A Review of the Evidence for the Temporal Transferability of Mode-Destination Models

James Fox (corresponding author)
tsjbf@leeds.ac.uk
Institute for Transport Studies
University of Leeds
United Kingdom
LS2 9JT

Stephane Hess
S.Hess@its.leeds.ac.uk
Institute for Transport Studies
University of Leeds
United Kingdom
LS2 9JT

6300 words + 2 figures + 3 tables

November 13, 2009

ABSTRACT

One of the main motivations for developing travel behaviour models is to use them to forecast future levels of transport demand. Given that the interest in transport planning is often in long term forecasts, with forecast horizons of up to thirty years, it is important to consider the transferability of travel behaviour models over time. The importance of model transferability has been recognised ever since disaggregate models were first applied in the late 1970s and early 1980s, but seems to have been largely forgotten about in recent years, where the focus has been on the development of ever more advanced models that better explain current behaviour, with a particular focus on the representation of taste heterogeneity. However, there are sufficient grounds to suspect that the model that best explains current behaviour may in fact not necessarily be the best tool for forecasting, not least due to the risk of overfitting to the base data. This paper aims to return the crucial issue of *temporal transferability* of travel demand models to the research agenda. The paper has three main components. We first discuss the notion of transferability, highlighting the potential impacts of violations of the assumption of transferability, and set out how transferability can be assessed. This is followed by what we believe to be the most complete review of existing work in this area to date. Finally, we set out a number of areas where future research should be directed.

INTRODUCTION

The main interest in the field of travel behaviour research lies in the development of models that are able to closely replicate choices made by travellers in real life settings. The development of these models has two main aims; the models are used to *understand* current travel behaviour, and the models are applied to generate *forecasts* of likely future behaviour. For the former, efforts mainly focus on producing accurate measures of willingness-to-pay indicators, e.g. for use in appraisal. For the latter, various different contexts arise, namely to forecast behaviour under hypothetical changes in scenarios (e.g. new transport infrastructure), to apply a model developed in one area to another area, and to use a model to produce long term forecasts of future behaviour. Notwithstanding the possibility of all three playing a role at the same time, it is the last of these, namely the long term forecast, that is at the heart of the issues discussed in this paper.

The importance to transport practice of producing accurate temporal forecasts should not be underestimated. Indeed, such forecasts are used by local and national government agencies to give an indication of likely future demand for the provision of transport services, and they help shape policy decisions for example in the context of new infrastructure developments. The complexity of this process is further increased by the need to take account of demographic changes, as well as the impact of changes in the transport infrastructure.

To make these forecasts, the approach that is typically followed is to develop models that represent a tractable simplification of current behaviour, and then use those models to forecast future behaviour. This means that the more advanced types of models, such as Mixed Multinomial Logit, are rarely used in this context, because the computational requirements they impose in application are too great. Indeed, while this cost may be justified in estimation, application relies on running the model in a potentially very large number of different contexts, often iteratively. The forecasting problem is further simplified by separating the key travel choice decisions on a given day, typically:

- travel frequency - whether to travel, and if so how many times

- mode of travel

- destination zone

- time of day

For each of these choices, separate models are usually developed by travel purpose, as experience has demonstrated that the factors influencing these choices vary according to travel purpose. The focus of this paper is on the mode and destination choice decisions, which may be modelled as sequential choices, or as a simultaneous choice.

In a forecasting context, mode-destination models are used to assess the effectiveness of different policies over forecasting horizons of up to 30 years. These models can include detailed socio-economic segmentation, enabling both a better fit to the estimation dataset, and an ability to predict the impact of trends in the input variables over time, such as increasing car ownership, or ageing of the population.

However, forecasting with such models relies on a significant assumption, namely that the parameters that describe behaviour in the base year can be used to predict future behaviour, an issue that is referred to as *transferability*. Over recent years, this issue has dropped off the radar, with

the majority of effort going into the development of ever more advanced models that better explain current behaviour, with a particular focus on the representation of taste heterogeneity. However, it is possible that the model that best explains current behaviour may in fact not necessarily be the best tool for forecasting, not least due to potential issues of overfitting.

The problem is that if the assumption of transferability is violated, then the future forecasts will be subject to uncertainty, irrespective of how well the models fit in the base year, how much segmentation they incorporate, and how accurately future model inputs can be forecast. As reflected in the discussions in this paper, the topic of transferability has received less and less attention in recent years. So while the use of models in forecasting remains one of the two main aims of travel behaviour research, this is not reflected in current research activity.

The issue of what is meant by *transferability* is explored further in the next Section. For the purposes of this introduction, it is useful to cite Koppelman and Wilmot (*1*), who define define a *transfer* as:

> *"...the application of a model, information, or theory about behaviour developed in one context to describe the corresponding behaviour in another context."*

This paper is concerned with the transferability of models, rather than underlying behavioural theories, in the context of model forecasting. In forecasting, models developed at one point at time are applied to predict behaviour at a future point in time. It is thus assumed that the models are *temporally transferable*, i.e. that the parameters that explain travel behaviour when the model was estimated will also explain future travel behaviour.

The aim of this paper is to return the crucial issue of *temporal transferability* of travel demand models to the research agenda. In particular, this paper examines the evidence for the transferability of mode-destination choice models, which are applied over forecasting horizons of up to 30 years, but with, as we demonstrate, little evidence for their transferability over such periods.

The paper has three main components. We first discuss the notion of transferability, highlighting the potential impacts of violations of the assumption of transferability, and set out how transferability can be assessed. This is followed by what we believe to be the most complete review of existing work in this area to date. Finally, we set out a number of areas where future research should be directed.

