The demographics of private motor vehicle kilometres travelled (VKT) in New Zealand

November 2025

Eilya Torshizian, Principal Economics Limited, Auckland

Eugene Isack, Principal Economics Limited, Auckland

Alina Fehling, Principal Economics Limited, Auckland

NZ Transport Agency Waka Kotahi research report 745

Contracted research organisation – Principal Economics Limited

NZ TRANSPORT AGENCY WAKA KOTAHI

Te Kāwanatanga o Aotearoa New Zealand Government

ISBN 978-1-991311-53-5 (electronic)
ISSN 3021-1794 (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., Isack, E., & Fehling, A. (2025). *The demographics of private motor vehicle kilometres travelled (VKT) in New Zealand* (Research report 745). NZ Transport Agency Waka Kotahi.

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

Keywords: demographic features; EV uptake; fleet composition; IDI analysis; motor vehicle registration; vehicle kilometres travelled; VKT.

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.

The views expressed in research reports 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 reports should not be construed in any way as policy adopted by NZTA or indeed any agency of the New Zealand Government. The reports may, however, be used by New Zealand Government agencies as a reference in the development of policy.

While research reports are believed to be correct at the time of their preparation[1], NZTA and 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 reports in isolation from other sources of advice and information. If necessary, they should seek appropriate legal or other expert advice.

In December 2023, the name of Waka Kotahi NZ Transport Agency (Waka Kotahi) was changed to NZ Transport Agency Waka Kotahi (NZTA). References published by the organisation prior to this date retain the previous name.

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.

Disclaimer

These results are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), which is carefully managed by Stats NZ. The opinions, findings, recommendations and conclusions expressed in this report are those of the author(s), not Stats NZ or other government agencies.

Access to the anonymised data used in this study was provided by Stats NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business or organisation. The results in this report have been made confidential to protect these groups from identification.

Careful consideration has been given to the privacy, security and confidentiality issues associated with using administrative and survey data in the IDI. Further details can be found in the privacy impact assessment for the IDI available from www.stats.govt.nz.

The results are based in part on tax data supplied by Inland Revenue to Stats NZ under the Tax Administration Act 1994. This tax data must only be used for statistical purposes. No individual information may be published or disclosed in any other form nor provided to Inland Revenue for administrative or regulatory purposes. Any person who has had access to the unit-record data has certified that they have been shown and have read and understood section 81 of the Tax Administration Act 1994, which relates to

[1] 'This research was conducted July 2023- December 2024.

secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes and is not related to the data's ability to support Inland Revenue's core operational requirements.

For more information about the IDI, please visit https://www.stats.govt.nz/integrated-data/.

Acknowledgements

The authors thank NZTA for commissioning this project through the National Land Transport Programme. We also thank the steering committee members Dr Malcolm Menzies (chair), Sandy Fong from NZTA, Thoa Hoang from the Ministry of Transport and Stephen Christie from the Wellington Transport Analytics Unit.

We want to thank the peer reviewers Professor Jacques Poot, Emeritus Professor at the University of Waikato, and Dr Adolf Stroombergen, Chief Economist of Infometrics.

We would also like to acknowledge Ruyi Jia's assistance with data analysis, the helpful inputs from Daniel Lawrence and Todd Wylie from NZTA and Kain Glensor from the Ministry of Transport on various datasets and Stats NZ's collaboration on our Integrated Data Infrastructure research process. We are also thankful for Ian Binnie's initial input on the scope of the project.

Abbreviations and acronyms

2SLS	two-stage least squares
AADT	annual average daily traffic
ABM	agent-based modelling
ACC	Accident Compensation Corporation
ANCAP	Australasian New Car Assessment Program
APC	administrative population census
CART	classification and regression tree
CBD	central business district
CO_2	carbon dioxide
COF	certificate of fitness
EV	electric vehicle
FUA	functional urban area
g/km	grams per kilometre
GDP	gross domestic product
GHG	greenhouse gas
GIS	geographic information system
GPS	Global Positioning System
GTFS	General Transit Feed Specification
HAPINZ	Health and Air Pollution in New Zealand
HH/hhd	household
HNZC	Housing New Zealand Corporation
HTS	Household Travel Survey
IDI	Integrated Data Infrastructure
inc	household income
IR	Inland Revenue
LPV	light private vehicle
ML	machine learning
MOE	Ministry of Education
MOH	Ministry of Health
MSD	Ministry of Social Development
MSM	Macro Strategic Model
MVR	Motor Vehicle Register

