Modelling the dynamics of bus use in a changing travel environment using panel data

Kiron Chatterjee

Centre for Transport & Society, University of the West of England, Bristol, BS16 1QY, UK

e-mail: Kiron.Chatterjee@uwe.ac.uk

tel: +44 117 32 82032

fax: +44 117 32 83002

Abstract Panel data offers the potential to represent the influence on travel choices of changing circumstances, past history and persistent individual differences (unobserved heterogeneity). A four-wave panel survey collected data on the travel choices of residents before and after the introduction of a new bus rapid transit service. The data shows gradual changes to bus use over the four waves, implying time was required for residents to become aware of the new service and to adapt to it. Ordered response models are estimated for bus use over the survey period. The results show that the influence of level of service (LOS) is underestimated if unobserved heterogeneity is not taken into account. The delayed response to the new service is able to be well represented by including LOS as a lagged variable. Current bus use is found to be conditioned on past bus use, but with additional influence of lagged LOS and unobserved heterogeneity. It is shown how different model specifications generate different evolution patterns with the most realistic predictions arising from a model which takes into account lagged responses to change in LOS and unobserved heterogeneity. The paper demonstrates the feasibility of developing panel data models that can be applied to forecasting the effect of interventions in the travel environment. Longer panels - encompassing periods of both stability and change - are required to support future efforts at modelling travel choice dynamics.

Keywords Travel mode choice; dynamics; panel data; unobserved heterogeneity; state dependence; initial conditions

Introduction

Models are required which can accurately represent people's travel choices in response to changes in the travel environment. Travel choice models are usually developed from cross-sectional survey data. These models reflect between–person variation in travel choices at a single point in time where it is assumed that travel choices are consistent with prevailing travel conditions, i.e. it is assumed an equilibrium state holds. For forecasting the impact of policy options adjustments are made to explanatory variables to reflect the scenario of interest and new travel choices are predicted, i.e. a new equilibrium state is forecasted. Forecasts are not possible of the path of evolution towards the new equilibrium. Goodwin (1998) claimed that the concept of equilibrium is acting as a barrier to sound policy advice and recommended that attention should be placed on dynamic analysis of travel choices.

Panel data involves repeated observations over time for the same individuals and can be used as an alternative basis for developing travel choice models. Models estimated using panel data are able to reflect within-person variation in travel choices over time, as well as between-person variation in travel choices. Inter-temporal dependence, or the dynamics, of travel choices is therefore able to be considered. This paper seeks to contribute to the practical development of methods to model the dynamics of travel choices from panel data. It presents an analysis of panel data collected for residents in an urban area in England before and after a new bus rapid transit service was introduced. Alternative model specifications for incorporating dynamics are tested and their merits assessed. An exposition is provided next of how panel data models can incorporate dynamics.

Exposition of possibilities of panel data

The exposition is based on a general specification for modelling discrete choices using panel data initially proposed by Heckman (1981a) and slightly modified by Kitamura (2000).

$$y_{it}^{*} = \beta' x_{it} + \sum_{l=1}^{t-1} \gamma_l y_{i,t-l} + \phi \sum_{s=1}^{t-1} \prod_{l=1}^{s} y_{i,t-l} + \alpha_i + \varepsilon_{it} \quad (1)$$

$$y_{it}^{*} = \begin{cases} 1 \; if \, y_{it}^{*} \geq 0 \\ 0 \; if \, y_{it}^{*} < 0 \end{cases} \quad (2)$$

where y^{*}_{it} is a latent continuous random variable for choice made by individual *i* at time period *t* and i = 1,…,N and t = 1,…,T.

The first term on the right hand side of Eq. 1 represents the contemporaneous effect of explanatory variables (*current circumstances*) where x_{it} is a vector of explanatory variables and β' is a vector of parameters to be estimated. This term is the same as that included in models estimated from cross-sectional data, but the advantage of panel data is that it collects multiple observations from the same individuals meaning that, as well as variation in x_{it} between individuals, it can incorporate variation in x_{it} for the same individuals between time periods (within-person variation)[1]. A generalization can be made to allow the β' parameters to vary between time periods to represent changes in travel choice sensitivity over time.

[1] Stated choice surveys can also be used to collect multiple observations of y^{*}_{it} for the same individuals with experimenter-specified values for x_{it} (representing hypothetical policy scenarios). This provides within-person variation in choices but the time frame in which choices are made is ambiguous.

As noted by Heckman (1981a), the first term on the right hand side of Eq. 1 can be expanded to include historic information for explanatory variables (*past circumstances*). For example, past values of explanatory variables (referred to in this paper as lagged explanatory variables) such as x_{it-1} could be included in a model to examine whether travel choices are influenced by past as well as current circumstances. In the case of an intervention, this might be used to see whether travel choices adapt immediately or remain (partially) influenced by previous conditions[2]. Future values of explanatory variables could also be included if there is reason to believe that travellers anticipate future conditions and take these into account in travel choices.

The second term on the right hand side of Eq. 1 represents state dependence and the effect of the history of past travel choices (*past behaviour*). Specifically, the coefficient, y_l, represents the effect of the choice l time periods ago on the current travel choice. The third term on the right hand represents duration dependence, a similar concept to state dependence. Through the coefficient, φ, it represents the effect of the length of time spent in state 1 for those individuals for whom the current state is 1. The rationale for state and duration dependence is that experience of behaviour may lead to altered familiarity, perceptions, preferences and constraints that are not able to be represented through other explanatory variables. They can be interpreted to represent habitual behaviour.

The fourth term on the right hand side (α_i) represents an individual specific error term. This can be treated as a fixed constant, or as a random variable, and is possible to incorporate into a model when repeated observations are available for individuals. It allows persistent differences in individual choice propensities to be recognized (without explaining them via structural variables). The term therefore represents *unobserved heterogeneity*. It is also theoretically possible with panel data to incorporate further unobserved heterogeneity by allowing parameters β', y_l and φ to vary by individual or groups of individual. Models

[2] Models which predict the dependent variable as a weighted function of present and past values of explanatory variables are called distributed-lag models (Grilliches 1967).

which include unobserved heterogeneity in parameters are referred to as random parameters models.

The final term on the right hand (ε_{it}) is a random error term that is independently distributed over *i* with arbitrary serial correlation. A generalization that can be made is to allow serial correlation in the random error term[3]. Serial correlation models capture state dependence through an unobserved component rather than structural variables.

It can be appreciated from the above exposition that the dynamic properties of travel choices, such as a persistent tendency of individuals towards a particular choice, can be specified in different ways and it is important to be led by theoretical considerations in specifying models. Bradley (1997) noted that panel surveys have usually been used for monitoring travel choices rather than estimating models for use in forecasting. Research is needed to confirm that potential advantages of panel data can be realized in practice. Before presenting new empirical work, the experience to date with using panel data to develop travel choice models is reviewed.

Review of travel choice modelling with panel data

The review pays particular attention to the assumptions made about dynamics in models and is organised according to the time frame in which the studied travel choices are made. The review concentrates on models of single travel choices made at repeated, discrete time intervals. Structural equation modelling (Golob, 2003) and duration modelling (Hensher, 1997) have also been used to analyse panel data, but these are outside the scope of the review.

Long-term dynamics

[3] Serial correlation is specified by specifying $\varepsilon_{it} = \rho\varepsilon_{it} + \zeta_{it}$, where ρ is the correlation and ζ_{it} is a random error term that is independent and identically distributed.

