FORECASTING LONG DISTANCE PASSENGER TRANSPORT

Matt FRANCKI
Principal Transportation Planner
Gutteridge Haskins & Davey Pty Ltd
Sydney

Ann EYLAND
Senior Lecturer, School of Economic and
Financial Studies
Macquarie University
Sydney

ABSTRACT: *The paper describes an application of the Gravity Model to forecast annual airline passenger traffic at a particular airport. The advantages are the exclusive use of published statistics and the application of regression analyses to select the variables and calibrate the model. The method is recommended for use in strategic planning for air transport. Its extension to other modes of passenger transport is discussed.*

INTRODUCTION

The purpose of forecasting passenger transport can be broadly classified into one of two categories which, in turn, influence the choice of forecasting procedure in terms of model used and accuracy required. The two categories are here termed "strategic" and "analytical"; they are not necessarily mutually exclusive.

Strategic planning generally has to do with development of a transport network or the services on it. By the nature of the task to be performed, the planning is usually for the medium and long term. In air transport, the planning often concerns a part of the network; a single airport and its feeder services. A number of aggregate models, regression equations and gravity models among them, are suitable for strategic planning.

Tactical planning, as the name applies, often has to do with underlying forces behind the demand for travel. Its purpose is often to determine policy or the consequences of policy. It is frequently short term and concerns the choice of mode. Disaggregate models are appropriate to this type of planning.

In the study reported herein, the planning was strategic and the immediate purpose was to forecast the demand for air services at Maroochy airport, Queensland. At present, Maroochy is a relatively minor airport suitable for use by the Fokker Friendship, F27, prop aircraft. The question was whether development would be to the F28, small jet standard or whether the facilities should be staged for full jet service on the domestic network.

A gravity model was used to prepare a cross-sectional estimate of passenger demand at five representative regions in Queensland for 1980/81. These were Coolangatta, Townsville, Cairns, Rockhampton (to include Gladstone) and Mackay (to include Proserpine). They were chosen to represent the five major regional centres on the Queensland coast, all of which have a substantial tourist trade and, with the possible exception of Coolangatta, serve a hinterland with agricultural and mining industries.

The fitted model was then verified by simulating air travel at the five representative regions in 1976/77 before being used to forecast travel at Maroochy airport which, in ten years' time, could serve a regional population on the Sunshine Coast similar in size to that currently served by Coolangatta on the Gold Coast.

CHOICE OF A MODEL

The task in hand had to do with "strategic" planning of development for Maroochy airport which serves the fast growing region of the Sunshine Coast. Modelling was necessary because projection of past trends would not suffice even in the medium term, as the base is too small.

Modelling of a demand for air services is not well established in Australia. Overseas, it generally takes one of two approaches. The first uses a time series analysis of the variation in demand over time between individual zone pairs, an example of which is reported by Oberhausen and Koppelman (1982). The second uses a cross-sectional analysis of differences in volumes among a number of zone pairs at one time, an example of which is reported by Su and Huffman (1975). An interesting combination of the two approaches has been reported by Moore and Soliman (1981).

In seeking a simple model which would forecast demand for air services to an "order of" accuracy, allow sensitivity analysis, and use published data which itself allows reasonable forecasting of independent variables, the Quandt-Baumol (1966) formulation of the gravity model was selected. This formulation of the model has previously been used in Australia, as by Francki and Lajovic (1974) to model inter-city bus travel in Queensland.

The Quandt-Baumol model states that:

$$V_{ij} = K \prod_{\ell} M_{i\ell}^{u_\ell} \prod_{m} M_{jm}^{v_m} \prod_{n} H_{ijn}^{w_n}$$

or, in logarithmic form:

$$\log V_{ij} = K + \sum_{\ell} u_j \log M_{i\ell} + \sum_{m} v_m \log M_{jm} + \sum_{n} w_n \log H_{ijn}$$

where:

V_{ij}	is the number of passengers demanding travel between zones i and j in a certain time period.
$M_{i\ell}$	is the ℓth measure relating specifically to zone i.
M_{jm}	is the mth measure relating specifically to zone j.
H_{ijn}	is the nth measure relating specifically to zone i-j.
i	ranges over the subscripts of all representative regions.
j	ranges over the subscripts of all feeder regions.

Advantages of the model are that the logarithmic transformation allows a unique but simple calibration to be performed using least squares techniques; the significance of each variable in explaining observed variations can be tested on the basis of which variables can be included or excluded from the final model; origin/destination data need only be obtained for a sample of zone pairs.

