136

T. R. NO. 136

Technical Report

LOAD-FACTOR MEASUREMENT
FOR ROAD TRANSPORT CORPORATIONS

by

Nitin R. Patel

विद्याविनियोगाद्विकासः
IIM
AHMEDABAD

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD

LOAD-FACTOR MEASUREMENT
FOR ROAD TRANSPORT CORPORATIONS

by

Nitin R. Patel

T. R. No. 136
October 1976

INDIAN INSTITUTE OF MANAGEMENT
Ahmedabad

ABSTRACT

In order to plan its time-tables, a road transport corporation needs to have reliable information on the load-factors on various bus-trips that are currently operated. Existing information does not permit determination of load-factors on segments of trips, but provide only over-all trip-wise load-factors. This paper outlines a system for estimating segment-wise load-factors. Such load-factors need to distinguish between local and through traffic. The conceptual separation of these types of traffic is discussed. An approach to estimating segment load-factors is described which includes development of results for simultaneous confidence intervals for the estimates.

LOAD-FACTOR MEASUREMENT FOR ROAD TRANSPORT CORPORATIONS

- Nitin R. Patel -

Introduction

One of the principal items of information required by a road transport corporation is the load-factor experienced on its various trips. Load-factor estimates form a vital component in the information needs of the corporation in determining its time-table. In this context the main use of load-factor computations is to support the following types of decisions:

1 to add extra trips on a given route;

2 to drop trips from a given route;

3 to add trips over a segment of a given route; and

4 to break up a trip into smaller trips by eliminating operation over a segment (or segments) of a given route.

This paper arose out of an information-system study conducted at one of the largest State Road Transport Corporations in India. The corporation currently operates over 20,000 trips daily with a fleet of about 5,000 vehicles. One of the objectives of the study was to examine ways of improving load-factor estimates for decisionmaking. This paper analyses one approach that could lead to substantial improvement in the quality of information available to decision-makers.

The nature of the problem

For purposes (1) and (2), the load factor as measured presently [i.e. Actual Earnings / (Number of seats x Start to Finish Fare)] is adequate. It is a reasonable approximation of the ratio of seat kilometers utilised to seat kilometers available, especially in view of the fact that the fare is proportional to the distance travelled above a minimum value. (The minimum fare was 30p. and the charge was 5p. per km. for distances above 6 kms.)

However, for objectives (3) and (4) the load factors can be misleading. For instance, there may be a segment of a trip which is heavily overloaded, but the rest may be lightly loaded, yet the load factor may seem reasonable for the trip as a whole. To assess the implications of adding an extra trip over a segment of the route or breaking the trip up and dropping a segment it is necessary to know the through as well as the local passenger traffic on the segment separately. This is illustrated in the examples given below:

Example 1

Let us suppose that a particular trip operates on a route consisting of three segments. Further suppose each segment is 50 km in length and the bus has 50 seats (total capacity). Then the available seat kilometers in each segment is 2500. Suppose load factors for each segment (defined as $\frac{\text{used seat-km.}}{\text{available seat-km.}}$ for the segment) are 0.5, 0.9, 0.5. Prima facie it may appear that we should consider adding a trip on segment 2 to bring down the excessive crowding in that segment. However, the following may be the composition of seat kilometers utilized:

	Seat kilometers utilized in segment		
	1	2	3
Persons getting on in 1 and off in 1	100	-	-
Persons getting on in 1 and off in 2	950	950	-
Persons getting on in 1 and off in 3	200	200	200
Persons getting on in 2 and off in 2	-	150	-
Persons getting on in 2 and off in 3	-	950	950
Persons getting on in 3 and off in 3	-	-	100
	1250	2250	1250
Load factors	0.5	0.9	0.5

The effect of adding a trip on segment 2 will simply be on the local traffic (150 seat-km) and not on the through traffic (2100 seat-km). Thus, adding a trip would probably only reduce the load factor on segment 2 from 0.9 to $\frac{2100}{2500}$, i.e. 0.84.

Example 2.

Suppose in the previous example, the load factors were 0.7, 0.3, 0.7, it may seem prima facie that the trip should be broken up into a trip over segment 1 and a trip over segment 3, dropping the operation over segment 2.

However, suppose the composition of utilized seat-kms is as below:

	Seat kilometers utilised in segment		
	1	2	3
Persons getting on in segment 1 and off in 1	1000	-	-
Persons getting on in segment 1 and off in 2	-	-	-
Persons getting on in segment 1 and off in 3	750	750	750
Persons getting on in segment 2 and off in 2	-	-	-
Persons getting on in segment 2 and off in 3	-	-	-
Persons getting on in segment 3 and off in 3	-	-	1000
	1750	750	1750
Load factors	0.7	0.3	0.7

Then dropping operation over segment 2 would also drop load factors in 1 and 3 to $\frac{1000}{2500}$ = 0.4.

