Estimating Average Distance Travelled from Bus Boarding Counts

Word Count: 3330 + 10 figures + 3 tables = 3330 + 2500 + 750= 6580 equivalent words

A. J. Richardson
The Urban Transport Institute
PO Box 363, Alexandra, Victoria, 3714, Australia
Phone/fax number: 61 3 5774 7617
e-mail address: tony.richardson@tuti.com.au

Abstract. A common task involved in the monitoring of public transport services is the measurement of passenger boardings and the estimation of distances travelled by those passengers. This is often done by means of the conduct of sample surveys of passenger boardings and alightings and the calculation of the average distance travelled by passengers of various types. This paper describes the development and testing of a different method of calculating average distance travelled on a bus route that uses data obtained from automatic ticketing machines rather than from sample surveys. Importantly, the method relies only on the locations at which passengers board, and does not need to know where each passenger alights. Rather, a simplifying assumption is made that most bus trips are round trips and that passengers board the bus for their return journey at about the same place as where they got off on their forward journey. The method has been tested against trip lengths estimated from a comprehensive origin-destination survey carried out on National Bus Company buses in Melbourne, Australia in April-May 1994, and has been shown to give very good estimates of average distance travelled. The error in estimating total distance travelled by adult full-fare and adult concession passengers is less than 0.5%, while the error in estimating total distance travelled by student passengers is about 3%. The larger error for students is primarily due to the smaller passenger flows for students. Given these results, it is concluded that the "Up-Down" method of calculating average distance travelled based on boarding data obtained from automatic counts provides a better, cheaper and more flexible method of calculating average distance travelled than using sample surveys which measure boarding and alighting location.

BACKGROUND

The National Bus Company is the largest private provider of bus services in Melbourne, Australia. For several years, the Transport Research Centre (TRC) was involved in monitoring patronage levels on National Bus Company (NBC) services and in calculating the reimbursements due to NBC from the State Government in return for the carriage of various types of passenger already holding multi-modal tickets issued on government-run train and tram services.

Two major tasks were involved in this monitoring process. Firstly, each month the TRC received electronic copies of the records of all passenger boardings as recorded by the NBC drivers on on-board ticketing machines. After editing of these records, the TRC calculated the reimbursement due to NBC based on the number of passengers (of various types) carried and the distance that these passengers were carried. The second task of the TRC was to carry out sample surveys to periodically update the number of sections travelled by the different types of passenger on each route. This task was deemed necessary because the on-board ticketing machines recorded only the boarding of passengers and did not record (except for those passengers buying cash NBC tickets) how far they travelled. The TRC sample surveys recorded the boarding point and alighting point of all passengers and calculated the average number of sections travelled by subtracting the average boarding point of the passengers (of each type) from the average alighting point of these passengers.

There were, however, some misgivings about the methods used to calculate the average sections travelled. These included:

- the results were calculated from sample surveys, and hence were subject to the normal levels of sampling error associated with sample surveys;
- the frequency of conducting the surveys was variable, ranging from once per quarter for the more heavily loaded routes to once per year for the lightly loaded routes;
- on any surveyed day, not all runs on a route were surveyed and this raised questions as to whether the surveyed runs were truly representative of the entire day's runs on that route;
- surveys were conducted only on weekdays, thereby giving no indication of whether trip distances were different on weekends; and
- surveys were conducted only during "normal" times of the year, with no surveys conducted on public holidays or during school holidays.

For these reasons, it was considered desirable to find an alternative way of calculating the average sections travelled, which would account for the issues raised above but which would not involve asking questions directly of the passengers involved.

The estimation of transit travel patterns from passenger count data has received surprisingly little attention in the literature, most attention having been given to estimating road-based travel patterns from traffic count data (e.g. 1, 2, 3). Wong and Tong (4) describe the estimation of time-dependent origin-destination matrices for transit networks, using Hong Kong Mass Transit Railway data to empirically validate their methods, while Friedrich, Mott and Noekel (5) use a fuzzy logic approach for the continuous updating of transit origin-destination matrices from passenger count data for the Munich S-Bahn system.

Both Wong and Tong (4) and Freidrich et al. (5) were interested in the estimation of origin-destination matrices, given counts of passenger boardings and alightings. The emphasis in this paper is slightly different, in that it is concerned with estimating average distance travelled from boarding counts only. This is because the allocation of revenue in multi-operator public transport systems is usually based on the number of boardings and the total distance travelled on each of the operators. Subsequent papers will describe the success in estimating distributions of travel distance and origin-destination matrices from boarding counts only.

THE BASIC CONCEPT

To estimate average distance travelled from boarding counts only, a method was devised which was deceptively simple in its basic concept but which, as shall be seen later, is remarkably accurate in its ability to predict the average number of sections travelled by passengers of various types.

The basic concept underlying the method is an assumption that passenger flows on most public transport routes are reversible over the course of the day. This is essentially the same as the symmetry assumption used by Navick and Furth (6), when they state that "The boarding pattern for a route in one direction is equivalent to the alighting pattern in the opposite direction over the course of the day". That is, it is assumed that most trips are return trips from A to B, and later back from B to A. Under this assumption, it is assumed that the place where a passenger alights from a bus is most likely to be the same as, or very close to, the place where they get on the bus for the return trip. Therefore, although the alighting point of passengers is not recorded in the NBC ticketing-machine data, it can be approximated (on average) by the boarding points for trips in the reverse direction on the same route.