TRANSFERABILITY
Defining Transferability

Koppelman and Wilmot (*1*) provide the following definition of transferability which is, in the authors' view, the best definition provided in the literature:

> *"First, we define transfer as the application of a model, information, or theory about behaviour developed in one context to describe the corresponding behaviour in another context. We further define transferability as the usefulness of the transferred model, information or theory in the new context."*

The first part of this definition can be interpreted quite broadly. For example, applying a model based on principles of utility maximisation assumes that those principles apply in the context in which the model is applied, as well as in the context in which the model is developed. However,

the focus of the transferability literature, and this paper, is on *model transferability*. That is to say, assessing the ability of models developed in one context to explain behaviour in another context, under the assumption that the underlying behavioural theory on which the model is based is equally applicable in the two contexts.

Somewhat surprisingly, none of the other papers reviewed attempted to set out their own definition of transferability, and indeed in many cases the term is used under the assumption that its meaning is known to the reader.

Temporal and Spatial Transferability

A key distinction made in the literature is between *temporal transferability* and *spatial transferability*. Temporal transferability is concerned with the application of models developed using data collected at one point in time at another point in time, whereas spatial transferability is concerned with the application of models developed using data from one spatial area in another spatial area. Usually temporal transfers take place within the same spatial area, and spatial transfers take place at or around the same point in time. However, is some cases a model is transferred over both time and space and so the two categories are not mutually exclusive.

To consider temporal and spatial transferability in the context of disaggregate mode destination choice models, it is useful to define in summary form the utility functions used in the models:

$$U_{md} = \beta\, X + \varepsilon_{md} \tag{1}$$

where: U_{md} is the utility of mode-destination alternative md
β is a vector of model parameters
X is a vector of observed data
ε_{md} is the random error term

In model development, the objective is to identify model parameters that best explain the observed data. Thus, as a model is developed, and its ability to explain the observed choices increases, the term $\beta\, X$ increases in importance, and the term ε_{md} decreases in importance. Nonetheless, mode destination models do not perfectly explain the observed choices, and so some random error remains. The mean effect of this term is captured in the mode specific constants, which in a mode choice context will capture effects such as the relative reliability of modes, levels of comfort, climate and hilliness for walking and cycling, and so on.

In a spatial transfer at the same point in time, the transferability of the model will depend on the relevance of the parameters in the transfer context, for example the degree of similarity in sensitivities to travel time and cost, and on the appropriateness of the alternative specific constants. Models would be expected to be transferable for areas that have similar characteristics, such as similarities in mean travel times and costs, levels of highway and public transport reliability, climate, hilliness and so forth.

For a temporal transfer in a given area, the considerations are different. The effect of area to area differences is not present, instead the key issue is whether the parameters remain constant over time. Stated more explicitly, the issue is whether within a given population segment, the sensitivities to the different variables that form the utility functions, and the mean contribution of unmeasured effects as measured by the alternative specific constants, remain constant over time. In some instances, the ratio between model parameters is also important. For example, the value-

of-time implied by the ratio between the cost and time parameters in a model, which will change over time if there are changes in the cost and time parameters.

Thus temporal and spatial transferability are not the same thing. A model might be temporally transferable within a given area, but contain a specification that does not transfer well to other areas. Another model might contain a detailed specification that transfers well to other spatial areas, but does not transfer well over time.

Spatial transfers typically involve a *transfer sample*, a sample of choices observed in the transfer context, which may allow a locally estimated model to be developed for comparison with the model transfer. When a model is applied to forecast future behaviour, this is a transfer of the model to a new temporal context. However, unlike many spatial transfers, no transfer sample is available. There is, therefore, an important practical difference between temporal and spatial transfers.

Temporal transferability can be assessed, however, by using two datasets collected at different points in time from the same spatial area. Typically one dataset is historical, one is contemporary. Models estimated from the two samples can be compared to make assessments of model transferability, and from these, conclusions can be drawn about the temporal transferability of similar models used for forecasting. Such an assessment however also needs to take into account that while models may be *temporally* transferable in one area, this may not be the case in another area, or in a different context. From this perspective, such assessments need to be based on the use of contexts that are as similar as possible to the study context.

This paper is concerned with the temporal transferability of mode destination models over long-term forecasting horizons. It is worth emphasising that over such forecasting horizons, key model inputs, such as population, employment and travel times and costs on the networks, will be subject to considerable uncertainty, and different assumptions can have substantial impacts of the predictions of future travel behaviour. For example, a perfectly transferable model might be fed with poor predictions of input variables and consequently produce poor quality forecasts. Thus, as illustrated later on in this paper, temporal transferability is a factor in producing the best possible forecasts of future behaviour, but is certainly not the only consideration.

Conditions for Transferability

A theme in a number of the early papers on the transferability of disaggregate models was a belief that disaggregate models, which represent choice at the individual level, should be more transferable than aggregate models, which typically represent choices at the zonal level. In some cases, claims were made for the models without much supporting evidence. For example, Ben-Akiva and Atherton (*2*) claimed, in the context of spatial transferability, that:

> *"A second major advantage of the disaggregate demand modelling approach is that it is transferable from one urban area to any another. It has been hypothesised that, because disaggregate models are based on household or individual information and do not depend on any specific zone system, their coefficients should be transferable between different urban areas."*

Although the second sentence of this quote concedes transferability is a hypothesis, the first seems to treat it as a given for a transfer to any area. The argument about the zone system seems to have been made in reference to aggregate modelling approaches, which typically operate at the zonal

level, but the arguments were not set out. More generally, while a number of these early papers in the transferability literature claim that disaggregate models are more transferable than aggregate techniques, only Watson and Westin (*3*) empirically demonstrated that claim.

