

Factor Analysis for the Study of Determinants of Public Transit Ridership

*Sharfuddin J. Syed, Government of North West Territories, Canada
Ata M. Khan, Carleton University*

Abstract

Behavioral studies based on attitude survey questionnaires with numerous variables may be tainted with repetitions and correlations. To overcome these deficiencies, a factor analysis approach is demonstrated that produces clusters of uncorrelated factors. From 47 observable variables contained in the Ottawa-Carleton Transportation Commission (OC Transpo) attitude survey, only 8 underlying factors have emerged. Bus information service is the most important factor. In addition to factor analysis, this article reports on a logistic regression model, based on key factors, for estimating the odds of ridership.

Introduction

Public transportation agencies undertake attitude surveys in order to identify the factors that affect public transit ridership. The questionnaires are designed to elicit responses on a large number of variables. However, analysis of data collected may be deficient due to repetitions and correlations. In order to obtain useable answers from a large number of survey responses, advanced statistical analyses are required.

In this research, a factor analysis approach was defined for removing anomalies and to produce clusters of uncorrelated factors using a large number of observed variables. The identification of the key factors that serve as determinants of public transit ridership would assist in developing policies, strategies, and tactical measures for attracting riders. This article describes the factor analysis methodology and reports its application to OC Transpo's attitudinal survey data. Additionally, the development of a logistic regression model is described that can explain ridership. The identified key factors are used as independent variables to estimate the odds of ridership.

Methodological Framework

The study methodology consists of six components. At the outset, the public transportation characteristics of the National Capital Region (Canada) are introduced. The attitude survey data collected for the OC Transpo are described next. The third section covers the theory of factor analysis and describes the rationale for its application to this research. Factor extraction and interpretation are noted next. The analysis of a large database and tests for reliability and validity are explained. In this part of the research, each individual variable was coded, missing values were discarded, and response-oriented variables were identified. This process yielded 47 variables. By use of factor analysis, these variables were reduced to a small number of individual entities called "factors" and were given labels to suit the current knowledge regarding bus service. The factor analysis results yielded the order of importance of these determinants of ridership.

In the fifth step, the newly labeled factors were used as independent variables for the study of the dependent variable of ridership. The logistic regression analysis enables the estimation of the odds of bus ridership and the relative importance of factors included in the model. In the final section, conclusions are reported and actions to be taken are described.

Public Transportation in the National Capital Area (Canada)

The National Capital Area (NCA) consists of the Regional Municipality of Ottawa-Carleton (RMOC) and the Outaouais Urban Community (OUC). In the RMOC, the OC Transpo provides public transit services for about 0.7 million persons (Figure 1). As compared to other transit agencies serving medium-size urban areas, OC Transpo ranks favorably in terms of transit trips per capita.

A bus rapid transit system is available based mainly on an exclusive bus roadway, the Transitway (Zargari and Khan 1998). In the central business district, exclusive bus lanes are provided in mixed-traffic environments. The Transitway is an access-controlled, grade-separated, two-way rapid transit facility. It enables travelers to minimize the number of transfers through the extensive use of sub-urban express services in peak periods. Stations are designed to provide a comfortable, sheltered, and secure environment for passengers while they are waiting for buses.

