Revealed Preference Mode and Destination Choice Modelling: A Comparison of Household Survey and Census Estimations

Dr William Brazil
Adjunct Assistant Professor/Transport Modeller
Trinity College Dublin/National Transport Authority

Dr Peadar Ó Súilleabháin
Senior Transport Modeller
National Transport Authority

Mr Barry Colleary
Head of Modelling/ Adjunct Assistant Professor
National Transport Authority/ Trinity College Dublin

Dr Páraic Carroll
Assistant Professor
University College Dublin

Mr John Nott
Senior Transport Planner
National Transport Authority

Dr James Fox
Senior Associate Director
Jacobs

Abstract

The National Transport Authority's Regional Modelling System (RMS) provides strategic modelling coverage for the entire Irish state, via the five regional multimodal modes. These models can broadly be considered to fall within the structure of the classic four stage model. However, the most recent iteration of the model departs from this structure, and instead utilises a combined mode and destination multinomial choice model. This multi-regional and highly segmented estimation provides trip purpose and region specific generalised time and mode specific preference parameters.

This paper outlines the estimation of these parameters, based upon the use of revealed preference observations from both the census derived POWSCAR dataset and the Authority's National Household Travel Survey (NHTS), for the Eastern Region of Ireland. We provide a comparison of model parameters estimated from travel survey and census data, across four commuting segments, highlighting the similarities and differences between the respective estimated parameters.

Using segments of in excess of fifty thousand observations, a data set of over 1.5 million revealed preference observations, and nine thousand unique mode-destination choice pairs, this study presents the results of what may be the largest transport choice modelling exercise ever undertaken within an Irish context. In addition we highlight the fundamental response characteristics of the most comprehensive multimodal transport model within the state, while also enabling a comparison of parameters estimated from population and sample level data.

Results provide mode specific generalised time sensitivity parameters, highlighting communalities between non-public transport modes, and differences with respect to mode specific constants, noting how such constants vary across segments and data sources.

1. Introduction to Regional Modelling System

The National Transport Authority (NTA) updates and maintains extensive strategic forecasting modelling assets, primary of which is the Regional Modelling System (RMS). This system can be considered to be analogous to the traditional four stage model, with the National Demand Forecasting Model (NDFM) largely undertaking trip generation, and the region specific Regional Multimodal Models undertaking trip distribution, mode choice, and assignments. While the NDFM operates on a national scale, the other three stages are estimated and applied for geographical regions, specifically: the Easter, Western, Mid-Western, South-Western, and South-Eastern regions. These regions are centred upon the major urban settings within the Irish state. Figure 1 outlines the geographic distribution of zones and models.

The RMS operates upon the principle of generalised cost/time, whereby all of the measureable costs and travel time incurred by an individual travelling between a given origin and destination are measured and quantified. These generalised time estimates are based upon network flow estimations, as well as previous studies providing conversion factors for components, such as public transport crowding or parking charges, into units of generalised time (gen. minutes).

As with any modelling system, it is not possible to quantify every element that informs an individual's mode or destination choice, and to account for this mode constant terms are used to ensure model calibration. These constants also account for individual preferences for specific modes.

This paper presents the results the estimation of both the mode/alternative specific constants (ASC) and the generalised time sensitivity parameters for the Eastern Region of Ireland. The estimation of these parameters is vital to the performance of the overall RMS, as in essence, the coding of any my service or piece of infrastructure into the model, can fundamentally be considered to be an alteration of the generalised time equilibrium. The response to this change is controlled by the estimated parameters, and is therefore the main driving force accounting for the modelled changes in mode, route, and destination share.

