HOW CAN WE MODEL TRAVEL FREQUENCY? A CRITICAL REVIEW OF CURRENT PRACTICE

Kaveh Jahanshahi
Rand Europe
Andrew Daly
Rand Europe
Bhanu Patruni
Rand Europe
Charlene Rohr
Rand Europe

1. INTRODUCTION

In forecasting travel demand, a key part are the models predicting the numbers of trips or tours that travellers will make. It is the output of these models that forms the input to models of mode, destination etc. choice which give the detailed demand for travel on specific parts of the transport networks.

Early work on trip generation, as it was called, used methods such as growth factors, linear regression at zonal or household level, or category analysis at household or person level (see the presentation by Ortúzar and Willumsen, 2011). But these methods, even though they differ from each other, can all be criticised on three important grounds:

- first, they treat trip generation as a mechanical process and do not give any recognition to the behavioural basis that causes people to travel;
- second, by aggregating the data they lose sight of its nature, in that observations of numbers of trips or tours can take only the values 0, 1, 2 etc. and not negative or fractional values;
- third, again because of aggregation, they lose both statistical efficiency and insight into the variations across socio-economic dimensions that drive differences in travel frequency.

These criticisms have led analysts to look instead at 'travel frequency choice', developing models that take explicit account of the behavioural basis of the choice whether or not to travel and of the 'count' nature of the data that is generated by this process (i.e. non-negative integers).

An approach that takes account of the choice nature of this behaviour was set out by Daly (1997) and Daly and Miller (2006), introducing the 'stop-go' model and setting it in the context of utility maximisation so that it could be linked properly with the models of mode and destination choice etc.. This link is central, in particular, to the recognition and estimation of the role of accessibility in determining travel frequency. A different approach, using negative binomial models that also deal properly with count data, was introduced by Jahanshahi et al. (2009). These different approaches, together with others noted by those authors, pose the question of the relationship of these approaches and their links to behaviour.

A recent paper by Paleti (2016) sets a number of count models in the utility maximising framework, so that it can be claimed that several count models, dealing properly with disaggregate data, are consistent with reasonable models of behaviour. However, not all of the models that can reasonably be considered are discussed by Paleti, while it is also not clear how accessibility can be incorporated in all of these models.

In the following section of this paper, we extend the model types beyond those considered by Paleti, providing mathematical specifications for two general types of model: those that assume that the entire population can choose whether or not to travel and then model the number of trips made by those who do choose to travel; and those that assume that a sub-population is not able to travel and that the remainder choose how many trips to make.

Section 3 focusses on the different insights given by the different models and their suitability for forecasting. In particular, we present different parameter values obtained with different model specifications to predict the trip frequency for rail travel and discuss the implications of these.

A final section summarises the findings and makes recommendations for work in practice and further research.

2. MODEL SPECIFICATIONS

This section provides a comparison between Hurdle type count models (Mullahy, 1986), such as the stop-go model for trip frequency analysis (Daly, 1997) and zero-inflated count models (more specifically zero-inflated Poisson, ZIP, and negative binomial models, ZINB); both Hurdle and zero- inflated count models contain two trip generation processes: binary and count model components. However, Hurdle models separate the modelling of zeros from that of counts (i.e. the count model is truncated at 0) while zero-inflated models predict zero counts from both binary and count processes. Conceptually, the Hurdle models assume that individuals make a decision on whether to make a trip or not and then the number of trips to be made, conditional on making at least one trip; the zero inflated count models, however, separate a subset of observations originating from a subpopulation (or state) that can only have zero trips (known as “structural” zeros); the remaining zeros come from the count model subpopulation (we call these “functional zeros”).

Below we provide mathematical specifications for the two model types:

2.1 Stop-go hurdle model

The stop-go model (Daly, 1997, gives the model specification and reports on a number of previous studies using it) first predicts the probability of making a trip (or not) and then the choice of the number of trips being made conditional on making at least one trip. As explained in Daly (1997), the first stage (i.e. probability of making a trip) is predicted by the logit model as:

$$\mathrm{Pr}(y_i > 0) = \frac{1}{1+\exp(-v_i)} \quad \text{or} \quad \mathrm{Pr}(y_i = 0) = \frac{1}{1+\exp(v_i)} \tag{1}$$

where v_i is the utility function of making one or more trips given by:

$$v_i = \gamma_0 + \gamma_1 z_{1i} + \cdots + \gamma_m z_{mi} = Z_i' G \tag{2}$$

where Z_i is the matrix of explanatory variables (i.e. behavioural components such as household socioeconomic characteristics etc) and G is the array of unknown coefficients which need to be estimated.

