# MODELLING TIME PERIOD CHOICE IN LARGE-SCALE HIERARCHICAL DEMAND MODELS: SOME PROBLEMS AND A SOLUTION

Andrew Gordon, Mott MacDonald
Andrew Daly, RAND Europe and Institute for Transport Studies, Leeds
John Bates, John Bates Services
Frederic Oladeinde, Department for Transport

## 1. INTRODUCTION

The hierarchical logit model is the cornerstone of many real life applications of transport demand modelling around the world. It can be used to model a variety of traveller choices, including those of mode, destination, time period and trip frequency.

The hierarchical logit model is the form recommended in recent Web-based Transport Analysis Guidance (WebTAG) published by the UK Department for Transport (see http://www.webtag.org.uk), which also recommends an incremental application of the model. The guidance is intended for use by practitioners in the UK who are undertaking modelling as part of the appraisal of a transport scheme.

This paper describes the difficulties encountered in fully complying with certain aspects of the guidance, and the solution we have developed to overcome those difficulties.

## 2. BACKGROUND

During much of the 1980s and 90s transport modelling in the UK focused on traffic assignment modelling, with rerouting assumed to be the only response to changes in the generalised cost of travel. From the mid-1990s variable demand modelling became more widespread as a result of the publication of the report “Trunk roads and the generation of traffic” by the Standing Advisory Committee on Trunk Road Appraisal (SACTRA) (SACTRA, 1994). However, these models mostly used so-called simple elasticity functions where the demand for travel for a particular zone pair is a function of the cost of travel for that zone pair only.

The UK Department for Transport issued new advice on variable demand modelling in June 2006 (DfT, 2006). This recommends that full variable demand modelling should be used for the modelling and appraisal of most transport schemes. This entails explicit representation of destination, mode and time period choice responses. The incremental hierarchical logit (IHL) is the preferred model form. The work described in this paper concerns two specific recommendations:

- Demand modelling should be undertaken with trip matrices in production-attraction (PA) format, rather than origin-destination (OD)

- Time period choice, if modelled, should *not* appear *below* destination choice in the hierarchy

However, it turns out that no model in the UK has satisfactorily managed to combine 'traditional' PA-based demand modelling *and* modelling time period choice above destination choice in the hierarchy in a way that takes proper account of outbound and return travel costs, and ensures consistency between outbound and return modes and destinations.

This was a difficulty that had to be resolved for implementation in the DIADEM software (http://www.diadem.org.uk) that has been developed by DfT to support the WebTAG guidance.

Before discussing the details of the problem we first set out the difference between OD and PA-based demand modelling and then describe the 'traditional' approach to PA-based demand modelling.

Suppose a traveller goes from home to their workplace and back home again. Using the OD representation these are two independent trips: one from the home zone to the workplace and another from the workplace zone to home. Under the PA representation this is two PA trips, with the home as the production and the workplace as the attraction zone for each trip, i.e. the linkage between the two trips is retained.

If demand modelling were done on an O-D basis, the outbound and return would not be linked, so it would be possible for behaviour to change for one leg (say a change of mode or destination) without the corresponding change in the return trip. This is a significant drawback of the approach. Nor is it convenient to model the impact of conditions on the return leg, e.g. a one-way toll, on the choice of mode and destination for the outbound leg.

On the other hand all network assignment models require an OD-based representation of trip making, so in PA models conversion from PA to OD is required. Conversely, OD-based costs from the assignment need to be converted to PA for use in the demand model.

For the sake of illustration we start with a hierarchical model of mode and destination choice, with mode choice above destination choice (reflecting the recommended hierarchy in WebTAG).

The basic model is:

$$T_{ijm} = T_i p_{m|i} p_{j|im}$$

where

$T_{ijm}$ is the number of trips from production zone $i$ to attraction $j$ by mode $m$

$T_i$ is the total number of trips from production zone $i$ (perhaps derived from a separate trip generation model)

$p_{m|i}$ is the conditional probability of choosing mode $m$, given the trip is produced in zone $i$

$p_{j|im}$ is the conditional probability of choosing destination $j$, given the trip is produced in zone $i$ and uses mode $m$

