Journal of Transport Geography 19 (2011) 615–622
Contents lists available at ScienceDirect
Journal of Transport Geography
journal homepage: www.elsevier.com/locate/jtrangeo

R-Tresis: developing a transport model system for regional New South Wales

Stuart Bain, David A. Hensher *, Zheng Li

Institute of Transport and Logistics Studies (ITLS), Faculty of Economics and Business, The University of Sydney, NSW 2006, Australia

ARTICLE INFO

Keywords:
Regional transport
Integrated transport and land use models
Multi-modal
Spatial models
Data specification and sourcing
3SLS

ABSTRACT

This paper sets out a demand modelling framework for the development of a regional transport and land use model system (R-Tresis), to be implemented for New South Wales (Australia). Traditionally, the focus of such a model system has been major metropolitan areas such as Sydney, where we have developed Tresis (Hensher, 2002). Given the growing concern about regional accessibility to many service classes, there is a need for a modelling capability that can be used to prioritize and guide policy decisions in regions that are often described as remote, rural, low density and small town. In developing a framework that is capable of integrating both demand and supply elements of transportation and land use activity, we recognized the challenges in developing primary data sources, and the high likelihood of a reliance on secondary data sources. This suggested an alternative approach to demand modelling that was not dependent on choice models; namely a suite of continuous choice models in which we capture the actual activities undertaken by each mode on both the demand and supply side at high spatial resolution.

1. Introduction

Regional and rural multi-modal transport models are somewhat rare in contrast to metropolitan-wide systems (the latter overviewed by Wegener (2004). There are some exceptions, such as the National Dutch model, the Italian decision support system, and the National model of Thailand, all described in Daly and Sillaparchasrn (2008). In one sense these are regional models in that they represent large geographical jurisdictions, and hence are informative in evaluating policy instruments beyond metropolitan areas. No such capability exists in Australia.[1]

A number of these national models have collected primary data on passenger and freight activity, and developed a number of interconnected travel choice models using multinomial and nested logit model forms. This luxury is not available in Australia, and has required a rethink of methods to capture transportation activity in particular. Recent research by Collins et al. (2010) on the demand for air travel between airports throughout Australia, involved a set of simultaneous equations in which there was endogeneity with respect to a number of influencing factors on air travel (e.g., air fares, number of competitors and even passenger movements). Their aviation sector model system focused on four models, respectively for: passenger demand, number of competitors, fares and number of flights between each airport pair. Secondary data collected from disparate published sources provided the complete data set used to estimate a three stage least squares (3SLS) system of equations. The success of this model system provided the impetus for selecting such the 3SLS model system for each mode (i.e., car, coach, train and plane) for the regional model system developed below.

An advantage of structuring the demand and supply side in this way is that we can introduce feedback between the alternative modes, through endogenous and/or exogenous variables defined in terms of other available modes. Such a use of 3SLS is rare in transportation studies (see Ozbay et al. (2006)).

This paper sets out an initial model system for a multi-modal regional context, specifying data needs and relevant exogenous and endogenous variable required to develop scenario based applications relevant to key agendas such as accessibility and greenhouse gas emissions. The following sections describe the overall model system, the approach adopted to construct relevant data given its availability at different levels of spatial resolution (i.e., tourism region vs. statistical local area vs. local government area), model estimation and interpretation. The focus is on passenger activity, although extensions can incorporate the freight sector, subject to sourcing data.

* Corresponding author.
E-mail addresses: S.Bain@itls.usyd.edu.au (S. Bain), D.Hensher@itls.usyd.edu.au (D.A. Hensher), Z.Li@itls.usyd.edu.au (Z. Li).

[1] There is relatively little research on developing regional (or rural) travel demand model systems, except the Statewide travel demand models common in the USA which are essentially simplified versions of the metropolitan models with data readily available, albeit at a higher level of spatial aggregation. Although we could cite this broad set of studies, this adds little to the challenges we face with such paucity of data at any level of spatial resolution in regional and rural Australia (the last National Travel Survey being in 1976).

0966-6923/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtrangeo.2010.06.018

2. Spatial challenges in data identification

Regional transport is described by a multi-modal network linking spatial jurisdictions throughout a pre-defined region. In general, greater spatial disaggregation is preferred to higher levels of aggregation, in terms of behavioral representation; however the decision on what level to adopt is determined primarily by a mix of data availability and application requirements. In the Australian context, secondary data is available at a number of levels, including: postcode, local government area (LGA), statistical local area (SLA), and tourism region (TR). As we go from the highest level of spatial aggregation (i.e., tourism region), to the least spatially aggregate (i.e., postcode), key data items are increasingly missing, especially car passenger trips between origins and destinations (with traffic counts of vehicles at specific locations being the only source).

Although the model system set out below is sufficiently general that it can accommodate data at any of the spatial levels, our preferred level is LGA. There are 119 zones of which the majority are LGAs, with the only exception being the four amalgamated Sydney zones and the two for the ACT (metropolitan Canberra and regional) zones. These zones were amalgamated from LGAs because the geographic size of LGAs within the Sydney metropolitan area is significantly smaller than those elsewhere in the state and the focus of trip activity related to Sydney and Canberra can be appropriately represented by the four main centroids of the Sydney metropolitan area and the one centroid for the ACT.

