Pivoting in Travel Demand Models

Andrew Daly[1], James Fox[1], Bhanu Patruni[1], Frank Milthorpe[2]

[1]RAND Europe, Westbrook Centre, Milton Road, Cambridge, CB4 1YG, United Kingdom

[2]Bureau of Transport Statistics, Transport for NSW, 18 Lee Street, Chippendale NSW 2008

Email for correspondence: Frank.Milthorpe@transport.nsw.gov.au

Abstract

In forecasting travel demand it is quite common to base future-year forecasts on an accurately known pattern of base-year observed flows. The "accurate" base-year flows generally do not come directly from strategic travel models. By focussing the modelling effort on predicting changes it is possible to make significant reductions in the expected forecast error. Following Manheim (1979), the process of taking a fixed base point and making forecasts relative to that is called pivoting.

In much travel demand modelling practice, the key issue is to derive forecast matrices of trips, between origin zones and destination zones, which can then be assigned to highway or public transport networks. In this context, the role of pivoting is to exploit observed matrices to improve the accuracy of forecast matrices.

The methods used for pivoting are diverse, sometimes giving substantially different results for the same inputs and little has been published to help analysts choose an appropriate method for their particular study. An important insight of recent work is that pivot methods need to be adapted to the context in which they are used. In particular, the nature of the information available about the base and its relationship to the structure of the forecasting model can lead to choices of different pivoting methods.

This paper reviews the limited literature that exists on pivoting, and assesses the strengths and weaknesses of different approaches. We will set out the basic '8 case' method and apply it to several different forecasting situations. We will also set out criteria by which different pivoting methods can be judged and recommend improvements which may be useful in different contexts.

1. Introduction

In forecasting travel demand it is quite common to base future-year forecasts on an accurately known pattern of base-year observed flows; this approach is even required practice in the UK. By focussing the modelling effort on predicting changes it is possible to make significant reductions in the expected forecast error. Following Manheim (1979), the process of taking a fixed base point and making forecasts relative to that is called pivoting.

In much travel demand modelling practice, the key issue is to derive forecast matrices of trips, between origin zones and destination zones, which can then be assigned to highway or public transport networks. In this context, the role of pivoting is to exploit observed matrices to improve the accuracy of forecast matrices.

The methods used for pivoting are diverse, sometimes giving substantially different results for the same inputs and little has been published to help analysts choose an appropriate method for their particular study. The Daly *et al.* paper (2005), largely by the present authors, gave a review of some of the literature and set out the methods we recommended at that time. However, further literature has come to our attention and our own methods have

improved through recent experience. In the following section of the paper we review the literature of which we are aware, though it has to be said that not all of the relevant material is in the public domain.

An important insight of recent work is that pivot methods need to be adapted to the context in which they are used. In particular, the nature of the information available about the base and its relationship to the structure of the forecasting model can lead to choices of different pivoting methods. In Section 3 of the paper, therefore, we set out the basic '8 case' method that we employ and the flexibility that is available within the general specification of that method to deal with different forecasting situations.

Section 4 sets out a set of criteria by which different pivoting methods can be judged. These criteria build on our experience with the 8-case method, but are intended to be relevant to other pivoting techniques that may be employed.

Finally, Section 5 presents a summary and provides some recommendations for studies that use the 8-case pivoting method. Again while the recommendations are based on experience to the 8-case method, they may be relevant for other pivoting approaches.

2 Literature review

The general pivoting principle set out by Manheim (1979) is clear, but practical applications of the concept were slow in starting.

Bates *et al.* (1987) set out a nested incremental logit model, referencing earlier work by Kumar (1980), and show mathematically how this model can be applied to forecast changes in mode choice. The model is developed as a 'marginal' model that predicts changes relative to the base situation as a function of increments in utility (or generalised cost). The paper also demonstrates how the nested incremental approach can be extended to deal with a new mode. A procedure is described for inferring the mode constant of new modes by rating the mode on a series of attributes which can be compared against existing models and their attributes.

Abraham *et al.* (1992) described the development of an incremental four-stage model for London to evaluate major rail schemes. The model was developed from the existing LTS model, a conventional four stage model applied in an absolute fashion. It was important that the model produced results consistent with the LTS model, while at the same time producing results more rapidly than LTS, which had large run times. The incremental formulation allows the development task to be broken into individual stages, although in this case there is no incremental trip generation model, rather forecasts are taken directly from LTS.

Williams and Beardwood (1993) set out a 'residual disutility' approach to incremental transport models. They start by setting out why we would want to pivot, namely that there will be local aspects that influence travel choices for specific zones pairs that will not be fully represented in the variables in the mode split mode, and these will lead to differences between modelled and observed mode split. Rather than ignore these differences completely, it would be better to carry their influence through as a constant element through the modelling process.

In the residual disutility approach, constraint to match the base matrix in base conditions is achieved by adding a value, the residual disutility adjustment, to the base travel disutility. Once calculated for the base year application, matrices of residual disutility can be carried forward to apply in future years. Williams and Beardwood suggest that the advantages of this approach relative to the incremental approach are that it may be easier to introduce to existing transport modelling packages, it is more efficient with data storage, as only one matrix of residual disutilities is required rather than base trips and base disutilities, and that

the residual disutilities can be analysed for relationships with other potential explanatory variables as a means for further model development. The paper demonstrates how the residual disutility approach can be applied to mode choice, to both singly and doubly constrained trip distribution models, as well as to a more complex application for the APRIL strategic transport model for London where an iterative residual disutility approach was required.

Ortúzar and Willumsen (2011) discuss incremental modelling in a chapter on simplified demand models in their widely used Modelling Transport text, giving the following formula, which can also be found in Manheim (1979):

$$p_k^{'} = \frac{p_k^0 \exp(V_k - V_k^0)}{\sum_j p_j^0 \exp(V_j - V_k^0)} \tag{2.1}$$

where: p'_k is the proportion of trips using alternative k;

p^0_k is the original proportion of trips by alternative k; and

$(V_k - V^0_k)$ is the change in utility of using mode k.