Thus ideally for any segment we need to know not only the total seat-km. utilized on the segment, but also the through seat-km. (i.e. seat-km. on the segment due to passenger boarding or alighting or both boarding and alighting outside the segment). The local seat-km. is of use in making decisions of type (3) above, and the through seat-kms. are of use in making decisions of type (4) above.

The conductor's way-bill records the number of tickets in each denomination issued at each stage on the trip. If no combination tickets were issued, it would be a simple matter to calculate local and through seat-kms utilised for any segment of the trip. However, the issue of combination tickets is substantial as can be inferred from the fact that the nineteen ticket denominations in use were 5, 10, 15, 20, 25, 30, 40, 45, 50, 60, 65, 75, 90 paise and rupees 1, 2, 3, 4, 5, 10. Most fares would therefore require issue of combination tickets. Upto 4 tickets may need to be issued for a fare (eg. Rs 6.35). It would also be possible to work out the local and through seat-kms utilised if combinations were recorded as such at every stage on the way-bill. However, experiments with this approach have shown that it is not a practicable scheme.

Suggested approach

The approach suggested and analyzed in this paper is concerned with trips operated on routes upto 225 km in length. Such routes account for more than 97 percentage of the total number of routes. The approach envisages a change in the ticket denominations to the following values:

Paise	:	5, 10, 15, 20, 25, 30, 50, 65
Rupees	:	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

We could print 5p, 10p, 15p, 20p, 25p and 65p tickets in a different colour (say red) from the others (say black). Also we could require the conductor to always issue only one black ticket to each passenger. Combination tickets would be made up of one black ticket and as many red tickets as may be required. The maximum number of red tickets that would need to be issued to a passenger would be three. Thus the maximum number of tickets that would need to be issued to a passenger is the same as in the present scheme. The purpose of this arrangement is two-fold:

1 The number of passengers can be easily calculated (equals number of black tickets issued). In fact the conductor then can be spared the work of entering the number of passengers at each stage on the way-bill.

2 The maximum error that can be made in calculating the seat-km used by any passenger would be 10 km. Thus, reasonably accurate estimates of through and local traffic for any segment (however defined) can be made using data from the way-bill.

<u>Mathematical model for load-factor estimation</u>

Consider a trip which operates on a route with n stands. Number the stands 1, 2.........n with the starting point for the trip denoted by stand 1 and the ending point for the trip denoted by stand n.

Let N_{ij} = Number of passengers who travel from stand i to stand j over t operations of the trip.

Let us suppose we are interested in an arbitrary segment of the trip starting at stand p and ending at stand q $(p < q)$. If C is the capacity of the bus, and d_{ij} the distance between stands i and j, the quantities of interest are given by the following formulae:

Load factor for through passengers on segment p, q (L_1)

$$= \frac{\sum_{j=p+1}^{q} \sum_{i=1}^{p-1} d_{pj} N_{ij} + \sum_{j=q+1}^{n} \sum_{i=p}^{q-1} d_{iq} N_{ij} + \sum_{j=q+1}^{n} \sum_{i=1}^{p-1} d_{pq} N_{ij}}{d_{pq} C.t}$$

Load factor for local passengers on segment p,q (L_2)

$$= \frac{\sum_{j=p+1}^{q} \sum_{i=p}^{q-1} d_{ij} N_{ij}}{d_{pq} C.t}$$

These formulae are linear functions of N_{ij}; in other words,

$$L_1 = \sum_{j=i+1}^{n} \sum_{i=1}^{n-1} a_{ij} N_{ij} \qquad \text{........} \quad (1)$$

$$L_2 = \sum_{j=i+1}^{n} \sum_{i=1}^{n-1} b_{ij} N_{ij} \qquad \text{........} \quad (2)$$

Where $a_{ij} \geq 0$; $b_{ij} \geq 0$ are known constants.

The difficulty, of course, is that N_{ij} are not directly observable but have to be inferred from the sales of tickets of various denominations.

Let N_i^k be the number of black tickets (B.T.'s) of denomination k issued at stand i $(k = 1, 2 \ldots\ldots \bar{K};\ i = 1, 2 \ldots\ldots n-1)$

Let us denote by S_{ik}, the set of all possible destinations (stands) that a passenger buying a denomination k ticket at stand i can reach. Let $s_{ik} = |S_{ik}|$ = number of elements in S_{ik}. Let t_{ik} be the smallest member of S_{ik} i.e. $S_i^k = \{t_{ik}, t_{ik} + 1, \ldots\ldots\ldots\ldots\ldots, t_{ik} + s_{ik} - 1\}$.