CHOICE OF INDEPENDENT VARIABLES

The two criteria for selection of independent variables for testing were their availability and the reliability with which they could be forecast. Three variables were tested describing a given zone and three describing the relationship between a pair of zones.

Variables Relating to a Given Zone ($M_{i\ell}$), (M_{jm})

P_iP_j Populations for each zone were obtained from the Australian Bureau of Statistics (ABS). Projections for use in forecasting were made using trend analysis to long term totals obtained from the Borrie report on Australia's population (1975).

A_iA_j Areas of catchment are also available from the ABS and, of course, do not change with time. The purpose of using area was to account for the low density or isolation of many catchments which could reduce the impedance of time and distance.

G_iG_j Guest rooms in hotels/motels are also available from the ABS and were anticipated to reflect both population and the level of tourist or business activity. Some problems were identified with changes in Census definition over time. Forecasts were made by applying a guest room rate to population forecasts. In future, it may be worth taking occupancy rates into account.

Y_iY_j Income was not included due to the difficulty of reliable forecasting of income by locality. Moore and Soliman (1981) included this parameter in their time dependent model.

Variables Relating to Pairs of Zones (H_{ijn})

T_{ij} Travel time is the most common measure of spatial separation between origin and destination. It was measured from timetables and hence included waiting time for connections.

F_{ij} Cost or fare for the journey has a linear relationship with travel distance. Because of the correlation with distance, fares were used by themselves in the analysis as well as in ratios such as cost per unit journey time, F_{ij}/T_{ij}.

R_{ij} Intermodal competition was ignored in the model of air travel and where competition from surface modes exists due to relative proximity of airports, the origin/destination pair was excluded. This need not be so; where competition from another mode is substantial, a dummy variable R_{ij} can be introduced to take on a value of 2.718 ($\log R_{ij} = 1$) where the alternative mode exists between i and j, or 1.0 ($\log R_{ij} = 0$) where it does not, Francki and Lajovic (1974).

N_{ij} Frequency of available services connecting i and j in a 24 hour period was tested in the model of air travel.

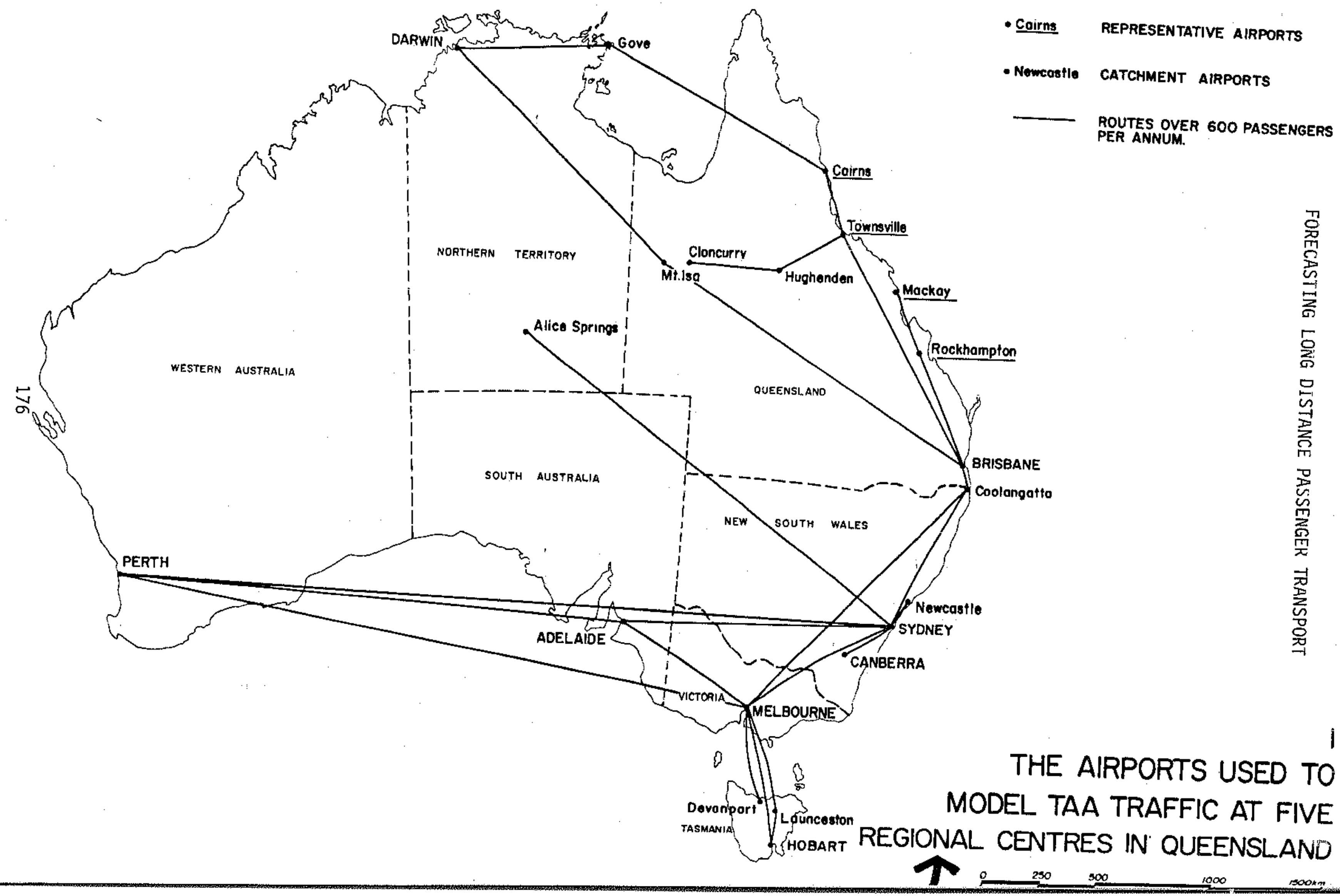
X_{ij} Community of interest may exist between localities based on social, economic or administrative bonds, as in the model by Moore and Soliman (1981) which readily available parameters do not reflect. This problem can, at times, be solved by stratification as in the study of bus travel by Francki and Lajovic (1974).

