Australasian Transport Research Forum 2011 Proceedings
28 - 30 September 2011, Adelaide, Australia
Publication website: http://www.patrec.org/atrf.aspx

Modelling toll roads - where have we gone wrong?

Peter Davidson[1]

[1]TransPosition, Level 5 199 George St, Brisbane, 4000

Email for correspondence: peter@transposition.com.au

Abstract

Transport modelling has made a rare intrusion into the attention of the media and the public due to inaccuracies in recent projections done for toll roads in Australia. Generally the modelling of toll roads in Australia has been fairly poor, with many roads attracting only half of their projected volumes and some roads, such as Brisbane's Clem 7, attracting much less. However this is not a uniquely Australian phenomenon, with similar outcomes around the world.

A number of papers have been written to explore the reasons for this, and some have come to the conclusion that the most likely cause is biased analysis, either inadvertent or deliberate. Genuine modelling uncertainly, it is argued, would lead to as many under-predictions as over-predictions, and would tend to have reducing errors as models improve. However there are some specific modelling problems that could lead to similar patterns of errors. In this paper we review the approaches used in recent Australian toll road projects, examine their underlying theory, and explore the ways in which the models may lead to over predicted demands.

The paper concludes with an examination of alternative approaches to modelling toll roads and their practicality in an Australian context, and some recommendations for improving modelling outcomes.

Disclaimer

The author of this paper is a practicing transport modeller, and has done peripheral work on some of the projects discussed in this paper. Much of the work done on toll roads is closely covered by confidentiality agreements and is treated as commercial-in-confidence. None of the confidential information that the author has had access to has been used in the preparation of this paper – all of the referenced material is publicly available and, where possible, web links are provided. As discussed in the paper, there is limited detailed information publicly available on the modelling approaches used for toll roads and so this paper focuses heavily on those projects that did release information, particularly Brisbane City Council's Hale St Link Project (now the Go-Between-Bridge) Business Case and the Hills M2 Upgrade Environmental Assessment. The arguments presented here should not be taken as criticisms of those particular projects, their proponents or their modellers – in fact they should be commended for making the information available.

The nature of a paper such as this one is that not every argument can be supported by evidence, and some issues are contentious. It almost goes without saying that the unsupported views expressed in this paper are simply the author's opinions and expressed in the interest of improving the state of the modelling profession.

1. Introduction

Transport modelling usually has little prominence in the public's mind and is rarely mentioned on news reports, current affairs programs or newspaper editorials. In the past few years, however, this had changed. A sequence of prominent failures in Australian toll roads have led to an increased scrutiny of the planning process for toll roads, and the models that provide the foundation for predictions.

As well as discussions in the media, there have been a number of papers, both in Australia and internationally, that have sought to demonstrate and explain the prediction errors that have plagued toll road projects around the world (Bain 2009; Flyvbjerg et al. 2005; Goldberg 2006). The National Cooperative Highway Research Program conducted a broad review of estimating toll road demand and revenue, which included an analysis of performance across 26 toll highways throughout the United States (Kriger et al. 2006). There has also been a recent Symposium on Patronage Forecasting organised by the Bureau of Infrastructure, Transport and Regional Economics (BITRE) which included presentations by Robert Bain (RBConsult), Steve Kanowski (GHD) and a review of Traffic Forecasting Performance by BITRE staff lead by Mark Harvey (Bain 2011; Kanowski 2011; Harvey 2011).

The general consensus of these studies is that there is a real problem; (Bain 2009) show that the mean ratio of actual to modelled demand is 77% (i.e. on average toll roads attract 23% less traffic than forecast), with a standard deviation of 26% - this is in contrast to a mean of 96% for un-tolled roads. (Flyvbjerg et al. 2005) show that half of the 183 road projects reviewed had errors of at least +/- 20%, with a quarter having errors greater than +/- 40%. BITRE (reporting Li and Hensher, 2010) show that for Australian toll roads the record is worse – on average traffic volumes have been 45% below the forecast levels. This average has incorporated the standout problems in Sydney's Lane Cove Tunnel (37% lower than predicted), Sydneys' Westlink M7 (50% below predictions), but was prepared before Brisbane's Clem 7 (currently 80% lower than predicted) (Harvey 2011).

There is also a common view that the problem is not just a technical one. (Flyvbjerg 2008) argues that technical problems would tend to lead to symmetrically distributed errors – with as many over-predictions as under-predictions. Furthermore model and data quality should be improving over time, so the level of error should be dropping. But the evidence is that the majority of studies overstate the likely demand on new projects, and that the inaccuracy in forecasts has stayed fairly constant over time. From this he concludes that psychological and political-economic explanations are more relevant than technical ones. He points particularly to optimism bias (which is a form of self-deception) and strategic misrepresentation (which is a politically or economically motivated deception of others).

However we would argue that, while the psychological and political-economic explanations for error are undoubtedly important, we should not be too quick to discount systematic flaws in our modelling processes.

This paper attempts to give one modeller's perspective on the possible technical causes of error. Because the issues raised are of concern to non-modellers, the arguments are sometimes preceded by background information on the modelling approaches. This background material is by necessity superficial and abbreviated. Some of the discussion is addressing issues that were exhaustively examined long ago in the academic literature of discrete choice modelling, but many of the lessons of the theoretical study have not made their way into the work of practicing modellers. This paper is focused on what has actually been done to produce the modelling inputs into investment decisions and what mistakes may have been made.

2. Toll model approaches

Toll modelling is a specialised form of general demand modelling, and so it makes use of many of the same processes and tools as other demand modelling. In its simplest form it can be done using rules of thumb and spread-sheet analysis, but for major projects it is usually done in the context of a strategic transport model. All of the major toll road projects that have been done recently in Australia have been analysed using some form of strategic transport model, in most cases based on the most commonly used transport model for the city in question (such as the Brisbane Strategic Transport Model, or the Melbourne Integrated Transport Model). These models have almost all been traditional four step models (FSM), so most toll modelling has been done by including toll considerations at one or more steps within the FSM.