Kitamura and Bunch (1990) estimated ordered probit car ownership models from 1984-1987 Dutch National Mobility Panel (DNMP) data. As well as examining the influence of explanatory variables such as income and household size, they investigated the influence of unobserved heterogeneity and state dependence. They found that the importance of unobserved heterogeneity depended on the specification used. State dependence was found to be strongly significant, although its magnitude of effect depended on the assumptions made for initial conditions. The initial conditions problem occurs for panel data with a finite, small number of observation periods. The authors concluded that distinguishing the relative importance of unobserved heterogeneity and state dependence is problematic and further research is required. Kitamura and colleagues developed further models from DNMP data but noted that there remained unresolved issues about the specification of models to address serial correlation in error terms (Kitamura, 2009).

Dargay and Hanly (2007) used eleven years of data from the British Household Panel Survey to estimate an ordered probit car ownership model and a binary choice commute mode (car and non-car) model. They used a random effect specification (to allow for unobserved heterogeneity). They included state dependence variables (last year's car ownership or last year's commute mode) to test the possibility that households which are otherwise similar will have different choice probabilities in the current year depending on their car ownership or commute mode in the previous year. The results showed the importance of both state dependency and unobserved heterogeneity (after taking into account other factors such as household income and fuel prices).

Woldeamanuel et al. (2009) used data from the German Mobility Panel for 1996-2006 to estimate linear regression models of car ownership. They used a two-way random effects model to simultaneously examine the importance of unobserved heterogeneity between households and between time periods. Their results showed the presence of unobserved heterogeneity between households but not between time periods.

Srinivasan and Bhargavi (2007) investigated changes in mode choice of workers in Chennai, India. Respondents were asked in a household survey about travel behaviour and socio-economic characteristics at the time of the survey and five years previously. Dynamics of mode choice were examined through time-varying explanatory variables, state dependence variables and by allowing differences in users' sensitivity to explanatory variables for the two time periods. It was found that dynamic models provided a substantial improvement in model fit and failing to account for state dependence led to inflated estimates of level of service variables.

Short-term dynamics

Bhat (2000) used multi-day travel diary data for commuters in the San Francisco Bay area and estimated multinomial logit mode choice models with a random parameters specification to examine the effect of observed and unobserved heterogeneity in mode preferences and level of service parameters. He showed that ignoring unobserved heterogeneity in model parameters led to underestimating sensitivity to policy interventions.

Ramadurai and Srinivasan (2006) used one-day travel data from the San Francisco Bay Area Travel Survey to model within-day mode choice decisions. They used multinomial logit model specifications to examine inter-dependencies between mode choice decisions made during the day. The models considered time-varying explanatory variables, state dependence variables, lagged explanatory variables and unobserved heterogeneity. They found that each of these sources of dynamics was statistically significant, although lagged explanatory variables only to a limited extent.

Medium-term dynamics and changing travel environments

Bradley (1997) tested static and dynamic multinomial logit model specifications when looking at the effect on mode choice of a new rail commuter line in the Netherlands. Using before and after two-wave panel data for 475 commuters, he found improved model fit for dynamic model specifications (either accounting for response lags or state dependence) and found that forecasts are quite different if dynamic specifications are used instead of static specifications. He acknowledged limitations in the models that could be estimated with only two waves of data and concluded that, to understand and model the impacts of changes in the travel environment, 'multiple "after" periods are necessary to determine whether policies grow, diminish or remain stable over time'.

Yanez, Mansilla and Ortúzar (2010) collected panel data for about 300 residents in Santiago, Chile, before and after the introduction of Transantiago, a reorganized public transport system in February 2007. Mode choice was modelled using a multinomial logit random parameters framework with individual specific errors and individual specific parameters for explanatory variables (Yanez and Ortúzar 2009). The effect of the intervention was modelled by allowing the parameters for independent variables to be modified after the intervention through a shock effect, *S*. It was tested whether the effect was constant, whether it varied randomly according to individuals or systematically according to socio-economic group and whether it varied according to transport mode alternatives. Results showed that the shock effect varied among individuals and transport mode alternatives. The authors acknowledge that their model specification is unsuited to forecasting due to its representation of the intervention through a shock effect, rather than through changes to the values of explanatory variables. The authors refer to planned future work to incorporate inertia using a modelling framework introduced by Cantillo, Ortúzar and Williams (2007). This framework includes inertia thresholds which imply that a change of travel choice depends on whether the change in utilities between time periods (the change in the combined effect of explanatory variables) exceeds a specific value.

The review has shown that there is limited experience with developing dynamic travel choice models from panel data in the context of changing travel environments. The following section of the paper introduces the survey and data that is used in the analysis.

Panel survey and data

The context for the case study and the panel survey are described in detail in Chatterjee and Ma (2009) and are summarized in this section. The Fastway bus rapid transit system began operating in the Crawley and Gatwick Airport area in Southern England in 2003. The Fastway buses travel in dedicated lanes and guideways along significant parts of their routes. The first Fastway service (Route 10) was introduced in September 2003 and the second service (Route 20) was introduced in August 2005. The overall length of the Fastway routes is 24kms. The Route 20 service provides the case study for this paper. It provides a direct connection between neighbourhoods in the town of Crawley (Broadfield, Southgate Three Bridges) and Gatwick Airport and Horley. A Fastway network map can be found at http://www.fastway.info/.

A four-wave panel survey was conducted to obtain information on travel choices over a period of time before and after the introduction of the Route 20 service. Wave one took place one month before the introduction of the Route 20 service and the subsequent three waves took place at two month intervals after this. The target population for the panel survey was residents living close to the route of the new Route 20 service in Broadfield and Three Bridges. The panel survey used self-administered postal questionnaires as the survey instrument. Respondents were asked to provide personal and household information and frequency of use of different transport modes. 554 complete responses were received to the first wave of the survey and 186 residents provided complete responses for all four waves. The sample used for the analysis is the 186 residents who participated in all four waves. It is recognized that the analysis sample is not representative of the population from which it was drawn due to survey non-response bias. The data analysis was not intended to seek statistical generalizations for the survey population.

Duration modelling has been used to analyse the time taken for residents to start using the Route 20 service and the factors influencing this (Chatterjee and Ma, 2009). The analysis here focuses on the changing frequency of bus use of residents over the survey period. Traditionally, mode choice is analysed at the level of individual trips with the dependent variable being the transport mode chosen from the set of possible alternative options. The panel survey did not include a travel diary and hence insufficient information was available about specific trips to enable such an analysis.

The dependent variable in the analysis is the frequency of using the bus reported by residents (0 = not at all, 1 = less than once per week, 2 = 1 to 2 days a week, 3 = 3 to 4 days per week, 4 = 5 days a week or more). It is a discrete, ordered variable with five possible response values and can be analysed using ordered probability models.

A net increase in the number of bus users over the four waves is shown in Fig. 1. The increase is less prominent between waves one and two than between waves two and four. This implies time was required for residents to become aware of the new service and to respond to it. In terms of gross changes (turnover) in bus use between survey waves, the same number of residents increased bus use as reduced bus use between waves one and two, but 18 more residents increased bus use than reduced bus use between waves two and three and five more residents increased bus use than reduced bus use between waves three and four.

[Fig. 1 here]

Summary statistics for explanatory variables that were tested in the analysis are shown in Table 1. The explanatory variables include personal and household characteristics of the residents (mostly in the form of categorical variables) and three bus level of service (LOS) variables. These were each calculated with respect to a popular local destination (Gatwick Airport). Their values varied across residents according to home location and were modified by introduction of the Route 20 bus service.