There is no reason why the above assumptions are necessarily correct. It is possible that some public transport trips are one-way only and not associated with any return trip. It is also possible that the return trip is different to the forward trip in that the passenger may access a different stop on the return trip from the one at which they alighted on the forward trip. However, it is probably reasonable to assume that most public transport trips are reversible. This assumption will be tested empirically later in this paper.

TESTING THE CONCEPT

The concept outlined in the previous section could fortunately be tested empirically because of a data collection exercise conducted on NBC buses in April-May 1994. For 52 days in this period, all passengers boarding the buses running on all 38 NBC routes were also asked to state their destination. While this process generated a considerable amount of information, it also caused considerable interference to the boarding process for passengers. Hence, in late May, the process was discontinued. Nonetheless, during this period, information had been gathered on the origins and destinations of over one million passenger trips on all NBC routes.

This database could therefore be used to test the new "up-down" method of calculating the average number of sections travelled, and then compare the results with the actual sections travelled as measured in the NBC survey data. Since this period also included weekends and school holidays, it was possible to see whether any variations in trip length occurred in these periods.

The Method of Analysis

Every passenger boarding an NBC bus during the period from April 1 through May 22, 1994 had their boarding point recorded (as is normal with all on-board ticketing machine recordings). In addition, they were asked to state where they would be alighting. From this information, the actual length of their trip was calculated, and the average could then be calculated for all passengers on the route. In addition, given that their boarding point was recorded, the new "up-down" method of calculating average trip lengths could be used to estimate the average trip length for all passengers on the route.

An Example - Route 1

As an example of the methods used in this analysis, we will consider a single route (Route 1 - this route number is an alias for an actual route to preserve confidentiality) and later generalise to the entire population of NBC routes being operated in April 1994. For the purpose of this analysis, we will consider only those passengers travelling on MET tickets (the multi-modal tickets used on the Government MET system of trains, trams and buses), since NBC-ticket passengers are either not subject to reimbursement calculations or, for NBC adult concession passengers, they continue to have both origin and destination recorded on the on-board ticketing machines in the process of buying a ticket.

Calculating Actual Sections Travelled: The first step in the process was to calculate the actual sections travelled using the full information on boarding and alighting location recorded over the 52 days of the O-D survey period. The first stage in this process was to calculate the number of sections travelled for each trip (noting that a trip which starts and ends in the same section is counted as a 1 section trip, a trip which starts and ends in adjacent sections is a 2 section trip, and so on). For each day of the survey period, the distribution of sections travelled was then calculated. The next step converted the absolute frequencies into relative frequencies and also recognised that

different trip patterns may exist on different types of day within the survey period. For the purposes of this analysis, the 52 days were split into 24 working days that were not school holidays, 17 weekend and public holiday days, and 11 working days that were also school holidays. The reason for this split was to see whether different numbers of sections were travelled on different types of day, bearing in mind that the existing sample survey procedures were only conducted on working days that were not school holidays. For each of these groups, the average proportion for each number of sections was calculated, as was the standard deviation of this proportion across these days. These calculations were performed in both the Up and Down directions for each type of passenger for each type of day. The results for Route 1 for full adult MET passengers are shown in Figure 1 for normal working days, weekends/public holidays, and school holidays.

Similar calculations were performed for adult concession MET passengers and student concession MET passengers for each type of day. The results of the entire analysis are summarised numerically for Route 1 in Table 1.

It can be seen that, across all passengers, the average number of sections travelled did vary by type of day for Route 1, with travel on weekends and public holidays being longest overall at 2.91 sections, followed by travel in school holidays at 2.71 sections, with travel on normal working days being the shortest at 2.56 sections. These trends apply across all types of passenger, although the actual number of sections travelled does vary by passenger type. It therefore appears that there may be a case for separately estimating sections travelled on weekends and holidays, if an accurate representation of total sections travelled is to be obtained. Even though fewer people, of all types, travel on weekends, public holidays and school holidays, those that do travel tend to travel longer distances than those travelling on normal working days. Using the "Up-Down" method enables these differences to be observed without the need for costly sample surveys on all types of day.

Estimating Average Sections Travelled using the "Up-Down" Method: The previous section has calculated the actual number of sections travelled using the boarding and alighting sections for each passenger trip as recorded by the NBC drivers using the on-board ticketing machines during the survey period. This section of the paper now discards the alighting information, and attempts to estimate average sections travelled using only the boarding location information for each trip. The first step was to tabulate the distribution of boarding locations for trips in the Up direction for each day of the survey period. The next step was to tabulate the distribution of boarding locations for trips in the Down direction for each day of the survey period.

The average boarding location in each direction was then calculated by taking the weighted average of the boarding locations for each day. The boarding locations in the Up and Down direction are now compared to estimate the average sections travelled, by making the assumption that passengers travelling in the Up direction will alight, on average, at the same location as passengers travelling in the Down direction board the bus for the return journey. The results for all types of passengers on all types of day on Route 1 are summarised numerically in Table 2.