There are newer transport modelling approaches, including the author's 4S model (Davidson 2011) which have not been used as widely. This paper will briefly mention these approaches, but will focus on the assumptions and limitations of the more commonly used methods.

It is difficult to get good, publicly available information on the structure of the models used for toll assessment. Particularly when projects are developed as Public Private Partnerships (PPP) most if the information is commercial in confidence, even after the contracts have been awarded. There are a number of public reports that describe the structure of the toll model used, but in most cases, detailed information on the model is rare, and very few reports give model coefficients. There are exceptions, at least with some government studies. This paper makes use of two quite detailed reports –Appendix C of the Hale St Link Feasibility Study (Business Case Traffic and Transport Report) prepared by SKM/Connell Wagner for Brisbane City Council and the Hills M2 Upgrade Environmental Assessment (prepared by AECOM for the RTA). The approaches used in these studies are believed to be representative of typical toll road studies in Australia. It should be noted that both of these studies were done by government agencies, rather than private sector proponents. Typically government sponsored modelling is more conservative that that done by bidders for PPP's.

The Hale Street Link (now named the Go Between Bridge) is a new toll bridge across the Brisbane River, opening on 5 July 2010. It was developed by the Brisbane City Council as one element of the TransApex plan. The Hale Street Link Business Case predicted Mar 2010 demands of 1.7m trips/quarter with an annualization of 340, which equates to around 20,000 veh/day with a toll of $2.20 ($2 in 2006$), rising to around 2.5m trips/quarter or just less than 30,000 veh/day in 2021 (Hale Street Link Integrated Project Team 2006, p.81; SKM Connell Wagner Joint Venture 2006b, p.39) . Note that the 2010 traffic estimates did not incorporate the effect of traffic ramp-up (these effects were incorporated into the Revenue Forecasting model) so the opening numbers would be expected to be 30% lower than those given here (14,000), rising to the full estimates over 2 years (Hale Street Link Integrated Project Team 2006, p.103). After Clem 7's disappointing results, but before the bridge itself opened, the projections were revised to 12,500 on opening, 17,500 by 2011 and 21,000 by 2021 (Moore 2010). These results reflected a change in the configuration of the connections at either end of the bridge with fewer grade-separated movements; according to the Project Modification Report, the changes had little impact on traffic volumes (Brisbane City Council 2008, p.29). Once the bridge opened, it carried 13,000 at the end of 2010 with a toll of $1.50, and one year after opening (July 2011) it carries around 14,000 trips with a toll of $2.42 (Moore 2011b). The Council predicts that the bridge will carry more than 20,000 veh/d within the next five years, and state that the bridge continues to hit traffic targets (Moore 2011a). Since many of these demand estimates come from newspaper reports and press releases, it is unclear if the numbers are average weekday traffic (AWDT) or average daily traffic (AADT) so direct comparison is difficult. Nonetheless, it does appear that the Hale Street Link Business Case over-predicted demand, and that its level of over prediction is similar to that observed by Bain and Flyvbjerg.

The Hills M2 Upgrade project is for the widening and upgrade of Sydney's M2 Motorway, including four new ramps. The upgrade is underway but will not be completed until 2013.

3. Toll approach: Inclusion in assignment model

The simplest way to model toll choice is to treat the decision to use the toll road the same as all other decisions on which road to use. The toll road is included as a standard link in the network, and the toll is included as a cost component on that link. This is usually done using a generalised cost route choice model, where paths are built based on minimising aggregate cost (including value of time and vehicle operating costs as well as tolls). It is also possible to incorporate tolls within a minimum-time path build, by converting the toll into equivalent travel time.

The traffic assignment model will assign traffic to the toll road whenever it is a part of the shortest path for a particular travel segment (trip between an origin and a destination). In the absence of congestion, this would be an all-or-nothing process. This can lead to discontinuities in demand (and elasticities) with varying toll levels. A link that is on the shortest path will be assigned all relevant demand, and this will be unaffected by increasing toll levels until a critical point is reached where the road stops being on the shortest path. At this point all relevant demand will switch to an alternative route. Depending on the characteristics of the travel market for the road this can lead to sudden drops in demand, as a key market segment switches to the next best route.

The inclusion of congestion reduces this somewhat, as an equilibrium process will ensure that traffic will switch to alternative routes as the shortest path route becomes congested. This can reduce discontinuities but is unlikely to eliminate them.

It should be noted that this model includes no allowance for random variability – it is fully deterministic. The only thing causing any spread in behaviour in route choice is congestion, which causes traffic to equilibrate between similar routes.[1]

This process is very easy, and can be included in any transport model. It also has the advantage of incorporating the influence of congestion without any extra looping other than the standard iteration of the equilibrium assignment process.

The difficulty is that it often produces low estimates of demand for toll roads. The reason is that the conversion between money and time required for the simple route choice trade-off is done using average values of time. However the users of toll roads may be expected to have higher than average values of time. To allow for this, the analysis is sometimes done with artificially reduced toll charges. For example, here is the description of the Commercial Vehicle toll choice model in the Hale St Link analysis.

> *Commercial Vehicles are assigned as a separate class, alongside the non-tolled private vehicles in the multi-class equilibrium assignment during each iteration, with access to tolled routes at 50% of the actual toll. This approach was tested and gave an acceptable level of CV volumes on toll roads. (SKM Connell Wagner Joint Venture 2006b, p.12)*

In this case it appears that the simpler approach was used due to lack of behavioural data for commercial vehicles, and it should be noted that more complex modelling was used for other vehicle classes.

Nonetheless, the description above shows one of the key problems with this approach. If we accept that the use of full tolls gives unrealistic outcomes, then the question is how to determine what discount to the toll is appropriate. It is possible that simple rules of thumb could be used, but it is not clear that these would be transferable. Without some basis for determining the discount, this approach lends itself to merely producing the expected outcomes. And a model which is calibrated to give expected, acceptable outcomes is particularly vulnerable to optimism bias.

A recent US study found that of the 12 major toll road projects surveyed, 4 of them included tolls as a component of generalised cost in the volume-delay function. (Kriger et al. 2006, p.99).

