The External Validation of the NTM

Hugh Gunn, HGA Ltd
Stephen Miller and Peter Burge, RAND Europe

1. Introduction

This paper describes an on-going project to assist in the development and use of the UK National Transport Model (NTM), under the Department for Transport (DfT). The overall project has two main objectives:

- to assess and report the forecasting performance of the NTM, based on external evidence, and
- to propose key improvements in the modelling process to enhance forecasting performance.

A decision has been taken to re-base the NTM distribution module on explicit disaggregate behavioural models, leading to an updated version which we shall refer to here as NTM+. The initial validation objective, testing the version of the NTM operational in 2002, has been extended to include work on the newer version, NTM+. The work was accordingly split into two separate stages. Stage 1, reported here, examines trip rate inputs for stability (these being common to both NTM and NTM+), and also the mode/distance/destination-type distribution module of the NTM.

'Backcasting' was judged the most valuable source of validation evidence of the general forecasting reliability (strengths and weaknesses) of the NTM. Accordingly, in the early stages of the project there was a search for appropriate historical data from which to generate the necessary inputs to the model, and from which to take observed measurements of key statistics to compare with model outputs.

Two backcast periods are under investigation. The closest to the estimation and base years is the year 1991. Evidence from this ten-year retrospective is of particular interest, since a ten-year prospective has been the first major use of the model, and it is expected that such a horizon will continue be of crucial planning interest for many policy objectives.

The furthest backcast that has been proposed is a twenty-five year retrospective, to 1975. In practice, much data exists for 1976, and a major source (the National Travel Survey)
The paper is set out in six sections. This initial section, Section 1 provides an introduction. Section 2 describes the expectations set for the model, against which its performance should be judged. Section 3 sets out the methodology adopted for the validation process. Section 4 describes the databases which have been assembled to support the Validation Process. Section 5 presents the results for an analysis of trip generation as backcast to 1991 and 1976 by the NTM, focussing on the trip purposes commute and social, and the results of backcasting mode and ‘destination’ models from NTM to 1991, once again focussing on commuter and social travel purposes. Finally, section 6 sets out some conclusions from the work so far.

2. Expectations for the Forecasting Model

2.1 Introduction

The starting point for the analysis is an understanding of how the National Transport Model is used, since this defines the scope of what should be covered within any validation exercise.

This Section provides an overview of the user requirements from the model along with details on how the model is typically used, and information on the model scope and its perceived strengths and weaknesses.

2.2 Use of the NTM

There are two main purposes for the model: providing forecasts of demand for travel, usually over the next 10 – 15 years but, in the case of the White Paper, 25 years; and looking at the potential impacts of different policies.

Until recently, forecasts were being produced for 2010, although there is interest in 2015, particularly for projects with a ten-year lead time. The model is calibrated to long-term elasticities and it is recognised that the NTM is not appropriate for short-term forecasting. The length of any forecast is measured relative to the base year and the advice is that a forecast of 5 years is the shortest sensible to produce, with 10 years being a more typical forecasting period.

There is an interest presently in producing trajectories from the NTM using the current observed figures and the future year point-forecasts, but the model, at present, does not lend itself to taking account of dynamic effects needed to calculate these. In terms of longer forecasts, the NTM has recently been used to forecast up to 2025; the general feeling is that the model is capable of producing forecasts 25 years from the base, as long...
as the analyst is prepared to make some assumptions for the inputs and to acknowledge the uncertainty that may exist in the forecasts.

Another way in which the model is used is illustrated in the Road Pricing Feasibility study. The Road Pricing Feasibility study looked at different charging schemes and the likely changes in traffic and congestion considering both the impact on roads and on the environment.

The forecasts that are produced and disseminated from the model tend to be at a very aggregate level.

### 2.3 Assumptions used in forecasts

There are a number of exogenous assumptions made in the model:

- population and employment is input from GAD projections;
- GDP is obtained from Treasury assumptions and feeds into the value of time;
- fuel efficiency is provided by the DfT Cleaner Fuels Vehicles division and reflects existing EU agreements for improvements in efficiency;
- car fuel prices: fuel duty is assumed to remain constant;
- regulated rail fares are assumed to increase at RPI +1; and
- bus fares are provided from the DfT TAS bus market model.

All prices are adjusted to the 1998 price base with some assumptions on changes in the RPI.

In conclusion there is a well-established ‘niche’ for the model, well-understood by the users; it is used to generate ‘scenario’-type demand consistent with input assumptions, and operates at a national level, with no expectations that its predictions will be relevant at a detailed geographic level.

### 3. The Validation Methodology Adopted

#### 3.1 Background

As stated above, the NTM validation project is targeted at the trip rate models and mode/distance band models for personal travel, and separate analyses have been devised for each set of models. The aim is to demonstrate the accuracy of the NTM models, and by analysis to identify how this can be improved. The approach is to some extent innovative (certainly in this context), and we will refer to it a ‘Disaggregate Validation Approach’ (DVA). It consists of a comparison between model predictions and actual
observed individual data from an historic national travel diary survey, in this case the NTS. Further validations are planned for the NTM+ under development, using flow data and count data in conjunction with the PASS3 assignments.

For the trip rate models, the DVA takes the incidence of trips made by each individual in the NTS survey for each purpose (0, 1, 2, 3 etc) and compares overall patterns with the predictions of a Poisson model in which the NTM supplies an initial estimate of average trip making. From this analysis, the overall scale of any discrepancy can be measured, and also related to possible explanatory variables.