Later works, building on empirical findings that the disaggregate models were not always transferable, were more measured in their claims. Daly (*4*) set out three conditions for spatial transferability:

- *relevance*, does the local model give any information on travel behaviour in the transfer area?

- *validity*, is the transfer model acceptably specified for the transfer area?

- *appropriateness*, is it appropriate to use the transferred model in the target area?

Thus models are only expected to be transferable under certain circumstances.

An important finding from the literature review was that while some authors had attempted to set out conditions for spatial transferability, to the authors' knowledge no corresponding research exists for temporal transferability.

Impacts of Violation of the Transferability Assumption

If temporal transferability does not hold, what are the implications for forecasting? When the model is used to forecast future behaviour, the forecasts will be subject to error, due to differences between the model parameters and the true model parameters in the future year. This will add error into the forecasts, and the magnitude of this error would be expected to be larger the longer the forecast period, i.e. the longer the interval over which the model is transferred.

It should be emphasised that temporal transferability is not stated here as the only condition that must be satisfied to produce accurate forecasts, rather it is a factor that is is often neglected. By contrast, significant effort may go into predicting the composition of the future population, and other model inputs.

Figures 1 and 2 seek to illustrate this point. In this simple example, the base car share in 2000 is 40%, which is forecast to grow steadily to 70% by 2030. However, due to uncertainty in the input variables, the uncertainty in this prediction is $\pm 10\%$. In the first figure, the model is taken to be perfectly transferable, and therefore the overall uncertainty in the 70% mode share for car is $\pm 10\%$. In the second example, uncertainty due to input variables is again taken to be $\pm 10\%$, but uncertainty due to the transferability of the model is also $\pm 10\%$. Thus, the overall uncertainty in the forecasts is $\pm 20\%$.

What these figures aim to illustrate is that model transferability may add further uncertainty to model forecasts. One approach modellers use to deal with uncertainly in the future input variables is to run models for different scenarios, for example by running low, medium and high growth scenarios. However, understanding the uncertainty introduced into forecasting by both the input variables *and* model transferability would give a more complete picture of the true levels of uncertainty associated with future forecasts, and the the relative importance of these two effects.

Assessing Transferability

In a temporal forecast context, testing for transferability is not possible in advance. Indeed, we are producing forecasts for a future period and the accuracy of these forecasts can only

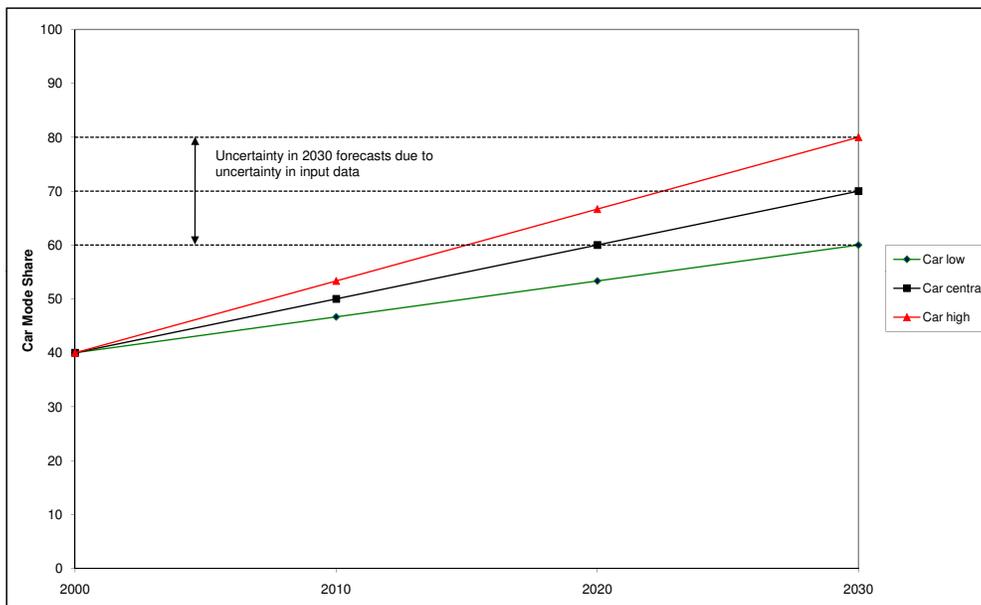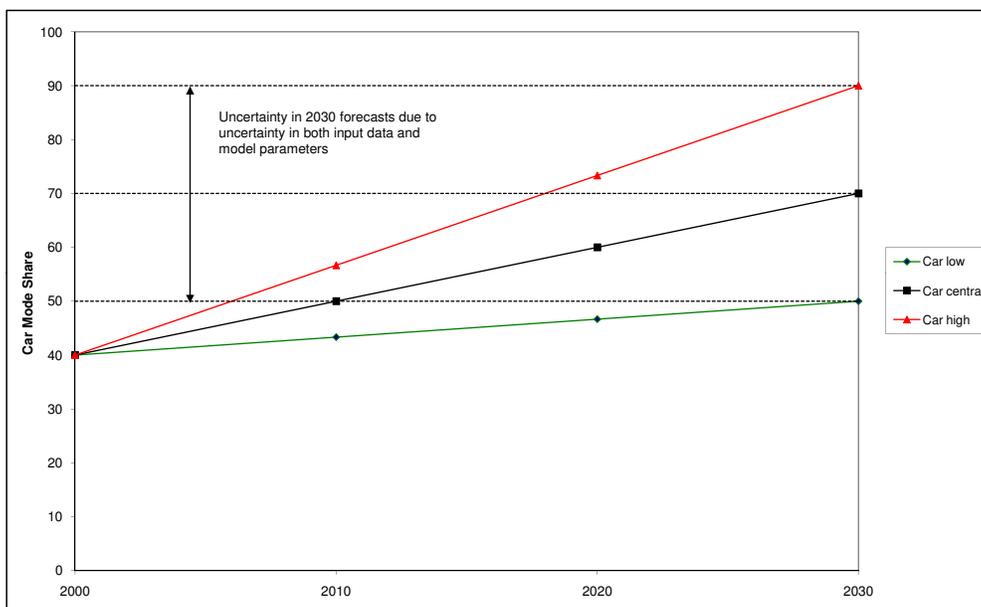Figure 1: Uncertainty in Model Inputs Only

