New Zealand Transport Outlook

Household Travel Model SAS Programmes

November 2017

Short name

Household Travel Model SAS Programmes

Purpose of the model

The Household Travel Model SAS programmes are used to supply two inputs to the Postprocessor:

1) Base year (2012/13) estimates of travel (trips, kilometres travelled, and hours of travel) by region by mode; and
2) Projections of New Zealand-wide travel by mode by five-year intervals from 2012/13 to 2042/43.

Software used

SAS

For questions and comments:

transportoutlook@transport.govt.nz

Transport Outlook Household Travel Model SAS Programmes Documentation

1. At a high level, what does this model do?

In the case of the Transport Outlook Household Travel Model SAS Programmes, there is a discrepancy between what the programmes were originally designed to do and what they are actually used for. The programmes were designed to be a complete household travel model, providing projections of household travel by region and mode by five-year intervals from 2012/13 to 2042/43. Projections of three measures of travel are provided: number of trips, number of kilometres travelled, and duration of travel in hours. The model is based heavily on data from the New Zealand Household Travel Survey (see www.transport.govt.nz/research/travelsurvey/).

While the programmes did produce projections that initially looked reasonable, closer inspection revealed some anomalies that could not be easily explained, especially in the smaller regions. Further exploration and testing led us to the conclusion that the model was functioning as designed, and that the anomalies were due to the relatively small sample size of the Household Travel Survey, resulting in random variations in the demographic characteristics of the sampled population. In some regions, it appeared that any true average behavioural differences between regions were being overwhelmed by simple random variation in the sampled population. For this reason, we did not consider the model results to be satisfactory, and moved to a Plan B, which was to develop the separately documented Transport Outlook Household Travel Model Postprocessor.

At this point, the SAS programmes are used only to supply two inputs to the Postprocessor:

1) Base year (2012/13) estimates of travel (trips, kilometres travelled, and hours of travel) by region by mode; and
2) Projections of New Zealand-wide travel by mode by five-year intervals from 2012/13 to 2042/43.

Item (1) is, of course, mostly the same as the published results of the Household Travel Survey (see http://nzdotstat.stats.govt.nz/WBOS/Index.aspx?DataSetCode=TABLECODE7431), so the SAS programmes are not really needed for this role. They do, however, supply a break-out of public transport between train, bus, and ferry, which is not in the published results. Item (2) is, however, still a unique result of the SAS programmes and a necessary Postprocessor input. It is obtained by summing the regional results across all regions.

2. Where do I find the model results?

The most readable place to find the SAS programmes results are in three tabs of the Postprocessor Excel workbooks for each scenario. These tabs are “Total Trip Tables Original” (number of trips), “Total Distance Tables Original” (number of kilometres), and “Total Duration Tables Original” (number of hours). These three tabs contain the complete SAS Programmes output, including the regional breakdowns that are no longer used.

3. What are the inputs to this model and where do they come from?

Appendix A lists the individual SAS Programmes and their inputs. There are two sets of programmes.

The first set works mainly with New Zealand 2013 Census data and Statistics New Zealand population projections to obtain estimates of projected population growth by demographic "cells", which are combinations of demographic characteristics. This set of programmes also projects vehicle ownership, based mainly on the results of a Census question asking the number of vehicles owned by each household.

The second set works mainly with the New Zealand Household Travel Survey data, applying the appropriate demographic growth projections to each individual trip sampled, and then expanding the sample to produce travel projections for future years.

Our original aspiration with the SAS Programmes was to be able to make travel projections at the lowest possible level of geographic detail, which would be by local board in Auckland and by territorial authority (district, city or unitary councils) elsewhere, so as to be able to capture differences in travel behaviour between central cities, suburbs and rural areas in our projections. Unfortunately, the sample size and sample design of the Household Travel Survey proved to be grossly inadequate for the task. As a legacy of these experiments, the SAS Programmes are probably more complicated than they need to be, working with data and projections at the local board or territorial authority level, then aggregating to the regional level, when the same data and projections are available at the regional level.

The individual data files from Stats NZ have been downloaded as Excel files from the Stats NZ website and are available in the \Census Data Files subdirectory for each case. These are then used to generate .CSV files, which can be read by the SAS programmes.

There are three Stats NZ files that are used.