---

[1] Note that there are stochastic assignment techniques (Daganzo & Sheffi 1977), but the author is not aware of them being used for toll modelling.

4. Toll approach: Market segmentation

One of the key problems with including toll choice in the assignment model is that, more than other aspects of travel, the trade-off between tolls and travel time is strongly influenced by variations in value of time. As discussed earlier, the use of a single value of time will tend to under-estimate toll road use, since the market for toll roads is generally those with a higher value of time. So a simple solution is to use multiple values of time.

The diagram shown below is from the Hills M2 Upgrade Environmental Assessment (RTA 2010). The report notes that

> *The trip matrices are split by vehicle type and purpose. The private car purposes (commute, business and other) are further segmented by 3 household income groups being high, middle and low. This further segmentation allows for further divisions of value of travel time (VOT), and by increasing the number of segments allows a more detailed assessment of tolled versus non-tolled choice across the network. The segmentation occurs at a zonal level and is based on household income data from the 2006 census. The sensitivity to geography is important as it acknowledges that there are lower income and higher income suburbs across the Metropolitan area.*

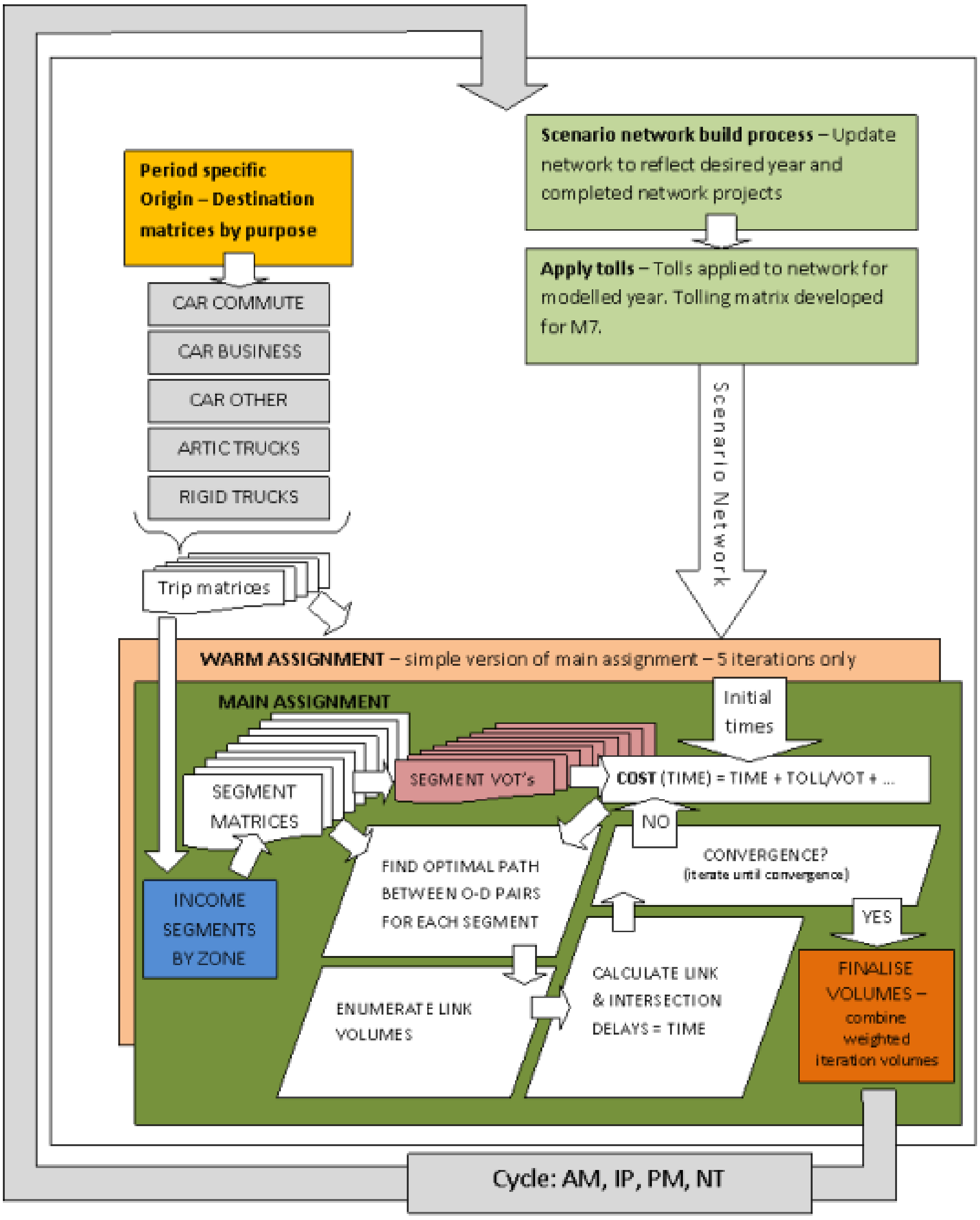
By segmenting the travel matrices (for car travel at least) by income, the route choice model can incorporate some degree of value of time variation. Each cell of the matrix is effectively assigned three times (in this case) and each assignment can use a different path. Of course the assignment can still incorporate congestion (through a multi-class equilibrium assignment).

Note that the Hills M2 model (based on the Transurban Sydney Strategic Traffic Model) used fixed demand matrices, which were segmented by income post-distribution. Thus the model cannot incorporate the variation in overall demand due to the project, nor the differential effects on different income groups. So it would, for instance, assume that the distribution of demand by income between two locations is completely unaffected by the presence of a toll road. But we would expect that the destinations reached by the toll road would be relatively more attractive for the higher income travellers, and so the demand for those destinations would be higher for those with high values of time.

It is not strictly necessary that the segmentation be done by income, or that demographic variations be incorporated into the factors. In fact there is good reason to think that even within an income cohort there will be some variation in value of time. In terms of allowing differential response to toll roads, the approach would work almost as well if fixed factors were used, since this would still allow for variation. The use of income allows other data sources to be used to determine the proportions, most importantly the Population Census. However it would be possible to determine proportions based on behavioural surveys.