This paper is structured as follows

- An outline of the NTA models generally, and the Mode and Destination Choice models in particular
- An overview of the data used for the purposes of model parameter estimation
- The methods used, including the model structure and assumptions
- The results of the commute and education purpose parameter estimations for the Eastern Regional Model
- A discussion of the results and their impacts for choice modellers in transport model estimation

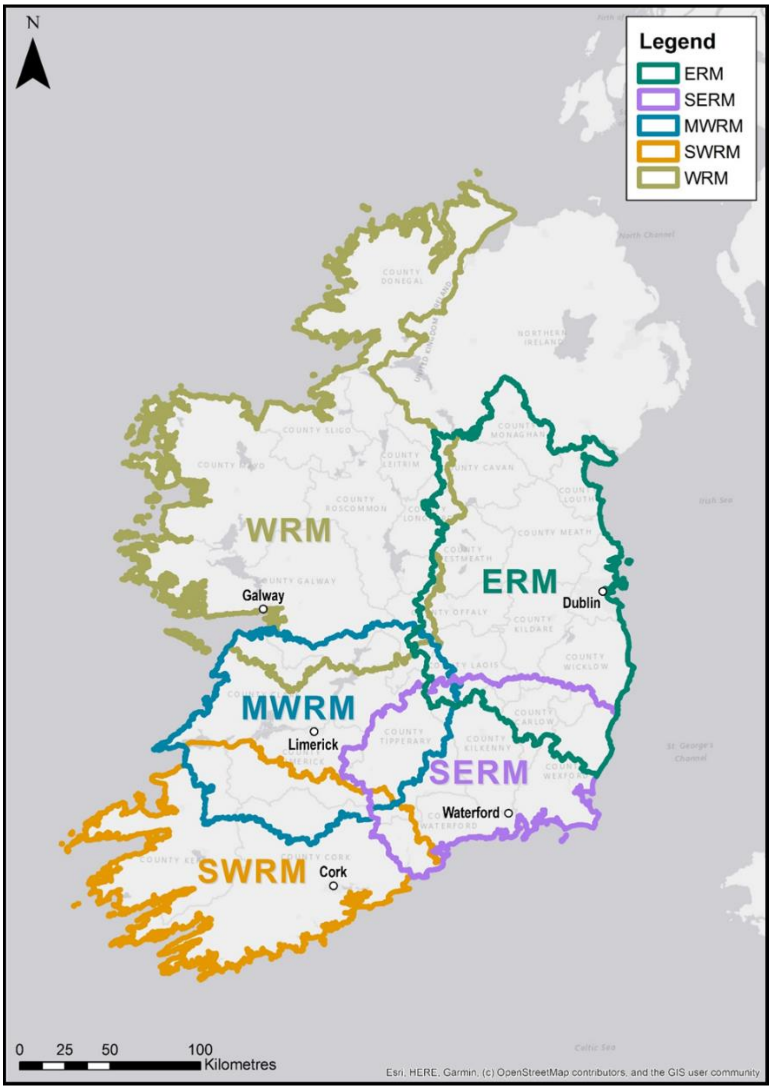

Figure 1: Regional Multimodal Models

2. Model Structure and Choice Model

The Regional Modelling System operates region specific multimodal models that are best understood through the lens of the tradition four stage model. However, the NTA RMS operates a combined mode and destination choice model in estimation form, essentially undertaking a simultaneous trip distribution and mode choice estimation. Each mode destination pair is represented as a unique choice, defined by its dual properties, rather than a nested approach whereby either mode is selected first and then destination, or destination and then mode. So for example, travelling to Zone 1 by Bus, Zone 2 by Bus, Zone 1 by Car, and Zone 2 by Car are all separate and unique alternatives available to the trip maker.

Within the Eastern Regional Model, there are 1907 destination zones and 4 estimated modes, resulting in 7,628 unique choices being available. The choice model is estimated in a highly segmented fashion, with 33 unique demand segments being represented, taking account of both trip purposes and car availability. Mode and destination choice parameters are estimated for each of these segments. This segmentation is designed to reflect the implementation of parameters in the regional multimodal models, taking account for known differences in travel behaviours in terms of value of time and propensity to take certain modes, based upon trip purpose.

