Projections of regional air passenger flows in New Zealand 2018-2043

Tim Hazledine

Professor of Economics at the University of Auckland

Final Report

May 17, 2016

For the Ministry of Transport, Wellington

Acknowledgements

Dr Haobo Wang provided excellent overall management of this project, and he, Tim Herbert, Ralph Samuelson, Helen Huang, David Ryan, Sonya van de Geer, Richard Cross and John Macilree gave guidance and comments, and were generous in sharing their expert knowledge of the NZ air transport sector.

1. Job Specification

Purpose

The purpose of this modelling work is to produce long-term projections of regional air passenger counts in New Zealand, which include domestic air passengers travelling from one region to other regions, and counts of international air passenger departures from each region. The time horizon will be five-year intervals from 2018 to 2043.

Definition of regions

The regions to be used correspond to the Regional Councils of New Zealand; however Nelson and Tasman will be combined into one region (Nelson-Tasman), and the Chatham Islands will be combined with Canterbury. So, the modelling will be based on the following 15 regions: Northland, Auckland, Waikato, Bay of Plenty, Gisborne, Hawke's Bay, Taranaki, Manawatu, Wellington, Nelson-Tasman, Marlborough, West Coast, Canterbury, Otago, and Southland. It is noted that some regions have more than one airport.

Scope

The modelling work is to generate two main outputs:

(1) Projections of number of annual domestic air passengers travelling from each region to each of the other 14 domestic regions. The projections will reflect the passengers' actual origin and destination, without regard to any intermediate connections that may have been required. In other words, the projections will reflect true passenger origins and destinations, not "legs". Thus, these projections will exclude passengers travelling on the domestic legs of international journeys, and will not pick out arrivals/departures at airports not the endpoints of multi-stop domestic itineraries, so that total national passenger departures calculated by this method will not match with total national counts of airport domestic departures.

(2) Projections of international air passenger departures from each region, disaggregated to Australia and Rest of World destinations. Again, each passenger will be counted only once, at his or her true origin, without regard to any connections required. No estimation of international arrivals is expected.

Procedure

The job is basically in two parts: econometric modelling of actual passenger flows over the 2009-15 period, followed by application of the achieved econometric model to generate forecasts through to 2043.

Outline of this Report

Section 2 will explain and review the literature on the "gravity model" which has become standard in empirical work explaining economic flows across time and space, and which will here be developed into an econometric model explaining quarterly variations (2009-15) of airline passenger numbers across routes within New Zealand (and to Australia and the rest of the world). This model will underpin the future projections.

Section 3 will define and source the data to be used in the econometric model. Section 4 will show and discuss the estimated model. Section 5 will explain and show the projections of future domestic travel up at five-yearly intervals through to 2043. Section 6 reports models of outbound travel, to Australia and to the rest of the world. Section 7 concludes briefly.

2. The gravity model

The basis of the econometric investigation into the determinants of inter-regional air travel in New Zealand is the gravity trade model, which has been successfully used in hundreds of econometric models of inter- and intra-national spatial flows – of goods, services, capital, information. [1]

The essence of the gravity model is the hypothesis that flows between two geographic entities (country/region/city) are monotonic and possibly non-linear functions of the GDP and/or populations of the two end-points (the "attractors") and the distance between them, which is expected to deter flows.

The gravity model is typically estimated with the flows and the GDP/population and distance variables in natural logs, so that the estimated coefficients can be interpreted as elasticities. Other variables can be added. If these are continuous variables with an expected monotonic effect they can be logged as well. If they are non-continuous, then they will generally be entered linearly, for example as dummy variables.

The gravity model applied to international trade

The gravity model has been remarkably successful at explaining trade flows, and has become the dominant model in empirical trade research over the past twenty or so years. In econometric estimation, it even tends to mimic the strictly proportional effects predicted by the physical gravity equation, although from an economic theory perspective, this is actually a problem: those effects are "too large". In particular, the natural rationalisation of the effect of distance on trade as being due to transport costs does not add up: why, for example, would increasing distance from, say, 500 kilometres to 1000 kilometres wipe out the profitability of about one half of the

[1] See Hazledine (2009), (2014); Gillen and Hazledine (2015) and references therein for an entrée to the gravity modelling literature.

potential trade flows, purely on the basis of the added costs of moving goods the additional kilometres? Grossman (1998) plugged in plausible numbers for transport and other costs, and found that the distance effect thereby predicted is only about one thirtieth of the proportional gravity model effect!

This puzzle – the over-importance of distance – can now be seen as setting the stage for John McCallum's (1995) research, which plugged an empirical gap in models of merchandise shipment flows, which in previous gravity model research had always been confined to trans-border transactions, for the simple reason that it is these flows that are routinely recorded for customs purposes. McCallum made use of an unusual Statistics Canada database tracking shipments of merchandise goods between Canadian provinces and between provinces and U.S. states in 1988. Estimating the gravity model on these shipments data, McCallum found that he needed an enormous dummy variable to enable it to encompass the two types of shipment flows. Basically, for given values of incomes and distance, trade between the provinces is much more likely than trade across the border. For example, a business in British Columbia was more than *twenty times* more likely to import goods from Ontario than a business in Washington State, even though both are about the same distance from Ontario. So, there really is `too little trade' -- it is not (just) a matter of distances between agents.

The problem of a lack of data on internal flows of goods -- the Statistics Canada database used by McCallum and others is just about unique -- can be partially dealt with using micro-level Census data on the location of people within countries, as Helliwell and Verdier (2001) were the first to demonstrate. Applying their method to Australia-New Zealand trade flows, Hazledine and Lipanovic (2004) found another huge border effect between these two countries, even though the economic barriers between them are actually lower than they are between Canada and the United States.