Figure 2: Uncertainty in Model Inputs and Model Parameters

be assessed in the future. Evidence on the temporal transferability of particular types of models can however be produced by looking at historical studies, i.e. studies where we are in position to compare the forecasts to what actually occurred in reality. Specifically, temporal transferability can be assessed by using datasets that have been collected at two points in time in the same geographical area. Provided identical, or similar, variables are collected in the two cases, it is possible to use the sets of data to develop identically specified models at both points in time, and make assessments of model transferability. This generally makes the assumption that the actual *model type* is transferable, and that transferability is only influenced by the specification of the utility function. This assumption is largely based on the fact that in applied work, the same basic model structures are generally used throughout, but the possibility that different model structures may be more appropriate at different times, and the impact of this on forecasting abilities, is an interesting area for future research.

The measures of transferability used in the literature can be placed into two broad categories. First are tests of parameter equality. These represent strict statistical tests of the hypothesis of parameter transferability, and were the key measures of transferability employed in the early literature. Many of these tests rely on the availability of a transfer sample, which is used to develop a *locally estimated model*, and then the transferred model is assessed relative to this locally estimated model.

The second category is predictive measures, which are assessments of the predictive ability of a model in the transfer context. Predictive measures can be used to make assessments of model transferability, but they do not necessarily directly measure transferability, and so need to be interpreted with caution. They are however arguably less reliant on the assumption that the same model structure applies in both contexts.

Tests of Parameter Equality

A frequently used statistical test in the literature is the Transferability Test Statistic (TTS), which assesses the transferability of the base model parameters β_b in the transfer context t, under the hypothesis that the two sets of parameters are equal:

$$TTS_t(\beta_b) = -2(LL_t(\beta_b) - LL_t(\beta_t)) \tag{2}$$

where: $LL_t(\beta_b)$ is for the base model applied to the transfer data
$LL_t(\beta_t)$ is for the locally estimated model

TTS is ξ^2 distributed with degrees of freedom equal to the number of model parameters. It can be seen that this test is the same as the standard likelihood ratio test but applied to pairs of log-likelihood values in a different context.

The Transfer Index (TI) was devised Koppelman and Wilmot (1), and measures the predictive accuracy of the transferred model relative to a locally estimated model, with an upper bound of one. A reference model is used in the calculation of TI, typically a market shares model in the case of mode choice.

$$TI_t(\beta_b) = \frac{(LL_t(\beta_b) - LL_t(\beta_t^{ref}))}{(LL_t(\beta_t) - LL_t(\beta_t^{ref}))} \tag{3}$$

where: β_t^{ref} is the reference model for the transfer data; and
$LL_t(\beta_t) \geq LL_t(\beta_b) \geq LL_t(\beta_t^{ref})$

Unlike the TTS, the TI does not either accept or reject the hypothesis of model transferability. Rather it provides a relative measure of model transferability. Within a given study area, the TI can be used to directly assess different sets of models. When comparing between different studies, the TI still provides insight if the same reference model specification is used, but does not have a general scale in a formal sense.

The statistical measures discussed above are concerned with the overall fit to the data, and are the measures that have been used in the literature to assess transferability. It is also possible to analyse differences in individual parameter values, using information on the significance of the parameter in the base and transfer models. For example, the cost and time parameters in a model are key to the forecast responses to policy, and so changes in these parameters over time are of particular relevance.

Predictive Measures

As was discussed in the introduction to this section, predictive measures were increasingly used to assess transferability as the transferability literature developed. For example, Lerman (*5*) argued that the early transferability literature had used an over-restrictive definition of transferability, with an over-emphasis on statistical tests, and argued that transferability should not be seen as a binary issue, but rather that the extent of transferability should be explored. In the same book, Ben-Akiva (*6*) argued that achieving perfect transferability is impossible, as a model is never perfectly specified, and therefore pragmatic transferability criteria are required in addition to standard statistical tests.

Predictive measures need to be interpreted carefully when making assessments of model transferability. In cases where both base and transfer samples are available, then provided both datasets provide accurate samples of individual choices, the ability of the base model to predict choices in the transfer context is a direct test of the transferability of the model.

However, in many studies that validate model predictions against observed outcomes, a detailed transfer sample is not available, and the model forecasts are validated against aggregate shares. In these studies, the predictions of the model depend on the accuracy of the assumed inputs as well as the transferability of the model itself. So, a model may be highly transferable, but if fuel prices dramatically increase during the forecast period, and this was not anticipated when the future inputs where assembled, the model predictions may be some way off the observed outcomes. Care needs to be taken to distinguish input errors from transferability errors, and in some cases it is not possible to disentangle the two effects.