Ortúzar and Willumsen claim that such incremental forms are not difficult to develop or implement, and have the advantage of preserving the base matrices in application. For simple logit models, this approach is equivalent to the approach set out in Bates et $al.$, and adopted by Abraham et $al.$, above.

Ortúzar and Willumsen also discuss absolute models applied incrementally, an approach they (incorrectly) state to be less rigorous than incremental modelling. For absolute models applied incrementally they present both factor and additive forms (the meaning of these two terms is discussed further below) and note that the factor form is equivalent to a matrix of K-factors. They note some of the issues with the approach, such as base-year zeros, and the potential for additive forms to forecast negative trips.

Daly et $al.$ (2005) presented a brief summary of current UK pivoting practice at that time. The approaches used for pivoting in a number of the UK multi-modal studies undertaken around the turn of the century were summarised, as was the UK Department for Transport's Variable Demand Modelling Advice (VaDMA).

UK transport planning advice is now best described by the UK Department for Transport's web-based guidance, WebTAG. WebTAG discusses pivoting in Section 1.5 of Unit 3.10.3, Variable Demand Modelling – Key Processes, which distinguishes three methods of applying models:

1. absolute models, which directly estimate numbers of trips in each category;
2. absolute models applied incrementally, which use absolute model estimates to apply changes to a base matrix; and
3. pivot-point models, which use cost changes to estimate the changes in the numbers of trips from a base matrix, often described as incremental or marginal, such as the model described in Bates et $al.$ (1987) and Ortúzar and Willumsen (2011) above.

Both Methods 2 and 3 are forms of pivoting in that they use the demand models to predict changes to the base matrices, which are taken as the best estimate of base year travel patterns.

WebTAG Unit 3.10.2 strongly recommends that demand models are developed on a Production-Attraction (PA) rather than Origin-Destination (OD) basis. To apply an incremental pivot-point model (Method 3), the base matrix also needs to be defined on a PA

basis, which can be seen as a restriction of this approach as base matrices may only be available at an OD level. However, when applying an absolute model incrementally (Method 2), the demand model can be developed on a PA basis, and the changes can be applied to base matrices defined on either PA or OD bases by introducing a transposition step to convert from PA to OD format immediately prior to pivoting. An important further advantage of Method 2 is that the model may be more detailed, for example in terms of the travel purpose and person type segmentations, than the base matrix. Aggregation over the model categories can be made prior to pivoting and it is shown in the Appendix that this yields 'natural' forecasts. A final advantage of Method 2 is that 'synthetic' model forecasts are produced, which can be examined to see whether they are reasonable in key respects, e.g. mean trip lengths.

WebTAG Unit 3.10.3 notes that that pivot-point models may face difficulties where too many, or key, cells of the base matrix are zero, but does not give any guidance as to how to address these issues. The Unit also notes that the differences can be applied by using the demand models to define either growth factors to apply to the base matrices (factor pivoting) or differences in trips to apply to the base matrices (additive pivoting). In fact, pivoting techniques may use a combination of these two approaches to make best-estimate forecasts of growth. It may seem that an additive approach avoids problems with zero cells, but there is a compensating new problem that negative cell values may be predicted which require correction; simply setting these to zero may introduce other inconsistencies.

A zero value can arise in the observed matrix either because such trips are impossible (a 'structural' zero), perhaps because no connection exists by the relevant mode, or because trips were possible but not observed, perhaps because of the observation method employed, e.g. sampling. A straightforward application of an incremental method would never give any trips for a base-zero cell, even if there are substantial developments, e.g. on a 'green field' site. Zero values can arise in modelled matrices, either in the base or in the future or both, for example when the relevant network does not offer a path, and these can occur in the presence of an observed zero or of observed trips. Again a straightforward application of an incremental method can be problematic.

Section 3 details how a combination of factor and additive pivoting is used in the 8-case method used by RAND Europe for applying absolute models incrementally. This approach is specifically designed to deal with the issue of zero cell values.

3. The 8-case approach

The context in which we developed this method, described in our 2005 paper, was one of making incremental applications of an absolute model. The method set out from the start to deal with matrices that might contain zeros.

3.1   The Basic Method

The 8-case approach is based on RAND Europe's experience with a number of pivot-point models used in their transport demand forecasting systems. Some of these models have special procedures to adjust the calculation when the growth in a specific cell is considered to be 'extreme' (some of the models identify 'extreme' cells automatically, while some others rely on manual selection). Extreme growth situations can often be identified with the development of greenfield sites. Our preferred approach involves automatic selection of the 'extreme' cases and is therefore reflected in the method set out in this section.

The 8-case method implements factor pivoting as the default, termed 'normal growth', but uses additive pivoting for 'extreme growth' cases and cases where the model gives zero trips

4

in the base case. The method is carried out at matrix cell level. That is, for a specific origin, destination, mode, time of day and purpose, adjustments are made relative to the corresponding cell in a base matrix.

For normal growth cases, the approach is to apply the ratio of model outputs for base and forecast situations as a growth factor to the base matrix, i.e. in a given cell the predicted number of trips P is given by

$$P = B \cdot \frac{S_f}{S_b} \tag{3.1}$$

where: S_f gives the modelled, i.e. 'synthetic' trips for a future year;
S_b gives the synthetic trips for the base year;
B is the observed (base) matrix.

However, two considerations mean that it is not always possible to apply this calculation as simply as stated.

First, any combination of the three components on the right hand side of this equation may be zero (or very small) making the calculation impossible or meaningless. Eight possible cases arise (combinations of zero values) and these are dealt with separately.

Second, particularly when there is a land-use change affecting a currently undeveloped zone, the change may be quite extreme and strict application of the formula above can lead to an 'explosion' in the number of trips. In these cases it is better to pivot by applying the difference method, i.e. adding $(S_f - S_b)$ to the base matrix, rather than using a factor as shown above.

The eight possible cases and the recommended treatments taken from Daly *et al.* (2005) are set out in Table 1.