Now if we assume that each passenger issued with a BT of denomination k is equally likely to go to any stand in the set S_i^k, then

$$p(N_{it_{ik}}, N_{it_{ik}+1}, \ldots\ldots\ldots N_{it_{ik}+s_{ik}-1}) = \frac{N_i^k\,!}{\prod_{j=t_{ik}}^{t_{ik}+s_{ik}-1} N_{ij}\,!} \left(\frac{1}{s_{ik}}\right)^{N_i^k}$$

where $N_{ij} = 0, 1, \ldots\ldots\ldots N_i^k$ with $\sum_{j \in S_{ik}} N_{ij} = N_i^k$ ...... (3)

(This is a special case of the multinominal distribution)

We will now consider the problem of estimation of L_1, L_2. Reasonable estimators $\hat{L}_1$, $\hat{L}_2$ would be

$$\hat{L}_1 = E(L_1), \quad \hat{L}_2 = E(L_2)$$

To work out $E(L_1)$ it is useful to rewrite (1) in the following form:

$$L_1 = \sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} \left(\sum_{j \in S_{ik}} a_{ij} N_{ij} \right)$$

This is true because $\sum_{j=i+1}^{n} a_{ij} N_{ij} = \sum_{k=1}^{\bar{K}} \sum_{j \in S_{ik}} a_{ij} N_{ij}$

$$\therefore \quad \hat{L}_1 = E(L_1) = \sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} \sum_{j \in S_{ik}} a_{ij} E(N_{ij})$$

$$= \sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} \left(\sum_{j \in S_{ik}} \frac{a_{ij}}{s_{ik}} \right) N_i^k = \sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} \bar{a}_{ik} N_i^k$$

$$\text{where} \quad \bar{a}_{ik} = \frac{1}{s_{ik}} \sum_{j \in S_{ik}} a_{ij} \qquad \ldots\ldots (4)$$

Similarly

$$\hat{L}_2 = \sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} \bar{b}_{ik} N_i^k \quad \text{where} \quad \bar{b}_{ik} = \frac{1}{s_{ik}} \sum_{j \in S_{ik}} b_{ij} \qquad \ldots\ldots (5)$$

We will next develop confidence intervals for the estimates $\hat{L}_1$ and $\hat{L}_2$. This is useful in itself but is even more important to guide the choice of sample size. The number of trips operated per day is 24,000, so that if load factors are estimated monthly without sampling the volume of processing becomes very large. To develop confidence limits for $\hat{L}_1$, $\hat{L}_2$, it will be necessary to know the variances of L_1, L_2 and the co-variance between L_1 and L_2. We first derive Var (L_1).

The first point to note is that

$$Var\ (L_1) = \sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} Var \left(\sum_{j \in S_{ik}} a_{ij}\ N_{ij} \right)$$

This assumes that the distribution of the destinations of passengers issued with denomination k tickets at stand i are independent of destinations of other passengers, i.e. passengers issued with other denominations at stand i and passengers originating at other stands.

We shall next develop a result that will be useful in calculating $Var \left(\sum_{j \in S_{ik}} a_{ij}\ N_{ij} \right)$

<u>Result 1</u>

If $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$ is a multinominally distributed random vector with parameters n and $p_1, p_2, \ldots\ldots p_m = \frac{1}{m}$

(i.e. $p_r(\underline{x}) = \frac{n!}{x_1!\, x_2!\, \ldots x_m!} \left(\frac{1}{m}\right)^n$

$x_1, x_2, \ldots\ldots x_m = 0, 1, \ldots\ldots n$

and $x_1 + x_2 + x_m = n$)

and $y = \underline{a} \cdot \underline{x}$ where $a = (a_1, a_2 \ldots\ldots a_m)$, then

$$\mathrm{Var}(y) = \frac{n}{m}\left[\sum_{i=1}^{m} a_i^2 - \left(\frac{\sum_{i=1}^{m} a_i}{m}\right)^2\right]$$

= n x variance in elements of $\underline{a}$.