Like the assignment based toll choice model, this approach is quite simple, and can be added to existing models without too much effort. But it does have the benefit of allowing for some variation in values of time. Because of this it does not need any of the adjustments to toll levels referred to above to allow for the fact that toll roads tend to be preferentially used by those with a higher than average travel time. There are no obvious biases in this approach, and simplicity of the approach allows it to be readily understood.

Figure 1 : Typical structure for market segmentation approach



Source: (Transurban 2010, p.114)

5. Toll Approach: Logit toll choice

From what little information has been made available on projects across Australia, it appears that the most common approach to toll modelling in recent years is the logit toll choice model. Most of the models are similarly structured and are simple extensions of the standard multi-nomial logit model used widely in many modelling applications, in particular mode choice models.

The logit choice model has been used for

- Brisbane North South Bypass Tunnel/Clem 7 (Maunsell 2005, p.75; Maunsell AECOM 2006, p.4)
- Airport Link (SKM Connell Wagner Joint Venture 2006a, pp.13-232)

- Hale St Link/Go-Between Bridge (SKM Connell Wagner Joint Venture 2006b, p.11)

It is understood that this approach has also been used for other toll road projects in Australia, but it is difficult to find public documents that definitively state this. A recent US study found that of the 12 major toll road projects surveyed, 6 of them used some sort of diversion curve or toll elasticity, most likely based on the logit model. (Kriger et al. 2006, p.99).

The argument for the use of a logit toll choice model is that it allows for variability in the response of individuals to toll roads, and it has a form which is familiar to modellers, and which is readily implemented with the common modelling tools.

Logit toll choice models incorporate a number of parameters that must be estimated from behavioural data. The simplest type of data to use for calibration is stated preference data, where a range of hypothetical alternatives are presented to survey subjects, usually in groups of two or three. People are asked to select their preferred alternative, and this provides a sample point for a discrete choice calibration. There are potentially problems with this data source, as discussed in the section below on Stated Preference Surveys.

Because the use of the logit toll choice model is so pervasive, it is worth considering the underlying assumptions to this approach, and the potential problems, in more detail.

5.1 Theoretical background

The discussion here is a simplified introduction to the logit model that seeks to explain enough to allow non-modellers to follow the argument. Most introductory modelling textbooks will have a much more thorough treatment, and a much deeper treatment can be found in the excellent text by Kenneth Train "Discrete Choice Methods with Simulation" (Train 2003), freely available on-line.

The logit model is an application of the Random Utility Model is based on a number of simple assumptions.

1. People make choices by determining the desirability of each alternative (its utility) and choosing the one that is most desirable.
2. There is variation in people's assessment of utility, due to both individual variability and modeller ignorance and so utility can be understood as a random variable, with parameters such as variance and mean that show how it is distributed across a population.
3. The utility variable is made up of two parts – a deterministic component that is (to some extent) observable, and a random error. That is, for each alternative i, the utility is given by $U_i = V_i + \epsilon_i$, where U_i and ϵ_i, are random variables.
4. Many models can be constructed from the preceding assumptions, but the logit model further assumes that the random error ϵ is i.i.d. – that is it is independently and identically distributed across all alternatives. This means that if we know exactly how a person feels about one alternative it tells us nothing about how they feel about another alternative (independent). Further, the degree of variation in how people perceive one alternative is the same across all alternatives; no option is more or less random than any other (identically distributed).
5. Finally the logit model assumes that the error term takes a particular random distribution, called the Gumbel or Type I extreme value distribution. It is chosen because it has a similar bell-shaped curve to the normal distribution, but a much simpler probability density function (it's cumulative distribution function is given by $F(\epsilon) = e^{-e^{-\epsilon}}$) which makes all of the maths much easier.

From these assumptions a number of conclusions can be formed. Firstly, because choice is determined only by which alternative had the highest utility the absolute value and the scale of utility does not make any difference. We can add any constant to the utility of all alternatives, or we can multiply the utility of all alternatives by a constant amount without changing their relative ranking. In other words only differences in utility matter and the scale of utility is arbitrary.

Secondly, because of the random error term, people do not always choose the option with the highest deterministic utility; there is always a chance that the random factors will make up for any lack in deterministic utility that an option has. Thus the probability of choosing an option is equal to the probability that the random factors are enough to make it the most attractive option. From this it is obvious that the "all-or-nothing" approach discussed earlier can be seen as a special case of the logit model where there is no random factor (or the variance of the random variable ϵ is zero).

Working through the maths, it is possible to use the assumptions given above to derive the familiar equation for logit choice.

$$P_i = \frac{e^{U_i}}{\sum_j e^{U_j}}$$

In the case of a binomial choice, when there are only two options A and B, this simplifies further to

$$P_A = \frac{1}{1 + e^{U_B - U_A}}$$

Thus the probability of choosing an option depends only on the difference between the utilities.

Because the scale of utility is arbitrary we can normalise the utility in any way that we like by multiplying by any desired constant. But whatever we do to the fixed utility must also be done to the random error. Sometimes normalisation is done to set the variance of the error term to a convenient constant. But to better visualise the assumptions of the model it is useful to normalise to a convenient scale, such as minutes or dollars. By doing this we can understand the variance (or standard deviation) of the random error component in familiar units.

Again by working through the maths it is possible to show that the standard deviation of the random error term is dependent on the scale that we use for utility (β) through the following equation.

$$\sigma = \sqrt{\frac{\pi^2}{6\beta^2}}$$

Using this equation it is possible to interpret model coefficients in terms of the degree of variability in the error term. It should be remembered that the error term ϵ is completely unrelated to any aspect of the alternatives – it is a measures of all of the random influences that we do not know anything about. The bigger the variance of the error term, the more the model is saying that the deterministic factors do not matter and random chance is the dominant influence on choice. In the extreme case, where the variance of ϵ is infinite, the deterministic factors have no effect, and each alternative is equally likely to be chosen.