[Table 1 here]

Model specifications

The following model specifications were tested with the panel data.

Pooled model

The simplest specification tested was a static (or Bernoulli) model which pools the data and assumes each observation is independent from other observations:

$$y^*_{it} = \beta' x_{it} + \varepsilon_{it} \tag{3}$$

where y^*_{it} is a latent ordinal variable for individual i at time period t; x_{it} is a vector of explanatory variables; β' is a vector of parameters to be estimated; and ε_{it} is a random error term that is assumed to be normally distributed across observations with mean = 0 and variance = 1 (leading to the ordered probit model).

With this equation the observed ordinal data, y_{it}, are defined as:

$$y_{it} = \begin{cases} 0\ if -\infty < y^* \le \tau_0 \\ 1\ if \tau_0 < y^* \le \tau_1 \\ 2\ if \tau_1 < y^* \le \tau_2 \\ 3\ if \tau_2 < y^* \le \tau_3 \\ 4\ if \tau_3 < y^* \le +\infty \end{cases} \tag{4}$$

where τ are threshold parameters that define *y*. The τ parameters are simultaneously estimated with the other model parameters. The software used for model estimation was Limdep version 8.0 (Greene 2002) where τ_0 is normalized to zero.

In Eq. 3 bus use can vary across individuals, *i*, based on inter-individual differences in explanatory variables and can vary across time periods, *t*, based on intra-individual differences in explanatory variables. It thus represents the effect of current circumstances.

Random effects model

Advantage can be made of the repeated observations available in a panel to recognise unexplained persistent differences in the bus use of individuals (unobserved heterogeneity). The introduction of an individual specific error (individual effect) term into Eq. 3 leads to:

$$y_{it}^{*} = \beta' x_{it} + \alpha_i + \varepsilon_{it} \tag{5}$$

where α_i is an error term, the value of which varies between individuals but is invariant over time. It is assumed that α_i is independent of x_{it}[4] and is normally distributed across individuals with mean = 0 and variance = 1.

Response lag model

[4] The assumption of independence between α_i and x_{it} is questionable and can introduce omitted-variable bias. Following Mundlak (1978) it is customary to parameterize the individual effect as a linear function of the mean of time varying independent variables as follows: $\alpha_i = \alpha\,\bar{x}_i + \eta_i$ where $\bar{x}_i$ is mean of time varying independent variables (instrumental variables), α is a vector of estimable parameters and η_i is independent of x_{it}. Correction for correlation between α_i and x_{it} is not shown in the subsequent model specifications but was tested for all random effects models and not found to be required.

Lagged explanatory variables are tested in the analysis to investigate if there is a delayed behavioural response to the introduction of the Route 20 bus service.

$$y_{it}^{*} = \beta' x_{it} + \sum_{r=1}^{R} \beta_r' x_{it-r} + \alpha_i + \varepsilon_{it} \qquad (6)$$

where R is a positive integer and β_r' is a vector of parameters to be estimated for lagged explanatory variables.

In the following analysis only bus LOS is included as a lagged explanatory variable. The effect of past circumstances is also modelled by testing the effect of change in bus LOS since the previous time period.

State dependence model

A simple version of the state dependence model was tested where only the behavioural state at the previous time interval was taken into account (Markov model).

$$y_{it}^{*} = \beta' x_{it} + \gamma y_{it-1} + \alpha_i + \varepsilon_{it} \qquad (7)$$

where y_{it-1} is the observed ordinal variable for individual i at time period t-1 (state variable) and γ is a parameter to be estimated.

Serial correlation models capture state dependency though an unobserved component, but interest in this analysis was on explicitly modelling state dependence and hence this was not tested[5].

Kitmaura (2000) notes that the state dependence model of Eq. 7 can be viewed as a special case of a distributed-lag model (where only the lag for the previous time period is included), since y_{it-1} would be expected to be a function of the prevailing values of explanatory variables at *t-1*. However, it is to be noted that y_{it-1} is a non-stochastic variable (the previous observed behavioural state), rather than a predicted modelled value, and hence there is an important distinction. In fact, lagged explanatory variables can be included in Eq. 7 in addition to state variables to examine if there is a separate effect of past prevailing circumstances to that of past observed behavioural state.

Initial conditions problem

The initial conditions problem arises when applying Eq. 7 to panel data with a finite, small number of observation periods where the choice process is not observed from the start. The simplest solution is to drop the first set of observations for each individual with the value of the state variable in the second observation (now the first observation that is modelled) the observed value for the behavioural state at the first observation. This is only a reasonable assumption if the behavioural process is being observed from the start (and the first observation represents the first period at which the behaviour takes place), or if the process is in equilibrium at the time of the first observation. Without these conditions applying (which is normally the case with empirical data) it is likely that the initial state variable, y_{i1}, will be correlated with the individual effect, α_i, and inconsistent estimators will be obtained.

[5] Roy, Chintagunta and Haldar (1996) found that serial correlation was not statistically significant after state dependence and unobserved heterogeneity were included in panel data models for ketchup purchasing.

Different methods have been put forward to address the initial conditions problems. Hsiao (2003; 208-211) describes a two-step method suggested by Heckman (1981b) which is based on first approximating the probability of the initial state variable, y_{i1}, as a function of x_i (as x_i varies across whole observation period) and any pre-sample variables that are available and, second, applying Eq. 7 to model y_{it} for subsequent time periods (time = 2,…n), whilst allowing the error term from the first step to be freely correlated with ε_{it}. This method has not been widely applied, since standard software is not able to be used to estimate it. Two other solutions proposed by Orme (2001) and Wooldridge (2005) are more easily applied in standard econometric modelling software. Orme's method uses the same first step as Heckman and obtains an estimate of the error term[6] that is then added to Eq. 7 which can be estimated as a standard random effects model. The first step is:

$$y_{i1}^{*} = \lambda' z_i + \eta_i \tag{8}$$

for t = 1, where y_{i1} is a latent ordinal variable for individual i at time period 1 (first observation), λ' is a parameter to be estimated, z_i is a vector including explanatory variables relevant to time period 1, pre-sample information and means of time varying explanatory variables x_{it} and η_i is an individual effect which has correlation with α_i of ρ. The second step is:

$$y_{it}^{*} = \beta x_{it} + \gamma y_{it-1} + \delta e_i + \alpha_i + \varepsilon_{it} \tag{9}$$

for t = 2, … ,N.

[6] Generalized residual is calculated as: $e_i = (2y_{i1}-1)\varphi(\lambda' z_i)/\Phi(\{2y_{i1}-1\}\lambda' z_i)$ where φ and Φ are the normal density and distribution functions respectively and λ' are estimated values of parameters in Eq. 8.

Instead of starting by specifying a distribution of the probability of y_1 given α_i, Wooldridge's method specifies an approximation for the probability of α_i given y_1.

$$\alpha_i = \alpha_0 + \alpha_1 y_{i1} + \alpha_2 \bar{x}_i + u_i \qquad (10)$$

where $\bar{x}_i$ is means of time varying explanatory variables and u_i is a random error term that is assumed to be normally distributed across observations with mean = 0 and variance = 1 and is independent of y_{i1} and $\bar{x}_i$.