The next stage involves estimating the number of trips conditional on $\mathrm{Pr}(y_i > 0)$; the standard stop-go uses a recursive approach to estimate number of trips: first the choice is made on whether exactly 1 trip or 2 or more trips will be made; then, given that 2+ trips are to be made, the choice is whether exactly 2 trips or 3+ trips is to be made, etc.

Assuming $\mathrm{Pr(stop)}$ is the probability of stopping at every level (i.e. stop at 1 relative to go for 1+ trips, stop at 2 relative to go for 2+ trips, etc) and is the same for all levels, the stop-go process becomes a geometric count model. The probability of making y_i trips will be then given by:

$$\mathrm{Pr}(y_i | y_i > 0) = \mathrm{Pr}(y_i > 0)\, \mathrm{Pr(stop)}\, (1 - \mathrm{Pr(stop)})^{y_i - 1} \tag{3}$$

where

$$\Pr(\text{stop}) = \frac{1}{1+\exp(v_{gi})} \quad (4)$$

and v_{gi} is the utility of 'go' (i.e. making more trips, relative to 'stop' at certain level) given by:

$$v_{gi} = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} = X_i'B \quad (5)$$

In equations (2) and (5), z_{i1} to z_{im} are a set of m regressor variables (vector of Z_i) explaining the probability of making one or more trips and x_{i1} to x_{ik} are a set of k explanatory variables (vector of X_i) forming the utility of making more trips (go) when an individual makes 1 or more trips. Note that the Zs and the Xs may or may not include terms in common.

Considering $\exp(v_{gi})$ as the expected value (μ_i) in a geometric count model[1], we can also write equation 3 as:

$$\Pr(y_i|y_i > 0) = \Pr(y_i > 0)\left(\frac{1}{1+\mu_i}\right)\left(\frac{\mu_i}{1+\mu_i}\right)^{y_i-1} \quad (6)$$

where the γ's and β's are to be estimated by maximising the loglikelihood function LL_1:

$$LL_1 = \sum_{i=1}^{n} LL_{1i}$$

where

$$LL_{1i} = \ln\left(\frac{1}{1+\exp(Z_i'G)}\right) \qquad \text{if } y_i = 0$$

$$= \ln\left(\frac{1}{1+\exp(-Z_i'G)}\right) + \ln\left(\frac{1}{1+\exp(X_i'B)}\right) + (y_i - 1)\ln\left(\frac{1}{1+\exp(-X_i'B)}\right) \qquad \text{if } y_i > 0$$

(7)

In these models v_i and v_{gi} are the utilities of making one or making more trips than the current number of trips (respectively). Accessibility can be incorporated in these functions, for example with a logsum from a mode-destination choice model.

2.2 Zero-inflated count models

The probability distribution of general zero-inflated count models can be written as:

$$\Pr(y_i = t) = \begin{cases} (1 - p_i) + (p_i)g(y_i = 0), & \text{if } t = 0 \\ (p_i)g(y_i), & \text{if } t > 0 \end{cases} \quad (8)$$

where p_i is the probability of trips being generated from the count model (i.e. $1 - p_i$ is the probability of structural zeros) given by the logistic formulation below[2]:

$$p_i = \frac{\lambda_i}{1+\lambda_i} \quad (9)$$

and $g(y_i)$ is the probability distribution function of the count model given by:

For ZIP: $$g(y_i) = \Pr(Y = y_i|\mu_i) = \frac{\mu_i^{y_i}}{\Gamma(y_i+1)}\exp(-\mu_i) \quad (10)$$

For ZINB: $$g(y_i) = \Pr(Y = y_i|\mu_i, \alpha) = \frac{\Gamma\left(\frac{1}{\alpha}+y_i\right)}{\Gamma(y_i+1)\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \quad (11)$$

where Γ is the standard 'gamma' function.