The relative error measure (REM) has been used in the literature to assess model transferability. It assesses for the prediction the ability of a model to predict the choice frequency in some aggregate group as follows:

$$REM_{mg} = \frac{(P_{mg} - O_{mg})}{O_{mg}} \tag{4}$$

where: P_{mg} is the prediction for alternative m in group g
O_{mg} is the observed choices for alternative m in group g

It should be noted that g is often dropped, i.e. predicted and observed alternative (e.g. mode) shares are compared. As the REM measure is self-scaling, it can be applied both to probabilities, and to aggregate choice predictions such as numbers of individuals choosing m and g.

LITERATURE REVIEW

The literature on temporal transferability has been broken down into three sub-sections. The first two sub-sections discuss studies using disaggregate mode choice models, and thus are more directly relevant than the other literature to the focus of this paper on models of mode and destination choice. The final sub-section then presents evidence from other model types, in most cases aggregate models of trip generation.

The mode choice studies are further broken down into direct tests of model transferability, where both base and transfer models have been developed allowing formal statistical tests of transferability to be made, and validation studies, where model predictions are compared to aggregate statistics on mode share, often after substantial changes to travel times and/or costs. It should be noted that these validation studies use data collected in the transfer context to define the inputs to the models, which removes the complication of combinations of errors in the input data discussed in the Predictive Measures section. A number of the papers present both comparisons of base and transfer models, and use the transfer data to validate the performance of the base model in forecasting, and so are discussed in both sub-sections.

Mode Choice Transferability Studies

Four studies of the transferability of mode choice models have been reviewed. Train (*7*) compared models developed before, in 1972, and after, in 1975, the opening of the Bay Area Rapid Transit (BART) system in San Francisco. Silman (*8*) compared a model for Tel-Aviv between 1972 and 1976. McCarthy (*9*) also analysed pre-BART data, in his case from 1973-74, with the post-BART data from 1975. Badoe and Miller (*10*) developed models from two large household interviews, collected 22 years apart. All four studies analysed home-work trips only.

In addition to these four mode choice studies, two studies have investigated the transferability of models of simultaneous mode and destination choice, the exact focus of this paper. Karasmaa and Pursula (*11*) used Helsinki data from 1981 and 1988, and Gunn (*12*) investigated models for the Netherlands using 1982 and 1995 data. Like the four mode choice studies, Karasmaa looked at home-work trips only, but Gunn ran analyses for home-work, home-shopping and home-social and recreational.

Overall, the mode choice studies supported the hypothesis that model parameters are reasonably stable over time, although this finding was not universal with two of the six studies reporting substantial changes over time. Silman and McCarthy both used the Transferability Test Statistic (TTS), and were able to accept the hypothesis of temporal parameter stability at a 10% confidence interval, although McCarthy rejected the hypothesis at a 5% confidence level. The Badoe & Miller study is noteworthy, as it the only study that considers a long term forecasting interval. Badoe & Miller rejected the hypothesis that the parameters were equal over a 22 year period, but for some model specifications Transferability Indices (TI) of almost 0.9 were obtained. Thus a transferred model from 1964 used to predict 1986 behaviour had 90% of the predictive ability of a local model estimated on 1986 data.

Neither of the mode-destination studies (Karasmaa, Gunn) calculated TTS or TI values. Gunn's findings of general parameter stability were consistent with the mode choice studies, however in Karasmaa's analysis there were significant differences between the base and transfer parameters.

Badoe and Miller made an interesting assessment of the impact of model specification on model transferability by testing seven different model specifications, ranging from simple market

shares models, and models with mode constants and level-of-service variables only, through to models with detailed market segmentation. For all model specifications, the TTS rejected the hypothesis of parameter stability at a 5% confidence interval. The TI increased from 0.132 for the simple market shares model, to 0.894 in the level-of-service variables only model, although interestingly more detailed specifications with market segmentation had lower TI values, despite higher log-likelihood values.

This finding raises an interesting question as to whether, for long term forecasting, there is an optimum level of complexity to ensure the predictive ability of the model is retained over time. It may be that adding detailed market segmentations improves the fit to the base data, but that this is a case of over-fitting, and gives less robust forecasts over the longer term.

Mode Choice Validation Studies

Parody (*13*) assessed the impact, over a one year period, of a free bus service accompanied by substantial increases in parking changes at the University of Massachusetts at Amherst. Ben-Akiva and Atherton (*2*) predicted the impact of preferential lanes on bus usage and car pooling along the Shirley Highway in Washington D.C.. Train (*7*) validated the ability of pre-BART models to predict demand when BART was introduced, and then investigated how the forecasting performance of the models varied with model specification in Train (*14*). Silman (*8*) used a model for Tel-Aviv developed using 1972 data to predict behaviour in 1976. Milthorpe (*15*)'s study had a different focus, providing a comparison of the forecasts of a four-stage model developed in the early 1970s to observed data from around 2001.

The general pattern from these studies is that the mode choice models were able to predict the impact of often substantial changes in level-of-service on mode share with reasonable accuracy. This finding is reassuring for the application of mode choice models over periods of up to five years, but it does not provide any direct evidence about the transferability of the models over the longer term.

Parody's analysis used panel data, and in one test assessed the impact of substantial increases in parking charges. In this test, a full model specification with socio-economic parameters performed substantially better than a model with level-of-service parameters alone. This suggests that an improved model specification yielded more transferable level-of-service parameters. Train's 1979 analysis also concluded that improving the model specification resulted in improvements in the model predictions.

It seems that the improvement in the predictive performance of the models that results from adding socio-economic parameters is a result of improved estimates of the key level-of-service parameters, rather than the impact of changes in socio-economics, given that most of these model tests have been undertaken over short term forecasting horizons of up to five years. These improved estimates then enable the models to better predict the impact of changes in level-of-service. Silman explicitly noted this pattern, by observing that when socio-economic parameters were added, the significance of the key cost and time variables in his models were improved.