For the mode/distance band models, the DVA is based on model predictions (from the NTM) for the probability that, in a random sample of choices made amongst multiple categories (mode, distance band), one category will be picked, and compares this with a validation sample of individual choices which record the choice which was actually picked (from the NTS). Both model and validation database generated probabilities which sum to unity; here the major result is an analysis of the different patterns of probabilities, once again relating these to potential explanatory variables.

The DVA works by building a likelihood function for the validation data set item by item, trip by trip, in which the predictions of the NTM model are taken as a core prediction, but are then modified by adding the effects of any available covariates in the validation data, each with an associated, unknown coefficient. Essentially, if the model is sufficient to predict the validation data, the co-efficients associated with the additional explanatory covariates should be statistically insignificant from zero (or unity, if they are scaling parameters), saying that they do not add to the explanation.

The background to the DVA is the work done on the transferability of disaggregate models, using an approach proposed by Ben-Akiva which itself is a multivariate, disaggregate extension of the calibration techniques suggested in MAU Note 149 (see Gunn et al 1985\(^1\)). Essentially, a disaggregate model fitted to one area, in one time period, was used as a ‘core’ prediction for a different area and time period, and then modifications to this core (scaling terms, or additional effects) were sought to judge the degree of transferability of the original model.

The DVA differs only to the extent that the ‘core’ predictions (of choice probabilities from the available set of distance band, mode, time of day options) come from the NTM, not from another disaggregate model.

The ‘scaling approach’: experiments carried out in the Netherlands were part of the basis for the development of the Dutch National Model (NTM), which for some fifteen years used the ‘core’ of the OVD models with scaling adjustments taken from a non-geo-coded national travel survey (Gunn and Pol, 1986\(^2\)).

Finally, since the focus of the validation study is on the trip rate and mode/destination models, the backcasts that have been performed have used the actual reported car-ownership patterns in the target years (1991 and 1976), along with reported population

---


levels and locations of trip attractors such as employment, and best-supported estimates of ‘Value-of-Time’. The results are described in separate Reports.

3.2 Trip Rate Analysis: 1991 and 1976

3.2.1 Overview

A first step in the analysis was the comparison of the back-dated 1991 NTEM inputs to the 1991 NTS sample, and the derivation of adjustment factors that bring the inputs in line with the size of the sample and the distribution of demographic groups.

In the analysis of NTM trip rates, these factors have been applied to the NTM outputs to ensure that they are comparable with the NTS data: after adjustment the NTM outputs should form a forecast of the trip making behaviour of the people surveyed in the NTS, over the duration of the survey (seven days). In short, if the model forecasts are ‘correct’ they should match the NTS observations sufficiently closely that differences can be explained in terms of an appropriate sampling model. An appropriate model in this case is the Poisson distribution, with NTS trip count observations (by flow category) being drawn from a Poisson distribution whose mean is the adjusted NTM trip rate forecast:

$$T_{seg}^{NTS} \sim Po\left(\alpha_{seg} \cdot T_{seg}^{NTM}\right)$$ (0.1)

In reality, and as acknowledged in the context of the validation project, the NTM forecasts are not expected to be ‘correct’ and so the trip rate analysis seeks to describe the differences between model outputs and observed behaviour, and identify any patterns of error, by fitting a Poisson regression of observed trip rates against model forecasts and categorical variables.

3.2.2 Coding of trip-rate categories

Table 1 presents the numbering of NTM flow categories based on purpose, person and household type and distance band.
The distribution of trips over distance bands is examined as part of the trip distribution validation, so this component of the flow identifier has been discarded for the trip rate analysis. Predicted flow volumes have been aggregated for each combination of origin zone and (distance-free) flow number.

There are certain pairs of origin zones that cannot be distinguished in the NTS data: zones 1 & 2, 4 & 6, and 5 & 7 are indistinguishable in 1991. In 1976, zone 3 is also indistinguishable from 1 & 2. Therefore, NTM predicted flows for these pairs of zones have been aggregated. For non-home-based (NHB) trips, origin zone cannot be determined at all; so predicted flows have been aggregated over all origin zones.

3.2.3 Applying adjustment factors

Adjustment factors were calculated by region, person type, and household type for adults and children. These have been applied to the NTM flow predictions to allow fair comparison with the NTS, but it has been necessary in some cases to apply factors to aggregate categories. For example, for home-based work purposes an adjustment factor is necessary for all people except full-time workers.

