EJTIR

Issue 16(4), 2016
pp. 573-599
ISSN: 1567-7141
tlo.tbm.tudelft.nl/ejtir

The Danish National Passenger Model – model specification and results

Jeppe Rich[1]
DTU Management Engineering, Technical University of Denmark, Denmark.

Christian Overgaard Hansen[2]
COH Aps, Copenhagen, Denmark.

The paper describes the structure of the new Danish National Passenger model and provides on this basis a general discussion of large-scale model design, cost-damping and model validation. The paper aims at providing three main contributions to the existing literature. Firstly, at the general level, the paper provides a description of a large-scale forecast model with a discussion of the linkage between population synthesis, demand and assignment. Secondly, the paper gives specific attention to model specification and in particular choice of functional form and cost-damping. Specifically we suggest a family of logarithmic spline functions and illustrate how it is applied in the model. Thirdly and finally, we evaluate model sensitivity and performance by evaluating the distance distribution and elasticities. In the paper we present results where the spline-function is compared with more traditional function types and it is indicated that the spline-function provides a better description of the data. Results are also provided in the form of a back-casting exercise where the model is tested in a back-casting scenario to 2002.

Keywords: *National passenger travel demand, discrete choice, cost-damping, back-casting, functional form.*

1. Introduction

The paper serves two objectives with the aim of improving predictive modelling practise. Firstly, at the general level it describes the design of a large-scale forecast model and the modelling principles behind the various model components and how these are linked. Although the design perspective and the linkage of models involve many non-trivial considerations such perspective is rarely covered in academic papers where the focus is predominately on methodological improvements of specific model components. Secondly, at the more specific level, we look into model specification issues and in particular that of “cost-damping” and why this arises and how it can be modelled using a logarithmic spline function. The spline function approach presented in the paper is novel and can be seen as a “super flexible” functional form, which can accommodate a flexible cost damping pattern in which the marginal disutility of cost declines with the distance. An additional topic which is elaborated in the paper is that of ‘model validation’. Specifically, we present a back-casting exercise where the model is tested backwards in time. Although this is helpful for understanding model limitations and shortcomings it is rarely done due to the significant effort required. Although the national model includes a separate freight model this is not considered in the paper.

[1] Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark, T: +45 4525 1536 F: +45 4593 6533 E: rich@dtu.dk
[2] COH ApS. Toldbodgade 37 A 2, 1253 København k, Denmark, E: coh_aps@mail.dk

1.1 Past models

Historically, past large-scale national and regional models can be divided in two epochs: The time before and after the introduction of the micro-econometric random utility framework (McFadden 1981, 1978; Ben-Akiva and Lerman, 1985). In the time before this epoch modellers applied gravity models (e.g., Wilson et al., 1977) and other meta-type models and addressed heterogeneity and substitution patterns in simplified ways. The period after, which we will consider in more details, has evolved into different schools of modelling practise. The first school may be referred to as the "European econometric school" and has materialised from early work by The Hague Consultancy Group, which in turn seems to have emerged from the development of the first version of the Dutch national model. This school predominately represents an econometric approach to model building and has inspired the model building practise in Europe as well as in the US. The second school can be referred to as the "The activity-based modelling school" which has focused on "activity-based modelling".

'The European Econometric School'

The Hague Consulting Group (HCG) was responsible for the first version of the Dutch National Model (HCG, 1990; 1992). In many ways, this model framework can be seen as a forerunner of many of the proceeding urban and national models in Europe as it applied disaggregate modelling techniques and introduced the concept of "tour-based modelling". In addition, the Dutch model applied "matrix pivoting" (Daly et al., 2005) with respect to baseline matrices and used advanced forecasting methods known as "prototypical sample enumeration techniques" as described in Daly (1998). Daly (2007) gives an excellent overview of the development of national models from the mid-seventies. The further development of the Dutch national model is today served by other entities and it is worth mentioning the work by Significance (e.g., Willigers and Bok, 2009). Other examples of national models include the models in Norway, Sweden, and in the UK for which there has been a long-standing tradition of using national models aimed at planning support at the national level. A distinction of these models is the segmentation into models for shorter and longer distances. For the Norwegian model this is described in Hamre et al. (2002), Fox et al. (2003) and in Rekdal (2009). A common challenge in the Norwegian models is the lack of reliable income information in the underlying travel survey on which the models are estimated. This is undesirable as it may affect the balance between cost and time sensitivity across different groups and modes. The Swedish SAMPERS model, which does incorporate income, has been described in Beser and Algers (2002) and has been applied for transport planning in Sweden for many years. More recently non-linear utility functions with the aim of achieving cost-damping behaviour were considered in the SAMPERS long-distance model (Börjesson, 2010; WSP, 2011). A common characteristic of the Norwegian and Swedish models is that these models do not use matrix pivoting in forecasting. The UK model (Department for Transport, 2009) represents a traditional mode, destination and frequency choice framework but seems to be limited by the nature of the data with respect to destination choice and segmentation among individuals, which in both cases, seems to be coarsely defined. The work by Rohr et al. (2010) on a long-distance model for UK represents a good reference for a joint long-distance national model for mode, destination and frequency model. In addition to the work on national models, the literature on urban and regional models is relevant. The model for West-Midland presented in Fox et al. (2014) represents a recent model framework, which in addition to the choice of mode and destination includes departure time choice. Also the model for Paris described in Tuinenga and Pieters (2006) represents a model based on the disaggregate model principles of the Dutch model and include time-of-day choice as well a matrix pivoting. In addition, Fox et al. (2003) gives a brief overview of four urban models in Sydney, Paris, Stockholm and The Netherlands (The Landelijk Model Systeem). The OTM model for Copenhagen has been particularly relevant as it represents a model for the Greater Copenhagen region of more than 2 million citizens and has been based on a largely similar data foundation as the national model, namely the TU data (Christensen and Skougaard, 2015). A general description of the OTM model is provided in Vuk and Hansen (2009) in an application for the new Metro City line in Copenhagen. More details related to the OTM demand model are provided in Fox et al. (2006) and Fox (2005).

'The activity-based modelling school'

A key reference is the work by Bowman and Ben-Akiva (2000) on the Portland model whereas a more detailed description of the activity-based model development in the US can be found in Vovsha et al. (2005). Although much work on activity-based modelling has emerged from the US, there are significant European contributions to activity-based modelling as well. This includes early work by Axhausen and Gärling (1992) and later Arentze and Timmermans (2004) on the ALBATROSS model. The latter model represents a rule-based approach to modelling rather than a random utility approach and has not been applied on a large scale for policy analysis in The Netherlands. In addition there is now a growing European literature on large-scale agent-based models, which although in a simplified way, includes activity-based demand-side principles (Balmer et al, 2006; Charypar and Nagel, 2005). The activity-based modelling tradition covers a relatively rich set of dependent variables and often operates on very detailed spatial resolution levels (Bradley, et al. 2010). Often these models apply micro-simulation techniques to derive an agent-based demand, rather than having a probabilistic allocation of demand for a set of prototypical individuals as has been practise in many European models. The Flanders model is an example of a comprehensive model framework which is based on micro-simulation (Verlinden et al., 2015; de Bok et al., 2015). It includes an advanced model for the population synthesis based on micro-simulation. Although it shares many similarities with the micro-simulation approach in the Danish National model it is more advanced in that it involves a dynamic updating of the population.

1.2 The Danish transport market in brief

It is relevant to briefly consider the composition of the Danish transport market. In Table 1 and Table 2 we present the distribution of trips and mileage across modes in the model and three aggregated trip purposes. The tables are based on baseline matrices for 2010, which has been calibrated to traffic counts. As a result, it includes the total traffic of Danish citizens and foreigners operating on the network.

Table 1. Distribution of trips between modes and main trip purposes in 2010

	Purpose			
Modes	Commuting	Business	Other	Total mode share
Walk	6%	2%	18%	14%
Bike	24%	5%	12%	15%
Car driver (car)	42%	70%	39%	41%
Car passenger (carp)	9%	16%	25%	19%
Public transport (pub)	18%	6%	7%	10%
Aviation (air)	0%	1%	0%	0%
Total	100%	100%	100%	100%

As can be noted, bike and walking represent a significant share of the national trips, although in terms of mileage, the share is much lower. For urban areas such as Copenhagen, the share of biking is much higher and will when measured for trips inside Copenhagen be as high as 50% in certain parts of the city. For public transport the pattern is somewhat similar with high market shares for urban areas and very low market penetration in rural areas.

The main transport corridor in Denmark is the East-West bound corridor across the Great Belt and across the Island of Funen. The traffic growth on the Great Belt from the opening year to present has greatly surpassed the underlying growth in the population and the economy.

Table 2. Distribution of transport mileage between modes and main trip purposes in 2010

	Purpose			
Modes	Commuting	Business	Other	Total mode share
Walk	0%	0%	1%	1%
Bike	6%	1%	3%	4%
Car as a driver (car)	58%	67%	46%	52%
Car as a passenger (carp)	10%	21%	39%	26%
Public transport (pub)	26%	8%	10%	16%
Aviation (air)	0%	4%	1%	1%
Total	100%	100%	100%	100%

Table 3. Traffic on the Great Belt corridor

Type (Million)	Before opening	1998	2010	2014
Vehicles	2.8	6.8	10.5	11.4
Rail passengers	4.4	6.7	8.4	8.6

Further projects for upgrading this corridor are being discussed and in particular an upgrade of the rail service. Hence, traffic across this corridor is highly relevant when developing and testing a national model and we will consider this in more detail with respect to the back-casting exercise in Section 5.4.

1.3 Structure of the paper

Firstly, in Section 2, we describe the overall model structure of the Danish national model and in particular the linkage between demand and supply and the model segmentation. In Section 3 we give a brief presentation of data sources and the zone system. Section 4 includes a detailed description of the model structure for the weekday transport demand model. In Section 5 we present selected model results with emphasis on parameters, elasticities and back-casting. Finally, we offer a conclusion in Section 6.

2. Overall model structure

The entire model framework consists of several separate models. This includes models for population synthesis, household synthesis, car ownership, transport demand and assignment models for the choice of route. The demand model represents choice of transport mode, destination and frequency of trips for more than 15 model segments divided by different purposes, durations and whether the trips are internal Danish trips or include a foreign destination.

The demand models are for the main parts based on a microscopic estimation and simulation framework. We start by predicting the distribution and size of the entire Danish population in a given year. This population is then grouped into households using a micro simulation approach. In a next stage we feed the population to the demand model, which then calculates (and partly simulates) demand for the entire population (a list of 5.4 million individuals grouped into 2.4 million households). The model is then iterated with a supply model (the assignment model) until convergence is obtained as illustrated in Figure 1.