In the Maroochy study, the community of interest between the representative and the catchment regions was calibrated by entering a dummy variable for the three major capital cities; Brisbane, Melbourne and Sydney, and allowing the model to estimate the extent of the community of interest. In effect, the variable did not relate to a pair of zones but to individual catchment zones, X_j, where j took three specific values. The community of interest with the major capital cities relating to Queensland's regional centres is rationalised by the additional travel that is generated by what are significant administrative or commercial ties.

CALIBRATION OF THE MODEL

The flow of passengers between different zones was modelled for 1980/81 using the Quandt-Baumol model. The zones used were five representative regions and 16 catchment or feeder regions from the rest of Australia. These included all those having an interchange of at least 600 trips per year with one or more of the five representative regions. The Brisbane-Coolangatta route was excluded from the analysis as being too short and being dominated by road traffic. In all, the feeder regions represented in the model contained 60% of Australia's population. The set of zones used in estimating the model is shown in Figure 1.

Multiple regression was used to fit the model, the particular program being that in the MINITAB statistical package, release 81.1, from the Statistics Department, Pennsylvania State University, U.S.A. Automatic step-wise regression procedures were not used for variable selection. Instead, variables were discarded on the basis of failure to meet the partial F-test criterion.



THE AIRPORTS USED TO MODEL TAA TRAFFIC AT FIVE REGIONAL CENTRES IN QUEENSLAND

Combinations of variables based on logical grounds were considered. This meant that discarded variables would be reconsidered with other subsets of variables. The final model had to meet the usual goodness of fit criteria, viz, the global fit as measured by the coefficient of determination, R^2, had to be high and the internal fit as measured by the size and distribution of the residuals had to be good.

Further, predictions of total passenger flow and number of passengers per flight in 1980/81 were considered as a measure of fit. Validation of the model over time was by predictions of 1976/77 passenger flow.

Three alternative formulations were considered to reasonably represent air travel to and from the representative regions:

1st Alternative ($R^2 = 0.61$):

$$V_{ij} = 1.815\ (P_i P_j)^{0.524}\ A_j^{0.127}\ F_{ij}^{-1.35}$$

2nd Alternative ($R^2 = 0.84$):

$$V_{ij} = 2.25 \times 10^{-4}\ G_i^{1.45}\ G_j^{0.868}\ A_i^{0.209}\ T_{ij}^{-0.635}$$

3rd Alternative ($R^2 = 0.92$):

$$V_{ij} = 1.72 \times 10^{-2}\ G_i^{1.415}\ G_j^{0.335}\ T_{ij}^{-0.2714}\ e^{2.6C_1 + 1.8C_2 + 1.42C_3}$$

where variables are as previously stated, and

C_1 = 1 if j is Brisbane, 0 otherwise

C_2 = 1 if j is Sydney, 0 otherwise

C_3 = 1 is j is Melbourne, 0 otherwise

The third alternative model was finally selected, partly because of a better overall fit, and partly because the first and second models, while balancing out the errors, substantially miscalculated the travel associated with the three major capital cities. The details of the fitted model are given in Table 1.