It soon became evident that the border effect was not, empirically, an anomaly. Frankel *et al* (1995) had already reported the 'remarkable' finding that countries with a common linguistic/colonial heritage tend to trade 65% more with each other than they would in the absence of these links. And there is the equally remarkable over-sized distance coefficient, noted above. So what theory could synthesize and explain these relationships?

The answer came out of the emerging economics of imperfect information, missing from traditional theories such as Heckscher-Ohlin but now increasingly in the centre of how economists understand the world. Trade is a result of people doing business with each other. Doing business is a tricky, risky activity. Information is key, but information is both imperfect and unevenly distributed. Formal and informal institutions, norms and customs serve to mitigate the effects of imperfect information as a deterrence to commerce and trade. And it suddenly seemed highly plausible -- even, obvious – that such effects would be related to borders and distance. People within a country would share customs and institutions that would make it easier for them to do business with each other than with foreigners. And, both within and across national borders, distance between potential partners might well be linked with 'unfamiliarity'; with what we call the 'transaction costs' of doing business.

These considerations will be born in mind when we now move on to the enterprise of testing the gravity model on data on 'trade in people' -- passenger air travel.

Applying the gravity model to passenger air traffic

It would seem a natural fit to apply the gravity model to 'trade' in people; specifically to air passenger travel movements between cities or regions, and Matsumoto (2007) reports that the earliest such study was by Harvey (1951). However, despite this long lineage, and a similarly long lineage of gravity models of merchandise trade, the two literatures have been disjoint. The empirical trade studies never, to my knowledge, mention trade in air travel; and the papers modelling air travel have not, with two exceptions (Klodt, 2004; Hazledine, 2009), shown any awareness of the border effect that has in recent years preoccupied commodity trade modellers. The latter may be in part due to data limitations – the most readily available aviation databases are restricted to international passenger movements, such as the ICAO database used by Matsumoto, or to internal US flying activity, using the well known Department of Transportation DB1B 10% sample (Bhadra and Key, 2008). Another factor may be that the primary concern of aviation transportation economists has usually been with price, not quantity. For example, Dresner *et al* (1996) were interested -- as are many U.S. aviation specialists – in the impacts on market prices of the entry of Low-Cost Carriers (LCCs), such as Southwest Airlines. For this, they estimate an equation to explain price. But this equation has demand -- number of passengers -- in it, and demand is itself a function of price, so the appropriate thing from an econometric point of view is to also estimate a demand function, allowing for the simultaneity between price charged and quantity demanded.

That is, the demand equation, which is a gravity model in trade economics terms, is not the prime focus of attention – it is just there to clean up the estimation of the pricing model. This is something of a pity, since these air travel gravity models[2] are actually more advanced than the standard trade models, because they do include an important price variable -- the actual average price, or 'yield', achieved on a route.

Now consider the effect on passenger numbers of the distance between the two cities. In standard trade gravity models, as noted above, distance (usually between countries) is invariably a strongly negative factor in size of trade flows, often with an elasticity of around -1. Physical distance generates economic distance, it seems, to an extent which has caused much concern, since the elasticity is an order of magnitude larger than could easily be justified by the obvious explanation – transport costs, which are usually around 10% or less of the value of the goods transported. Since the airfare component of passenger air travel is, in general, a much larger fraction of the total

[2] Another such study is Reiss and Spiller (1989). Dresner *et al* (1996, footnote 3, p310) give references to other studies of the determinants of air traffic and/or prices.

cost of the transaction (in this case, a trip), and since airfares are strongly related to distance (see Gillen and Hazledine (2012) for Canada and trans-border airfares, and many other studies), we might expect air travel gravity models to show even larger distance coefficients, but they don't. For example, Hazledine (2009) finds distance elasticities between -0.2 and -0.4, which is a range consistent with normal price elasticities of demand, given that the elasticity of airfares with respect to distance is around -0.6 (Gillen and Hazledine (2012)) and allowing for the likely proportion air fares are of the total trip cost.[3]

In any case, in studies of air travel flows within North America, it is common to find that distance does not have the same repellent effect as it does for commodity trade, and, indeed, that distance may even somehow encourage air travel. The availability of ground transportation substitutes for air travel in continental North America – in particular, private automobiles – could plausibly explain a positive relationship between distance and air travel over shorter distances – say up to 1000kms -- as the attractiveness of air relative to road changes with longer journeys, but we would naturally expect that air travel does not have significant modal competition on longer routes, so that the "usual" negative effect of distance would show through eventually.

This consideration suggests using a non-linear specification of the role of distance: Dresner *et al* (1996) do this with a quadratic formulation, but (unremarked by them) the estimates of this have the "wrong" curvature: U-shaped rather than the inverse of this.

Chi and Baek (2012) estimate a fairly standard gravity model for quarterly aggregate US air travel. Their model includes the number of seats on offer, which can hardly be expected to be exogenous to demand shifters such as population.

Hsaio and Hansen (2011) explain US city-pair air passenger numbers with a gravity model augmented with airfare and frequency variables. Airfare is instrumented (by route distance and jet fuel price), but flight frequency is included as though an independent exogenous variable, which it is unlikely to be.

Wei and Hansen (2006) estimate an interesting model of demand in the hub-and-spoke network setting which is quite prevalent in the US, but not in New Zealand or anywhere else. Their model is generally quite successful, but shows some surprising coefficients: negative on income; negative on total trips made to a hub; positive on distance.

