LRT. 84.4-

749

Transpn. Res.-A Vol. 18A, No. 2, pp. 135-140, 1984
Printed in Great Britain.

0191-2607/84 $3.00 + .00

THE DETERMINANTS OF LIGHT-RAIL TRANSIT DEMAND—AN INTERNATIONAL CROSS-SECTIONAL COMPARISON

PETER GORDON and RICHARD WILLSON
University of Southern California, Los Angeles, CA 90089, U.S.A.

Abstract—An international cross-section of light-rail transit systems is examined. Demand is tested as a function of transport system, city, and population attributes. All tests confirm the standard hypotheses at high levels of significance. High levels of explanatory power support the notion of model transferability. The models developed in this paper are used to predict demand for new LRT systems now being installed in North American cities. Model outputs suggest that the official forecasts are very optimistic.

INTRODUCTION AND BACKGROUND

Though travel demand forecasting is by now one of the largest fields of applied urban research, it is still subject to criticisms. Various authors have noted conceptual problems (McGillivray, 1972) found in the conventional forecasting models; others have challenged the forecasts produced by these models (Wohl, 1976); Wachs (1982) has added the complaint that the context in which these models are used is open to certain ethical dilemmas. We agree with all three challenges and add that surprisingly little is as yet known about how the demand for transit trips (via the light-rail mode, here) interacts with city descriptors as well as with characteristics of the population and transit system referents.

This study uses an international cross-section of cities to investigate those relationships. A single-equation framework is used since there is little or no relationship between transit provision (supply) and the relevant economic variables: there is little chance of an identification problem since supply is independent of fares charged as well as the other descriptors. Rather, transit provision is assumed to respond to political factors. Put another way, fixed rail transit has been found to be so inefficient in the modern city that its existence can not really be rationalized by economic behavior (see Hilton, 1968). We are left with a simple relationship between the various system characteristics and levels of usage.

The wide-ranging international cross-section allows us to comment on the possible international transferability of the hypothesized relationships. Inded, some remarkably stable relationships are found to exist over a quite heterogenous sample. Further, the models developed in this study are used to make demand forecasts for light-rail systems now being developed or considered in six North American cities. The models allow for the computation of statistical confidence intervals on projected demand levels. It will be seen that these provide some perspective on the conventional forecasts: not only are they found to be very optimistic but many can be characterized as outliers with a very low probability of occurrence. Given the criticisms of conventional techniques already mentioned, this result enhances the credibility of our models.

To date, very little research has been directed toward the sort of international comparison suggested here. Pushkarev and Zupan (1980) do consider fixed guideway performance in a cross-section of North American cities. Yet, many of their statistical estimations are over a sample of diverse fixed guideway modes; they do not report testing whether this sort of pooling creates a misspecification. In addition, many of their statistical correlations involve simple two-variable models, suggesting the great likelihood of bias due to omitted variables. Finally, much of their work appears to be polemical tract masquerading as research; they take the re-centralization of cities as a presumed goal and look for transit schemes that would implement that goal. In so doing, they are pre-judging a major controversial issue: just what are the costs and benefits of centralization vs dispersed settlement patterns?

Webster and Bly (1980) offer results from an international collaborative study of public transit. Yet, that study dwells on a comparison of averages for various countries and is not really comparable to the investigation reported here.

Gordon (1982) conducted a study very similar to the present investigation but for the heavy-rail mode. Indeed that study came to many of the same conclusions which are reached here: conventional predictors of demand, such as per capita incomes, urban population densities, etc., explain a large proportion of demand and these relationships are remarkably stable over a diverse cross-section of cities. Further, predictions from these estimations appear as more reasonable than those derived via conventional travel demand modeling.

DATA AND METHODOLOGY

Light-rail transit modes are usually defined as all railbound modes serving intra-urban surface travel demand, sometimes involving exclusive rights-of-way or elevated or underground sections of track. The mode is distinguished from "heavy rail" by much lower passenger carrying capacities, lower speeds, the capacity to operate at grade intersections, and lower operating costs.