By comparison with the actual sections travelled shown in Table 1, it can be seen that the "Up-Down" estimation method gives good results for Route 1 for each type of day and for each type of passenger. Overall the estimated sections travelled for Route 1 is 2.61, compared to the actual value of 2.64 (an error of approximately 1%).

The Results for All Routes

The same analysis as described above was repeated for all 38 of the NBC routes. The results of this analysis in terms of the flows on each route, the actual average sections travelled by MET passengers of various types and the "Up-Down" estimates of sections travelled are shown in Table 3 (note that the route numbers are aliases, while the flows have been randomly factored to preserve confidential information).

It can be seen that, in general, the estimated and actual sections travelled are in good agreement. Overall, by comparing the weighted average number of sections travelled across all the routes, the adult full fare results differ by about 1.5%, the adult concession results differ by 1%, while the student concession results differ by about 4.5%.

The results shown in Table 3 can be shown graphically by plotting the estimated versus actual sections travelled as shown in Figures 2, 3 and 4.

As can be seen from the graphs, and the slopes of the regression lines fitted to the data, the estimated "Up-Down" values appear to be a reasonably good fit to the actual values. However, some of the points appear to lie a significant distance away from the line of best fit, especially for the student fares. Examination of the ratios of the estimated and actual values shows that these outliers are correlated with the passenger flows on the routes, as shown in Figures 5, 6 and 7.

It can be seen that the highest and lowest ratios occur for low flow routes. As the flows increase, the ratios approach a value of unity (i.e. the estimated sections travelled equals the actual sections travelled). It thus appears that, as would be expected from sampling theory, the results are more accurate for those routes with larger sample sizes.

The above result is confirmed if one plots the total number of sections travelled on each route, rather than the average number of sections, as shown in Figures 8, 9 and 10. In practice, it should be remembered that it is the total number of sections travelled, not the average sections travelled, which govern the overall level of reimbursement paid to NBC.

As shown in Figures 8, 9 and 10, the estimated and actual total sections travelled on each route are very close. Figure 10, however, shows that while the student fares are the worst predicted, the overall level of passenger flows for student fares are much smaller than for adult full fares or adult concession fares. Therefore, any error in the estimation of average student sections travelled will have a smaller effect on total sections travelled. In addition, since student fares are only reimbursed at half the level of adult fares, the impact on total reimbursement amounts will be even further reduced.

CONCLUSIONS

This paper has sought to demonstrate the value of a different method of calculating average sections travelled on a bus route. Importantly, the method relies only on the locations at which passengers board, and does not need to know where each passenger alights. Rather, a simplifying assumption is made that most bus trips are round trips and that passengers board the bus for their return journey at about the same place as where they got off on their forward journey.

The method has been tested against trip lengths estimated from a comprehensive origin-destination survey carried out on NBC buses in Melbourne in April-May 1994, and has been shown to give very good estimates of average and total sections travelled. The error in estimating total sections travelled by adult full-fare and adult concession passengers is less than 0.5%, while the error in estimating total sections travelled by student passengers is about 3%. The larger error for students is primarily due to the smaller passenger flows for students.

Given the above results, and bearing in mind the flexibility offered by calculating average sections travelled on a continuous basis (thereby explicitly accounting for the distribution of school holidays, weekend days, public holidays and normal working days within any particular month), it is concluded that the "Up-Down" method of calculating average sections travelled is better than the conventional method of calculating average sections travelled from sample surveys on selected days of the year.

This method of calculating the length of passenger trips is applicable to any public transport service that has a continuous record of boardings from Automatic Ticketing Machines, and where it is reasonable to make an assumption that most trips are two-way trips between origin and destination. Importantly, these results confirm the findings of Navick and Furth (6), and support their conclusion that "this finding should be confirmed using data from other cities". The "Up-Down" method has now been used since 1997 as the primary method of calculating average sections travelled in the NBC bus surveys.

While the “Up-Down” method as described in this paper does not provide a trip length for any individual passenger, it does provide very good estimates of average trip length across a group of passengers. However, further research, to be described in subsequent papers, will describe the success in estimating distributions of travel distance and origin-destination matrices from boarding counts only.

REFERENCES

1. Van Zuylen, H.J. and Willumsen, L.G. (1980). “The Most Likely Trip Matrix estimated from Traffic Counts”. Transportation Research B, 14, pp 281-293.
2. Bell, M.G.H. (1991). “The Estimation of Origin-destination Matrices by Constrained Generalised Least Squares”. Transportation Research B, 25, pp 13-22.
3. Yang, H., Iida, Y. and Sasaki, T. (1991). “The Equilibrium-Based Origin-Destination Matrix Estimation Problem”. Transportation Research B, 28, pp 23-33.
4. Wong, S.C. and Tong, C.O. (1998). “Estimation of Time-Dependent Origin-Destination Matrices for Transit Networks”. Transportation Research A, 32, pp. 35-48.
5. Friedrich, M. Mott, P. and Noekel, K. (2000). “Keeping Passenger Surveys up-to-date – A Fuzzy Approach”. Paper presented to the 79th Annual Meeting of the Transportation Research Board, Washington DC.
6. Navick, D.S. and Furth, P.G. (2002). "Using Location-Stamped Farebox Data to estimate Passenger-Miles, O-D Patterns, and Loads". Paper presented to the 81st Annual Meeting of the Transportation Research Board, Washington DC.