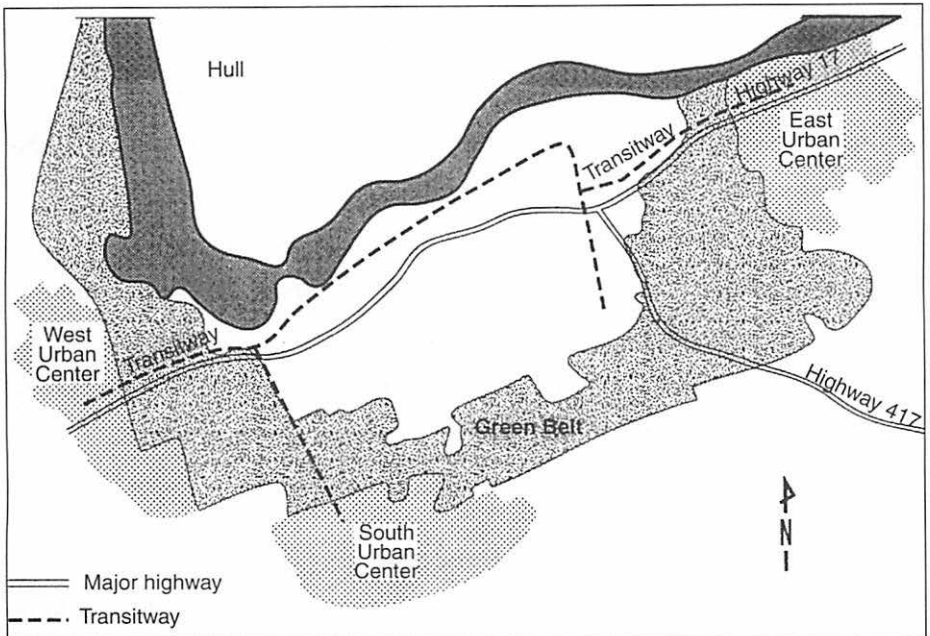


Figure 1. Transitway in the regional municipality of Ottawa-Carleton (Canada)

Bus information is provided through printed schedules and maps, video screens, and dial access to OC Transpo's public information office (Gault 1995). Bus arrival information provided to the public is based on scheduled times. Although in recent years, OC Transpo has installed an automatic vehicle location and control (AVLC) system on selected routes, the information provided by this system is not available to potential users through a state-of-the-art advanced traveler information system.

Attitude Survey Data

Surveys have been used by the OC Transpo since 1988 for tracking attitudes in support of policy formulation, planning, and marketing of services (OC Transpo 1995). The surveys feature questions designed to yield information on user characteristics and attitudes toward services and facilities provided by the transit authority. The OC Transpo's analyses produce weighted frequencies of responses. In this research, the 1995 Attitude Survey data were selected for detailed analysis based on advanced statistical methods.

The survey was designed and conducted by the Carleton University Survey Centre. Trained interviewers fluent in English and French were employed. The sample was extracted from telephone numbers. The survey design took into account representative samples across age groups (18 years and older), and care was taken to ensure proper representation from all parts of the RMOC. Prior to actual use, the English and French versions of the questionnaire were tested. A total of 2,000 telephone interviews were completed. According to the Survey Centre, the sample was representative of residents living in areas served by public transit and is accurate to 3 percent (19 times out of 20).

Factor Analysis: Technique and Rationale

Factor analysis as an advanced statistical technique was described by a number of authors. See, for example, Kim and Mueller (1981) and Harman (1976). It is an efficient performer of data analysis and achieves variable reduction by

defining a set of factors extracted from the variables in the database. The principal component analysis was chosen for this purpose. Factor loadings are obtained by squaring the correlation coefficient of the variables concerned. In a factor model, the factor loadings and their signs are important items of information. The highly loaded factors provide a reliable basis to explain transit ridership. Therefore, factors with high loadings are retained in the model structure and the lightly loaded are discarded. The threshold for such retention or discarding depends on the percentage of variance explained by the respective indicators.

The attitude toward ridership of public transit is dependent on variables that in the original version may be tainted with shallow perception of the questions answered, indifference, or mere capitulation to a “sought-after” question clearly manipulated. Some of these questions and answers may simply be vague. Factor analysis is expected to overcome these anomalies by creating clusters of observed variables.

Factor Extraction and Interpretation

For factor extraction, Cattell’s Scree Test was used. As shown in Figure 2, it is a plot of unrotated factors versus the proportion of total variance or eigenvalues. In the scree plot, there is a negative slope, decreasing from factor number 1 to 6 and then it seems to level off and slope further down and level off again. The point corresponding to the first scree may initially determine the number of factors that should emerge. However, subjective judgment is required to determine the level-off point. Details on factor extraction process are presented below.

Tables 1 through 8 show clusters for selected factors. The factors, their respective share of loading (variance), the original variables, and their factor loadings are also shown. In these tables, 35 out of 43 variables are accounted for by these factors.