Aggregate adjustment factors are also needed where flow for a certain purpose is not segmented along a given dimension, and for origin zones that cannot be identified with a region. The aggregate factors are calculated by taking a weighted average of the factors.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Person type</th>
<th>SEG</th>
<th>1 adult / 0 car</th>
<th>1 adult / 1+ car</th>
<th>2+ ad / 0 car</th>
<th>2+ ad / 1 car</th>
<th>2+ ad / 2+ car</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB Work</td>
<td>Full time emp</td>
<td>High</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rest of pop’n</td>
<td></td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>HB EB</td>
<td>Full time emp</td>
<td>High</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rest of pop’n</td>
<td></td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>HB Educ</td>
<td>Child (0-15) Full time emp</td>
<td></td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other 16-64</td>
<td></td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pensioner</td>
<td></td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>HB PB / Shop</td>
<td>Child (0-15) Full time emp</td>
<td></td>
<td>61</td>
<td>62</td>
<td>63</td>
<td>64</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other 16-64</td>
<td></td>
<td>66</td>
<td>67</td>
<td>68</td>
<td>69</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pensioner</td>
<td></td>
<td>71</td>
<td>72</td>
<td>73</td>
<td>74</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>76</td>
<td>77</td>
<td>78</td>
<td>79</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>HB Rec / Visiting friends</td>
<td>Child (0-15) Full time emp</td>
<td></td>
<td>81</td>
<td>82</td>
<td>83</td>
<td>84</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other 16-64</td>
<td></td>
<td>86</td>
<td>87</td>
<td>88</td>
<td>89</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pensioner</td>
<td></td>
<td>91</td>
<td>92</td>
<td>93</td>
<td>94</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>96</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>HB Hols / Day trips</td>
<td>All persons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1301</td>
</tr>
<tr>
<td>NHB EB</td>
<td>All persons</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1321</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1341</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1361</td>
</tr>
<tr>
<td>NHBO</td>
<td>All persons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1381</td>
</tr>
</tbody>
</table>

Table 1: Pass1 Factors and flows for distance band 1 (band 2 = +200, band 3 = + 300, etc.)
for individual segments, in proportion to the population before adjustment – that is, in proportion to the population in the NTEM inputs.

3.2.4 Analysis of trip counts - introduction

The following output shows the results of a Poisson regression of observed trip counts against adjusted predictors and purposes.

Analysis: example output, illustration only

```
.poison observed p2-p8, exposure(adjp)
Iteration 0: log likelihood = -3393749.5
Iteration 1: log likelihood = -130817.53
Iteration 2: log likelihood = -77751.786
Iteration 3: log likelihood = -76818.465
Iteration 4: log likelihood = -76817.934
Iteration 5: log likelihood = -76817.934
Poisson regression Number of obs = 1169
LR chi2(7) = 86711.15
Prob > chi2 = 0.0000
Log likelihood = -76817.934 Pseudo R2 = 0.3608

------------------------------------------------------------------------------
observed | Coef. Std. Err. z P>|z| [95% Conf. Interval]
-------------+----------------------------------------------------------------
p2 | .2876131 .0065574 43.86 0.000 .2747609 .3004653
p3 | -.4169785 .0041708 -99.98 0.000 -.4251532 -.4088038
p4 | -.0935035 .0028765 -32.51 0.000 -.0991413 -.0878656
p5 | -.1306432 .0029211 -44.72 0.000 -.1363683 -.124918
p6 | .1155896 .0066524 17.38 0.000 .1025512 .128628
p7 | .1731527 .0057525 30.10 0.000 .1605512 .1857552
p8 | -.957914 .0042015 -227.99 0.000 -.9661488 -.9496792
_cons | .5224496 .0021927 238.27 0.000 .5181519 .5267472
adjp | (exposure)
------------------------------------------------------------------------------
```

Figure 1: Regression by purpose

These results can be interpreted with reference to the following model:

\[
T_{NTS}^{Seg} \sim Po\left(\alpha_{Seg} \cdot T_{NTM}^{Seg} \cdot \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n)\right)
\]

Here \(\alpha_{Seg} \cdot T_{NTM}^{Seg}\) is the adjusted NTM flow predictor for a particular flow segment, \(x_1, \ldots, x_n\) are the purpose indicator variables and \(\beta_0, \ldots, \beta_n\) are the coefficients in the report above. \(\beta_0\) is the constant term "cons" and the other coefficients p2 – p8 apply additionally to purposes 2 – 8 through indicator variables. Purpose 1 (HB Work) has no
separate indicator variable – observed trip counts for this purpose are explained by the constant term alone.

The coefficients can be used to calculate incidence rate ratios (IRRs) for each purpose by evaluating the exponential term in equation (0.2). As a check that the model is understood correctly, these can be compared against direct ratios of observed NTS trips vs. adjusted NTM predictors for each purpose:

An IRR greater than one implies that the model underestimates trip making for a particular purpose. The advantage of fitting a model and calculating IRRs is that multiple dimensions can be analysed simultaneously, as will be seen below.

3.3 Mode/Distribution Analysis

3.3.1 Introduction

This sub-section lays out the methodology to be used for the analysis of trip distribution in the NTM, and later with the NTM+. The NTM outputs will be validated against the NTS sample for the appropriate year, using a disaggregate approach to search for systematic biases in the NTM, and to measure the significance of additional variables that appear in the NTS but not in the NTM outputs.

- The NTS will be treated as an unbiased sample;
- The NTM outputs will be adjusted to match trip rates by population segment and purpose observed in the NTS (although, using the described methodology, there should be no effect on the trip distribution analysis);
- Each NTM travel purpose will be treated separately.
- Since geo-coded trip destination is not recorded in NTS for the year ‘91, only the distribution over modes and distance bands can be analysed.

The analysis will be performed by developing a disaggregate model of mode and distance band choice in which, for each trip observed in the NTS, the trip maker faces a choice between \(6 \times 12 = 72\) mode and distance band alternatives (6 modes, 12 distance bands)\(^3\). The ratio of (adjusted) NTM outputs will give a base choice probability for each alternative based on the traveller’s population segment, home zone (for home based travel purposes) and the purpose. Bias terms, to be fitted, will explain systematic differences between the base probabilities and observed behaviour.