LIST OF FIGURES

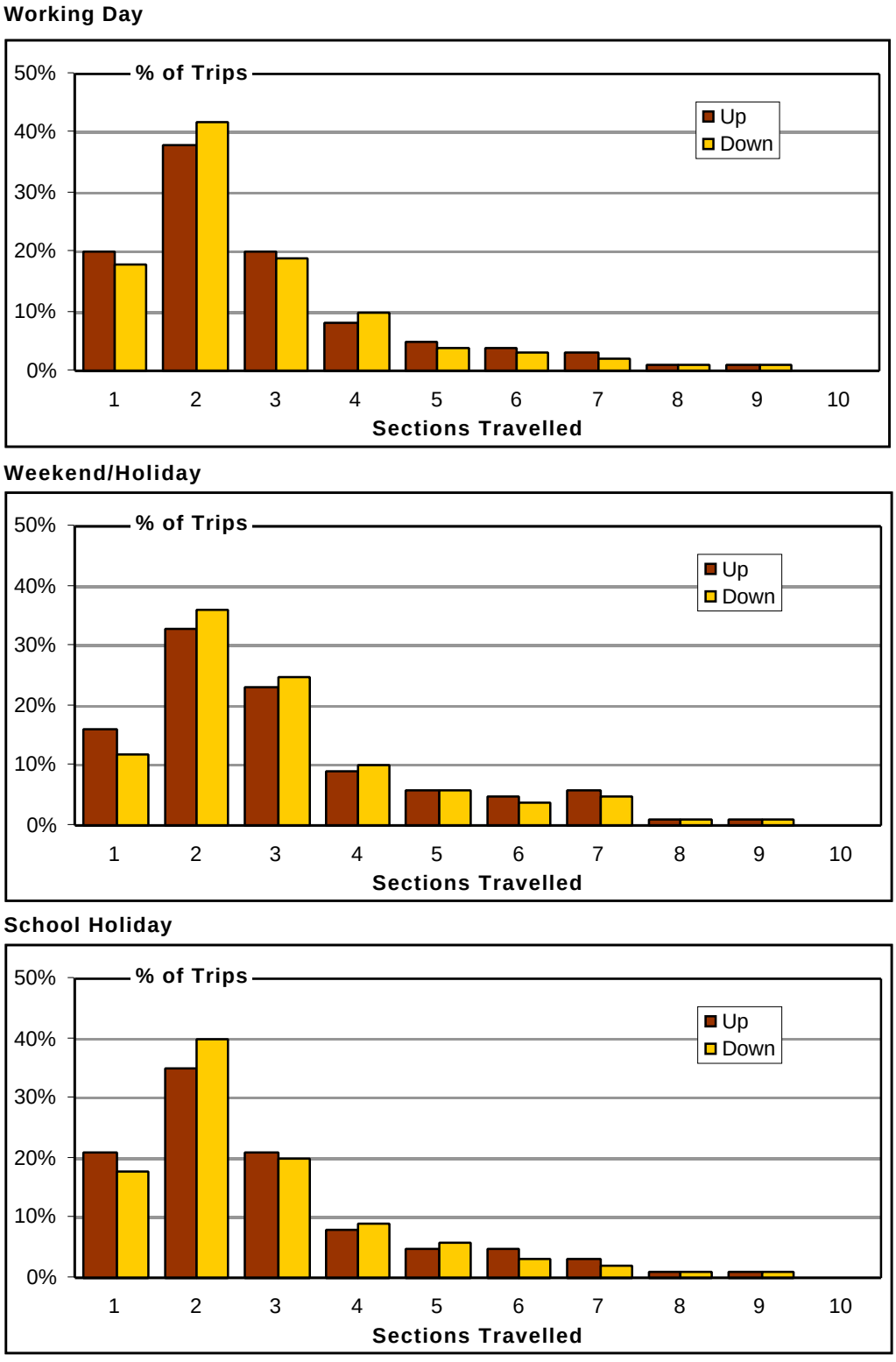

FIGURE 1 Sections Travelled by Full-Fare Adult Passengers on Route 1

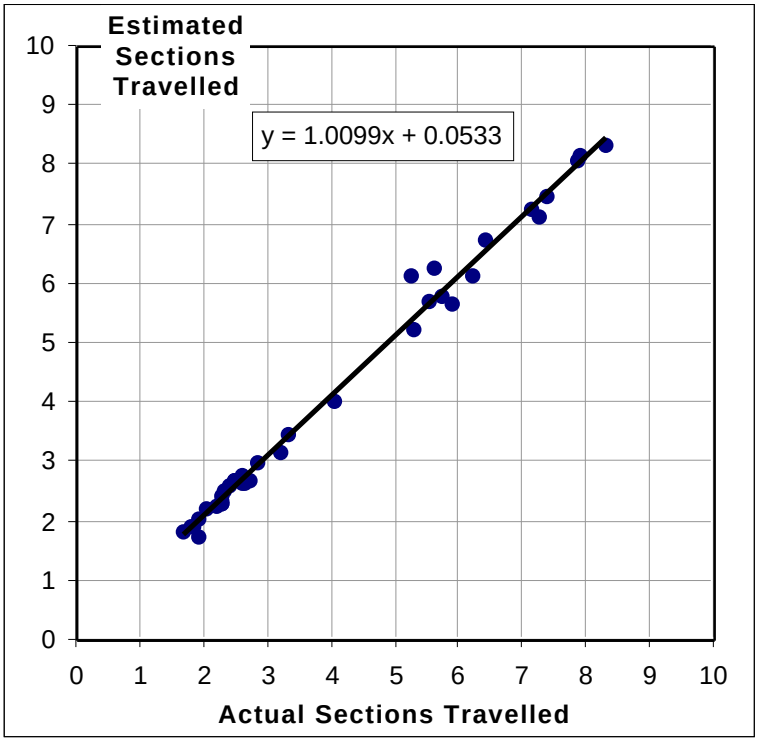

FIGURE 2 Estimated vs Actual Average Sections - Full-Fare Adults

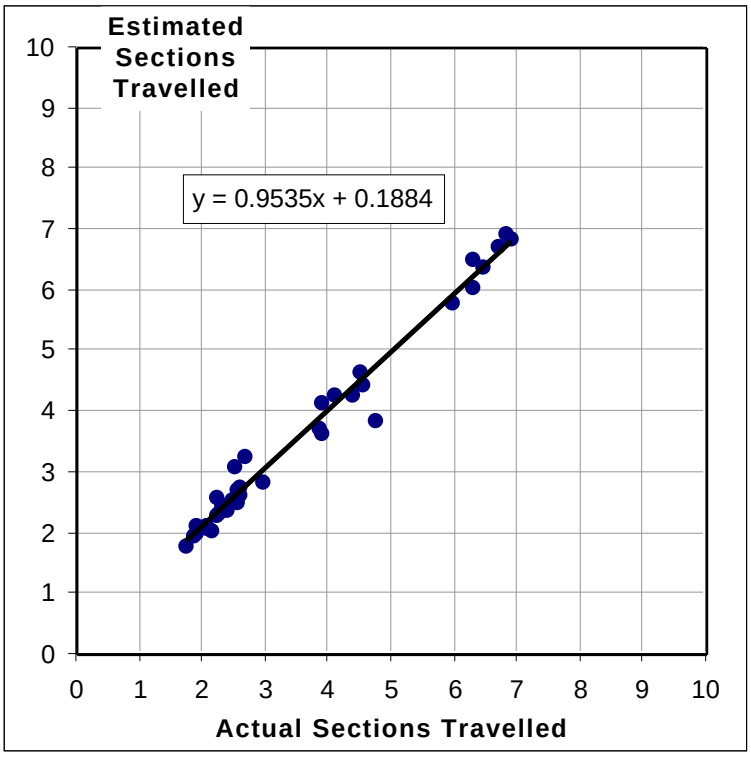

FIGURE 3 Estimated vs Actual Average Sections - Concession -Fare Adults

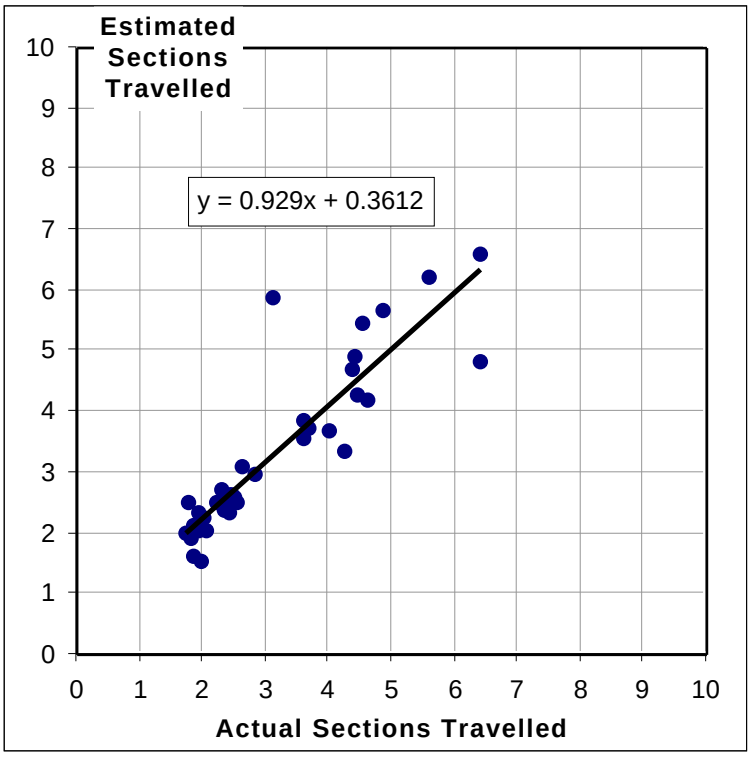

FIGURE 4 Estimated vs Actual Average Sections - Student Fares

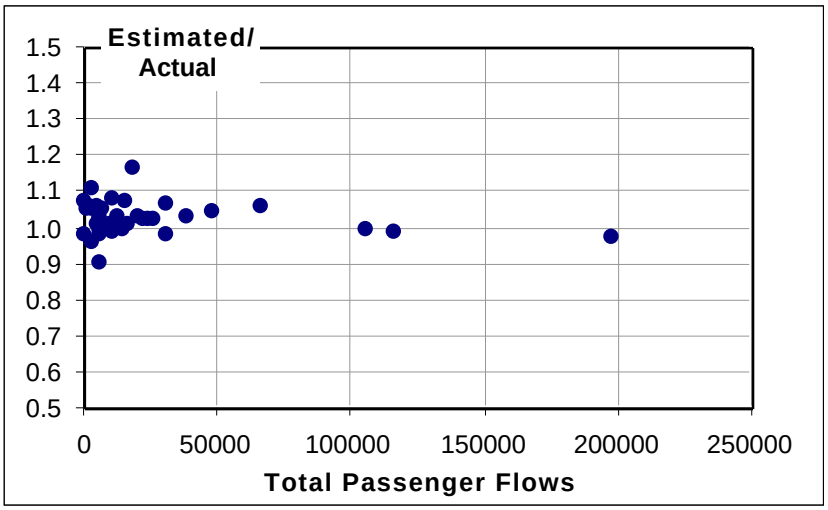

FIGURE 5 Error Ratios vs Passenger Flows - Full-Fare Adults

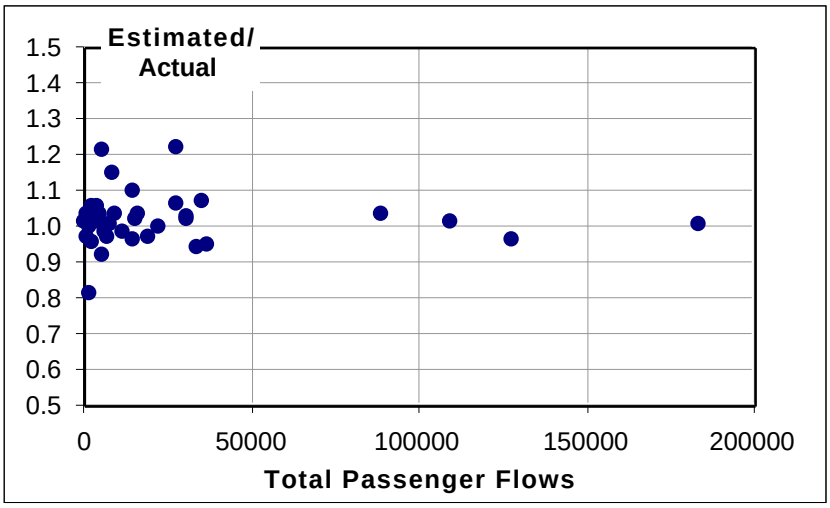

FIGURE 6 Error Ratios vs Passenger Flows - Concession-Fare Adults

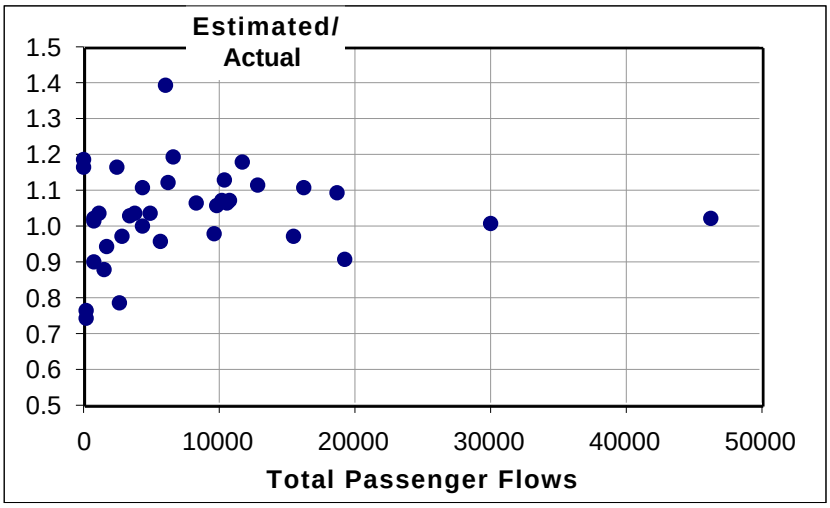

FIGURE 7 Error Ratios vs Passenger Flows - Student Fares

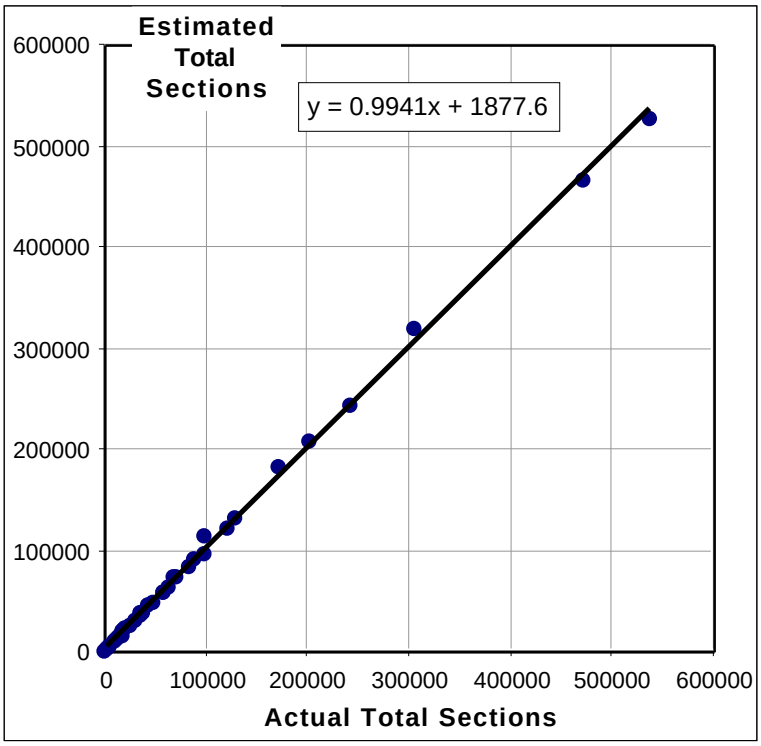

FIGURE 8 Estimated vs Actual Total Sections - Full-Fare Adults