NHI	National Health Index
NO_2	nitrogen dioxide
NVED	National Vehicle Emission Dataset
NZDep	New Zealand Deprivation Index
NZIMD	New Zealand Index of Multiple Deprivation
NZQCF	New Zealand Qualifications and Credentials Framework
OLS	ordinary least squares
PHO	Primary Health Organisation
PM_{10}	particulate matter with a diameter of 10 microns or less
$PM_{2.5}$	particulate matter with a diameter of 2.5 microns or less
PT	public transport
RAMM	road assessment and maintenance management
RLTDM	Regional Land Transport Demand Model
SA2	statistical area 2
SDM	Spatial Durbin Model
SELM	Spatial Error Linear Model
SEM	Spatial Error Model
SLM	Spatial Lag Model
SM	simplified model
SQL	structured query language
TLA	territorial local authority
TMS	Traffic Monitoring System
UCSR	Used Car Safety Ratings
VEMT	Vehicle Emissions Mapping Tool
VEPM	Vehicle Emissions Prediction Model
VFM	Vehicle Fleet Model
VKT	vehicle kilometres travelled
VMT	vehicle miles travelled
VSRR	Vehicle Safety Risk Ratings
WOF	warrant of fitness
XL	extra large
XS	extra small

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

To impact transport users' behaviour and provide an efficient transport system, we need evidence on travel patterns and the role of household features, fleet composition and network characteristics. To date, our understanding of these factors is partial and based on a few data points. This report addresses the gap in knowledge of vehicle kilometres travelled (VKT) profiles across New Zealand and provides information on:

- the demographic factors influencing VKT, the role of fleet composition and contextual factors
- variations in VKT across regions and by the features of roads and the public transport (PT) network
- development of a publicly available VKT database that will be available for future research.

We further investigated the impact of VKT demographics on private vehicle ownership and battery electric vehicle uptake. VKT profiles will help to improve the effectiveness of pricing (and mode shift) policies, which has been identified as critical for estimating the generalised costs between modes. Hence, the findings of the current report will be critical for any assessment of the effectiveness of pricing policies and the other drivers of VKT. Technically, we identified a crucial role in VKT demographics and various modelling frameworks used for informing policy decisions.

We used frontier methodologies to gather various datasets together

We explored the available literature and identified that the current studies are not based on a comprehensive dataset. We used the best available methods to address the complexities of constructing the dataset using various sources, including the Ministry of Transport's Motor Vehicle Register, Stats NZ's administrative population census, NZTA's Rightcar safety ratings, environmental ratings and others. The report explores the technical complexities of gathering this data and describes our solutions. The dataset covers odometer readings between April 2018 and June 2024.

The literature shows significant variation in the impacts of VKT determinants

Our review of the available literature identified a long list of VKT determinants (demographic, vehicle, economic and spatial features) with varying estimated impacts from one study to another. This highlights the importance of establishing an evidence base and local data for studies of VKT patterns across New Zealand regions and communities.

Determining what demographic features and other factors are important to VKT depends on the purpose analysis

If the purpose is monitoring VKT and a small set of variables is desired, our results suggest that fleet composition provides a better prediction of VKT. If the purpose is to model VKT in detail, we identify significant correlations with demographic factors, vehicle and spatial features, and the features of public and road transport networks. If the purpose is to assess policy initiatives (ex-post or ex-ante), all factors matter, as well as further interaction of policy-targeted features of VKT with other factors.

What we found

Consistent with common knowledge, our **descriptive statistics** (not controlling for other factors) show that VKT positively correlates with the number of dependent children and household and individual income. Other observations are as follows:

- The least-deprived groups have higher VKT per household, but deprivation group 10 (most-deprived group) also has higher VKT per household than average.
- The households with one or two people have a lower VKT than the mean.
- VKT is negatively correlated with the age of the oldest member of the household.
- Spatial patterns of VKT indicate a close negative correlation with PT coverage.