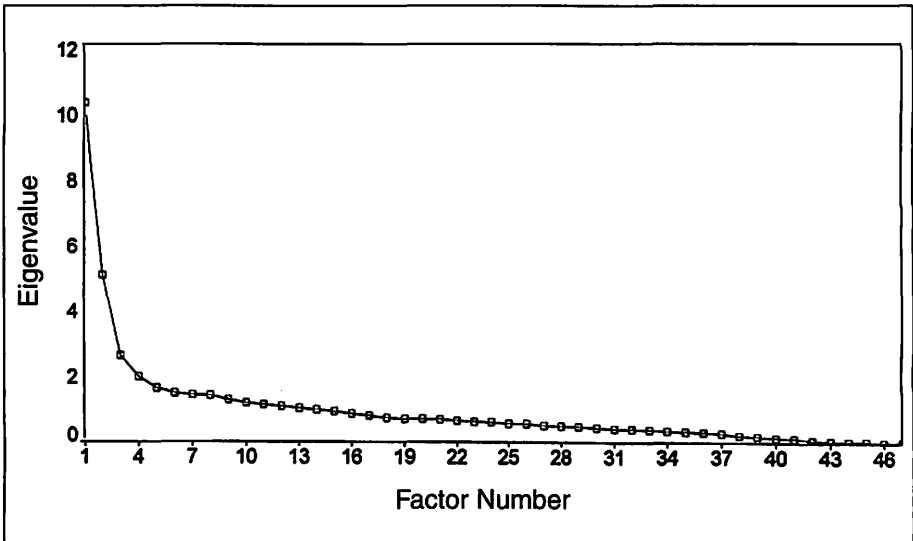


Figure 2. Factor scree plot

Table 1		
Bus Information Cluster		
<i>Factor and % Loading</i>	<i>Original Variable Descriptions</i>	<i>Factor Loading</i>
Bus information 22.1% ^a	Do you call customer service?	0.93805
	Do you use the timetable?	0.91546
	Called OC Transpo Information Centre	0.89160
	Drivers describe stops	0.88163
	Use TV screen	0.87487
	Have system map	0.87453
	Pick up other information	0.87021
	Heard of Easier Access Program	0.87012
	Use 560 telephone system	0.86284
	Use printed schedules at stops	0.85484
	Bus information	0.70452
	Information accurate	0.59465
	Yes—seen bus kneeling ^b	0.54846
	How often (560 telephone system)?	0.41828

a. Percentage of variance explained by the factor.

b. Kneeling is a hydraulics function whereby the bus is lowered down so that its first step lines up with the curb level for easier access.

Table 2 On-Street Service Cluster		
<i>Factor and % Loading</i>	<i>Original Variable Description</i>	<i>Factor Loading</i>
On-street service 10.9%	Bus service unreliable	0.73673
	Bus frequency adequate	0.73021
	Bus routes well planned	0.58765
	Taking bus frustrating	0.58233
	Bus not easy to use	0.57367
	Transfers are convenient	0.46835

Table 3 Customer Service		
<i>Factor and % Loading</i>	<i>Original Variable Description</i>	<i>Factor Loading</i>
Customer service 5.6%	Information accurate	0.87169
	Information courteous	0.86992
	Information Centre easy to get through	0.73616

Table 4 Cleanliness Cluster		
<i>Factor and % Loading</i>	<i>Original Variable Description</i>	<i>Factor Loading</i>
Cleanliness 4.3%	Transitway stations are not kept clean	0.69639
	Bus shelters clean	-0.66766
	Buses not kept clean	0.74036

Table 5
General Attitude Cluster

<i>Factor and % Loading</i>	<i>Original Variable Description</i>	<i>Factor Loading</i>
General attitude 3.5%	Taking buses benefits the environment	0.80120
	It is reasonable for people to stand sometimes on rush-hour buses	0.59456
	No matter where I live, I will always avoid taking the bus	-0.5090
	The Transitway has improved bus service for me	0.42273

Table 6
Transitway Station Safety Cluster

<i>Factor and % Loading</i>	<i>Original Variable Description</i>	<i>Factor Loading</i>
Transitway station safety 3.2%	Transitway station safe	0.83010
	Why not use Transitway station	0.80022