Gillen and Hazledine (2015) analyse supply of air services to regional towns or cities in six countries: NZ, Australia, Canada, Sweden, Norway, and three US states (Arizona, Colorado, Oregon). Lacking data on actual passenger numbers, they proxy

[3] It is however probably slightly disturbing that the Hazledine (2009) elasticity estimates are twice as large for business than for leisure travellers.

this with the number of seats offered on point-to-point routes. Although it is reasonable to use number of seats as a proxy for number of passengers using the seats, a limitation of their approach is that it cannot distinguish point-to-point travel from trips originating or terminating at a third airport (ie, one- or two-stop itineraries). Gillen and Hazledine (2016) focus on pricing of airfares on regional routes in Australia and New Zealand

In summary, whereas in the modelling of commodity and service trade flows estimated distance effects are always “too large”, in the air travel gravity models it has proved difficult to find any significant or sensible distance effects at all. Apart from this issue, the literature reveals that the gravity model performs well empirically in both settings – in particular, the attractors of population and incomes show through strongly, and we now proceed to gathering and analysing the data needed to model passenger air travel within New Zealand.

3. Variables and Data

Table 1 gives descriptive statistics for the variables used in the econometric modelling of domestic passenger flows. Note that, over the data period from Quarter 3, 2009 through to Quarter 2, 2015, 28 NZ towns or cities had, at some time in this period, scheduled air service to at least one of the three large “trunk” cities[4]. This implies a possible 478 (=28x27/2) routes. However, some city pairings were not reported in the data[5], and others were eliminated from the econometric modelling, as noted below, for having too few passengers, leaving observations on passenger numbers on 231 routes, most of these requiring one- or two-stop journeys.

PAX measures quarterly origin/destination travel between each city (or town) pair in NZ with scheduled air service, from the 3rd quarter of 2009 through the 2nd quarter of 2015. There are 14028 rows in the original database supplied by the Ministry of Transport which after “cleaning” is reduced to 13237 rows.[6] Table 1 shows that the average number of quarterly passengers in each direction is 4,444, and that the busiest route (Auckland-Wellington) had in one quarter 284,360 fare-paying passengers.[7] There are many routes – all of them with one- or two-stop itineraries – with very few passengers noted in the database. In the econometric modelling we will simply restrict

[4] Paraparaumu gained service to Auckland in 2011; Masterton and Wanaka lost service (to Auckland and Christchurch, respectively) in 2013.
[5] For example, only 11 of the 27 destinations that can be accessed from Timaru had data.
[6] Cleaning involves (i) eliminating routes with just three or four quarterly observations, (ii) eliminating a few routes with suspiciously high airfare numbers, (iii) eliminating various routes not currently served by scheduled air transport (Oamaru, Mt Cook ….), (iv) eliminating quarters with zero passengers reported.
[7] This was the number who flew from Auckland to Wellington in Q3, 2014.

the estimation to routes with at least 50 quarterly passengers in one direction. This will reduce the sample size to 10,940.[8]

Note that these data are "origin-destination" in the sense that a passenger leaving one airport (the origin) is counted as travelling to wherever they finally step off a plane (the destination), whether or not this involves changing planes en route. What these data do not tell us is whether, on each observed trip, the passenger is outward or inward bound. That is, these data – impressive and valuable though they are – do not give us the true origins of inter-city travel – meaning, basically (in most cases), where travellers live.

The data are for travel in both directions – that is, we have two observations for each quarter/route, being passenger numbers in each direction. It is immediately noticeable in the data that travel within NZ is not perfectly balanced, directionally, though it is generally quite close to balance. To take a quite extreme example, in the 3rd quarter of 2009, 29,221 people flew from Christchurch to Queenstown, and 30,655 flew in the other direction. This could mean that rental car companies have to deal with a surplus of North-South travellers on this route. However, we here will in essence be treating travel in each direction on a route as just multiple observations of the same thing. That is, we will not be distinguishing direction of travel in the econometric modelling, and nor in the forecasting exercise. A "route" is a city-pair, in no particular order of end-points. However, we will, for each route sort the two population numbers into larger and smaller.

WPOPLARGE and **WPOPSMALL** are derived from quarterly population estimates for the likely catchment area for each airport, broken down by urban and rural residents. These data were calculated by linear interpolation and extrapolation from 2006 and 2013 National Census normally-resident population counts at the TLA (Territorial Local Authority) level. This was not a trivial exercise, especially in the case of Wellington and Queenstown airports.[9] It was found that the best econometric results were yielded by weighting each rural resident at just 0.25 of each urban resident, indicating a considerably lower propensity for rural dwellers to use commercial aviation services for domestic travel.

For each route, the larger of the two end-point populations is distinguished from the smaller. This will allow us to identify any tendency for the larger population to exert more (or less) influence on travel numbers than the smaller.

GDPPOPD and **GDPPOPO** are regional per capita GDP measured by Statistics NZ in 2013. That is, these are cross sectional variables picking up systematic differences in regional incomes. Preliminary econometric work showed no significant difference

[8] The model can be estimated on all the observations without much effect on coefficients and their significance, though the overall goodness of fit is lower.

[9] Queenstown took on the other half of the Wanaka population after the cessation of direct service (to Christchurch) in 2013Q1; Wellington also takes on (most of) Masterton and the Wairarapa after 2013Q1, and loses half of Kapiti Coast after Auckland-Paraparaumu service began in 2011Q4.

in the coefficients on these variables when entered separately (and nor would we have any reason to expect such), and it is the product of the two income measures that will appear in the regression model.

QGDPPOP is quarterly regional per capita GDP. Such data are not presented by Statistics NZ. The series was therefore calculated using annual regional data on per capita GDP, with national quarterly data used to disaggregate this quarterly at the regional level. This procedure therefore assumes that within-year cyclical and seasonal variations are similar across regions.

DISTANCE is the direct ("as the crow flies") distance between the end-points of each route, in kilometres, and supplied by the Ministry of Transport. The shortest route is Blenheim-Wellington; the longest is Kaitaia-Invercargill, which of course was not flown directly.