One way of looking at it is to imagine that people make an assessment of each alternative, give them a score and then roll some dice and add its result to the score. They then choose the alternative with the highest score. If the standard deviation of the error term is small then the values on the dice are small compared with the scores – for example we might be scoring out of 100 and then rolling 2 dice. If the standard deviation of the error term is large then the values on the dice are large compared with the scores – for example we might be scoring out of 10 and then rolling 5 dice. In the second case the scores still have an impact, but the roll of the dice is the biggest factor.

5.2 Typical toll model structure

So what does this mean for toll models? Well the most common approach is to assume that when people are choosing their route they make a high-level choice between the best tolled route and the best non-tolled route. This is treated as a binomial logit choice, where the deterministic part is the generalised cost of each route.

The publicly available information on the Brisbane City Council's Hale St Link project gives us a rare level of detail on an operating toll choice model, including not only the model structure, but also model parameters. The structure of the model is similar to that used in other studies, and the calibration was done using standard techniques on a stated preference survey. SKM/Connell Wagner, the consultants who did the work for BCC describe the model structure using the diagram below.

The diagram shows that there is some additional complexity that arises because of the interaction of multiple toll roads, and issues to do with convergence, but that the heart of the model is the binomial logit choice between the best tolled route and the best free route.

According to the technical report (SKM Connell Wagner Joint Venture 2006b, p.14), the utility function for each alternative route is given by

$$U_k = \beta_1\, Toll + \beta_2\, TravelTime + \beta_3\, DelayTime + ASC$$

Where $TravelTime$ is the number of minutes spent travelling in (relatively) unconstrained conditions, whereas $DelayTime$ is the number of minutes travelling at speeds less than 40km/h.

As discussed earlier, utility values can be scaled by a constant factor or have a constant factor added to them. But any scaling will also affect the variance of the error term. The scaling is somewhat arbitrary, and sometimes scaling is done to give a specified variance. However it is easier to understand what the parameters mean if the utilities are scaled to understandable units. Thus we can scale the equation above to give utilities in monetary units by setting the factor on the toll to one. This gives

$$U_k = Toll + \frac{\beta_2}{\beta_1}\, TravelTime + \frac{\beta_3}{\beta_1} DelayTime + \frac{ASC}{\beta_1}$$

If we call $\frac{\beta_2}{\beta_1}$ the value of time θ

Then

$$U_k = Toll + \theta\, TravelTime + \gamma\theta DelayTime + ASC'$$

5.3 Implicit assumptions in toll choice models

The Hale St Link modelling used the following parameters (SKM Connell Wagner Joint Venture 2006b, p.18):

$\beta_1 = Scale\ Factor = -0.0029$ /c

$\theta = Value\ of\ Time = 18c/m$ = \$10.80/hr

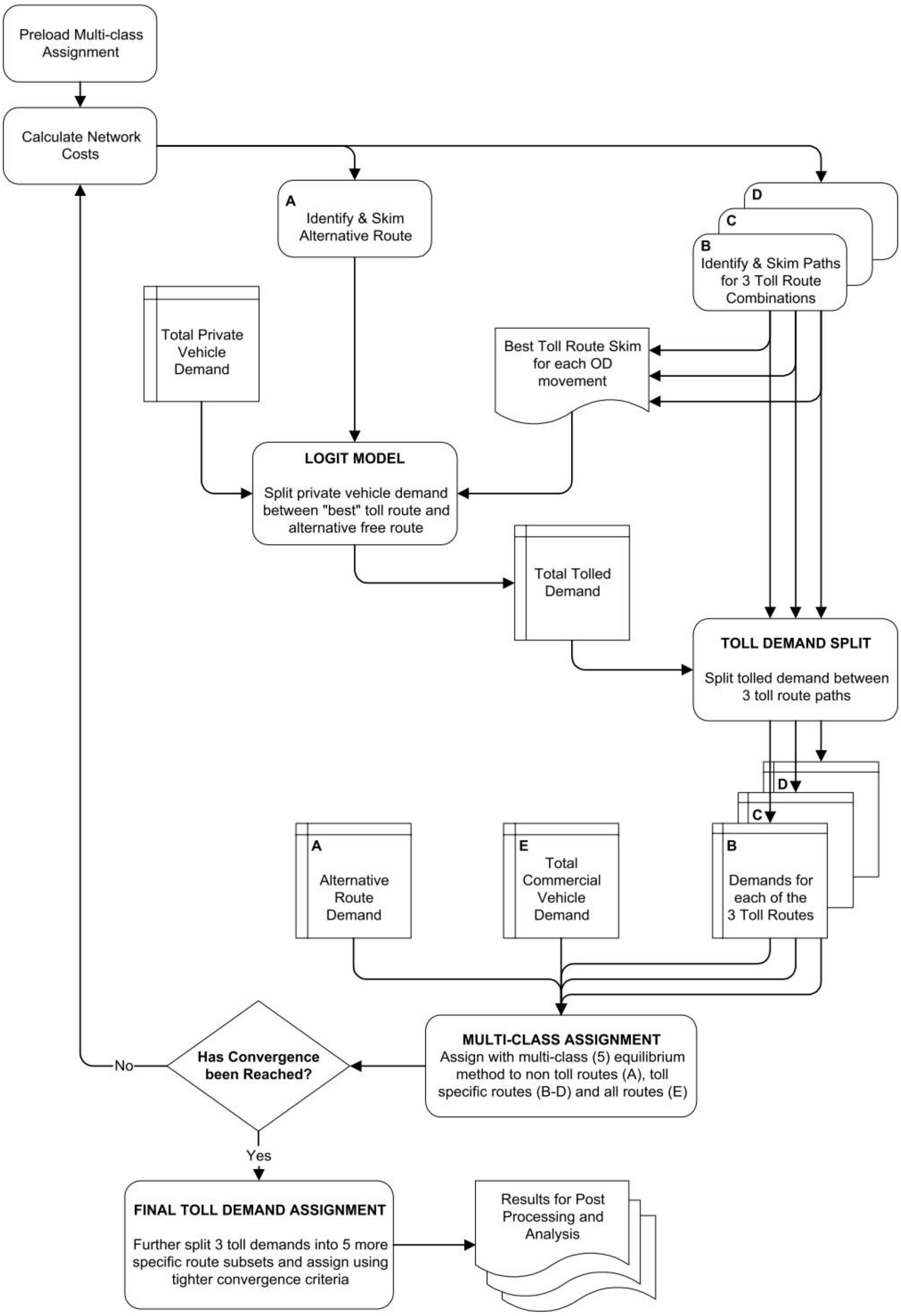
$\gamma = Delay\ factor = 1.65$

Using the equation given above we can see that the scaled standard deviation of the error term is

$$\sigma = \sqrt{\frac{\pi^2}{6\beta^2}} = \frac{\mathbf{1.282}}{|\boldsymbol{\beta}_1|} = 442.2c = \$4.42$$

