Regional Land Transport Demand Model – a research note

Technical notes for practitioners

July 2024

Eilya Torshizian, Principal Economics Limited, Auckland
Alina Fehling, Principal Economics Limited, Auckland

NZ Transport Agency Waka Kotahi research note 012
Contracted research organisation – Principal Economics Limited

NZ TRANSPORT AGENCY WAKA KOTAHI
Te Kāwanatanga o Aotearoa
New Zealand Government

ISSN 2703-5654 (electronic)

NZ Transport Agency Waka Kotahi
Private Bag 6995, Wellington 6141, New Zealand
Telephone 64 4 894 5400; facsimile 64 4 894 6100
NZTAresearch@nzta.govt.nz
www.nzta.govt.nz

Torshizian, E., & Fehling, A. (2024). *Regional Land Transport Demand Model: Technical notes for practitioners* (NZ Transport Agency Waka Kotahi research note 012).

Principal Economics Limited was contracted by NZ Transport Agency Waka Kotahi in 2023 to carry out this research.

Keywords: Hybrid transport model; Regional Land Transport Demand Model (RLTDM)

An important note for the reader

NZ Transport Agency Waka Kotahi (NZTA) is a Crown entity established under the Land Transport Management Act 2003. The objective of NZTA is to undertake its functions in a way that contributes to an efficient, effective and safe land transport system in the public interest. Each year, NZTA funds innovative and relevant research that contributes to this objective.

Research notes are the output of research generally undertaken within short timeframes in response to a specific issue or development and the outputs are not independently peer reviewed. The views expressed in research notes are the outcomes of the independent research and should not be regarded as being the opinion or responsibility of NZTA. The material contained in the research notes should not be construed in any way as policy adopted by NZTA or indeed any agency of the New Zealand Government. The research notes may, however, be used by New Zealand Government agencies as a reference in the development of policy.

While research notes are believed to be correct at the time of their preparation[1], NZTA and the agents involved in their preparation and publication do not accept any liability for use of the research. People using the research, whether directly or indirectly, should apply and rely on their own skill and judgement. They should not rely on the contents of the research notes in isolation from other sources of advice and information. If necessary, they should seek appropriate legal or other expert advice.

Please note:

This research was conducted under a previous policy context. For example, the research was developed and/or undertaken under the 2021-24 Government Policy Statement for Land Transport. Consequently, references contained in the report may be to policies, legislation and initiatives that have been concluded and/or repealed. Please consider this in your reading of the report and apply your judgement of the applicability of the findings to the current policy context accordingly.

[1] This research was conducted January 2023-August 2023.

Acknowledgements

We would like to thank Ernest Albuquerque for his comments on this project.

We also acknowledge data and inputs received from Haobo Wan, Dan Jenkins and Kane Swift (Ministry of Transport); and Moses Kakrada, Kieran McMaster and Kerry Greig (NZ Transport Agency Waka Kotahi).

Abbreviations and acronyms

ACC	The Accident Compensation Corporation
CPI	Consumer price index
ETS	Emissions trading scheme
GDE	Gross domestic expenditure
GDP	Gross domestic product
GUI	Graphical user interface
LCV	Light commercial vehicle
LPV	Light passenger vehicle
MC	Motorcycle
NLTDM	National Land Transport Demand Model
PT	Public transport
PV	Private vehicle
RLTDM	Regional Land Transport Demand Model
RUC	Road user changes
VAR	Vector autoregression
VEC	Vector error correction model
VKT	Vehicle kilometre travelled

Contents

Please note:

This research was conducted under a previous policy context. For example, the research was developed and/or undertaken under the 2021-24 Government Policy Statement for Land Transport. Consequently, references contained in the report may be to policies, legislation and initiatives that have been concluded and/or repealed. Please consider this in your reading of the report and apply your judgement of the applicability of the findings to the current policy context accordingly.

Executive summary

With policy discussions focusing more on equity, emissions and inclusion, there is an increasing demand for more flexible policy models and tools. The Regional Land Transport Demand Model (RLTDM) provides a useful framework for current policy discussions. However, it was developed using the MATLAB software package, which is rarely used by practitioners. New Zealand Transport Agency Waka Kotahi (NZTA) commissioned Principal Economics Limited to recode the RLTDM in Stata, using its matrix language (Mata), and provide further notes for practitioners on how to use the model.

Abstract

This report provides a technical description of the Regional Land Transport Demand Model (RLTDM), which is a hybrid approach to forecasting transport demand across New Zealand regions. The model's outputs include deterministic and stochastic forecasts of a wide range of economic and transport series. The model has been re-coded in Stata and Mata, which will make it more accessible to practitioners.

1 Introduction

1.1 The National Land Transport Demand Model

The National Land Transport Demand Model (NLTDM) was developed between 2011 and 2012. It was intended to be a tool for considering how transport demand may evolve, rather than a tool for providing point estimates of demand.

The model takes a hybrid approach to forecasting transport demand. This simplifies the relationship between transport demand and macroeconomic aggregates. However, it combines top-down relationships with additional details of behavioural parameters and often includes reduced (that is, simplified) forms of conventional regional transport models.

The hybrid approach means the NLTDM can be manipulated by people with differing degrees of modelling expertise. This enables researchers and policy advisors to further investigate transport-demand factors, which is useful given the inherent uncertainty with transport modelling. The model evaluates transport-demand scenarios 30 years into the future, taking account of mega-trends in:

- population growth
- spatial demography
- technology
- income and economic growth
- industrial composition
- policy and prices in relevant areas (such as fuel price escalation and volatility, and the environment) (Stephenson & Zheng, 2013).

1.2 The Regional Land Transport Demand Model

In 2016, Stephenson enhanced the NLTDM for the purpose of regional analysis. This model is called the Regional Land Transport Demand Model (RLTDM) (Stephenson, 2016). The RLTDM illustrates the sociodemographic and transport-demand projections in 12 regions. It contains sufficient sociodemographic detail to connect long-term transport demand to primitive drivers of demand, such as a region's population, age groups and industrial composition. The RLTDM can be used to measure the uncertainty associated with future transport demand, using observable historical drivers of economic uncertainty.

In contrast with the NLTDM, the RLTDM covers migration, travel demand, mode choice and freight (Stephenson, 2016). These developments to the RLTDM led to the NLTDM being upgraded to construct regional scenarios. These upgrades include:

- improving how inter-regional migration is modelled, based on origin−destination age-specific migration probabilities
- capturing the effects of intra-regional density, land use and transport demand (congestion), using stylised relationships between population growth and density, and their functional relationships to congestion and travel costs
- introducing mode of regional travel choices, based on regionalised conditional logit models of discrete choice
- introducing inter-regional freight flows, based on origin–destination matrices for freight flows.

