Journal of Air Law and Commerce

Volume 21 | Issue 3 Article 3

1954

A Regression Analysis of Airline Costs

Jesse W. Proctor

Julius S. Duncan

A REGRESSION ANALYSIS OF AIRLINE COSTS

BY JESSE W. PROCTOR AND JULIUS S. DUNCAN[1]

Mr. Proctor, Institute of Industrial Relations, University of California at Los Angeles. B.B.A. 1953, M.A. 1954, University of New Mexico; *Formerly*, Administrative Assistant, Atomic Energy Commission.

Mr. Duncan, Chairman, Department of Economics, University of New Mexico. B.A. 1918, M.A. 1919, University of Mississippi; B.D. 1924, Emory University; Ph.D. 1932, Columbia University; *Formerly*, Economic Analyst, Interstate Commerce Commission; Instructor, Babson Institute; Assistant Professor, St. John's College; Lecturer in Economics, Bryn Mawr College.

ANALYSIS of the variables which affect airline costs has extensive policy implications. This paper describes a study of the following cost influencing variables: (1) capacity of the plane; (2) length of the flight; (3) ton-mile load factor; (4) number of hours per day planes are utilized; (5) metropolitan population served; (6) average speed of planes; and (7) net assets of each firm. Investigation indicates that only three of these variables have significant correlation with cost per revenue ton-mile. These three variables are: ton-mile load factor; capacity of the plane; and number of hours per day planes are utilized. The four remaining variables mentioned above were found to be insignificant. This paper describes the regression analysis which gave rise to the above conclusions.

It also seeks to evaluate the results of this regression anaylsis in the light of the theory of costs, operating conditions, and other factors which do not appear in the equations. The conclusion is reached, first, that length of flight stage is still an important variable. Cost theory suggests that, other things being equal, the line having the longer flight stage will have lower costs per revenue ton-mile. Different types of planes tend to obscure differences in cost of operation. Furthermore, the present average length of flight stage is greater than the average distance between cities served by the airline. Second, that daily flight time per aircraft is an important variable which has been neglected in existing cost studies. Third, that the number of variables affecting airline costs is so great as to justify more use of regression analysis as a supplement to conventional cost accounting methods.

The relative significance of these variables is important when the

[1] The authors are indebted to Professor Carleton Eugene Buell of the Department of Mathematics of the University of New Mexico for criticism of the statistical techniques employed in this study. Our thanks are also due to Professors Nathaniel Wollman, Mervyn Crobaugh and David Hamilton of the Department of Economics for suggestions about manner of presentation. We also wish to express our gratitude to the University of New Mexico Committee on Research for two financial grants which encouraged us to go on with this investigation.

following questions are under consideration: airline corporate mergers; relative emphasis to be placed on coach traffic and first-class service; and the extent to which charges per passenger-mile or per ton-mile should decrease with increasing distance. Identifying and measuring the factors responsible for costs may have a significant bearing on the kind of air service the public receives and the amount of subsidy required to get it.

The investigation of these factors was made by means of a least-squares analysis. The purpose of this article is to examine the policy implications that result from the regression analysis.

It is recognized that these variables are not the only variables which can significantly affect airline costs. Also, one of the variables used is of the omnibus type. That is, a number of elements enter into the load factor, for example, the top 97 pairs of cities served significantly affects the load factor.[2] It also seems likely that there is a relationship between the number of top 97 pairs of cities served and metropolitan population. The absence of late data on the top 97 pairs of cities served made it inadvisable in this particular experiment to attempt to measure the relative role of the number of the top 97 pairs of cities served.

I. Selection of Variables

The criterion for the selection of the other variables was their implications for air transport policy. Capacity of plane is important because, it costs almost as much to buy and almost as much to operate a small plane as it does a large plane.[3] The number of hours per day that a plane is used is important, because a substantial percentage of an airline's investment is in its aircraft, and therefore, the greater the utilization the lower the unit costs will be.[4] Theoretically, unit cost may be expected to be inversely correlated with the size of the firm. It therefore is necessary to consider the relative significance of the size of the firm. Metropolitan population served by an airline furnished an important clue to its market. Speed of the plane has been and is an important factor in the growth of airline travel.

