ANALYSIS OF THE UNCERTAINTIES IN DAY-TO-DAY DYNAMIC MODELS FROM OBSERVED CHOICE RESPONSES

Ronghui Liu
Institute for Transport Studies, University of Leeds
Leeds LS2 9JT
United Kingdom

1. INTRODUCTION

Recent advances in traffic network modelling have seen the development of day-to-day models that explicitly incorporate drivers’ dynamic/behavioural responses (on routes, departure-time, mode, etc) to network changes and policy instruments (such as responsive signal control, dynamic route guidance). These models have a far greater number of adjustable parameters than traditional equilibrium network models, making it far more important and difficult to validate their behaviour components. This paper examines drivers’ day-to-day route choice behaviour through analysis of hyper-route data reconstructed from partial number-plate survey. A hyper-route is defined as a series of time-stamped locations passed on an individual journey. Whilst practically difficult to collect data for origin-destination routes, hyper-routes can be used to study the drivers’ route-choice behaviour on part of their journey (in the study area) through which insights on drivers’ behavioural responses can be gained and compared with the equilibrium hypothesis of route choice.

An extensive registration plate survey at a number of locations, both on-street and in parking areas, have been collected in the city of York, England. These observations were made on a number of days, over a number of weeks, both before and after a planned capacity change (e.g. road work) to the network. The reason for observing a number of days and weeks is in order to separate ambient daily variability from a systematic response to the capacity change. Hyper-routes are reconstructed by the matching of these registration plates observed at a series of locations taking into account of journey times of the moving vehicles.

The matching of partial-registration plates at two locations is a commonly-used technique in a variety of traffic studies, most notably in obtaining information on the origin to destination movements of vehicles through a traffic network (Ortuzar and Willumsen, 1990). A number of statistically based methods have been proposed to reduce “spurious matches” in which different vehicles with the same partial registration number observed at two points are paired together (e.g. Makowski and Sinha, 1976; Maher, 1985). Watling and Maher (1992) developed an algorithm which combined journey time data (the times at which vehicles pass the observation points) into the matching process. The method made significant improvement in reducing spurious matches and was shown to be fairly efficient computationally. This paper extends Watling and Maher’s method to three observation points which yields a method to construct hyper-route data.

This paper presents the statistical method and solution algorithm, demonstrates how this estimation procedure can be carried in an efficient simulation software and provides example results of running the procedure in matching the observed data in the York survey.

2. A STATISTICAL MODEL FOR MATCHING THREE OBSERVATION POINTS
2.1 Three observation station survey

We first consider the arrangement displayed in Fig. 1: of the vehicles passing one station, some leave system before passing any subsequent station(s); there are additional streams of vehicles (with mean flows $\lambda_1$ and $\lambda_2$) entering the system between stations 1 and 2, and between stations 2 and 3 respectively. The data collected at each station $s$ over a period $[a_s, a_s+T_s]$ consists of:

- $n_{sk}$, the number of vehicles of partial plate $k$, observed at station $s$;
- $t_{ik}$, the time vehicle $i$ of partial plate $k$ passed point station 1;
- $s_{jk}$, the time vehicle $j$ of partial plate $k$ passed point station 2; and
- $r_{lk}$, the time vehicle $k$ of partial plate $k$ passed point station 3; for $(k=1,2,...,K; s=1,2,3; i=1,2,...,n_{1k}; j=1,2,...,n_{2k}; l=1,2,...,n_{3k})$

where $K$ is the maximum number of different partial plates observed, $a_s$ the beginning and $T_s$ the duration of the survey at station $s$.

![Figure 1. A system with three observation points.](image1)

![Figure 2. The modelled system, with the six paths indicated as P1, P2, ..., P6.](image2)

2.2 The statistical matching method

The survey of Fig. 1 can be modelled as a system shown in Fig. 2. It is assumed that:
• vehicles entering the system between stations 2 and 3 follows path 1 (P1);
• of the vehicles entering the system between stations 1 and 2 and reached station 2, a proportion \( \beta \) later pass station 3 (P2) whilst a proportion \((1-\beta)\) left the system after station 2 (P3); and
• of the vehicles passing station 1, a proportion \( \alpha_x \) later pass both stations 2 and 3 (P4), a proportion \( \alpha_y \) later pass station 2 then leave the system before reaching station 3 (P5), the remaining proportion \((1-\alpha_x-\alpha_y)\) leave the system before station 2 (P6).