The number of quarterly visitors arriving in New Zealand from a foreign port (**VISITORS**) is quite seasonal **--** peaking in quarters 4 and 1 – and also trended upwards, with an average annual increase of about 3%. These are Statistics NZ data.

SWITCHBACK, **FERRY**, are "dummy variables" taking the values 0 or 1. Although New Zealand might be characterised as being geographically a rather "linear" country, laid out NE to SW, there is enough width that nearly one third of the possible routes are "switchbacks" involving a trip in one direction followed by another coming back more than 180 degrees in the opposite direction. An example is Napier-Gisborne, which must be flown via Auckland.[10] Nearly one half of the routes have end-points on different islands, and so would require a Wellington-Picton **FERRY** journey across Cook Strait, if undertaken by car, bus or train.

The number of nonstop routes (**NNONSTOP**) is zero in most cases[11], and is at its highest between Auckland and Wellington, for which route Air New Zealand and Jetstar between them put on about 23 daily flights in each direction.

FARENZ is the average airfare paid on all tickets sold in a quarter on a particular directional route, converted from $US in the data to their (original, in most cases) NZ dollar values at quarterly exchange rates. These data (and the passenger numbers) were supplied to the Ministry of Transport as proprietary "Sabre Global Demand Data". Although the data overall seem plausible, there may be some sampling or other errors, and the tiny number of very high average airfare numbers could be examples of such. In the econometric modelling, airfare is deflated by the quarterly Consumer Price Index (CPI)'

Note that the econometric modelling will be at the actual city-pair route level, not the "regional" level at which the eventual forecasting exercise will be carried out.

[10] Auckland-Gisborne and Auckland Napier/Hastings are 337 and 330 km flights; Gisborne-Napier would be a 132 km flight, and is somewhat longer by road (218kms).
[11] Just 18% of all domestic routes have non-stop (direct) service.

Sometimes – eg Auckland-Blenheim – this is the same thing, but seven of the fifteen NZ regions have two scheduled carrier airports, and the city- or town-pair routes to and from these airports are not in general close substitutes, and need to be modelled individually. For example, the Auckland-Dunedin routes will for the vast majority of travellers not be equivalent to Auckland-Queenstown; nor Tauranga-Wellington with Rotorua-Wellington (though there is more of a case for combining Paraparaumu and Wellington routes).

For forecasting, the up to four routes (including, of course, one-stop itineraries) between Region A and Region B will be aggregated to the regional level.

4. Econometric results

Models were estimated by Ordinary Least Squares on the Eviews8 package. The model was built up in a reasonably small number of stages. It follows standard gravity modelling practice by taking natural logarithms of the dependent variable, PAX, and of those explanatory regressor variables which are (a) continuous, and (b) expected to have a monotonically positive or negative effect on passenger numbers. Other variables, including dummy variables, are entered as exponentials, which can accommodate zero values.

Notable omissions from the final model are quarterly dummy variables. Perhaps surprisingly, there is no significant seasonality to passenger air travel in New Zealand, at least at the quarterly level.

Also perhaps surprisingly, we are unable to detect any cyclical variation in domestic air travel, at least using the quarterly regional per capita GDP variable QGDPPOP. This may indeed be how things are in New Zealand, but it should be noted (a) that our data cover just six years of data, all being years of moderate macroeconomic recovery from the 2008-09 Global Financial Crisis; (b) that imposing national quarterly variation on the constructed quarterly regional GDP data will introduce error, which will be amplified when origin and destination per capita GDP data are multiplied together, as they are in a gravity model set-up.[12]

Regression 1 on Table 2 shows the fully specified model with no allowance for any linkages between the regressors. The goodness of fit of the model is very high by the standards of cross sectional gravity models[13], and so are the t-statistics of every one of

[12] Of course, for the eventual purpose of long-run forecasting at five-yearly intervals, the presence or lack of seasonal and cyclical effects in the model are not an issue.

[13] Although this in format a panel database, with 24 quarterly observations, most of the “action” is generated by the large cross sectional differences in the variables across the routes. That is, the estimation is probably best seen as involving repeated observations of a cross sectional model, something which will be reflected in the very high t-statistics estimated for most variables’ coefficients, these in turn obviating concern about heteroscedasticity, which does not in general bias cross sectional coefficient estimates, but which can bias (in either direction) the size of standard errors/t-stats. In this cross sectional or near-cross sectional context, it is common in the literature to only show R^2 and adjusted R^2 as measures of the overall fit of an OLS model.

the coefficients. However, there is an inherent problem with this specification, which is of particular relevance given the ultimate use of the model for forecasting, in which projections of regional population growth will be key.

The problem stems from two characteristics of passenger air travel -- actually, characteristics of public transport/transit (PT) systems in general -- which are not factors (or have been in any case ignored) in the commodity trade gravity modelling literature.

The first special characteristic of PT is that demand is not independent of supply. Specifically, the attractiveness of a PT service may be notably affected by the frequency (and speed) of the service, and this in turn is largely determined by the size of the market -- the realised demand. Because of fixed costs and "lumpiness" constraints on aircraft size, the number of daily flights that can be economically offered by an airline depends above all on the number of people who might thereby be induced to fly. Our Model 1 shows that the number of nonstop flights is a strong predictor of passenger numbers, but this number (varying from zero upwards) can itself be expected to be significantly determined by end-point populations, which are themselves regressors in the model. That is, NNONSTOP is likely to be crowding out some of the ultimate effect on PAX of the population variables, which is a problem because it is the latter – not NNONSTOP -- which will be forecastable.