The main data source for light-rail usage and system characteristics is the *UITP Handbook of Urban Transit*, which reports data for systems in 152 cities world-wide.† The *Handbook* offers a consistent data source for a large international sample; the choices of predictor variables are constrained in some instances in that many values for many cities remain unreported.

The composition of the sample provided by the *UITP Handbook* reflects the widespread use of light-rail transit in Europe. Western European countries account for 61 observations, while Eastern European countries account for 36 observations. The remainder of the sample consists of the following: USSR–22; Japan–12; North America–10; others–11.

The dependent variable of this study is demand represented by passengers per day per kilometer of line. The predictor variables are summarized in Table 1. While we did consult other sources to complement the *UTIP Handbook*, some major indicators of urban structure (CBD floorspace and employment, for example) are still not at hand. Results reported here must be regarded as exploratory, therefore. The strong statistical associations found, however, justify attempts to broaden the data base and carry on. Whereas city population density is so far the main urban form descriptor used, we were able to include variables which describe transit system characteristics, and measures which describe the availability of competing modes.

There are several caveats to keep in mind as usual. The main stipulation here is that although "corridor" population densities (once "corridor" can be defined such as to be consistently useful) are preferable to our (available) citywide populations densities, we are able to use our models to forecast ridership for light-rail lines to be located in "well-placed" alignments. That is, the observations of the sample can be presumed to reflect route alignments which have been advantageously placed *and* which have had the time to attract some complementary proximate land uses, including high density development. Thus, the estimated coefficients of our city population density variable must be interpreted accordingly.

†Union International des Transports, *Modern Light Rail*, 44th International Congress, Brussels, 1981.

MODEL RESULTS

The results of the estimations are summarized in Table 2. Models 1 and 2 are in linear form, while models 1A and 3 are in semi-logarithmic form. All four models show coefficients with expected signs and t-values associated with high levels of statistical significance. The coefficients of determination are high despite the wide-rangeing international nature of the sample.

Model 1 is dominated by city population density, which is positively related to ridership, as expected. Gross national product per capita is negatively related to demand, as expected, and represents the opportunity to use an alternative mode of transportation, such as the automobile, Average stop distance is a measure of quality of service, and is negatively related. The U.S. dummy variable is also negatively related, reflecting cultural and historical factors, including inexpensive fuel. Missing values limit the sample size to 52.

Model 1A used a log transformation on the dependent variables. We note significantly higher t-values on many independent variables, apparently accounting for a non-linearity that exists in Model 1.

Per capita automobile registration was substituted for GNP in Model 2, and in combination with city population density and average stop distance, also produced expected signs and significant t-values. Per

Table 1. Variable definitions[1]

PASSKM:	passengers per kilometer of route per day
CDEN:	city population density
GNP:	gross national product per capita[2]
PCCAREG:	per capita city automobile registration[3]
STOPDIS:	average distance between stops, measured in kilometers[4]
US:	dummy variable denoting U.S. cities
COMM:	dummy variable denoting East-bloc observations

Notes: 1) data collected from the 1980 UITP Handbook, which compiled data available to transit operators in late 1978

2) gross national product per capita data is national data; U.S. cities use regional income data

3) automobile registration data were also collected for metropolitan areas; vehicle registrations were collected for city and metropolitan areas

4) other system descriptor variables collected include: percent exclusive right-of-way; total daily system passengers; average speed; existence of a subway system; and maximum and minimum fare.