Substituting Eq. 10 in Eq. 7 gives:

$$y^*_{it} = \beta x_{it} + \gamma y_{it-1} + \alpha_1 y_{i1} + \alpha_2 \bar{x}_i + u_i + \varepsilon_{it} \qquad (11)$$

for t = 2, … ,N.

Arulampalam and Stewart (2008) compared the approximation methods of Heckman, Orme and Wooldridge and found similar results were obtained for each method. The Orme and Wooldridge methods are tested in the following analysis.

Modelling results

Four wave models

The first set of models presented in Table 2 is estimated using all four waves of the data. A model using only wave one data is shown for comparison. Explanatory variables included in the models are those that are statistically significant across the majority of models tested with the addition of age and gender, which are not statistically significant but are retained for interest.

[Table 2 here]

Bus use frequency is higher for residents who live in the Broadfield residential area (AREA), do not have a household car (HHC_0), have a bus pass (BPASS) and who commute to work (COMMU). Bus use frequency is lower for residents who have a driving licence (LICEN) and have two household cars (HHC_2). The LOS variable included in the models is DIFF_CAR, the difference in total travel time by bus compared to car. Its statistical significance is higher than the other two LOS variables tested. It has the expected sign where the longer the journey by bus compared to car the lower the frequency of bus use predicted.

The random effects model (model 3) provides a considerable improvement of model fit over the pooled model (model 2) and highlights a high degree of persistence in bus frequency across observations that is not explained by explanatory variables. It is also notable that in the random effects models the statistical significance of DIFF_CAR is higher than in the pooled model. This shows that taking account of unobserved heterogeneity enables the influence of DIFF_CAR to be better estimated. Including DIFF_CAR as a lagged variable by one time period (model 4) improves model fit and the variable has higher statistical significance than when included as a contemporaneous variable. When including the full set of prevailing and historical information for DIFF_CAR (representing a distributed lag model) it is only the one period lagged variable that is statistically significant. The implication is that there is a delayed response by one time period (two months) between Route 20 introduction and bus use.

The effect of past circumstances was also modelled by testing variables for (i) the absolute change in DIFF-CAR; and (ii) the occurrence or not of a change in DIFF_CAR exceeding a threshold value. It was also appropriate in these models to include base values for DIFF_CAR (wave one values prior to introduction of Route 20). It was found that including a dummy variable for a threshold reduction of travel time exceeding 10 minutes (DIFF_CAR_RED10) was statistically significant with expected sign and the base value of DIFF_CAR was also statistically significant with expected sign (see model 5). However, the model fit was inferior to model 4.

The possibility of unexplained differences in bus use at different time periods which are consistent for individuals (time effects) was tested by adding dummy variables for waves 2, 3 and 4. These were not found to be statistically significant when included in models with DIFF_CAR (models 3 to 5).

It is difficult to compare parameter estimates between different ordered probit models due to differences in estimates of threshold parameter values from one model to another. Furthermore, estimated parameters for the random effects models are not directly comparable to those for the pooled model due to different scaling of the error variance. Marginal effects can be used to compare parameter estimates. They are computed at mean values for explanatory variables with the random effect term set to zero.

Calculation of marginal effects for the wave one model shows that a one minute increase in DIFF_CAR is associated, *ceteris paribus*, with a 2.4% increased probability of not using the bus at all and a 0.2% reduced probability of using the bus five days a week or more. A lower sensitivity to LOS is found for the panel data models. With model 4 a one minute increase in DIFF_CAR is associated with a 0.5% increased probability of not using the bus at all and a 0.1% reduced probability of using the bus five days a week or more. The greater sensitivity to LOS evident for the wave one model is likely to be due to wave one travel choices having been made in a stable context where they have adjusted to the travel environment, whereas the four wave models encompass a period when travel choices are in the process of adjusting to the changed travel environment.

The source of the persistence in the random effects models is unclear without further model tests. It could be due to state dependence. Kitamura (2000) refers to a test that has been proposed to distinguish whether the persistence is due to true state dependence or spurious state dependence. It involves including lagged explanatory variables along with non-lagged explanatory variables in a model and comparing whether predicted outcomes are modified. Conducting the test showed that including lagged as well as non-lagged LOS variables for DIFF_CAR resulted in improved model fit. Predicted outcomes are therefore affected by inclusion of lagged variables and it is shown that there is a dynamic response to the intervention and hence there is state dependence. It is therefore worthwhile to test models with state dependence.

Three wave models

State dependence models are now considered. Due to the initial conditions problem these could only be estimated for three waves of data (waves two to four). For reference purposes pooled and random effects models similar to those previously discussed are shown in Table 3 (models 6 to 8). Parameter estimates are similar, except DIFF_CAR is not statistically significant unless included as a lagged variable.

[Table 3 here]

Three state dependence models are shown: naïve model (no correction for initial conditions problem, i.e. y_{i1} being correlated with α_i) (model 9); Orme model (model 10); and Wooldridge model (model 11). In the naïve model the state variables dominate and the only other statistically significant variables are HHC_0 and BPASS. In the first step of the Orme model, z_i includes values for the explanatory variables at time period 1, one variable representing pre-sample information and means of time varying independent variables x_{it}.

The pre-sample variable measured at wave 1 was a variable for attitude to bus use ('buses provide a realistic option for some of my journeys') which was intended to represent experience of using buses prior to the survey. The results for the first step model are not shown but a number of variables were statistically significant (similar variables to those shown in Table 2 for four wave data), including DIFF_CAR (with expected sign, t-value = -2.48). The pre-sample variable was statistically significant at very high significance level (t-value = -5.29).

A similar result is obtained for the Orme second step model (model 10) as the naïve model (model 9). The individual effect is not statistically significant in either models, implying persistence in bus use is due to state dependence and not unobserved heterogeneity. Lack of statistical significance of the generalised residual coefficient in the Orme model suggests that y_{i1} is not correlated with α_i after controlling for pre-sample experience and that initial conditions can be treated as exogenous.

The Wooldridge model results tell a different story with initial state variables dominating previous time period state variables (which are not statistically significant). The estimated coefficients for the initial period observations show a positive gradient which implies a positive correlation between the initial period observations and individual effect. The lagged variable for DIFF_CAR is close to statistical significance and the individual effect is statistically significant. Calculation of marginal effects for state variables shows that using the bus 5 days a week or more in first wave increases the probability of using the bus 5 days a week or more' by 77.1% and decreases the probability of not using the bus at all by 58.8%. Using the bus 5 days a week or more in the previous wave increases the probability of using the bus 5 days a week or more by 0.6% and decreases the probability of not using the bus at all by 15.5%.

State dependence is shown to be important in all of the models but there are different interpretations. The naïve and Orme models suggest that without explicitly accounting for the effect of initial choices (but implicitly taking into account unobserved initial differences in the

Orme model) previous choices strongly determine current choices, but there are no persistent differences due to unobserved factors. With the Wooldridge model, bus use is strongly conditioned on initial state variables. The Wooldridge model leads to a better fitting model (after taking into account the extra three degrees of freedom used). After explicitly accounting for the effect of initial choices, it suggests there are persistent differences in travel choices (from wave 2 onwards) that are due to unobserved factors, rather than bus use at previous period, and LOS is influential (with one-period lagged response). The results for the Wooldridge model are considered to provide the best evidence and indicate that there is both state dependence and unobserved heterogeneity.