The second special and problematic characteristic of PT which is not a major feature of goods trade is that the cost of the journey itself -- eg, the airfare – will generally be a substantial (often the major) component of passengers' total travel costs, and this cost will be increasing in distance, because of the variable costs of supplying travel services (fuel and labour, in particular.) Thus, we can expect the FARENZ variable in regression 1 to be interfering with the coefficients estimated for DISTANCE. The latter variable is of course invariant, and so forecasting issues are not directly raised here, but for the sake of a well-specified model it is desirable to sort the direct and indirect distance effects on PAX.

We deal with these problems as follows. For each of NNONSTOP and FARENZ we regress the variable, just as it enters regression model 1, on a suitable subset of the other regressors, with these specified *exactly* as they appear in the PAX model. The residuals from these subset regressions will therefore be generated by exogenous factors uncorrelated with the other regressors, so that the full effect (direct and indirect) of these regressors can be estimated by re-running the PAX model with the residuals from the NNONSTOP and FARENZ regression models used instead of their actual values.

So, look at regression 4 on Table 2. The Log of FARENZ is regressed on DATE, DISTANCE and DISTANCE-squared, SWITCHBACK, and NNONSTOP, all of which are significant explanatory variables for airfares. Note that, if we were primarily interested in modelling airfares, we could probably do better than

Regression 4: most airfare models (eg Gillen and Hazledine, 2015) enter distance in log form, to yield a constant distance-price elasticity, and might also include other cost shifters, such as jet fuel prices and airline-specific cost efficiency. However, our purpose here is just to isolate those determinants of airfares which appear elsewhere in the PAX model.

Table 1: Descriptive Statistics

	minimum	mean	maximum
PAX	0.3	4444	284360
DATE	1	12.4	24
WPOPLARGE	5118	257809	1462110
WPOPSMALL	1261	37505	479490
GDPPOPD	34602	47700	80297
GDPPOPO	34825	48018	80297
DISTANCE	51	523	1338
VISITORS	469787	668848	937673
SWITCHBACK	0	0.32	1
FERRY	0	0.49	1
NNONSTOP	0	0.89	23
FARENZ	76	281	1020

Table 2: Regression models (OLS with Eviews 8.0; number of observations = 10940								
	1		2		3		4	
	dependent = LOG(PAX)				dependent = NNONSTOP		dependent=LOG(FARENZ/CPI)	
	coefficient	t-statistic	coefficient	t-statistic	coefficient	t-statistic	coefficient	t-statistic
C	-22.862	-15.43	-30.728	-54.94	-44.101	-26.05	-1.949	-271.43
DATE	0.017	20.13	-0.004	-4.03			0.011	54.75
LOG(WPOPLARGE)	0.576	77.90	0.846	125.03	1.031	52.95		
LOG(WPOPSMALL)	0.472	68.88	0.671	104.41	0.763	38.79		
LOG(GDPPOPD*GDPPOPO)	0.626	23.51	0.936	35.40	1.186	14.89		
DISTANCE	0.002472	19.89	0.00096	8.29	0.000726	2.22	0.001	40.15
DISTANCE-squared	-1.73E-06	-19.49	-1.63E-06	-18.74	-2.15E-06	-8.33	0.000	-20.31
LOG(VISITORS)*ZQN	0.069	33.36	0.092	45.07	0.089	14.34		
LOG(VISITORS)*ROT	0.021	10.41	0.021	10.44				
LOG(VISITORS)*KKE	0.031	14.44	0.058	27.02	0.103	16.23		
LOG(VISITORS)*TUO	0.020	8.89	0.020	8.93				
SWITCHBACK	-0.602	-31.21	-0.737	-38.95			0.072	20.21
FERRY	0.114	6.29	0.114	6.23				
LOG(FARENZ/CPI)	-1.879	-39.85						
LOG(RFARENZ/CPI)			-1.879	-39.67				
NNONSTOP	0.196	56.35					-0.035	-67.67
RNNONSTOP			0.261	83.58				
R-squared	0.8555		0.8555		0.4102		0.6381	
adjusted R-squared	0.8553		0.8553		0.4099		0.6379	

As for NNONSTOP, regression 3 shows that the combination of population, income and distance variables as deployed in the PAX model do indeed have their expected effect on the provision and extent of nonstop air service on routes, as do two of the four "visitor" variables.

Then, the residuals from regressions 3 and 4 are used as regressors, in the PAX model 2.[14] Note that there is absolutely no change in the R-squared of the model, compared with model 1, and no change in the coefficients of regressors not involved with the production of residual variables in regressions 3 and 4. This is because all we are doing here is shuffling around linear combinations of some regressors, in exactly the same format (logs and/or products) as they appeared in the first regression.

Nor does the coefficient on the log of airfares change, but the coefficient on NNONSTOP does alter, because this variable was one of the regressors in the FARENZ regression 4.

[14] For NNONSTOP, which enters linearly, the residual is the actual minus the predicted value. For FARENZ, which enters in log form, the residual is the ratio of actual to predicted value.

Notably, the coefficients on the main forecasting exogenous variables --population and incomes -- have increased, now that we have allowed them to, in effect, claim credit both for the direct effect of population and incomes on the willingness to travel, and the indirect effect through the effect of more demand on the quality of service (provision and frequency of nonstop flights).

Also, the estimated time trend coefficient, which was a worryingly large positive number in Model 1, implying an annual trend growth in passenger numbers of around 6%, is now a small, negative number. The difference is due to the positive trend in real airfares found in Model 4.

It is interesting to discuss the effect on the distance variables. Note first that the quadratic formulation here is to take account of two opposing effects of distance on demand for air travel. Over smaller distances in particular, airlines compete with road (and, less so in NZ, rail) as transport modes. The longer the route distance, the more attractive is air travel , with its faster journey speeds and larger endpoint airport ground fixed time costs. So, this effect is positively related to distance but with a decreasing effect as distances become very large (and air travel thus becomes the dominant mode).