Table 2. Estimations for 152-observation sample (t-values in parentheses)

MODEL	DEPENDENT VARIABLE	PREDICTOR VARIABLES						R^2	$\bar{R}^2$	N
		CDEN	GNP	PCCAREG	STOPDIS	US	COMM			
1.	PASSKM	0.311 (7.26)	-0.255 (-1.97)	-	-974.531 (-1.58)	-1348.137 (-1.82)	- -	0.72	0.69	52
1A	LPASSKM	4.46×10^{-5} (3.34)	-9.78×10^{-5} (2.21)	-	-0.466 (-2.21)	-0.859 (-3.30)	-	0.58	0.54	52
2	PASSKM	0.355 (6.90)	-	-7155.437 (-2.67)	-1260.264 (-1.63)	-	-	0.76	0.74	37
3	LPASSKM	4.27×10^{-5} (2.94)	-8.99×10^{-5} (-2.15)	- -	- -	-0.740 (-2.88)	0.440 (2.51)	0.45	0.43	91

*an L preceeding any variable refers to its natural logarithm

capita *city* auto registration was found to have more explanatory power than comparable *metropolitan* data. Using the auto registration variable reduced the size of the sample to 37 because of missing values.

Model 3 is in semi-logrithmic form and uses city population density, GNP per capita, as well as dummy variables for United States and East-bloc observations. These new variables have coefficients with the expected signs and are highly significant.

Other data collected for the study included variables beyond those indicated in Table 2. The convention adopted was to run each regression with other possible independent variables and delete those with t-values of less than one. Dhrymes (1970) reports that this procedure has a good chance of maximizing adjusted R-squared values.

Maximum or minimum fares, when used as a replacement for GNP, had the expected negative sign but did not have a large enough t-value. Average distance between stops was found to be superior to average speed as a measure of quality of service. Proportion of line with exclusive right-of-way, or measures of other available modes of transit were also tested but did not have as much explanatory power as average stop distance. Sufficient data were not available to test more detailed measures of city structure, such as corridor or CBD densities.

Model 1A directly yields elasticity estimates. They are: 0.2 for city density; -0.53 for income; -0.28 for average distance between stops; and -0.58 for the U.S. dummy variable. Elasticities in the other models were found to be similar to those in Model 1A. Auto registration had an elasticity of -0.54 in Model 2, while the dummy variable for East-bloc observations had an adjusted elasticity of 0.55.† These results are similar to the elasticities found by Gordon (1982).

†The elasticity of the dummy variable has been adjusted to permit proper interpretation in a semi-logarithmic equation (Halvorson and Palmquist, 1980).

INFLUENCE STATISTICS

We were able to compute a variety of regression diagnostics in an attempt to learn more from our sample. The simplest of these is the studentized residual, formed by re-estimating the model with dummy variables introduced for each observation, there being as many re-estimations as there are observations. Four outliers were found by this technique. They were: Hong Kong (t-values of 3.12), Kosice (2.79), Bern (2.32), the Hague (-2.03). All other similar values were substantially lower.

Are there valid statistical grounds for detecting outliers in order to delete them from the sample? Additional diagnostic statistics shed light on just why these four observations were outliers from the estimated model. For example, Belshley, Kuh, and Welsch (1980) report that various "DF-Betas" are computable which measure the effect that any single observation has had on the values of coefficients estimated in the original regression. They also note that "cutoff values" equal to $2/\sqrt{N}$ (where N is the sample size) can be used to judge which DF-Beta values signal a significant observation.

Using this information and returning to the case of Hong Kong, we see that of the four coefficients estimated, city population density has the very high

DF-Beta value of 1.19 (vs a "cutoff" of 0.28). In fact, this was the largest of all DF-Beta's computed in all of our work on this sample. Looking at Hong Kong's actual population density, we find it is 5.5 standard deviations above the sample mean; Hong Kong daily passengers per kilometer (the dependent variable) was also 4.8 standard deviations above the mean. Analysis of the regression residuals, then, revealed that this extreme case had a significant effect on the estimated coefficient of city population density; further study of the data revealed why this may have been so.

We concluded that Hong Kong should either be dropped from the sample or that it should be included only when the specification is non-linear such that extraordinarily high transit use as a consequence of extraordinarily high population densities can be explained by the model. Model 1A shown in Table 2 is in a semi-logarithmic form and takes into account the non-linearity introduced by Hong Kong.

Likewise, Bern, Switzerland, was an outlier, as evidenced by the studentized t-statistics. In this case, the DF-Beta for GNP per capita received a large positive value, indicating that large ridership is associated with large per capita incomes, an anomaly for the linear model. The prescription is, again, either specify a non-linear model which can account for the behavior of the Bernese or drop them from the linear model.