In this paper we will mainly focus on the demand model and in particular the weekday demand model representing more than 99% of the trips made by Danish citizens. The Population synthesis and household simulation model is described in detail in Rich and Jensen (2015) and applies an iterative proportional fitting algorithm to fit a base matrix of prototypical individuals (Rich and Mulalic, 2012) combined with a simulation model for groupings into households. The demand models are linked with an assignment model (stochastic user equilibrium) in order to properly represent congestion effects and the fact that increased demand is counteracted by lowered accessibility due to congestion. Methods for improving system convergence have been considered in details in Rich and Nielsen (2015).

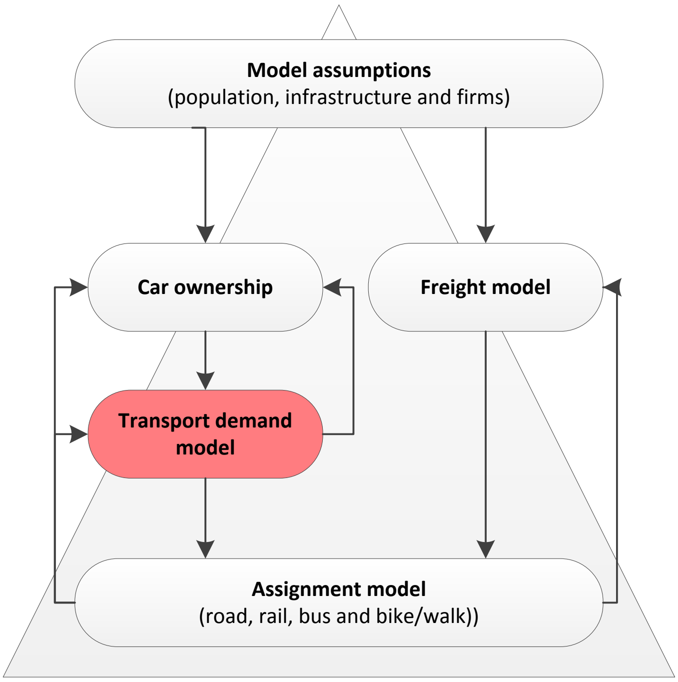

Figure 1. Overview of the model framework in the Danish National Model.

The demand model is linked in a traditional random utility framework[3] of nested logit models in which the car ownership at the upper level condition the choice process at lower levels and where lower levels feed accessibility measures (e.g. logsum variables) to the car ownership. More specifically, the choice process is decomposed into; (i) long-term choice represented by the choice of car ownership and (ii) transport related choices including trip frequency, destination choice and choice of mode. The entire passenger model framework consists of four groups of transport demand model segments as illustrated below. These are:

- The Danish weekday model represents more than 99% of the trips although in terms of mileage, it is less dominant. The weekday model consists of several sub models (14 segments) and involves trips between all parts of Denmark including rather long trips between the eastern and western parts.
- The international day model is largely focused on; (i) business transport (mainly by air), (ii) border traffic between Copenhagen and Skaane (mainly commuting), and (iii) border traffic between Southern Denmark and Northern Germany (mainly shopping and leisure).
- The overnight model is concerned with trips involving an overnight stay. These trips may be international or domestic.
- The Transit model is mainly concerned with trips from Scandinavia to Northern Germany.

[3] A set of discrete choice models where the utility of individuals in each model is decomposed into a deterministic indirect utility function and a random term.

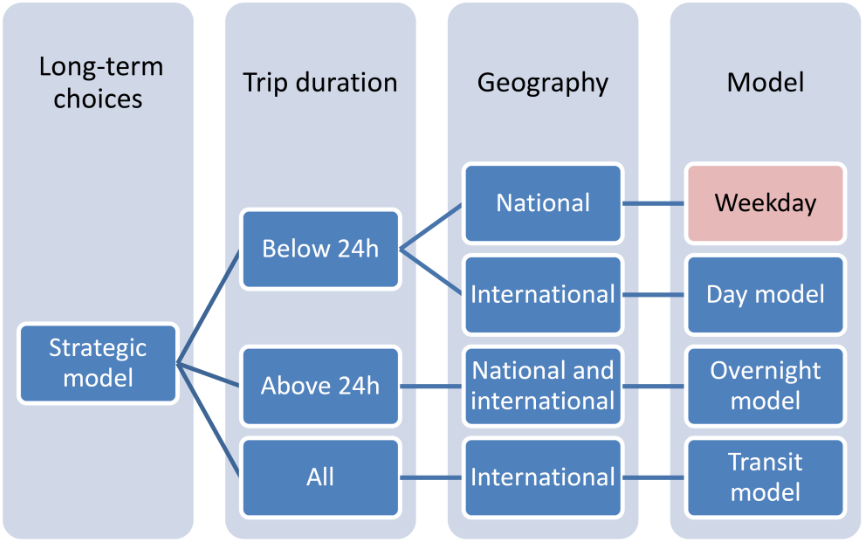

Figure 2. Overview of the model framework in the Danish National Model

It is not possible to fully describe all models in a single paper. As a result, we will focus on the weekday model. The main segmentation of the weekday model is by trip purpose and covers commute, education, escort (e.g. bringing or picking someone up), shopping, leisure and business. Business that relates to logistics is dealt with in the freight model. The modes covered are identical to the modes presented in Table 1 and Table 2.

3. Data

The demand estimation is based on a sample of the Danish national travel survey. The survey has been continuously collected over more than 15 years and provides a representative picture of the transport behaviour of the Danish population in that it monitors trip frequencies, trip chaining behaviour, time-of-day choice, destination choice (at the level of L_3 zones, see below), trip purpose, and choice of mode. A complete description of these data is available from Christensen and Skougaard (2015).

3.1 Socio-economic classification

The national travel survey contains detailed knowledge about the socio-economic characteristics of the population. However, as the model is to be used for forecasting, the socio-economic profile of individuals is limited by the details represented in the population synthesis. The structure of the synthesised population is shown below in Table 4 with further details regarding the different variable categories provided in Appendix B.

The fully spanned population matrix represents more than 4.3 million (10×2×10×6×2×2×907) entries and leaves on average 1.25 individuals per cell. Based on this matrix, which for each matrix entry represents a group of prototypical individuals, we produce a micro-list for all Danish individuals. The individuals in this list are then grouped into households using micro-simulation. For instance, we use spouse-matching in order to take into account that certain pairs are more likely to form a household than others.

Table 4. The socio-economic classification in the population synthesis

Description	Classes
Age groups	10
Gender	2
Personal income	10
Labour market association	6
Single families (dummy)	2
Children (dummy)	2
Zone ID	907

3.2 Geographical zones

The Danish zone system is defined according to four different aggregation levels, which are all internally consistent in the sense that the more disaggregate levels add up to the more aggregate levels. As shown in Table 5 and Table 6, the most aggregate level is the municipality level (L0) consisting of 98 zones, whereas the most disaggregate level (L3) consists of 3,670 zones. The current model is estimated on the L2 zone system with 907 Danish zones.

Table 5. Number of zones for the different aggregation levels

Level	Description	Zealand	Jutland/Funen	Total
L_0	Municipality level	45	53	98
L_1	Strategic level	70	106	176
L_2	***National level***	***530***	***377***	***907***
L_3	Regional level	2234	1436	3670

Table 6. Size of the zones in the different systems

Level	Region	Avg. addresses	Avg. size (km^2)
L_0	Zealand	19,522	204.5
L_1		12,919	131.5
L_2		***2,355***	***24.4***
L_3		612	15
L_0	Jutland and Funen	27,753	636.2
L_1		14,260	318.1
L_2		***2,832***	***63.6***
L_3		667	6.4

When constructing the zone system at levels 2 and 3, the aim was that the zones should be homogenous in terms of the population and work places, that the zones could be connected unambiguously to the road network, and that cities could be distinguished from rural areas (down to 3,000 addresses at level 2 and 1,000 at level 3). Major transport terminals (airports, harbours, transport centres) are defined as individual zones. Islands typically are defined as separate zones and will typically be smaller in size. If possible, zones are further classified according to land use purpose (e.g., industry, apartments and urban centres), especially at level 3. In addition, zones have been designed to conform to prior zone systems and administrative borders.

3.3 Land-use variables

One of the most important land-use variables is clearly the residential location, which is covered in the population synthesis as described in connection to Table 4. Hence, aggregation of the population table for any given year across one or several socio-economic dimensions will render the number of individuals and households per zone and possibly divided into various socio-economic classes. Land-use variables in addition to the population are represented by jobs by sector. Labour market data has been achieved from Statistics Denmark in a standard classification of 127 sectors. In the model, they have been aggregated to 36 sectors, which are further aggregated and applied in the passenger model as well as in the freight model. In the passenger model, the 36 sectors have been aggregated to 11 aggregated sector categories as shown in Table 7. In addition to this, also the population size and the geographical size of the zones can be applied as land-use variables.

Table 7. Land-use data used in the demand model

DST Sector codes	Description
45001-45002	Wholesale, car sale, car repair, ...
47001-47003	Retail shopping (daily consumables)
47004-47008	Retail shopping (non-daily consumables)
55000-56000	Hotel, restaurant, cafes, …
85001-85002	Elementary schools and high schools
85003-85004	Higher education
86001-87000	Hospital, doctors, …
88000-88000	Kindergarten and day-care, …
90000-93002	Amusement parks, fitness, museums, cinemas, …
96000-97000	Service jobs, hairdressers, laundries, ...

The sectorial division in Table 7 may seem coarse, however, as for the classification of the population there is a clear trade-off between the need for detailed information and the ability to forecast such details in a proper way. In particular it is generally a challenge to forecast a very detailed sectorial and geographical division of the labour market and rather than pretending this is possible, it is in our view better to simplify the input data.

3.4 Level-of-service

The level-of-service variables are calculated in the assignment models. These cover a range of different travel-time components including free-flow travel time and congestion time for cars and in-vehicle time, waiting time, and transfer time for public transport and air transport. For public transport, the choice of routing in the public network is determined in the assignment model. Hence, the choice whether to choose a bus or a rail line is a choice which is modelled internally in the assignment model. In the demand model, only the choice whether the main mode for the tour is public transport is determined and this is then based on a weighted average level-of-service across all public sub-modes on the given tour. From an application point of view this way of designing the model is preferable because it makes it easy to integrate and calibrate a complete public schedule (which represents all public transport modes) into the model. However, from a more theoretical point of view it could be questioned as it limits the flexibility of the utility functions somewhat.