The conditional probabilities are given by:

$$p_{j|im} = \frac{\exp(U_{ijm})}{\sum_k \exp(U_{ikm})}$$

and

$$p_{m|i} = \frac{\exp(U_{i*m})}{\sum_n \exp(U_{i*n})}$$

where $U$ represent the utilities of travel:

$$U_{ijm} = \lambda^{dest} c_{ijm}$$

$$U_{i*m} = \theta^{mode} \log \sum_k \exp(U_{ikm})$$

$\lambda^{dest}$ is a parameter (<0) to convert the generalised cost of travel $c_{ijm}$ (usually measured in minutes) to the utility scale for destination choice, which is defined by the error relating to this choice

$\theta^{mode}$ is a scaling parameter (>0 and ≤1) and $\theta^{mode} = \lambda^{mode} / \lambda^{dest}$, where $\lambda^{mode}$ converts generalised cost to the utility scale for mode choice

For simplicity of presentation the absolute, not incremental, version of the model is given above and we have not included segmentation by traveller type or trip purpose, though we expect such segmentation to be used in almost all applications. Similarly we have not explicitly included any alternative-specific constants or size variables and assume that they have been taken up in the cost $c$.

Note that neither trip numbers nor generalised cost includes a time dimension; both are assumed to be 24 hour quantities. This means that $T_{ijm}$ refers to total trips over the 24 hour period and that the costs are average 24 hour costs. Typically the latter may be a flow-weighted average of travel costs over the day, which includes an average over outbound and return directions of travel.

In the context of PA modelling when we talk about the time period in which a trip takes place, there are really two distinct time periods - the time period of the outbound trip and the time period of the return trip. We use the subscripts *r* and *s* to refer to the outbound and return time periods respectively.

If time period choice appears at the bottom of the hierarchy then it is possible to have separate models for the outbound and return time periods and to take the sum of the resulting logsums as the 24 hour utility to be passed to the 24 hour mode/destination model described above:

$$p_{r|ijm} = \frac{\exp(U_{ijmr})}{\sum_t \exp(U_{ijmt})}$$

$$p_{s|ijm} = \frac{\exp(U_{ijms})}{\sum_t \exp(U_{ijms})}$$

$$U_{ijm} = \theta^{dest}\left(\log\sum_r \exp U_{ijmr} + \log\sum_s \exp U_{jims}\right)$$

where $U_{ijmr} = \lambda^{time} c_{ijmr}$, $U_{ijms} = \lambda^{time} c_{ijms}$

$\theta^{dest} = \lambda^{dest} / \lambda^{time}$, $0 < \theta^{dest} \leq 1$

$\lambda^{time}$ is the scaling factor to convert generalised cost to the utility scale for time period choice.

Adding logsums to obtain the 'all day' utility $U_{ijm}$ is not completely consistent with the theory of utility maximisation that underlies the logit model. Nevertheless, it has previously been considered an acceptable approximation and has been used in applications such as the Dutch National Model (Daly et al, 1990).

Alternative approximations have been used elsewhere, for example where $U_{ijm}$ is based on the flow-weighted average outbound and return costs of travel. However, this is also inconsistent with utility theory.

The accuracy of these approximations is not explored any further in this paper. The important point is that they can only be applied when time period choice is at the bottom of the hierarchy.

However, with time period choice above destination choice (as recommended in WebTAG) it is more difficult. Assume that time period choice sits below mode and above destination choice in the hierarchy. As input to the time period choice model for outbound legs (for a given mode), we require a 'logsum' over all the potential destinations for that period. Similarly a logsum back from all the potential destinations is required as input to the time period choice for return legs. The problem with this is that the logsums each imply a distribution over destinations and there is no guarantee that these are the same. Only if the time period model considered outbound and return legs simultaneously could this problem be solved, otherwise we may be implying that travellers are returning from different destinations from those they went out to.