Table 1:   The 8 Cases

Base	Synthetic Base	Synthetic Future	Predicted		Cell Type
(B)	(S_b)	(S_f)		(P)	
0	0	0		0	1
0	0	>0		S_f	2
0	>0	0		0	3
0	>0	>0	Normal growth	0	4
			Extreme growth	$S_f - X_1$	
>0	0	0		B	5
>0	0	>0		$B + S_f$	6
>0	>0	0		0	7
>0	>0	>0	Normal growth	$B \cdot S_f / S_b$	8
			Extreme growth	$B \cdot X_2 / S_b + (S_f - X_2)$	

In cases (4) and (8) the standard (factor) function is used initially, for values of S_f up to the limit when S_f is X_1 (case 4) or X_2 (case 8), and from that point an absolute growth is applied. In case (4) the starting point for absolute growth is 0, in case (8) it is $B \cdot X_2 / S_b$.

To complete the specification of the calculation it is necessary to specify the X variables and to define when a cell is considered to be zero (our experience has led us to use a test value of 10^{-3}).

The definitions for X_1 and X_2 defined by parameters k_1 and k_2 were given by Daly *et al.* (2005) as:

$$X_1 = k_2.S_b \qquad\qquad (3.2)$$
$$X_2 = k_1.S_b + k_2.S_b.\max(S_b /B, k_1/k_2) \qquad\qquad (3.3)$$

Common values for the parameters k_1 and k_2 are k_1=0.5, k_2=5. Note that the specification given here corrects that paper in clarifying the intention that $X_2 \to \infty$ as $B \to 0$ but X_1 does not depend on B.

As is described below, the specification we currently recommend is as above but simplifying (3.3) to give $X_2 = X_1$ as in (3.2).

A variant of the 8-case method is used in The Netherlands by NEA (NEA, 2004a, 2004b) in applications for freight matrices. In this work, the switching points between normal and extreme growth formulae are determined by the distribution of growth factors over cells for each specific policy; cells with growth factors in the 'tails' of the distribution (both high and low) are assigned to additive pivots. Moreover, base matrices are 'seeded' with small values to avoid the problems of zeros. The use of the policy-specific distribution seems unwelcome and we have not adopted it for our own work, nor have we adopted the seeding approach which runs the risk of arbitrariness, as the results depend on the seed value used; also the use of additive pivoting for small growth is unnecessary and reduces accuracy. However, the different context of freight applications may affect the appropriate choice of method. It is also claimed that this approach avoids the arbitrariness of fixing X_1 and X_2, and identifies 'extreme' cells more reasonably, though defining the tails is also arbitrary.

Another pivoting technique that has been applied in the UK by URS Scott Wilson is a GEH approach. GEH is widely used in transport planning to measure differences between traffic flows. However, it can also be used for pivoting. In the GEH method, a GEH measure is calculated in the base for each OD pair:

$$GEH = \sqrt{\frac{(S_b - B)^2}{0.5(B + S_b)}} \qquad\qquad (3.4)$$

Note that the matrices should be scaled in this and the following formulae to match the number of actual observations on which B is based. The GEH formula is useful in itself in indicating a degree of match between the matrices.

Then for pivoting, the value of the predicted matrix P is obtained by varying P in order to achieve the same GEH measure. Thus for each OD pair:

$$\sqrt{\frac{(S_b - B)^2}{0.5(B + S_b)}} = \sqrt{\frac{(S_f - P)^2}{0.5(P + S_f)}} \qquad\qquad (3.5)$$

It can be seen that the GEH method also applies a combination of factor and difference pivoting, because

$$GEH = \sqrt{(S_b - B)\frac{(S_b - B)}{0.5(B + S_b)}}$$
(3.6)

The first term in the GEH calculation can be seen as deriving from a difference pivot, while the second is a relative pivot, so that the GEH represents a geometric mean of the two.

In a study recently undertaken on behalf of the UK Department for Transport to develop a model of demand for long-distance travel (Rohr *et al.*, 2010), both the 8-case and GEH measures were tested. The predictions from the two approaches were compared for two simple test examples. The conclusion from these tests was that the 8-case method performed better, in particular lower and more plausible growth was predicted in a test case where the observed base matrix was significantly smaller in magnitude than the base prediction of the demand model. Based on the judgment of the study team the 8-case method was adopted as the pivoting approach for that study. This finding, however, may depend on specific features of the long-distance study.

An analysis of potential error suggests that difference pivoting should be used when errors in the model are independent of the number of trips, while factor pivoting should be used when the errors are proportional to the number of trips. In these cases the positive correlation of Sb and Sf brought about by errors in the model itself work to reduce the error in the pivoted result. In general, errors in the logit-type models used for forecasting travel demand will be proportional to the numbers of trips, giving a general preference for factor pivoting. The 8-case method therefore uses factor pivoting wherever possible, switching to difference pivots only when that is necessary.

3.2 Experience with the 8-Case Method

Since the 8-case method was set out in Daly *et al.* (2005), RAND Europe have gained further experience on the workings of the method from three studies:
- the PRISM model for the West Midlands region of the UK;
- a model of long-distance demand for the UK Department of Transport;
- a model for the Greater Sydney region in Australia.

In the PRISM model, around 900 zones are used and the base matrices, derived from expanded sample surveys, were considerably sparser than the synthetic matrices, which caused a number of issues for the pivoting process, in particular for public transport modes. The sparsity issue and the corresponding large expansion factors meant that when both B and S_b are non-zero, B was larger on average than Sb because a significant volume of Sb occurs in cells where B is zero. This feature is problematic, because if Sb/B < k1/k2 (= 0.1 for PRISM) the switch point X2 for extreme growth is equivalent to testing whether Sf > Sb, so that all positive growth is classified as extreme. Following analysis of the matrices, a simple modification to the extreme growth formula was made to specify X2 as:

$$X_2 = k_2.S_b$$
(3.7)

Note from Equation (3.2) that this means $X_1=X_2$. This formulation overcame some of the issues with pivoting for the PRISM case, and means that the trigger point for extreme growth is no longer dependent on the value for the base matrix B. This makes the formulation better able to deal with sparse base matrices, as the switch point for extreme growth is no longer dependent on the ratio S_b/B. Base matrix sparsity typically varies between modes, for

example matrices for car are often less sparse than those for public transport modes. Using a formulation for X_2 that is independent of sparsity ensures the approach works reliably for all modes.