The RLTDM includes 12 regions:

1. Northland.

2. Auckland.
3. Waikato.
4. Bay of Plenty.
5. Gisborne-Hawke's Bay.
6. Taranaki.
7. Manawatu-Wanganui.
8. Wellington.
9. Upper South Island.
10. Canterbury.
11. Otago.
12. Southland.

1.2.1 How the Regional Land Transport Demand Model has been used

Stephenson (2016) used the RLTDM to assess regional sensitivity to travel costs, which reflects several factors, including population composition and relative incomes. Retired people who live alone and solo-parent families tend to be more income-constrained than other households, and exhibit greater sensitivity to travel costs. Northland households are the most sensitive to changes in travel costs; Taranaki and Auckland households are the least sensitive.

In densely populated and urban regions, where vehicle speeds are relatively low, increasing the speed has a positive effect on the number of journeys. In rural areas, where vehicle speeds are relatively high, increasing the speed reduces the number of journeys.

Other key findings from Stephenson (2016)'s study include:

- in urban areas with high population density and better access to public transport (PT), such as Auckland and Wellington, people change transport modes more easily
- in regions with lower incomes, people are most sensitive to changes in travel costs (see section 4)
- in Wellington, people's transport-mode choices are remarkably unresponsive to changes in travel costs.

Stephenson (2016)'s analysis shows that population size appears to be the most important driver of freight flows − a 10 percent increase in population is associated with a 16.8 percent increase in freight tonne-kilometres. While a 10 percent increase to population has a lower impact on freight flow than a 10 percent increase to share of services in gross domestic product (GDP), the former is easier to come by than the latter. In fact, there is a negative relationship between GDP at origin and gross freight flows. While this appears counterintuitive, it may reflect that New Zealand industry's growth margin is in services and growing regions are becoming better able to service their own needs.

Stroombergen et al. (2018) use the RLTDM to assess the impact that socioeconomic changes have on New Zealand's land transport system in a range of scenarios. Their analysis suggests that a lower fuel price increases use of PT and private vehicles (PV), because its income effect is greater than its substitution effect.

To analyse the impact that higher population density has on the land transport system, Stroombergen et al. (2018) increased the population density of Auckland and Wellington regions by 10 percent more than would occur endogenously. They held the New Zealand population constant, so that density in other regions decreased. Their results suggest that this change would have these effects:

- Use of PT and active modes of transport would increase in Auckland and Wellington, while vehicle kilometre travelled (VKT) by PV would increase in other regions (other regions experience a corresponding decline in density). Nationally, the demand for PV travel would increase by 3.2 percent, or two billion kilometres, by 2050. The increase in VKT would be dominated by Auckland and Wellington. Stroombergen et al. calculate that, by 2050, VKT would be 18.3 percent above the base case (business as usual). This means that other regions' contribution of VKT to the national total would be 4.1 percent less than the business as usual.
- National demand for PT would increase by 12.4 percent, which is significantly more than the increase in demand for VKT. Demand in Auckland and Wellington would increase by 16.2 percent, which implies that demand in other regions would reduce by 4.8 percent.

1.3 About this report

The RLTDM was developed using the MATLAB software package. As practitioners rarely use this software, the model has not been well used. NZTA commissioned Principal Economics Limited to recode the model with a more useful programming language. The model can be found here: www.nzta.govt.nz/resources/research/notes/012/.

This report does not repeat technical documentation on the RLTDM in NZTA research report 584 (Stephenson, 2016) or other information on the model's structure in NZTA research report 520 (Stephenson & Zheng, 2013). Instead, we aim to provide practitioners with a useful explanation of the model, and tools and guidelines to help them apply it.

2 Technical details of the model

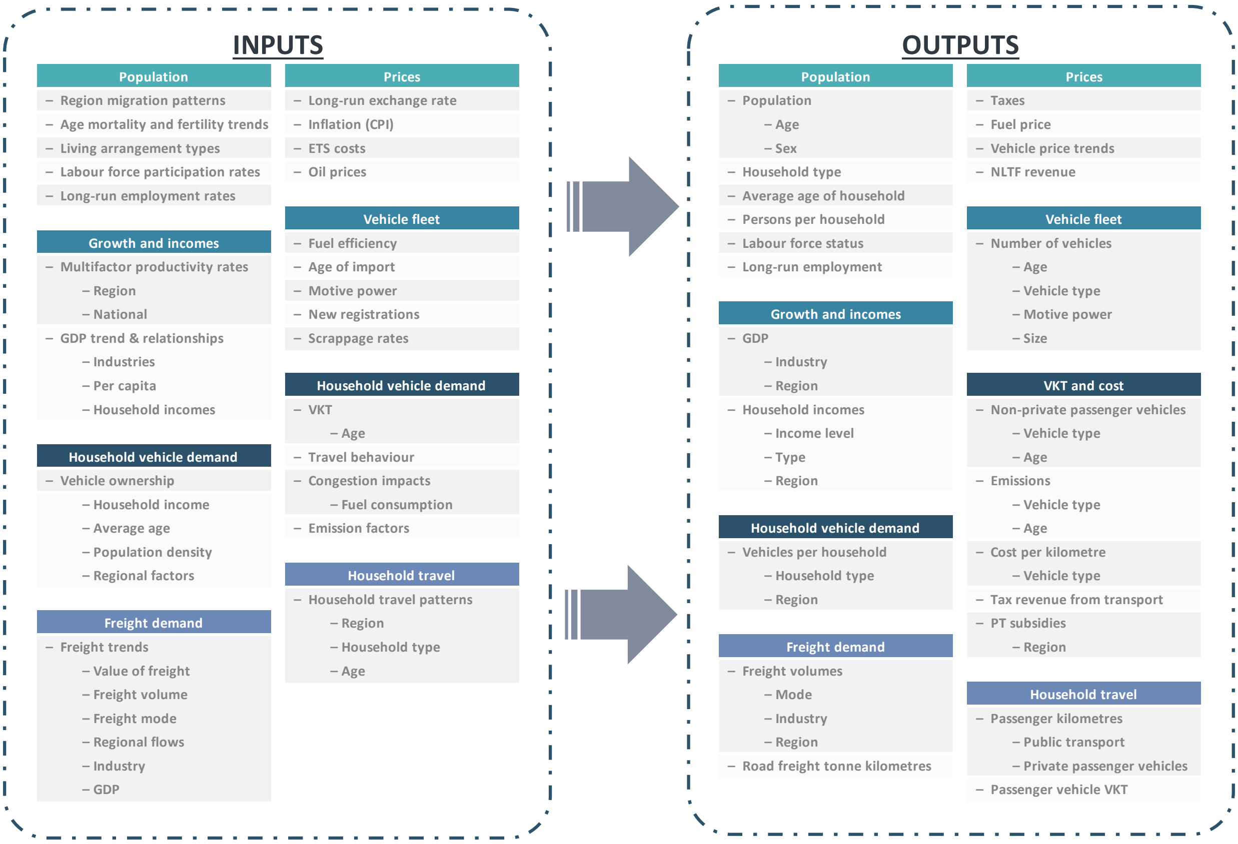
The RLTDM consists of eight sub-models:

1. Population.
2. Economic growth and income.
3. Household vehicle demand.
4. Freight demand.
5. Prices.
6. Vehicle fleet.
7. VKT and cost
    a. mode choice.
8. Household travel.

The outputs of each sub-model inform the mode choice sub-model. Therefore, we start this section by explaining the mode-choice sub-model. Next, we look at the population sub-model, which also defines the RLTDM's stochastic feature. Then we present the economic growth and income sub-model, VKT and cost sub-model and household travel sub-model. We briefly describe the freight demand, prices and vehicle fleet sub-models in Section 3.

Transport models share a common logical flow (see Figure 2.1) between inputs and outputs. In this section, we explain the inputs and outputs for each sub-model. For example, inputs into the population sub-model are regional migration patterns, age mortality and fertility trends, living arrangement types, labour force participation rates and long-run employment rates. These inputs, combined with the economic and demographic modelling framework (see Section 3.2.3) produce the population sub-model's outputs. The outputs include projections of population by age and sex and household types across regions.

Figure 2.1 Flow from inputs to outputs in the RLTDM (inspired by Stephenson, 2016, p. 14)



Note: The RLTDM is based on various econometrics analyses, which are beyond the scope of this report. However, since the estimated coefficients from the econometric analyses are used to inform the RLTDM in Stata, we provide details of the equations in this section, to explain what drives the estimated effects. These details draw on NZTA research report 584 (Stephenson, 2016).

In this report, we provide the most important information for practitioners. While mode choice is not one of the RLTDM's eight sub-models, it has a separate Do-file[2] and sums up most information collected from other sub-models in its logit function. Therefore, we recommend considering how information is used in the mode choice sub-model.

2.1 Mode choice sub-model

The are seven modes: cycle, drive, passenger, walk, bus, taxi (Auckland and Wellington only) and train (Auckland and Wellington only).

A major part of the mode choice sub-model is the extrapolation of choice probabilities, which are estimated using Equation 2-1.

$$P^r_{n,i,t} = \int L^r_{n,i,t}(\beta^r_n) f(\beta). d(\beta) \qquad \text{(Equation 2-1)}$$

Where

$P^r_{n,i,t}$	=	the probability that a person or household (*n*) in region (*r*) will choose a mode of transport, given a particular choice situation (*i*) and period (*t*)
L	=	the logit probabilities, which are based on a mixed, random distribution.

[2] A Do-file lists and executes Stata commands.

For more details and results of the econometric analysis of mode choice, see Tables 4.1, 4.2 and 4.3 in Stephenson (2016), pp. 33–34.

2.2 Population sub-model

The RLTDM uses a modified cohort component method, which disaggregates population by size and age, and evolves based on a transition matrix (*T*)[3] plus net migration (*i*–*m*) (see Equation 2.2).

$$pop_t = T.pop_{t-1} - i_t - m_t$$

$$T = \begin{bmatrix} f_0 & f_1 & f_2 & \dots & f_{\omega-2} & f_{\omega-1} \\ s_0 & 0 & 0 & \dots & 0 & 0 \\ 0 & s_1 & 0 & \dots & 0 & 0 \\ 0 & 0 & s_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & s_{\omega-2} & 0 \end{bmatrix} \qquad \text{(Equation 2-2)}$$

Where

pop_t	=	population at time *t*
i_t	=	outward migration at time *t*
m_t	=	inward migration at time *t*
fw	=	the age-specific fertility rate (*w* represents the age group)
S_w	=	the probability of a person shifting from one age group at time *t* + 1
T	=	changes as age-specific fertility and mortality rate change.

Net migration is projected using an ARIMA(1,0,1) process (see Equation 2.3).

$$N_t = \rho.N_{t-1} + \theta.\varepsilon.N_{t-1} + \varepsilon.N_t + \mu.N_t \qquad \text{(Equation 2-3)}$$

Where

Nt	=	net migration at time *t.*

As we explain in Section 2.3, for stochastic projections the error (*ε*) is sampled randomly from a normal distribution, with standard errors from historic estimates.

For more details of the population sub-model, see Section 3.1 in Stephenson and Zheng (2013).

2.2.1 Stochastic feature

For the stochastic (random) feature, the model uses a vector error correction (VEC) model. This captures the endogeneity of economic activity (gross domestic expenditure (GDE) per capita) and net migration (outward migration minus inward migration) (see Equation 2-4).

[3] *T* evolves over time with changes to age-specific fertility rates, based on autoregressive time-series forecasts; and changes to age-specific mortality rates, based on Standards New Zealand's 2009 base-year medium scenarios.

$$\Delta y_t = A.y_{t-1} + \sum_{i=1}^{3} B.\Delta y_{t-i} + \quad _t \qquad \text{(Equation 2-4)}$$

Where

y_t	=	a vector that observes three variables: GDE per capita, immigrant arrivals and immigrant departures
A	=	a matrix that contains the parameters that reflect long-term equilibrium (cointegrating) relationships between variables
B	=	a matrix that reflects parameters that govern the system's adjustment back towards equilibrium
ε_t	=	a random error term.

The stochastic feature is captured in the RLTDM Do-file's loop for the number of runs (the default number is 20 but can be changed). The impact is captured through the population Do-file's VEC model of arrivals and departures, which is defined based on normal (Gaussian) random variates. The flow-on effect of variations is captured through the other sub-models' responses to population projections.

2.3 Economic growth and income sub-model

To project growth trends, the RLTDM uses a growth-accounting model based on a conventional (log-transformed) Cobb-Douglas production function. Long-run growth potential (Y) is a function of growth in multifactor productivity (A), capital stock (K) and employment (L).