Access to a car, as either a passenger or driver, has long been known to play a major role in determining the mode choice of individuals. If an individual does not have access to a car, regardless of the associated utility, it is assumed that they cannot take this mode. To account for this, many strategic transport models undertake segmentation of trips based upon factors such as car ownership and car availability. The NTA's modelling system models car availability on a trip level, informed by the levels of car ownership and household car competition within the origin zone, based upon CSA level data from sources such the Central Statistics Office's SAPS and POWSCAR datasets, and the Authority's National Household Travel Survey.

Separate estimations are carried out for the various socioeconomic cohorts and trip purposes based up estimated zonal levels of car availability. For example commute trips, are sub-divided both based on the traveller's profession being classified as either Blue or White Collar, and the levels of car availability present within the origin zone.

The MDC utilises a standard multinomial choice model (MNL). Long accepted as the workhorse of the choice modelling sector (Yates & Mackay, 2006), the MNL has the advantage of being a relatively efficient model to estimate when using large datasets, in contrast to more advanced choice models such as Mixed Logit. Latent Choice, or Latent Variable choice models. For details of the choice model is provided in Section 4. Given that ERM segments contain the aforementioned 7,628 unique choices, some segments contained revealed preference observations numbering in excess of half a million records, and required numerous iterations in respond to changes in network costs during model calibration, the requirement for efficiency in the estimation was considered to be more important than the increased sophistication offered by the more advanced models.

3.Data

The Place of Work, School or College - Census of Anonymised Records (POWSCAR) is a dataset constructed from the Irish Census of 2016, and represents one of the most comprehensive datasets for transport within the state.

Of specific interest for this work, the dataset contains data regarding trip origins, destinations, the attractions located in each zone, and the respondent's usual mode of transport. Treating this data as revealed preference observations, each POWSCAR record can be considered to be a choice to travel to a given destination via a certain mode of transport, and also a choice not to travel to all other destinations via the modes available.

As this data is taken from the census, rather than a transport specific survey, the data is somewhat limited, in that mode choice only refers to the stated usual mode, with no data regarding frequency or variability. This limitation is balanced by the comprehensive nature of the database, and the near universal level of coverage it provides across the Irish state.

For more details on POWSCAR please see: https://www.cso.ie/en/census/census2016reports/powscar/.

The National Household Travel Survey (NHTS) is a household level travel survey undertaken by the National Transport Authority in Ireland, for the purposes of gathering travel data, supporting the Authority's main functions, and enabling the calibration and re-calibration of the Authority's suite of transport models.

The NHTS is a two part survey designed to have national coverages, and consisting of a household level survey collecting data on the characteristics of the household, and a travel diary collecting data on specific trips made by the household members over a 3 day period. The NHTS can be considered to be a relatively standard household travel survey and diary, as carried out in numerous similar jurisdictions.

The survey includes 5,906 households, 10,289 diaries, and 62,307 trip records, of which 2,956 are commuting trips falling within the Eastern Region, and being wholly contained in that region. Inter-regional trips are not considered for estimation purposes and are accounted for in another part of the regional modelling system. The 2017 iteration of the survey utilises a sampling approach based upon the RANSAM methodology for Ireland (Whelan, 1979), a well-established method of collecting nationally representative observations within the Irish state.

For more information on the NHTS, the authors refer the reader to https://www.nationaltransport.ie/planning-and-investment/transport-modelling/national-household-travel-survey/.

Table 1 outlines a comparison between the two respective data sources for the commuter trips within the Eastern Region of the Regional Modelling System. POWSCAR is population level data and the main advantage of this dataset is its size, as it contains the vast majority of commuter trips made in Ireland. However, as this is census data rather than transport specific revealed preference data, the mode choice relates to their usual mode of transport, rather than being trip specific. In addition, the census only provides citizens with the ability to select one mode, resulting in the inability to collect data on multimodal trips.

Table 1:Dataset Comparison

POWSCAR Eastern Region	NHTS Eastern Region
Population Data	Survey Data
Individual Level	Trip Level
Usual Mode	Trip Specific
Single Mode	Multi Modal
Commuting and Education Trips	All Trip Types
1,559,514 commute observations	2,956 commute observations
2016 Data	2017 Data
Census Derived Data	Transport Specific Survey

Table 2 outlines the sample sizes for the respective segments of commute trips. These are divided between blue and white collar jobs, due to modelling requirements, and whether or not the individual had a car available for the trip.