The cost influence of length of the flight has been prominent in recent discussions of airline costs. The certificated airlines argued in the Transcontinental Air Coach case that the principal reason why the

[2] See Koontz, Harold D. "Economic and Managerial Factors Underlying Subsidy Needs of Domestic Trunk Line Air Carriers," JRL. OF AIR LAW & COM., Spring 1951, pp. 127bf.; "Domestic Airline Self-Sufficiency: a Problem of Route Structure," *American Economic Review*, March, 1952. pp. 118-120.

[3] Koontz, Harold D., "Airline Self-Sufficiency: Rejoinder," *American Economic Review*, Vol. XLIII, No. 9, June, 1953. p. 375. It may be that research could develop a small cheap plane for the thinly trafficked routes. See Carter, John P., "Domestic Airline Self-Sufficiency: Comment," *Ibid.*, p. 370. One difficulty in the use of a small cheap plane, assuming that it has been developed, is that it is difficult to imagine the Civil Aeronautics Board allowing a common carrier plane to operate without two people capable of piloting it, and, pilots' salaries are an important part of airline costs.

[4] Cherington, Charles R. "The Essential Role of Large Irregular Air Carriers." *Future of Irregular Airlines in United States Air Transportation Industry*, Hearings Before a Subcommittee of the Select Committee on Small Business, U. S. Senate, March 31 through May 8, 1953 (Washington, Government Printing Office, 1953), p. 578.

non-certificated and non-scheduled airlines were able to make their low rates for trans-continental coach service was their long flights.[5] American Airlines and Eastern Airlines made the matter of length of flight a specific issue in setting standard fares for the transportation of aircoach passengers between New York and Washington, D. C.[6]

II. Data and Method

Perhaps it will be easier for the reader to follow the somewhat mathematical material which follows if a non-mathematical tabulation of the most important single finding of this paper is presented here. The preparation of this table was suggested by the results of the regression analysis and prepared after the mathematical section was completed. This tabular presentation compares a ranking of airlines on the basis of daily flight time per aircraft and ranked from low to high with a ranking of the same airlines on the basis of total operating expense per revenue ton-mile. The rank order of operating expense per revenue ton-mile, unlike that of average daily flight time per aircraft, is from high to low. The table follows:

RANK CORRELATION BETWEEN DAILY FLIGHT TIME PER AIRCRAFT AND TOTAL OPERATING EXPENSE PER REVENUE TON-MILE

Airline	Daily Flight Time per Aircraft (from low to high)	Total Operating Expense per Revenue Ton-Mile (from high to low)
Wiggins	1	1
Lake Central	2	3
Bonanza	3	4
West Coast	4	11
Central	5	2
Empire	6	8
Wisconsin	7	5
Southwest	8	14
Mohawk	9	12
Trans-Texas	10	6
All American	11	9
Southern	12	4
Northeast	13	13
Pioneer	14	16
Frontier	15	7
Piedmont	16	15

Source: Table I

All the data used in this study applies to the year 1951. Table I gives operating and cost data for 31 airlines, and for the following variables: length of the interstation flight (in miles); speed of the plane (in miles per hour); plane utilization (average number of hours per

[5] CAB Docket 3397 (Transcontinental Coach Type Service Case) et al., *Opinion* and *Report* of William J. Madden, Examiner.

[6] *Docket No.* 6098 *Et Al.* (The Short Haul Coach Fare Investigation) See particularly Supplemental Exhibit No. AA-e-2 of CAB Docket No. 3663.

day each plane is used); metropolitan population served; and operating expense per revenue ton-mile (in cents).

Table II shows the method by which aircraft capacity was computed. The figures were derived by dividing revenue tons per aircraft mile by ton-mile load factor. Table III presents assets employed which figure was obtained by subtracting investments and special funds from total assets.

Logarithms of each of the variables here presented were taken, with the exception of ton-mile load factor. In the case of the ton-mile load factor, the raw data were used in the form in which they appear in Table II.

The logarithms of the variables previously described are represented by the following notations:

x_1 = Cost per revenue ton-mile (in cents)
x_2 = Capacity of the plane (in tons)

The airlines listed above all have flight stages of one hundred miles or less. The inverse correlation is not perfect, but it appears to be close enough to be regarded as significant. The material which follows seeks to measure more precisely the relative influence of each one of the variables considered.