One solution is to move to tour-based modelling, which has been used in many large-scale demand models around the world. To start with we only consider simple tours, i.e. like PA-based modelling we assume the traveller goes from home to work then straight home again. The model is as follows

$$T_{ijmrs} = T_i p_{m|i} p_{rs|im} p_{j|imrs}$$

$$p_{j|imrs} = \frac{\exp(U_{ijmrs})}{\sum_k \exp(U_{ikmrs})}, U_{ijmrs} = \lambda^{dest} c_{ijmrs}$$

$$p_{rs|im} = \frac{\exp(U_{i*mrs})}{\sum_{rs} \exp(U_{i*mrs})}, U_{i*mrs} = \theta^{time} \log \sum_j \exp(U_{ijmrs}),\ \theta^{time} = \lambda^{time} / \lambda^{dest}$$

$$p_{m|i} = \frac{\exp(U_{i*m**})}{\sum_m \exp(U_{i*m**})}, U_{i*m**} = \theta^{mode} \log \sum_{rs} \exp(U_{i*mrs}),\ \theta^{mode} = \lambda^{mode} / \lambda^{time}$$

The 'tour' cost $c_{ijmrs}$ is the sum over the outbound OD cost in time period *r* and the return OD cost in time period *s* for *ijm*. Alternatively the average of the outbound and return costs can be used, provided a corresponding adjustment is made to $\lambda^{dest}$.

There is not a clear cut distinction between PA modelling and simple tour-based modelling, but for our purposes tour-based modelling explicitly links the outbound and return time periods of trips. In other words with tour modelling we know, for example, how many trips go out in the AM peak and return in PM peak. With PA modelling we know the totals going out in the AM and the totals returning in the PM, but without any linkage between the two. This can be represented as a matrix of the following form:

**Table 1.** Tour and PA modelling data requirements.

| | | Return time period | | | | |
|---|---|---|---|---|---|---|
| | | AM | IP | PM | OP | Total |
| Outbound time period | AM | | | | | |
| | IP | | | | | |
| | PM | | | | | |
| | OP | | | | | |
| | Total | | | | | |

Here there are four time periods: AM peak, inter-peak, PM peak and off peak. Tour modelling involves the individual matrix cells, PA modelling just the row and column totals (shaded).

This matrix shows an important disadvantage of the tour-based approach – the amount of data required. In the incremental version of the model it is necessary to know, for the base matrix, the number of trips in each cell of the above matrix, for each zone pair and mode (and any other segmentation such as traveller type or trip purpose) for each *ijm*.

It was considered that the data requirements for tour modelling would be too onerous for many users of the guidance and the associated DIADEM software. Typically the required information is only available from local household interview surveys/travel diaries which are undertaken relatively infrequently in the UK. Intercept surveys such as roadside interviews, which are the basis of many models in the UK, do not usually collect this data.

A method had to be found that combined the benefits of tour-based modelling with the lower data requirements of PA modelling.

## 3. SOLUTION

The solution turned out to be relatively straightforward. As suggested by Table 1 the problem is essentially one of estimating individual matrix cells, given the row and column totals. This is analogous to the problem of updating trip matrices when only the new row and column (trip end totals) are known. The furnessing procedure, also known as Iterative Proportional Fitting, is commonly used to do this (Furness, 1965). This process requires an initial estimate of the matrix which can be adjusted to meet the required trip ends. For the UK we are able to obtain our initial estimate from the National Travel Survey (NTS (DfT, ongoing)), which is an annual travel diary survey of over 8400 households. This provides detailed trip making information, including the outbound and return times of all trips made.

Rather than trip numbers we work with the proportions of all trips for each *ijm*.

More formally, the method can be described as:

$$\Omega_{ijmrs} = a_{ijmr} b_{ijms} \omega_{ijmrs}$$

where *a* and *b* are factors calculated using a furnessing procedure with the constraints:

$$\sum_s \Omega_{ijmrs} = Pout_{ijmr}$$

$$\sum_r \Omega_{ijmrs} = Pret_{ijms}$$

$\omega_{ijmrs}$ are prior estimates from NTS of the proportion of trips for *ijm* that go out in time period *r* and return in time period *s* ($\sum_{rs} \omega_{ijmrs} = 1$)

$\Omega_{ijmrs}$ are the best estimates of the proportion of trips for *ijm* that go out in time period *r* and return in time period *s*

$Pout_{ijmr}$ is the locally observed proportion of trips for *ijm* that go out in time period *r*

$Pret_{ijms}$ is the locally observed proportion of trips for *ijm* that return in time period *s*

The initial tour proportions $\omega_{ijmrs}$ include spatial dimensions *i* and *j* for completeness (in the notation and in the DIADEM software). In practice the values are likely to be the same for all *ij,* particularly where they have been calculated from national data sources.