3.3.2 Form of the model

The choice model for each purpose will predominantly be a multinomial logit (MNL) model, although it is possible that a nested logit (NL) model could be used to demonstrate

\(^3\) For 1975, it is necessary to aggregate two of the distance bands in order to ensure consistency between the NTM and NTS definitions; there will therefore be 11 distance bands and 66 alternatives.
different levels of substitution between sets of alternatives, e.g. between modes, between distance bands, or between groups of modes (slow modes, car modes, PT modes etc.)

In the case of an MNL mode $l$, the choice probabilities are modelled as follows. For a traveller in population segment $i$, the probability of choosing mode $j$ and distance band $k$ for a particular purpose is:

$$P_{jk} = \frac{\exp(V_{ijk})}{\sum_{j'k'} \exp(V_{j'k'})}$$

(0.3)

where $V_{ijk}$ is the utility associated with mode $j$ and distance band $k$ for that purpose. The base choice probabilities can be recovered by setting

$$V_{ijk} = \log_e (\rho_{ijk})$$

(0.4)

where $\rho_{ijk}$ is the flow volume predicted by the NTM for population segment $i$, mode $j$ and distance band $k$ for the purpose in question.

Setting the utility functions in this way creates a deterministic choice probability for each alternative. To identify bias, it is necessary to add additional terms to the utility functions, to be estimated (calibrated) against the NTS sample. A full utility function might take the form

$$V_{ijk} = \log (\rho_{ijk}) + \alpha_j + \beta_k + \gamma_{jk} + \alpha'_{ij} + \beta'_{ik} + \gamma'_{ijk}$$

(0.5)

Here, $\alpha_j$ would be a bias term for mode $j$, while $\alpha'_{ij}$ would be a bias term specific to population segment $i$. Similarly $\beta_k$ and $\beta'_{ik}$ would capture distance-band biases, and $\gamma_{jk}$ and $\gamma'_{ijk}$ are cross terms specific to a particular mode and distance band. Ultimately, we did not use cross-terms as the results became difficult to interpret.

Not all of these terms need occur in the model, and constraints may be imposed upon them. For example, there may be a single bias term for long rail journeys, encompassing more than one distance band.

The population segmentation $i$ may also cover variables that do not appear in the NTM output, such as gender, but that are recorded in the NTS. Thus it would be possible to identify gender-specific mode choice effects that are not currently modelled in the NTM (or at least, are not represented in the output).

3.3.3. Estimation of the model

The bias terms (parameters) will be fitted by the method of Maximum Likelihood Estimation (MLE). The parameter estimates are treated as statistics – with an associated sampling distribution – so as well as generating absolute values for the bias terms, MLE produces standard errors, which can be used to test whether the terms are significantly different from zero and from each other; and a covariance matrix which can reveal any major correlations between the parameter estimates.
Estimation is aided by the use of ‘apply tables’, in which the predictions of a partially developed model can be compared with observed behaviour, and tabulated against segmentation variables that do or do not appear in the model. Therefore, the presence (or absence) of a difference in behaviour in the NTS according to a variable such as age or gender can be identified easily, without the need to estimate a model using that variable. If an effect is identified, the model can then be estimated.

### 3.3.4 Variables of interest

NTM predictions are segmented according to the following variables, and it will be necessary to test for any systematic biases that can be associated with them:

- **Household type**: Five segments, incorporating car availability and number of adults;
- **Person type**: Specifically children, full time workers, other adults of working age, and adults aged 65+;
- **SEG of head of household**: Low, medium and high;
- **Origin zone (Pass1 zone)**: 17 zones. This can be identified in the NTS for home-based travel, although certain groups of zones must be aggregated. Identification may not be possible for NHB purposes.

In addition, the NTM population inputs are further segmented into:

- An additional three household types, making a total of eight;
- A total of eleven person types, incorporating gender and more detailed employment status for adults.

The following variables, which appear in the NTS (based on 1991) but not in the NTM inputs or outputs, may be of interest:

#### Household and local variables

- **Address type**: Detached house, semi-detached etc.;
- **Number of bicycles**;
- **Household income**;
- **Public transport accessibility**: E.g. walk time to bus stop, walk/bus time to railway station, access times to shops. (These accessibility variables appear at the Pass1 zone level but not at a local level in the NTM inputs);
- **Presence of an OAP bus scheme**;

#### Individual variables

- **Relationship to head of household / marital status**;
- **Age** (more precise bands);
Journey variables

- Start time;
- Day of the week.

3.4 Conclusions

The focus of the validation is, of course, guided by the intended use of the model output, and for this the conclusions of the previous Chapter, ‘Expectations for Performance’ are crucial. The major tests should be on the broad picture – kilometres of travel by mode and purpose for private travel, and kilometres of travel by vehicle class for freight. Within that, however, the distribution of travel between traveller-types is important for evaluation purposes, as are trip-length distributions as indicators of road-types used. In particular, the reproduction of patterns of behaviour by different income classes is an important starting-point for any pricing policies. The model does not claim detailed geographic focus, but does attempt to generate representative totals for major area types (eg urban areas within certain size categories). This too will be checked.