Table 2: Segment Comparison

Segment	POWSCAR Observations	NHTS Observations
Blue Collar Car Available	176,612	311
White Collar Car Available	1,212,017	1,975
Blue Collar Non Car Available	27,834	114
White Collar Non Car Available	143,051	556
Total	1,559,514	2,956

The utility derived by the trip maker from each unique mode destination pair can, in simplest terms, be considered to be comprised of a two components. These are the attractiveness of the destination, in terms of the jobs, educational places, shopping opportunities etc., and the generalised time/cost, expressed in units of generalised minutes, experienced when travelling from Zone I to Zone j by Mode m. Generalised time, much like utility, is an economic construct designed to account for both the tangible and measureable costs of travelling, as well as the less quantifiable ones such as crowding, boarding penalties etc. It is often referred to interchangeable as generalised time or generalised costs.

The value of generalised time is dynamic in nature, as the more trips assigned to a given route or service will impose increased generalised minutes on said route/service. As the calibration of the regional model progressed, iterative steps were taken in terms of assigning trips to modes and routes, resulting changes to the respective generalised time values. While such assignments tend to stabilise as the calibration process progresses, there is a need to account for such changes, requiring the repeated re-estimation of the model parameters.

It must be noted that the all origin-destination pairs have generalised time values associated, regardless if the mode would actually be concerned realistic. For example if there is no realistic public transport service between two zones, there will still be an assigned time, reflecting the deterrence of walking to the nearest public transport stop, even if this adds a very large generalised time element.

Estimation was repeated until the calibration has meet the pre-determined criteria and is judged to be of acceptable standard. Discussion of overall regional model calibration is beyond the scope of this paper, for more details the authors suggest referring to the regional multimodal calibration reports (NTA, 2022).

As the RMS requires the estimation of separate parameters where there was a car available to choice maker, the data is segmented in Car Available and Non Car Available segments. For the NHTS data, a question was included in the travel survey where, if the mode take was not car/van, respondents were asked to indicate if there had been a car available to them for the trip. It should be noted that car availability is defined at point of origin, therefore even if there is no destination parking a trip would still be considered car available. For the POWSCAR data this information was not present. In lieu of other information, trips originating in a household with a car present were deemed to be car available, whereas those originating in a household without a car where not. As a result of this, roughly 12% of trips in the POWSCAR data is defined as being Non Car Available, compared to 29% in the NHTS. This may be expected as in certain households simultaneous trip makers may be competing for the car, and the NHTS data captures this, while POWSCAR does not. As this difference is not a function of generalised time, destination attractions, or the intrazonal value, it was expected to be reflected in the mode specific constant.

4. Methodology

Mode and destination choice models for the RMS were estimated using an MNL based upon the theory of randomly distributed utility. The same specification was used for both the POWSCAR and NHTS estimations. This is a standard methodology in transport choice models, for further details please see (Hensher et al, 2015), (Louviere et al, 2002) for more details. Specifically, the deterministic component of the utility is considered to be derived from a given mode destination pair can be considered to be a balance between the disutility of travel associated with the mode, and purpose specific attractions (jobs and educational places) located at the destination. There is also an intrazonal element to the utility, as irrespective of generalised time, as there is often a positive utility contribution associated with remaining within your origin zone.