TABLE I

OPERATING AND COST DATA

	(1) Length of the Flight (Flight stage in Miles)	(2) Speed of the Plane (Average Miles per Aircraft Hour)	(3) Plane Utilization (Daily Flight Time per Aircraft)	(4) Metropolitan Population Served (000)	(5) Cost (Total Operating Expense per Revenue Ton-Mile, in cents)
All American	57	133	6:06	20,200	116.3
American	270	216	6:56	56,928	43.0
Bonanza	100	140	4:27	183	141.5
Braniff	176	182	6:36	11,869	50.6
Capital	142	167	7:28	41,097	51.0
Central	51	134	4:40	1,757	318.5
C & S	175	175	8:36	18,000	59.2
Colonial	112	150	6:52	13,500	77.0
Continental	131	179	6:30	3,831	62.3
Delta	174	191	7:36	13,119	45.3
Eastern	182	187	9:30	44,000	42.6
Empire	59	143	4:50	451	112.4
Frontier	81	141	7:28	2,500	125.2
Lake Central	73	142	3:51	5,405	169.3
Mid-Continent	144	167	6:12	6,725	64.8
Mohawk	79	137	5:53	9,250	100.5
National	199	207	8:17	23,431	42.9
Northeast	94	150	6:46	16,000	81.1
Northwest	271	202	6:53	27,000	56.7
Piedmont	90	153	8:08	3,362	75.4
Pioneer	89	150	7:03	2,050	71.3

Southern	67	147	6:09	2,337	150.1
Southwest	54	127	5:51	5,143	78.9
Trans-Texas	78	147	6:05	2,600	130.3
TWA	293	193	6:24	51,500	46.2
United	270	207	6:38	48,913	42.3
West Coast	58	135	4:30	1,850	103.3
Western	172	181	6:40	6,250	44.2
Inland	147	158	9:23	1,000	64.7
Wiggins	45	116	2:21	2,500	820.9
Wisconsin	69	142	5:43	6,500	130.9

Source: *Recurrent Report of Mileage and Traffic Data and Recurrent Report of Financial Data*, CAB where available, (Figures rounded). The remainder from *World Airline Record*. (Chicago, Roadcap and Associates, 1952.)

TABLE II

TON-MILE LOAD FACTOR AND AVAILABLE TONS PER AIRCRAFT MILE

Company	(1) Revenue Tons per Aircraft Mile	(2) Ton-Mile Load Factor*	(3) Available Tons Per Aircraft Mile, Column (1) Divided by Column (2)
All American	.96	.400	2.40
American	3.98	.689	5.77
Bonanza	.79	.358	2.21
Braniff	2.57	.557	4.61
Capital	2.68	.510	5.25
Central	.35	.167	2.10
C & S	2.17	.558	3.89
Colonial	1.68	.505	3.32
Continental	1.70	.537	3.17
Delta	2.74	.598	4.58
Eastern	3.07	.528	5.81
Empire	.69	.313	2.20
Frontier	.72	.398	1.81
Lake Central	.48	.212	2.26
Mid-Continent	1.53	.565	2.71
Mohawk	1.02	.476	2.14
National	3.12	.567	5.50
Northwest	4.30	.570	7.54
Piedmont	1.08	.449	2.41
Pioneer	1.16	.483	2.40
Southern	.60	.317	1.89
Southwest	1.18	.491	2.40
Trans-Texas	.65	.372	1.75
TWA	3.62	.670	5.40
United	3.75	.630	5.95
West Coast	.82	.421	1.95
Western	2.68	.656	4.09
Inland	1.54	.575	2.68
Wiggins	.07	.166	.42
Wisconsin	.81	.430	1.88

* Converted to decimal form.

Source: *Recurrent Report of Mileage and Traffic Data* and *Recurrent Report of Financial Data*, CAB where available, (Figures rounded). The remainder from *World Airline Record.* (Chicago, Roadcap and Associates, 1952.)