The car assignment (and trucks) is a stochastic user equilibrium model, which due to the capacity dependence is iterated with the demand model. The public as well as the air assignments in the model are all schedule-based assignments without capacity restrictions. Hence, these are calculated only once at the start of every model run. More details on which level-of-service variables are applied in the demand model is described in Section 4 below. For walk and bike modes we use a simple all-or-nothing approach.

4. Model description

To describe the underlying mathematical model structure (refer to Table 15 for a notation list) we will start with the final output, the OD matrices, and trace the calculation of these back to the various demand models, which we will then consider in more detail.

The final OD matrix $OD_s^{Final}(i,j,m)$ for a given segment s, choice of mode m, origin i and destination j zone is fed to the assignment model in the iterative model scheme as illustrated in

Figure 1. As was illustrated in Figure 2, the model consists of four different model components. Each of these models (WD=Weekday, ON=overnight, ID=international day[4] and TR=transit) will produce OD matrices which together represents the accumulated OD demand, hence;

$$\begin{aligned} OD_s^{Final}(i,j,m) &= Q\left(OD_s^{Model}(i,j,m)\right) \\ &= Q\left(OD_{s,}^{WD}(i,j,m) + OD_s^{ON}(i,j,m) + OD_s^{ID}(i,j,m) + OD_s^{TR}(i,j,m)\right) \end{aligned} \tag{1}$$

Here Q represents the "pivot mapping" from the modelled synthetic OD matrix to a pivoted version of this matrix. The mapping Q follows Daly et al. (2005) with a specific parameterization suited for the national model and with a built-in normalisation in order to maintain model sensitivity after pivoting. In the following we will focus on the calculation of the $OD_s^{WD}(i,j,m)$ matrix in (1) which constitutes more than 99% of all trips[5].

The modelled matrix $OD_s^{WD}(i,j,m)$ is generated in the tour-based passenger model and represents a list of primary trips (PTL) and a list of secondary trips (STL) summed over individuals n and households h populated by these individuals. Hence,

$$\begin{aligned} OD_s^{WD}(i,j,m) = \sum_{h=1}^{H} W_h \sum_{n\in h} W_{n|h} \Big(& PTL_{s,n|h}(m,j|i) + PTL_{s,n|h}(m,i|j) \\ & + k\big(STL_{s,n|h}(m,j|i) + STL_{s,n|h}(m,i|j)\big)\Big) \end{aligned} \tag{2}$$

W_h is a household expansion factor used to expand a given sample population to the entire Danish population. If the model is run on a 100% sample (which is the default setting) then $W_h = 1$. If the model, however, is based on a 10% sample, W_h will be 10 on average. However, W_h is constructed in such a way that the expanded population reflects the true population exactly at the level of the municipalities and by age groups. It means that W_h will generally be different from the naïve weight in order to account for accumulated sample bias. $W_{n|h}$ is a secondary expansion factor which adjusts the weight of the individuals in order for the model to be consistent with population forecasts. It is worth stressing that the population synthesis (Rich and Jensen, 2015), which is partly based on micro simulation, is designed so that the random drawing of the population can be replicated when using the same seed number.

In (2) we refer to $PTL_{s,n|h}(m,j|i)$ as a model of trips for mode m between zone j and i. This way of expressing the model refers to the underlying probability model for choice of mode and destination conditional on the origin zone. In other words, from an estimation point of view we would consider $PTL_{s,n|h}(m,d|i)$ where d is the zone destination and i is the origin zone. Most often, the origin zone would be equal to the residential zone.

In (3) and (4) below $Pr_h(c=c')$ represents the probability model for car ownership c for household h. When the car ownership changes at the level of the household, it will then have impact on the demand because demand is conditional on the car ownership, which in (3) and (4) is weighted by $Pr_h(c=c')$. The calculation of $PTL_{s,n|h}(m,d|i)$ is given by

$$PTL_{s,n|h}(m,d|i) = \sum_{c'=1}^{C} TP_{n,s}(c=c'|i) Pr_{n,s}^{(P)}(m,d|c=c',i) Pr_h(c=c') \tag{3}$$

$$STL_{s,n|h}(m,d|i) = \sum_{c'=1}^{C} TS_{n,s}(c=c'|i) Pr_{n,s}^{(S)}(m,d|c=c',i) Pr_h(c=c') \tag{4}$$

[4] An international trip without an overnight stay.
[5] Clearly, when measured in terms of produced mileage, the mileage contribution of the other models is substantially larger as these trips are generally much longer.

In (3) and (4) $TP_{n,s}$ and $TS_{n,s}$ represent the number of primary and secondary tours as determined by the frequency models. Hence, $TP_{n,s}$ is essentially the expected sum over frequency probability outcomes $Pr_{n,s}^{(P)}(f|c)$ multiplied by the frequency f. This is done for primary and secondary tours, hence

$$TP_{n,s}(c) = \sum_{f'=1}^{F_P} Pr_{n,s}^{(P)}(f = f'|c) * f' \tag{5}$$

$$TS_{n,s}(c) = \sum_{f'=1}^{F_S} Pr_{n,s}^{(S)}(f = f'|c) * f' \tag{6}$$

4.1 Transport cost and time specification

In total, there are twelve tour models in the weekday demand model with six primary tour models for commute, education, escort (e.g. bringing or picking someone up), shopping, leisure and business and six secondary tour models with an identical segmentation. All of these are expressed in a generic model frame as will be described below.

In the estimation process the following cost functions are applied for all models.

$$cost_{n,m=car}(d|i) = c_{car}(d|i)\left[\sum_g \frac{1_{n,g}}{VoT(g)}\right] \tag{7}$$

$$cost_{n,m=carp}(d) = 0 \tag{8}$$

$$cost_{n,m=pub}(d|i) = c_{pub}(d)\left[\sum_g \frac{1_{n,g}}{VoT(g)}\right] \tag{9}$$

In (7) $c_{car}(d)$ defines the car cost to destination d from a given origin. Moreover, in (7), $1_{n,g}$ is an indicator variable indicating if individual n belongs to income group g and $VoT(g)$ represents the value-of-time for individuals in this group. So rather than expressing costs in terms of monetary units it is expressed in terms of time units as we deflate the cost component by the value-of-time $VoT(g)$. In the model, we apply a pre-defined $VoT(g)$ based on a previous value-of-time study. This is to avoid identification problems in the balancing between cost and time. The value-of-time differs between segments in that business travel is different from the other purposes.

Time for cars (and passengers) as shown in (10) is a combination of free-flow time $fft_{car}(d)$, congestion time[6] $ct_{car}(d)$, ferry-sailing-time $fst(d)$ and ferry-waiting-time $fsw(d)$. All of these attributes are calculated in an assignment model.

$$time_{car}(d) = [fft_{car}(d) + \gamma_1 ct_{car}(d)] + fst(d) + \gamma_2 fwt(d) \tag{10}$$

Public transport time (calculated in a schedule-based assignment model) consists of three weighted components, vehicle-time $Inv_{pub}(d)$, number of transfers ns_{pub} and waiting time wt_{pub}.

$$time_{pub}(d) = Inv_{pub}(d) + \vartheta_1 ns_{pub} + \vartheta_2 wt_{pub} \tag{11}$$

In the model, time and cost are joined in a generalised time measure GTT_m where the n index has been suppressed to simplify the notation.

$$GTT_m(d) = time_m(d) + cost_m(d) \tag{12}$$

In most cases we will not use $GTT_m(d)$ directly in the model but use different transformations of $GTT_m(d)$ depending on the segment and the mode. An important transformation is the natural

[6] The extra time that is caused by congestion.

logarithm, which can be combined with the linear function (Daly, 2010) into a "hybrid function" with added flexibility. Hence, essentially a form

$$G\big(GTT_m(d)\big) = \beta_1 GTT_m(d) + \beta_2 \ln(GTT_m(d)) \tag{13}$$

However, as suggested in Rich and Mabit (2015) a number of other hybrid functions may be defined, of which many will perform better than (13). For instance, it was suggested to use sequences of log-power polynomials as in (14) below.

$$G\big(GTT_m(d)\big) = \sum_{q=1}^{Q} \beta_q \ln\big(GTT_m(d)\big)^q \tag{14}$$

Later in Rich (2016) these log-power polynomials were generalised into a logarithmic spline class in order to tackle cost-damping challenges. This approach will be described in more details in Section 4.2 below.

4.2 Cost-damping and a logarithmic spline class

Cost damping involves decreasing sensitivity with respect to cost and time. As discussed in Rich and Mabit (2015) there are numerous reasons why cost damping is likely to exist. This includes human preferences not being entirely linear with distance (e.g. start-up costs), potential unobservable attributes such as car-occupancy rates, heterogeneity in the error-terms and selection bias. In its nature, cost-damping results from scaling effects which are not dealt with properly. These will tend to drive up elasticities for longer distances. There are two types of 'medicine' for such scaling effects: i) to segment the data into different scale-segments (long and short distances), or ii) to choose a functional form, which for longer trips prevents the scaling. As discussed earlier in the paper the first option has not been considered an alternative in this context. However, the second option is by no means trivial. Although equation (13) and (14) is a step in the right direction in that it mixes a damped and non-damped component it will not prevent scaling effects. The function will always be dominated by the non-damped part on longer distances. An alternative to this is to develop a function class with more structure imposed on it. An example is the spline-function in (15) below which was presented as a possible function candidate in Rich (2016).

$$\mathcal{F}\big(GTT_m(d)\big) = \sum_{q=1}^{Q} 1_q(GTT_m(d)) \left[\theta_q ln\big(GTT_m(d)\big)^{Q-q+1} + \alpha_q\right] \tag{15}$$

The indicator function $1_q(x)$ is defined such that $1_q(x) = 1 \Leftrightarrow x \in \left[c_{q-1}, c_q\right]$ and zero elsewhere. The function is connected in $Q-1$ knot points $c_1, \dots, c_{Q-1}$. The function is defined in such a way that the $\theta_1 \ln(x)^Q + \alpha_1$ function operates on the first part of the curve where $c_0 \le x < c_1$. For $c_1 \le x < c_2$ we apply the function $\theta_2 \ln(x)^{Q-1} + \alpha_2$ and continue successively such that the tail of the function is modelled using a pure logarithmic form. It is clear that the class can be extended easily by relaxing the requirement that q are integers. In that case we would simply require that $q_1 > q_2 > \cdots > q_Q$. This would also mean that the tail-function $F_Q(x)$ could be different from the pure logarithmic form. However, in the national model we have applied $q = 1, \dots, 3$ only (e.g., $Q = 3$).