For the NTA regional models, the deterministic component of utility was as follows:

$$V_{md} = ASC_m + \beta_m * GenTime_{md} + Intrazonal\,(0_{not\,intra}, 1_{intra}) + \ln(att_d) \quad (1)$$

Where:

V_{md} is the utility derived from travelling to destination d by mode m

ASC_m is the mode specific utility constant for mode m

β_m is the mode specific generalised time sensitivity parameter

$GenTime_{md}$ is the generalised time in minutes experienced travelling from a given origin zone to destination d via mode m

Intrazonal is the intrazonal constant that only enters the equation if the origin and destination are the same zone

att_d is the number of segment specific attractions in the destination zone

Equation 2 outlines the relationship between the utility derived from a given mode and destination pair, and the probability of it being selected, with respect to the utility of the other mode and destinations.

$$P_{mdi} = \frac{e^{v_{mdi}}}{\sum_{j=1}^{J} e^{v_{mdj}}} \quad (2)$$

Where:

P_{mdj} is the probability of selecting mode and destination pair i

V_{mdi} is the deterministic component of the utility for mode and destination pair i

J is the number of mode and destination pairs within the choice set

In the implementation of the mode and destination choice model in the ERM, this probability translates into the proportion of overall trips within a given segment that is assigned to the mode and destination pair.

The mode and destination choice model within the ERM is based upon the concept of mode specific generalised time, with this variable being the primary driver of mode and destination changes within the model. Table 3 outlines the components considered as part of the generalised time equation for each of the respective modes within the model estimation. This also highlights the relative asymmetry, in terms of the complexity of the generalised time equations, for the respective modes. While this is common across the industry, it does result in the better representation of some modes than others. Generalised times values are estimated from the ERM assignment model, and therefore must be updated following the re-estimation of the choice parameters. The generalised time values for a given OD pair is identical for a given mode-segment between POWSCAR and NHTS estimations. The generalised time equation contains its own specific parameters, but as these are well established features of transport forecasting, they are considered outside the scope of this paper, other than being acknowledged as forming an input for the choice model estimation.

Table 3: Generalised Time Components

Public Transport	Car	Cycling	Walking
Access Time	Parking	Distance	Distance
Egress Time	Operating Costs	Route Quality	Segment Speed
Wait Time	Tolls	Segment Speed	
In Vehicle Time	Distance		
Fare	Travel Time		
Boarding Penalty	Parking Search Time		
Transfer Penalty			
Crowding Penalty			

It should be noted that public transport generalised functions are the most complex in the modelling process, and as they included access and egress legs, some of these trips can actually be considered to be multimodal.

For each segments and journey purposes, the utility derived by travelling to any destination zone by a given mode can be understood to be a trade-off between the generalised time incurred when travelling and attractions located in that zone. Attractions are comprised of the number of segment appropriate of job and educational places located in destination zone. These attractions are calculated based upon equations derived from the POWSCAR data. The utility equation considers the natural log of the number of destination attractions, as once this enters Equation 1, it has the effect of ensuring that the probability of selecting a given zone is proportional to its attractions. The same attractions were used in both the POWSCAR and NHTS estimations.

5. Results

Results are provided in terms of a comparison between the two distinct estimations (POSWCAR and NHTS) and the generalised time sensitivity parameters and the Alternative Specific Constants.

Table 4 outlines the respective generalised time parameters $\beta_{m,}$ for each of the respective demand segments and the four modes available. First considering car, values of between -0.082 and -0.070 have been estimated across the two segments and two data sources, with the NHTS Blue Collar segment being the greatest outlier. For Public Transport values of between -0.018 and -0.021 are estimated from the POWSCAR data, while a greater range of between -0.015 and -0.032 are produced from the survey data. Walking generalised time sensitivities are estimated to be between -0.066 and -0.078 for the various POWSCAR segments and between -0.055 and -0.084 for the NHTS. Finally cycling parameters are estimated to be -0.068 and -0.75 for POWSCAR and -0.019 and -0.072 for NHTS.

In general, there seems to be agreement between the parameters estimated for the various modes and data sources for car and walking, with estimated values falling close to -0.07, and this is also reflected for the POWSCAR cycling estimations, but less so for cycling in the NHTS. In contrast, while estimated parameters for public transport are somewhat similar between the data sources, with values roughly centred around -0.020, these are markedly and consistently different than those estimated for the other three modes. These may be a result of public transport being more complex to more, and the parameters for the other modes mainly being driven by sensitivity to travel time alone.