TABLE 1

FINAL REGRESSION MODEL (TAA PASSENGERS)

PARAMETER	ESTIMATE	STANDARD DEVIATION	"t"
Constant	- 4.062	1.478	- 2.75
log (rooms-model)	1.4149	.1989	7.11
log (rooms-feeder)	.3348	.1052	3.18
time	- .27141	.06332	- 4.29
"Brisbane"	2.6006	.3939	6.60
"Melbourne"	1.4238	.3202	4.45
"Sydney"	1.8094	.3960	4.57

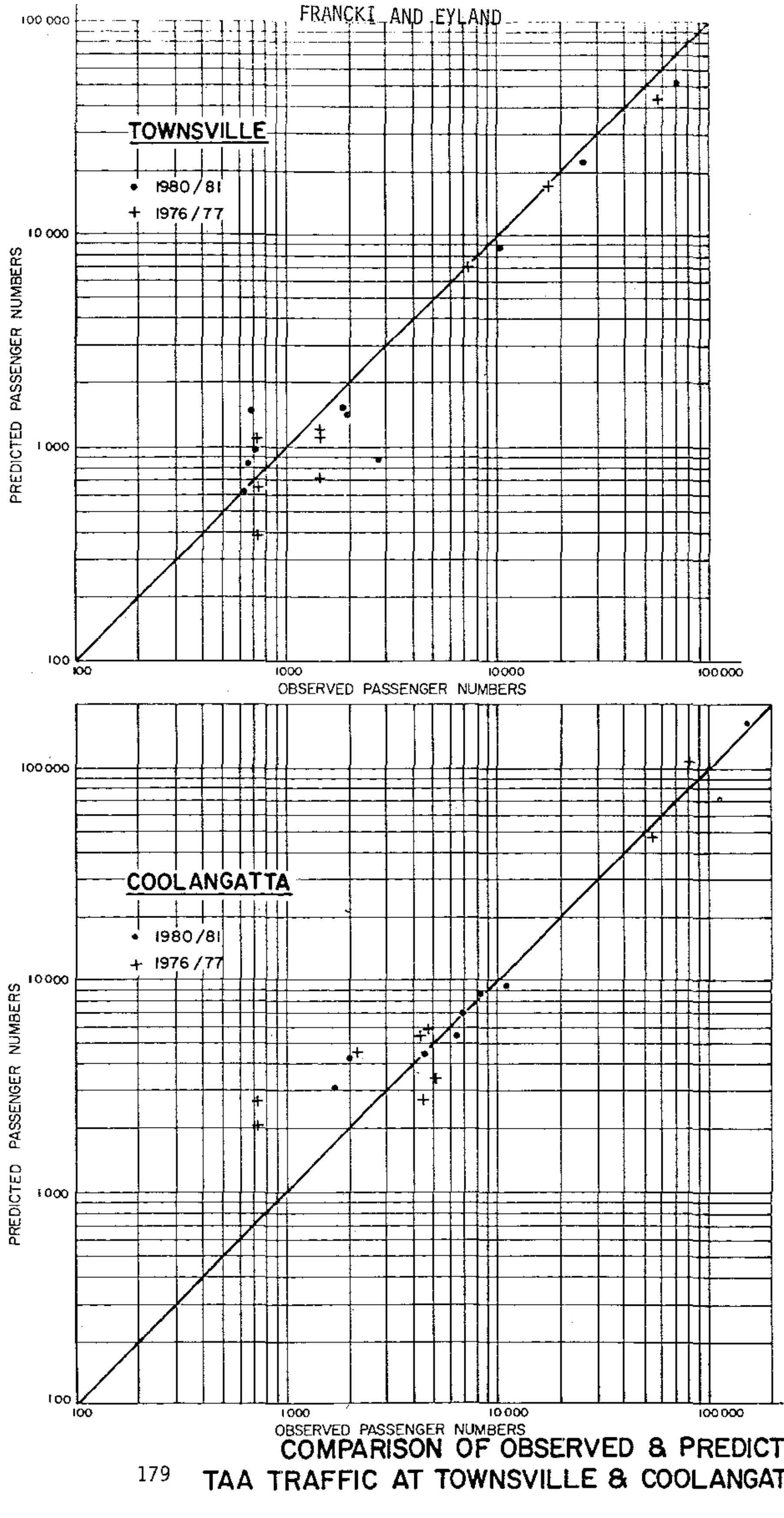
R^2 = 91.7% Std. dev. about line = .4939 D.f. = 34

The plot of predicted versus observed is shown in Figures 2 and 3 and a comparison of observed readings, together with estimates based on the fitted model is shown in Table 2.