In fact rather than working directly with NTS data we calculated $\omega$ using data from the DfT's National Trip End Model (NTEM) which itself is based on NTS data. The results for the three standard home-based trip purposes used in WebTAG are presented in the following tables:

**Table 2.** Initial tour proportions for home-based work.

| | | | Return time | | | | |
|---|---|---|---|---|---|---|---|
| | | | AM | IP | PM | OP | Total |
| Car | Outbound time | AM | 2.5% | 13.6% | 45.2% | 5.6% | **66.9%** |
| | | IP | 0.0% | 3.8% | 7.0% | 3.4% | **14.2%** |
| | | PM | 0.2% | 0.1% | 3.0% | 3.3% | **6.6%** |
| | | OP | 1.6% | 4.0% | 3.1% | 3.7% | **12.3%** |
| | | Total | **4.3%** | **21.5%** | **58.3%** | **16.0%** | **100.0%** |
| Bus | Outbound time | AM | 2.7% | 14.5% | 48.0% | 5.9% | **71.1%** |
| | | IP | 0.0% | 3.8% | 6.9% | 3.3% | **14.0%** |
| | | PM | 0.2% | 0.1% | 2.6% | 2.8% | **5.6%** |
| | | OP | 1.2% | 3.0% | 2.3% | 2.7% | **9.3%** |
| | | Total | **4.0%** | **21.4%** | **59.7%** | **14.8%** | **100.0%** |
| Rail | Outbound time | AM | 3.2% | 17.1% | 56.8% | 7.1% | **84.2%** |
| | | IP | 0.0% | 2.5% | 4.5% | 2.2% | **9.2%** |
| | | PM | 0.1% | 0.0% | 1.1% | 1.3% | **2.5%** |
| | | OP | 0.5% | 1.3% | 1.0% | 1.2% | **4.1%** |
| | | Total | **3.8%** | **20.9%** | **63.5%** | **11.7%** | **100.0%** |

**Table 3.** Initial tour proportions for home-based employer's business.

| | | | Return time | | | | |
|---|---|---|---|---|---|---|---|
| | | | AM | IP | PM | OP | Total |
| Car | Outbound time | AM | 3.5% | 14.3% | 25.2% | 5.5% | **48.5%** |
| | | IP | 0.0% | 14.9% | 13.5% | 3.5% | **31.8%** |
| | | PM | 0.0% | 0.0% | 4.1% | 6.0% | **10.1%** |
| | | OP | 0.8% | 0.6% | 0.7% | 7.4% | **9.5%** |
| | | Total | **4.2%** | **29.8%** | **43.5%** | **22.5%** | **100.0%** |
| Bus | Outbound time | AM | 3.2% | 13.2% | 23.2% | 5.1% | **44.7%** |
| | | IP | 0.0% | 16.1% | 14.5% | 3.7% | **34.4%** |
| | | PM | 0.0% | 0.0% | 4.3% | 6.3% | **10.6%** |
| | | OP | 0.8% | 0.6% | 0.8% | 8.0% | **10.3%** |
| | | Total | **4.0%** | **29.9%** | **42.8%** | **23.2%** | **100.0%** |
| Rail | Outbound time | AM | 3.5% | 14.4% | 25.3% | 5.5% | **48.7%** |
| | | IP | 0.0% | 16.5% | 14.9% | 3.8% | **35.3%** |
| | | PM | 0.0% | 0.0% | 3.1% | 4.6% | **7.8%** |
| | | OP | 0.7% | 0.5% | 0.6% | 6.4% | **8.2%** |
| | | Total | **4.2%** | **31.4%** | **44.0%** | **20.4%** | **100.0%** |

 