In the other direction, there is the standard gravity model effect of distance as a deterrent to travel because people are less likely to have reason to visit other people the further they are away. Also, longer flights take more time in themselves, which will be a deterrent to travel at all, in particular for people who don't enjoy travel as a time-using activity.

So, the net result is a concave quadratic distance effect on PAX, as indeed is estimated in both regressions 1 and 2. It is of interest to calculate the turning point, where the function becomes downward-sloping. In regression 1, the turning point comes at 714kms route distance, whereas the more reliable regression 2 implies a turning point at 294kms – say, about the distance of a four hour trip on NZ roads.

Now let us examine the other coefficients of regression 2. The elasticities of passenger numbers with respect to route end-point populations are quite large, and sum to more than one. This implies that an evenly distributed increase in population (ie, of both cities) of x% will generate a larger than x% increase in passenger air traffic between the cities or towns. The reasons for this are not obviously clear, which may not matter for forecasting, but which may be a fruitful urban economics topic for further investigation.

The exponent on the product of end-point regional per capita GDP, if interpreted as the income elasticity[15], implies that, for New Zealanders, air travel is a luxury good –

[15] For business travel, interpretation of the GDP exponent as an income elasticity is probably not appropriate

an evenly spread increase of y% in per capita incomes will have nearly double the effect on passenger numbers.

Numbers of visitors arriving in NZ each quarter was not a significant regressor by itself -- it seems that travel on the great majority of domestic routes is not predominantly or even largely tourist travel. However the VISITORS variable shows up significantly when interacted with dummies denoting that at least one endpoint of the route was Queenstown (ZQN), Rotorua (ROT), Kerikeri (Bay of Islands -- KKE), or Taupo (TUO). Not surprisingly, Queenstown appears to attract the largest share of visitors.

The prospect of a lengthy switchback itinerary is quite discouraging of air travel. The need for a surface travel itinerary to include a voyage in the Cook Strait Ferry is moderately encouraging for the air travel alternative.

The highly significant coefficient of 0.261 on the residual from the NNONSTOP model implies that each exogenous increase by one in the number of nonstop daily flights offered on a route increases travel demand by about 30% -- a sizeable response to improvements in service.[16]

Finally, the estimated price elasticity of demand for air travel in New Zealand is quite large – nearly -1.9 – perhaps reflecting the fact that road transport is a very popular alternative to air travel on all but the lengthiest journeys.

5. Forecasting Domestic Inter-regional Passenger Numbers

We are now ready to move to forecasting, which is the desired end of this exercise. We proceed in the following steps:

1. Use June 30, 2015 data and the PAX regression model 2 in Table 2 to predict quarterly PAX (ie at 2015Q2). Instead of the Census-based population numbers used in the estimation, use the – slightly different SNZ regional numbers that will be used for forecasting. These numbers, for 2015 and beyond, are kept on sheet PAX FORECASTS That is, the end-point WPOP numbers for each city as used in the regressions are adjusted so that, for each region, the sum for one or two cities/towns in the region adds up to the SNZ June 30, 2015 total for that region. These numbers are worked out on sheet EXOGENOUS FORECASTS.

[16] The true relationship may not be linear: in particular, going from no direct service to one daily nonstop flight seems likely to have a bigger effect on demand than, say, going from 3 to 4 daily non-stops. Because our focus here is on getting a good forecasting model (and it is unlikely to be possible to forecast provision of non-stop services years or decades ahead), we do not attempt to explore this possibility.

2. Add up the predicted numbers for PAX on the one to four city-pair routes linking each region.[17] This number will not match the actual inter-regional PAX flows, for three reasons:

   (1) The population numbers plugged in to the city-pair estimated model are not identical to the numbers used in estimation
   (2) City-pair routes with fewer than 50 quarterly passengers have not been explicitly modelled
   (3) The regression model is not perfect – there will be modelling error for each city-pair

3. Adjust the intercept term for each of the one to four city-pair models by the same proportional amount such that the sum of the predicted values is now correct – it equals the actual total number of passengers between the two regions. The adjusted intercept terms for each city-pair model will be retained henceforth for the future forecasts. Both steps 2 and 3 are carried out on sheet STACKED 2015

4. From sheet EXOGENOUS FORECASTS the spreadsheet will extract the numbers needed for each year (2018….2043) for

   - Population (weighted for rural/urban split)
   - Inbound visitors
   - Regional GDP per capita (GDPPOP), assumed to grow equally for all regions at the Treasury forecast rate of increase of national real GDP

5. For each year to be forecast, the spreadsheet will plug in the exogenous forecast values on the appropriate sheet (STACKED 2018, etc). Forecast PAX numbers will be automatically generated and copied onto sheet PAX FORECASTS. This worksheet is the "bottom line" of the project, for domestic passenger forecasts. The output is attached to this Report.

Note that it is very simple for any user of this forecasting model to try out different values for population, income and visitor numbers forecasts. These need simply to be entered on worksheet EXOGENOUS FORECASTS, and the spreadsheets will automatically generate the appropriate passenger number forecasts.

Other exogenous events can also be dealt with, albeit not so simply. For example, a surprising occurrence over 2015/16 has been the expansion of Jetstar's activities in New Zealand to include service to four NZ regional cities (Nelson, Napier/Hastings, New Plymouth, Palmerston North).

[17] If both regions have only one airport, there will be just one inter-regional route. If one region has two airports there will be two interregional routes. If both regions have two airports, there will be four interregional routes.