The model assumes that capital stock evolves at a constant rate, equivalent to growth in employment. Making assumptions about labour-force rates, capital-stock growth and the approach to modelling the unemployment rate avoids explicitly modelling business cycles and limits uncertainty in economic growth to one factor – multifactor productivity. This makes it easier to evaluate uncertainty and allows users to easily control growth when they run scenarios. It also implies that the growth-forecasting equation (this is a simple Cobb-Douglas production function of labour, capital and technology) was reduced to a function of population growth, employment and multifactor productivity. This is consistent with the Treasury's assumption for its long-term fiscal model.

Uncertainty in long-term growth is modelled as shocks (ε) around trends in multifactor productivity growth (μ) (see Equation 2.5).

$$A_t = \mu A_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(o, \sigma) \qquad \text{(Equation 2-5)}$$

Regional GDP depends on historic trends in industry-specific GDP forecasts (see Equation 2.6).

$$Y_{r,t} = \sum_i [({}^{E_{i,r}}/_{E_i}) \times GDP_{i,t}]\, \mu A_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(o, \sigma) \qquad \text{(Equation 2-6)}$$

Where

$Y_{r,t}$	=	GDP of region r at time t
E	=	employment counts
i	=	industry

r	=	region
t	=	time (year)
$GDP_{i,t}$	=	GDP of industry *i* at time *t*
Σ	=	the sum of all industries
μ	=	multifactor productivity growth
ε_t	=	a random error term.

The industries in the model are based on an aggregation of these industry groups published by Standards New Zealand:

- Agriculture and food manufacturing (AG).
- Forestry, logging and wood-processing (FOREST).
- Mining, petroleum and chemicals industry (MINING).
- Manufacturing (excluding wood and food manufacturing) (OTHM).
- Construction and utilities (water, gas and electricity) industry (CONS).
- Wholesale, retail, food and beverages, and accommodation services (TRADE).
- Communications, finance, real estate and professional services (OTHS).
- Central and local government administration (PUB).

Household income (*INC*) is calculated using Equation 2.7.

$$INC_{i,j,t} = INC_{i,j,t-1} \times (1 + \left(\alpha_i \Delta GDP_{j,t}\right) + \beta \Delta AGE_{i,j,t}) + \varepsilon_t, \qquad \varepsilon_t \sim N(o, \sigma) \qquad \text{(Equation 2-7)}$$

Where

$INC_{i,j,t}$	=	household income for household type *i*, in region *j* at time *t*
i	=	household type
j	=	region
t	=	time (year)
$GDP_{j,t}$	=	GDP of region *j* at time *t*
$AGE_{j,t}$	=	average age of household in region *j* at time *t*
α, β	=	estimated relationship between change in GDP, and change in age and household income
ε_t	=	a random error term.

Household incomes follow GDP growth per working-age person (ΔGDP). This relationship considers differences in the pace of income growth due to lifecycle effects. The average age of households (ΔAGE) is also a predictor of income growth; it captures the variation in income due to factors other than household living arrangements.

2.4 VKT and cost sub-model

The VKT and cost sub-model connects fleet information and travel behaviour. It projects:

- cost of travel per kilometre, by vehicle type and age
- VKT, by vehicle type and age
- pollutant and greenhouse gas emissions, by vehicle type and age
- travel-related tax revenue from the transport sector.

This sub-model brings together information from some of the model's other sub-models. Cost of travel per kilometre is an important factor for mode choice. It is calculated using Equation 2.8.

$$COST_{i,j,k,y,t}/KM_{i,j,k,y,t} = EFF_{i,j,k,y,t} \times PUMP_{j,t} + RUC_{i,j,k,t} \times GST \qquad \text{(Equation 2-8)}$$

Where

$COST_{i,j,k,y,t}/KM_{i,j,k,y,t}$	=	cost ($) per kilometre of travel
i	=	vehicle class
j	=	fuel type
k	=	vehicle size
y	=	vehicle age
t	=	time
EFF	=	fuel efficiency estimates, taken from Table 9.1 of Stephenson and Zheng (2013), p. 51
PUMP	=	domestic fuel price (consumer costs at the pump) which are modelled as a function of the exchange rate adjusted for the international price of oil, a constant rate for the importer's margin (MARG), fuel tax rates and a constant rate of GST
RUC	=	the weighted average road user charge (RUC) paid per kilometre by vehicle weight, calculated from historic RUC-kilometres purchased, and type and weight of vehicle
GST	=	the goods and services tax (additional to RUC).

Projections of VKT are calculated using Equation 2.9. They are based on historical average VKT per vehicle, by vehicle type and age.

$$\Delta\left(VKT_{i,j}/VEH_{i,j}\right) = \alpha_i \Delta COST_{i,j,t} + \beta_i \Delta INC_{i,j,t} \qquad \text{(Equation 2-9)}$$

Where

$\Delta\left(VKT_{i,j}/VEH_{i,j}\right)$	=	the change (Δ) in VKT per vehicle (VEH)
i	=	vehicle class

j	=	the technology and age combination within each vehicle class
β_i	=	estimated impact of change in income on changes in VKT per vehicle
$\Delta COST$	=	the change in cost per kilometre for each vehicle
ΔINC	=	the change in average household income (light passenger vehicles and motorcycles) or percentage change in GDP per capita (other vehicle classes).

Table 2-1 shows the RLTDM's default price elasticities of demand for VKT for different vehicle classes. The default price and income elasticities for bus and freight vehicles are zero. This is because VKT per vehicle are assumed to rely solely on demand for bus and freight services, which are calculated in the freight demand and household travel sub-models.

Table 2-1 Default price elasticities of demand for VKT for different vehicle classes (reprinted from Stephenson & Zheng, 2013, p. 53)

Vehicle class	Price elasticity	Income elasticity
Light passenger vehicle (LPV)	−0.08	0.01
Light commercial vehicle (LCV)	−0.04	0.01
Motorcycle (MC)	−0.01	0.01

To calculate tax rates per kilometre travelled, the RLTDM uses fuel taxes from the prices sub-model, fuel use per kilometre per vehicle, and RUC rates per kilometre (for vehicles that are subject to RUC). These tax rates exclude GST and emissions trading scheme (ETS) costs, which are calculated separately. Tax rates per litre of fuel use include tax attributable to the National Land Transport Fund and Accident Compensation Corporation (ACC), and fuel monitoring and local authority levies. These per-kilometre tax rates, by vehicle, are multiplied by VKT to project revenue growth.

For ETS costs, fuel consumption (that is, VKT multiplied by fuel use per kilometre) is multiplied by the per-litre ETS costs calculated in the prices sub-model. Emissions are calculated by multiplying fuel use by emissions factors (see Table 2.2).