The spline-parameters $\{\theta_q, \alpha_q\}\,\forall q$ are functions of the knot-points and if we consider a log-power spline function of degree Q and normalise $\theta_1 = 1$ and $\alpha_1 = 0$ then unique spline parameters $\theta_2, \dots, \theta_Q$ and $\alpha_2, \dots, \alpha_Q$ exist and can be found from equation (16) and the recursive equation system in (17) below

$$\theta_q = \frac{Q}{Q-q+1} \prod_{r=2}^{q} ln(c_{r-1}), \forall q = 2, \dots, Q \tag{16}$$

$$\alpha_2 = \frac{1}{Q-1} \ln(c_1)^Q \tag{17}$$

$$\alpha_3 = \alpha_2 + \frac{2!}{Q-1}\ln(c_1)\ln(c_2)^{Q-1}$$
$$\alpha_4 = \alpha_3 + \frac{3!}{Q-1}\ln(c_1)\ln(c_2)\ln(c_3)^{Q-2}$$
$$\alpha_5 = \alpha_4 + \frac{4!}{Q-1}\ln(c_1)\ln(c_2)\ln(c_3)\ln(c_4)^{Q-3}$$
$$\ldots$$
$$\alpha_q = \alpha_{q-1} + \frac{(q-1)!}{Q-1}\ln(c_{q-1})^{Q-q+2}\prod_{r=1}^{q-2}\ln(c_r)$$

Potentially other function candidates than the log-power series could have been chosen[7]. It could be argued that this particular function with the assumption of connected log-power terms is generally restrictive. This is to some extent true although it is restrictive for a reason (to enforce a particular curvature for the elasticities). However, as we will see in the empirical section this type of function is supported by the data as it outperforms traditional functions. Another thing to note is that the spline function easily can be "blended" with other functions to accommodate deviation from the spline curvature.

The spline-class has been applied to the primary trips only. For the secondary trips, which are generally much shorter, we have not applied a spline formulation but a standard linear and logarithmic hybrid function as shown in (13). To account for differences between the transport market in and around Copenhagen and the rest of the country we allow for a separate parameterisation $r = 1$ for Copenhagen and the rest of the country represented by $r = 2$. In terms of parametrisation we have generally applied generic parameters for $GTT_m(d)$ terms. Hence, we do not allow mode-specific parameters, although for the business segment there is an exception. In addition, for the spline-function, we apply a single scaling parameter for the spline function and do not consider different scaling parameters for different intervals. The optimal spline-parameters are found by comparing the optimal values of the likelihood function for different sets of knot points. As a result, the standard form of the utility function is given below in (18).

$$\begin{aligned} V(m,d) = ASC_m + \ &\varphi z_{m,d} \\ &+ \sum_{r=1,2}\Bigg[\beta_r^{Li}GTT_m(d) + \beta_r^{Lo}\ln\big(GTT_m(d)\big) \\ &+ \beta_r^{Sp}\sum_{q=1}^{Q}1_q\big(GTT_m(d)\big)\Big[\theta_q ln\big(GTT_m(d)\big)^{Q-q+1} + \alpha_q\Big]\Bigg] \\ &+ ln(S_{d,1} + e^{\gamma_1}S_{d,2} + e^{\gamma_2}S_{d,3} + e^{\gamma_3}S_{d,4}) \end{aligned} \tag{18}$$

The term $\varphi z_{m,d}$ represents other variables including variables related to car ownership, parking, socio-economy (dummies for different groups choosing a certain mode) and regional dummies (by mode). As seen, the parametrisation of the entire spline-function is represented as a generic scaling β_r^{Sp} and depends in addition on the spline-parametrisation $\{\theta_q, \alpha_q\}$. The linear and log parameters β_r^{Li} and β_r^{Lo} represent potential blending parameters and parameters applied in the secondary trip model. The model also includes non-linear size terms $S_{d,1}, \ldots, S_{d,4}$ for the choice of destinations as described in (Daly, 1982). The definition of the size-terms differs with trip purpose. For commute it will be represented by total number of jobs (and in this case only a single size-term is included) whereas for leisure and shopping it may consist of several sector-specific employment variables in combination with the size of the population.

4.3 Nesting structure

---

[7] A power-series is another example as discussed in Rich (2016).

In the estimation of the models we apply a nested logit model. In the mode- and destination model there are two possible nesting structures as presented in (19) and (20) below.

$$P_n(m,d) = P(m|d)P(d) \tag{19}$$

$$P_n(m,d) = P(d|m)P(m) \tag{20}$$

In the first specification in (19) the choice of mode is assumed to be perfect substitutes (substitution a la the multinomial logit model), whereas destination choice is limited. In the second specification it is the opposite way around with limitations in the choice of mode and with destination choices being perfect substitutes. These two specifications have been tested for all segments using a Horowitz-type test (as these models are essentially non-nested) and the resulting specification is shown below in Table 8.

Table 8. Land-use data used in the demand model

Segment	Applied nesting structure	Logsum parameter
Commute	$P(m|d)P(d)$	0.815
Education	$P(d|m)P(m)$	0.517
Escort	$P(m|d)P(d)$	0.596
Shopping	$P(d|m)P(m)$	0.559
Leisure	$P(m|d)P(d)$	0.768
Business	$P(m|d)P(d)$	0.673

4.4 Frequency models

For the frequency models we apply a simple multinomial logit model for the choice of tours on the given day. These models are generally relatively simple and depend on socio-economic characteristics such as the number of children in the household, the age of the respondent, the car ownership status, and accessibility measures represented as the logsum from the mode and destination choice model. We will not consider this model in more detail in the present paper.

5. Results

5.1 General model specification

A number of different specifications have been tested, including various specifications related to the form of the GTT function. Below, we present a comparison of a simple linear specification, a specification with a linear and logarithmic hybrid, and a spline specification as implemented in the present model.

As can be seen, there are considerable performance impacts of choosing the spline-function to other alternative forms. In fact, as the models are having almost the same number of parameters[8], the likelihood increase from a linear/log specification as expressed in (13) to a spline formulation as in (15) more than out-weight the likelihood increase of going from a multinomial logit model to nested logit model specification. Hence, the performance increase is considerable.

The implemented spline-parametrisation is based on a definition of knots points which define the intervals on which the different spline functions operate. These are shown below in (10). As we saw in equation (16)-(17), the spline-parametrisation given by the scaling parameters θ_q and the intercept parameters α_q, is a direct function of the choice of knot-points.

These parameters are not based on model estimation as the underlying model is highly non-linear. Rather these parameters are based on testing and validation of several combinations. The spline function provides a significantly better description of the data compared to more

[8] If we do not count the knot-points as parameters, the linear/log model has one additional parameter compared to the model with the spline function. If they are included the models with the spline function has one parameter in excess. In any case, in this context it does not affect the goodness of fit in any significant way.

traditional functional forms despite the fact that the number of parameters is almost the same. To see this we apply the approach in Ben-Akiva and Swait (1984) based on the non-nested test in Horowitz (1983) in combination with the results in Table 9. The adjusted rho-squared is given by $\bar{\rho}^2 = 1 - \frac{L(\beta)-K}{L(0)}$ and the distribution of the difference between $\bar{\rho}^2$ for different (possibly non-nested models) is given by

$$\Pr(\bar{\rho}_2^2 - \bar{\rho}_1^2 > z) \leq \Phi(-[-2zL(0) + (K_2 - K_1)]^{1/2}) \tag{21}$$

Table 9. Goodness-of-fit primary model. The column to the right (spline) represents the final model

Segment	Nobs	$L(0)$	$L(\beta)$ (Linear)	$L(\beta)$ (Linear/Log hybrid)	$L(\beta)$ (spline)
Commute	22763	-120919	-74108	-74003	-73828
Education	9040	-48003	-23458	-23425	-23431
Escort	6335	-33690	-15339	NA	-14805
Shopping	16034	-84740	-37994	-37578	-37148
Leisure	16034	-89922	-53114	-52424	-51700
Business	2275	-12101	-8443	-8364	-8316

Table 10. Implemented knot-points in the model

Trip purpose	First knot point (KM)	Second knot point (KM)	Q
Commute	150	300	3
Education	100	200	3
Escort	50	150	3
Shop	50	150	3
Leisure	75	150	3
Business	150	300	3

To test the null hypothesis that the linear model is better the spline model and the log/linear model is better than the spline model, we compute the test statistics $z_1 = \bar{\rho}_{spline}^2 - \bar{\rho}_{lin}^2$ (to test if the linear model outperforms the spline model) and $z_2 = \bar{\rho}_{spline}^2 - \bar{\rho}_{loglin}^2$. Similarly, we compute the corresponding $Z_1 = -[-2z_1L(0) + (K_2 - K_1)]^{1/2}$ where $K_2 - K_1$ is the difference in the number of parameters between the spline and linear model. The results for the different models are presented in Table 11 below.

Table 11. Comparison of fit between models based on non-nested Horowitz type tests

Segment	$\bar{\rho}_{lin}^2$	$\bar{\rho}_{loglin}^2$	$\bar{\rho}_{spline}^2$	z_1	z_2	Z_1	Z_2	$Pr(z_1 \leq Z_1)$	$Pr(z_2 \leq Z_2)$
Commute	0.386	0.387	0.388	0.002	0.001	-23.580	-18.708	<0.0001	<0.001
Education	0.508	0.508	0.508	0.001	0.000	-7.071	3.464	<0.001	<0.001
Escort	0.540	NA	0.556	0.016	NA	-32.619	NA	<0.001	NA
Shop	0.550	0.555	0.560	0.010	0.005	-41.085	-29.326	<0.001	<0.001
Leisure	0.407	0.415	0.423	0.016	0.008	-53.141	-38.053	<0.001	<0.001
Business	0.288	0.295	0.299	0.010	0.004	-15.811	-9.798	<0.001	<0.001

As can be seen from Table 11, the linear model is in all cases the poorest representation of the data. Moreover, for all models except the model for education, the spline model provides a significantly better description of the data compared to the log/linear model. For education the log/linear model is actually slightly better and has been applied instead of the spline model.

For each segment there are as many as 170 parameters. Many of these are dummy variables taking the region of destination into account for each mode or the type of urbanisation level for the destination. It is neither possible nor relevant to consider all of these (interested readers might consider the background document), and we present only a selected set of parameter estimates in

Appendix C. However, the central parameters are generally estimated with very low standard error and are well identified. This is also the case for logsum-parameters which represent the different nesting structures of the different segments. For commute, escort, leisure and business the nesting structure is such that destination choice is over mode choice, whereas for education and shopping the opposite nesting structure is used.

5.2 Replication of distance profile

An important model validation test is to assess whether the model replicates the distance distribution for the different modes. We have tested this by looking at the replication on shorter distances but also on longer distances. The latter is relevant for large infrastructure projects where a significant share of trips is long distance. In Figure 3 below we illustrate the modelled and observed distance distribution for commuter trips. All other segments look similar and have not been included.