This is surprising because it has never before been shown that NZ regional routes were other than a natural monopoly for Air New Zealand. In any case, the immediate effect appears to have been an increase in the number of nonstop flights available on the served routes (because Air NZ appears to have not fully accommodated Jetstar by reducing its own flight frequency on a one-for-one basis), and lower airfares. An analyst who was confident that the new competition will persist at least through to 2018 could model this in the forecasts by going to the “STACKED 2018” worksheet, and replacing in columns Y and AG the zero and one values to which residual numbers of nonstop flights and residual price ratios have been normalised by whatever the analyst believes is appropriate to the new situation.

For example, at time of writing this Report, it seems that the total number of daily flights from Palmerston North to Auckland has increased from eight (Air NZ) flights pre-entry to ten (7 Air NZ; 3 Jetstar) post-entry. If this seems likely to persist , then the zero value in row 7 of the worksheet for column Y would be replaced by the number 2. And then the unit value in column AG for price might be changed to, say, 0.8, if the analyst believed that the additional competition was reducing airfares on the route by 20%.

6. **International Outbound Passenger Forecasts**

We also need to model and forecast outbound air travellers, at the minimally disaggregated level of travellers destined, at least in the first instance, to Australia, and those travelling to the rest of the world. These two flows are approximately similar in size: over the 2009-15 period, on average 575,000 people departed NZ for Australia, and 522,000 for other countries, each quarter. All NZ data here are at the regional level.

Those leaving are themselves in one of two categories: NZ residents and short-term visitors to this country. These numbers are, overall, not far apart, with quarterly overseas visitors around 20% more numerous than numbers of Kiwis departing the country.[18]

So, the outbound totals to each region (PAXOZ and PAXROW) each comprise two quite different types of traveller, and the econometric model must accommodate these, in so far as the data allow.

Nearly all overseas air travel is carried in or out of five airports: Auckland, Wellington, Christchurch, Queenstown, and Dunedin, and we can expect that probably

[18] NZ residents departing could be subdivided into short-term and long-term (emigrant) departures. However, the latter make up around 2% of departures or less, typically. See http://statistics.govt.nz/browse_for_stats/population/Migration/IntTravelAndMigration_HOTPFeb16.aspx

quite large numbers of overseas visitors will in our O/D data be recorded as leaving from one of these five airports, either because such was their only destination on their NZ trip (eg skiers to Queenstown), or because their last stopover on a multi-destination trip was, say, Auckland, or because, although not stopping over (ie for a night or more) they found their way to the outward port by surface transport (rental car, bus).

On the reasonable assumption that most overseas visitor stays are much shorter than three months in duration, we will use quarterly inward visitor numbers, multiplied by dummies for the four regions (Auckland, Wellington, Canterbury, Otago) to pick overseas-resident outward passengers of this type.

However, the data do not allow us to distinguish between residents and non-residents flying from a regional city or town to a gateway airport and thence out to the world – for example, people flying from Christchurch to Los Angeles via Auckland directly.

Outbound resident travel we would expect to be affected by the usual demand shifters of price and income, but our data do not enable us to identify these effects: we are unable to estimate sensible or stable coefficients on airfares or on regional per capita incomes. This seems likely to be due to a number of factors:

- The airfares are averaged from fares to many different destinations (especially, of course, fares to the rest of the world), the mix of which will differ across NZ regions
- There may be noise in the data from mixing of fares paid by domestic residents and foreigners
- Income effects will be blurred by the regional-resident travellers finding their way to a gateway airport by other than scheduled domestic service, and so getting mixed in with residents of the gateway airports
- The fact that we have here a cross section of just 15 regions (compared with the around 250 O/D pairs in the domestic passenger travel database) is a likely cause of instability and lack of significance for estimated coefficients.

We also find that neither seasonal nor trend effects are discernible in the 2009-15 data. Nor did exchange rates have any explanatory power. The only other successful explanatory variable is – not surprisingly – regional population.

Table 3: OLS Regression models (No. obs'ns = 360)				
	dependent = LOG(PAXOZ)		dependent = LOG(PAXROW)	
C	-2.381	-4.76	-3.254	-4.41
LOG(POP)	0.871	20.34	0.915	14.49
LOG(VISITORS)*AKL	0.188	16.21	0.221	12.95
LOG(VISITORS)*CHC	0.188	18.83	0.171	11.61
LOG(VISITORS)WLG	0.162	16.67	0.132	9.23
LOG(VISITORS)*OTAGO	0.173	19.72	0.122	9.44
R-squared	0.9176		0.8387	
Adjusted R-squared	0.9165		0.8364	

The stripped-down regression models are shown on Table 3. The elasticities of outward travel with respect to population are similar, and quite close to 1, meaning that there is little difference across more and less highly populated regions in the propensity of individual New Zealanders to take trips abroad.

The hypothesis that we should be able to track some overseas visitors when they leave the country is well supported empirically: in all four regions with gateway airports, the coefficient on visitor numbers is quite large and strongly significant.

The relative simplicity of the econometric models creates issues for the forecasting exercise. We can and should use the same exogenous projections for regional population and overseas visitor numbers that are used for the domestic PAX forecasts. We can also use the same per capita income forecasts, but we lack an estimated elasticity for this, and the extant literature is not much help. Assuming that most foreign travel by NZ residents is leisure travel, it would seem reasonable to assume that this is a luxury good, perhaps with an income elasticity around 1.5. If so, then we should use a number about half of 1.5 in the forecasts, given that only about one half of outbound passengers are NZ residents. An income elasticity of 0.75 will be used in the forecasts.

The empirical literature is more forthcoming with reasonable estimates for the own-price elasticity of demand, for which Mumbower *et al*'s (2014) survey and original results suggest that a number around -1.5 could be appropriate. However, it may be problematic to assume that foreign and local travellers pay the same price for international travel, and in any case I have no means of forecasting this price. So, and as for the domestic forecasts, I have made allowance for the user of the forecasting spreadsheet to insert changes in price levels when they feel these to be appropriate (and, of course, to alter the assumed price elasticity as they wish).