Table 7
Security Cluster

<i>Factor and % Loading</i>	<i>Original Variable Description</i>	<i>Factor Loading</i>
Security 3.1%	Personal security at night in multistoried parking lots	0.66647
	Taking bus at night no threat to security	0.63321
	Personal security at night while walking in neighborhood	0.56441

<i>Factor and % Loading</i>	<i>Original Variable Description</i>	<i>Factor Loading</i>
Reduced fare 2.6%	Peak-hour trip too expensive Bus pass good buy Good deal during off-peak hours	0.75583 0.42952 0.39000

Factor Extraction Process

By rule of thumb, the factors with eigenvalues equal to or greater than 1 are included in the factor model and their factor loadings are calculated. Another procedure to determine the number of factors for a factor model is by drawing a plot of eigenvalues as the ordinate with the respective variables as the abscissa. There will be a noticeable scree (i.e., a horizontal leveling off followed by another slope downwards). This corresponding number of factors is taken into the factor model.

Next, a factor matrix is created with all the variables sharing these factors. The coordinates are rotated orthogonally, say 45 degrees, for simplification of observation and interpretation. From the rotated table, if the loadings are sorted in order of magnitude and the lower values of loadings are blanked out, distinguishable clusters should emerge.

The premise of factor analysis is that the correlation between variables is due to a sharing of common factors. Factors may be orthogonal (i.e., they are located in an axis perpendicular to each other or they may be oblique, in which case their axes are not perpendicular with one another). In this research study, orthogonal analysis has been done using the Varimax rotation method.

The ridership index was analyzed for reliability. It shows an 89 percent scale reliability. Next, the underlying dimensions of ridership were examined. This process required the assessment of the correlation matrix of all 47 variables. To share common factors, the correlation between the observed variables must be

high. Another step was a sample adequacy test. The Kaiser–Meyer–Olkin (KMO) measure of sample adequacy is derived from correlation information and shows if the data are suitable for factor analysis or not. Values of this measure above 0.7 can be regarded as acceptable. The following results were obtained in this analysis:

KMO Measure of Sampling Adequacy = 0.79589

Bartlett Test of Sphericity = 12835.64; significance = 0.00000

The Bartlett Test of Sphericity examines the hypothesis that the correlation matrix is an identity matrix with a diagonal as 1 and off-diagonal values as 0. In this case, this hypothesis is not true. Therefore, factor analysis is feasible with the 1995 database.

At this stage of analysis, the factor extraction procedure commences. As noted previously, there are several methods of this extraction. Here, the principal component analysis was chosen and each variable was standardized to have a mean of 0 and a standard deviation of 1 and therefore variance of 1. The factor loading was obtained by squaring the correlation coefficient of the variable concerned, which denotes the variance of the variable. In the case of principal component analysis, total variance was calculated. After factor extraction, the communalities have been recalculated for finding the actual variance.

In the next step for factor extraction, Cattell's Scree Test was used (Figure 2). The Varimax Orthogonal rotation was carried out. Compared to the unrotated matrix, this rotated matrix has resulted in a more interpretable set of constructs. This was made possible by means of Varimax rotation, which converged in 15 iterations. The Varimax rotation used maximization of variance of squared loadings for each factor.

The final step was to calculate the factor scores. This was done by studying the product of the standardized value of a variable and its factor score coefficient,

and summing them up separately under each factor for each variable. Another test, the Direct Oblimin method, was used to determine if the fit between data and factor structure could be improved. However, no improvement occurred. In obtaining the structure, the lower magnitude (≤ 0.39) loadings were blocked out in the rotated matrix in order to show the clusters more distinctly.

Logistic Regression Analysis

A logistic regression model was developed to obtain the odds of ridership. This mathematical model for factor analysis is somewhat similar to a multiple regression equation. However, each variable is expressed as a linear combination of factors that are not actually observed (Pedhazur 1982).

In this model form, the dependent variable is binary in nature with two probable values (i.e., 1 = ridership occurring, 0 = not occurring); the multiple linear regression analysis and the discriminant analysis techniques pose difficulties (Norusis 1990). Therefore, logistic regression was chosen as the appropriate technique in this analysis.