It should be pointed out that the ability to address the initial conditions problem was compromised with this panel data due to the limited availability of explanatory variables which had within-person variation over the survey period. Within-person variation was present in the LOS variable, but this is an exogenous variable that does not contribute to correlation between y_{i1} and α_i. The Orme method can take advantage of pre-sample information (to use as instrumental variables), but the information that was available about residents prior to the survey period was limited. These two points mean that the Orme and Wooldridge methods were compromised in being able to distinguish state dependence and unobserved heterogeneity. Future attempts at modelling state dependence require longer period panels which encompass both pre-intervention stability and post-intervention change. Panel surveys should seek to include more variables with within-person variation and should seek retrospective information about behaviour prior to the first observation period.

Comparing predictions to observations

A comparison was made of how well different four-wave models predicted the overall change in bus use arising after introduction of the new bus service (three-wave models could not be used for this comparison as they were only estimated for waves two to four). The predictions demonstrated that the cross-sectional wave one model overestimates the initial number of non-bus users (predicting 156 non-users instead of observed 114 non-users) and overestimates the increase in bus use (predicting increase from 30 users to 67 users compared to observed increase from 72 users to 91 users). The cross-sectional model may be viewed as providing a long-run prediction, but it is important to note that in this study the travel environment had been stable prior to the panel survey being conducted. Where this is not the case, cross-sectional models are unlikely to provide valid predictions of long-term outcomes. The random effects models (models 3-5) are more accurate in their initial predictions but underestimate the increase in bus use that occurs.

In Table 4 a comparison is made for three-wave models of predicted sequences of bus use (run patterns) to observed sequences. Bus frequency values were recoded with values of 2, 3 and 4 recoded as 2 for the purposes of presenting a manageable set of run patterns in the table.

[Table 4 here]

Table 4 shows that the pooled model (model 6) predicts too many residents not using the bus at any period. The random effects model (model 7) provides run pattern predictions closer to the observations, but overestimates stability with few changes in bus use. This is unsurprising since no explicit dynamics is incorporated into the model (only unobserved heterogeneity). Model 8 with lagged LOS variable predicts 25 residents increase bus frequency between waves two and three compared to 21 observed increases in frequency. It only predicts two residents increase bus frequency between waves three and four compared to 17 observed increases. This arises due to its one period lag specification. Model 11 (Wooldridge state dependence) shows the closest match to observations, but only predicts that 11 residents increase bus frequency between waves two and four compared to the 27 observed (and 26 predicted by model 8). It overestimates stability in bus use. Model 8 best captures the response to the introduction of the new bus service.

Using the panel data models for forecasting

Forecasting implications are compared for the hypothetical scenario of a general 10 minute reduction in bus travel times. The forecasts were made for the analysis sample of 186 residents. It was of interest to compare the forecasts from a state dependence model with other models, hence models estimated from three waves of data were used. Socio-economic characteristics of the analysis sample were kept constant and set to observed wave one values. Initially DIFF_CAR was set to wave one values and then a 10 minute reduction was applied for three subsequent time periods. Predictions for models 1, 7, 8 and 11 are shown in Fig. 2.

[Fig. 2 here]

There is only any predicted change in bus use between time periods one and two for the wave one model (model 1) and random effects model with LOS variable (model 7). A substantially larger increase in bus use is predicted from the wave one model, but this model overestimates the number of residents who do not initially use bus. A delayed response to the intervention (between time periods two and three) is shown for the random effects model with lagged LOS variable (model 8). Modest change in bus use is shown for the Wooldridge state dependence model (model 11) with only slight change occurring between waves one and two and between waves two and three. State variables in the presence of changing explanatory variables mean that state dependence models can predict gradual change over time (as shown in Kitamura 2000; 124). The modest forecasted change in bus use for model 11 arises due to the dominance of the initial period state variables. With the context and data used in this study the most realistic forecasts arise from the random effects model specification which takes into account lagged responses to the change in LOS (model 8). The inclusion of state variables leads to exaggerated predictions of the stability of travel choices.

Summary and conclusions

The panel data enabled the effect on travel choices to be represented of changing circumstances, past history (both past circumstances and past behaviour) and persistent individual differences (unobserved heterogeneity). Cross-sectional data does not allow these sources of dynamics to be considered.

The analysis has shown that treating the panel data observations as independent observations does not result in a statistically significant estimate for the LOS variable. However, recognizing that the data consists of repeated observations and accounting for unobserved heterogeneity indicates a statistically significant effect of LOS. This highlights the necessity of recognizing unobserved heterogeneity in panel data.

It is found that LOS in the previous observation period provides a better explanation of bus use than LOS in the current observation period. This is consistent with the observed delay in response to the new bus service which presumably arises from residents requiring time to become aware and adapt to its introduction. Previous studies have shown the importance of past behaviour in determining current behaviour. This can be interpreted as habitual behaviour or commitment to the behaviour. The Wooldridge state dependence model suggests that bus use initially observed at the start of the survey period exerts a continuing influence on bus use. After taking into account initial bus use, subsequent bus use is found to be influenced by LOS (with a one-period lagged response) and persistent differences due to unobserved factors.

The predictions of different model specifications are compared. A wave one cross-sectional model underestimates initial bus use and predicts larger increases in overall bus use than observed. Panel data models slightly underestimate the increase in overall bus use compared to that observed, but examination of run patterns shows that their predictions are closer to observations than those from the cross-sectional model.

Panel data models allow dynamic evolution patterns to be forecasted. This is seen for the scenario of a general 10-minute reduction in bus travel times. Three different specifications are compared and show immediate responses to the intervention, delayed responses and gradual responses. The most realistic predictions arise from the model specification which takes into account lagged responses to change in LOS and unobserved heterogeneity. The state dependence model, which takes into account the influence of past behaviour, leads to exaggerated predictions of the stability of travel choices. However, this is a consequence of the short panel used and resulting difficulties in addressing the initial conditions problem. In principle, a model that recognises the effect of past behaviour (alongside other influences) would provide the best basis to represent the gradual responses to a change in the travel environment that occurred in this study.

While a variety of model specifications was tested, there are other promising specifications that could be tested in future work. Yanez and Ortúzar (2009) used a random parameters framework to allow heterogeneity in relationships (for example, in the extent of influence of past behaviour) to be examined. A latent class modelling approach would allow groups of individuals with similar relationships to be identified. These approaches increase sample size requirements by requiring the estimation of a larger number of parameters, but are attractive from a theoretical point of view since dynamic responses are likely to vary across the population.

Ordered response models were estimated in this study and future work should seek to develop panel data models for qualitative choices between travel models using traditional multinomial logit specifications. It would also be valuable to better understand the behavioural processes that generate the dynamics of travel choices. Structural equation modelling (SEM) has been used in some past studies (e.g. Golob, Kitamura and Supernak 1997) to explore the role of subjective factors (perceptions, attitudes, habits), their inter-dependency with travel behaviour and how this changes over time. Qualitative research would also be helpful to better understand why travel choices are affected, not only by current circumstances, but by past circumstances and behaviour.

To further develop the capabilities of incorporating dynamics into models it is recommended that, in areas where major transport initiatives are being contemplated (and therefore survey and monitoring resources can be reasonably justified), priority is given to incorporating a panel element in travel surveys. Opportunities should also be explored to take advantage of new technologies (smart cards, mobile phone records, on-line booking systems) which can automatically track travel behaviour over time.

Acknowledgements

The research reported in this paper was funded by a grant from the UK Engineering and Physical Sciences Research Council (EPSRC). The author wishes to thank his former colleague, Dr Kang-Rae Ma (Chung-Ang University, Korea), for his contribution to the research and his colleague, Professor Goodwin, for useful discussions about the data analysis.