The modelling results suggest that a lot of New Zealand travellers must find their way to a gateway airport by means other than internal air travel (which presumably would be picked up in the O/D data). Look, for example at the tiny numbers of Waikato residents who supposedly travel overseas. It seems very likely that large numbers of Hamiltonians, etc, drive or bus to Auckland airport to catch their flight.

The “bottom line” is that the Sabre data on passenger flows and average airfares, from which highly successful regression and forecasting models can be developed for point-to-point domestic air travel, are not able to support a fully specified model of outbound travel, at least at the high level of destination aggregation (Australia/RoW) limited to here.

7. Conclusion

This Report for the Ministry of Transport, Wellington, has used excellent quarterly proprietary data on actual passenger numbers and average airfares, and other data, over the 2009-2015 period, to successfully estimate an econometric model explaining differences and changes in passenger numbers at the city-pair route level. Aggregated to the regional level, the econometric model has been combined with forecasts of important exogenous variables (regional populations, regional incomes, aggregate overseas visitor numbers) to generate forecasts or projections of inter-regional air travel at five-yearly intervals from census year 2018 through to census year 2043.

The procedures used to generate the passenger forecasts have been automated, such that it is reasonably simple for Ministry of Transport officials to update forecasts or experiment with ranges for exogenous variables.

References

Bhadra, D. and J. Kee (2008), 'Structure and dynamics of the core US air travel markets: a basis empirical analysis of domestic passenger demand', *Journal of Air Transport Management*, 14(1), January, 27-39.

Chi, Junwook, and Jungho Baek (2012), 'A Dynamic Demand Analysis of the United States Air Passenger Service', *Transportation Research Part E*, 48 (4), 755-761.

Dresner, M., J-S C. Lin and R. Windle (1996), 'The Impact of Low-Cost Carriers on Airport and Route Competition', *Journal of Transport Economics and Policy*, September, 309-328.

Frankel, J.A., S-J Wei and E. Stein, (1995), 'APEC and regional trading arrangements in the Pacific', in W. Dobson and F. Flatters (eds), *Pacific Trade and Investment: Options for the 90s', John Deutsch Institute*, Kingston Ontario.

Gerardi, K.S. and A.H. Shapiro (2009), 'Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry', *Journal of Political Economy*, 117, 1, 1-37.

Gillen, David and Tim Hazledine (2012), 'The New Pricing in North American Air Travel Markets: Implications for Competition and Antitrust', chapter 3 in James Peoples (ed), *Pricing Behavior and Non-Price Characteristics in the Airline Industry*, Volume 3 of *Advances in Airline Economics*, Emerald (UK and elsewhere).

Gillen, David and Tim Hazledine (2015), 'The Economics and Geography of regional airline services in six countries', *Journal of Transport Geography* 46 (June), 129-136.

Gillen, David and Tim Hazledine (2016), "Pricing of Regional Airline Services in Australia and New Zealand, 2011-15", *Economic Papers*, forthcoming

Grossman, G.M. (1998). 'Comment on Alan V. Deardoff 'Determinants of bilateral trade: does gravity work in a neoclassical world', in Frankel, J.A. (ed), *The Regionalization of the World Economy*, University of Chicago Press, Chicago.

Harvey, D. (1951), 'Airline passenger traffic patterns within the United States', *Journal of Air Law and Commerce*, 18, 157-165.

Hazledine, Tim (2009), 'Border effects for domestic and international Canadian passenger air travel', *Journal of Air Transport Management*, 15, 7-13.

Hazledine, Tim (2011), "Price Discrimination in Australasian Air Travel Markets", *New Zealand Economic Papers*, 45(3), 311-324.

Hazledine, Tim (2014), "Searching for the Sweet Spot: the determinants of direct services from Canadian airports", *Journal of Air Transport Management*, 41, October, 50-55.

Hazledine, T. and S. Lipanovic (2004), 'Australia-New Zealand Border Effects and Trans-Tasman Integration', Chapter 9 in J. Poot (editor), On the Edge of the Global Economy: Implications of economic geography for small and medium-size economies at peripheral locations, Cheltenham UK, Edward Elgar, 2004, pp 100-118

Helliwell, J. F. and Genevieve Verdier (2001), `Measuring internal trade distances: a new method applied to estimate provincial border effects in Canada', Canadian Journal of Economics, 34(4), November, 1024-1041.

Hsaio, Chieh-Yu and Mark Hansen (2011), 'A Passenger Demand Model for Air Transportation in a Hub-and-Spoke Network', *Transportation Research Part E,* 47 (6), 1112-25.

Klodt, H. (2004), 'Border effects in passenger air traffic', *Kyklos*, 57, 519-532.

Matsumoto, H. (2007), 'International air network structures and air traffic density of world cities', *Transportation Research Part E*, 43, 269-282.

McCallum, J. (1995), 'National borders matter: Canada-US regional trade patterns', *American Economic Review*, 85, 615-623.

Mumbower, Stacey, Laurie Garrow and Matthew Higgins (2014), 'Estimating flight-level price elasticities using online airline data: a first step towards integrating pricing, demand, and revenue optimization', *Transportation Research Part A,* 66, 196-212.

Reiss, P. C. and P. T. Spiller (1989), 'Competition and Entry in Small Airline Markets', *Journal of Law and Economics*, 32, 2, Part 2, October, pp S179-S202.

Wei, Wenbin and Mark Hansen (2006), 'An aggregate demand model for air passenger traffic in the hub-and-spoke network', *Transportation Research Part A*, 40, 841-51.