We now need to have some indicator variables to represent the genuine matches for various routes. We define indicator variables \( x_{ijk}, y_{ijk} \) and \( w_{jlk} \) to be 1 if a vehicle of type \( k \) was observed to follow route 4, 5 and 2 respectively, and 0 otherwise. The collections of the relevant variables can then be written as:

\[
X_k = \{x_{ijk} : i=1,2,\ldots,n_{1k}; j=1,2,\ldots,n_{2k}; l=1,2,\ldots,n_{3k}; k=1,2,\ldots,K\}
\]

\[
Y_k = \{y_{ijk} : i=1,2,\ldots,n_{1k}; j=1,2,\ldots,n_{2k}; k=1,2,\ldots,K\}
\]

\[
W_k = \{w_{jlk} : j=1,2,\ldots,n_{2k}; l=1,2,\ldots,n_{3k}; k=1,2,\ldots,K\}
\]

We use \( x_{.,.k} \) to denote the total of matches for \( X_k \), where a dot denotes summing over the subscript:

\[
x_{.,.k} = \sum_{i=1}^{n_{1k}} \sum_{j=1}^{n_{2k}} \sum_{l=1}^{n_{3k}} x_{ijk}.
\]

Similarly \( y_{.,.k} = \sum_{i=1}^{n_{1k}} \sum_{j=1}^{n_{2k}} y_{ijk} \) and \( w_{.,.k} = \sum_{j=1}^{n_{2k}} \sum_{l=1}^{n_{3k}} w_{jlk} \) denote the total of matches for \( Y_k \) and \( W_k \) respectively.

If we use \( v_k \) to represent the number of vehicles that entering the system between stations 1 and 2 and arrived at station 2 and \( v_k \) the genuine matches for route 1. From the constrains that: \( n_{2k} = x_{.,.k} + y_{.,.k} + v_k \) and \( n_{3k} = x_{.,.k} + w_{.,.k} + u_k \) the matches \( v_k \) and \( u_k \) can then be substituted as: \( v_k = n_{2k} - x_{.,.k} - y_{.,.k} \) and \( u_k = n_{3k} - x_{.,.k} - w_{.,.k} \).

The method further assumes that \( X_k \) and \( Y_k \) follows a trinomial distribution \( T(n_{1k}, \alpha_x, \alpha_y) \), \( V_k \) a Poisson process with mean \( \lambda_1 \), \( W_k \) a Binomial distribution \( B(w_k, \beta) \) and \( U_k \) a Poisson process with mean \( \lambda_2 \). The model further assumes that journey times for vehicles making movements between the pairs of observation points \((1, 2), (2, 3)\) and \((1, 3)\) follow normal distributions with means and variances equal to \((\mu_1, \sigma_1^2)\), \((\mu_2, \sigma_2^2)\) and \((\mu_3, \sigma_3^2)\) respectively. Then the distributional assumptions for the model are in terms of parameters \( \alpha_x, \alpha_y, \beta, \lambda_1, \lambda_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2 \) and \( \sigma_3^2 \) which are assumed to be fixed throughout the survey. The indicator variables for genuine matches to be estimated by the model are \( X_k, Y_k \) and \( W_k \).

A solution approach similar to that proposed in Watling & Maher (1992) is adopted here. The method is based on the use of indicator variables \( (X_k, Y_k, W_k) \) and works on the joint probability function. The method works in two iterative stages. In one stage, given the model parameters, the method determines the most probable combination of vehicle matches by maximising the joint probability density function with respect to the indicators, subject to certain constraints. In the other stage, maximum likelihood estimates are determined given the values of the indicator variables: the model parameters can be estimated by differentiating the log of joint probability function with respect each \( \alpha_x, \alpha_y, \beta, \lambda_1, \lambda_2, \mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2 \) and \( \sigma_3^2 \). It is reasonable to assume that the most probable matches correspond to the maximum likelihood estimations of the parameters.