A number of models were developed and tested through the likelihood ratio by introducing one factor at a time and noting the change in the log likelihood. The likelihood ratio tests the null hypothesis that the coefficient for the last variable added to the model is 0.

Table 9 shows the model that incorporates the top eight factors. The odds ratio is used to assess the effect of the independent variable on the dependent variable in terms of the likelihood of the event that is coded as 1 (in this case it is ridership). The value that is greater than 1 indicates a positive effect of an independent variable on the dependent variable. The value that is less than 1 indicates a negative effect. The maximum-likelihood method was used to find the coefficients for the factors included in the model.

The goodness of fit of the model was examined by deriving the Pseudo R^2 square and R^2 . The change in Pseudo R^2 , due to inclusion of a factor in the

Table 9
Logistic Regression Model (dependent variable: ridership)

Interpreted Factor Label	Odds Ratio
Bus information	91.221 ^a
On-street service	2.1712 ^a
Customer service	1.6885 ^b
Cleanliness	1.0769
General attitude	0.7376
Safety: Transitway station	1.7356 ^b
Safety: En route	1.0814
Reduced fare	1.3611
Pseudo R^2	0.800
R_L^2	0.523

a. $p < .05$; two-tailed- t test.

b. $p < .10$.

regression model, is an indicator of the unique information provided by that variable. These are used as indicators of relative importance of the respective factors as per their magnitudes (Table 9).

The high importance of a particular factor reflects respondents' perception of attributes of public transportation. The inclusion of questions on attributes is based on management's requirement for information. For example, bus information-related questions in the OC Transpo's attitude survey are due to the company's desire to learn about the impact of some of the new services provided by them. The high relative importance of that factor reflects attitudes of respondents. The combined results obtained from the factor and logistic models are mutually reinforcing in terms of establishing factor ranks.

Conclusions and Implications

This research has identified key factors on the basis of factor analysis for making urban transit bus ridership more attractive to RMOC residents. On the basis of percentage variance explained by various factors, the following is the hierarchy of factors:

- bus information, 22.1 percent;
- on-street service, 10.9 percent;
- customer service, 5.6 percent;
- cleanliness, 4.3 percent;
- general attitude, 3.5 percent;
- transitway station safety, 3.2 percent;
- personal security, 3.1 percent; and
- reduced fare, 2.6 percent.

On the other hand, the logistic regression results show a slightly modified order of importance: bus information, on-street service, station safety, customer service, safety en route, reduced fare, cleanliness, and general attitude. Taken together, a knowledge of these factors could be used to develop strategies, tactics, and operational measures to enhance ridership.

Bus information service is the most important factor cluster compared to others. The implication of this result is that all state-of-the-art user information systems that are not currently available in the Ottawa-Carleton region should be implemented on a priority basis. For example, real-time bus arrival information should be provided at stations, other locations such as shopping centers, and also through the telephone system (Eno Transportation Foundation 1998; *Urban Transport News* 1997; U.S. Department of Transportation 1996). This can be achieved by using Intelligent Transportation Systems (ITS) technologies, for example, based on Global Positioning Systems (GPS). Without a high-quality pervasive bus information system, it would be difficult even to keep regular bus users in the fold of bus transit.

On-street service is next in the order of relative importance. It includes reliability of service, adequate bus frequency, well-planned routes, convenience at transfers, level of service, etc. Potential riders, though unconsciously, will assess the conditions of the various components of this service before switching from cars to bus. To provide an excellent on-street service, the bus company should increase its fleet of buses so that adequate frequency of service can be maintained. All these new purchases should include a variety of buses to suit different demands at different times and places. Travel time should be within a threshold of tolerance, determined through surveys. Technological and institutional concerns, operational and environmental performance, financial performance, standard of service including passenger waiting time, walking distance to stops, the need to interchange between routes and services, passenger journey time, passenger travel expenditure, and affordability are to be incorporated in the on-street service planning. Expansion of the rapid transit service by building Transitway extensions and reduction of congestion in mixed-traffic environments in downtown would help in enhancing on-street service.