The result is straight forward once it is recognized that the co-variance matrix of $\underline{x}$ is

$$\begin{bmatrix} \frac{n}{m}(1-\frac{1}{m}), & -\frac{n}{m^2} & \ldots\ldots, & -\frac{n}{m^2} \\ -\frac{n}{m^2}, & \frac{n}{m}(1-\frac{1}{m}), & \ldots\ldots, & -\frac{n}{m^2} \\ -\frac{n}{m^2}, & \ldots\ldots & \ldots\ldots & \frac{n}{m}(1-\frac{1}{m}) \end{bmatrix} = \frac{n}{m} I - \frac{n}{m^2} U$$

where U is an mxm matrix with all elements = 1.

$$\mathrm{Var}(\underline{a}\,\underline{x}) = \underline{a}\left(\frac{n}{m} I - \frac{n}{m^2} U\right)\underline{a}'$$

$$= \frac{n}{m}\sum_{i=1}^{m} a_i^2 - \frac{n}{m^2}\,\underline{a}\,U\,\underline{a}'$$

$$= \frac{n}{m}\sum_{i=1}^{m} a_i^2 - \frac{n}{m^2}\;\underline{a}\cdot\underline{1}\cdot\underline{1}'\,\underline{a}' \quad \text{where}$$

$$\underline{1} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \quad \text{is an } m \text{ element column vector}$$

$$= \frac{n}{m}\sum_{i=1}^{m} a_i^2 - \frac{n}{m^2}\left(\sum_{i=1}^{m} a_i\right)^2$$

$= n \times$ Variance in elements of $\underline{a}$.

Applying this result, we obtain:

$$\mathrm{Var}\left(\sum_{j\in s_{ik}} a_{ij} N_{ij}\right) = N_i^k \left\{\frac{1}{s_{ik}}\sum_{j\in s_{ik}} a_{ij}^2 - \frac{1}{s_{ik}^2}\left(\sum_{j\in s_{ik}} a_{ij}\right)^2\right\}$$

$$= N_i^k \times v_{ik}$$

where v_{ik} is the variance in the elements of the set

$$\left\{ a_{ij} \mid j \in s_{ik} \right\}.$$

This gives us the following main result:

$$\text{Var}\ (L_1) = \sum_{i=1}^{n-1} \sum_{k=1}^{\overline{K}} N_i^k\ v_{ik} \qquad \text{...... (6)}$$

Similarly we can show that

$$\text{Var}\ (L_2) = \sum_{i=1}^{n-1} \sum_{k=1}^{\overline{K}} N_i^k\ w_{ik}$$ where w_{ik} is the variance in

the elements of the set

$$\left\{ b_{ij} \mid j \in S_{ik}. \right\} \qquad \text{...... (7)}$$

We next obtain an expression for Cov (L_1, L_2), the co-variance between L_1 and L_2. This is readily derived using our earlier results for the variance and the following identity:

$$2\ \text{Cov}\ (L_1, L_2) = \text{Var}\ (L_1 + L_2) - \text{Var}\ (L_1) - \text{Var}\ (L_2)$$

Now,

$$\text{Var}\ (L_1 + L_2) = \text{Var} \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (a_{ij} + b_{ij})\ N_{ij} \right)$$

$$= \text{Var} \left(\sum_{i=1}^{n-1} \sum_{k=1}^{\overline{K}} \left[\sum_{j \in S_{ik}} (a_{ij} + b_{ij})\ N_{ij} \right] \right)$$

$$= \sum_{i=1}^{n-1} \sum_{k=1}^{\overline{K}} \text{Var} \left[\sum_{j \in S_{ik}} c_{ij}\ N_{ij} \right]$$

where $c_{ij} = a_{ij} + b_{ij}$

Using the earlier results on variances,

$$\text{Var} \left(\sum_{j \in S_{ik}} c_{ij} N_{ij} \right) = N_i^k u_{ik}$$ where u_{ik} = Variance in the elements of the set c_{ij}, $j \in S_{ik}$

$$\therefore \quad \text{Var} \; (L_1 + L_2) = \sum_{i=1}^{n-1} \sum_{k=1}^{\overline{K}} N_i^k u_{ik}$$

$$\therefore \quad 2 \text{ Covar} \; (L_1, L_2) = \sum_{i=1}^{n-1} \sum_{k=1}^{\overline{K}} N_i^k (u_{ik} - v_{ik} - w_{ik})$$

$$= \sum_{i=1}^{n-1} \sum_{k=1}^{\overline{K}} N_i^k . \; 2 \text{ Cov} \; (a_{ij}, b_{ij} \mid j \in S_{ik})$$

$$\text{Cov} \; (L_1, L_2) = \sum_{i=1}^{n-1} \sum_{k=1}^{\overline{K}} N_i^k \; \text{Cov} \; (a_{ij}, b_{ij} \mid j \in S_{ik}) \qquad \ldots\ldots \quad (8)$$