1) *Census 2013 Projected Family and Household Types.* This is from the Stats NZ dataset "Subnational Family and Household Projections: 2013(base)–2038". The SAS Programmes use only Table 4 of this Excel file, showing projections of numbers of families and households by family and household types and by territorial authority and Auckland local board. Additional columns have been added to these tables to convert the data to estimated numbers of households in the five household types used by the SAS Programmes: one-person households, couples without children, single parents with children, all other single-family and multi-family households (mainly two parents with children), and other multi-person households (mainly unrelated flatmates).
2) *National Family and Household Projections.* This is from the Stats NZ dataset "National family and household projections, population by living arrangement type, age, and sex, 2013(base)-2038", showing projected number of people (not households) by living arrangement and age class. The original data downloaded is in the first sheet labelled "NZ.Stat Export". Additional columns and sheets have then been added to convert this data to estimates of number of people by the five household types used by the SAS Programmes and by the smaller number of age classes used by the SAS Programmes. The results are shown in the last sheet "Scaled to Original Pop".
3) *Census 2013 Special Vehicle Ownership Tabulation.* This is from a special tabulation performed for us by Stats NZ, showing the number of households in the 2013 Census owning no vehicle, one vehicle, two vehicles, or three or more vehicles for the five household types used by the SAS Programmes and for the five household income classes used by the SAS

Programmes, for all regions, territorial authorities, and Auckland local boards. The data is used as provided. This data is used not only as a source of vehicle ownership data, but also as a source of base year data on the number of households by household type, household income and region/territorial authority/Auckland local board.

There are three additional data files used with the Census data and Stats NZ data that come from sources other than Stats NZ:

1) *Vehicle Ownership Shift.* This is a file of modeller assumptions specific to each case showing potential future shifts in travel behaviour, which might occur as vehicle-sharing services catch on. The evidence suggests that one of the determinants of how much people travel by vehicle is their vehicle ownership: people who own a vehicle tend to travel more than people who don't. The model is designed to take this into account, with travel by vehicle growing as households increase their vehicle ownership. The model grows (or at least changes) vehicle ownership endogenously as incomes rise and household type compositions shift. However, what happens if vehicle-sharing services catch on? In this event, we consider that the access to a vehicle provided by vehicle-sharing services will induce people who do not own a vehicle to start acting as if they do own a vehicle. The vehicle ownership shifts shown in this file are not designed to be literal increases in vehicle ownership but they represent 'virtual' vehicle ownership as a result of access to vehicle-sharing services.

   The file shows three shifts in behaviour by region, household type, and year: households owning no vehicle who shift to behaving as if they own one vehicle; households owning no vehicle who shift to behaving as if they own two vehicles (typically applicable to household types including a couple), and households owning one vehicle who shift to behaving as if they own two vehicles (again, typically applicable to household types including a couple).

2) *TA to Region Conversion File.* This file assigns territorial authorities (TA) and Auckland local boards to the region(s) in which they lie. If a territorial authority lies in more than one region, then the file gives an estimated fraction of the population in the territorial authority that lies in each region. These estimates are based primarily on data provided by the article on each multi-region territorial authority in Wikipedia. They are approximate. This data file is used in aggregating Auckland local board and territorial authority data into regional data. Our tests showed that the estimated regional populations computed using this file matched the actual regional populations reasonably closely.
3) *Income class change probabilities.* This is a table showing the probabilities that a household in the given income class in the given year will move up to the next higher income class over the following five years. This table is used to model how the distribution of household incomes is likely to change over time. The probabilities in this table were modelled in a separate spreadsheet model "Example Income Distribution Analysis". The latter model relies on two simple assumptions: (1) that the distribution of household incomes is uniform over the range of incomes in the income class, and (2) an increase of x% in real GDP per capita leads to an increase in the real income of every household in the income class equal to x% of the midpoint of the income range for that income class.

Example: in the $30,001 to $50,000 household income class, the midpoint of the class is $40,000 and the width of the class is $50,000-$30,000=$20,000; a 5% increase in GDP is assumed to lead to an increase in income for people in this income class of $40,000 x 5% = $2,000; hence, the probability of someone in this income class moving up to the next income class given a 5% increase in GDP is $2000/$20,000 = 10%.

GDP assumptions are taken from the separately documented Population and GDP Assumptions Model. Empirically, the model in the "Example Income Distribution Analysis" gives an increase in mean income a bit lower than the increase in GDP per person, which is probably reasonable. Also, there is a tendency for households in the lowest income class not to move up as quickly as households in the higher income classes, since a given percentage increase in a low income yields a smaller dollar increase than the same percentage increase in an already high income. This, too, is probably a reasonable assumption.