Table 2-2 Emission factors (reprinted from Stephenson & Zheng, 2013, p. 54, Table 9.4)

Emission factor	Petrol (g/l)	Diesel (g/l)
CO_2	2,311.70	2,650.75
CH_4	0.65	0.15
CO	161.00	11.65
NMVOC	31.03	3.97
N_2O	0.05	0.14
NO_x	7.40	24.73
SO_2	0.07	4.01

2.5 Household travel sub-model

The outputs of the other sub-models inform the household travel sub-model, to account for region-specific travel-mode choices and vehicle-travel characteristics. The key outputs are:

- PT boardings and passenger-kilometres travelled
- PV travel and passenger-kilometres travelled.

2.5.1 Public transport boardings

In the regions, PT boarding is assumed to grow according to regional population growth (Δ*POP*), changes in average vehicle travel costs (Δ*COST*) and growth in average household incomes (Δ*INC*). Differences in travel mode are then adjusted, based on changes in the age composition of regional populations. The outputs are converted to per-capita passenger-kilometres travelled using Equation 2.10.

$$PASSKM_{j,t} / POP_{j,t} = BOARDINGS_{j,t} \times \alpha_j \left(\left(\frac{PASSKM_{j,t-1}}{BOARDINGS_{j,t-1}} \right) / POP_{j,t} \right) \quad \text{(Equation 2-10)}$$

Note: Equation 2.10 includes a coefficient for growth in passenger-kilometres travelled. However, by default, we have assumed that kilometres per boarding does not increase. We assigned per-capita passenger-kilometres travelled to age groups, based on the observed national shares.

Where

$PASSKM_{j,t} / POP_{j,t}$	=	per-capita passenger-kilometres travelled in region *j* at time *t*
$BOARDINGS_{j,t}$	=	PT boardings in region *j* at time *t*
α_j	=	the relationship with the last year's per-capita passenger boarding (kilometres per boarding are assumed unchanged)
j	=	region
t	=	time (year).

2.5.2 Private passenger-vehicle travel

To capture different effects of population growth on travel demands in different regions, the RLTDM uses constant parameters estimated for differences in VKT per vehicle (*VKT/VEH*) and vehicle occupancy, by region, relative to the national average (see Equation 2.11).

$$VKT_{j,t} = \alpha_j \left(\frac{VKT_j}{VEH_j} \Big/ \frac{VKT_{NZ}}{VEH_{NZ}} \right) \times \left(VKT_{NZ,t} / VEH_{NZ,t} \right) \times VEH_{j,t} \quad \text{(Equation 2-11)}$$

Where

$VKT_{j,t}$	=	VKT by region *j* at time *t*
α_j	=	projected changes in relative region VKT and national VKT by vehicle
$VEH_{j,t}$	=	the number of vehicles in region *j* at time *t*

VKT per capita, by region, is calculated. Age distributions (*DIST*) of the national average driver VKT, by age group, are then applied, to obtain age-adjusted region-specific estimates of VKT (Stephenson & Zheng, 2013, p. 56).

3 Technical details of programming

We have converted the RLTDM MATLAB codes to Mata, in Stata, because few researchers use MATLAB. We considered the pros and cons of various statistical packages and concluded that Mata in Stata provides the most appropriate framework for the RLTDM.

In this section, we describe the model's codes, and the dimensions of the matrices that practitioners need to consider when they use the model.

To execute the model, practitioners need to follow the steps described in section 4.1. This section also includes a description of Do-files, for users who are interested in these technical details.

3.1 Choice of software

We selected Stata from three alternatives to the current MATLAB language (see Table 3.1). In addition to the factors shown in Table 3.1, a range of other factors are important for developing codes (such as, the ease of matrix operations).

We focused on features of the software packages that would be most useful to the four main groups of RLTDM users:

1. Economists, who are highly likely to use the model.
2. Data and spatial analysts, who are highly likely to use the model.
3. Transport modellers, who are moderately to highly likely to use the model.
4. Data scientists, who are less likely to use the model.

To ensure that users can continue using the RLTDM in future, the programming language needs to have a fairly stable coding environment that does not significantly change as the software package is updated. This is important because the same codes should apply over time.

The most useful alternative programming languages to MATLAB are Stata and R. Stata provides a more stable coding environment than R. However, R provides a wider range of packages for further analyses, such as econometric or descriptive (illustrative) analyses. The project team and steering group decided to use Stata, because of its stability.

Table 3-1 Features of alternative programming languages to MATLAB

Language	User group	Stability	Ease of further analysis
Python	Data scientists, transport modellers	Stable	High
R	Transport modellers, data and spatial analysts	Unstable	High
Stata	Economists, data and spatial analysts	Stable	Medium

3.2 The model's Stata Do-files

The RLTDM codes are available in 12 Do-files. There is one Do-file for each of the RLTDM's eight sub-models; one Do-file for mode choice; one Do-file for the RLTDM functions; one Do-file for loading Excel data to Mata, and one Do-file for the main model, called 'RLTDM.do'. All the Do-files are available in the main project folder.

The scenario-modelling tools are all available in 'RLTDM.do'. Therefore, modellers do not need to use other Do-files unless they identify issues with the outputs or need to change the assumptions.

3.2.1 'RLTDM.do'

Programming starts from 'RLTDM.do', which defines all the file paths.

Excel files are loaded using 'Loading data from excel.do'. The Excel files consist of various workbooks contained in 13 spreadsheets:

1. Population data.
2. Economic growth data.
3. Labour force data used in the economic growth and income sub-model.
4. Fleet (vehicle ownership) data used in the household vehicle demand sub-model.
5. Travel data.
6. VKT data.
7. Price data used in the prices sub-model.
8. Freight-volume data used in the freight-demand sub-model.
9. Origin−destination data used in the freight-demand sub-model.
10. Freight betas used in the freight-demand sub-model.
11. Travel-panel data used in the VKT and cost sub-model.
12. Travel data used in the household travel sub-model.
13. Mode data used in the mode-choice sub-model.

The Excel files in the data folder, and earlier documents about the RLTDM, explain the various sources of data used by the model. The current version of the model was updated in March 2020, using the latest data available at that time. The first spreadsheet in each workbook displays the date of the most recent update. Some spreadsheets include information from econometric modelling that uses other primary sources of data. The econometric models are beyond the scope of this report, but are described by Stephenson and Zheng (2013) and Stephenson (2016).

'Loading data from excel.do' has a clear structure and provides a straightforward framework for future updates. However, updating the RLTDM remains a major task, given the model uses various sources of information.