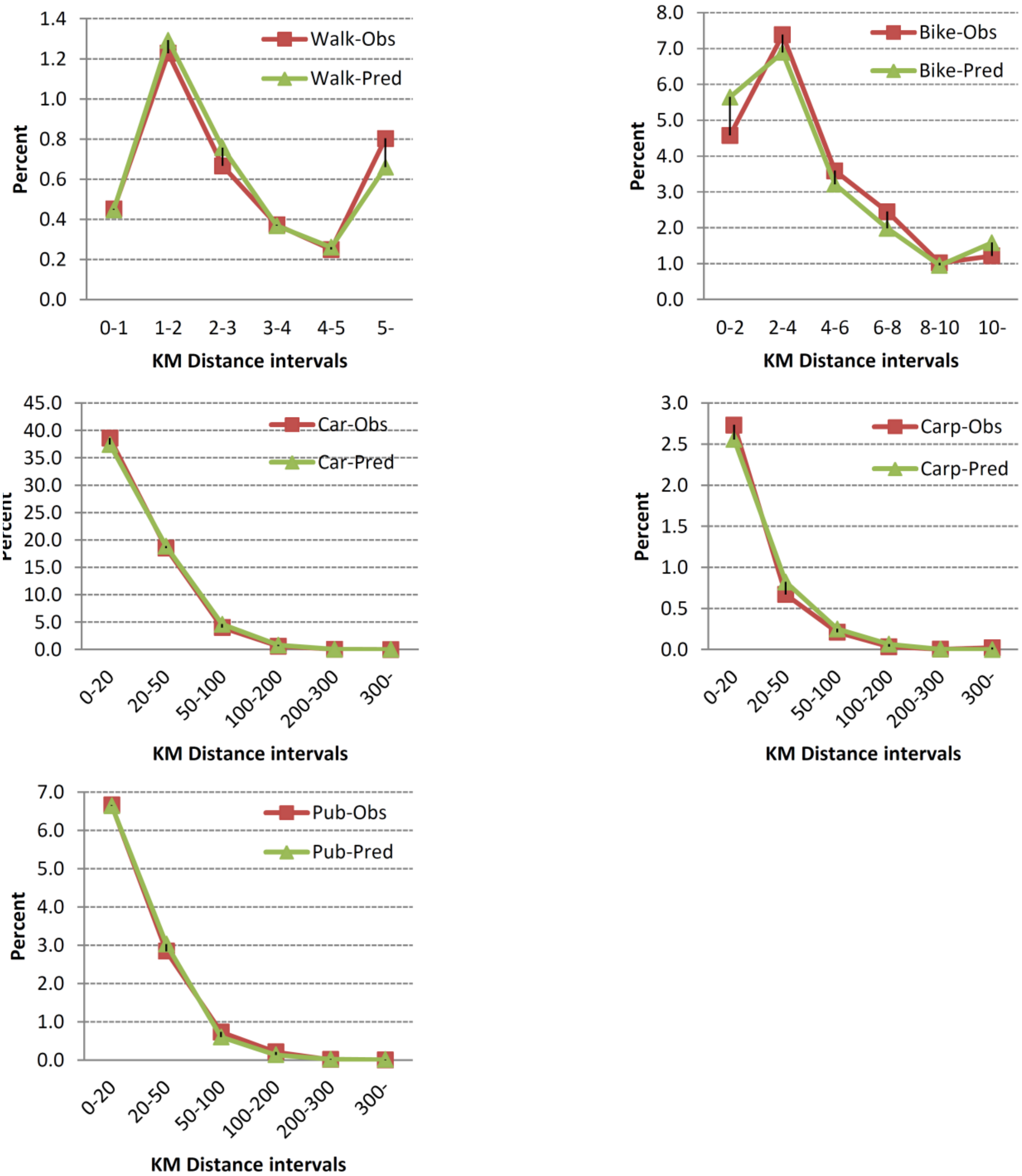

Figure 3. Assessment of the modelled and observed trip length distribution for commuter trips. Profile is for (very) long tails

Overall the distance profile is replicated nicely, even for very long distances. Prior to the introduction of the spline function a linear-log model was applied, however, this approach generally under-predicted the tail probabilities. However, after the introduction of splines we have had consistently well-behaved distance profiles. To illustrate the performance of the spline-function compared to the log-linear hybrid model in terms of replicating the distance profile, we have calculated the norm deviation $\|Dist^{obs} - Dist^{est}\|$ between the observed and modelled trip distribution for each mode. Hence, for each model we calculate

$$\|Dist^{obs} - Dist^{est}\| = \sqrt{\sum_{i=1}^{I} \left(N_i\left[Dist_i^{obs} - Dist_i^{est}\right]\right)^2} \tag{22}$$

N_i represents the observed observations for each of the $i = 1, \dots, I$ distance band. Then to make a simple benchmark we calculate the ratio τ of the norm deviation of the log-linear and the spline model. Hence

$$\tau = \frac{\|Dist^{obs} - Dist_{LL}^{est}\|}{\|Dist^{obs} - Dist_{SP}^{est}\|} - 1 \tag{23}$$

Here $Dist_{LL}^{est}$ is the predicted distance profile for the log-linear model and distance and $Dist_{SP}^{est}$ for the spline model.

Table 12. Distance replication across models and segments measured as in terms of τ

	Walk	Bike	Car driver	Car passenger	Public	Total
Commute	4%	0%	37%	32%	-15%	30%
Education	54%	54%	52%	94%	-18%	14%
Escort	10%	82%	97%	54%	-50%	92%
Shopping	12%	-26%	92%	0%	38%	74%
Leisure	9%	-25%	60%	96%	54%	50%
Business	18%	-23%	50%	-15%	25%	48%
Total	16%	7%	62%	61%	2%	45%

In (12), positive percentages represent cases where the spline function performs better than log-linear model. The general picture is that the spline-model performs much better. Specifically, all of the column and row totals are positive although there are exceptions where the log-linear model performs better. This is not a misspecification but the "cost" of having a generic functional form across modes. After all, the spline model as shown in Figure 3, replicates the distance profile well across all modes. Secondly, observe that the "worst" segments, the escort-public curve, include only very few observations and should not be given much attention. The same is true for the bike-business segment.

5.3 Model elasticities

Elasticities by mode and overall trip purposes are presented in Table 13. These elasticities are based on a model-simulation with a 10% increase in the corresponding variables. Elasticities are before pivoting and without iterations with the assignment model.

For cars the costs include only kilometre based costs, whereas for time, only in-vehicle time is considered (e.g., ferry time is excluded). For public transport, costs only represent ticket costs and time is only in-vehicle time (e.g., possible waiting time and transfer time are included in other variables). The average elasticities have been stable across many different specifications and will not change significantly depending on the form of the GTT model.

Another issue related to model sensitivity is that the sensitivity may be affected by pivoting. In order to make sure that the sensitivity structure of the model after pivoting conforms to the sensitivity before pivoting, it is common to apply a row-normalisation technique (Daly et al., 2005). Still, because there can be substantial differences between the synthetic and observed matrices, there can be differences in the model sensitivity pattern before and after pivoting. However this has been tested and elasticities before and after pivoting have been found to be largely similar.

The difference between the spline-function and the linear/log function is how the elasticity-curve is damped as a function of distance. The difference is illustrated in Figure 4 where we present the elasticity curve for shopping trips.

Table 13. Mode and destination elasticities based on sensitivity tests with respect to: "CC" = car cost, "CT" = car time, "PC" = public cost, "PT" public travel time

Trip purpose	Mode	Measure	CC + 10%	CT + 10%	PC + 10%	PT + 10%
Commute	Walk	Trips	0.250	0.487	0.079	0.080
Commute	Bike	Trips	0.234	0.480	0.105	0.112
Commute	Car driver	Trips	-0.163	-0.253	0.054	0.066
Commute	Car passenger	Trips	0.412	-0.527	0.101	0.111
Commute	Public	Trips	0.273	0.571	-0.585	-0.679
Commute	Walk	Mileage	0.353	0.671	0.080	0.086
Commute	Bike	Mileage	0.315	0.636	0.116	0.135
Commute	Car driver	Mileage	-0.417	-0.519	0.052	0.068
Commute	Car passenger	Mileage	0.590	-1.114	0.114	0.134
Commute	Public	Mileage	0.403	0.785	-0.762	-1.050
Business	Walk	Trips	0.165	0.654	0.025	0.050
Business	Bike	Trips	0.146	0.623	0.042	0.079
Business	Car driver	Trips	-0.058	-0.113	0.015	0.045
Business	Car passenger	Trips	0.321	-0.471	0.049	0.097
Business	Public	Trips	0.206	0.922	-0.298	-0.798
Business	Walk	Mileage	0.217	0.842	0.026	0.053
Business	Bike	Mileage	0.191	0.806	0.047	0.092
Business	Car driver	Mileage	-0.221	-0.457	0.016	0.057
Business	Car passenger	Mileage	0.386	-0.985	0.059	0.131
Business	Public	Mileage	0.291	1.158	-0.389	-1.382
Leisure	Walk	Trips	0.166	0.537	0.078	0.040
Leisure	Bike	Trips	0.163	0.548	0.110	0.057
Leisure	Car driver	Trips	-0.258	-0.241	0.057	0.037
Leisure	Car passenger	Trips	0.255	-0.636	0.086	0.048
Leisure	Public	Trips	0.192	0.640	-1.065	-0.606
Leisure	Walk	Mileage	0.238	0.764	0.077	0.044
Leisure	Bike	Mileage	0.225	0.744	0.108	0.061
Leisure	Car driver	Mileage	-0.487	-0.425	0.049	0.034
Leisure	Car passenger	Mileage	0.316	-1.150	0.073	0.043
Leisure	Public	Mileage	0.262	0.816	-1.111	-0.836

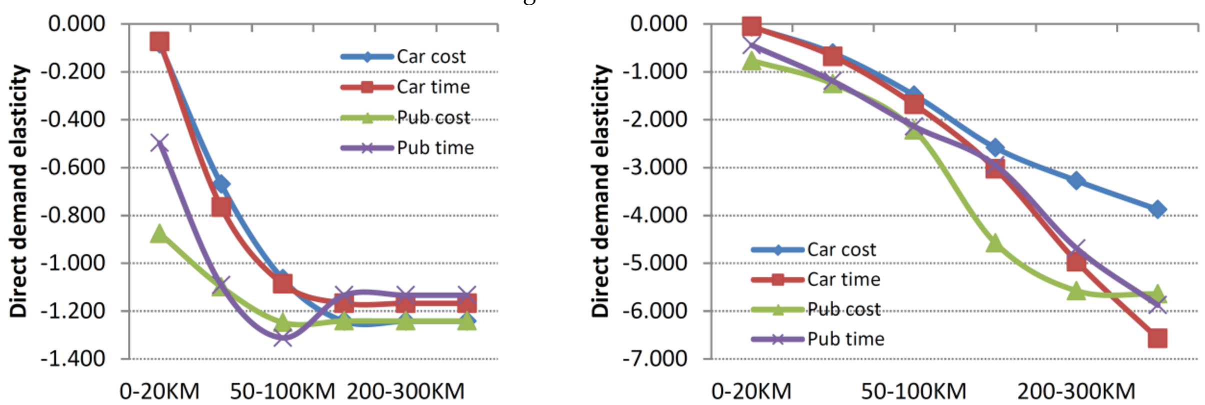

Figure 4. Typical elasticity curve by distance for the spline-function (to the left) compared to a linear/log function (to the right). The example is for shopping trips.