Further advantages of the formula (3.7) is that it is considerably simpler than (3.3) and that the equality with (3.2) eliminates possible anomalies for small values of B, when the case (4) pivot could be very different from the case (8) pivot when one or both were classed as extreme.

In both the PRISM and UK long-distance models some form of normalisation has been applied. Equation (3.2) ensures that at a cell level, synthetic growth in trips is reflected in the pivoted number, but because demand is distributed across cells differently between B and S_b there is no guarantee that growth will match when summed across a mode, or across all modes.

In the PRISM model, a row normalisation has been applied so that growth in trips from a given origin zone predicted by the generation models is exactly matched after pivoting. Daly *et al.* (2005) show that, for a simple logit model, this approach is equivalent (for case (8) cells) to incremental modelling. In the UK long-distance model, a 'double normalisation' was applied, which ensured growth in trips predicted by the frequency component of the demand models was exactly matched after pivoting, and that the growth in trips by mode was predicted closely after pivoting. The situation with nested logit models, as in the long-distance model, is considerably more complicated.

RAND Europe have been investigating the performance of the pivoting process for the Sydney Strategic Travel Model (STM, see Fox *et al.*, 2011), which is also a nested logit model, in detail, as part of work to estimate and implement updated and extended demand models for the STM. A number of performance criteria have been specified to assess the performance of the pivoting process, and these are summarised here.

4 Pivoting Criteria

Building on our experience with the 8-case method, this section sets out criteria by which different pivoting methods can be assessed. Based on these criteria, we can then make concrete recommendations for improved pivoting methods.

4.1   Comparison of synthetic and predicted growth

It is useful to compare the proportional growth predicted by the demand model, the *synthetic* growth, to the growth predicted after pivoting, termed the *predicted* growth:

$$\text{Synthetic growth:} \qquad \frac{(S_f - S_b)}{S_b} \qquad\qquad (4.1)$$

$$\text{Predicted growth:} \qquad \frac{(P - B)}{B} \qquad\qquad (4.2)$$

Growth can be compared across all modes, for specific modes, or for particular groups of zones. Often users of model outputs will look at growth in trips by mode, and so this is a useful level of analysis.

Synthetic and predicted growth should be compared at different levels. In particular, a given pivoting method might ensure that they match exactly for a given cell in a matrix, but that is no guarantee that they match when demand is summed over matrices, such as total demand for a given mode.

A particular issue that can arise when comparing synthetic and predicted growth is that of sign change, where synthetic growth and predicted growth have different signs, a result that is difficult to explain to users of model outputs. The simple example in Table 2 for a single mode with two destination zones D1 and D2 illustrates how sign change can occur.

Table 2:   Sign Change Example

	D1	D2	Total
S_b	10	10	20
S_f	9	12	21
$(S_f - S_b)/\ S_b$	-10%	+20%	+5%
B	15	5	20
P	13.5	6	19.5
(P - B) / B	-10%	+20%	-2.5%

The difference between synthetic and predicted growth can be reduced by using a *normalisation* so that synthetic and predicted growth match exactly at the aggregate level of interest. Normalisation can be defined at different levels, for example growth in total demand for a mode, growth in demand for a mode and origin combination, growth in total demand for a given origin, and so on. In general, normalisation should always be applied when pivoting.

However, simple normalisations are not totally effective in nested models, in that normalisation at one level can adversely affect the normalisation at a higher or lower level in the model. In the Appendix, it is shown that a top-down normalisation can be defined that gives a model that ensures matches between synthetic and predicted outputs at all levels in the model. Moreover, this model is consistent with utility theory, as shown in the Appendix, so that it can be guaranteed that sign changes (as in Table 2) and other drastic divergences between synthetic and predicted forecasts cannot occur. The issue is that top-down normalisation can be difficult to implement in some model implementations.

4.2   Sparsity

As discussed for the PRISM example, matrix sparsity can cause issues for pivoting as it results in different distributions of demand over cells between the base matrix B and the synthetic base S_b. A sparsity measure can be specified as:

$$Sparsity\ Index = \frac{number\ of\ cells[S_b > 0]}{number\ of\ cells[B > 0]} \qquad (4.3)$$

An index close to one indicates a similar distribution of positive values of S_b and B over cells, a situation where the pivoting procedure would be expected to work well. High indices indicate B to be sparser than S_b. Often the base matrix is derived by expanding count data, and so when the number of zones is large, and/or the sampling rate is low, high sparsity indices are expected, because S_b is derived by considering the probability of travelling to each available destination and therefore widely distributed over cells, whereas B is only non-zero for those OD pairs where a trip is observed.

To give an idea of how sparsity can affect the correlation of synthetic and forecast results, Table 3 presents test results from the commute model in the Sydney STM for each of the five modes included in the pivoting process, presenting for each mode the sparsity index and the ratio between predicted and synthetic growth. These comparisons are presented for two

different base matrices with different levels of sparsity: the first, Journey to Work (JTW) data which is effectively fully observed, being derived from the 2006 census; the second, expanded Household Travel Survey (HTS) data where the mean expansion factor is approximately 180.

Table 3:   Matrix Sparsity and Growth Comparison

Mode	Synthetic Growth	JTW Base Matrices		HTS Base Matrices	
		Predicted / synthetic growth	Sparsity Index	Predicted / synthetic growth	Sparsity Index
Car Driver	40.5 %	0.70	19.1	0.26	949.3
Car Pass.	12.6 %	0.52	77.4	0.22	3770.6
Rail	86.0 %	0.48	42.5	0.34	2373.5
Ferry	9.4 %	0.26	9.0	0.49	371.4
Bus	74.3 %	0.31	42.2	0.20	1848.9

It can be seen from Table 3 that predicted growth is systematically lower than synthetic growth, and that these differences generally increase as the base matrix sparsity increases. A particular feature of these tests is the large number of model zones, 2690, which means that even the fully observed JTW base matrices are much sparser than the probability-based synthetic matrices.