The Household Travel Survey data comes in three types of files: data on households, data on people, and data on trips. There is one more data file used with the Household Travel Survey data, which is the Stats NZ 2013 Meshblock Correspondence file. This file shows various geographic data for each Census meshblock, including the territorial authority and local board. Since the Household Travel Survey household data shows the meshblock where the household resides, this file can be used to add additional geographical data to the record. However, while this data facilitated earlier experiments, it is not currently used in the model.

4. How does this model derive its results?

As noted above, the model works in two steps. The first set of programmes works mainly with New Zealand 2013 Census data and Stats NZ population projections to obtain estimates of projected population growth factors for each future model year by demographic 'cells'. These cells represent people with a given combination of region or residence, household type, household income class, household vehicle ownership (0, 1, 2, or 3+ vehicles), and age class. An intermediate step in this process is to project the number of households by region, household type, household income, and household vehicle ownership, from which the total number of vehicles in each region can be estimated.

The second set of programmes adds demographic data to each New Zealand Household Travel Survey trip record, including region of residence, household type, household income class, household vehicle ownership, and age class, thus allowing each sampled trip to be assigned to a demographic "cell". This data is added to the original trip information in the record, which includes mode, distance travelled, duration of trip, purpose of trip, time of day, day of week and a weighting factor. The latter is pre-assigned by the New Zealand Household Travel Survey such that, when trips, distance, or duration for each trip record are multiplied by the weighting factor and summed across sample trip records, they yield the estimated total trips, total distance and total duration travelled represented by the sample trip records.

Computing a future travel projection corresponding to a given set of population growth factors, thus, becomes a straightforward process of multiplying the weighting factor for each sample trip

record by the growth factor corresponding to its particular demographic cell. The weighting factors may then be used to estimate total trips, total distance or total duration represented by the sample trip records, as usual. In a similar fashion, assumed changes in trips of a particular type, such as by particular modes or for particular trip purposes, may be represented by growing the weighting factors for sample trip records of that particular type.

Originally, our intention was to model alternative scenarios completely within the SAS Programmes by running a series of 'adjuster' programmes to perform the appropriate modifications to the weighting factors. However, given that we now have the Postprocessor, it is simpler to make most changes in the Postprocessor, and usually necessary to do so in any case, given that any region-specific changes made in SAS Programmes would be averaged across all regions by the Postprocessor. Currently, the only 'adjuster' programme still used in the SAS Programmes is one for adjusting New Zealand-wide travel by trip purpose (work and non-work) to represent the impacts of improved communications technologies. This cannot be done in the Postprocessor, since data by trip purpose is not passed to the Postprocessor.

Appendix A – Household Travel Model SAS Programme Files

Geographical terminology: In most parts of New Zealand, the level of local government below region is a territorial authority, abbreviated as 'TA' here. Many TAs are referred to as 'districts'. Auckland is legally both a region and a single TA. However, it does have a lower level of local government known as 'local boards' or (in Stats NZ terminology) 'community boards'. In the documentation below, and in the SAS code, we refer to the level of geographical division below the region (local boards in Auckland and TAs elsewhere) as 'DLBs', which stands for 'district/local board'. Technically, not every TA outside Auckland is a district—some are cities or unitary councils—but, in our terminology, DLBs are a broad term for all TAs outside Auckland together with all local boards in Auckland. A few TAs are split across two or more regions.

Reading this documentation: There are two sets of SAS files for each case to be run, which are stored in their own subdirectories: the Census Data Files and Trip Data Files. SAS code files in each set are listed below within each set in the order in which they are generally run. Each dataset read or created is listed in non-italicised type and assigned a code in parentheses the first time it is referenced below. Census Data Files have codes ending in 'D', while Trip Data Files have codes ending in 'T'. For brevity, subsequent references to these datasets refer simply to the code. Some of the actual CSV files may additionally contain the name of the case and/or the date the file was created in YYYYMMDD format in their filename, in addition to the file names shown here.