TABLE III

ASSETS EMPLOYED

	Total Assets (00000's)	Investments and Special Funds (00000's)	Adjusted Assets (Total Assets less inv. and special funds) (00000's)
All American	21.13	3.21	18
American	1436.53	165.22	1271
Bonanza	6.65	.01	6.6
Braniff	160.30	5.81	154
Capital	195.02	6.06	189
Central	14.02	.01	14
C & S	114.16	3.11	111
Colonial	35.34	.79	35
Continental	49.74	3.49	46
Delta	174.32	3.87	170
Eastern	1042.58	187.84	855
Empire	4.71	.02	4.7
Frontier	17.66	.12	18
Lake Central	6.32	.01	6.3
Mid-Continent	76.12	2.44	74
Mohawk	14.54	.09	14
National	181.27	16.14	165
Northeast	49.38	2.11	47
Northwest	471.50	7.98	464
Piedmont	21.60	3.10	19
Pioneer	19.11	.01	19
Southern	12.44	.01	12
Southwest	16.06	.08	16
Trans-Texas	11.10	.42	11
TWA	1217.98	120.07	1098
United	1127.25	79.98	1047
West Coast	10.98	.01	11
Western	137.39	17.02	120
Inland	16.61	.03	17
Wiggins	2.03	.62	1.4
Wisconsin	10.80	.01	11

Source: *Recurrent Report of Mileage and Traffic Data,* and *Recurrent Report of Financial Data,* CAB where available, (Figures rounded). The remainder from *World Airline Record.* (Chicago, Roadcap and Associates, 1952.)

x_3 = Flight stage (length of the non-stop flight in miles)
x_4 = Ton-mile load factor (in decimal form)
x_5 = Number of hours per day planes are utilized
x_6 = Metropolitan population served (in hundreds of thousands)
x_7 = Average speed of planes (in miles)
x_8 = Net assets of each firm (in hundreds of thousands of dollars)

The following linear equation arises from the logarithms (or raw data, as the case may be) of the above data.

$$x_1 = B_2x_2 + B_3x_3 \ldots B_8x_8 + K$$

The B's in this equation, commonly referred to as regression coefficients, may be obtained by the method of least squares. Let

$$Z_{ij} = n\sum_{i=1}^{n}(x_i - \bar{x}_j)(x_j - \bar{x}_j)$$

The best estimate of x_1 may be given in terms of the remaining variables obtained by selecting the regression coefficients so as to satisfy the following simultaneous linear equations.[7]

$$\begin{aligned} B_2Z_{22} + B_3Z_{23} + \ldots + B_mZ_{2m} &= Z_{21} \\ B_2Z_{32} + B_3Z_{33} + \ldots + B_mZ_{3m} &= Z_{31} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ B_2Z_{m2} + B_3Z_{m3} + \ldots + B_mZ_{mm} &= Z_{m1} \end{aligned}$$

When the logarithms of the data previously described were arranged in a form corresponding to the above set of equations, the result was

TABLE IV

SIMULTANEOUS EQUATIONS

$$\begin{aligned}
319.05B_2+246.32B_3+57.03B_4+107.87B_5+50.97B_6+73.20B_7+82.86B_8&=-310.33\\
246.32B_2+284.64B_3+59.40B_4+87.90B_5+46.62B_6+76.78B_7+83.36B_8&=-272.48\\
57.03B_2+59.40B_3+18.01B_4+26.06B_5+11.29B_6+16.58B_7+18.62B_8&=-77.07\\
107.87B_2+87.90B_3+26.06B_4+73.58B_5+16.13B_6+26.62B_7+28.99B_8&=-141.12\\
50.97B_2+46.62B_3+11.29B_4+16.13B_5+18.92B_6+13.89B_7+19.07B_8&=-49.19\\
73.20B_2+76.78B_3+16.58B_4+26.62B_5+13.89B_6+24.87B_7+24.73B_8&=-81.75\\
82.86B_2+83.36B_3+18.62B_4+28.99B_5+19.07B_6+24.73B_7+29.40B_8&=-88.12
\end{aligned}$$

Techical note: In obtaining the above moments, the logarithm of the variable x_8 (net assets of each firm) was divided by 10. This had the effect of dividing Z_{88} by 100 and of dividing Z_{8j} and Z_{1j} by 10. An adjustment was made to the final regression coefficient by dividing it by 10. See Klein, Lawrence R., *Econometrics* (Row, Peterson and Company, 1953), p. 146, for the logic of this adjustment.