$\therefore$ If the covariance matrix for the elements a_{ij}, b_{ij}, $j \in S_{ik}$ is denoted by D_{ik}, and the covariance matrix for L_1, L_2 by D,

$$D = \sum_{i=1}^{n-1} \sum_{k=1}^{\overline{K}} N_i^k D_{ik} \qquad \ldots\ldots \quad (9)$$

[In fact this result can be readily generalised to any arbitrary number of load factors L_1, L_2 .... L_p]

From the multivariate Central Limit Theorem, it is reasonable to consider the vector $\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$ to be distributed according to the bivariate-normal distribution with mean $\mu = \begin{bmatrix} E(L_1) \\ E(L_2) \end{bmatrix}$ and co-variance matrix $\Sigma = D$. If we make this approximation, we can use a result derived by Morrison* based on a procedure suggested by Roy and Bose@ for obtaining simultaneous confidence intervals for any linear combinations of L_1 and L_2. We will be interested in such simultaneous confidence intervals because we will be interested mainly in estimating L_1, L_2 and $L_1 + L_2$ (the total load factor over the segment p, q) simultaneously.

The special form of the result that we will use can be summarized as follows:

If $x: \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is distributed according to the bi-variate normal distribution with mean μ and covariance matrix Σ, and if $\theta = (\theta_1, \theta_2)$ is any two-dimension (row) vector,

$$\Pr\left(\frac{[\theta(x-\mu)]^2}{\theta'\Sigma\theta} \leq T^2_{\alpha:2} \right) = 1-\alpha$$

Where $T^2_{\alpha:2}$ is the upper α percentile point for the χ^2 distribution with 2 degrees of freedom.

Applying this result we get the following simultaneous confidence intervals for L_1, L_2 and $L_1 + L_2$ at the α level of confidence.

$$L_1 : \left(\hat{L}_1 - \sqrt{\left(\sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} N_i^k v_{ik}\right) T^2_{\alpha:2}},\ \hat{L}_1 + \sqrt{\left(\sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} N_i^k v_{ik}\right) T^2_{\alpha:2}} \right) \quad \ldots(10)$$

$$L_2 : \left(\hat{L}_2 - \sqrt{\left(\sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} N_i^k w_{ik}\right) T^2_{\alpha:2}},\ \hat{L}_2 + \sqrt{\left(\sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} N_i^k w_{ik}\right) T^2_{\alpha:2}} \right) \quad \ldots(11)$$

$$L_1 + L_2 : \left(\hat{L}_1 + \hat{L}_2 - \sqrt{\left(\sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} N_i^k u_{ik}\right) T^2_{\alpha:2}},\ \hat{L}_1 + \hat{L}_2 + \sqrt{\left(\sum_{i=1}^{n-1} \sum_{k=1}^{\bar{K}} N_i^k u_{ik}\right) T^2_{\alpha:2}} \right) \quad \ldots(12)$$

In fact it is possible to construct a confidence interval for an arbitrary combination $\theta_1 L_1 + \theta_2 L_2$ which will simultaneously hold for the α level of confidence for any arbitrary number of θ_1, θ_2 pairs. The general confidence interval is

$$\left(\hat{\theta_1 L_1} + \theta_2 \hat{L}_2 - \sqrt{(\theta_1, \theta_2)\, D \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \cdot T^2_{\alpha:2}}, \right.$$

$$\left. \theta_1 \hat{L}_1 + \theta_2 \hat{L}_2 + \sqrt{(\theta_1, \theta_2)\, D \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \cdot T^2_{\alpha:2}} \right) \quad \ldots (13)$$

Conclusion

The results developed in this paper, namely formulae (4), (5), (10), (11) and (12) allow us to estimate the local, through and total load-factors for any segment of a trip. Further, they enable us to assess the reliability of these estimates by means of confidence intervals around these estimates. Although the results look complex, they are not difficult to compute since they involve basically calculating linear combinations of the N_i^{k}'s. The various 'weights' involved in the formulae, namely $\bar{a}_{ik}$, $\bar{b}_{ik}$, v_{ik}, w_{ik}, u_{ik} involve calculating means and variances of sets of numbers which are pre-computable in the sense that they do not depend on the variable data namely N_i^{k}'s. Once these 'weights' are calculated, they can be applied routinely week after week to the pertinent values of the N_i^{k}'s in a straight-forward manner.

ACKNOWLEDGEMENTS

This paper arose out of work being conducted by a project team at IIMA consisting of Professors VL Mote, M Raghavachari, JL Saha (co-ordinator) and myself. Many of the ideas discussed here emerged under the stimuli of discussions held by the project team, and I would like to record my indebtedness and gratitude to all the members of this project team.