'RLTDM.do' defines the model's dimensions and the matrices the model needs to collect outputs from its sub-models. To choose between deterministic and stochastic scenarios, 'RLTDM.do' has a mode scalar that needs to equal 0 (deterministic) or 1 (stochastic). For stochastic scenarios, the number of runs can be chosen (the default number of runs is 20). However, with the current version of the RLTDM, we focus on deterministic outputs, as the stochastic outputs need to be checked for robustness and the code may need to be revised.

3.2.1.1 Scenario analysis

To run different scenarios, the most relevant part of the RLTDM for practitioners is 'A. ASSUMPTIONS to be varied by scenario or stochastic simulation'. To shock the model, users can select from these different instruments:

- Population and labour market (A.1)[4]
  - stochastic parameters

[4] The references in brackets have been used in the Stata programming codes.

- probability of migration cycles.
- National and regional employment (A.1.3 and A.1.4).
- Domestic migration (A.1.5).
- Industry and income outlook (A.2)
  - annual growth rate and cycles of multifactor productivity assumptions
  - industry GDP shares (A.2.2)
  - industry MFP growth rate (A.2.3).
- Freight demand (A.3)
  - freight intensity growth (A.3.1)[5] for road, rail and sea freight.
- Prices (A.5)
  - CPI growth
  - exchange rate (long-run rate and smoothing factor)
  - fuel cost
    - oil prices
    - electric-vehicle efficiency and cost
    - emissions prices and taxes[6]
    - local taxes
  - road user charges (RUC)
    - light vehicles' RUC growth rate
    - heavy vehicles' RUC growth rate.
- Vehicle technology assumptions and age assumptions (A.6)
  - light passenger vehicles
    - new vehicle shares
    - growth in used-vehicle import ages
    - alternative-vehicles share and penetration[7]
  - light commercial vehicles
    - new vehicle purchases
    - growth in used-vehicle import ages
    - long-run shares of alternative vehicles in registrations
  - trends in overall fleet fuel-efficiency improvement.
- Travel demand (A.7)
  - regional variations in vehicle and private passenger kilometres travelled
  - regional (underlying) growth in public transport demand
  - VKT elasticities
    - light passenger vehicles

[5] Freight intensity growth is growth in freight per unit of volume GDP.

[6] Emissions prices and taxes affect the prices sub-model.

[7] Alternative vehicles include hybrids, electric vehicles and plug-in hybrids. For more details, see Equation 8.4 in Stephenson and Zheng (2013).

    - light commercial vehicles
    - motorcycles
    - heavy commercial vehicles
    - bus
    - public transport (PT) passenger demand elasticities
  - travel demand scenarios (A.8)
    - density multiplier
    - congestion multiplier
    - PT (train and bus) costs relative to driver travel
    - growth in driver (standardised) journey times
  - household income distribution shock (A.9).

After a user defines the scenario, the RLTDM runs its various sub-models. If the scenario is deterministic, each sub-model is executed, and results are produced after the model is run for one round. If the scenario is stochastic, each sub-model is executed for the defined number of runs (the default number of runs is 20).[8]

After running every sub-model, the RLTDM receives the outputs from the deterministic or stochastic scenario and writes them to the Excel spreadsheets.

3.2.2 'population.do'

'population.do' has this structure:

1. Defining dimensions.
2. Initialising the matrices to define the matrix sizes.
3. Defining the assumptions
   a. migration forecasts
   b. labour force.
4. Constructing regional forecasts.
5. Calculating regional forecast aggregates.
6. Calculating household forecasts
   a. forecast by features of population (gender and age).
7. Collect regional household and population number/aggregates and labour-market data.

The RLTDM's population sub-model contains various types of useful information, but these are the main outputs that inform the other sub-models:

- Regional population and household numbers.
- Regional labour-force and employment figures.
- Population of each household type.

3.2.3 'growth_and_incomes.do'

'growth_and_incomes.do' has a similar structure to that of the RLTDM's other sub-models. It starts by defining the required dimensions and initialising the matrices. The structure then contains the following parts:

[8] For each run, all sub-models are executed, but the inputs are not loaded again.

1. Projection of national growth.
2. Projection of regional GDP growth.
3. Projection of national household income.
4. Projection of regional household income.

The RLTDM projects regional household income and regional GDP per capita, which it uses as inputs into the other sub-models.

3.2.4 'HH_vehicle_demand.do'

'HH_vehicle_demand.do' uses population projections and income-growth projections. The file contains exogenous inputs on the probabilities of vehicle holdings by region and type of household living arrangement. These inputs include:

1. historical census data, with trends interpolated
2. income-elasticity parameters, calibrated using a binary (GLM) model
3. region-specific density parameters, calibrated using the same GLM model when the parameters are significant.

The RLTDM uses Equation 3.1 to calculate what simulates demand for vehicles.[9]

$$P_{hh}(VEH = x|VEH = y < x) = \frac{e^{\beta x}}{e^{1-\beta x}}, \qquad \beta_x = f(INC, AGE, WEL\_DUM) \qquad \text{(Equation 3-1)}$$

Equation 3.1 implies that the probability of a household owning x vehicles, conditional on owning y vehicles, is a function of:

- real household income (*INC*)
- the density of the region where the household lives (*DENS*)
- the average age in the household (*AGE*)
- whether or not the household is in Wellington (*WEL_DUM*).

3.2.5 'freight_demand.do'

Freight demand is modelled based on GDP growth by industry (*i*), freight intensity by industry (*tkm*), and a distance/tonne of freight multiplier for marginal increase or decrease in distance travelled by tonne of freight (see (Equation 3.2).[10]

freight_vol(*i*, *j*, *k*, *t*) = GDP(*i*, *j*, *k*, *t*) x freight_intensity(*i*, *k*, *t*) x distance_coefficient(*j*, *k*) (Equation 3-2)

Where

i	=	industry
j	=	region
k	=	mode (sea, rail or road)
t	=	time.

[9] For more details, see Equation 5.1 in Chapter 5 of Stephenson and Zheng (2013).

[10] Freight volume is measured in terms of inflation-adjusted economic value of freight, rather than physical measures, such as tonne-kilometres.

The RLTDM is calibrated to account for differences in distances travelled between regions. This assumes that the value of freight intensity by industry is the same for every region.[11] The model applies commodity-based measures of distances travelled. Distance parameters describe the marginal increase or decrease in distance travelled by a tonne of freight, given the demand for freight in a particular region. See Paling (2008) for these parameters.