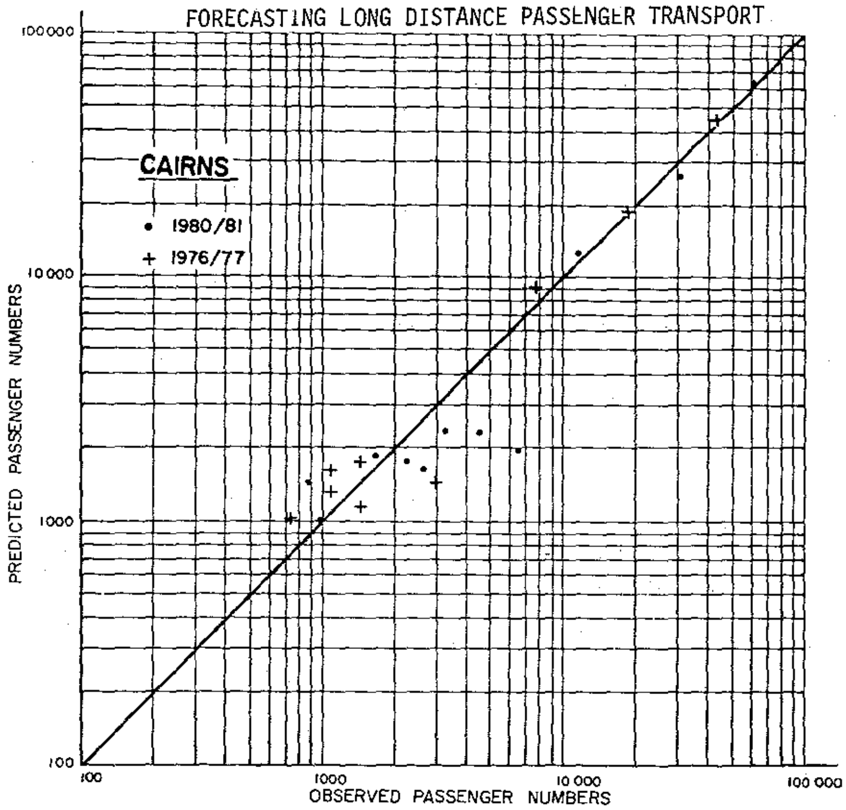
Some of the estimates for particular routes are substantially different from actual observed values. This was not important, since prediction for individual routes was not required. What was required was an estimate of the traffic volume at particular airports from all routes feeding into that airport. Comparison of estimated and observed total numbers of passengers and number of passengers per flight is given in Table 3.

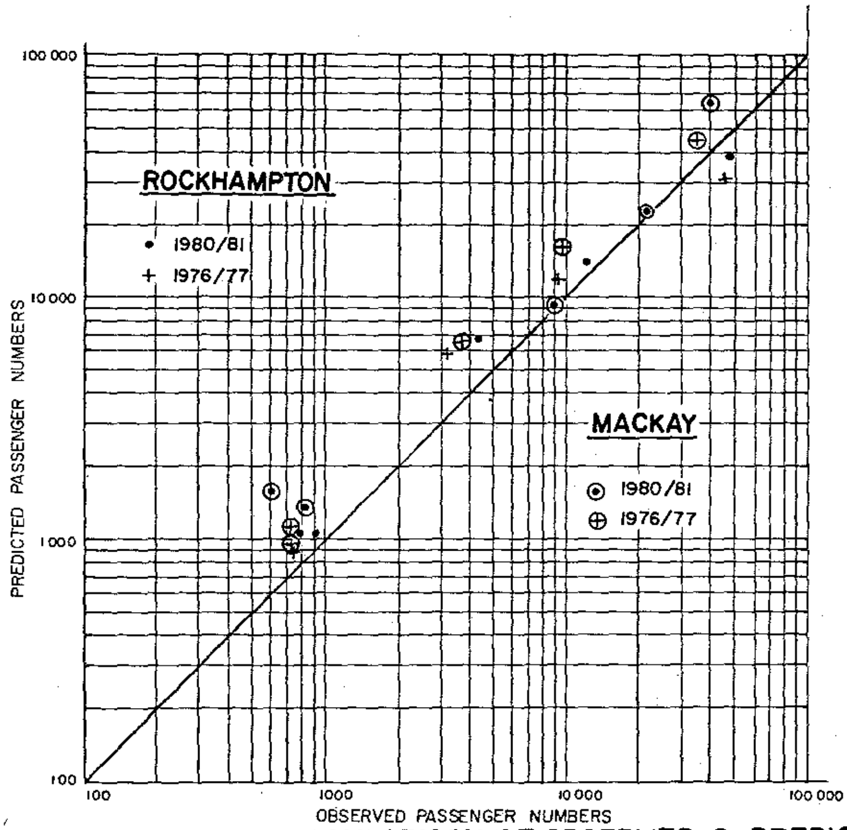
DISCUSSION OF RESULTS

The magnitudes of the exponents in each model agree with "a priori" assumptions. The novel feature is the assymmetry of the selected model. This model is intended to represent only a limited number of rows (departures) and columns (arrivals) of the total matrix of air travel. The model can therefore be asymmetrical in order to be more sensitive to the modelled terminal about whose development more detail will be known.