4. Available Databases for Validation by Backcasting

4.1 Introduction

There are a number of data sources that are available for validation of the model backcasts, this chapter discusses some of the sources that have been identified. Not all have been used in the validation of the NTM, given the transfer of resources to the validation of the NTM+.
4.2 National Travel Survey

The NTS is the primary transport statistics data source for the UK; this started as a periodic survey in 1965, moving to a continuous survey in 1988. This is a dataset that most transport professionals are familiar with and as such will only be briefly discussed.

When the survey was commissioned its main objectives were:

- to estimate the distribution of car ownership and the variation in car utilisation, and their dependence on demographic, socio-economic and other factors;
- to determine personal and household travel generation rates and the relationship between these rates and a wide range of demographic, socio-economic and other variables;
- to provide data affording an examination of the modal split for journeys of different types, to determine in what ways and what circumstances public transport is competitive with the private sector; and
- to provide information to fill gaps in national transport data derived from other sources.

The NTS data set is particularly useful as there is data available for the years 1975/6 and 1988-2000. An issue that would require consideration in the use of validation of the NTM is the independence of the data used as model inputs and the data used to validate the outputs.

Data collected

The survey collects data on household travel behaviour on a national basis. Separate tables are provided for data at the household, individual, journey, journey stage and vehicle level; these have common identifiers to allow cross-referencing between the various tables.

Data collected on trips includes the mode of transport used, the journey purpose, the number of journeys made, the time of day, distance and length of time of journeys, and the travel costs. In terms of socio-demographic data, the survey collects age, gender, occupation, socio-economic status, driving licence holding and car ownership and availability.

Data documentation

The NTS is very well documented and there are a wide number of users that can be consulted on the strengths and weaknesses of the data.

4.3 Other data sources

4.3.1 County Surveyors Society Trip Rate Data Bank
The CSSTRDB was a repeated cross sectional survey, undertaken over a period of six years spanning 1974-1981; as a result it provides a useful additional source of data for the 1975 backcast year. The primary aim of this survey was to obtain a trip rate data bank across a range of counties that could then be used as the basis for research and analysis of trip making.

**Method of data collection**

The data was collected through face-to-face interviews and included travel diaries for all trips made over the course of one day for each member of the households sampled. Each of the counties participating chose between 600 and 1000 households for inclusion in the sampling frame, of which 200 were interviewed in two clusters during each year of the survey.

**Data collected**

From the surveys a data bank was compiled which records the trips made by each member of the household. Each trip is broken down into stages with records of the start time, finish time, origin, destination, type of destination, mode of travel, and purpose of journey. In addition there is a record of the number of household members participating in each trip.

In terms of socio-demographic data, the survey records the age, gender, occupation, household status, employment status and driver licence holding for each individual. There is also a record of the number of cars, motorcycle and scooters available to those in the household.

### 4.3.2 Long Distance Travel Survey

The LDTS was a survey of long distance travel sponsored jointly by British Rail and the Department of Transport. As with the CSSTRDB, this was a repeated cross-sectional survey undertaken on an annual basis between 1973 and 1979.

There are a number of notable differences between the CSSTRDB and the LDTS. The CSSTRDB collected data on all trips, including walking trips, whereas the LDTS focused only on journeys greater than 25 miles. In addition, the CSSTRDB collected data from a number of different area types with various criteria applied for density and age of housing stock, whereas the LDTS had a slightly narrower focus restricting it to the major conurbations. This restriction is, however, one of the LDTS strengths; it provides rich data on the travel patterns between regions.

**Method of data collection**

The data was collected through postal surveys, with a target to collect 30,000 records. Each household provided travel diaries detailing their long distance travel behaviour.

**Data collected**

The survey collected data on the number of long distance journeys made, their origin and destination, the mode of transport used and reason for choice of mode, the purpose of the journey, and information on car sharing. In addition a number of attitudinal and behavioural questions were collected.

In terms of socio-demographics, the survey collected the traveller’s age, gender and occupation, along with information on car ownership and car availability.
4.3.3 RHTM – a brief history

The Regional Highways Traffic Model project was one of the largest and most expensive modelling projects ever carried out, with project costs in excess of six million pounds in 1978 currency. Its principal aim was to update the then current DTp (DfT) “procedures for estimating traffic flows for existing and proposed Trunk Roads”\(^4\). Discounting the pre-planning and post-project analyses phases, the project took place between 1975 and 1977, assembling a database for the year 1976.

Crucially for our purposes, it focussed solely on motorised road traffic, cars, vans, light and medium goods vehicles and heavy goods vehicles. The targeted flows were 24-hour, average working-week-day for 1976.

This database covered England, Wales and Scotland in some 3,613 zones, and a substantial amount of the work went into establishing the basic building blocks of a model - a digitised zoning system based on postcodes, corresponding planning data and networks.

Traffic data was collected from

- Household interviews
- Roadside interviews
- Manual Classified counts
- Automatic Traffic Counts.

For the purposes of the Backcasting project, and in a later stage of the project, we are proposing to make use mainly of the cordon and screenline data built up from the roadside interviews, augmented by the manual and classified counts (which were used for expansion, and completion of cordon leaks).

4.3.4 Data availability: summary

The study team acquired the data sets for the National Travel Survey, the County Surveyors Society Trip Rate Data Bank and the Long Distance Travel Survey from the UK Data Archive. The RHTM cordon-crossing data has been retrieved from original documentation of the study that had been archived by members of the study team. The data discussed in this section is therefore available for analysis as and when it is required within the validation project.