3.2.6 'prices.do'

The RLTDM projects vehicle-operating costs related to fuel prices and taxes. These costs are based on world oil-price scenarios and tax-rate assumptions. The model also calculates vehicle prices as a function of:

- New Zealand exchange rate assumptions
- the consumer price index (CPI)
- domestic-transport fuel prices.[12]

For details of the equations used in the model, see Stephenson and Zheng (2013, p. 43).

The model assumes that the rates of fuel tax and RUC, which appear in the model's VKT and cost sub-model, and the price of vehicles will grow at the rate of inflation (average inflation is assumed to be 2 percent per annum).

The RLTDM models domestic-transport fuel prices (the cost paid at the pump) as a function of:

- the exchange rate, adjusted for the international price of oil
- a constant rate for the importer's margin (*MARG*)
- fuel-tax rates
- a constant rate of GST.

The default oil-price assumption is entered in nominal US dollars, so that user-defined assumptions about oil price are not confused with assumptions about the exchange rate and inflation, which jointly determine the cost of oil in New Zealand dollars.

The Mata code is constructed using the matrices described in this section. This code is used in the model's VKT and cost sub-model, to calculate the cost per kilometre of each travel mode. These costs are used to calculate the probability of mode choices in the mode-choice sub-model.

3.2.7 'vehicle_fleet.do'

The vehicle fleet (vehicle demand) sub-model projects the number of vehicles by age, class, and motive technology or fuel type. The fleet (*V*) evolves according to a transition matrix (*W*), which varies by technology type (*i*) and over time (*t*) (see Equation 3-3).

$$V_{i,t} = W_i \, V_{i,t-1} + e_{i,t} \qquad \text{(Equation 3-3)}$$

Equation 3-3 is based on an autoregressive model. The main vehicle types are:

- bus
- heavy commercial vehicle (HCV)
- light passenger vehicle (LPV)

[11] Freight intensity includes the quality of freight, such as whether it is refrigerated.

[12] The base prices of petrol and diesel are the import prices including freight. Petrol and diesel taxes are unit charges (such as excise and levies) and include non-NLTF taxes.

- light commercial vehicle (LCV)
- motorcycle (MC).

'vehicle_fleet.do' initialises the parameters, and then projects the scrappage, registration numbers and fleet numbers for each travel mode.

3.2.8 'VKT_and_cost.do'

'VKT_and_cost.do' is complex for several reasons:

- The file has higher dimension arrays, which are complex to manage, because Mata does not provide an easy way to structure multi-dimensional arrays.[13]
- Mode choice is a sub-model in this Do-file.
- This Do-file switches the regions within matrices, to correct for inconsistencies in how the regions are ordered in the input sources.

3.3 The dimension of the matrices

A major task in Mata is ensuring that matrices are processed accurately, especially because, unlike other programming languages (such as MATLAB), Mata does not provide an easy way to structure multi-dimensional arrays. We have converted multi-dimensional arrays to two-dimensional matrices; however, this complicates multiplication of the matrices required in the RLTDM.

Due to data being available at different times, the datasets have different starting times, which also complicates the matrix operations.

The most common matrix size in 'population.do' is time (t) by age (a).[14] However, the model contains some three-dimensional matrices[15] that have been converted to two-dimensional matrices, using the dimension t by a for rows and regions (j) for columns: (t x a) x j.

[13] Mata does provide multi-dimensional arrays, but the arrays are not easy to use.

[14] Age is between 0 and 100.

[15] reg_pop_f_t_res is an example of one of these matrices.

4 An introduction to using Mata

Stata uses the programming language Mata to perform interactive matrix calculations and add new features to the software.

Mata is different to Stata's common commands, so Stata users need to familiarise themselves with it. Mata's manual provides guidance (StataCorp, 2023), but may be too detailed for users who intend to use only the RLTDM outputs. Therefore, in this section we explain how to get started with using Mata.

In section 5, we recommend developing an 'ado' file, which will provide a unique command and help file for Stata users and will minimise their interaction with Mata.

4.1 Initial setup

4.1.1 When the RLTDM Do-files are already loaded on your computer

To simply run the RLTDM, users can open 'RLTDM.do' and replace line 8 of the Do-file, with the local (current) working directory. Do this by replacing the text between the quotation marks with the directory where 'RLTDM.do' is located on your computer.[16]

4.1.2 When the RLTDM Do-files are sent by email

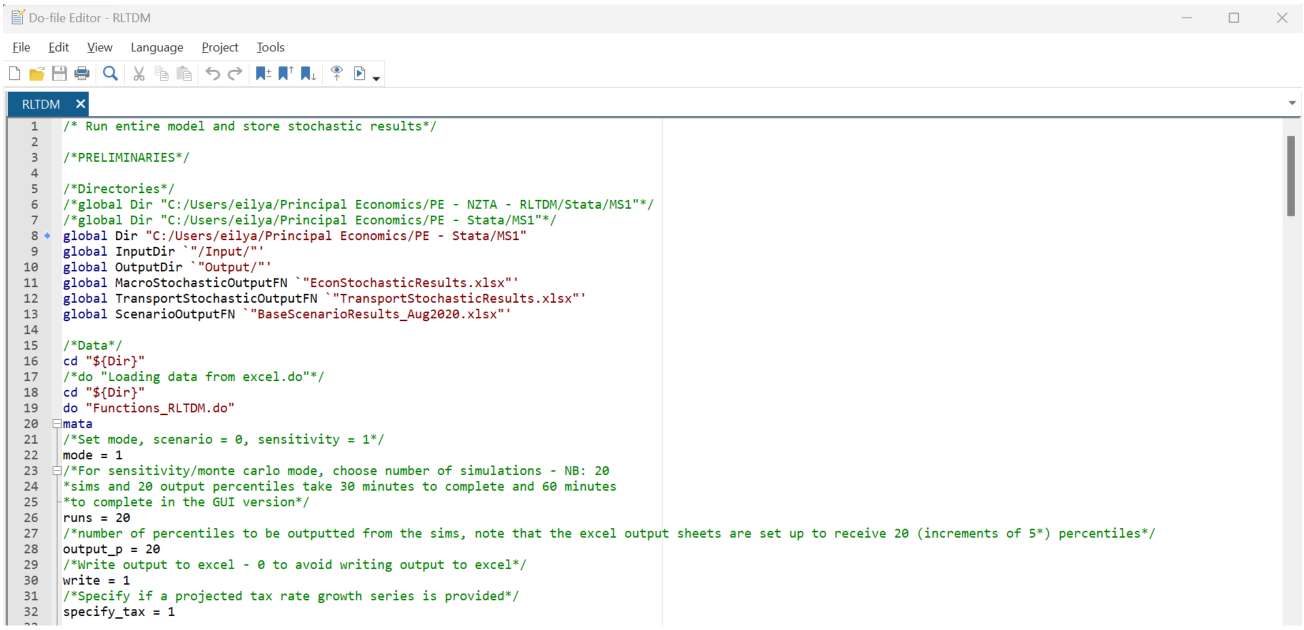
If you have received the RLTDM Do-files as an attachment in an email, follow these instructions:

1. Right click on the attachment in the email (this attachment is called 'RLTDM.zip') and choose 'Save as'.
2. Choose the folder where you want to save the files. For example, C:\users\[your name]\Documents\.
3. Open the folder where you saved the files, using File Explorer.
4. Find 'RLTDM.zip'. Right click on the file and choose 'Extract all'. This will open the folder containing all the RLTDM Do-files.
5. At the top or side of the File Explorer window, select the address bar. Right click on it and choose 'Copy'.
6. Click on 'RLTDM.do' to open the commands in Stata. This will reveal the window illustrated in Figure 4.1.
7. On line 8 of the command, select the text between the quotation marks and replace it with the file path that you copied in step 5.
8. Execute the model by pressing Ctrl + D or using the execute icon at the top of the command window.[17]

[16] For further instructions on using the change directory (CD) command, see www.stata.com/manuals/dcd.pdf

[17] This is the execute icon 

Figure 4.1 Command editor window



4.2 Transfer of Mata matrices to Stata

Mata matrices can be transferred into Stata as Stata matrices or variables. The project team and steering group consider that it is more useful to convert the outputs to variables. However, it is challenging to convert Mata matrices to Stata variables, because the matrices are all different sizes. For example, the rows of one matrix may be years (such as 1992 to 2020), while the rows of another matrix may be regions (such as region 1 to region 12). Therefore, to bring information from Mata to Stata, users should first investigate the size of matrices.

In Stata, the 'getmata' function converts Mata matrices into Stata variables. For example, the following command converts the information in Mata matrix 'pop_by_reg' into variables for 12 regions:

```
getmata ("nor" "akl" "wai" "bop" "tar" "man" "wel" "can" "otg" "sth" "gis" "usi") =
pop_by_reg
```

After converting a Mata matrix into Stata variables, users can process the data using common Stata commands (see an example of this in Table 4-1).

Table 4-1 Example of Stata commands

Command	Action
`g year = _n+1991`	Generates a variable that contains information on time ('g year')
`egen population = rowtotal(nor akl wai bop tar man wel can otg sth gis usi )`	Calculates the national population, using the 'egen population' command
`tsset year`	Introduces the timeseries data to Stata, using the 'tsset' command
`tsline population`	Illustrates the data and projections, using the time series two-way line plot command ('tsline')

5 Recommendations and next steps

Updating the econometric analyses

We did not update the econometric analyses that inform the multipliers used in different parts of the RLTDM, as this was outside the scope of this project. This is particularly relevant to analysing mode choice. For example, the current version of the RLTDM assumes the same cost per kilometre values of each travel mode from 2014 to 2020.

Given that travel behaviour has changed significantly in the past decade, we suggest updating the econometric analyses that inform the model.

Adding travel modes

The RLTDM currently includes an alternative-fuel vehicle category that consists of hybrids, electric vehicles and plug-in hybrids. The model also provides details on the share of alternative-fuel vehicles that are electric vehicles, and the timing of the uptake of electric vehicles. Given the increased emphasis on electrification over the last few years, we suggest making electric vehicles a separate mode to hybrids.

With the recent increase in teleworking, we suggest that teleworking is added to the model's modes. This would, however, increase the computational burden, which is already extensive and time consuming.

Replacing loops for modelling mode choice

We also suggest replacing the loops used to model mode choice with another solution, such as forks. The current loops need extensive computational power, which delays scenario modelling.

Changing the Mata code to an 'ado' file

The current Mata code needs to be executed manually. We suggest changing this to an 'ado' file, which is like a MATLAB graphical user interface (GUI). This would mean that users could shock the RLTDM using the command line, instead of using code. That would make the model more useful for users who are less familiar with Stata or Mata.

Adding an accessibility module to show outputs for different socioeconomic groups

The RLTDM provides a rich platform for modelling modular transport and land use. To align it further with the latest strategic developments and plans, we suggest the model should include an accessibility module that shows outputs for different socioeconomic groups.

Conducting further research on travel behaviour in New Zealand

Policy priorities have changed from the traditional approach – mobility outcomes – to the recent approach − environmental and equity outcomes. This change requires us to generate information about households and incorporate uncertainty into our analyses.

Earlier research shows that transport demand changed significantly following periods of disruption (such as the global financial crisis and COVID-19 pandemic). These types of disruption have significant implications for the success of policy initiatives. Therefore, we recommend further New Zealand-specific research into travel behaviour, including the long lasting versus transitory effects that disruption has on individuals' travel behaviour.

Updating the inputs to make the model more useful for practitioners

When we started this project, some data was only available up to 2020. Given the vast changes that occurred during the COVID-19 pandemic, we recommend updating the model's data to make it more useful for current research. This is particularly important given that recent data shows VKT has increased beyond expected levels. We have already updated the model with MBIE's national GDP and population figures, and the Ministry of Transport's 2022 VKT statistics. We recommend that the RLTDM is updated again, once the 2023 census data is available.[18]

[18] Stats NZ is expected to release the 2023 census data by October 2024.

References

Paling, R. (2008). *National freight demands study*. Ministry of Transport. https://www.transport.govt.nz/assets/Uploads/Report/FreightStudyComplete.pdf

StataCorp. (2023). *Stata 18 Mata reference manual*. College Station, TX: Stata Press. www.stata.com/manuals/m.pdf

Stephenson, J. (2016). *Regionalisation of the National Land Transport Demand Model* (NZ Transport Agency research report 586). www.nzta.govt.nz/assets/resources/research/reports/586/586-regionalisation-of-the-national-land-transport-demand-model.pdf

Stephenson, J., & Zheng, L. (2013). *National long-term land transport demand model* (NZ Transport Agency research report 520) https://nzta.govt.nz/assets/resources/research/reports/520/docs/520.pdf

Stroombergen, A., Bealing, M., Torshizian, E., & Poot, J. (2018). *Impacts of socio-demographic changes on the New Zealand land transport system* (NZ Transport Agency research report 646) www.nzta.govt.nz/assets/resources/research/reports/646/646-Impacts-of-socio-demographic-changes-on-the-NZ-land-transport-system.pdf