As can be seen from the curve to the left in Figure 4, up to the first knot-point the curve has a relative steep progression. After the second knot-point the curve starts to flatten out and after the third knot point the curve drops to the log-function and the curve tends to be almost horizontal. Other elements related to the choice of mode may affect the demand function and the curves may fluctuate somewhat as seen for the public-time component. However, it should be noted that there are only very few public shopping trip above 100 KM and fluctuations in competing modes with higher market shares may result in substitution effects. The shape of the elasticity curve for the linear/log to the right is steeper as it tends to be dependent on the scaling from the linear function. The curve to the right in Figure 4 illustrates a curve, where we have estimated the reference model with a linear/log function specification.

The above elasticities are point-estimates based on a general change in the corresponding attributes. Although the elasticities provide a basis for understanding how sensitive the model is to various exogenous changes, it does not provide a basis for validating whether the estimated model sensitivity is consistent with reality. To look more into this, a back-casting analysis, which is described below, has been carried out. Another relevant question is how the elasticities compare to the literature. We have looked at several sources including De Jong and Gunn (2001), Wardman (2014) and elasticities from the Danish OTM model as reported in Fox and Sivakumar (2006) and the conclusion is that average elasticities (e.g., the elasticities in Table 13) are consistent with the literature. An observation point however, it the difference in sensitivity between car drivers and car passengers with respect to travel time. This is a consequence of the zero-cost specification for passengers and we are currently working on a cost-sharing version to further investigate this. Whether the cost-damping behaviour in Figure 4 corresponds to finding from other models and the literature in general is not possible to say as such figures are never published. However, according to Daly (2010) and others there is strong empirical evidence for cost-damping behaviour in many models and this suggest that curves should tend to flatten out as distance increases.

5.4 Back-casting validation on specific corridors from 2010 back to 2002

During the validation process of the model, a back-casting exercise was carried out where the model was back-casted to year 2002. This was to test whether the model could replicate the 2002 situation.

It should be stressed that the back-casting exercise was not a full-scale back-casting as this would require an enormous amount of work in terms of identifying network changes, pricing structures (of which many are local), and not least public transport schedules for the entire country for 2002. In other words, the back-casting can be characterised as a scenario for 2002 which is based on major changes only. More specifically, the back-casting involved the flowing elements:

- Fitting of population based on exact “targets” for the population in 2002 from register data.
- Sector employment from Statistics Denmark.
- General adjustment of variable car costs to take fuel efficiency and prices into account.
- Specific adjustment of prices for major bridges and ferries due to changes of specific pricing structures.
- Change of public ticket prices by type of ticket and with particular emphasis on specific corridors, in particular the east-west corridor across the Great Belt (refer to Table 3).
- Major changes of the public rail network (e.g., removal of Copenhagen Metro line and removal of specific S-train lines).
- Ad-hoc adjustment of the level-of-service between east-west Denmark in order to account for changes in the rail schedule.

- Removal of more than 16 major road infrastructure projects including several motorways.
- Adjustment of major ferry routes in terms of sailing time and frequency.

All of these changes were then coded and a new road and public assignment was run as well as a population synthesis for 2002.

The overall development in the number of trips and mileage between 2002 and 2010 model has been based on various traffic counts and statistical transport indices and is shown in Table 14 below.

Table 14. Observed development in trips and kilometres per weekday in Denmark in 2002 and 2010

	2002	2010	% increase	2002	2010	% increase
Mode	Trips			Million KM		
Car	6,109,000	6,550,000	7.2%	94,744	103,889	9.7%
Car passenger	2,975,000	3,003,000	0.9%	50,567	51,157	1.2%
Public	1,569,000	1,547,000	-1.4%	28,926	29,828	3.1%
Air	4,900	5,100	5.6%	1,316	1,386	5.4%
Bike	2,247,000	2,290,000	1.9%	7,274	7,346	1.0%
Walk	2,017,000	2,079,000	3.1%	1,480	1,532	3.5%

In Figure 5 we compare the road mileage to observed mileage. The phrase "observed" could be debated and reality is that the observed mileage from 2002 is partly constructed from a range of data sources including network counts, GPS traces and up-weighted TU data. However, the way these data has been constructed is very much data-driven and entirely independent from the national model and they provide a good reference for validation. The overall growth in traffic is modelled quite satisfactory although there are some discrepancies in the growth on motorways.

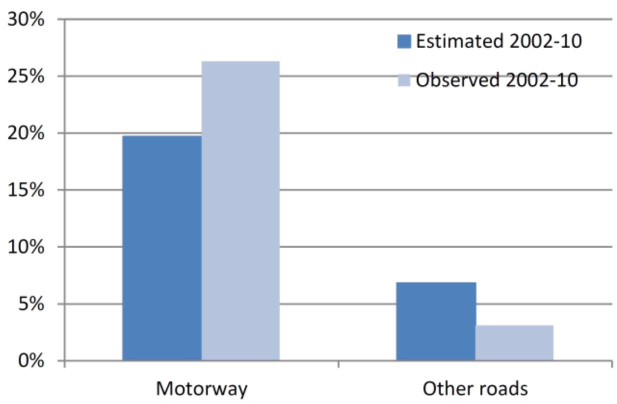

Figure 5. Comparison between modelled and observed mileage on road types for 2002-2010

When looking more into these discrepancies it is not related to the Copenhagen region where the distribution is matched quite well but mainly to Jutland and primarily to the southern part of Denmark close to the border. We believe this related to the fact that the model does not handle the growth in the international traffic which is particular relevant for the motorway system in this region. The overall comparison between observed and modelled rail and bus traffic is presented in Figure 6 below.

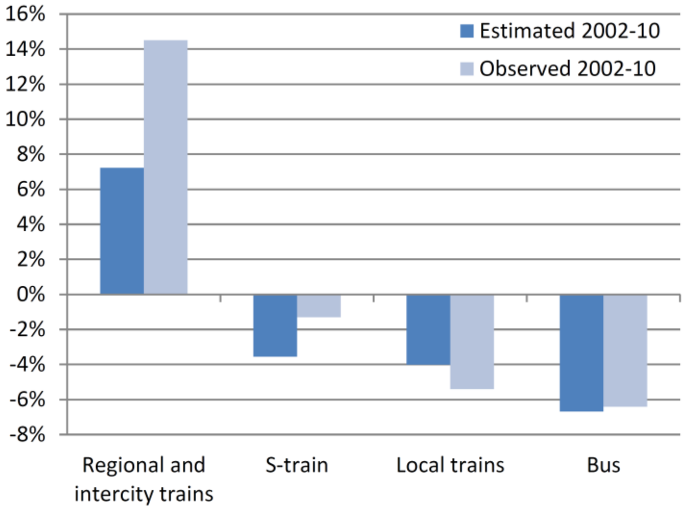

Figure 6. Comparison between modelled and observed mileage for public transport for 2002-2010

In particular for regional and intercity traffic the growth projection is not very good. However, this is a problem that relates particularly to rail operations in Mid- and Western Jutland which has increased quite dramatically (22%) in the period due to a much improved rail supply. This particular change is not included in the definition of the 2002 scenario and explains the difference.

A more interesting comparison is the comparison between observed and modelled traffic across the Great Belt. In this case "observed" is indeed observed as it is based on link-statistics for the bridge. For this particular corridor we have invested more time on getting the assumptions right. The reason for this is that the corridor represents a mixture of short and long trips for which scaling effects of elasticities could be a challenge. A particular useful element is that the corridor experienced a price-chock in 2005 where fares for cars were reduced by 20% overnight. This price variation gives us the opportunity to validate the performance of the spline-function. In Figure 7 below we can see that the match between observed and modelled trips is generally quite good.

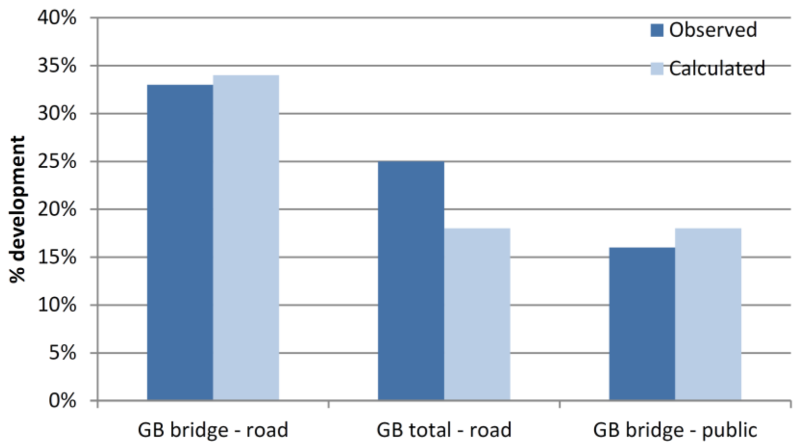

Figure 7. Comparison between modelled and observed trips between East and West Denmark for 2002-2010."GB total" means total crossings across the belt including ferries, whereas "GB bridge" represent only crossings on the bridge

The number of trips on the Great Belt Bridge is very much "spot-on" whereas for cars, there is a tendency that ferries cause some problems. This is not very surprising as it is very difficult to assess the exact pricing structure for these. The pricing structure is influenced by the time of the year and may change due to campaigns for which we do not have data. Secondly, the assignment model tends to be rather sensitive to whatever pricing structure is applied, which presumably is not far from the truth. However, it means that if the pricing assumptions are not exactly correct we will see fluctuations as those seen in Figure 7.

The overall impression from the back-casting is that the model does very much what we expect and replicates the development in 2002 quite well. This despite that the period from 2002-2010 is probably one of the most challenging periods we could have chosen with elements such as a financial crisis, an oil price span of 50-120$ per barrel, significantly reduced car prices for smaller cars in Denmark, the introduction of a green tax reform for cars and rapidly rising house prices which have led to derived impacts on the private economy of Danish households.