In equations (8) to (11), λ_i and μ_i are the expected (positive) values given by the exponential linking functions below:

$$\lambda_i = \exp(\gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \cdots + \gamma_m z_{mi}) = \exp(Z_i'G) \quad (12)$$

$$\mu_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}) = \exp(X_i'B) \quad (13)$$

where z_1 to z_m are a set of m regressor variables (vector of Z_i) explaining the probability of structural zeros and x_1 to x_k are a set of k explanatory variables (vector of X_i) for the number of trips (including 0) generated by the count model. The parameters γ, β and α, are estimated by maximizing the log likelihood function (see equation (17) below for the log likelihood function LL_2 of ZINB).

ZINB is a more general form of ZIP which relaxes the assumption of the same value for the mean and variance in the Poisson count model[3]. The term $\frac{1}{\alpha}$ in equation 11 above is called the dispersion parameter indicating the degree of difference between the mean and variance in the negative binomial count model. Equation 11 collapses to equation 10 for large values of $\frac{1}{\alpha}$ (i.e. when α approaches zero).

Also, when α tends to 1, equation 11 can be written as:

$$g(y_i) = \Pr(Y = y_i|\mu_i) = \left(\frac{1}{1+\mu_i}\right)\left(\frac{\mu_i}{1+\mu_i}\right)^{y_i} \quad (14)$$

which is the geometric distribution of the number of trips with μ_i representing the expected number of trips which is an exponential function of attributes.

As such equation (8) above when $\alpha = 1$ can be written as:

$$\Pr(y_i = t) = \begin{cases} (1 - p_i) + p_i\left(\frac{1}{1+\mu_i}\right), & \text{if } t = 0 \quad \text{(a)} \\ (p_i)\left(\frac{1}{1+\mu_i}\right)\left(\frac{\mu_i}{1+\mu_i}\right)^{y_i}, & \text{if } t > 0 \quad \text{(b)} \end{cases} \quad (15)$$

The first part of equation 15a (i.e. $1 - p_i$) is the probability of generating structural zeros and the second part of 15a is the probability of making no trips in the count model (i.e. $\left(\frac{1}{1+\mu_i}\right) = \left(\frac{e^0}{e^0+e^{X_i{}'B}}\right)$). The sum of the two parts corresponds to the hurdle type stop-go model where structural and functional zeros are not separated.

The differences between equation (15b) and equation 6 are:

1. p_i in equation (15b) is the probability of generating trips from the count model (as opposed to the probability of structural zeros) while $\Pr(y_i > 0)$ in equation (6) is the probability of making at least one trip and as such $\Pr(y_i > 0) < p_i$.
2. In equation (6) the count model is truncated at 0; hence we have the exponent $y_i - 1$ when y_i starts from 1. Equation (15b), however, can take zero trips generated from the count model, so we have y_i which starts from 0.

The points above can be shown mathematically as:

1. Solving for p_i in equation (15a) we obtain:

   $$p_i = (1 - p_0)\left(\frac{\mu_i}{1 + \mu_i}\right)^{-1}$$

   where, p_0 is $\Pr(y_i = 0)$: the probability of making zero trips (either structural or functional);

2. Substituting $\mathrm{p_i}$ in (15b) would give:

$$\Pr(\mathrm{y_i} > 0) = (1 - \mathrm{p_0})\left(\frac{1}{1+\mu_i}\right)\left(\frac{\mu_i}{1+\mu_i}\right)^{\mathrm{y_i}-1} \qquad (16)$$

Equation (16) is equivalent to equation (6) which is the probability of making $\mathrm{y_i}$ trips when at least one trip is made in the stop-go model. This means that the standard stop-go model (where the specification and the coefficients in $\mathrm{v_{gi}}$in equation 5 are the same for each number of trips) is a simplified version of ZINB with the logistic regression truncated at 0 and with the negative binomial distribution dispersion parameter fixed to one.