Table 4: Generalised Time Parameters

Generalised Time Sensitivity Parameters	POWSCAR/Census	NHTS/Survey
Car Blue Collar Car Available	-0.072	-0.082
Car White Collar Car Available	-0.070	-0.070
PT Blue Collar Car Available	-0.020	-0.020
PT White Collar Car Available	-0.021	-0.015
PT Blue Collar Non Car Available	-0.019	-0.032
PT White Collar Non Car Available	-0.018	-0.022
Walk Blue Collar Car Available	-0.066	-0.080
Walk White Collar Car Available	-0.078	-0.055
Walk Blue Collar Non Car Available	-0.069	-0.084
Walk White Collar Non Car Available	-0.075	-0.084
Cycle Blue Collar Car Available	-0.071	-0.025
Cycle White Collar Car Available	-0.068	-0.019
Cycle Blue Collar Non Car Available	-0.073	-0.126
Cycle White Collar Non Car Available	-0.075	-0.072

Table 4 provides the ASC values estimated for all segments and for both data sources. ASC estimation requires a reference category as it is a dummy variable, and therefore values are presented relative to public transport (the reference mode). Positive values represent greater perceived utility for a given mode with respect to public transport and negative values lesser perceived utility, all other things being equal.

Car ASCs are considerably higher for the NHTS segments than POWSCAR with values of 3.353 and 3.110 for Blue Collar and White Collar respectively. However, as the NHTS has a stricter definition of Car Availability, it is likely some trips in the POWSCAR segment were actually not car available in reality, and therefore this is accounted for in the model in terms of a smaller utility contribution from the Car ASC. In essence, Car appears less attractive in POWSCAR than NHTS, as it contains trips where Car was not actually an option. This appears to be one drawback of using census derived data, in comparison to travel survey observations.

While public transport is the reference category, and therefore all ASCs are set to zero, it is still possible to calculate the relative ASC for public transport with respect to other modes. For example when the Car ASC equals 1.023, then the public transport ASC is -1.023 with respect to car.

Walk ASC estimates are relatively consistent across segments for POWSCAR, with values of between 1.553 and 1.750, but vary more for the NHTS estimations with a range of between 0.9 and 3.863. In all but the NHTS Car Available segments, walk has the largest positive associated ASC value.

Cycle ASC are estimated to be negative in all segments, and it is also the only mode that produces a negative ASC with respect to the reference mode, public transport. This indicates that, ceteris paribus, cycling is the mode with the lowest level of preference, given the generalised time specifications used.

Table 5: Alternative Specific Constants

Alterative Specific Constants	POWSCAR/Census	NHTS/Survey
Car Blue Collar Car Available	1.023	3.353
Car White Collar Car Available	0.687	3.110
PT Blue Collar Car Available	0.000	0.000
PT White Collar Car Available	0.000	0.000
PT BC Blue Collar Non Car Available	0.000	0.000
PT White Collar Non Car Available	0.000	0.000
Walk Blue Collar Car Available	1.553	3.863
Walk White Collar Car Available	1.676	0.900
Walk Blue Collar Non Car Available	1.716	1.279
Walk White Collar Non Car Available	1.750	1.797
Cycle Blue Collar Car Available	-0.675	-1.400
Cycle White Collar Car Available	-1.003	-2.327
Cycle Blue Collar Non Car Available	-0.695	-1.140
Cycle White Collar Non Car Available	-0.931	-0.390

As the ASC enters the utility equation as a constant, it is possible to take the exponent of the ASCs and estimate a mode specific propensity for each segment and dataset. This represents the likelihood of selecting a given mode all other things being equal. As Public Transport is the reference mode, and the exponent of 0 is 1, results should be interpreted with to a higher or lower underlying propensity than is observed for Public Transport. This is presented in Table 6. In this format it is easier to see the markedly difference estimates produced between the NHTS and POWSCAR datasets. While some of these differences way be accounted for by considering the different definitions of car availability used, it is unclear as how that would impact estimates for cycling and walking.