Table IV. Solutions to the simultaneous equations given in Table IV were computed by means of the Doolittle method. This method of computation is outlined by Klein.[8]

The Doolittle method enables the computer to obtain the inverse moment matrix simultaneously with the regression coefficients. Elements in the diagonal of the inverse moment matrix are in turn used to establish the significance levels of the regression coefficients.[9]

Computations gave the regression coefficients, standard deviations, and significance levels shown in Table V.

III. SIGNIFICANCE OF COMPUTATIONS

The application of student's *t* distribution in Table V indicates that there are only three variables with significant coefficients, if it be as-

[7] See, for example, Harald Cramér, *Mathematical Methods of Statistics*, (Princeton University Press), p. 303.

[8] Klein, Lawrence R., *Econometrics* (Row, Peterson and Company, 1953), pp. 144-158.

[9] Kenney and Keeping, *Mathematics of Statistics* (D. Van Nostrand), p. 314.

sumed that a 5% level of significance is necessary before a regression coefficient is accepted as significant. These three variables are: capacity of the plane; ton-mile load factor; and plane utilization (number of minutes per day planes are used). The remaining variables are believed to be insignificant. Both capacity of the plane and load factor are significant at the 1% level; while plane utilization is significant at the 2% level. Coefficients of the magnitude of those obtained for length of flight, metropolitan population, speed of the plane and size of the firm would have been obtained by chance one time out of 5.

TABLE V

REGRESSION COEFFICIENTS

	Capacity of the plane (in tons)	Length of the flight (in miles)	Ton-mile Load Factor	Flight Time	Metropolitan population served	Average speed of planes	Net Assets of each firm
Regression Coefficients	-.423	+.229	-2.583	-.476	+.545	-.752	+.068
Authors' *t* ratio	3.39	1.20	5.09	2.49	1.17	1.23	.082
Significance Levels*	.01	.2	.01	.02	.20	.20	.50 (approx.)

* A confidence level of .01 indicates this regression coefficient will arise by chance once in a hundred trials.

It therefore appears that the true regression coefficient of each of these variables is zero. Even if we should decide to regard any one of these variables as significant, its level of significance is so low in comparison with the three "significant" variables that we should probably be safe in ignoring it.

A question arises as to whether or not the simultaneous omission of four variables from a regression equation might not cause a significant decrease in the ability of the equation to make predictions. We have shown that the omission of certain variables singularly would not significantly affect the equation. As a matter of fact, a further test, showed that simultaneous elimination of four variables would not significantly decrease the efficiency of the equation.[10]

If all variables except those found to be significant are ignored, a new least-squares equation results. This equation will have regression coefficients which are slightly different from the regression coefficients for the same variables previously computed. Further information on these variables is shown in Table VI.

Note that the decimal for the regression coefficient of the load factor has been shifted in Table VI. This permits us to use the load factor in percentage form instead of the somewhat unusual decimal form.

[10] For a simple and understandable explanation of the test used for this purpose, see McNemar, Quinn, *Psychological Statistics* (John Wiley & Sons, 1949), p. 266.

TABLE VI

	Ton-mile Load Factor (in percentage)	Capacity of the plane (in tons)	Plane Utilization (minutes per day)
Regression coefficient	-.022784	-.376317	-.559280
Standard deviation of regression coefficient	.096	.004	.189
Confidence limits of regression coefficients	.014 to .030	.18 to .58	.2 to .9
Coefficient of partial correlation	.733	.607	.500

When the regression coefficients of Table VI are properly fitted into an equation, the following expression results:

$$Y = (e^{9.25\ -.02t})\ (R^{\ -.4})\ (S^{\ -.6})$$

In this equation, Y is the cost per revenue ton-mile (in cents); e is the base of natural logarithms; R is the capacity of the plane (in tons); S is a measure of plane utilization (average flight time per plane per day); and t is the one-mile load factor (in percentage).

The multiple correlation coefficient of the above equation was found to be in the neighborhood of .96. By Snedecor's table[11] this equation would have reached the 1% level of significance with a correlation of .573. This means that the correlation here obtained is significant far beyond the 1% level.

A multiple correlation coefficient of .96 indicates that 92% of the variation is due to the three variables emphasized if it be assumed that changes in them are not the result of other factors. The remaining 8% is due to all other factors. The other factors may consist of managerial skill, variations in accounting methods, variations due to length of the haul, variations due to regional differences, etc.