2
COMPARISON OF OBSERVED & PREDICTED
TAA TRAFFIC AT TOWNSVILLE & COOLANGATTA





3
COMPARISON OF OBSERVED & PREDICTED TAA TRAFFIC AT CAIRNS, ROCKHAMPTON & MACKAY

TABLE 2

OBSERVED vs PREDICTED: MODEL OF TAA PASSENGERS 1980/81

MODEL REGION	FEEDER REGION	CALIBRATION 1980/81 OBSERVED	CALIBRATION 1980/81 PREDICTED	VALIDATION 1976/77 OBSERVED	VALIDATION 1976/77 PREDICTED
Coolangatta	Sydney	150,268	161,325	81,395	107,078
	Melbourne	111,998	71,161	55,848	47,569
	Adelaide	11,073	9,317	4,745	5,898
	Canberra	6,413	5,407	5,110	3,475
	Newcastle	8,461	8,548	4,380	5,506
	Hobart	6,853	6,958	2,190	4,599
	Launceston	4,576	4,511	4,460	2,736
	Perth	2,017	4,269	730	2,771
	Devonport	1,718	3,133	730	2,083
Townsville	Brisbane	70,155	52,208	57,305	42,263
	Sydney	25,608	22,015	17,885	17,765
	Melbourne	10,232	8,879	7,300	7,216
	Adelaide	1,965	1,460	1,460	1,124
	Canberra	1,873	1,561	1,460	1,219
	Newcastle	658	849	730	665
	Darwin	2,722	890	1,460	728
	Alice Springs	680	1,503	730	1,180
	Perth	1,028	498	730	393
	Hughenden	701	995	n.a.	812
	Cloncurry	626	628	n.a.	486
Cairns	Brisbane	59,909	62,993	42,340	45,672
	Sydney	30,209	26,562	18,250	19,198
	Melbourne	11,556	12,849	7,665	9,353
	Adelaide	2,627	1,682	1,460	1,160
	Mt Isa	3,418	2,732	1,460	1,740
	Canberra	1,670	1,883	1,095	1,318
	Newcastle	867	1,474	730	1,034
	Darwin	6,444	1,978	2,920	1,449
	Alice Springs	4,500	2,321	1,095	1,633
	Perth	976	1,033	730	306
	Gove	2,231	1,765	3,285	1,273
Rockhampton	Brisbane	46,935	38,053	45,260	33,513
	Sydney	12,136	14,047	9,125	12,333
	Melbourne	4,327	6,777	3,285	5,992
	Adelaide	789	1,067	730	894
	Canberra	905	1,063	730	903
Mackay	Brisbane	39,190	61,775	34,675	45,352
	Sydney	21,363	22,743	9,490	16,644
	Melbourne	8,807	9,173	3,650	6,761
	Adelaide	823	1,379	730	963
	Newcastle	603	1,581	730	1,123

TABLE 3

TOTAL ESTIMATED FLOW INTO REPRESENTATIVE REGIONS

(a) Passengers/IAA Flight

Model Region	Predicted	Observed
Coolangatta	692	715
Townsville	209	255
Cairns	316	402
Rockhampton	307	329
Mackay	504	363

(b) Total IAA Passengers 1980/81 : Calibration

Model Region	Predicted	Observed	Predicted / Observed
Coolangatta	274,626	303,381	0.91
Townsville	91,487	116,248	0.79
Cairns	117,271	124,410	0.94
Rockhampton	61,005	65,090	0.93
Mackay	96,650	70,785	1.37

(c) Total IAA Passengers 1976/77 : Validation

Model Region	Predicted	Observed	Predicted / Observed
Coolangatta	181,715	159,588	1.14
Townsville	31,588	31,755	0.99
Cairns	38,688	38,464	1.01
Rockhampton	53,635	59,130	0.91
Mackay	70,843	49,275	1.44

The first model of air passengers is substantially symmetrical and of the classical form. The exponent 0.5 of population, P_iP_j, has a similar value to that found in Gravity Model simulations of travel. The exponent -1.35 of the variable measuring impedance, F_{ij}, also has a value within the range commonly encountered. The influence of the variable A_j with an exponent 0.127, on the estimate of passengers is small.