Aggregating the zones on which matrices are defined, so that the majority of cells would be expected to have non-zero values for both the base matrix and the synthetic base matrix, would reduce sparsity. Tests with the Sydney matrices have demonstrated that this approach improves the performance of the pivoting process by giving a better correspondence between synthetic and pivoted growth. Table 4 illustrates how the correspondence between predicted and synthetic growth improves for the two sets of Sydney matrices illustrated in Table 3. The cells values are the ratios of predicted to synthetic growth. The aggregate pivot is performed using an 80-zone system.

Table 4:   Impact of Zonal Aggregation on Predicted/Synthetic Growth

Mode	JTW Base Matrices		HTS Base Matrices	
	Original pivot (2690 zones)	Aggregate pivot (80 zones)	Original pivot (2690 zones)	Aggregate pivot (80 zones)
Car Driver	0.70	0.98	0.26	0.95
Car Pass.	0.52	0.87	0.22	0.79
Rail	0.48	0.85	0.34	0.58
Ferry	0.26	-0.38	0.49	-1.21
Bus	0.31	0.60	0.20	0.37

The advantage of aggregation may perhaps be better understood if we rewrite equation (3.1)

$$P = B \cdot \frac{S_f}{S_b} = S_f \cdot \frac{B}{S_b} \qquad\qquad (4.4)$$

The fraction B/S_b is calculated at zonal level in the disaggregated pivoting approach. In the aggregated approach, this fraction is calculated over aggregate zones, which obviously gives a more reliable estimate. Effectively, the model output S_f is used to distribute demand over

the detailed zones. The loss of accuracy given by aggregating B is set against the gain in reliability. In choosing aggregate zones, account needs to be taken of the likely model error and the likely sampling error in B. It is not possible to determine objectively an optimal aggregation level, because of the need to maintain consistency across the study area, travel purposes and modes, but an informed judgement on the level can be made.

With the exception of the ferry mode, which is affected by the assignment procedure rather than the choice model, the correspondence between synthetic and pivoted growth is significantly improved by zonal aggregation. Further improvements are obtained when normalisation is applied, but it is desirable to obtain the best possible results before normalisation and that is the objective of the aggregation.

4.3 Distribution of demand over cases

A useful technique for assessing the performance of a pivoting method is to examine the distribution of demand over the different cases used for the possible combinations of zero and non-zero cells in each of B, S_b and S_f. Examination of aggregate results summed over all cases may hide anomalies with how the pivot process is working for particular cases. One issue that may arise is that a high fraction of synthetic demand occurs in cases 2, 3 and 4 normal, where the predicted growth is zero, and so growth in synthetic demand has no impact on the model predictions. Another issue is that too high a fraction of growth may be classified as 'extreme' in case 8 (for example the PRISM experience cited above), and this can result in predicted growth falling significantly below synthetic growth for a particular mode.

The following Sydney example illustrates the distribution of demand over cases for car driver using the HTS base matrices. The sparsity index for these matrices is 949.3, as in Table 3.

Table 5:   Example Distribution over 8 Cases, 2690 zones

Case	B	S_b	S_f	B > 0	$S_b > 0$	$S_f > 0$	P > 0
1	0	0	0	0.0 %	0.0 %	0.0 %	0.0 %
2	0	0	> 0	0.0 %	0.0 %	0.3 %	0.4 %
3	0	> 0	0	0.0 %	0.0 %	0.0 %	0.0 %
4n	0	> 0	> 0	0.0 %	95.8 %	84.5 %	0.0 %
4e	0	> 0	> 0	0.0 %	1.2 %	12.4 %	10.8 %
5	> 0	0	0	0.1 %	0.0 %	0.0 %	0.1 %
6	> 0	0	> 0	0.1 %	0.0 %	0.0 %	0.1 %
7	> 0	> 0	0	0.0 %	0.0 %	0.0 %	0.0 %
8n	> 0	> 0	> 0	28.4 %	1.6 %	1.3 %	23.1 %
8e	> 0	> 0	> 0	71.4 %	1.5 %	1.5 %	65.7 %
Total				100.0 %	100.0 %	100.0 %	100.0 %

Table 5 demonstrates that a high fraction of synthetic demand is occurring in case 4 normal, where the predicted demand is zero.

Table 6 shows how the distribution of demand over the 8 cases changes when the matrices are aggregated from the original 2690 model zones to the 80 zones used for the aggregate pivot. Here the sparsity index is reduced to 4.06.

Table 6:   Example Distribution over 8 Cases, 80 zones

Case	B	S_b	S_f	B > 0	S_b > 0	S_f > 0	P > 0
1	0	0	0	0.0 %	0.0 %	0.0 %	0.0 %
2	0	0	> 0	0.0 %	0.0 %	0.0 %	0.0 %
3	0	> 0	0	0.0 %	0.0 %	0.0 %	0.0 %
4n	0	> 0	> 0	0.0 %	15.0 %	14.9 %	0.0 %
4e	0	> 0	> 0	0.0 %	0.1 %	0.3 %	0.1 %
5	> 0	0	0	0.0 %	0.0 %	0.0 %	0.0 %
6	> 0	0	> 0	0.0 %	0.0 %	0.0 %	0.0 %
7	> 0	> 0	0	0.0 %	0.0 %	0.0 %	0.0 %
8n	> 0	> 0	> 0	94.5 %	83.7 %	81.4 %	91.8 %
8e	> 0	> 0	> 0	5.5 %	1.3 %	3.4 %	8.1 %
Total				100.0 %	100.0 %	100.0 %	100.0 %

Zonal aggregation has significantly reduced the percentage of synthetic demand that occurs in case 4 normal, and shifted that demand to case 8 normal. Moreover, the percentage of base demand in case 8 extreme has also been significantly reduced. As a result the percentage of predicted demand P that occurs in the standard case 8 normal, where the best-quality pivoting is likely to occur, has increased from 23% to 92%.

These issues relate to base matrix sparsity, as with sparse base matrices a high fraction of cells will be classified into the cases where base matrix demand is zero where different pivoting rules are used; base matrix sparsity can also impact on the rules used to transition between normal and extreme growth. The redefinition of the transition points between normal and extreme growth may also help in maximising the amount of demand that is forecast using more reliable factor pivoting.