In the analyses reported below, only the NTS will be used. This follows changes in the plans for the future development of the NTM, which include additional backcasting on a new model system.

5. Trip Rate Analysis 1991 and 1976, Distance and Mode Splits for 1991

5.1 Commute trips: trip rates for 1976 and 1991

Figure 1 summarises the results of the back-projection for 1976 to the London area. Looking first at the relationship of the different worker groups’ incidence rate ratios (IRR’s) to the 1.0 line (model predicting accurately) some major trends emerge. Full-time workers’ commute trip rates are under-predicted in four cases out of five, whilst others’ (part-time and casual labour) is consistently over-predicted. Amongst the full-time workers, it is the non-car-owning households which are worst underpredicted – by over 25%.

![Figure 1: Commute trip rates 1976: IRRs by population segment, zones 1-3](image)

Figure 2 extends this across the country, at an aggregate level. Broadly, trip rate forecasts are consistent with the picture for London, but relative to London there is over-prediction in five areas, by around 10%, off-setting the London under-prediction.

We can conclude that the trip-rate modelling could be improved in respect of the influence of car-ownership, and that non-full-time-workers commuting behaviour has changed (possibly due to type of job, and influences on the prevalence of shorter daily working hours over fewer working days, but this of course needs investigation).

In passing, it is worth noting that LMS1, the original Dutch National Model, had a module specifically changing commute trip (tour) rates according to input assumptions on the hours of the working week and the vacation allowance.
Figure 2: Commute trip rates 1976: IRRs by zone, based against zones 1-3

Figure 3 presents comparable information for 1991 (more segmentation is possible in that year). The general trend of over-prediction in the London area is evident for all worker groups other than full-time, high SEG workers, whose commute trip rates are under-predicted. This compares with over-prediction in 1976 for non-full-time workers and under-prediction for full-time workers without cars. In other words, there is no completely consistent picture emerging as to the aspects which could improve the model’s temporal stability, but there is evidence that there is important variation in the commute trip-rate data.

Figure 4 extends the results across the country, and the general picture is that, relative to London (in which the model over-predicts most commuter groups trip rates), the model is providing lower
forecasts in other parts of the country (IRRs greater than one, observed greater than prediction). Thus the effect seems to be worst in London, and actually worst in outer London.

5.2 Commute trips: mode and distance class predictions for 1991

Figure 5 presents a set of scatterplots derived from the 1991 NTM. The points on the various plots represent the results of analysing the model predictions for a given set of sub-categories, and the NTS observed results. The sub-categorisations vary from plot to plot, as do the models examined.

Briefly, the simplest plot is of the '000' model, essentially the unadjusted NTM, and its performance is represented by checks on its ability to match four different categorisation of trip. Firstly, it is checked for reproduction of the trips in the broad categories of mode and distance class (top left plot, "Overall"), the for ability to match a categorisation by mode, distance and person type (top right), then for performance in predicting trips in mode, distance and household type categories (bottom left), and lastly in predicting a mode, distance and origin zone categorisation of trips.
Figure 5: 1991: Scatter plots for Commute model 000: Unadjusted NTM

Fit (observed vs predicted trips) by mode/distance band combination

<table>
<thead>
<tr>
<th>Overall: $\Sigma z^2 = 320.2$</th>
<th>by Person Type: $\Sigma z^2 = 23534.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Household Type: $\Sigma z^2 = 1238.0$</td>
<td>By Pass1 Origin Zone: $\Sigma z^2 = 8507.0$</td>
</tr>
</tbody>
</table>

© Association for European Transport and contributors 2006
Clearly, the model is working very well indeed at an overall level, and when checked at household level. There is evidence of more variation at the geographic ‘zone’ level, and even more when we look how the trips are distributed between person types.

Figure 6: 1991: Scatter plots for Commute model 001: Factors fitted to 1991 Mode and Distance data

<table>
<thead>
<tr>
<th>Fit (observed vs predicted trips) by mode/distance band combination</th>
<th>Overall: Σz² = 266.9</th>
<th>by Person Type: Σz² = 23858.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Household Type: Σz² = 1201.6</td>
<td>By Pass1 Origin Zone: Σz² = 8548.2</td>
<td></td>
</tr>
</tbody>
</table>
The comparison of Figures 5 and 6 reveals that the fitted log-linear model introducing only mode and distance band parameters does little to improve the overall fit (despite having mainly statistically significant parameters).

Figure 7: 1991: Scatter plots for Commute model 015: mode, distance, person, household, origin parameters in model

<table>
<thead>
<tr>
<th>Fit (observed vs predicted trips) by mode/distance band combination</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall:</strong> $\sum z^2 = 665.9$</td>
</tr>
<tr>
<td><strong>by Person Type:</strong> $\sum z^2 = 10071.2$</td>
</tr>
<tr>
<td><strong>By Household Type:</strong> $\sum z^2 = 2438.9$</td>
</tr>
<tr>
<td><strong>By Pass1 Origin Zone:</strong> $\sum z^2 = 3203.3$</td>
</tr>
</tbody>
</table>

© Association for European Transport and contributors 2006
Figure 7, based on one of the most extensive models tested so far, shows that the data supports the need for adjustments to include parameters for person, household and origin-type of the trip.