- While the most-deprived areas indicate high emissions, there are variations in the emissions observed for the least-deprived areas across the regions.
- Auckland has a relatively high VKT per vehicle for low-safety star vehicles.

After **controlling for confounding factors**, a comparison between our regression results and the descriptive statistics suggests the following:

- While descriptive statistics show a positive correlation between VKT and income, this correlation changes to negative after controlling for other demographic factors. This suggests that income and VKT correlate with other confounding factors and indicate that, while income is not necessarily the cause of high and low VKT, it is a useful target/summary measure for VKT policy.
- The higher VKT of self-employed people drives the initially positive correlation with employment. The correlation with employment became negative after controlling for other demographic factors and being self-employed.
- VKT is higher for households located in the urban core and in areas outside of the metropolitan area, but this relationship changes after we control for local features, particularly PT coverage and access to amenities.
- After fixing the PT coverage (assuming the same level of PT coverage across suburbs and households), petrol vehicle VKT decreases.
- VKT has a significant negative correlation with PT coverage.

VKT is positively associated with being male**,** Māori and Pacific, low-income, older, having more dependent children and having higher education. More interestingly, our **regression analysis** suggests the following:

- An increase in the number of cars is associated with lower VKT. One additional car per household decreases VKT by 6%, and this relationship is decreasing – the VKT is declining below 6% for a third or more vehicles.
- Lower VKT in dense areas is driven by access to amenities and facilities.
- VKT is positively correlated with travel time to the nearest town centre using PT, walking and cycling and negatively correlated with travel time to the nearest town centre using a car.
- The highest VKT due to regional features is for Waikato, Bay of Plenty, Canterbury and Hawke's Bay.
- VKT is lower in most tier 3 urban environments (Gisborne, West Coast and Tasman).

We explored **electric vehicle (EV) patterns** and identified that decreasing the distance to EV chargers by 340 m increases the likelihood of EV adoption by 0.07% – this implies an elasticity of 0.007.

- This relationship decreases with density and income.
- Being in an area with a high population density is associated with a lower likelihood of EV adoption. This is after controlling for compounding factors.
- A 1% higher population per square kilometre (equal to 19.1 people) is associated with a 0.05% decrease in the likelihood of EV adoption.

These **vehicle features** are more significantly correlated with VKT (than demographics) and after controlling for other factors:

- Newer cars travel more, and the relationship is (non-linear and) decreasing.
- Station wagons have higher VKT than other body types, and convertibles have the lowest VKT.
- Electric cars travel more than petrol cars (with the above-mentioned caveats).
- While diesel vehicles seem to travel more than petrol cars initially, the correlation becomes negative after controlling for demographic features.
- Larger cars travel more than other vehicle sizes.

While important, demographic factors do not explain **car ownership** significantly. Although we identify significant impacts on car ownership, the low goodness of fit suggests that a broader range of factors may affect the car ownership decision. This finding helps reduce car ownership policy decisions. We also tested the correlation between these factors and vehicle emissions, excluding VKT. The results suggested a very low goodness of fit (at around 4%). This finding may indicate that users do not consider vehicle emissions in their vehicle ownership decisions.

We established a publicly available anonymised VKT dataset for future research

Using the synthpop R package, we generated a synthetic unit-record file of annual VKT for individuals and their main vehicle relationship that preserves the statistical properties of our original data. Variables included are year of observation, sex of owner, age of owner, usual residents in the household aged under and over 15, household income, region code, household size, highest NZQCF qualification of owner, ethnicity, annual VKT specific to individual and vehicle, vehicle age, vehicle body type, engine size and fuel type. The publicly accessible VKT demographics dashboard is available at https://principaleconomics.com/vkt-dashboard/.