So the calibration done for the Hale Street Link resulted in an error term with a standard deviation of $4.42 (with utilities specified in dollars).

Figure 2: Typical structure for logit approach



Source: (SKM Connell Wagner Joint Venture 2006b, p.6)

On page 78 of the Draft Business Case we find that the predicted travel time savings due to the toll road are up to 16 min in the peak period in 2021. If we assume that most of the time saved was in the Delayed Time category (less than 40km/h) then the value of the travel time saving is 1.65 x $10.80 x 16/60 = $4.72. With a toll of just under $2, the systematic difference between the two alternatives is $2.72. Given the fact that the Hale St Link is a relatively short bridge, it is likely that the travel time on the tolled route is quite low – say 5 min, with a value of around $1.50. So the most compelling market for the toll has U_{free} = -$6.22 and U_{toll} = -$3.50, with a utility difference of $2.72.

5.4 Problems with these assumptions

What this means is that the standard deviation of the error term is of a similar scale to the utilities themselves, and significantly higher than the utility savings. In terms of the analogy given above, it is as though we are giving the alternatives a score out of 10 and then rolling two dice; the final scores are significantly impacted by completely random factors.

In effect, the model is saying that people's choice between tolled and un-tolled routes is only mildly influenced by savings in travel times or tolls, and is largely influenced by completely random factors. In the binary choice formulation it effectively says that some people just like toll roads and some people just like free roads and the details of the alternatives do not matter too much.

This last point is demonstrated through some simple calculations. Consider the market mentioned above, where the travel time savings are 16 min and the toll is $2. The logit model gives the percentage choosing the tolled option at

$$p(toll) = \frac{1}{1+e^{-0.0029(472-200)}} = 68\%$$

Now consider a different market, where the toll is still $2, but instead of saving time it causes a delay of 16 min.

$$p(toll) = \frac{1}{1+e^{-0.0029(-472-200)}} = 12\%$$

So even though the tolled route makes no sense at all for this market (paying money to be delayed), a reasonable proportion of the people would still (according to the model) choose to pay the toll.

Now the problem with this comes through a consideration of how the model is applied. Each zone pair is assessed separately, with a separate toll/non-toll split calculated for each cell in the matrix. For a fairly small number of cells, the tolled route will be very compelling, with significant time savings. The model will allocate most (but not all) of these to the tolled route. But there will be a much larger number of cells where the tolled route will be fairly marginal, or not at all attractive. The logit model will still allocate a significant portion of these to the tolled route, even if it offers nothing but extra travel time.

Since there are likely to be many more cells that it does not work for, than cells for which it is attractive, the model will tend to overestimate total demand.

6. More advanced discrete choice models

There are alternative formulations of discrete choice models that overcome the deficiencies of the logit model. These include older approaches, such as the probit model, and more recent approaches, such as mixed logit or latent class models. The more advanced discrete choice models are more complex to implement, particularly because that generally have no closed-form expression, and so numerical or sampling methods must be used to find approximate solutions. To the author's knowledge, there have been no major toll roads analysed in Australia using these approaches, although some academic models have been developed, for example (Greene & Hensher 2003).

7. Overuse of stated preference data

The development of an understanding of how people respond to tolls needs some empirical data, particularly if numerical models are to be developed. There are two main ways in which we can obtain information about travel behaviour – we can watch or ask what they do (revealed preference) or we can ask them what they think they would do in certain situations (stated preference). It is generally much easier to obtain a response to a range of circumstances through stated preference surveys. This is because we can ask people about a whole range of different options in a single sitting. It is much more difficult to find and collect data on people's actual behaviour across a wide range of options, particularly when some of those options do not exist.

There is a growing body of literature on the problems with stated preference surveys, most notably the incidence of hypothetical bias. Even with simple choices, there is evidence that when people are asked what they would hypothetically do, they are often wrong. For simpler product-based choices a range of studies have found that respondents overstate their willingness to pay, often by a large amount (Murphy et al. 2005). It is not clear how much worse people are when asked to make hypothetical preferences between complex choices, such as is the case in many stated preference surveys for toll road preference.

It is instructive to consider how much cognitive effort goes into people's daily travel decisions. People generally spend an hour or more per day travelling, and it is likely that a not insignificant fraction of that time is spent thinking and reviewing their travel choices. People experiment with different modes and routes, and watch how fast or slow other travellers go. People often take months to fine tune their routes, or time of day preferences. Also much of this assessment is done sub-consciously and intuitively, and will include visual cues and emotional responses. For less frequent travel the decisions process would obviously be much simpler and shorter, but will often be founded on the framework built up for more frequent travel (for example, people may travel to an unknown destination by following most of a known route and then departing from it near the destination).

Compare this to the presentation of a hypothetical scenario in a typical stated preference survey. Two or more scenarios will be presented, with a list of numerical attributes (such as travel time, fares, level of congestion). People are asked to review the numbers, imagine what the various alternatives would be like, and then make a preference. To do this really effectively would take a high level of imagination, numeracy and self-knowledge. And then once one question is answered it is time to move onto the next.

Some analysis has been done of the effect of survey complexity on responses (Hensher 2006; Louviere et al. 2008) which generally show that more complex surveys lead to less consistency in people's responses. For complex transport choices it is much harder to assess whether respondents are giving answers to hypothetical questions which are consistent with their response to real world situations.