4.4   Consideration of zero cells and continuity

A final issue to be considered is that of *continuity*. Different rules may be defined for different combinations of matrices with zero values, but what happens if a given matrix increases in value slightly so that it moves from one pivoting rule to the next? Are the outputs from the pivoting process consistent at the transition points between cases, or can sudden jumps in trips occur which could lead to strange results when comparing different policies?

Detailed examination of the 8 cases of Table 1 shows the following.

1. No discontinuities arise for changes in S_f, either between the even and odd-numbered cases or between normal and extreme growth in cases 4 and 8. This is a key feature of the system, as it means that small differences between policies will not lead to large differences in forecasts.

2. When B changes, there are no discontinuities, providing $X_1 = X_2$.

3. When S_b changes, there is potential for discontinuities between some cases with $S_b = 0$ and small positive values of S_b. This of course relates to the presence of S_b in the denominator of the pivot formula.

It does not seem possible to eliminate the discontinuities relating to S_b. However, once the base synthetic forecasts are fixed, they would not normally be changed and these discontinuities will not affect the comparison of policy.

4.5   Conclusion for Sydney application

The Sydney analysis concluded with three recommendations to improve the performance of the pivoting process.

The first was that pivoting should be undertaken at a more aggregate zonal level, which will reduce the sparsity of the base matrices, and has been shown to lead to a better correspondence between synthetic and predicted growth.

The second recommendation was that some form of normalisation should be applied, as in the UK models discussed above, for example to ensure that synthetic and predicted growth match exactly for each mode. The Appendix indicates that a top-down normalisation can be applied that would match exactly at all levels, but this is not necessarily straightforward to implement in specific software environments.

The third recommendation was to revise the definition of the switch-point for extreme growth to use the simplified form given in Equation (3.7), which tests show is better able to work with sparse base matrices. This change also improves the continuity properties of the pivoting process and the consistency of the model with utility theory, improving its performance as indicated in the Appendix.

These recommendations have been implemented for the Sydney STM.

5 Summary and recommendations

Pivoting can improve the accuracy of forecasting in transport demand models as the model is used only to predict changes from a base that is known with better accuracy than is given by the model.

A number of different methods can be used for pivoting. The different methods can be classified into two types. The first are absolute models applied incrementally to estimate changes to a base matrix, such as the 8-case method discussed in Section 3, or the GEH method. The second are models that use cost changes to estimate changes in the number of trips from a base matrix, and are often described as incremental or marginal models. We prefer the absolute method, because

- it allows base matrices to be specified at a lower aggregation level, of purpose, person type or spatially, than the model;
- it allows base matrices specified on an OD basis to be used with a model specified on a PA basis;
- it allows procedures to be developed, such as the 8-case method, to deal with zero values; and
- it produces 'synthetic' matrices which can be examined to see that they are reasonable, e.g. with respect to trip length.

Levels of aggregation, e.g. by purpose can be specified flexibly to determine the best approach in each data context.

Our specific experience is with the 8-case method and our recommendations here are based on experience with that approach. However, they may be useful to consider when testing other pivoting techniques used to apply absolute models incrementally.

An important insight from our experience with the 8-case method is that any pivoting method needs to be adapted to the context in which it is used, and in particular the similarity or otherwise between the base matrices and the synthetic matrices from the demand model. For a given study, analysis of the operation of any pivoting technique is important to check that the method is giving reasonable forecasts. Section 4 provided four criteria that can be used to provide a framework for assessment. Our experience is that pivoting can often go wrong and so using any one approach directly without checking the results carefully is dangerous!

Failure of pivoting is often associated with a lack of match between the base matrix and the synthetic base forecast. Of course, if these were the same, no pivot would be needed, but a large difference between the matrices is likely to be a sign of potential problems. Objective measures of correspondence between these matrices can and should be developed; the GEH measure defined in (3.4) could be the basis for a measure.

For the 8-case approach, our recommendation for models that use detailed zone systems is to pivot at a more aggregate zonal level, and then disaggregate the pivoted demand back down to the detailed zone level. Pivoting at an aggregate level reduces the impact of base matrix sparsity, and gives an improved correspondence between synthetic and predicted growth.

To specify the switch point for normal growth X_2, set out in Equation (3.3), our recommendation based on experience from two recent studies is to use the simplified version for the switch point given in Equation (3.7). This simplified formulation has been found to be more robust for working with sparse base matrices, where demand is concentrated in a much smaller number of cells than in the synthetic model forecasts.

Finally, normalisation should always be applied, so that at the aggregate level of interest the synthetic growth predicted by the demand models is matched exactly after pivoting. Normalisation can be applied at different levels in nested models, such as growth in total demand for a mode, growth in demand for a mode and origin combination, and growth in total demand for a given origin. A judgement needs to be made for a particular study as to the appropriate level to normalise. The Appendix shows that an overall normalisation can be applied in nested models, consistent with utility theory if applied 'top-down'. This may however be complicated to apply in particular circumstances.

An area for research is the combination of pivoting with matrix estimation in which information such as traffic counts is used to improve the estimate of B.

Acknowledgements

References

Abraham, H, N .Shaw and L Willumsen (1992), A Micro-Based Incremental Four Stage Transportation Model for London, Proceedings 20th PTRC Summer Annual Meeting, UMIST, September 1992, England.

Bates, J., D. Ashley and G. Hyman (1987), The Nested Incremental Logit Model: Theory and Application to Modal Choice, 15th PTRC Summer Annual Meeting, University of Bath, England.

Daly, A., J. Fox and J.G. Tuinenga, (2005), Pivot-point Procedures in Practical Travel Demand Forecasting, European Congress of the Regional Science Association, Amsterdam.

Fox, J., Daly, A., Patruni, B. and Milthorpe, F. (2011), Extending the Sydney Strategic Model to represent toll road and park-and-ride choices, ETC.

Kumar, A (1980), Use of Incremental Form of Logit Models in Demand Analysis. Transportation Research Record 775, pp 21-27.

Ortúzar J. de D., L. G. Willumsen (2011), Modelling Transport, Fourth Edition, John Wiley & Sons Ltd, Chichester, UK.