**Table 4.** Initial tour proportions for home-based other.

| | | | Return time | | | | |
|---|---|---|---|---|---|---|---|
| | | | **AM** | **IP** | **PM** | **OP** | **Total** |
| **Car** | **Outbound time** | **AM** | 5.9% | 12.2% | 3.5% | 0.3% | **21.9%** |
| | | **IP** | 0.0% | 28.4% | 10.4% | 1.5% | **40.3%** |
| | | **PM** | 0.0% | 0.0% | 10.2% | 7.4% | **17.7%** |
| | | **OP** | 0.3% | 0.3% | 0.1% | 19.5% | **20.1%** |
| | | **Total** | **6.2%** | **40.9%** | **24.2%** | **28.7%** | **100.0%** |
| **Bus** | **Outbound time** | **AM** | 9.3% | 22.0% | 7.7% | 0.2% | **39.2%** |
| | | **IP** | 0.0% | 34.8% | 11.9% | 1.5% | **48.3%** |
| | | **PM** | 0.0% | 0.0% | 3.6% | 2.5% | **6.1%** |
| | | **OP** | 0.1% | 0.1% | 0.0% | 6.2% | **6.4%** |
| | | **Total** | **9.4%** | **57.0%** | **23.2%** | **10.4%** | **100.0%** |
| **Rail** | **Outbound time** | **AM** | 6.5% | 15.1% | 5.3% | 0.2% | **27.0%** |
| | | **IP** | 0.0% | 36.3% | 13.0% | 1.9% | **51.2%** |
| | | **PM** | 0.0% | 0.1% | 6.0% | 4.2% | **10.3%** |
| | | **OP** | 0.2% | 0.2% | 0.1% | 11.1% | **11.5%** |
| | | **Total** | **6.6%** | **51.7%** | **24.3%** | **17.4%** | **100.0%** |

The results are mostly as expected with, the most common combination of timing for commuting being to go out in the AM peak and returning in the PM peak. A few slight anomalies appear in the data, for example trips returning in earlier time periods than they go out. This particular behaviour can be attributed to overnight trips.

## 4. MODELLING RESULTS

To demonstrate the difference between the OD and the modified PA modelling approaches a number of tests have been carried out on an existing transport model. The scenario tested was the same in each case and involved introducing a cordon-based road pricing scheme around a medium sized urban area. The charge only applied in the AM peak period and in the inbound direction. The test involved using both the OD and PA approaches, with and without time period choice, making a total of four model runs. Trip frequency and distribution (destination choice) were modelled in each case. All tests used the incremental version of the model.

Figures 1 and 2 show the impact of the charge on total vehicle flows crossing the cordon in each direction, by time period, for each of the four model runs. Figures 3 and 4 show the impact on the trip matrix in terms of OD trips with a destination within the cordon and those with an origin within the cordon.