In the sections above, we have noted the large range of other background variables that are available for checking for an influence on the model performance (age, sex, income, etc). The exploration of the 1991 trip distribution data can be extended greatly by the use of such variables, and the statistical significance of the results derived from the usual tests for log-linear models.

5.3 Commute trips: distance distribution predictions for 1991

Figure 8 displays information on the commute distance-band distribution for a number of sources:

- The unadjusted NTM model (run conditional on historic car-ownership, population statistics etc, corrected for NTS trip rates)
- The model adjusted for differences between the NTS and ode/destination NTM patterns, using simple scaling models
- The NTS 1991 data, raw results
- The 1998 NTS raw results.

95% confidence bounds for the adjusted model are also displayed.

The overwhelming impression of this figure is that, at this level of aggregation, the model and the data are virtually indistinguishable. It can then be seen that the data in 1991 is virtually identical to the 1998 model – in other words, at a high level of aggregation, trip distributions have not changed to any significant extent – after adjusting the model to ensure car-ownership levels, population distribution etc are in accord with the NTS database.
To some extent, this could have been expected. Given the controls in place, the influences we are trying to determine are those of cost and travel time, with other factors controlled to focus directly on the mode-destination component. Travel times were slightly longer in 1991 (+5% for longer distance car trips), costs slightly lower (-9% for car). That the net result should be negligible should not be surprising. That said, if the test of the NTM is against its ability to operate at an aggregate level, the distance band choice aspects have certainly done that in the backcast – albeit where very little difference in stimulus factors (cost and time) existed, and such that there were, were offsetting.

Figure 9 breaks down the distance distribution by gender. This is only one of the many analyses that have been done, but illustrates a point: the aggregate performance of the model excellent as it is, results from an averaging of quite significant variation in the population. Of course, since the NTM was never designed to have such a focus, it should not be taken as a criticism of that model to perform the task required of it – only an indication that the evident aggregate stability derives from an averaging of many different divergent patterns.

That said, and even given the fact that the changes in stimuli between 1998 and 1991 were very small, figure 8 must be taken to validate the mode-destination component of the NTM as having been a reliable (back)predictor of travel patterns, at an aggregate level.
We see from figure 9 that gender is clearly an influential variable: we look next at geographic variations.

Figure 10 displays the same information for London. The trip length distributions, predicted and observed, have the same ‘sort’ of shape (decreasing with distance), but there are significant differences in the ranges 0 – 2 miles and 5 – 10 miles.
Figure 11 presents the same information for “Urban Medium” zones: once again, a mismatch in the short trips (underestimated up to 5 miles) and an overestimate of middle distance trips (in this case, 10 – 15 miles). These variations offset the imbalances in the London data. It is from such offsetting variations that the aggregate stability emerges.

![Figure 11: Commute model 005 – Zone 16 (Urban Medium)](image)

5.4 Commute trips: mode class predictions for 1991

Figure 12 displays the overall mode choice proportions. Once again, we see excellent matching to a distribution little changed from 1998, the year to which the model was calibrated.
Figure 13 breaks this down by gender, and here we start to see some differences. The NTS data testifies to differences in mode usage between the sexes (connected to types of job and consequent trip length) which show increased female usage of the ‘slow’ modes and bus, and less car-driver trips (both distance related effects).

Turning next to the geographic stability of the model, figure 14 shows the match for the London area (central and inner). The biggest problems are for the slow modes and bus, with the model over-estimating slow modes and under-estimating bus.
Figure 15 shows the same data for the “Urban Medium” zones: there are variations, but it has to be said that the match is excellent.

The plots shown above are only a small sub-sample of the information that the project has generated, but they serve to illustrate the methods and some major results.
5.5 Social trips: trip rates for 1976 and 1991

In the 1976 commuter data, and focussing on the London area, we saw the model under-predicting trips for full-time workers and over-predicting for other workers. In the records relating to social trips, everyone, regardless of household car-ownership and size, is over-predicted by the model. Non-employed adults are generally the worst over-predicted. Overall, the over-prediction is by around 30%.

Figure 16: Social trip rates 1976: IRRs by population segment, zones 1-3

Figure 17 generalises the results to the remainder of the country. Relative to London, which was over-predicted, most areas have to be scaled down by around 15%, leaving them over-predicted still. The North-East metropolitan areas are over-predicted even relative to London.
In Figure 18 we see the London results for 1991. The pattern here is also for consistent over-prediction, by more than 30%, in social trip making. Pensioners are now the worst predicted group, although only slightly worse than adult non-workers.

Figure 19 gives the scaling factors for the rest of the country. The over-prediction in Inner and central London does not persist, or at least not to anything like the same level. Crucially, the local variations across the country are substantial.

© Association for European Transport and contributors 2006
5.5 Social trips: mode and distance class predictions for 1991

As with the Commute trips, scatterplots have been produced for other purposes, and here we display the results for the social trips.