Next steps

Overall, this report improves our knowledge of VKT patterns significantly. The outcomes provide a rich dataset for exploring a wide range of policy and technical questions:

- Applying a similar approach to establish a VKT dataset for freight using Stats NZ’s Longitudinal Business Database will provide an invaluable data source for freight policy.
- Further analysis of the endogeneity of the factors of VKT. This report used a large dataset to establish the correlations between VKT and various aspects. The dataset’s large size and the parameters’ persistence across regression models indicate the robustness of the established correlations. Further analysis will be required to understand the impact of each identified factor. For example, an additional assessment of the correlation between proximity to EV charging stations on VKT suggested a significant role for density, household income and location features. The established dataset provides the information needed for future research on different aspects of VKT.
- A granular analysis of own and cross-VKT elasticities is unavailable in New Zealand. This would be particularly useful for understanding the impact of pricing policies such as tolling.
- Further assessment of the drivers of EV uptake. This report provides initial results to inform EV uptake policies. A future report should focus on specific hypotheses and consider various policy levers.
- Further assessment of local factors of modal share under different policies. The granular data is beneficial for understanding the impact of policies on the combination of price and quantity (revenue).
- Time-of-use pricing policies can be analysed by further disaggregating the explanatory variables and linking the established VKT dataset with other Integrated Data Infrastructure datasets.
- Improving forecasts of VKT and fleet composition by linking the established dataset with an advanced socio-economic model such as the Regional Land Transport Demand Model. This is particularly useful as an input to fleet composition forecasts.
- To inform that modelling (and forecasts of fleet composition) is helpful to improve our understanding of preference changes and social norms that have been identified as critical factors in shaping preferences and ultimately matter significantly to VKT over time.
- The dataset could be further used to investigate the impact of changes in preferences such as working from home, which are critical for VKT forecasts. Another essential factor that could be investigated is the impact of infrastructure availability such as PT coverage and the low-hanging fruit (least-costly policies) for future investment to achieve different policies such as mode shift.

- The impact of policies is often realised with delay. The frequent data available on motor vehicle registration and its linkage to address data could be used to establish an informative local dashboard of VKT that will monitor policy impacts promptly.
- The low explanatory power of demographics, built environment and regional context for vehicle ownership motivates further research into the impact of other factors such as travel demand management policies and the role of social norms on vehicle ownership.

Abstract

This report provides a comprehensive understanding of the vehicle kilometres travelled (VKT) of households across New Zealand, focusing on their demographic and spatial characteristics. We use Stats NZ's Integrated Data Infrastructure to link different datasets, including the Ministry of Transport's Household Travel Survey and Motor Vehicle Register, Stats NZ's administrative population census, NZTA's Rightcar safety ratings, environmental ratings and others to gather the most granular information about households' travel patterns at the household and suburb level. This dataset addressed complexities associated with the timing of recording odometers and derived over 10 million records from 2016–2022. We then provided a range of descriptive statistics on VKT patterns. We observed a positive correlation with having more children, being employed, being Māori and Pacific and having an electric vehicle (EV). We used a Poisson regression model, simplified to a log-log model, to control for demographic, economic, vehicle and spatial features. The results were different from descriptive statistics. For example, the positive correlation with income changed to negative. Also, the higher VKT of EVs compared to petrol vehicles changed to not statistically different after controlling for (fixing) the public transport coverage. We identified that vehicle features are more closely related to VKT than other determinants because vehicle choice indicates the use and purpose of travel. Our preliminary results of an analysis of EV adoption suggest a 0.07% increase in adoption with 340 m improved proximity to charging infrastructure – this is equal to an EV charger elasticity of 0.007. For future research, we generated a synthetic unit-record file of annual VKT for individuals and their main vehicle relationship that preserves the statistical properties of our original data.

1 Introduction

1.1 Improving our understanding of VKT in New Zealand

NZTA commissioned Principal Economics Limited to research and identify the relationship between household demographics and light vehicle kilometres travelled (VKT). The project proceeded in two phases:

- The initial phase of the study consisted of reviewing existing literature on VKT measures and their correlation with household demographics and socio-economic features. During this phase, we identified relevant demographic factors of VKT in addition to existing frameworks and available data for modelling/allocating private light VKT to households' features. The findings from the review provided information about the data required from the Stats NZ Integrated Data Infrastructure (IDI) and other sources and about the preferred demographic characteristics.
- In the second phase, we linked odometer readings to personal and household demographics held in the IDI and constructed VKT profiles for different socio-economic groups.

The findings of this report provide information about:

- the demographic factors influencing the fleet composition and, subsequently, transport outcomes
- the VKT characteristics of different geographic areas, which will help guide targeted interventions and policy development in areas where improvements in public transport (PT) coverage may improve network efficiency
- linkages between demographic factors, spatial attributes, fleet composition and VKT.