It is easy to believe that survey respondents adopt simple heuristics and rules-of-thumb to answer these surveys. Compounding this problem, respondents may think about why they are being asked these questions and what are the likely consequences of their answers. This is similar to self-completed television viewing surveys, where respondents claim to watch a surprisingly large number of documentaries and current affairs programs, and surprising few soap operas and reality TV shows. If people want the project to go ahead, they may give answers that they think will make that more likely. In contrast, people may stubbornly state that they would not use a toll road no matter how much time they would save.

These possible problems with stated preference surveys compound with the problems identified above with logit toll choice models. As discussed earlier, the calibration of a logit model is concerned (in part) with finding the scale of the error term in the utility function. If there is a lot of variation then the model becomes relatively insensitive to the alternatives themselves, the logit curve is fairly flat, and generally toll demands will be too high. But if the calibration data itself has lots of noise (uncorrelated randomness in people's responses) then the calibration will almost certainly produce a fairly flat distribution.

The issue of hypothetical bias in these sort of surveys is something that is being addressed in the literature and should be addressed in more detail, particularly given the scale of decisions made using these surveys. But it may be possible to be more imaginative in finding new ways of collecting revealed preference data. One way is to look for natural experiments that present a range of alternatives, or that change conditions. For example changes can be observed and recorded every time a toll road changes its price, or when planned road works change conditions for an extended period, or when fuel prices change significantly in a short time. Current behaviour could be analysed on alternative routes where people make a trade-off between speed and distance, or between flat and hilly roads, or between congested routes and uncongested routes.

All of this revealed preference data could be obtained more easily if longitudinal data is automatically collected, using technology such as in-car or personal GPS, automatic number plate surveys, or blue-tooth tagging.

8. Segmented stochastic slice simulation (4S) model

The author has developed a new approach to modelling, called the 4S model (Segmented Stochastic Slice Simulation). This approach (described in Davidson, 2011) offers a number of advantages over traditional models, particularly in its ability to reflect complex and varied behaviour, and in its ability to model high levels of spatial detail over a wide geographical area.

It works with a unified choice model which incorporates the desire to travel, destination choice, time-of-day choice, mode choice, and route choice (which includes toll choice). Like the Market Segmentation approach discussed above, the 4S model looks at a range of values of time. But instead of these being fixed, they are taken as Monte-Carlo draws from random variables, along with other parameters that determine travel behaviour. Since all aspects of choice are integrated, the model allows destination choice and mode choice to be influenced by value of time and road tolls. Thus the model will properly incorporate the effects of induced demand on toll road forecasts, and will take into account the differential impacts across income ranges. The model also manages time explicitly, and allows for time-varying demands. This makes it possible to consider tolls that vary across the day, and the differential competitiveness of toll roads for different travel purposes.

The 4S model is still fairly new, but it has been used for a number of projects, including one major toll road investigation (the Toowoomba Bypass). It is the author's hope that it will provide one way forward for improving the technical aspects of toll demand forecasts.

9. Conclusion

As discussed in the introduction to this paper, Flyvbjerg has argued that technical explanations alone are inadequate to explain the widespread errors in toll demand forecasts. This is because technical problems would be expected to cause roughly symmetric distribution of inaccuracies (with under-prediction just as likely as over-prediction) and would lead to improvement over time as error sources were identified and corrected. However an empirical examination of projection errors shows that toll demands are usually overestimated and error rates are not improving. From this Flyvbjerg concludes that the key problems are systematic bias – optimism bias and strategic misrepresentation.

Bain (2011) emphasises the importance of general prediction error, and notes that even if bias is corrected, there is still the need for large confidence intervals attached to any future estimates[2]. He also notes that there is a temptation to assume that bias is entirely man-made, but that some elements of bias could be model related.

The preceding discussion on the problems with logit toll choice models could be one area where a systematic prediction error could occur without being due to deliberate or unintentional bias. There are also potential problems with the use of stated preference surveys, where there is the possibility that people are not very good at understanding alternatives, or predicting their behaviour, or they may even deliberately adjust their answers to get the outcome they want. This could compound the logit problems.

Flyvbjerg's basic argument probably still holds – the problems identified in this paper are unlikely to be present in all projects, and it is very believable that studies are subject to conscious or unconscious bias.

However where technical problems can be addressed then it would make sense to do so. These are the author's recommendations to improve the technical aspects.

Work to eliminate secrecy (where possible) and encourage peer review.

Most studies are done by teams brought together for the project, and peer review (if it occurs) is often internal, or done as part of the modelling contract (with an obvious conflict of interest). Working papers and survey data are usually locked up in commercial-in-confidence data rooms, even after contracts have been finalised and funds have been raised. Minimising the scope and duration of confidentiality agreements as far as is commercially feasible, routinely making technical reports available, and ensuring that they include detailed descriptions of models (including parameters) would improve transparency and reduce the risk of bias. This would be further enhanced by routine, independent and thorough peer review which should include access to networks, models and all assumptions.

Look for more sources of revealed preference data

Stated preference surveys are simple to implement, easy to calibrate from, and are widely used. But it is not clear that they provide reliable information, and studies of simpler product choice surveys raise the possibility that hypothetical bias could be significant. The wider use of longitudinal travel surveys, often supported by newer technologies, could make revealed preference surveys a practical alternative. This is particularly true in Australia, now that we have a range of existing toll roads that can provide on-the-ground natural experiments.

Look for alternatives to logit based toll choice models

This paper has addressed some of the risks associated with using logit toll choice, particularly when the calibrated curves are fairly flat. The logit model does not allow for variation in value of time, but only for a completely random error that is unrelated to any aspect of the alternatives. If a simple alternative is sought, then the segmented approach would seem suitable, with a reasonable number of income/value of time segments. Alternatively, one of the newer approaches could be used, such as mixed logit (which does allow for variation in value of time), or the author's 4S model.