The loglikelihood function of the more general ZINB can be given as:

$$\mathrm{LL_2} = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{LL_{2i}}$$

$$LL_{2i} = \sum_{i=1}^{n} \left\{ \ln\left(\frac{1}{1+exp(Z_i'G)}\right) + \ln\left(1 + \left(\frac{1}{1+\alpha exp(X_i'B)}\right)^{\frac{1}{\alpha}} \exp(Z_i'G)\right)\right\} \qquad if\ y_i = 0$$

$$LL_{2i} = \sum_{i=1}^{n} \left\{ \ln\left(\frac{1}{1+exp(-Z_i'G)}\right) + \ln\Gamma\left(\frac{1}{\alpha}+y_i\right) - \ln\Gamma(y_i+1) - \ln\Gamma\left(\frac{1}{\alpha}\right) + \left(\frac{1}{\alpha}\right)\ln\left(\frac{1}{1+\alpha\ exp(X_i'B)}\right) + y_i \ln\left(1 - \frac{1}{1+\alpha \exp(X_i'B)}\right) \right\} \qquad if\ y_i > 0$$

(17)

Daly (1997) highlights the necessity of linking the count models to the utility maximization framework in modelling trip generation. It is already shown that the probability of zero trips being generated from the count model in ZINB are given by logistic regression (see equation 9); for the probability distribution function of the count model (i.e. Negative Binomial model), we are using Paleti (2016) specifications:

Following Paleti (2016), the negative binomial probability function (i.e. equation 11) can be written as:

$$\mathrm{g(y_i)} = \Pr(\mathrm{Y} = \mathrm{y_i}|\mu_i, \alpha) = \frac{\dfrac{\Gamma\left(\frac{1}{\alpha}+\mathrm{y_i}\right)}{\Gamma(\mathrm{y_i}+1)\Gamma\left(\frac{1}{\alpha}\right)}\left(\dfrac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{\mathrm{y_i}}}{\sum_{\mathrm{k=0}}^{\infty}\dfrac{\Gamma\left(\frac{1}{\alpha}+\mathrm{k}\right)}{\Gamma(\mathrm{k}+1)\Gamma\left(\frac{1}{\alpha}\right)}\left(\dfrac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{\mathrm{k}}} = \frac{\mathrm{e}^{\mathrm{v_i}}}{\sum_{\mathrm{k=0}}^{\infty}\mathrm{e}^{\mathrm{v_k}}}$$

(19)

where $v_k = \ln\left(\frac{\Gamma\left(\frac{1}{\alpha}+k\right)}{\Gamma(k+1)\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^k\right)$ and μ_i is the expected value of count shown in equation 13.

It can also be shown that the stop-go model (using geometric series expansion) can be written as:

$$\Pr(Y = y_i) = \Pr(go)^{y_i-1}\big(1 - \Pr(go)\big) = \frac{\Pr(go)^{y_i-1}}{\frac{1}{(1-\Pr(go))}} = \frac{\Pr(go)^{y_i-1}}{\sum_{k=1}^{\infty}\Pr(go)^{k-1}}$$

$$= \frac{e^{v_i}}{\sum_{k=0}^{\infty} e^{v_k}}$$

(20)

where $v_k = \ln\left(\frac{\mu_i}{1+\mu_i}\right)^{k-1}$ and $\mu_i = \exp(v_{gi})$ (see equation 5).

Again, it can be seen that equation 19 simplifies to equation 20 when $\alpha = 1$.

Moreover, the way in which terms for accessibility can be incorporated in the stop-go model can be emulated in the ZINB model (and hence in ZIP).

3. CASE STUDY

To gain practical understanding of the influence of analysing methods on modelling trip frequency, this section compares appliocation of three altenative models – i.e. stop-go, ZIP, and ZINB. For the case study, we use the UK National Travel Survey (NTS) data which Patruni et al (2017) have assembled for analysing rail trip frequency for commuting and business tours. However, to maintain compatibility, we do not use the full model presented in Patruni et al., because of its constants for specific numbers of tours, which are difficult to emulate in the other model types.

We use identical variables for all three alternative models to ensure like with like comparison. As shown in Section 2, the models are nested and as such we are able to compare directly their loglikelihood; we also compare more general goodness of fit measures (AIC for this paper).

Table 1 and Table 2 below present the findings from comparing the three model forms for commuting and home based business purposes. Panels 1a and 2a compare the loglikelihood and goodness of fit statistics. The Akaike Information Criterion (AIC) which is reported here deals with the trade-off between the goodness of fit and the complexity of the model. Panels 1b and 2b show the influences on the choice of zero tours for stop-go and structural zero tours for ZIP and ZINB; panels 1c and 2c report the influences on number of tours made. The last shaded row in each table reports the dispersion factor which is estimated in the ZINB models.