6. Summary and conclusion

In the paper we have presented the model structure of the Danish National passenger model and how the demand model, which is the main focus in this paper, links to other sub-models in the model framework. In particular, we described the link between the population synthesis and the demand model and how final OD matrices are calculated from a list of tours from all individuals in Denmark. We also described how individuals are grouped into households, which at a household decision stage decide on household decisions such as car ownership, and at an individual tour decision stage, decide on their individual travel pattern.

Compared to other national models the Danish model differs in that it encompasses both short and long distance trips in its weekday model for Danish transport. In Norway, Sweden and the UK these models have been exogenously segmented into short and long distances. Although this segmentation on the one hand makes the estimation simpler as scaling-effects for longer trips are less prominent, it does give rise to challenges in how to model trips across corridors which represent a mixture of short and long trips. This is particularly true for Denmark where the eastern and western parts of Denmark are separated by the Great Belt Connection. Due to the wide distance domain from very short trips up to trips in excess of 400 kilometres, scaling effects have been a central issue. Due to this, specific attention has been given to the issue of functional form and in particular cost-damping. As it is well acknowledged in the literature, it is generally challenging to cope with cost and time attributes for wide distance domains as scaling effects tends to drive up elasticities in the tail of the distribution. Previous solutions have often been to apply segmentation and more simple functions such as linear or linear/log hybrid functions. However, none of these solutions were found to be acceptable in our case. Due to this, a novel logarithmic spline-function has been developed which is defined as a sequence of logarithmic power functions connected in knot-points. The model makes it possible to have a more flexible damping profile and the empirical results are encouraging. The model gives rise to a significant improvement in goodness-of-fit and the shape of the elasticity curve is fundamentally different from traditional functions, which tend to have proportionality between elasticities and distance.

The ability of the model to project into the future and to support policy analysis has been tested during 2015 in a back-casting exercise from 2010 to 2002. Overall, the model performance is found to be quite promising. For those areas where the model gave rise to discrepancy between the 2002 observed baseline and the projection, it was closely linked with known shortcomings in the underlying level-of-service assumptions. In particular the model did not incorporate a substantial improvement in the rail supply for the Mid-Jutland region and this led (not surprisingly) to a corresponding under-prediction of rail demand for this area. On the other hand, for those areas where more effort was put into the assumptions (in particular the traffic between the eastern and western parts of Denmark) the model was able to project a growth rate quite similar to the observed growth rate. Overall the performance of the back-casting was encouraging and this despite the projection period was particular challenging with structural changes in the economy (e.g., the financial crisis and the oil price increase) and national taxation reforms for cars.

The presented model represents a first version of the Danish model. During the next years it will be further developed and will include additional features. A main upgrade to be commenced during 2016-2017 will upgrade the assignment model to a dynamic path based assignment model. This model will provide a better basis for the modelling of intersections, "spillbacks" and dynamics during the day.

References

Arentze, T.A, Timmermans, H.J.P. (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B* 38:613–633.

Axhausen, K., Gärling, T. (1992). Activity-based approaches to travel analysis: conceptual framework, models and research problems. *Transport Reviews* 12(4): 323-341.

Ben-Akiva M, Lerman S (1985) Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge.

Beser, M. H., Algers, S. (2002) SAMPERS – The New Swedish National Travel Demand Forecasting Tool. In Lundqvist, L., Mattsson, L.-G., (ed), *National Transport Models – Recent Development and Prospects*, Advances in Spatial Science, Springer-Verlag, Berlin, pp.101-118.

de Bok, M., De Jong, G., Baak, J., Helder, E., Puttemans, C., Verlinden, K., Borremans, D., Grispen, R., Liebens, J. and Van Criekinge, M. (2015). A Population Simulator and Disaggregate Transport Demand Models for Flanders. In Transportation Research Procedia, Current practices in transport: appraisal methods, policies and models – 42nd European Transport Conference Selected Proceedings, Vol 8:168–180.

Bowman J.L., M.E. Ben-Akiva, 2000. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A* 35(1):1–28.

Bradley, M., Bowmanb, J.L., Griesenbeckc, B. (2010). SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution. *Journal of Choice Modelling* 3(1):5-31.

Börjesson, M. (2010) Forecasting demand for high-speed rail. Proceedings of the 38th European Transport Conference, Glasgow, Scotland.

Charypar, D., Nagel. K. (2006). Generating complete all-day activity plans with genetic algorithms. *Transportation* 32(4):369-397.

Daly, A.J. (1982). Estimating Choice Models Containing Attraction Variables, *Transportation Research Part B* 16(1):5-15.

Daly, A. (1998). Prototypical sample enumeration as a basis for forecasting with disaggregate models. *Transport Planning Methods* 1(D):225-236.

Daly, A. (2007). National Traffic Forecasting Models in Europe and Elsewhere. In Hensher, D.A., Button, K.J., (ed.) *Handbook of Transport Modelling*, Oxford, Great Britain, Elsevier, 2007.

Daly, A. (2010). Cost damping in travel demand models. Technical report, Rand Europe. Url: http://www.rand.org/pubs/technical_reports/TR717.html

Daly, A., Fox, J., Tuinenga, J.G. (2005) Pivot Point Procedures in Practical Travel Demand Forecasting. ERSA paper. Url: http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa05/papers/784.pdf

Department for Transport, 2009. National Transport Model High Level Overview. Report for the Transport Ministry. Url:http://webarchive.nationalarchives.gov.uk/20110202223628/http://www.dft.gov.uk/pgr/economics/ntm/highleveloverview.pdf

Fox, J., Daly, A., Gunn, G. (2003). Review of RAND Europe's Transport Demand Model Systems. Document MR-1694-TRL. Url: http://www.rand.org/pubs/monograph_reports/MR1694.html

Fox, J., Daly, A., Gunn, H. (2003). Review of RAND Europe's Transport Demand Model Systems. Rand Report MR-1694-TRL. Url: http://www.rand.org/pubs/monograph_reports/MR1694.html

Fox, J., Patruni, B., Daly, A., Patil, S. (2014). PRISM 2011 Base: Mode-Destination Model Estimation. Rand Report RR-186-MM. Url: www.rand.org/content/dam/rand/pubs/research_reports/RR100/RR186/RAND_RR186.pdf

Fox, J. (2005). OTM 5 Demand Model Estimation. Report Prepared for Danish Transport Research Institute. Report PM-2139-DTF.

Fox, J., Sivakumar. A. (2006). OTM 5 Demand Model Estimation: Mode-Destination Modelling. Report Prepared for Danish Transport Research Institute. Report PM-2140-DTF.

Gerard de Jong and Hugh Gunn (2001). Recent Evidence on Car Cost and Time Elasticities of Travel Demand in Europe. Journal of Transport Economics and Policy, 2001, vol. 35, issue 2, pages 137-160.

Hamre, T.N., Rekdal, J., Larsen, O.I. (2002). Development of the Norwegian National Transport Model in phase 5. Part B: Estimation of models. Oslo, Institute of Transport Economics. TØI report 606/2002. ISBN 82-480-0297-7. (In Norwegian, with English summary.)Url: http://www.ntp.dep.no/Transportanalyser/Transportanalyse+persontransport/_attachment/502748/binary/813641?_ts=1400bcc8700

Hjalmar, C. and Skougaard, B. Z. (2015). Documentation of the Danish National Travel Survey. Report 10, DTU Transport. Url: http://www.modelcenter.transport.www7.sitecore.dtu.dk/english/-/media/Centre/Modelcenter/tu/TU_english%20version%202015.ashx?la=da

Horowitz, J. (1983). Statistical Comparison of Non-Nested Probabilistic Discrete Choice Models. *Transportation Science* 17(3):319-350

Manski, C.F., McFadden, D.L. (1981). Structural Analysis of Discrete Data and Econometric Applications. Cambridge, The MIT Press.

HCG (1990). The Netherlands National Model 1990. Hague Consulting Group, Netherlands.

HCG (1992). The national model system for traffic and transport. Rotterdam: Ministry of Transport and Public works.

Michael Balmer, Kay Axhausen, Kai Nagel (2006). Agent-based demand-modeling framework for large-scale microsimulations. *Transportation Research Record*: *Journal of the Transportation Research Board*, pp.125-134.

Rekdal, J. (2009). Evaluation of the Norwegian long distance transport model (NTM5). Report 0609 for Statens vegvesen. Url:http://www.moreforsk.com/publications/reports/transport-economics/evaluation-of-the-norwegian-long-distance-transport-model-ntm5/1197/375/

Rich, J., Mulalic, I., 2012. Generating synthetic baseline populations from register data. *Transportation Research Part A* 46:467–479.

Rich, J., Nielsen, O. A. (2015). System convergence in transport models: algorithms efficiency and output uncertainty, *European Journal of Transport and Infrastructure Research* 15(3):317-340.

Rich, J., Jensen, T. (2015). The Danish National Transport Model: Model forecast and population synthesis. Report 35243-006.

Rich, J., Mabit, S., 2015, Cost damping and functional form in transport models, *Transportation*, DOI: 10.1007/s11116-015-9628-8.

Rich, J. (2016). Spline function classes suitable for transport demand models Model. In submission. Temporary version published DOI: 10.13140/RG.2.1.5094.6003. Url:www.researchgate.net/publication/282848874_A_spline_function_class_suitable_for_transport_demand_models

Rohr, C., Fox, J., Daly, A., Patruni, B., Patil, S., Tsang, F. (2010) Modelling long-distance travel in the UK. Proceedings of the 38th European Transport Conference, Glasgow, Scotland.

Rohr, C., Fox, J., Daly, A., Patruni, B., Patil, S., Tsang, F. (2010) Modelling long-distance travel in the UK. Proceedings of the 38th European Transport Conference, Glasgow, Scotland.

Swait, J.V., Ben-Akiva, M. (1984). Incorporating Random Constraints in Discrete Choice Models: An Application to Mode Choice in Sao Paulo, Brazil. *Transp. Research Part B* 21(2):103-115.

Tuinenga, J.G., Pieters, M. (2006). ANTONIN: Updating and Comparing a Transport Model for the Paris Region. Proceedings ETC 2006. url:http://abstracts.aetransport.org/paper/download/id/2542

Verlinden, K., Puttemans, C., de Bok, M., de Jong, G., and Helder, E. (2015) "Micro-simulation with discrete choice models: application in Flanders", these ETC proceedings.

Vovsha, P., M. Bradley, and J. Bowman (2005) Activity-Based Travel Forecasting Models in the United States: Progress since 1995 and Prospects for the Future. In "Progress in Activity-Based Analysis", H. Timmermans (ed.), Elsevier Science Ltd, Oxford, UK, pp.389-414.