Manheim, M. L. (1979), Fundamentals of Transportation System Analysis. The MIT Press, Cambridge, Mass.

NEA Transportonderzoek en –opleiding (2004a), Groeifactormodule SMILE+ (SMILE+ Growth Factor Module), report.

NEA Transportonderzoek en –opleiding (2004b), Vrachtautomatrix 2000 (2000 Freight Vehicle Matrix), report.

Rohr, C., J. Fox, A. Daly, B. Patruni, S. Patil, F. Tsang (2010), Modelling Long-Distance Travel in the UK, European Transport Conference, Glasgow.

WebTAG Unit 3.10.2, Variable Demand Modelling – Scope of the Model, April 2011.

WebTAG Unit 3.10.3, Variable Demand Modelling – Key Processes, October 2009 (in consultation).

Williams, I., J. Beardwood (1993), A Residual Disutility Based Approach to Incremental Transport Models. PTRC 21st Annual Summer Meeting, Proceedings of Seminar D, PTRC, London.

Appendix

This Appendix shows that it is possible to:

- amend a tree-nested logit model, defined using the RUM1 convention, by the addition of constants to the utility functions, so that the model remains consistent with utility theory;
- because the model remains consistent with utility theory, it is guaranteed to give intuitive results;
- the amended model is entirely equivalent to a model that is pivoted at each level from the top downwards, with normalisation at each stage;
- natural extensions can be made to cover multiple segments, applications of PA models to OD matrices and to cover all 8 cases.

Calculations within a utility maximisation approach

Following Daly (1987) we define a tree logit model using the RUM1 ("non-normalised") specification, which is used, for example, in the STM model. The specification depends on a tree function that gives the ancestor node t_j for each node j in the model.

A tree logit model (RUM1) is defined recursively by the following

$$\log p_{k|t_k} = V_k - \log \sum_{t_h = t_k} \exp V_h \qquad (A1)$$

where V gives the measured utility for each alternative.

For composite alternatives, i.e. those that are the ancestor of some other node, utility is transmitted through the 'logsum'

$$V_j = \theta_j \log \sum_{t_h = j} \exp V_h \qquad (A2)$$

where $0 < \theta_j \leq 1$.

The intention is to set up a pivoted model, with the same form as the original model but with utility functions amended. Terms in this model are indicated in the equations by a superscript, i.e. p^* and V^*, so we get

$$\log p^*_{k|t_k} = V^*_k - \log \sum_{t_h = t_k} \exp V^*_h \qquad (A3)$$

$$V^*_j = \theta_j \log \sum_{t_h = j} \exp V^*_h \qquad (A4)$$

The amendment to the utility function is required to serve two functions

- it must be consistent with the utility transmission (A4) and

- it must give the pivoting that is required at each level of the tree.

It turns out that the amendment that is needed is:

$$V^*_h = V_h + K_h + L_{t_h} \qquad (A5)$$

where $K_h = \log \left(\frac{b_h}{p_h} \right)$;

b_h is the observed fraction of total trips choosing alternative h (note that this is the *marginal* fraction, not conditional on t_h, in parallel to the marginal forecast probability p_h);

$$L_g = \sum_{i \in A_g} K_i (1 - \theta_i) \frac{\Pi_{l \in A_{t_i}} \theta_l}{\Pi_{l \in A_g} \theta_l};$$

A_g is the set of ancestor nodes of g: $\left\{g, t_g, t_{t_g}, \dots i \mid t_i = r\right\}$; note that this includes the node itself but excludes the root r of the tree.

The effect of the fraction in L_g is to divide by the θ parameters for all nodes in the ancestor set from g to i inclusive.

We prove the two requirements for the amendment, consistency and pivoting, in that order.

Because of the way it is defined, L_{t_h} in (A5) involves only t_h and higher nodes, so when we calculate the conditional probability for h that term disappears and the probabilities (using equations A3 and A5) are determined by $(V + K)$ only

$$\log p^*_{h \mid t_h} = (V_h + K_h) - \log \Sigma_{t_k = t_h} \exp(V_k + K_k) \tag{A6}$$

Similarly, when we apply the recursive formula (A4) to get the utility for composite alternative j, L_j is constant and can be taken outside the logsum as a constant to obtain

$$V_j^* = \theta_j \log \Sigma_{t_h = j} \exp V_h^*$$

$$= \theta_j \log \Sigma_{t_h = j} \exp(V_h + K_h + L_j)$$

$$= \theta_j L_j + \theta_j \log \Sigma_{t_h = j} \exp(V_h + K_h) \tag{A7}$$

Moreover, we can calculate the first term in (A7) further, cancelling out the factor θ_j

$$\theta_j L_j = \theta_j \Sigma_{i \in A_j} K_i (1 - \theta_i) \frac{\Pi_{l \in A_{t_i}} \theta_l}{\Pi_{l \in A_j} \theta_l}$$

$$= K_j(1 - \theta_j) + \Sigma_{i \in A_{t_j}} K_i (1 - \theta_i) \frac{\Pi_{l \in t_i} \theta_l}{\Pi_{l \in A_{t_j}} \theta_l}$$

$$= K_j(1 - \theta_j) + L_{t_j} \tag{A8}$$

The second term in (A7) can also be calculated:

$$\theta_j \log \Sigma_{t_h = j} \exp(V_h + K_h) = \theta_j \log \Sigma_{t_h = j} \left(\frac{b_h}{p_h}\right) \exp V_h \tag{A9}$$

Now if $t_h = j$, we can use (A1) and (A2) to calculate

$$p_h = p_j \cdot p_{h \mid j} = p_j \cdot \frac{\exp V_h}{\Sigma_{t_k = j} \exp V_k} = p_j \cdot \frac{\exp V_h}{\exp \frac{V_j}{\theta_j}} \tag{A10}$$

so we can substitute in (A9) to obtain

$$\theta_j \log \Sigma_{t_h = j} \exp(V_h + K_h) = \theta_j \log \Sigma_{t_h = j} \left(\frac{b_h \cdot \exp \frac{V_j}{\theta_j}}{p_j}\right)$$

$$= \theta_j \log \left(\exp \frac{V_j}{\theta_j}\right) + \theta_j \cdot \log \Sigma_{t_h = j} \left(\frac{b_h}{p_j}\right)$$

$$= V_j + \theta_j \log \left(\frac{b_j}{p_j}\right) = V_j + \theta_j K_j \tag{A11}$$

because $\sum_{t_h=j} b_h = b_j$.