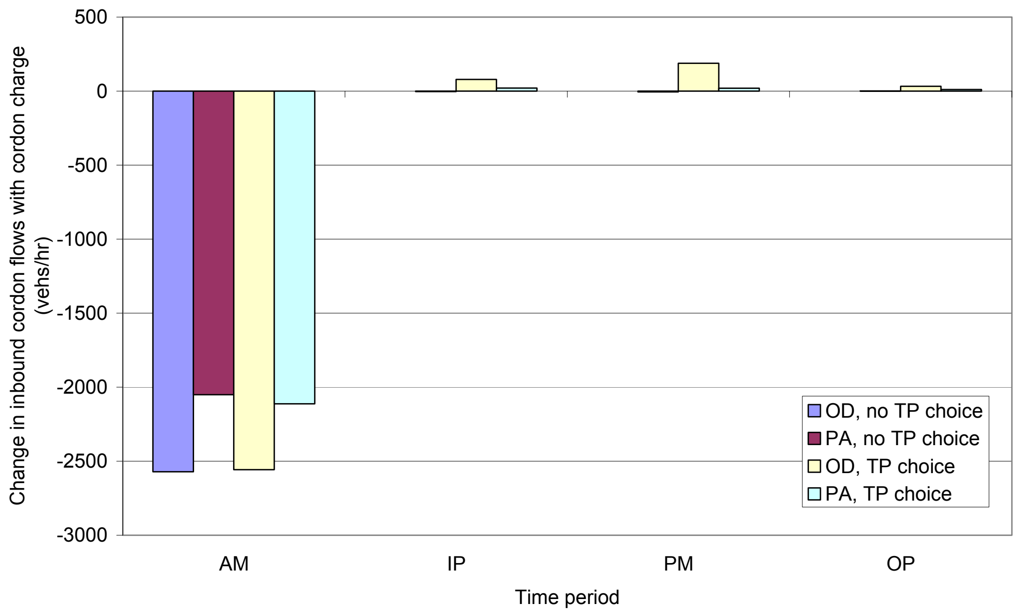


**Figure 1.** Change in inbound cordon flows.

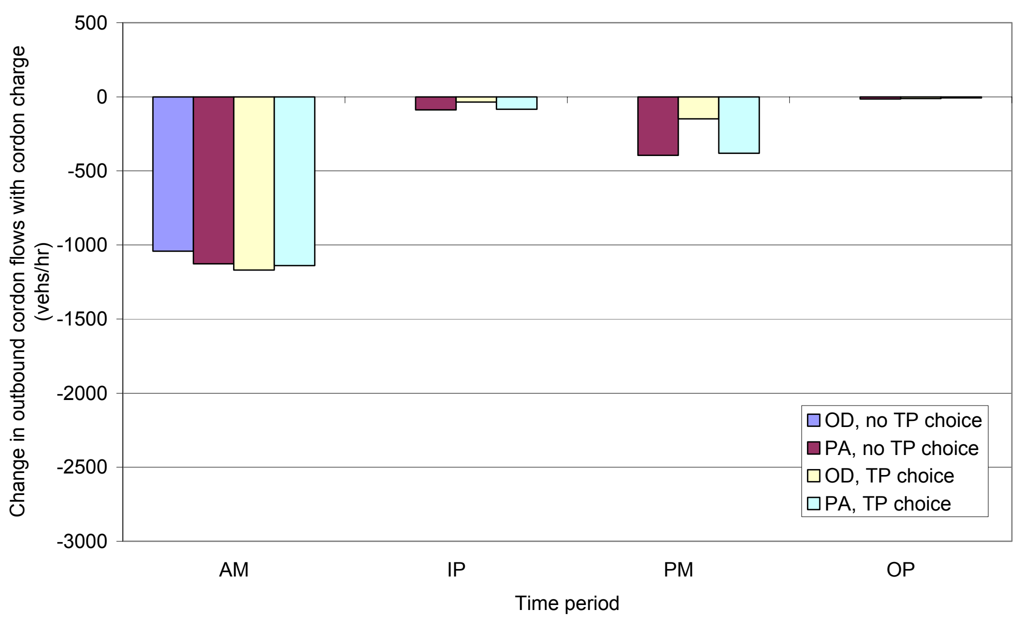


**Figure 2.** Change in outbound cordon flows.

 