Vuk, G. Hansen, C.O., Fox, J. (2009). The Copenhagen Traffic Model and its Application in the Metro City Ring Project. *Transport Reviews* 29(2):145-161.

Wardman, Mark (2014). Price Elasticities of Surface Travel Demand A Meta-analysis of UK Evidence. Journal of Transport Economics and Policy (JTEP), Volume 48, Number 3, September 2014, pp. 367-384(18)

Williges, J., Bok, D.M. (2009). Updating And Extending The Disaggregate Choice Models In The Dutch National Model. ETC proceedings 2009. Url:http://www.significance.nl/papers/2009-ETC-Updating&Extending-the-Disaggregate-CM.pdf

Wilson, A.G., Rees, P.H., Leigh, C.M. (1977). Models of cities and regions (Editors), Chichester and New York: John Wiley.

WSP (2011). Höghastighetståg – modellutveckling. Forskningsrapport, WSP Sverige AB. Url: http://fudinfo.trafikverket.se/fudinfoexternwebb/Publikationer/Publikationer_001401_001500/Publikation_001404/HHT%20rapport_110622.pdf

Appendix A

The notation of the most important elements is shown below in Table 15.

Table 15. Notation

Notation	Description
$OD_s^{Final}(i,j,m)$	Final pivoted OD matrix between zone i and j for segment s and by mode m.
$OD_{s,m}^{Model}(i,j,m)$	Final model OD matrix for segment s and by mode m and destination d before pivoting.
$OD_{s,m}^{WD}(i,j,m)$	Internal Danish Week Day trip matrix. Only between Danish zones.
$OD_{s,m}^{ON}(i,j,m)$	Overnight trip matrix, from all zones to all zones (inclusive foreign zones).
$OD_{s,m}^{ID}(i,j,m)$	International day matrix, from Danish zones to foreign zones under 24H duration.
$OD_{s,m}^{TR}(i,j,m)$	International transit matrix, from Foreign zones to foreign zones.
$PTL_{s,n\|h}(m,d\|z_0)$	Primary trip list for segment s and person n in household h. Trips are classified according to mode m and destination d and conditional on origin z_0.
$STL_{s,n\|h}(m,d\|z_0)$	Secondary trip list for segment s and person n in household h. Trips are classified according to mode m and destination d and conditional on origin z_0.
$TP_{n,s}(c=c')$	Number of primary trips for segment s and person n for car status $c=c'$.
$TP_{n,s}(c=c')$	Number of secondary trips for segment s and person n for car status $c=c'$.
$Pr_{n,s}^{(P)}(m,d\|c=c')$	Mode m and destination d probability of person n for segment s for (P) primary trips.
$Pr_{n,s}^{(S)}(m,d\|c=c')$	Mode m and destination d probability of person n for segment s for (S) secondary trips.
$Pr_h(c=c')$	Probability of car status c' for household h.
$GTT_m(d)$	Generalised time for mode m and destination d.
$\mathcal{F}(GTT_m(d))$	Spline-function class for $GTT_m(d)$.
$VoT(g)$	Value-of-time for income group i.
fft_{car}	Free-flow travel time for car.
fwt_{car}	Ferry waiting time for car.

ct_{car}	Congestion travel time for car.
$fst(d)$	Ferry sailing time for car.
$Inv_{pub}(d)$	Onboard time for public transport.
ns_{pub}	Number of transfers in public transport chain.
wt_{pub}	Waiting time for public transport.
$g = 1, \ldots, G$	Income intervals.
$c_1, \ldots, c_Q$	Spline knot points
$\theta_q, \alpha_q \ \forall q$	Spline parameters
$S_1, \ldots, S_4$	Attraction variables

Appendix B

Table 16. Gender classes

Gender	Description
1	Male
2	Female

Table 17. Age classes

Age groups	Description
1	0-7 years
2	8-14 years
3	15-17 years
4	18-24 years
5	25-29 years
6	30-54 years
7	55-64 years
8	65-74 years
9	75-84 years
10	>=85 years

Table 18. Income classes measured in 1000 DKK after tax

Income	Description
1	0-99 DKK
2	100-199 DKK
3	200-299 DKK
4	300-399 DKK
5	400-499 DKK
6	500-599 DKK
7	600-699 DKK
8	700-799 DKK
9	800-999 DKK
10	>=1000 DKK

Table 19. Classes for labour market association

Labour market association	Description
1	Full-time employed
2	Part-time employed (32 hours/week)
3	Students
4	Retired
5	Unemployed
6	Other people out of job

Table 20. Children class

Children	Description
1	0
2	1 or more

Table 21. Single class

Single	Description
1	Single household
2	Two adults

Appendix C

Table 22. Selected parameter estimates for the primary mode and destination model. Numbers in parentheses are standard errors

Parameter	R	Commute	Edu	Escort	Shop	Leisure	Business
GTT	1		-0.02025				
$(m = car, carp, pub)$			*(0.002945)*				
$\ln(GTT)$	1	-0.21217	-1.01859		-0.9217	-1.0716	
$(m = car, carp, pub)$		*(0.073522)*	*(0.106563)*		*(0.103792)*	*(0.086418)*	
$\ln(GTT)$	1						-0.21161
$(m = car)$							*(0.054832)*
$\mathcal{F}(\mathrm{GTT}\|Q = 3)$	1	-0.04495		-0.1522	-0.06688	-0.0446	-0.03771
$(m = car, carp, pub)$		*(0.002336)*		*(0.004797)*	*(0.00405)*	*(0.002907)*	*(0.002489)*
$\mathcal{F}(\mathrm{GTT}\|Q = 3)$	2	-0.06522	-0.05174	-0.12217	-0.08786	-0.0773	-0.05247
$(m = car, carp, pub)$		*(0.000601)*	*(0.00097)*	*(0.003234)*	*(0.000919)*	*(0.001042)*	*(0.00143)*
Walk time	1	-0.18309	-0.29843	-0.26278	-0.3424	-0.2128	-0.10319
$(m = walk)$		*(0.010708)*	*(0.020144)*	*(0.01325)*	*(0.014314)*	*(0.008631)*	*(0.033115)*
Walk time	2	-0.08737	-0.21232	-0.1469	-0.17167	-0.10601	-0.11738
$(m = walk)$		*(0.004858)*	*(0.01118)*	*(0.007176)*	*(0.006654)*	*(0.003661)*	*(0.01902)*
Bike time	1	-0.15768	-0.24936	-0.37141	-0.33584	-0.25893	-0.18703
$(m = bike)$		*(0.003325)*	*(0.008385)*	*(0.018138)*	*(0.012594)*	*(0.007023)*	*(0.018791)*
Bike time	2	-0.14833	-0.21002	-0.23514	-0.20229	-0.16558	-0.15431
$(m = bike)$		*(0.002528*	*(0.005762)*	*(0.010332)*	*(0.006752)*	*(0.00368)*	*(0.014324)*
Station proximity (origin)		-0.15155		-0.46096	-0.37115	-0.33065	-0.13991
$(m = pub)$		*(0.034476)*		*(0.233655)*	*(0.102815)*	*(0.050978)*	*(0.133691)*
Station proximity (dest)		-0.2731		-0.37869	-0.82818	-0.33237	-0.26827
$(m = pub)$		*(0.055578)*		*(0.27245)*	*(0.145407)*	*(0.066216)*	*(0.194581)*
Public first waiting time		-0.02142	-0.01179	0.005835	-0.00744	-0.00517	-0.03237
$(m = pub)$		*(0.004527)*	*(0.002617)*	*(0.005041)*	*(0.005369)*	*(0.003477)*	*(0.016616)*
Public walk time		-0.01884	-0.08015	-0.11875	-0.07047		-0.02043
$(m = pub)$		*(0.003635)*	*(0.00327)*	*(0.021284)*	*(0.006849)*		*(0.012092)*
ASC walk		3.811015	0.784658	7.021756	5.01564	1.483321	1.042202
$(m = walk)$		*(0.389379)*	*(0.750001)*	*(1.974992)*	*(0.896874)*	*(0.503887)*	*(1.220783)*
ASC bike		3.53375	-0.88668	5.436954	1.000297	0.502105	3.629345
$(m = bike)$		*(0.285789)*	*(0.609005)*	*(1.96207)*	*(0.85591)*	*(0.471202)*	*(0.839508)*
ASC car		-2.39394	-14.6563	0.37755	-4.61466	-3.73585	-1.77203
$(m = car)$		*(0.29184)*	*(0.900594)*	*(1.963008)*	*(0.844749)*	*(0.472513)*	*(0.877394)*
ASC car passenger		-1.14518	-6.68922	-0.20545	-4.78016	-2.29759	-1.15854
$(m = carp)$		*(0.29914)*	*(0.6347)*	*(1.982118)*	*(0.844898)*	*(0.466927)*	*(0.822474)*
Parking cost		-0.00361	-0.00379	-0.00529	-0.00701	-0.00602	-0.00706
$(m = car)$		*(0.000559)*	*(0.002418)*	*(0.002466)*	*(0.001379)*	*(0.001613)*	*(0.002373)*

CarOwnership – one car	1.269278	3.897128	1.213589	3.044758	1.60133	0.966457
$(m = car)$	*(0.050938)*	*(0.368298)*	*(0.116306)*	*(0.156269)*	*(0.066157)*	*(0.189696)*
CarOwnership – two cars	2.489803	5.406144	1.520704	3.758883	1.938452	1.420538
$(m = car)$	*(0.064981)*	*(0.429634)*	*(0.135183)*	*(0.199254)*	*(0.077918)*	*(0.240512)*
CarOwnership – one car	0.511541	1.153783	0.432365	1.999521	0.76391	0.1301
$(m = carp)$	*(0.096624)*	*(0.203715)*	*(0.200687)*	*(0.149553)*	*(0.062206)*	*(0.268255)*
CarOwnership – two cars	0.640973	1.665885	0.655218	2.467181	1.173884	0.336137
$(m = carp)$	*(0.131989)*	*(0.220027)*	*(0.232602)*	*(0.200692)*	*(0.074778)*	*(0.337303)*
S_2 (size parameters)		-0.94929		3.740636	2.905068	
		(0.089502)		*(0.079606)*	*(0.06604)*	
S_3		-4.59383			-0.38707	
		(0.482586)			*(0.298515)*	
S_4		-3.25392				
		(0.656554)				
Logsum (nest-parameter)	0.815446	0.517531	0.596962	0.559653	0.768723	0.673558
	(0.007398)	*(0.021865)*	*(0.015341)*	*(0.019778)*	*(0.010743)*	*(0.020587)*