Following from the discussion of the danger of over-fitting to the base data given above, there is clearly a need to find the appropriate balance in terms of the level of detail in the model. Adding socio-economic parameters has been found to improve the estimates of the core level-of-service parameters, however there is a danger that adding too much detail leads to over-fitting, and less robust forecasts over the longer term. Further empirical analysis of this issue would be valuable.

Other Transferability Studies

A number of other studies provide insight into the temporal transferability of models. The following paragraph summarises the various papers reviewed, and following that there is a discussion of the findings.

Hill and Dodd (*16*) used 1956 and 1964 household interviews from Toronto to assess zonal regression trip generation models. Kannel and Heathington (*17*) investigated the stability of household regression trip generation models estimated from 1964 and 1971 household surveys in Indianapolis, Indiana. Downes and Gyenes (*18*) compared the predictive performance of three trip generation techniques, zonal regression, category analysis and household regression, using data from Reading collected in 1962 and 1971. Yunker (*19*) analysed the predictive performance of trip generation and distribution models using 1963 and 1972 data from Wisconsin. Smith and Cleveland (*20*) investigated the time-stability of household regression trip generation models for Detroit, using household interviews from 1953 and 1965. Doubleday (*21*) analysed the same set of Reading data as Downes in order to investigate the temporal stability of category analysis trip generation models. Elmi and Miller (*22*) used Toronto data to investigate the temporal stability of aggregate trip distribution models, using data from 1964, 1986 and 1996. Cotrus and Shiftan (*23*) investigated the transferability of trip generation demand models using data from the 1984 and 1996/97 Israeli National Travel Habits survey.

Most of these studies are concerned with generation modelling, and typically used aggregate modelling approaches, based on regression, household classification and gravity model techniques. As such, any findings with respect to model transferability have to be interpreted with caution for the mode-destination modelling context. Nonetheless, general findings are of interest to the broader question of whether models developed at one point in time can be used to predict behaviour at a future point in time. These studies also have the advantage that they have tended to consider longer forecasting intervals, typically around 10 years, compared to the mode choice studies.

Few of these studies made formal statistical tests of model transferability. Elmi concluded that the parameters in his trip distribution models were statistically different between 1964 and 1986, although the 1964 models were able to predict 1986 behaviour well. Cotrus also rejected the hypothesis of temporal stability, both in Haifa and in Tel Aviv, over a 12/13 year period.

The assessments of the predictive performance of the generation models are supportive of the hypothesis of model transferability, with five of the seven studies reporting the models predicted future trip generations well. It should be noted however that, as discussed earlier, accurate aggregate predictions do not necessarily indicate transferability at the individual parameter level.

A noteworthy feature of many of the tests of the generation models is that the intervals of analysis often covered substantial changes in population, whereas the mode choice validation studies were typically concerned with the impact of substantial changes in travel cost and times. For example, Hill and Dodd's analysis covered a period when the population of the Greater Toronto area increased by 33%, and total car ownership rose by 45%. The good predictive performance of the models under these conditions provides some evidence for the temporal stability of socio-economic parameters that capture variation in behaviour across the population.

Elmi's analysis of work trip distribution models investigated the impact of improving the model specification, and, consistent with the mode choice studies, he concluded that improved model specification resulted in improved model transferability. Elmi obtained Transferability Indices as high as 0.84 for predicting 1996 behaviour with 1964 models, and 0.97 for predicting 1996

behaviour with 1986 models. An interesting result noted by Elmi was that the disutility of travel time reduced over time, from a value of -0.13 in 1964 to -0.08 in 1996. Elmi suggested that this reflected changes in spatial structure, and consequent increases is the mean distance to work.

Elmi's hypothesis that changes in model parameters might be related to changes in spatial structure may give an approach for forecasting how model parameters change over time. If evidence were assembled across studies of how model parameters had changed over time, it would be possible to investigate whether the changes in model parameters could be explained in terms of aggregate variables describing changes in spatial structure, such as the size of the urbanised area.

Summary and Critique

To draw the findings from the review of temporal transferability together, it is useful to summarise the key findings from the groups of studies. These summaries are presented in Tables 1 to 3 at the end of this section.

Overall, the direct tests of transferability summarised in the first table are supportive of the hypothesis that mode choice models can be transferred over time, with four of the six studies concluding the models tested were transferable. Furthermore, some of the validation studies demonstrate the models are able to predict the impact on mode share of substantial changes in level-of-service over short periods.

That said, these findings are specific to the evidence base that has been analysed. Considering the direct tests of temporal transferability summarised in the first table, it can be seen that the evidence is nearly all from commuting studies. Furthermore, all the validation studies summarised in the second table, and many of the generation studies summarised in the third table, are also based on commuter travel. Commuting travel might be expected to be more transferable than other purposes, as the journey to work is a regular trip, and as such would be expected to be accurately recorded with a higher degree of accuracy than less regular trips.

Another feature of the evidence base is that much of it is based on short-term forecast of up to 10 years. However, many transport models are applied over forecast periods of up to 30 years, and it seems reasonable to hypothesise that over longer time intervals, transferability would be less likely to be accepted. That said, the single body of evidence on longer term transferability, the studies from Toronto that developed mode choice models and distribution models, is supportive of model transferability.