The second model uses guest rooms, G, as the indicator of size and activity at each trip end. It is asymmetrical with an exponent 1.45 at the modelled end and 0.868 at the feeder end. This has the logic of reflecting the character of the trip which is largely dependent on the modelled location. The influence of spatial impedance is relatively low with travel time, T_{ij}, having an exponent of -0.635.

The third, selected model allows for the community of interest between the representative regions and Brisbane, Sydney and Melbourne. The strongest community of interest is associated with Brisbane where the resulting multiplier is $e^{2.6}$ or 13.46. Brisbane is the capital city of the State within which all the representative regions are located and, as such, must generate considerable business and administrative travel as well as travel generated by social and family ties. The multiplier for Sydney and Melbourne is 6.11 and 4.12 respectively; the relative value is understandable as both Melbourne and Sydney are Australia's major centres but Sydney tends to be the business capital for development activity in the north of Australia.

The presence of the community of interest factors associated with the three capital cities which tend to dominate air travel on the Queensland coast, reduces the importance of the variable measuring the impedance of distance; T_{ij} now has an exponent of -0.271. Similarly, the influence on the model of the parameter describing the feeder region, G_{ij}, is also reduced to an exponent of 0.335.

In total, the model has predicted TAA passenger movement at Cairns and Rockhampton within 10%, both in 1976/77 and in 1980/81. The total prediction for Coolangatta and Townsville has been more variable over the two periods. Mackay has been consistently and substantially over-estimated. Over the period of the study, passenger numbers at Mackay's airports for both domestic airlines have shown a decline, although TAA has increased its share of the market to show a nett increase; the guest rooms per 1,000 population have remained high and above those in the remaining representative regions with the exception of Coolangatta, as shown in Table 4.

TABLE 4

RELATION OF GUEST ROOMS TO POPULATION AT JUNE 1976

AIRPORT	REGIONAL POPULATION	GUEST ROOMS	GUEST ROOMS/ 1,000 POPN.
1. Coolangatta	119,423	2,388*	20.0*
2. Townsville	145,461	1,051	7.2
3. Cairns	124,663	1,345	10.8
4. Rockhampton	66,743	786	11.8
5. Mackay	77,037	1,159	15.0
6. Sydney	3,021,982	11,260	3.7
7. Brisbane	957,746	2,975	3.1
8. Melbourne	2,604,037	6,655	2.6
9. Adelaide	900,434	2,716	3.0
10. Mt Isa	27,633	220	8.0
11. Canberra	215,463	2,013	9.3
12. Newcastle	363,011	1,194	3.3
13. Darwin	39,193	853	21.8
14. Hobart	162,062	1,471	9.1
15. Alice Springs	21,673	442	20.4
16. Launceston	109,571	805	7.4
17. Perth	1,000,019	5,181	5.2
18. Gove	13,546	224	16.5
19. Devonport	88,861	699	7.9
20. Hughenden	4,317	25	5.8
21. Cloncurry	5,504	26	4.7

* Estimate

RELATION OF GUEST ROOMS TO POPULATION AT JUNE 1980

AIRPORT	REGIONAL POPULATION	GUEST ROOMS	GUEST ROOMS/ 1,000 POPN.
1. Coolangatta	165,250	3,232	19.6
2. Townsville	156,650	1,239	7.9
3. Cairns	134,030	1,714	12.8
4. Rockhampton	70,300	873	12.4
5. Mackay	86,730	1,464	16.9
6. Sydney	3,231,650	11,321	3.5
7. Brisbane	1,028,930	2,963	2.9
8. Melbourne	2,759,700	6,549	2.4
9. Adelaide	934,000	3,145	3.4
10. Mt Isa	27,300	254	9.3
11. Canberra	246,700	2,229	9.0
12. Newcastle	384,950	1,313	3.4
13. Darwin	50,612	823	16.3
14. Hobart	170,170	1,497	8.8
15. Alice Springs	24,403	482	19.8
16. Launceston	115,490	1,059	9.2
17. Perth	1,109,190	5,566	5.0
18. Gove	14,000	227	16.2
19. Devonport	93,750	699	7.5
20. Hughenden	2,850	26	9.1
21. Cloncurry	4,250	29	6.8

The correct inference may be that the number of guest rooms should be modified by occupancy rates to provide a more reliable indicator of air passenger generation. Further, it would clearly be desirable to model total domestic demand by both Ansett and TAA and their subsidiaries. Only TAA's origin/destination statistics were available to the authors.