CENSUS DATA FILES – This set of SAS files adds various data to the Census household projections to prepare them for use in making vehicle number and travel projections.

read_census_household_projection

> *creates file (1C*) HHProjection1 *of projected number of households by DLB and household type by year through 2043; as Stats NZ does not provide a projection for 2043, make the 2043 projection equal to the 2038 projection + (2038 projection-2033 projection)*
>
> *from Stats NZ: (2C):* Census 2013 Projected Family and Household Types *(CSV file)*

read_ta_to_region_conversion_file

> *create file (3C)* ta_to_region showing the DLBs in each region and, for DLBs that are split between two or more regions, the approximate percentage of their population in each region
>
> *from a spreadsheet created mostly from Wikipedia data (4C)* TA to Region Conversion File *(CSV file)*

add_region_to_census_household_projection (inputs 1C and 3C)

> *creates file (5C)* HHProjection3 *of projected number of households by portion of a DLB (referred to as 'area' in the comments in the SAS code) in each region by household type and year; prints out a summary by region and year*

read_special_census_tab (input 5C)

creates file (6C) special_tab2 of 2013 *household vehicle ownership data, showing the number of households in each DLB broken out by the number of vehicles owned, household type, and income class*

f*rom (7C)* Census 2013 Special Vehicle Ownership Tabulation (CSV file), *the special census tabulation of household vehicle ownership*

create_income_probabilities_table (input 6C)

creates file (8C) outlook.hhincomeprob4 *of projected income class probabilities by DLB, household type, and year from the special census tabulation data; these income class probabilities may change over time based on the data in (9C)* Income Class Change Probabilities *(CSV file)*

add_income_class_to_census_household_projection (input 5C and 8C)

creates file (10C) hhprojection5 of *projected number of households by portion of a DLB (referred to as 'area' in the code) in each region by household type, income class, and year*

create_ownership_probabilities_table (input 6C)

creates file (11C) hhownerprobs3 *of probabilities of household vehicle ownership (that is, probability of owning 0, 1, 2 or 3+ vehicles) by DLB, household type, income class, and year; contains a now unused feature for shifting these vehicle ownership probabilities over time, which has been replaced by a similar feature in the next programme* add_vehicle_ownership_to_census_household_projection *below*

add_vehicle_ownership_to_census_household_projection (inputs 10C and 11C)

creates file (12C) outlook.hhprojection9 of projected *number of households by region, household type, income class, household vehicle ownership and year; also reads file (13C)* Vehicle Ownership Shift *(CSV file) showing fraction of households owning no vehicle moving to one vehicle, fraction of households owning no vehicle moving to two vehicles and fraction of households owning one vehicle moving to two vehicles by region, household type and year;*

This programme is usually run twice. It is run the first time in the sequence shown here with a null Vehicle Ownership Shift file (taken from the Base Case) as an input to producing projections of actual vehicle ownership in the next two programmes. Then, for scenarios with vehicle-sharing services, it is run once again after summarize_vehicle_ownership CSV with a file of assumed vehicle ownership shifts appropriate to the scenario, so as to model travel behaviour consistent with 'virtual' vehicle ownership induced by vehicle-sharing services.

summarize_vehicle_ownership (inputs 12C)

produces files (14C) vehicle_ownership_breakout, *showing projected number of vehicles by region, household type, income class, household vehicle ownership, and year (15C)* outlook.vehicle_ownership_summary, *showing total number of vehicles by region and year, and*

(16C) outlook.vehicle_ownership_summary_allNZ, *showing projected number of vehicles nationwide by year*

summarize_vehicle_ownership CSV (inputs 14C, 15C and 16C)

produces CSV files vehicle_summary_region_dist *showing the projected number of vehicles by region, household type, income class, household vehicle ownership, and year,* vehicle_summary_region *showing the projected number of vehicles by region and year, and* vehicle_summary *showing the projected number of vehicles nationwide by year, which may be loaded into Excel*

read_projected_age_by_household_type

creates file (17C) PopAgeHHType2 *of national population by age group, household type, and year; since Stats NZ does not produce a projection for 2043, we again make the projection for 2043 equal to 2038 projection + (2038 projection-2033 projection)*

from Stats NZ (18C) National family and household projections *(CSV file); the* original data was named National family and household projections, population by living arrangement type, age, and sex, 2013(base)-2038

create_growth_factors_table (inputs 1C, 13C and 17C)

solves for nationwide people/household by household type, age group, and year; uses this data to create table of projected population by region, household type, income class, vehicle ownership, age, and year; then computes file (19C) outlook.growthfac *of growth factors (future year population/2013 base population) by region, household type, income class, vehicle ownership, age, and year*