Safety and security concerns should be addressed in the Transitway stations, in the access and egress routes, and inside the bus itself, especially at night. This would improve ridership. Safety concerns in Transitway stations, inside buses, and in the neighborhoods can be further improved by implementing in full, measures such as emergency push-button telephones at all stations, vigilant security patrols at night in remote areas, and improved street as well as platform lighting at current sites as well as new ones.

Customer service improvements are necessary to make riders happy and to improve general attitudes toward bus transit service. Bus operators, ticket sellers, telephone operators, and others responsible for customer service should be courteous to passengers. Customer service should be given a high profile while planning strategies to enhance bus ridership.

According to the model, a reduced fare should improve ridership consider-

ably. Although it is difficult to implement, given that public transit is a subsidized service, it deserves a closer examination. Keeping fares low through efficiency improvements is a worthwhile goal. Employers may offer transit fares as incentives instead of free parking. Fare reduction may also be achieved by transitioning to privately run, high-efficiency transit services.

The cleanliness factor has a number of implications. Riders prefer to travel in a bus that is sprightly clean. Also, the bus should be plying on roads that are kept clean from snow, dust, and dirt. Additionally, buses should be perceived as environmentally clean. This could be achieved by using low-pollution, alternative-fuel buses.

The general attitude can be positively influenced through carefully designed marketing means such as advertisements, which portray the numerous benefits of using public transit (Deka 1996). The benefit that comes to the forefront is the environmental one. The public's general attitude is also influenced by, among other things, comfort and convenience and other level-of-service (LOS) features. For example, if people perceive that using mass transit not only frees road space but also removes emissions, noise, and accident hazards, ridership should get a boost.

The public's general attitude cannot be changed easily. For this, direct intervention will be necessary. For example, advertising the benefits of using public transit can motivate people to use this mode. Also, information services can incorporate some advertisement features through ITS technologies. Imposing charges on single-occupancy-vehicle users can also turn around a negative attitude toward public transit.

Acknowledgments

The Natural Sciences and Engineering Research Council of Canada provided financial support for the research reported in this article. The authors gratefully acknowledge the availability of the OC Transpo's 1995 Attitude Survey data. The views expressed in this article are those of the authors.

References

- Deka, Devajoyti. Fall 1996. Public transit in the social marketing framework. *Journal of Public Transportation* (1).
- Eno Transportation Foundation Inc. January 1998. European transit philosophy—Customer is the king. News.
- Gault, Helen, E. 1995. Complementing transit in Ottawa–Carleton—Intermodal partnerships for bus rapid transit in the nation's capital. *Proceedings*. Canadian Transportation Research Forum.
- Harman, Harry, H. 1976. *Modern factor analysis*. University of Chicago Press.
- Kim, Jae-On, and Charles W. Mueller. 1981. Factor analysis: Statistical methods and practical issues. University of Iowa, a Sage University paper.
- Norusis, Marija. 1990. *SPSS-based system users' guide*. SPSS Inc.
- OC Transpo. 1995. *Attitude survey 1995*. Data Centre, Department of Journalism, Carleton University, Ottawa. Canada.
- Pedhazur, Edgar, J. 1982. *Multiple regression in behavioural research: Explanation and Prediction*, 2nd ed. New York University.
- Urban Transport News*. December 1997. Transit agencies could be partners in new traffic information services 25 (25).
- U.S. Department of Transportation. 1996. *Traveler information systems: Available information allows better choices: Intelligent transportation infrastructure*. Washington, DC: Federal Highway Administration.
- Zargari, S. A., and A. M. Khan. 1998. A simulation model of the bus transitway (TRNSIM). *Journal of Advanced Transportation* 32 (2).

About the Authors

SHARFUDDIN J. SYED received his master's degree in civil engineering (transportation) from Carleton University, Ottawa, Canada. He worked for OC Transpo for a number of years. At present, he is a project engineer for the Inuvik Region, Department of Transportation, Government of North West Territories.

ATA M. KHAN (ata_khan@carleton.ca), received his doctorate from the University of Waterloo (Canada) in civil engineering (transportation). He is a professor in the Department of Civil and Environmental Engineering and director of the Transportation Research Centre, Carleton University (Canada).