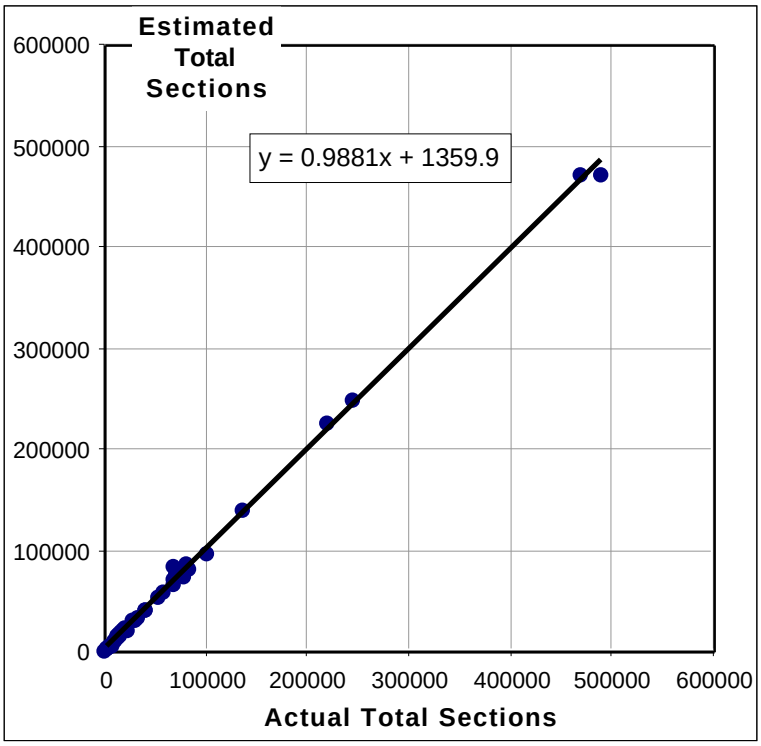

FIGURE 9 Estimated vs Actual Total Sections - Concession -Fare Adults

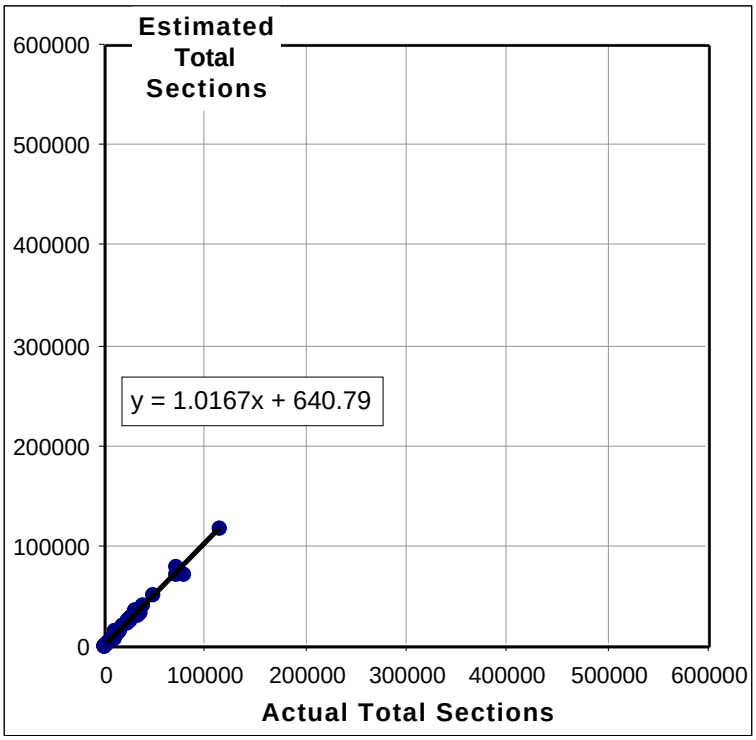

FIGURE 10 Estimated vs Actual Total Sections - Student Fares

LIST OF TABLES

TABLE 1 Summary of Actual Sections Travelled on Route 1

	Working Day		Weekend/Holiday		School Holiday		All Days	
	Daily Flow	Average Sections	Daily Flow	Average Sections	Daily Flow	Average Sections	Daily Flow	Average Sections
Adult								
UP	2563	2.72	792	2.99	1751	2.82	1812	2.78
DOWN	2567	2.62	729	2.97	1545	2.70	1755	2.68
Both	2565	2.67	761	2.98	1648	2.76	1781	2.73
Adult Concession								
UP	2350	2.50	838	2.83	1544	2.69	1686	2.59
DOWN	2306	2.46	796	2.81	1408	2.58	1623	2.54
Both	2329	2.48	817	2.82	1477	2.64	1654	2.56
Student/Child								
UP	777	2.41	109	2.99	147	2.73	425	2.48
DOWN	734	2.41	110	3.22	153	2.89	407	2.52
Both	756	2.41	109	3.10	150	2.81	417	2.50