Table 1 reveals some interesting results. For commuting rail tours – where the dispersion factor is estimated to be zero – ZIP and ZINB report the same loglikelihood value, with an AIC which is slightly better for ZIP; the marginal improvement is due to ZIP having one fewer estimated variable (i.e. the dispersion factor), which is insignificant in ZINB. This finding is in line with our expectations; zero dispersion factor reduces ZINB to a ZIP model; and we note also the identical estimation of coefficients in panel 1b and 1c.

The stop-go model shows weaker fit to the data with systematically over-estimated influences on trip frequency (see panel 1c).

Table 1 Comparison of model results across three models for commuting rail frequency

	Stop-Go	ZIP	ZINB
Panel 1a Goodness of fit statistics			
Number of Observations	125,658	125,658	125,658
Log-Likelihood	-27,864	-26,945	-26,945
Parameters estimated	24	24	25
AIC	55,777	53,938	53,940
Panel 1b. Influence of socioeconomic profile on the choice of zero			
Personal Income	-0.021***	-0.021***	-0.021***
Household Income	-0.018*	-0.018*	-0.018*
Number of HH Cars	0.415***	0.416***	0.416***
Company car in the household	0.142**	0.117	0.117
Full driving licence	0.505***	0.507***	0.507***
Full time worker	-0.767***	-0.704***	-0.704***
Part time worker	-0.378***	-0.378***	-0.378***
year of survey	-0.028***	-0.028***	-0.028***
Working in manufacturing sector	0.770***	0.772***	0.772***
Working in wholesale business	0.490***	0.491***	0.491***
Working in finance sector	-0.778***	-0.772***	-0.772***
Working in health/social care sector	0.58***	0.58***	0.58***
Age	0.022***	0.022***	0.022***
No car in HH	0.185***	0.183***	0.183***
People in managerial, professional and administrative occupations	-1.069***	-1.071***	-1.071***
Intercept	4.16***	4.07***	4.07***
Panel 1c. Influence of socioeconomic profile on the trip frequency			
Full time worker	0.400***	0.329***	0.329***
Working in finance sector	0.120**	0.095***	0.095***
Age	0.002	0.002	0.002
Company car in the household	-0.174**	-0.14***	-0.14***
Age below 25	0.116**	0.101**	0.101**
Age between 26 to 35	0.097***	0.081***	0.081***
Intercept	0.35**	0.749***	0.749***
Dispersion factor	N/A	N/A	0.000

*** significant at 99% confidence level, ** 95%, * 90%.

For home based business rail tours (Table 2) the dispersion factor is estimated to be significant and well above 1. Here ZINB shows the best fit to the data followed by the stop-go model; ZIP has the weakest fit. These results indicate that the goodness of fit has a direct relationship with the dispersion factor. Where the data is highly over-dispersed, the stop-go model, which assumes a dispersion factor of 1, fits better to the data than ZIP which assumes a dispersion factor of zero.

The differences across the three model structures are also observable from panel 2b and 2c which shows bigger variations across the models.

Table 2 Comparison of model results across three models for home-based business rail frequency

	Stop-Go	ZIP	ZINB
Panel 2a Goodness of fit statistics			
Number of Observations	125,658	125,658	125,658
Log-Likelihood	-12,243	-12,326	-12,191
Parameters estimated	21	21	22
AIC	24,529	24,695	24,428
Panel 2b. Influence of socioeconomic profile on the choice of zero			
Personal Income	-0.022***	-0.025***	-0.039***
Male	-0.126	0.155	0.309***
Number of HH Cars	0.174***	0.187***	0.244***
Company car in the household	-0.315***	-0.816***	-0.823***
Full driving licence	0.053	0.056	0.078
Full time worker	0.327***	0.319***	0.282***
Part time worker	-0.035***	-0.007	-0.005
year of survey	0.430***	0.464***	0.608***
Working in manufacturing sector	0.502***	0.52***	0.581***
Working in wholesale sector	0.452***	0.481***	0.606***
Working in construction b sector	0.210**	0.228**	0.295**
Working in health/social care sector	-0.393***	-0.443***	-0.65***
Working in real estate sector	0.002	-0.006*	-0.004
Age	0.330***	0.363***	0.518***
Full access to car in HH	0.321***	0.319***	0.282***
People in managerial and professional occupations	-1.355***	-1.372***	-1.489***
Intercept	5.277***	4.539***	3.263***
Panel 2c. Influence of socioeconomic profile on the trip frequency			
Male	0.370***	0.348***	0.405***
Time	0.039***	0.037***	0.037***
Age	-0.011***	-0.011***	-0.009**
Company car in the household	-0.668***	-0.557***	-0.384***
Intercept	-1.035***	-0.431**	-1.795***
Dispersion factor	N/A	N/A	3.153***