Hence we have a large origin–destination (O–D) matrix, 119 by 119. Each O–D pair is an observation unit for model estimation. As shown in the next section, the absence of reliable LGA by LGA trip data has required us to develop a model at the Tourist region level to obtain predictions of LGA by LGA modal trip activity. Once the modal demand is predicted down to an LGA by LGA level, we need to add in the modal attributes, at the LGA by LGA level, and the socio-demographic data (at the LGA level), to then be able to estimate the travel demand models of interest in R-Tresis.

Existing approaches used in developing O–D matrices for modal levels of service and costs common in metropolitan contexts, and some Statewide settings, were not able to be used herein due to the seriously deficient nature of data. There are five such approaches that have been used in the past in metropolitan-wide studies for obtaining information on attribute values (adapted from Hensher et al. (2005)), summarized here as background but unavailable to this study. Perhaps the most commonly practiced approach is to use network skim values obtained from transportation planning models and to use forecasted attribute values (by O–D pair) to impute unobserved attributes in the survey. Skims are generated to determine the cost of travel (e.g., time, distance) on the ‘cheapest' route from zone i to zone j. The skims are based on the computation of the minimum time paths between zones based on free-flow link speeds. The skims generated are then blended using weighted averages to replicate the actual costs or travel times. The problems with this approach include lack of sufficient sample sizes for certain O–D pairs by mode (e.g., Jou et al., 2006), lack of resolution within a traffic analysis zone (TAZ) (e.g., the bus travel time within a TAZ is assumed constant, whereas there can be tremendous variability in walk times within a TAZ, especially in suburban and rural TAZs), and lack of dynamic realism regarding congestion effects on travel time by time-of-day (e.g., traditional peak and off-peak assignments capture two time-of-day congestion effects). A second approach involves imputing unobserved attribute values with the average attribute values of observed alternatives. A major limitation of this approach is that sample sizes may be quite small or even non-existent for calculating certain origin–destination (O–D) pairs—a network with 500 one-way potential O–D pairs, for example, has 499,000 travel times to estimate in a modest four mode choice context. Moreover, imputed values are constant within an O–D pair, as travel times or distances within a TAZ are not differentiated. This method also does not preserve the variance in the underlying variable and as such produces inconsistent estimates (Brownstone, 1998). A third method involves sampling across the distribution of choosers to identify the proportion having chosen multiple alternatives, and substituting the means of these attribute values for non-chosen alternatives. Again, samples may be small or not exist for many O–D pairs, thus leaving the analyst to amend the approach with exogenous information. A fourth approach involves the use of stated attribute values, based on the notion that capturing perceptual data yields realistic behavioural models that represent the choice context in which choices are made. The limitations of the stated attribute approach include increased respondent load, and the significant challenges associated with forecasting future perceptions needed to support transportation planning activities. For example, how does one forecast what high occupancy vehicle travel time will be perceived to be across individuals in the future? A fifth approach developed by Washington et al. (2009) is to synthesise the data using known exogenous information such as travel distances or other socio-demographic characteristics, and to condition the synthesized data on these constraints. This approach, referred to as the Bayesian Imputation Multinomial Logit (BI-MNL) method, incorporates elements of Bayes' theorem, the multinomial logit choice model, and sampling based estimation to synthesise or impute unobserved data. The approach is motivated by a desire to: (1) reduce the reliance of model calibration on network skim values and their noted limitations; (2) to obtain within-zone variation in unobserved attributes, such as travel times by walking, which may be important to assess policies and programs that may differentially impact travellers within a zone, and (3) to provide a robust analytical alternative to imputation based approaches.

3. Developing a regional travel demand data capability to *Predict* LGA by LGA trip activity

The development of a regional modelling capability requires both demand- and supply-side information about transportation movements and networks. Furthermore, this information must be available in the same geographic units as used in the model. Whilst much supply-side information is publicly available (for example, census data on vehicle ownership, public transport timetables, airline route schedules, etc.), demand-side data is more difficult to obtain. The reasons for this are many: trips by personal vehicle are not centrally recorded; commercial operators may consider their patronage figures to be commercially sensitive, and governments may consider figures from nationalised operations to be politically sensitive.

In the absence of actual demand-side figures, travel surveys become one of the main methods of obtaining demand-side data. The collection of primary survey data is a costly process however, particularly if the survey is to be conducted at a high level of geospatial disaggregation. In Australia, there is extensive demand-side information available about localised transport movements (such as for journeys to work) and also information about the broader transport behaviour of persons resident in major metropolitan areas (such as from the Sydney Household Travel Survey). Financial constraints eliminated the possibility of collecting primary survey data specifically for this project, and the most recent National Travel Survey conducted in Australia occurred in 1976. The requirement for current demand-side data about long distance travel movements therefore necessitated a different approach.