All Passengers	5649	2.56	1687	2.91	3274	2.71	3851	2.64

TABLE 2 Summary of Estimated Sections Travelled on Route 1

	Working Days		Weekends/Holidays		School Holidays		All Days	
	Daily Flow	Sections Travelled	Daily Flow	Sections Travelled	Daily Flow	Sections Travelled	Daily Flow	Sections Travelled
Adult	2565	2.63	761	2.85	1648	2.68	1781	2.67
Adult Concession	2329	2.51	817	2.67	1477	2.64	1654	2.56
Student/Child	756	2.43	109	3.24	150	3.09	417	2.55
All Passengers	5649	2.55	1687	2.79	3274	2.68	3851	2.61

TABLE 3 Summary of Actual and Estimated Sections Travelled

	Total Passenger Flows			Actual Sections Travelled			Up-Down Sections Travelled		
Route	Adult	Adult Concession	Student / Child	Adult	Adult Concession	Student / Child	Adult	Adult Concession	Student / Child
1	12468	4811	2421	7.15	6.29	4.86	7.22	6.48	5.66
2	12731	16402	16245	5.54	4.10	4.42	5.71	4.25	4.90
3	30246	35314	12806	2.27	2.30	2.23	2.43	2.46	2.48
4	15079	27101	11756	2.34	2.54	2.63	2.51	3.10	3.10
5	3235	3528	5696	5.63	3.92	4.47	6.24	4.15	4.27
6	6060	551	40	2.64	2.58	1.96	2.64	2.63	2.32
7	38308	36412	15516	3.34	2.14	2.09	3.45	2.03	2.03
8	18508	18703	10771	5.26	4.39	4.40	6.14	4.28	4.70
9	5952	1512	2600	7.90	6.69	4.25	8.14	6.71	3.35
10	10402	15367	5955	2.03	2.07	1.77	2.19	2.11	2.47
11	366	323	172	7.25	6.82	6.43	7.11	6.91	4.79
12	14443	14793	19241	5.73	4.57	4.04	5.76	4.41	3.66
13	105650	109371	18702	2.30	2.24	2.04	2.30	2.27	2.23
14	2810	1353	116	5.90	4.74	3.15	5.65	3.86	5.85
15	1580	2258	1773	2.51	2.56	2.45	2.67	2.71	2.31
16	23802	30184	10388	1.84	1.88	1.75	1.88	1.94	1.98
17	363	458	134	2.48	2.26	1.98	2.66	2.34	1.52
18	14692	6202	6534	8.32	6.45	4.57	8.31	6.36	5.45
19	7099	14163	6305	1.81	1.91	1.88	1.90	2.10	2.11
20	5822	5644	4316	5.29	3.91	3.70	5.20	3.61	3.71
21	5463	7349	3345	1.92	2.07	1.95	1.73	2.08	2.01