With regard to the first stage, if we take the log of the joint probability density function and ignore the terms that are not involving \( x_{ijk}, y_{ijk} \) or \( w_{jlk} \) it can be shown that the problem becomes a maximisation problem:
Maximise
\[-\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{l=1}^{n_3} (d_{ijlk} x_{ijlk} + e_{ijk} y_{ijk} + f_{jlk} w_{jlk})\]

Subject to:
\[\sum_{i=1}^{n_1} x_{ijlk} \leq 1, \quad \sum_{j=1}^{n_2} x_{ijlk} \leq 1, \quad \sum_{l=1}^{n_3} x_{ijlk} \leq 1,\]
\[\sum_{j=1}^{n_2} y_{ijk} \leq 1, \quad \sum_{i=1}^{n_1} y_{ijk} \leq 1, \quad \sum_{l=1}^{n_3} w_{jlk} \leq 1, \quad \sum_{l=1}^{n_3} w_{jlk} \leq 1\]
\[x_{ijlk} = 0 \text{ or } 1, \quad y_{ijk} = 0 \text{ or } 1, \quad w_{jlk} = 0 \text{ or } 1 \quad (i = 1, 2, \ldots n_{k1}; j = 1, 2, \ldots n_{k2}; l = 1, 2, \ldots n_{k3})\]

Where
\[d_{ijlk} = \ln\left[\frac{\sqrt{2\pi}\theta \lambda_s (1-\alpha_x - \alpha_y)}{\alpha_s} (\sigma_1 + \sigma_2 + \sigma_3)\right]\]
\[+ \frac{1}{2} \frac{(s_{jk} - t_{ik} - \mu_1)^2}{\sigma_1} + \frac{1}{2} \frac{(r_{lk} - s_{jk} - \mu_2)^2}{\sigma_2} + \frac{1}{2} \frac{(r_{lk} - t_{ik} - \mu_3)^2}{\sigma_3}\]
\[e_{ijk} = \ln[\frac{\sqrt{2\pi}}{\alpha_s} (1-\alpha_x - \alpha_y)\sigma_1] + \frac{1}{2} \frac{(s_{jk} - r_{lk} - \mu_1)^2}{\sigma_1}\]
\[f_{jlk} = \ln[\frac{\sqrt{2\pi}}{\lambda_1 \sigma_2}] + \frac{1}{2} \frac{(r_{lk} - s_{jk} - \mu_2)^2}{\sigma_2}\]

---

**Figure 3.** Flow chart for the solution algorithm.
The two-stage iterative process is shown in Figure 3, where \( a \) denotes the set of parameters \( \{\alpha_x, \alpha_y, \beta, \lambda_1, \lambda_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_3^2\} \). At each iteration a new matching is solved based on the current parameter values. The new matches are then used to estimate a new set of parameters. The procedure starts with an initial set of estimates of parameters which are the best possible prior values available. The process ends until there is no difference in the parameter estimates between two successive iterations.

A computer software has been developed to implement the solution algorithm and a method for automatically enumerating the types of matches (Clegg, 2002). Demonstration of the algorithm through applications to a number of simulated data sets as well as the survey data will be presented.

3. FURTHER RESEARCH

Following the matching and reconstruction of hyper-route data, various summary measures will be computed, such as day-to-day reappearance rates on routes; daily variation of “departure time” (entry time to the study area); and mean and variance in link flows; travel times and park-and-ride use (if appropriate), both before and after the intervention. These measures, together with a stated preference self-completion survey conducted in parallel and trip and count information as collected by the Local Authority, will be used to investigate a number of behavioural hypotheses to explain the dynamics to driver route choice. This will be achieved by comparing the above mentioned measures with the output measures produced by the equilibrium model SATURN (Matzoros et al., 1987) and dynamic models DRACULA (Liu et al., 1995) and STEER (Clegg et al., 1995). A complimentary paper submitted to this Workshop (Clark & Batley, 2002) will present an initial analysis on drivers’ habitual route choice behaviour using such hyper-route data.

ACKNOWLEDGEMENTS

The author would like to thank the UK Engineering and Physical Sciences Research Council who supported this work with a research grant and David Watling and Richard Clegg for most helpful discussions.

REFERENCES