Figure 20 sets out the scatterplots for the social purpose, comparable to the commute displays above. The correspondence is clearly excellent, and also very acceptable in respect of person type and origin-zone categories. Given that the unadjusted model is showing such good correspondence, it is unremarkable that fitting extra parameters to correct to the 1991 data results also in excellent fit in each of the later models. Statistically significant effects were identified, but in practical terms, at an overall level they add little.
Figure 20: Scatter plots for Social model 000

Fit (observed vs predicted trips) by mode/distance band combination

<table>
<thead>
<tr>
<th>Overall: $\Sigma z^2 = 391.9$</th>
<th>by Person Type: $\Sigma z^2 = 10877.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Household Type: $\Sigma z^2 = 1065.1$</td>
<td>By Pass1 Origin Zone: $\Sigma z^2 = 3660.5$</td>
</tr>
</tbody>
</table>
Figure 21: Scatter plots for Social model 001

Fit (observed vs predicted trips) by mode/distance band combination

Overall: $\Sigma z^2 = 135.9$

by Person Type: $\Sigma z^2 = 9946.4$

By Household Type: $\Sigma z^2 = 807.1$

By Pass1 Origin Zone: $\Sigma z^2 = 3251.4$
Figure 22: Scatter plots for Social model 015

Fit (observed vs predicted trips) by mode/distance band combination

<table>
<thead>
<tr>
<th>Overall: $\Sigma z^2 = 180.9$</th>
<th>by Person Type: $\Sigma z^2 = 5156.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Household Type: $\Sigma z^2 = 1091.4$</td>
<td>By Pass1 Origin Zone: $\Sigma z^2 = 1122.0$</td>
</tr>
</tbody>
</table>
5.6 Social trips: distance distribution predictions for 1991

Figure 23, comparable to Figure 8 for Commute trips, shows an amazing correspondence between the 1998 model (fitted to NTS), the 1991 data and the 1991 NTM model results. At an aggregate level, then, little or nothing has changed, and the model reflects this.

![Figure 23: Social model 001](image)

Figure 24 examines the data and the models for gender biases: it is clear that, unlike the Commute trips, any differences in the trip-length distribution of male and female travelers is trivial.
In Figure 25, the Central and Inner London trip-length distributions are displayed, models against data.

The correspondence here is not quite as identical as for the country as a whole, but is still impressive. There are no clear biases to be seen.
In Figure 26, we see the same display for the Urban Medium zones. This is a less-close correspondence, although still clearly a good fit. Shorter distance trips are somewhat under-estimated by the model, with the medium-distance slightly over-estimated.

In general, the evidence of this section is that the NTM, although fitted at an aggregate level, is not introducing serious errors in respect of either gender or location differences.

5.7 Social trips: mode class predictions for 1991

Figure 27 displays the unadjusted NTM model results for 1991, showing many of the features of the Commute analysis. The 1991 forecasts vary little from the 1998 model, fitted to 1998 data. There are no major anomalies in the fit of the backcast to the NTS data. At the aggregate level, the predictions of mode split are highly acceptable for practical purposes.
Figure 28 breaks the forecast down by gender, where both model and data testify to differences in mode usage. By and large, males are more likely to be using the car-driver mode and less likely to be passengers, than females: the model reflects this adequately.

Turning to geographical variations, Figure 29 suggests a bias of the model towards car usage in London, away from slow modes and public transport. Overall though, the model fit is reasonable.
Figure 30 looks at another type of origin zone, Urban Medium, as discussed above in respect of commuting trips. Here again, minor variations can be seen (mainly between car-driver and car-passenger) but the overall fit is good.
6. Conclusions

Final conclusions from this first stage of the Validation project are currently in preparation; in the meantime our major conclusions from the analyses reported above, and others not reported here, are that

- The methodology developed here is potentially extremely useful. There is a large resource in terms of historic data available to both validate new models, and ultimately to allow the incorporation of past reported behaviour into model estimation, identifying time-varying effects which cannot be captured by cross-sectional data.

- The NTM model, in respect of mode and destination choice, has performed well in reproducing national patterns of behaviour back to 1991.

- This conclusion must be qualified by the observation that the model has predicted little change in a situation in which there has been little change in the factors affecting mode and destination choice at an aggregate level. In other words, the evidence for ability to respond to larger changes (such as can occur at local levels, or through new policies such as road-pricing) is weak. Note also that errors in earlier parts of the modelling have been deliberately corrected to focus on the choice of mode and destination alone.

- The NTM mode and destination module cannot match the detailed person-type and location-specific variations in the NTS data. This seems to have been anticipated by DfT, and accepted as a consequence of the basic choice of modelling approach. More flexible approaches will be needed if the model is to address policies which have differential effects between sub-groups (e.g. gender-specific, or income-class specific impacts), or to have outputs which are sufficiently geographically specific to identify local congestion problems.

- The NTM trip-generation estimates are much less in accord with the evidence of the historic NTS data, and in particular in the case of the London area. Person-type, household-type and location-type variations have been found, suggesting that the trip-end estimation process should be extended to examine temporal and geographic variation not currently in the models.

- This study has not been conducted as an Audit, simply as a check of model output against historic data. The NTM model is currently being upgraded. We would recommend that the new version is subjected to an Audit at which the model structure and input assumptions can be scrutinised.

- This study has looked only at the trip-end and mode/destination aspects of the model. Other components, in particular the car-ownership and licence holding models, can be examined against historic data using methods similar to the ones developed here, and this would be useful.

- The entirety of the model, that is its actual forecasting accuracy as judged in back projection, could also be tested against historic data.
Disclaimer: the views expressed in this paper are those of the authors, and do not necessarily represent the opinions of the UK Department for Transport.

Further information can be obtained from the authors

- Hugh Gunn – hugh.gunn@hpgholding.nl
- Stephen Miller – smiller@rand.org
- Peter Burge – burge@rand.org