22	10120	6847	802	6.21	5.96	6.41	6.14	5.80	6.57
23	115623	127107	9714	4.07	3.85	3.61	4.02	3.71	3.53
24	6658	8783	10592	2.29	2.38	2.46	2.31	2.46	2.61
25	197673	183571	46199	2.72	2.56	2.49	2.66	2.57	2.54
26	4341	5254	3840	2.42	2.69	2.84	2.57	3.27	2.95
27	21855	3738	669	2.61	2.59	2.54	2.67	2.62	2.57
28	25657	11793	4283	7.86	6.92	5.62	8.06	6.84	6.22
29	47543	30269	8303	6.42	4.52	3.60	6.73	4.63	3.83
30	30375	33821	2856	3.20	2.98	2.58	3.14	2.81	2.50
31	9657	1002	21	2.61	2.58	2.33	2.64	2.50	2.72
32	66096	88667	29950	2.61	2.47	2.34	2.77	2.55	2.36
33	5500	8038	4868	2.21	2.22	2.48	2.26	2.56	2.57
34	16559	22184	9771	2.27	2.38	2.35	2.30	2.38	2.48
35	19918	27307	10196	2.86	2.59	2.42	2.96	2.75	2.60
36	1271	1956	1201	1.70	1.73	1.81	1.79	1.79	1.88
37	2713	4871	1434	1.91	1.91	1.85	2.01	1.98	1.62
38	4838	2161	692	7.38	6.30	4.62	7.45	6.02	4.17
			Average	3.55	2.95	2.89	3.60	2.97	3.02