References

Arulampalam, W., Stewart, M.B.: Simplified implementation of the Heckman estimator of the dynamic probit model and a comparison with alternative estimators. Warwick Economic Research Paper Series No. 884, Department of Economics, University of Warwick. http://econpapers.repec.org/paper/wrkwarwec/884.htm. Accessed 12 April 2010 (2008)

Bhat, C.R.: Incorporating observed and unobserved heterogeneity in urban work travel mode choice modelling. Transp. Sci. **34**(2), 228-238 (2000)

Bradley, M.: A practical comparison of modeling approaches for panel data. In: Golob, T.F. Kitamura, R., Long, L. (eds.) Panels for Transportation Planning, pp. 281-304. Kluwer, Boston (1997)

Cantillo, V., Ortúzar, J. de D., Williams, H.C.W.L.: Modelling discrete choices in the presence of inertia and serial correlation. Transp. Sci. **41**(2), 195-205 (2007)

Chatterjee, K., Ma, K.: Time taken for residents to adopt a new public transport service: examining heterogeneity through duration modelling. Transportation **36**(1), 1-25 (2009)

Dargay J., Hanly, M.: Volatility of car ownership, commuting mode and time in the UK. Transp. Res. Part A **41**(10), 934 – 948 (2007)

Golob, T.: Structural equation modeling for travel behavior research, *Transp. Res. Part B* **37**(1), 1–25 (2003)

Golob, T., Kitamura, R., Supernak, J.: A panel-based evaluation of the San Diego I-15 Carpool Lanes project. In: Golob, T.F. Kitamura, R., Long, L. (eds.) Panels for Transportation Planning, pp. 97-128. Kluwer, Boston (1997)

Goodwin, P.B.: The end of equilibrium. In: Garling, T., Laitila, T., Westin, K. (eds.) Theoretical Foundations of Travel Choice Modeling, pp. 103-132. Elsevier, Oxford (1998)

Greene, W.H.: LIMDEP version 8.0: Econometric Modelling Guide volume 2. Econometric Software, Plainview, NY (2002)

Grilliches, Z.: Distributed lags: a survey. Econometrica **35**, 16-49 (1967)

Heckman, J.J.: Statistical models for discrete panel data. In: Manski, C.F., McFadden, D. (eds.) Structural Analysis of Discrete Data with Econometric Applications, pp. 114-178. MIT Press, Cambridge, Mass (1981a)

Heckman, J.J.: The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In: Manski, C.F., McFadden, D.

(eds.) Structural Analysis of Discrete Data with Econometric Applications, pp.179-195. MIT Press, Cambridge, Mass (1981b)

Hensher, D.A.: The timing of change: discrete and continuous time panels in transportation. In: Golob, T.F. Kitamura, R., Long, L. (eds.) Panels for Transportation Planning, pp. 305-319. Kluwer, Boston (1997)

Hsiao, C.: Analysis of Panel Data, 2nd Edn. Cambridge University Press, Cambridge (2003)

Kitamura, R.: Longitudinal methods. In: Hensher, D.A., Button, K. J. (eds.) Handbook of Transport Modelling, pp. 113-129. Elsevier, Oxford, (2000)

Kitamura, R.: A dynamic model system of car ownership, trip generation and mode split: model development and simulation experiment. Transportation **36**(6), 711-732 (2009)

Kitamura, R., Bunch, D.S.: Heterogeneity and state dependence in household car ownership: a panel analysis using ordered-response probit models with error components. In: Koshi, M. (ed) Transportation and Traffic Theory, pp. 477-496. Elsevier, Oxford. (1990)

Mundlak, Y.: On the pooling of time series and cross section data. Econometrica **46**, 69-85 (1978)

Orme, C.D.: Two-step inference in dynamic non-linear panel data models. Manuscript, School of Economic Studies, University of Manchester. http://personalpages.manchester.ac.uk/staff/chris.orme/. Accessed 12 April 2010 (2001)

Ramadurai, G., Srinivasan, K.K.: Dynamics and variability in within-day mode choice decisions: role of state dependence, habit persistence, and unobserved heterogeneity. Transp. Res. Rec. **1977**, 43-52 (2006)

Roy, R., Chintagunta, P.K., Haldar, S.: A framework for investigating habits, “the hand of the past,” and heterogeneity in dynamic brand choice. Marketing Sci. **15**(3), 280-299 (1996)

Srinivasan, K.K., Bhargavi, P.: Longer-term changes in mode choice decisions in Chennai: a comparison between cross-sectional and dynamic model. Transportation **34**(3), 355-374 (2007)

Woldeamanuel, M.G., Cyganski, R., Schulz, A., Justen, A. Variation of households' car ownership across time: application of a panel data model. Transportation **36**(4), 371-387 (2009)

Wooldridge, J.W.: Simple solutions to the initial conditions problem in dynamic, non-linear panel data models with unobserved heterogeneity. J. of Appl. Econometrics. ***20***, 39-54 (2005)

Yanez, M.F., Ortúzar, J. de D.: Modelling choice in a changing environment: assessing the shock effects of a new transport system. Paper presented at First International Choice Modelling Conference, 30 March – 1 April, Harrogate, UK (2009)

Yanez, M.F., Mansilla, P., Ortúzar, J. de D.: The Santiago Panel: measuring the effects of implementing Transantiago. Transportation **37**(1), 125-149 (2010)

Author Biography

Kiron Chatterjee is a Senior Lecturer in Transport Planning at the University of the West of England, Bristol, where he is course leader for the MSc in Transport Planning. A major focus of Kiron's current research is on the study of travel behaviour change using longitudinal data. This includes studying the effects on travel behaviour of social change and of transport interventions.

Figures

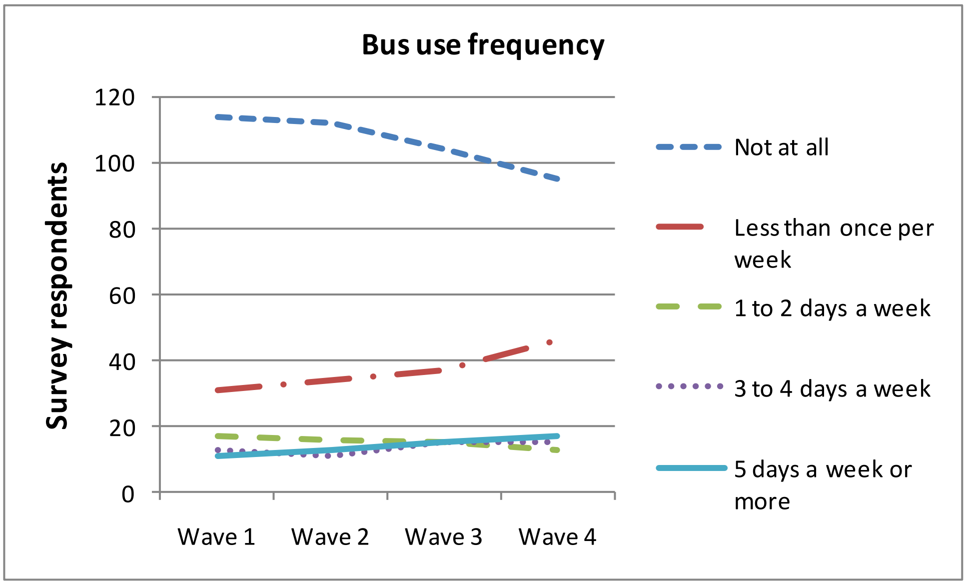

Fig. 1 Bus use frequencies reported by respondents (N=186)