Table 6: Alternative Specific Propensities

Alterative Specific Propensities	POWSCAR/Census	NHTS/Survey
Car Blue Collar Car Available	2.78	28.59
Car White Collar Car Available	1.99	22.42
PT Blue Collar Car Available	1.00	1.00
PT White Collar Car Available	1.00	1.00
PT BC Blue Collar Non Car Available	1.00	1.00
PT White Collar Non Car Available	1.00	1.00
Walk Blue Collar Car Available	4.73	47.61
Walk White Collar Car Available	5.34	2.46
Walk Blue Collar Non Car Available	5.56	3.59
Walk White Collar Non Car Available	5.75	6.03
Cycle Blue Collar Car Available	-0.675	0.25
Cycle White Collar Car Available	-1.003	0.10
Cycle Blue Collar Non Car Available	-0.695	0.32
Cycle White Collar Non Car Available	-0.931	0.68

6. Conclusions

This paper presents the results of the mode and destination choice models for the Eastern Region of the NTA's Regional Modelling System, and compares the results of parameter estimations undertaken using both transport specific survey data and travel data derived from the Irish census of 2016.

These estimations represent both the very large scale application of traditional choice modelling methods to revealed preference choice sets, specifically utilising census derived population level data, as well as a comparison of such an application with a more tradition travel survey based approach.

Results of the commute segment analysis demonstrate relatively consistent responses to generalised times across socio-economic class and car availability status for walking, cycling, and driving, with a smaller degree of sensitivity observed for changes in public transport values. These parameter estimates were found to be relatively stable across all segments for the three non-public transport modes, with values of approximately -0.07 emerging. This may be a results of the relatively simple composition of the generalised time values associated with these modes, as it is mainly dominated by travel time and distance. In contrast, generalised time parameters for public transport are estimated to be, in general, closer to -0.02. This may reflect the much more complex nature of the generalised time used for public transport, where elements such as boarding penalties, transfer penalties, and crowding must be accounted for. In terms of the comparison between POWSCAR and NHTS estimations, there is a general agreement between the parameters estimated. The only area where a major discrepancy occurs is with regard to cycling, which is by far the least popular mode within the study area.

The second parameter set of interest in this work is the ASC, or mode specific constant. The ASC accounts for the elements of travel that are not captured by the generalised time estimates. Overall, there appears to be a preference within the Irish population for walking and driving, followed by public transport, and finally cycling. These estimates are found to be less consistent across the segments and data sources than the generalised time parameters. It should be noted, that for car travel at least, this is likely a function of the necessary, but sub-optimal, need to use different Car Availability definitions, highlighting one of the drawbacks associated with non-transport specific census data.

Overall the estimations presented in this paper demonstrate that travel survey data has the ability to estimate parameters that are generally consistent with population level travel data, when the same explanatory variables are used. This could potentially provide modellers with additional confidence regarding parameter estimations where population data is not available, suggesting that travel surveys of a sufficient sample size can produce reliable population level parameters for generalised time. As ASC/constants are more likely to be adjusted in model calibration, the discrepancy between census and survey data is less important, but should still be noted.

References

Hensher, D., Rose, J., & Greene, W. (2015). Applied Choice Analysis (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9781316136232

Jordan J. Louviere, David A. Hensher and Joffre D. (2002),Stated choice methods: analysis and application, Swait, Cambridge University Press, ISBN: 0-521-78830-7

Brendan J. Whelan, 'Ransam - random sample design for Ireland', Economic and Social Research Institute, Economic and Social Review, Vol. 10, No. 2, 1979, 1979, pp169-174

NTA Eastern Regional Calibration Report Model, available at: https://www.nationaltransport.ie/planning-and-investment/transport-modelling/regional-modelling-system/regional-multi-modal-models/east-regional-model/

Yates, J., & Mackay, D. F. (2006). Discrete Choice Modelling of Urban Housing Markets: A Critical Review and an Application. Urban Studies, 43(3), 559–581.