*** significant at 99% confidence level, ** 95%, * 90%.

To better understand the scale of differences in practice, we undertook tests to evaluate the influence of a 10% increase in personal income on the probability of making business rail tours and the estimated number of tours. As the reference case, we evaluate a typical full time, managerial and professional male working in finance sector and facing car competition in their household (i.e. fewer cars than licence holders). We also consider the sample average income (i.e. £26.5k p.a.), age (i.e. 41 years old), and number of household cars (i.e. 1.7) in 2006. The results are shown in Table 3; panel 3a shows changes in the probability of making tours and panel 3b reports the percentage change in estimated number of tours as a result of 10% increase in personal income. It can be observed that in both cases, ZINB is more elastic than the other two models; ZINB shows 0.24% increase in the probability of making tours and 8.3% increase in the estimated number of tours when personal income is increased by 10%. These values are 0.19% and 5.9% for stop-go and 0.14% and 6.43% for ZIP.

Table 3 Comparison across three models –test of 10% increase in income on business rail tour-making

	Stop-Go	ZIP	ZINB
Panel 3a- Changes in probability of making tours for 10% increase in income			
Probability of making no tours	96.76%	97.89%	97.17%
Probability of making no tours - after 10% increase in individual income	96.57%	97.75%	96.93%
Change in probability	-0.19%	-0.14%	-0.24%
Panel 3b – changes in number of tours for 10% increase in income			
Estimated number of tours (in year 2008)	0.039	0.026	0.031
Estimated number of tours after 10% increase in individual income	0.041	0.027	0.034
%age change in estimated number of tours as a result of 10% increase in income	5.88%	6.43%	8.35%

We further estimate changes in the probability of making no business rail tours by individual income ranging from £14k to £90k (see Figure 1). It can be observed that the differences across the models increase by increase in income level.

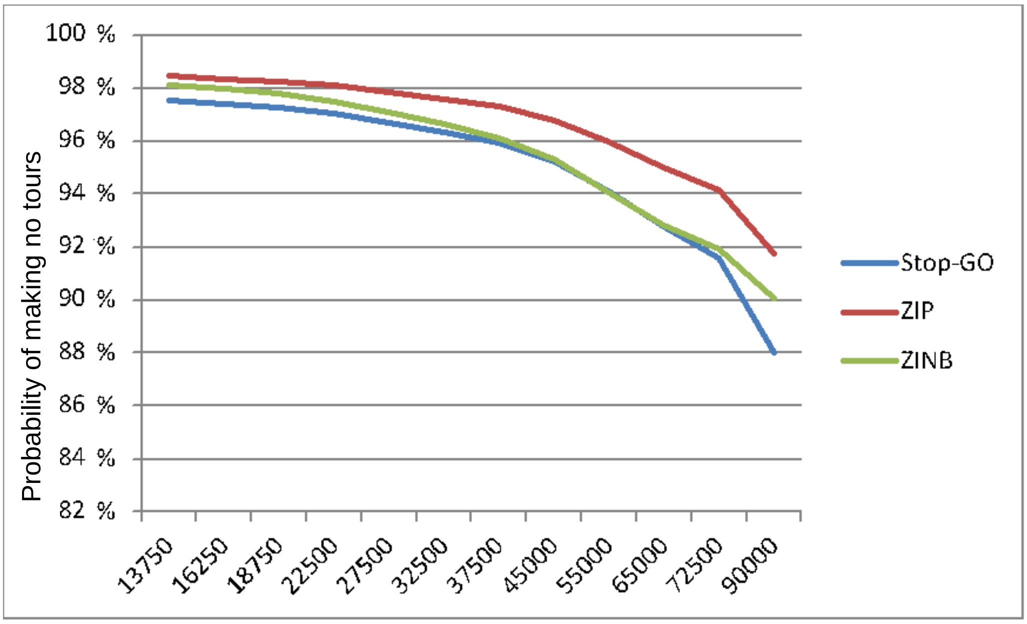

Figure 1 Changes in probability of making no rail tours by income (business trip)