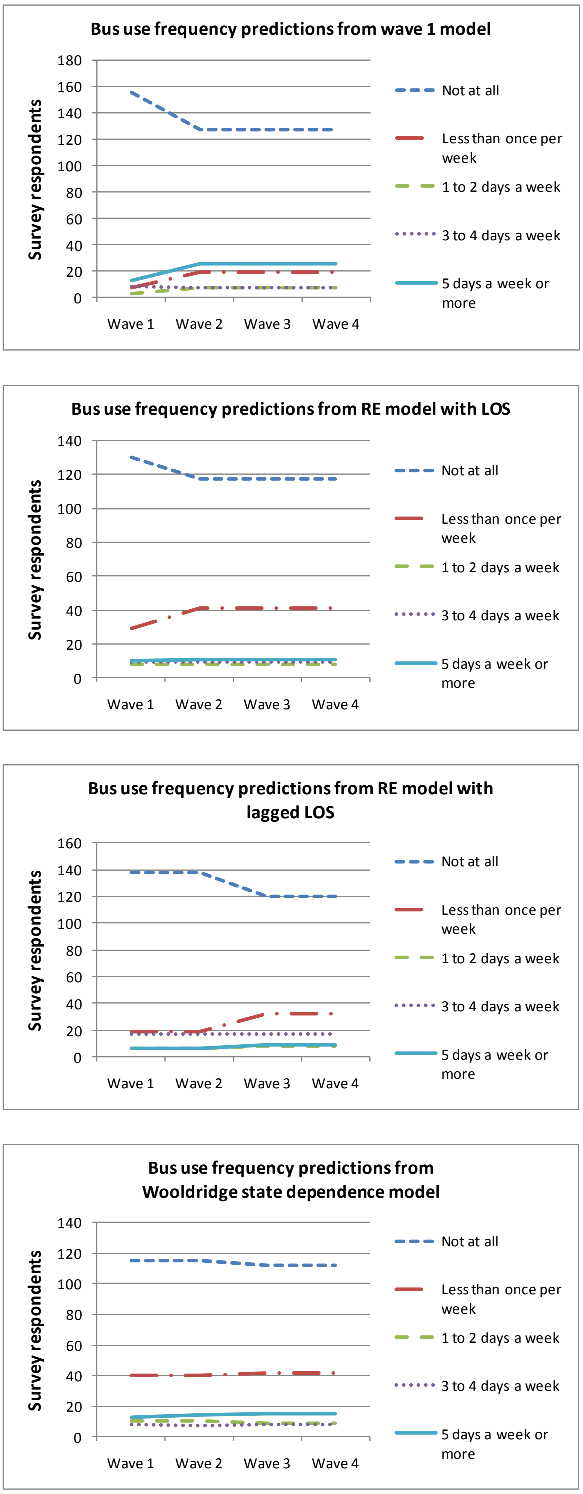

Fig. 2 Bus use frequency predictions from models 1, 7, 8 and 11 for 10 minute travel time reduction scenario (N=186)

Tables

Table 1 Explanatory variables

Explanatory variable	Categories	Total number of respondents at wave 1	Percentage of respondents at wave 1
Time non-varying categorical			
Residential area (AREA)	0 = Three Bridges	114	61
	1 = Broadfield	72	39
Gender (GENDER)	0 = Female	104	56
	1 = Male	82	44
New resident (NEW_RES)	0 = Not new resident	178	96
	1 = Less than one year	8	4
Live with spouse (SPOUSE)	0 = No	83	45
	1 = Yes	103	55
Time varying categorical			
Driving licence (LICEN)	0 = No	24	13
	1 = Yes	162	87
Full-time employed (OCC_FULL)	0 = No	96	52
	1 = Yes	90	48
Part-time employed (OCC_PART)	0 = No	151	81
	1 = Yes	35	19
Retired (OCC_RETI)	0 = No	150	81
	1 = Yes	36	19
Children in household (CHILD)	0 = No	150	81
	1 = Yes	36	19
Cars in household (HH_CAR)	0 = 0 car	19	10
	1 = 1 car	94	51
	2 or more cars	73	39
No car in household (HHC_0)	0 = No	167	190
	1 = Yes	19	10
2+ cars in household (HHC_2)	0 = No	113	61
	1 = Yes	73	39
Bus pass (BPASS)	0 = No	165	89
	1 = Yes	21	11
Commute to work (COMMU)	0 = No	60	32
	1 = Yes	126	68
Job change in last 3 months (JOB_CHA)	0 = No	178	96
	1 = Yes	8	4
Explanatory variable	**Mean**	**Minimum**	**Maximum**
Time non-varying continuous			
Age (AGE)	48.9	17	82

Walk access time to bus stop for bus to Gatwick (minutes) (WALK_BUS)	11.8 (wave 1) 5.8 (waves 2-4)	1 (wave 1) 1 (waves 2-4)	19 (wave 1) 12 (waves 2-4)
Total time to travel to Gatwick by bus (minutes) (TOT_BUS)	37.6 (wave 1) 26.0 (waves 2-4)	18 (wave 1) 13 (waves 2-4)	53 (wave 1) 40 (waves 2-4)
Difference in total time to travel to Gatwick by bus compared to car (minutes) (DIFF_CAR)	22.7 (wave 1) 11.1 (waves 2-4)	9 (wave 1) 3 (waves 2-4)	35 (wave 1) 21 (waves 2-4)

Table 2 Panel data models for four waves of data

Model **Parameter estimates (t-values in brackets)**	**1. Wave one only**	**2. Pooled**	**3. Random effects**	**4. Random effects with lagged LOS**	**5. Random effects with threshold change in LOS**
Explanatory variables					
Constant	0.969 (1.37)	0.269 (0.95)	0.216 (0.23)	1.681 (2.23)	1.420 (1.17)
AREA	1.224 (3.71)	0.648 (5.73)	1.686 (4.78)	1.112 (3.82)	2.473 (4.48)
AGE	-0.00237 (-0.29)	-0.00582 (-1.51)	-0.00334 (-0.25)	-0.00241 (-0.22)	0.00281 (0.20)
GENDER	0.082 (0.42)	0.023 (0.24)	0.162 (0.47)	-0.058 (-0.22)	0.184 (0.55)
LICEN	-0.712 (-2.49)	-0.561 (-3.96)	-1.608 (-3.51)	-2.031 (-5.37)	-1.357 (-3.28)
HHC0	1.016 (3.09)	1.064 (6.65)	2.942 (4.87)	2.718 (6.78)	2.621 (5.52)
HHC2	-0.509 (-2.40)	-0.454 (-4.52)	-1.385 (-3.75)	-1.758 (-5.52)	-1.381 (-3.67)
BPASS	1.627 (5.37)	1.590 (11.14)	2.670 (11.32)	2.149 (8.73)	2.325 (9.39)
COMMU	0.427 (1.80)	0.183 (1.62)	0.529 (1.40)	0.502 (1.62)	1.022 (2.62)
DIFF_CAR	-0.0627 (-2.52)	-0.00999 (-1.38)	-0.0234 (-2.79)	-	-
DIFF_CAR (one period lagged)	-	-	-	-0.0325 (-4.20)	-
DIFF_CAR (wave 1 value)	-	-	-	-	-0.124 (-3.09)
DIFF_CAR_RED10 (reduction in DIFF_CAR exceeding 10 minutes)	-	-	-	-	0.439 (3.06)
Threshold parameters					
τ_1	0.707 (6.64)	0.813 (14.93)	1.759 (18.32)	1.804 (17.45)	1.780 (17.45)
τ_2	1.312 (9.25)	1.312 (19.60)	2.817 (25.95)	2.907 (24.73)	2.824 (25.45)
τ_3	2.019 (10.06)	1.930 (21.38)	4.172 (33.33)	4.338 (35.77)	4.174 (33.22)
Random effect					
σ_u	-	-	2.168 (11.05)	2.618 (12.01)	2.246 (11.48)
$\rho = \sigma_u^2/(\sigma_\varepsilon^2 + \sigma_u^2)$	-	-	0.825	0.873	0.835
Goodness of fit statistics					
L(β) (Log lik.with τ & β)	-172.2	-736.7	-566.3	-563.1	-565.1
L(C) (Log lik.with τ)	-217.7	-916.0	-916.0	-916.0	-916.0
-2[L(C)- L(β)]	91.1	358.7	699.4	705.8	701.8
Degrees of freedom	9	9	10	10	11
No. observations	184	744	744	744	744