Comparison to Other Models

Despite a reasonable validation at a second time interval, a disadvantage of the model remains its cross-sectional character, whereas air travel is notoriously variable over time. Moore and Soliman (1981) have maintained a differentiation between pair effects and year effects and claim a significant improvement in the model as a result; naturally, the size of the data base increases substantially.

It is interesting to note that Moore and Soliman calibrated a gravity model using population, income and fare but improved it substantially by adding a dummy variable for combined area-pair and time effects which the authors would have called a community of interest factor. It is possible that the need for a dummy variable indicates the existence of an independent variable which remains to be identified.

Francki and Lajovic (1974) have successfully calibrated a gravity model for inter-city bus travel in Queensland. The model was symmetrical because the objective was to obtain passenger demand on a state-wide road system in order to plan bus services. The calibration was improved by stratification. The independent variables in all cases contained population, travel time and, with one exception, zone area averaged over the pair. Capital city trips were kept separate and, in addition to population and time, contained a variable of unit fare and a dummy variable to take account of competition from rail where such existed.

No comparisons with reports of disaggregate modelling of air travel were made. The time horizon for this study was such that there would have been no opportunity to take advantage of any greater accuracy of a 'disaggregate model' because of the difficulty in reliably forecasting the necessary independent variables.

The total cost of the aggregate modelling reported here was $9,000.

CONCLUSIONS

The Quandt-Baumol formulation of the gravity model is suitable for the simulation and forecasting of long distance passenger transport. Its major advantages are its ability to use published data for a sample of zone pairs and the simplicity and flexibility of its calibration.

The model's limitation is its requirement for clear spatial separation between zones, hence its inability, admittedly untried, to model travel in a relatively homogeneous area. Its use has generally been applied to a single mode although it could be combined with a mode-split model thereby losing its simplicity.

The model has been termed empirical although this should not detract from its use where the objective is the validity of the forecast and not the understanding of the forces behind the forecast. The model is deserving of greater use, and with use will come better selection of variables and a better understanding of the forces represented by them. Guest rooms in motels/hotels have provided one such variable in determining travel which is expected to be tourist and business oriented.

An asymmetric model has advantages when considering travel at the margin of a total set of zone interchanges. The model can be made more sensitive to parameters describing specific zones. This has the further advantage of simplifying sensitivity analyses by varying only the parameter value at the subject zones and leaving constant the values at all other zones in the catchment.

The gravity model shares the disadvantage with many others of cross-sectional analysis, particularly in forecasting air traffic which is notoriously variable over time. A combination of time series and cross-sectional analyses, as reported by Moore and Soliman (1981) promises to improve the model substantially without an excessive increase in complexity.

The authors recommend the use of the Quandt-Baumol formulation of the gravity model to practitioners involved in strategic or sketch planning of either the network or the services for long distance passenger transport. The simplicity of the model's use will fit the tightest of budgets and will give form and credibility where educated guesswork may often be the only feasible alternative.

ACKNOWLEDGEMENTS

The authors would like to acknowledge, in the first place, the help of Mr Brian Porter and staff of TAA (Trans Australia Airlines) for their cheerful assistance and the sample of origin/destination statistics without which the model would have been impossible to construct.

The study was carried out under the auspices of Gutteridge Haskins and Davey Pty Ltd, consulting engineers, for Maroochy Shire in Queensland. Previous work on the same model by Francki and Lajovic (1974) was carried out under the auspices of Rankine and Hill, consulting engineers, for the Department of Transport in Queensland.

Last, but not least, the authors would like to acknowledge the willing assistance of Beth O'Meara in the data preparation and the typing of the paper through many drafts.

REFERENCES

Borrie, W.D. (1975). "Population and Australia: demographic analysis and projection", First Report of the National Population Enquiry.

Francki, M., and Lajovic, T.E. (1974). "Modelling of intercity bus travel", Lecture, Institution of Engineers, Australia Transportation Branch (unpublished)

Moore, O.E., and Soliman, A.H. (1981). "Airport catchment areas and air passenger demand", Transportation Engineering Journal of ASCE, Vol.107, No. TE5.

Oberhausen, P.J., and Koppelman, F.S. (1982). "Time series analysis of intercity air travel volume", Transportation Research Record 840.

Quandt, R.E., and Baumol, W.J. (1966). "The demand for abstract transport modes: theory and measurement", Journal of Regional Science, Vol.6, No. 2.

Su, V., and Huffman, L. (1975). "The demand for air travel between New York city and other large cities", Transportation Research Record 529.