4. DISCUSSION AND CONCLUSIONS

This paper is motivated by the existence of different models for the prediction of the numbers of trips or tours made by individuals or households. The methods used in early transport planning exercises are rejected as they do not reflect that trip-making is the result of individual travellers' choices. Further, they do not reflect the nature of the data, specifically that the number of trips are integers, i.e. 0, 1, 2 etc. and not a fractional or negative number. Nevertheless, there remain several different approaches, including the 'stop-go' model, Poisson regression and negative binomial models, each of which can use a 'hurdle' or 'zero inflated' approach to account for the fact that observations of zero trips are more frequent than would be indicated by models without these features.

A recent paper by Paleti (2016) sets a number of count models in the utility maximising framework, which is particularly important because it allows accessibility to be included in the model as an explanatory variable, coming out of utility-maximising models of mode and destination choice. Moreover, setting the models in the utility maximising framework gives a basis for appraisals using the models, along the lines of Daly and Miller (2006). It is emphasised that the three models reviewed here fit this criteria.

The contribution of this paper is to make a detailed comparison of three of these models: the stop-go model (Daly, 1997), zero-inflated Poisson and zero-inflated negative binomial (Jahanshahi et al., 2009). It is shown analytically that zero-inflated Poisson (ZIP) relates closely to the stop-go model, providing the stop-go model formulation is not adjusted to replicate specific numbers of trips (e.g. that commuters are likely to go to work 5 times per week). Zero-inflated negative binomial (ZINB) is a generalisation of ZIP.

Tests of the three models on data previously analysed by Patruni et al. (2017), but with simplified implementations of the stop-go model to make it more comparable with ZIP and ZINB, illustrate these findings. On both data sets ZINB gives the best fit in terms of likelihood, but on commuter data it is no better than the simpler ZIP. On the commute data stop-go performs less well than ZIP, but on the business data it performs better, findings which relate to the distribution of numbers of tours. Specifically, the dispersion factor in ZINB is effectively constrained to 1 in the stop-go model and to 0 in ZIP; thus when the true value is closer to one of these extremes, the corresponding simpler model will perform better.

The elasticity tests and comparisons of predicted number of tours confirm analytical findings of differences in model outcomes when the distribution of data is different to that assumed by ZIP and Stop-Go. When the dispersion factor is larger than 1, ZINB is more elastic to changes in behavioural parameters

In future work we would recommend that methods be found for modelling in ZINB and ZIP where specific numbers of trips attract more or less choices than would be expected. Further, the identification with utility maximisation should be made closer, so that accessibility can be included and appraisal can be undertaken. It is our intention to work on these points in the near future.

BIBLIOGRAPHY

Daly, A. (1997) Improved methods for trip generation, Transportation planning methods Volume 11. Proceedings of seminar F held at PTRC European Transport Forum, Brunel University, England, P415.

Daly, A., Miller, S. (2006) Advances in modelling traffic generation, presented to European Transport Conference, Strasbourg.

Jahanshahi, K., Williams, I. N., Hao, X. (2009) Understanding travel behaviour and factors affecting trip rates, European Transport Conference, Leeuwenhorstm, the Netherlands.

Mullahy, J. (1986) Specification and testing of some modified count data models, Journal of Econometrics, 33, 341-365.

Paleti, R. (2016) Generalized Extreme Value models for count data: Application to worker telecommuting frequency choices, Transportation Research Part B: Methodological, 83, 104-120.

Patruni, B., Daly, A., Rohr, C. (2017). Understanding demand for rail travel in the UK, European Transport Conference, Barcelona.

NOTES

[1] Please note that the distribution of number of trips is completely geometric (including zeros), if Pr(y=0) = Pr(stop). The standard-go stop specification is geometric FLEX(0), where FLEX is a notation used by Paleti (2016) to capture terms pertaining to specific counts.

[2] It is also possible to use probit regression for estimating the probability of structural zeros

[3] See Jahanshahi et al (2009) for more discussion on the difference between Poisson and negative binomial regressions.