Similar considerations caused us to drop Kosice, Czechoslovakia, and the Hague, Netherlands, from the sample. Results from the restricted sample are displayed as part of Model 1B, where the statistical fit is, or course, superior to that of the other models. Model 1B:

$$\text{PASSKM} = \text{intercept} + \underset{(8.04)}{0.283(\text{CITY DENSITY})} - \underset{(-2.23)}{0.230(\text{GNP})} - \underset{(-1.98)}{936.3(\text{STOPDIS})}$$

$$- \underset{(-2.17).}{1227.81(\text{U.S.})} \qquad R^2 = 0.77, \bar{R}^2 = 0.75, N = 48$$

What is worth repeating is that this procedure for restricting the sample is not contrived as part of curve-fitting by brute force. Rather, cases where there is evidence that they depart from the linear specification for good reason are identified and removed. The results yield an estimation from a sample of cases where there is no evidence that the observations depart significantly from the hypothesis of linear association.

FORECASTS

The model's developed in this paper were used to forecast demand for a number of North American cities that are either constructing or planning for light-rail transit development. As pointed out in the second section, these forecasts are for a "well-placed" line in each of the respective cities; one would have to demonstrate unusual characteristics to justify a specific forecast that would exceed those provided by the models proposed in this paper.

Table 3 summarizes the values of the predictor variables, which have been obtained from transit publications and personal communications. The resulting comparative forecasts are shown in Table 4, and provide fairly consistent results. Calgary's light rail system is already in operation, and provides a good check on the performance of our models. Models 1A and 3 provide remarkably close estimates to the ridership experienced there. Calgary's original estimate of ridership is almost twice what has been achieved—this sort of over-prediction has been observed in many other North American cities for both light-and-heavy rail ridership forecasts. If we accept that the models proposed in this paper provide a more realistic general assessment of ridership potential, then it appears that the projections for cities now considering light-rail are also overstated.

Models 1 and 2, which are linear, indicate negative ridership projections for certain cities. Obviously, there should be a non-linear specification toward the very low end of predictor variable values suggested by some non-sample cities.

Confidence intervals have also been computed for the forecasts. Problems were encountered with the intervals computer for models 1 and 3 because they are linear and in some cases predict negative values for cities with low potential for light-rail. Table 5 shows the confidence intervals for Models 1A and 3 at the 95% level. Both the narrow intervals for mean forecasts and the wider intervals for individual forecasts are shown. We note that in many instances, the official yearly forecasts are beyond the upper edge of our models' predictions. We conclude that this forms a basis for the belief that many of the official forecasts are, indeed, unlikely to be realized.

CONCLUSIONS

Cross-sectional work in the social sciences is especially interesting and especially timely. It is worthwhile because the question of the transferability of various hypotheses between settings remains to be fully tested. It is timely because industrialization and urbanization in the rest of the Western countries has matched (and at times surpassed) U.S. levels only recently, making such comparisons possible.

This study makes the point that certain simple relationships between city, population, transportation system descriptors and light-rail transit demand are remarkably stable over a diverse inter-

Table 3. Values of predictor variables for selected North American light rail systems

City	City Density (per km^2)	Income per capita	City Auto Reg per capita	Stop Distance (km)
Calgary	1247	6930	0.42	0.81
Detroit	3616	7648	-	1.41
Los Angeles	2332	7699	0.59	1.16
Portland	1506	7505	-	0.96
Seattle	2287	7714	-	0.65
Vancouver	3627	6930	0.46	1.81

Notes: predictor variables obtained from the UITP Handbook; Modern Railway Annual Transit Digest; City and County Fact Book; and Census of Canada.