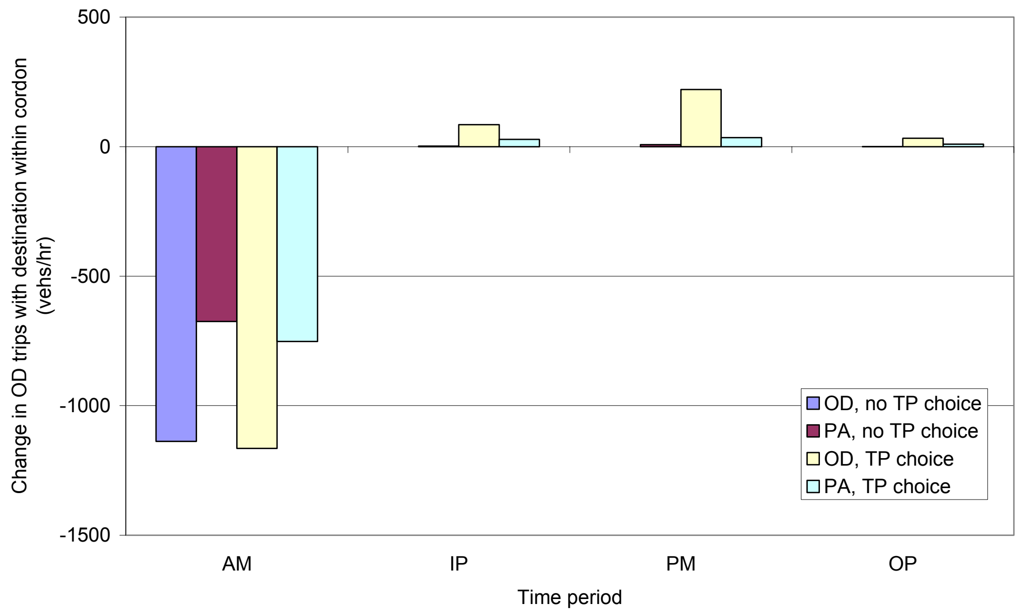


**Figure 3.** Change in trips with destination within cordon.

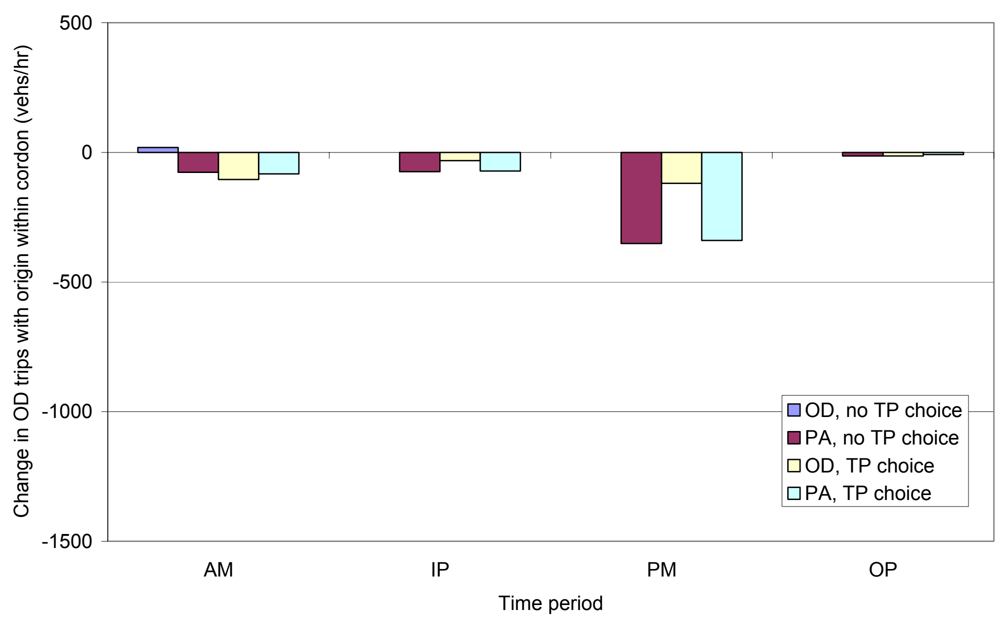


**Figure 4.** Change in trips with origin within cordon.

First we consider the results with no time period choice. As would be expected, for OD-based modelling the change in cost in the AM peak has no impact on other time periods. The modified PA modelling has a smaller impact in the AM peak. This is a result of using an average of outbound and return costs, so the toll (in one direction only) has less of an impact. However, the crucial difference is that the modified PA modelling shows an impact in other time periods, particularly a decrease in trips travelling outbound across the cordon. These are the return legs of tours which would have had to pay the cordon charge in the AM peak.

When time period choice is introduced there is the expected switch of inbound cordon trips away from the AM peak. This effect is larger with OD-based modelling than PA.

## 5. SUMMARY AND CONCLUSION

The UK Department for Transport's guidance on variable demand modelling presents technical difficulties for 'traditional' production-attraction modelling. These can be solved by the use of tour modelling. However, the data requirements of tour modelling are likely to be too onerous for many users of the guidance.

The work presented in this paper has shown how national data sources can be used in combination with local data to implement an approximation to full tour modelling. The *local* data requirements are no more onerous than for traditional PA modelling.

When tested on an AM peak tolling scheme the results of this method appear reasonable, and are significantly different from OD-based demand modelling which is still often used in the UK.

Having overcome these difficulties there should now be greater use of PA-based modelling, which is significantly superior to OD modelling in many situations.

The new method will be made available in next version of DIADEM software, which is developed and maintained on behalf of the UK Department for Transport.

## 6. REFERENCES

Daly, A., Gunn, H., Hungerink, G., Kroes, E. and Mijjer, P. (1990) Peak-Period Proportions in Large-Scale Modelling presented to *P.T.R.C. Summer Annual Meeting*.

DfT (ongoing). National Travel Survey.
http://www.dft.gov.uk/pgr/statistics/datatablespublications/personal/ Accessed 14 August 2007.

DfT (2006). Guidance on variable demand modelling. WebTAG units 2.9.1, 2.9.2, and 3.10.1-3.10.4. http://www.webtag.org.uk Accessed 14 August 2007.

Furness K. P. (1965). Time function iteration. **Traffic Engineering and Control 7**, 458-460.

SACTRA (The Standing Advisory Committee on Trunk Road Assessment) (1994). *Trunk roads and the generation of traffic*. HMSO, London.