By connecting existing administrative and census data, the outputs of this project also provide a base dataset for further research and data linkages for future research on VKT analysis, transport modelling, equity analysis and health impacts. The focus of this report is on light private vehicle VKT.

1.2 Demographic and spatial features of VKT

This report addresses the gap in our understanding of VKT profiles across New Zealand and provides information about the demographic and spatial characteristics of households that are associated with VKT. The outputs will provide information about households' residential location as opposed to where VKT is travelled. This will provide a useful dataset that can be used to determine potential correlations between household features and geographic location and VKT and will provide useful information for the estimation of emissions by household type. The outputs will be helpful in informing the design and helping with the evaluation of initiatives that alter light VKT and its equity impacts.

NZTA previously commissioned Principal Economics to investigate various transport policy levers and their VKT impacts (Torshizian et al., 2025b). The first objective of our analysis was to estimate the generalised costs necessary to facilitate the mode shift from cars/light passenger vehicles to PT. To understand the driver(s) of VKT and the impact of different policies (and their interactions), it is critical to consider the variation in generalised costs with the socio-economic features. Given the lack of this information, we used IDI analysis of own and cross-product elasticities across different socio-economic groups. Based on our extensive literature review (in that report), the cross-elasticities between light vehicles and other modes are very small (and are statistically insignificant).[2] An important finding of that report was that PT coverage is

[2] This is confirmed in New Zealand studies using demand system methodologies (AIDS and QUAIDS) and granular IDI data (Torshizian & Isack, 2020; Torshizian & Meade, 2020).

necessary for mode shift. A useful question motivated by the current report is how the provision of additional PT may impact VKT and VKT growth.

Technically, the combination of personal, household and location information can be used to create a statistical model that identifies the main characteristics of people and households and their vehicles associated with medium and high VKT.

In New Zealand, official data on national VKT is available from two primary sources, each with its own limitations, and their findings may not always align consistently with each other:

- The Ministry of Transport allocates VKT from odometer readings by locality based on the Motor Vehicle Register (MVR) residential address. However, if the region where the vehicle is registered does not match where it is inspected, it is allocated to the inspection location. This information is collected each time a vehicle is inspected for a warrant of fitness (WOF) or certificate of fitness (COF).[3]
- The second primary data source is collected and maintained by NZTA for the road assessment and maintenance management (RAMM) database using road traffic counts collected from the Traffic Monitoring System (TMS) for all public roads in New Zealand (Ministry of Transport, 2019).

As the RAMM database relies on traffic counts, modelling is undertaken to extrapolate traffic counts on each road to assume the total distance travelled. Additionally, traffic counts are not always consistently recorded at each site, leading to time lags between VKT estimates. For assessing the relationship between demographics and VKT, it is not possible to attribute traffic counts to individual vehicles or their owners. This limits the usefulness of using the RAMM dataset for assessing the relationship between demographics and VKT (and its related factors).

VKT estimates are produced quarterly and annually using MVR odometer data. As with NZTA's estimates, VKT estimates from these datasets are intended to reflect where they have occurred, as indicated by the regional reallocation undertaken by the Ministry of Transport. McKibbin et al. (2022) provided further details on the these estimates, noting that the regional observed VKT is estimated based on both odometer readings and RAMM data. Notable issues with the use of MVR odometer data are the inconsistent recordings of WOF odometer readings, particularly for new vehicles, which are tested less frequently than older vehicles (NZ Transport Agency Waka Kotahi, 2025).

The New Zealand Household Travel Survey (HTS) collects information on day-to-day travel in New Zealand, including kilometres travelled, travel mode, trip purpose, trip length, vehicle occupancy and demographic attributes. The survey has been conducted for the years 1989/90 and 1997/98 and then annually since 2003. For the last survey wave of 2019/20, 2,194 households were surveyed. The survey is conducted throughout 2 days of travel. Compared to other VKT collection methods, the HTS provides the most direct link between household light vehicle travel and household demographics. Despite the limited sample size, the advantage of the HTS data compared to the Stats NZ census is that it covers all travel purposes and not only travel to/from work and education.

