revenues were \$2,614,225 for January, \$2,485,342 for February, and \$2,637,972 for March. The calculations were done according to Equations 1-3 as follows:

Average cumulative percentage of January (1976 to 1984) = 8.3975,

TABLE 5 Standard Deviation by City and Month

City	Month										
	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.
Madison											
Standard deviation	7.43	5.75	5.57	5.21	4.61	4.18	3.45	2.88	1.98	1.31	0.86
95 percent interval	17.57	13.60	13.18	12.32	10.90	9.89	8.16	6.81	4.68	3.10	2.04
Green Bay											
Standard deviation	4.73	4.71	4.11	3.78	3.38	2.63	1.95	1.39	1.08	0.69	0.39
95 percent interval	11.20	11.13	9.71	8.94	8.00	6.22	4.61	3.29	2.55	1.63	0.75
Janesville											
Standard deviation	8.66	5.05	5.17	4.20	3.52	2.78	2.53	2.26	1.52	1.01	0.67
95 percent interval	20.50	11.94	12.21	9.94	8.32	6.58	5.98	5.35	3.59	2.39	1.58
Racine											
Standard deviation	4.27	3.20	2.43	2.10	2.33	2.16	2.19	2.01	1.88	1.44	0.66
95 percent interval	10.09	7.56	5.75	4.96	5.51	5.10	5.17	4.76	4.45	3.40	1.56
Kenosha											
Standard deviation	4.05	3.99	3.82	2.73	2.28	1.26	1.40	1.69	1.10	0.85	0.53
95 percent interval	9.58	9.44	9.04	6.45	5.39	2.98	3.31	4.00	2.60	2.01	1.25

Standard deviation = 0.2028, and
 Student's t-distribution * Mean standard error
 = $2.365 * 0.2028 = 0.4796$.

Annual projection = January revenue/Average cumulative percentage = $2,614,225/0.083975$
 = \$31,130,991.

The lower bound of the 95 percent confidence interval
 = January revenue/(Average cumulative percentage
 + 0.004796) = $2,614,225/(0.083975 + 0.004796)$
 = \$29,449,015.

The upper bound of the 95 percent confidence
 interval = January revenue/(Average cumulative
 percentage - 0.004796) = $2,614,225/(0.083975$
 - 0.004796) = \$33,016,729.

In the following table, the rest of the results are given (in dollars) for February and March. The predictions are decreasing from the January predictions.

<u>Predictions</u>	<u>January</u>	<u>February</u>	<u>March</u>
Low	29,449,015	28,713,778	28,067,103
Middle	31,130,991	30,032,785	29,580,575
High	33,016,729	31,912,183	31,235,019

Because the final 1985 revenues are not finalized, these predictions cannot be tested. However, if the trend has not varied much from the past 8 years, it is expected that starting in May, accurate predictions should be available.

CONCLUSIONS

Predicting total annual revenue based on partial-year revenue is an important mechanism that can be used by planning and budgeting personnel in the transit industry. Revenue collected from regular transit operations is often the largest component of overall system revenue. In addition, while other income is predetermined, the ridership levels cannot be controlled. Thus, accurate predictions of annual revenue can be extremely valuable. A methodology to project transit revenues was developed based on the variations of revenue of large, medium, and small Wisconsin cities. A prediction based on the cumulative percent average of each month was developed in addition to two upper and lower limits for this prediction.

Analyzing the results obtained from these cities indicated that this mechanism is applicable, and that accurate short-term predictions could be generated by using it. The results of the data analysis indi-

cate that reliable estimates, with less than 5 percent error, of future revenue for the MCTS could be predicted as early as May with 95 percent confidence. Similarly, reasonably accurate forecasts were obtained for Wisconsin's small cities with errors not exceeding 5 percent. Less accurate estimates were made for Madison as a result of the inconsistency of some revenue data (year 1982), and an apparent system change in 1980. Therefore, a practical model cannot be developed for the Madison system until a stable data base is available.

Applying this model will give budgeting personnel a clearer picture of their future revenue by giving them a clear date around when year-end shortfalls or surpluses can be detected. In addition, the possible error accompanying this projection can be estimated.

Additional research should be conducted to improve the validity of the model by exploring the year-by-year fluctuation. The years with the greatest error are those that changed most from the previous year and if a version of the model can be developed that takes this variation into account, earlier and more accurate forecasts may be possible. This will encourage more operators to consider using it as an ongoing management analysis tool.

ACKNOWLEDGMENTS

Necessary data were provided by the Milwaukee County Transit System, Madison Metro, Kenosha Transit, Racine Transit, Janesville Transit, and Green Bay Transit. Partial funding was provided by a grant from the National Science Foundation.

REFERENCES

1. F.S. Koppelman. Uncertainty in Travel Behavior Forecasting. In *New Horizons in Travel Behavior*, Lexington Books, Lexington, Massachusetts, 1981.
2. M. Godet. *The Crisis in Forecasting and the Emergence of the Prospective Approach*. Pergamon Press, Inc., New York, 1979.
3. E.R. Walpole and H.R. Myers. *Probability and Statistics for Engineers and Scientists*. The Macmillan Company, New York, 1972.

The authors are solely responsible for the content of this paper.

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