Properly account for induced demand

Many studies assess toll roads with the assumption that the overall pattern of demand is unchanged (fixed matrices). This is unrealistic, as the addition of a major new piece of infrastructure will change patterns of travel, including mode choice, destination choice and

---

[2] One way of assessing the size of these confidence intervals is to use Monte Carlo analysis of a range of model runs. This can be done directly in the 4S model, or done using multiple runs of a traditional model. See (Lemp & Kockelman 2009) for a discussion.

may lead to increase levels of total travel and changes to land use. These changes should be considered in the modelling. Also the economic assessment of projects should include consideration of induced demand - incorporating accessibility benefits rather than just travel time savings.

Extend the use of risk and scenario assessment

Most studies include some degree of scenario analysis, where key parameters are varied to investigate their effect. But these are often fairly limited in scope, and usually consider relatively small individual changes in a small number of variables. The wider use of Monte Carlo approaches (often included in the financial modelling) in transport models would show the compounded risk of multiple factors working together, and give a better understanding of the confidence intervals, or error bars that should be attached to all model results.

10. References

Bain, R. (2009) Error and optimism bias in toll road traffic forecasts. *Transportation*, 36(5), pp.469-482.

Bain, R. (2011) Error and Optimism Bias in Toll Road Traffic & Revenue Forecasts. Available at: http://www.bitre.gov.au/Info.aspx?ResourceId=802&NodeId=188.

Brisbane City Council (2008) Hale Street Link Project Modification Report.

Daganzo, C.F. & Sheffi, Y. (1977) On stochastic models of traffic assignment. *Transportation Science*, 11(3), p.253.

Davidson, P. (2011) A new approach to transport modelling - the Stochastic Segmented Slice Simulation (4S) model and its recent applications. *Australian Transport Research Forum*.

Flyvbjerg, B. (2008) Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice. *Europen Planning Studies*, 16(1). Available at: http://dx.doi.org/10.1080/09654310701747936.

Flyvbjerg, B., Holm, M.K.S. & Buhl, S.L. (2005) How (in) accurate are demand forecasts in public works projects?: The case of transportation. *Journal of the American Planning Association*, 71(2), pp.131-146.

Goldberg, J.L. (2006) The fatal flaw in the financing of private road infrastructure in Australia. In *Australasian Transport Research Forum*. Available at: http://patrec.org/web_docs/atrf/papers/2006/1471_Goldberg%20(2006).pdf.

Greene, W.H. & Hensher, D.A. (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8), pp.681-698.

Hale Street Link Integrated Project Team (2006) Hale Street Link Business Case (Working Draft). Available at: http://bi.mipo.jsadigital.com.au/Documents/Draft Business Case 13Nov06.pdf.

Harvey, M. (2011) Review of Traffic Forecasting Performance - Toll Roads. Available at: http://www.bitre.gov.au/publications/02/Files/Review_of_Traffic_Forecasting_Performance_Toll_Roads_040711.pdf.

Hensher, D.A. (2006) Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. *Environmental and Resource Economics*, 34(1), pp.7-44.

Kanowski, S. (2011) Case studies of recent Australian toll road projects. Available at: http://www.bitre.gov.au/publications/02/Files/GHDSymposium_Presentation_Final_submitted.pdf.

Kriger, D., Shiu, S. & Naylor, S. (2006) *Estimating toll road demand and revenue*, Transportation Research Board. Available at: www.trb.org/publications/nchrp/nchrp_syn_364.pdf.

Lemp, J.D. & Kockelman, K.M. (2009) Understanding and Accommodating Risk and Uncertainty in Toll Road Projects. *Transportation Research Record: Journal of the Transportation Research Board*, 2132(-1), pp.106-112.

Louviere, J.J. et al. (2008) Designing discrete choice experiments: Do optimal designs come at a price? *Journal of Consumer Research: An Interdisciplinary Quarterly*, 35(2), pp.360-375.

Maunsell (2005) North South Bypass Tunnel - Traffic and Transport.

Maunsell AECOM (2006) Traffic Forecasts for the North-South Bypass Tunnel included in the River City Motorway Product Disclosure Statement. Available at: http://www.rivercitymotorway.com.au/userfiles/file/Final%20Version%20from%20ASX%20website.pdf.

Moore, T. (2010) It's a long way to be back in black. *Brisbane Times*. Available at: http://www.brisbanetimes.com.au/business/its-a-long-way-to-be-back-in-black-20100609-xwi5.html [Accessed August 4, 2011].

Moore, T. (2011a) A birthday to celebrate? Go Between Bridge turns one but critics remain. *Brisbane Times*. Available at: http://www.brisbanetimes.com.au/queensland/a-birthday-to-celebrate-go-between-bridge-turns-one-but-critics-remain-20110705-1h0l0.html [Accessed August 8, 2011].

Moore, T. (2011b) Mayor predicts Go Between boost. *Brisbane Times*. Available at: http://www.brisbanetimes.com.au/queensland/mayor-predicts-go-between-boost-20110705-1h068.html [Accessed August 4, 2011].

Murphy, J.J. et al. (2005) A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, 30(3), pp.313-325.

SKM Connell Wagner Joint Venture (2006a) Airport Link Environmental Impact Statement - Traffic and Transport. Available at: http://www.airportlinkeis.com/OtherLinks/EIS/pdfs/Vol3-TechPapers/01_Traffic_and_Transport.pdf.

SKM Connell Wagner Joint Venture (2006b) Hale Street Link Feasibility Study, Appendix C - Business Case Traffic and Transport Report. Available at: http://bi.mipo.jsadigital.com.au/Documents/Appendix C - Part 1.pdf.

Train, K. (2003) *Discrete choice methods with simulation*, Cambridge Univ Press. Available at: http://elsa.berkeley.edu/books/choice2.html.

Transurban (2010) Hills M2 Upgrade Environmental Assessment- Technical Paper 1 - Traffic and Transport Impact Assessment - Appendix A. Available at: http://www.hillsm2upgrade.com.au/files/environmental_assessment/volume2_part1/techpaper1/m2ueavol2pt1ttiaappendixa.pdf.