Table 3 Panel data models for three waves of data

Model **Parameter estimates (t-values)**	**6. Pooled model**	**7. Random effects with LOS**	**8. Random effects with lagged LOS**	**9. Naïve state dependence**	**10. Orme state dependence**	**11. Wooldridge state dependence**
Explanatory variables						
Constant	0.124 (0.37)	0.622 (0.43)	0.172 (0.16)	-0.479 (-1.05)	-0.422 (-0.92)	0.425 (0.40)
AREA	0.453 (2.98)	1.843 (3.18)	1.507 (3.94)	0.235 (1.44)	0.210 (1.25)	0.442 (1.19)
AGE	-0.00664 (-1.49)	-0.00856 (-0.51)	-0.0120 (-0.77)	-0.00882 (-1.42)	-0.00942 (-1.49)	-0.0241 (-1.69)
GENDER	-0.016 (-0.15)	0.150 (0.35)	0.323 (0.86)	-0.007 (-0.05)	0.018 (0.13)	0.114 (0.38)
LICEN	-0.562 (-3.40)	-1.836 (-2.95)	-1.058 (-2.13)	-0.093 (-0.46)	-0.113 (-0.55)	-1.559 (-0.51)
HHC0	1.133 (6.14)	2.968 (3.88)	2.958 (5.28)	0.554 (2.08)	0.583 (2.10)	2.995 (0.64)
HHC2	-0.457 (-3.98)	-1.275 (-2.64)	-1.338 (-3.40)	-0.202 (-1.43)	-0.166 (-1.16)	-0.547 (-0.33)
BPASS	1.657 (10.05)	3.200 (6.22)	4.627 (10.76)	0.889 (4.44)	0.905 (4.44)	0.571 (0.21)
COMMU	0.124 (0.95)	0.454 (0.84)	0.670 (1.57)	0.009 (0.05)	-0.027 (-0.14)	0.143 (0.10)
LICEN mean	-	-	-	-	-	1.118 (0.34)
HHC0 mean	-	-	-	-	-	-1.898 (-0.42)
HHC2 mean	-	-	-	-	-	0.046 (0.03)
BPASS mean	-	-	-	-	-	1.317 (0.44)
COMMU mean	-	-	-	-	-	-0.404 (-0.27)
DIFF_CAR	0.0186 (1.20)	-0.0257 (-0.41)	-	-	-	-
DIFF_CAR (one period lagged)	-	-	-0.0334 (-3.02)	-0.00480 (-0.42)	-0.00462 (-0.41)	-0.0227 (-1.64)
State variables						
y_{it-1}=1	-	-	-	1.437 (9.20)	1.441 (9.27)	0.106 (0.29)
y_{it-1}=2	-	-	-	2.286 (12.11)	2.236 (10.76)	0.356 (0.60)
y_{it-1}=3	-	-	-	2.588 (11.10)	2.627 (10.09)	0.567 (0.91)
y_{it-1}=4	-	-	-	4.477 (18.70)	4.532 (18.39)	0.634 (0.48)
Initial state variables						
y_{i1}=1	-	-	-	-	-	2.171 (3.99)
y_{i1}=2	-	-	-	-	-	2.742 (3.77)
y_{i1}=3	-	-	-	-	-	2.494 (3.31)
y_{i1}=4	-	-	-	-	-	5.908 (2.49)

Generalised residual						
e_i	-	-	-	-	0.00441 (0.41)	-
Threshold parameters						
τ_1	0.859 (13.42)	2.029 (13.75)	2.011 (14.27)	1.344 (11.63)	1.365 (11.65)	1.933 (11.25)
τ_2	1.330 (17.35)	3.174 (19.61)	3.142 (19.39)	2.072 (13.89)	2.111 (13.83)	2.989 (12.82)
τ_3	1.928 (18.90)	4.538 (30.05)	4.541 (30.44)	3.065 (17.02)	3.127 (16.72)	4.363 (18.49)
Random effect						
σ_u	-	2.362 (9.31)	2.618 (12.01)	0.963×10^{-9} (0.00)	0.846×10^{-8} (0.00)	1.242 (4.82)
$\rho = \sigma_u^2 / (\sigma_\varepsilon^2 + \sigma_u^2)$	-	0.848	0.873	0.000	0.000	0.607
Goodness of fit statistics						
L(β) (Log lik.with τ & β)	-558.5	-441.2	-438.8	-405.3	-400.5	-380.8
L(C) (Log lik.with τ)	-696.7	-696.7	-696.7	-696.7	-696.7	-696.7
-2[L(C)- L(β)]	276.4	511.0	515.8	582.8	592.4	631.8
Degrees of freedom	9	10	10	14	15	23
No. observations	558	558	558	558	558	558

Table 4 Comparison of observed and predicted run patterns

Run patterns			**Observed**	**Model predictions**			
Wave 2	Wave 3	Wave 4		6. Pooled model	7. Random effects with LOS	8. Random effects with lagged LOS	12. Wooldridge state dependence
0	0	0	87	142	114	115	109
0	0	1	8	1	2	2	0
0	0	2	2	0	0	0	0
0	1	0	3	0	0	1	0
0	1	1	9	3	1	18	6
0	1	2	0	0	0	0	0
0	2	0	0	0	0	0	0
0	2	1	2	0	0	0	0
0	2	2	1	0	1	2	0
1	0	0	3	0	2	1	0
1	0	1	3	0	0	0	1
1	0	2	0	0	0	0	0
1	1	0	1	0	0	0	0
1	1	1	20	18	36	14	33
1	1	2	1	0	0	0	3
1	2	0	0	0	0	0	0
1	2	1	2	0	0	0	1
1	2	2	4	0	0	4	2
2	0	0	0	1	1	1	0
2	0	1	0	0	0	0	0
2	0	2	1	0	0	0	0
2	1	0	0	0	0	0	0
2	1	1	1	0	0	0	1
2	1	2	2	0	0	0	0
2	2	0	1	0	0	0	0
2	2	1	0	0	0	0	0
2	2	2	35	21	29	28	30
Total			186	186	186	186	186

Note: Bus use frequencies of 2, 3 and 4 are all recoded as 2 to assist comprehension