1.3 Uses for detailed information about VKT patterns

An improved understanding of the role of demographic factors, the features of the network and the vehicles on VKT is critical for improving the efficiency of the transport system, which has a significant role in productivity and economic growth. Various policy questions require a deep understanding of VKT factors. For example, a road tolling policy requires a deep understanding of the VKT response to understand the revenue raised from tolling and its possible implications for social welfare. Another example is an investigation of the

[3] COF is applicable to large motorhomes.

impact of policy levers for the adoption of electric vehicles (EVs) that requires evidence of the factors contributing to the adoption of EVs. In sum, all these factors aim at improving the welfare of society, which might be measured differently using different appraisal approaches, but a deep understanding of VKT will be required for understanding their impact with any approach.

1.4 Our methodology – literature review, big data analysis and econometric analysis of VKT drivers

For a systematic review of the literature, we used the following steps:

- Reviewing existing practices adopted locally and internationally for similar analyses of VKT and its impact on transport outcomes and the methods for disaggregating VKT by socio-economic features.
- Using independent research search engines and databases such as EconLit, Sage, ProQuest, EBSCO, Google Scholar and a general Google search to find other relevant academic and policy literature.
- Using inputs from the project steering committee and the peer reviewers.
- Contacting relevant organisations and authors of relevant research to learn about other (unpublished) available sources.

In our search for relevant literature, we used a range of keywords to find the available studies from different search engines. Then we applied the identified methods for data manipulation and constructed a large VKT dataset providing information about the features of households, vehicles and regions at the granular household level. We used this data to establish a model of VKT and provide preliminary results of drivers of VKT (with initial estimates of the factors of EV adoption). We also provided publicly accessible VKT data extracted from the household-level information.

1.5 Report structure

The remainder of this report is structured as follows:

- Chapter 2 provides an overview of our literature review on the demographics of VKT, its determinants and the transport outcomes.
- Chapter 3 identifies the linkage between VKT and household features and the methods available for the analysis of VKT.
- Chapter 4 describes our gathered VKT data in correlational terms.
- Chapter 5 models the VKT data to provide an understanding of the role of demographic, vehicle and regional features on VKT. This chapter provides preliminary results of an analysis of EV adoption.
- Chapter 6 describes our approach for anonymising and extracting VKT and demographic data. This is a valuable data source for future research.
- Chapter 7 concludes and discusses areas for future research.

Chapters 1–3 focus on the first phase of the project. After concluding the usefulness and feasibility of the project with the steering committee, we progressed with the VKT analysis in the second phase. At the end of chapter 3, we provide a summary of our findings in the first phase.

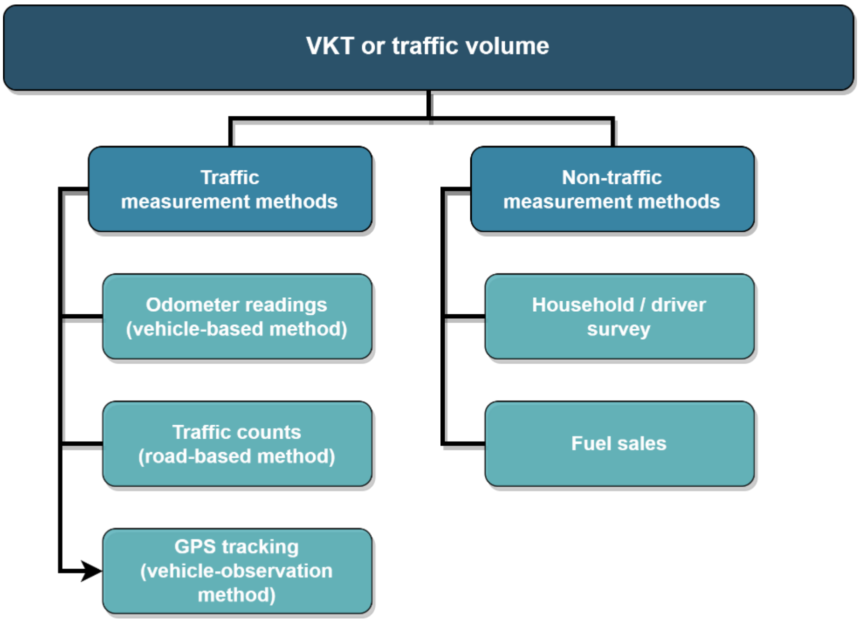
2 Literature review

This chapter provides a literature review of the general approaches to VKT analysis, the factors considered, the available data and transport models, the VKT profile measurement issues, and the feasibility and usefulness of the available data and methods.