Table VI indicates that the confidence limits of this equation are broad. This means that, although the regression coefficients fit the observed data well, these coefficients are subject to considerable error when they are applied to the universe of data similar to that from which the observations were taken. In other words, if we were to fit this equation to observations in a year other than 1951, we should expect considerable error.

IV. CONCLUSIONS

As previously stated, the indications from the regression analysis that length of flight stage is an insignificant factor in airline costs is viewed with great reserve. The empirical data does not yield results cost theory would lead us to expect. Stated in another way, a larger sample than one year is necessary before one is justified in concluding

[11] Snedecor, George W., *Statistical Methods* (Iowa State College Press, 1940) p. 286.

that length of flight stage is unimportant. Other things being equal, the line with shorter flight stages will at the minimum have to spend more time in idling its engines on the ground and lifting the plane to the desired operating flight level. Gasoline consumption per plane mile in reaching operating height is, of course, greater than that used at cruising altitudes. Furthermore, this regression analysis does not differentiate between types of planes. For example, a large number of the feeder airlines used in this comparison operate with DC3s. These planes are adapted to short-stage flights and operation costs are much lower than the later two and four engine planes used by the trunk airlines. Perhaps a still more important reason for believing that flight stage is a significant variable is that the average length of plane hop is now much greater than the average length of interstation flight. The increase in the number of through flights accounts for this difference. These longer flight stages are very important because on them the total travel time (in terms of elapsed time of transit from passenger's place of business or residence to destination business or residence) vividly shows the superiority of air transportation to competing media of transportation for such distances.

The most significant conclusion of this study is that daily flight time per aircraft is a significant variable in airline cost. The regression analysis of empirical data confirms the *a priori* hypothesis that unit costs, other things being equal, decrease with increased utilization. There is a point, of course, beyond which this is not true. Repair costs and fuel costs are, of course, also related to the number of plane miles or plane hours.

This finding supplements and clarifies the conclusion of Koontz as to the crucial role of the route structure in airline costs. An airline must have a route long enough to enable it to use its planes a reasonable number of hours per day. It seems unnecessary to labor the point that Wiggins Airline with a daily flight time per aircraft of only two hours and twenty-one minutes is going to be a high cost operation.

Our data indicate that any cost analysis which does not take into account number of hours per day of plane utilization as well as capacity of plane, and load factor is unrealistic. The existing flight stage cost studies, while specifying load factor and capacity of plane do not take into account explicitly daily flight time per aircraft.

Moreover, the study herein presented avoids the controversial assumptions involved in deciding how to allocate indirect costs to flight stages. The statistical approach, also, can bring into focus the most important cost influencing variables from the standpoint of the country as a whole. Regression analysis would seem to offer the possibility of a starting point from which to measure deviations from average. This regression analysis, for example, strongly confirms both the theory of costs and available empirical studies as to the crucial importance of the load factor. It partially confirms Koontz's finding that the size of the firm is relatively unimportant in airline costs. The indication is, however, subject to the reservation that the route structure should be long enough to make possible use of planes a reasonable number of hours per day. The large number of variables in airline costs would seem also to suggest the need of regression analysis as a workable tool in weighing the relative importance of these variables. Further work in this field would, however, require a larger sample than was available when this paper was prepared.

The extent to which the cost of rendering passenger service between, for example, Boston — New York — Washington is increased by time lost in waiting — part of the time with engines running — for tower clearance, and stacking merits further investigation.[12] The problem of the relationship between congestion of air traffic, duration of weather requiring instrument flying, length of flight stage and fares between densely populated Atlantic Seaboard cities, probably has not been finally disposed of despite the fact that American Airlines and Eastern Airlines have given up, at least for the time being, their efforts to get higher coach fares between Boston — New York — Washington on long-haul coach trips. The limitation by the CAB of the number of scheduled seat miles between congested short flight stage cities on the Atlantic Seaboard in such a manner as to insure high average load factors between such airports merits consideration. The unanimity of all studies, including this one, on the importance of the load factor is the basis for this suggestion. This does not rule out the validity of differential fares as another method of more closély correlating fares and costs between such centers.

---

[12] The element of circuity resulting from a multiplicity of air routes serving heavily populated areas appears to have been allowed for in counting scheduled ramp-to-ramp elapsed time.