An empirical finding from both mode choice and distribution studies is that improving model specification improves model transferability. Although the improvements in model specification described are often the addition of socio-economic parameters, this improvement in model performance seems to come about because the improved models provide better estimates of the key cost and time parameters that respond to short-term policy changes. Over a longer term forecasting horizon, substantial changes in the distribution of the population across segments would be expected, and so the findings in terms of model specification may be different, depending on the relative stability of level-of-service and socio-economic parameters over the longer term.

It is noted that only two studies of temporal transferability have considered simultaneous models of mode and destination choice, the focus of this particular paper. Gunn's study found a good level of temporal transferability, but in Karasmaa's work three out of four level-of-service parameters were not transferable.

The dates of the studies are noteworthy, with half (9 out of 18) published in the 1970s, and with only two papers published over the last decade. Clearly research efforts into the issue of

model transferability have been limited since the cluster of work in the 1970s and early 1980s, as already aluded to earlier. Additionally, the evidence that models of mode-destination choice are temporally transferable over forecasting intervals of up to 30 years is extremely limited. Given the importance of such long term forecasts in transport planning, this is a serious shortcoming in the field, and an important area for future research.

DIRECTIONS FOR FUTURE RESEARCH

It is clear from this review that further empirical evidence on the temporal transferability of mode-destination models would be valuable, and in particular give insight into their suitability for forecasting over longer term forecasting horizons.

Comparisons between model predictions and observed aggregate outcomes can provide valuable insight. However, it is difficult to disentangle the impact of model transferability from other factors, due in particular to the magnitude of errors in input assumptions over long term horizons. For example, Milthorpe (*15*) analysed forecasts for Sydney, Australia, from 1971 to 2001. Over this period, the population was predicted to grow by 55%, but actually grew by 35%. Clearly such errors have a substantial impact on the model predictions, irrespective of the transferability of the models.

The best approach for future research is to focus on cases where detailed disaggregate data, such as household interview data, is available over periods of 20-30 years. Provided reasonable levels of consistency exist between the data collected at each point in time, and consistent level-of-service data can be assembled for each point in time, then such datasets can be used to directly test the assumption of temporal transferability.

Three additional areas for future research are identified. The first of these goes back to the earlier point that even in studies testing for transferability, the assumption is generally made that the actual *model type* is transferable, and that transferability is only influenced by the specification of the utility function. Here, future work should look at the transferability of the actual model form in addition to the specification. Secondly, the existing literature focusses almost exclusively on the transferability of the most simplistic type of models, generally Multinomial Logit. Currently, nearly all forecasting models make of these simpler model forms, given the earlier point about the complexity and computational cost of more advanced models. However, it seems that an interesting avenue for research in this context would be to test whether the use of advanced models, such as Mixed Multinomial Logit, would improve transferability, which would make the higher computational cost more acceptable. Finally, though implicit in the background to most work, developing and applying tests for model transferability is clearly only a means to an end, and the over-arching aim of future research should be to provide guidance on how the transferability of models can be improved.

Table 1: Temporal Mode Choice Transferability Studies

Paper	Area	Purpose(s)	Time Frame	Degree of Transferability	Comments
Train (1978)	San Francisco, U.S.	Commute		LOS parameters more stable than other terms	
Silman (1981)	Tel-Aviv, Israel	Commute	4 years (1972-1976)	Good, time parameters particularly stable	
McCarthy (1982)	San Francisco, U.S.	Commute	1.5 years (1973/74-1975)	Parameters stable over short-term	Box-cox transforms used
Badoe and Miller (1995, 1998)	Toronto, Canada	Commuter mode choice	22 years (1964-1986)	Statistical differences between parameters but models broadly transferable in terms of predictive performance, ASCs and scale change over time	Level-of-service only models performed well, some highly segmented models less transferable
Karasmaa (1997)	Helsinki, Finland	Commute	7 years (1981-1988)	Poor: 3 out of 4 LOS parameters not stable	Mode-destination models
Gunn (2001)	Netherlands	Commute, personal business, shopping, social and recreational	13 years (1982-1995)	Good, particularly for level-of-service parameters	Mode-destination models, some evidence that transferability may vary with purpose

Table 2: Temporal Mode Choice Validation Studies

Paper	Area	Purpose(s)	Time Frame	Predictive Performance	Comments
Parody (1977)	Univ. of Amherst, Mass., U.S.	Commute	4 waves: 1. Autumn 72 2. Spring 73 3. Autumn 73 4. Spring 74	Good, substantial improvement when model specification improved with socio-economic terms	Large changes in modal costs over time period
Ben-Akiva and Atherton (1977)	Washington D.C., U.S., Santa Monica U.S. (application only)	Commute	D.C. 1970-74 S.M. 1974	Good in response to significant changes in LOS	Focus on car-pooling policies
Train (1978, 1979)	San Francisco, U.S.	Commute		Poor for transit due to problems with input data, predictions improve with improved model specification	Lack of info. for new BART mode, erroneous walk time data
Silman (1981)	Tel-Aviv, Israel	Commute	4 years (1972-1976)	Mixed - main car driver and bus modes predicted well, minor car passenger mode signif. over-predicted	