TRIP DATA FILES - This series of SAS files adds various data to the Household Travel Survey trip-leg data to make travel projections

create_income_categories

creates file (1T) outlook.personal_income_categories *of income classes and letter codes for them from instream data*

add_income_to_pe (input 1T)

Creates file (3T) outlook.pe10to12_with_income *of people participating in the Household Travel Survey with their approximate dollar personal incomes, sex, and age*

from household travel survey participant data for 2010, 2011 and 2012: (2T) outlook.pe10, outlook.pe11, outlook.pe12

read_census_meshblock_file

Creates file (4T) outlook.meshblock1 *of data on each census meshblock, including the TA or Auckland Local Board (referred to as 'Community Board' or 'CB' by Stats NZ) and Ward*

from (5T) Stats NZ 2013 Meshblock Correspondence File *(CSV), a spreadsheet downloaded from the Stats NZ website*

add_income_to_hh (inputs (3T) and (4T))

Creates file (7T) outlook.hh10to12_with_mesh *of data on households participating in the Household Travel Survey, with their approximate total dollar incomes and meshblock data, including TA or Auckland Local Board (or "CB" in the code) and ward; also totals the number of people in the household; prints out summary of number of records by TA, Auckland Local Board, and Ward*

From household travel survey (6T) household data outlook.hh10, outlook.hh11, outlook.hh12

create_tr10to12

creates file (9T) outlook.tr10to12 *of Household Travel Survey trip-leg data*

from household travel survey trip-leg data (8T) outlook.tr10 outlook.tr11 outlook.tr12

add_hh_data_to_trips (inputs 7T and 9T)

creates file (10T) outlook.tr10to12_with_hh_data *of Household Travel Survey trip-leg data with complete household data; also adds code for household type, household income class, distance class of trip, purpose of trip, mode, and district/local board (DLB)*

add_age_to_trips (input 3T and 10T)

creates file (11T) outlook.tr10t012_with_age_data *of Household Travel Survey trip-leg data with all data in 10T plus the sex and age class of the traveller*

StatsNZSummary (inputs 11T)

produces a summary file of Household Travel Survey trip-leg data that can be compared with the Household Travel Survey results on the Stats NZ website (http://nzdotstat.stats.govt.nz/wbos/Index.aspx?DataSetCode=TABLECODE7431)

grow_trips (inputs 11T and 19C)

duplicates the Household Travel Survey trip-leg data (still at a disaggregated level) for each projected year and multiplies the trips, distance, and duration of the sample record by the weighting factor for the record so as to produce a record showing the modelled number of trips, distances, and durations represented by the sample record; it then multiplies these expanded trips, distances and durations by the appropriate growth factors to produce a file of projected trip-leg data (12T) outlook.trips_by_projected_year; *also duplicates the 4-wheel driver and 4-wheel passenger records to produce records that, with appropriate scaling, can be used as a basis for projecting vehicle share trips*

one of more adjuster files (inputs 12T)

these files create whatever scenario we wish to create; currently, most of the adjustments we make to create scenarios are be done in the post-processor; the only exception are the adjustments by trip purpose (reducing the number of work trips and non-work trips); hence, currently only one set of adjustments is ever done: "adjuster by purpose 4"; this programme was originally designed to handle a variety of potential adjustments, so it is written in a much more general way than it really needs to be for its current purpose

"adjuster by purpose 4" reads a file of multipliers of trips, distances and durations by mode, trip purpose and year (13T) adjustmentsp4 *and multiplies the modelled trips, distances, and durations on each record of 12T by the appropriate multiplier to produce a file (14T)* outlook.trips_by_projected_years.

summarize_trips_low_level (input 14T)

produces file(15T) tripsum3 *of summarised trip data by region, household type, income class, number of vehicles, age class, sex, mode distance class and year*

summarize_trips_by_region (input 15T)

produces several files summarising the projected trip data: tripsum4 *(16T) of projected trip data by region, mode, distance class and year;* tripsum5 *(17T) of projected trip data by region, mode, and year;* tripsum6 *(18T) of nationwide data by mode and year;* tripsum7 *(19T) of projected trip data by region, age class, sex, mode and year;* tripsum8 *(20T) of projected trip data by region, household type, mode, and year*

summarize_trips_by_region_CSV (inputs 16T, 17T, 18T and 19T)

produces the same tables as the previous programme, but in .csv format: trip_summary_region_dist, trip_summary_region, trip_summary, trip_summary_region_ageclass_sex *and* trip_summary_region_hhtype