Given the results (A8) and (A11) we are now in a position to simplify equation (A7)

$$V_j^* = V_j + K_j + L_{t_j} \qquad (A12)$$

Equation (A12) is a repeat of (A5) at the next higher level, indicating that the change to the model is carried through to the next level and hence to all levels. **That is, we satisfy the first requirement of consistency in the amended model.**

In equation (A6) we noted that the marginal probabilities at each level in the tree are determined by $(V + K)$ only. We can then obtain the ratio of probabilities for two alternatives in the same nest

$$\frac{p_k^*}{p_h^*} = \frac{\exp V_k}{\exp V_h} \cdot \frac{\exp K_k}{\exp K_h} = \frac{p_k}{p_h} \cdot \frac{b_k/p_k}{b_h/p_h} = \frac{b_k}{b_h} \qquad (A13)$$

That is, the formulation of K ensures we obtain the observed proportions at each level in the model. Provided the total number of trips is correct, this implies that we obtain the correct number for each alternative. The utility correction (A5) thus **satisfies both the requirements to obtain a pivoted model**.

Extending to multiple segments

The above formulae apply for any single segment. For pivoting we require to aggregate over segments (person types, purposes) and to apply the correction factors to the aggregate trips. In this case we obtain a more general version of equation (A13) for segment s

$$\frac{p_{js}^*}{p_{hs}^*} = \frac{p_{js}}{p_{hs}} \cdot \frac{b_j/p_j}{b_h/p_h} = \frac{b_{js}}{b_{hs}} \qquad (A14)$$

where $b_{js} = b_j \cdot p_{js}/p_j$, which is the base matrix multiplied by the fraction of flow indicated by the model.

The fraction b_{js} appears to be the best indication we can get of the fraction of base flow that is due to segment s. Moreover, adding up over the segments gets back to the total base matrix, as required.

Dealing with all 8 cases

The formulae above are possible to calculate only for cases 8n and 7 because these are the only places we can calculate K=log(B/S$_b$), $\log\left(\frac{b_h}{p_h}\right)$ in the notation of the Appendix. Suppose for cases 1-6 we set K=0. Then we will get the right answer for cases 1-3, 7 and 8n, as can be seen by detailed consideration of the pivot formulae in Table 1. For the other cases:

- 5 and 6 are forecast consistently with utility because they just add B to S$_f$ and this is like assuming there is some fixed demand – captive, perhaps;
- 8e is forecast consistently with utility if B>S$_b$, because we are then just taking a positive demand X2*(B/S$_b$–1) as being fixed. The recommendation to simplify the formula for X2 = X1 implies that B<S$_b$ does not occur.

Case 4n, which is just set to zero, is also consistent with utility theory but the problem here is that the switch point between 4n and 4e depends on S$_f$ and may therefore differ between forecast scenarios. In this case there is no correct formula, because there is no utility function that will give the results we need, which are linear in S$_f$ but with a negative offset. For this reason there can be no complete guarantee that we get utility-consistent results.

So a proposal would be to make a two-step pivot. First, do the utility-based pivot, as set out above, with K=0 for cases 1-6. This will give us S_f^* which is utility-consistent for cases 1-3 and 5-8. For case 4, we then do a second-stage pivot, using S_f^* in place of S_f, to get the final result. For 5, 6 and 8e we just have to add the constant as above.

Dealing with OD matrices

Some base matrices are defined on an OD basis, rather than on a PA or tour basis. The problem with these matrices is that it is difficult to apply the notion of utility maximisation, because we do not know whether the traveller is choosing the origin or the destination.[1]

The most obvious approach, which seems to be consistent with other procedures discussed here, is to allocate OD matrices to PA in the proportions indicated by the model. This may seem quite onerous. But if we calculate a matrix split in this way, the 'observed' number of tours in one direction would be given by

$$B^1 = B \frac{S_b^1}{S_b^1 + S_b^2}$$

with the superscripts indicating the direction of the tours and B being the base trip matrix, which is assumed to be symmetric. If we then make a case 8n pivot, for tours in both directions, we would obtain:

$$P = B^1 \frac{S_f^1}{S_b^1} + B^2 \frac{S_f^2}{S_b^2} = B \frac{S_f^1 + S_f^2}{S_b^1 + S_b^2}$$

which is exactly what we would obtain if we followed a conventional procedure of calculating the synthetic trips, then making the pivot.

This approach is therefore recommended for the pivoting of OD matrices using a model that represents PA behaviour.

Practical procedures

Two alternative procedures seem to be available.

1. Using the utility formula (A5) directly

To apply these formulae directly we would calculate a matrix of K+L, setting K=0 for cases 1-6, then apply the corrections. Some thought is needed to do this efficiently.

2. Factoring with normalisation

What is shown above is that it is possible to set up a model that matches the base observations exactly and that is consistent with utility. This can be achieved equivalently by factoring or by adding utility. For factoring, we have to work top-down to ensure we have achieved consistency.

If we consider what happens when we change the utilities for forecasting, based on equation (A13)

$$\frac{p_k^1}{p_h^1} = \frac{\exp(V_k + \Delta V_k)}{\exp(V_h + \Delta V_h)} \cdot \frac{\exp K_k}{\exp K_h} = \frac{p_k^*}{p_h^*} \cdot \frac{\exp(\Delta V_k)}{\exp(\Delta V_h)} \qquad \text{(A15)}$$

where p_k^1 is the forecast proportion choosing k and

ΔV_k is the change in utility of alternative k.

---

[1]   There may also be a problem when dealing with detours (non-home-based) because again we do not necessarily know which trip end is chosen.

which is exactly the formulation used in incremental models, e.g. in the WebTAG advice (2009).

This approach is likely to be easier to implement and understand than procedure 1.