Table 4. Comparison of light rail forecasts (passengers per day per km)

SYSTEM	OFFICIAL FORECAST	Model 1	Model 1A	Model 2	Model 3
Calgary 12.9 kms	1782 (actual) 3100 estimated	1482	1610	242	1772
Detroit 23.9 kms	6485 estmated	103	531		871
Los Angeles 36.2 kms	1464 by 2000	-65	563	-1052	828
Portland 24 kms	1628 by 1900	-78	607		812
Seattle 2.1 kms	N/A	413	712		820
Vancouver 21.7 kms	4608 estimated	1247	1123	-561	1958

notes: Official Forecast derived from Modern Railroads' 1983 Annual Tranportation Digest.

national cross-section of cities. Certainly much work remains to be done since many of our predictor variables can be replaced by more specific indicators of settlement patterns, economic well-being, and transit system performance. Nevertheless, we confess to being astonished by the strong statistical associations that have been found. At this point it is possible to say that there is quite a bit which is transferable.

We have also tried to make the point that the relationships detected can serve as a simple and a fairly reliable forecasting tool. Perhaps these are the two attributes that are most needed given Wachs' indictment of current practice in the travel demand forecasting field.

Otto von Bismark is alleged to have said that those who enjoy legislation or sausage should avoid watching either one being prepared. Had he known about conventional travel demand forecasts, he might have included them as a third item to be wary of. The

Table 5. 95% confidence intervals (for mean and individual predictions, respectively)

SYSTEM	Model 1A Low	Model 1A High	Model 3 Low	Model 3 High
Calgary (mean)	1436	1791	1436	2145
(individual)	1836	3098	768	4085
Detroit (mean)	400	704	602	1260
(individual)	231	1218	316	2400
Los Angeles (mean)	437	724	602	1137
(individual)	252	1253	316	2166
Portland (mean)	479	770	615	1071
(individual)	277	1331	323	2040
Seattle (mean)	567	894	640	1034
(individual)	328	1546	341	1969
Vancouver (mean)	964	1612	1275	3004
(individual)	558	2788	670	5722

models developed in this paper seem to pose no such problem: the preparation of the forecasts is certainly presentable.

REFERENCES

Belsley D. A., Kuh E. and Welsch R. E. (1980) *Regression Diagnostics: Influential Data and Sources of Collinearity.* Wiley, New York.

Cahners Publishing Co. (1982) *Modern Railroads Rail Transit*: 1982 *City-by-City Transit Digest.*

Dhyrmes, P. J. (1970) On the game of maximizing $\bar{R}^2$, *Australian Economic Papers.* Volume 9.

Halvorson R. and Palmquist R. (1980) The interpretation of dummy variables in semilogarithmic equations *The Amer. Econ. Rev.* 70, 474–475.

Gordon P. (1982) The Determinants of Rail Transit Demand—an International Cross-Sectional Comparison, University of Southern California.

Hilton G. (1968) Rail transit and the pattern of modern cities: The California case. *Traffic Quarterly*, pp. 379–393.

McGillivray R. (1972) Some problems in urban transportation models, *Traffic Quarterly*, pp. 547–558.

Pushkarev B. and Zupan J. (1980) *Urban Rail in America: An Exploration of Criteria for Fixed-Guideway Transit.* U.S. Department of Transportation.

Statistics Canada (1977) *Census of Canada.* Ottawa, Ontario.

Transportation Research Board (1978) *Light Rail Transit: Planning and Technology.* Special Report 182, Washington, D.C.

Union Internationale de Transports (1981) *Modern Light Rail*, 44th International Congress, Brussels.

Union Internationale des Transports Publics (1981) *UITP Handbook of Urban Transport.* Brussels.

United States Bureau of the Census (1981) *City and County Fact Book.* Washington, D.C.

Wachs, M. (1982) Ethical dilemmas in forecasting for public policy *Public Management Forum.* November/December 1982, p. 562–567.

Webster F. and Bly P. (editors) (1980) *The Demand for Public Transport: Report of the International Collaborative Study.* Transportation and Road Research Laboratory, Crowthorne, U.K.

Wohl, M. (1976) The Case for Rapid Transit: Before and After the Fact, Conference Proceedings from *Transportation Alternatives for Southern California*, University of Southern California.