Table 3: Temporal Generation Model Studies

Paper	Area	Model Class	Purpose(s)	Time Frame	Evidence for Transferability?	Comments
Hill and Dodd (1966)	Toronto, Canada	Zonal regression	All purposes, all purposes peak hour	8 years (1956-1964)	Yes, after correcting for differences in data processing	Actual results after correction applied unclear
Kannel and Heathington (1973)	Indianapolis, Indiana, U.S.	Household regression	All purposes	7 years (1964-1971)	Yes Ű predicted trips within 2% of observed	Panel of household used, this may have influenced findings
Downes and Gyenes (1976)	Reading, U.K.	Zonal regression, category analysis, hh regression	All purposes, plus split into shop, work, other	9 years (1962-1971)	Yes, forecasting errors close to base year errors	
Yunker (1976)	S.E. Wisconsin, U.S.	Zonal regression analysis	Commute, shopping, other, non-home-based	9 years (1963-1972)	Good Ű predicted growth close to observed, larger differences by purpose	Observed trips grew by 25% in period
Smith and Cleveland (1976)	Detroit, Michigan, U.S.	Category analysis, hh regression	All purposes	12 years (1953-1965)	No, trip rates not stable Ű uniform growth over categories	Uniform growth likely to be income and/or accessibility
Doubleday (1977)	Reading, U.K.	Aggregate, category analysis	Regular (work) and non-regular	9 years (1962-1971)	Trip rates not stable, exception was employed males	Accessibility had an impact, and possibly income growth
Cotrus, Prashker and Shiftan (2005)	Haifa, Tel Aviv, Israel	Person level regression	All purposes	12/13 years (1984-1996/7)	Mixed, statistically rejected, but predictions good with 7% and 3% errors	

REFERENCES

[1] Koppelman, F. and C. Wilmot, Transferability Analysis of Disaggregate Choice Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 895*, Transportation Research Board of the National Academies, Washington, D.C., 1982, pp. 18–24.

[2] Ben-Akiva, M. and T. Atherton, Methodology for Short-Range Travel Demand Predictions. *Journal of Transport Economics and Policy*, Vol. 11, 1977, pp. 224–261.

[3] Watson, P. and R. Westin, Transferability of Disaggregate Mode Choice Models. *Regional Science and Urban Economics*, Vol. 5, 1975, pp. 227–249.

[4] Daly, A., *A Study of Transferability of Disaggregate Mode Choice Models from Grenoble to Nantes*. Cambridge Systematics Europe, 1985.

[5] Lerman, R., A Comment on Interspatial, Intraspatial, and Temporal Transferability. In *New Horizons in Travel-Behaviour Research* (A. M. Stopher, P and W. Brög, eds.), Lexington Books, Lexington, 1981, pp. 628–632.

[6] Ben-Akiva, M., Issues in Transferring and Updating Travel-Behaviour Models. In *New Horizons in Travel-Behaviour Research* (A. M. Stopher, P and W. Brög, eds.), Lexington Books, Lexington, 1981, pp. 665–686.

[7] Train, K., A Validation Test of a Disaggregate Mode Choice Model. *Transportation Research*, Vol. 12, 1978, pp. 167–174.

[8] Silman, L., The Time Stability of a Modal-Split model for Tel-Aviv. *Environment and Planning A*, Vol. 13, 1981, pp. 751–762.

[9] McCarthy, P., Further Evidence on the Temporal Stability of Disaggregate Travel Demand Models. *Transportation Research B*, Vol. 16, No. 4, 1982, pp. 263–278.

[10] Badoe, D. and E. Miller, Analysis of the Temporal Transferability of Disaggregate Work Trip Mode Choice Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1493*, Transportation Research Board of the National Academies, Washington, D.C., 1995, pp. 1–11.

[11] Karasmaa, N. and M. Pursula, Transferability Analysis of Disaggregate Choice Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1607*, Transportation Research Board of the National Academies, Washington, D.C., 1997, pp. 38–44.

[12] Gunn, H., Spatial and Temporal Transferability of Relationships between Travel Demand, Trip Cost and Travel Time. *Transportation Research E*, Vol. 37, 2001, pp. 163–189.

[13] Parody, T., Analysis of Predictive Qualities of Disaggregate Modal Choice Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 637*, Transportation Research Board of the National Academies, Washington, D.C., 1977, pp. 51–57.

[14] Train, K., A Comparison of the Predictive Ability of Mode Choice Models with Various Levels of Complexity. *Transportation Research A*, Vol. 13, 1979, pp. 11–16.

[15] Milthorpe, F., A Comparison of Long Term Sydney Forecasts with Actual Outcomes. *Australasian Transport Research Forum*, Vol. 28, 2008.

[16] Hill, D. and N. Dodd, Studies of Trends of Travel Between 1954 and 1964 in a Large Metropolitan Area. *Highway Research Record*, Vol. 141, 1966, pp. 1–23.

[17] Kannel, E. and K. Heathington, Temporal Stability of Trip Generation Relations. *Highway Research Record*, Vol. 472, 1973, pp. 17–27.

[18] Downes, J. and L. Gyenes, *Temporal stability and forecasting ability of trip generation models in Reading*. Transport and Road Research Laboratory, 1976.

[19] Yunker, K., Tests of the Temporal Stability of Travel Simulation Models in Southeastern Wisconsin. In *Transportation Research Record: Journal of the Transportation Research Board, No. 610*, Transportation Research Board of the National Academies, Washington, D.C., 1976, pp. 1–5.

[20] Smith, R. and D. Cleveland, Time-Stability Analysis of Trip-Generation and Predistribution Modal-Choice Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 569*, Transportation Research Board of the National Academies, Washington, D.C., 1976, pp. 76–86.

[21] Doubleday, C., Some Studies of the Temporal Stability of Person Trip Generation Models. *Transportation Research*, Vol. 11, 1977, pp. 255–263.

[22] Elmi, D. B., A. and E. Miller, Transferability Analysis of Work-Trip-Distribution Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1676*, Transportation Research Board of the National Academies, Washington, D.C., 1997, pp. 169–176.

[23] Cotrus, J. P., A. and Y. Shiftan, Spatial and Temporal Transferability of Trip Generation Demand Models in Israel. *Journal of Transportation and Statistics*, Vol. 8, No. 1, 2005.