2.1 General approaches to VKT analysis

The Bureau of Infrastructure, Transport and Regional Economics (2011) identified five methods of road VKT estimation, classified into two groups of traffic and non-traffic measurements (Figure 2.1). The traffic methods are odometer readings, traffic counts and Global Positioning System (GPS) data. The non-traffic methods are household travel surveys and data on fuel sales. In this section, we first consider the traffic measurement methods and then investigate the non-traffic measurement methods.

Figure 2.1 Road VKT or traffic volume methods (adapted from Bureau of Infrastructure, Transport and Regional Economics, 2011, p. 10)



These methods can be used in combination – for example, both GPS logging devices and paper travel diaries are used to record travel behaviour in the HTS (Ministry of Transport, 2022a).

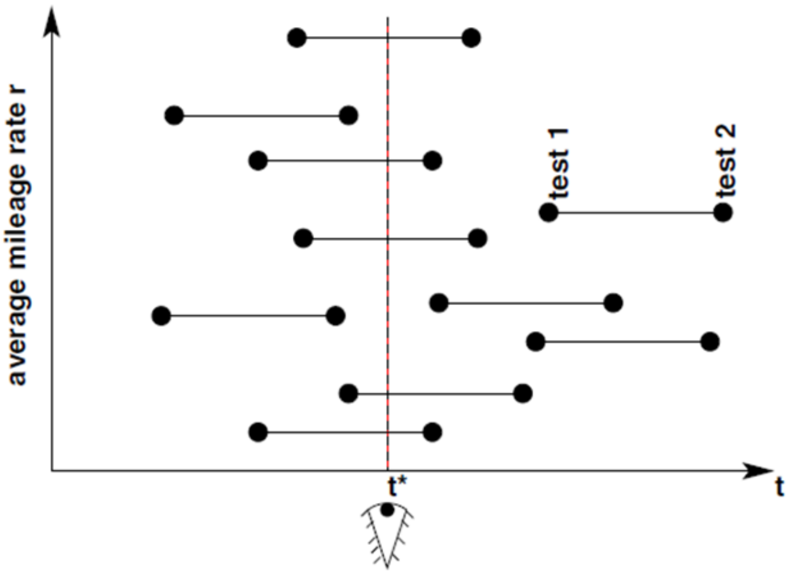
2.1.1 Traffic measurement methods

2.1.1.1 Odometer readings

Odometer readings taken from individual vehicles can provide a direct measurement of the distance travelled over a certain period. Diao and Ferreira (2014) and Aslanyan and Jiang (2021) adopt the Vehicle Census Dataset for Massachusetts (VCDM) in their analysis of vehicle miles travelled (VMT). This dataset consists of data on quarterly average daily passenger VMT and number of vehicles per census tract, which they adopt alongside spatially aggregated demographic attributes, including socio-economic features, vehicle characteristics, built environment and accessibility measures. Similarly, Holtzclaw (1994) and Holtzclaw et al. (2002) used a spatially disaggregate dataset from 1990 US census data that aggregated averages of VMT per household.

Odometer readings for the estimation of VKT are typically undertaken at aggregated scales (Azevedo & Cardoso, 2009; Task Force on Road Traffic, 2007; Wilson et al., 2013). The odometer data is used to calculate the average VKT per day (often annualised) based on the difference between readings. Obtaining a reading between intervals is defined as the straddle point (Wilson et al., 2012, 2013, 2015; Cairns et al., 2014). For each observational interval, the difference in odometer readings and the difference in test rates is computed. By dividing the difference in readings by the difference in intervals, a mileage rate for each interval and an average of all intervals across segments of interest is derived to use as the overall VKT rate. We note that the available literature analysed VKT at an aggregate level based on vehicle type, geographic areas and demographic features. As noted by Wilson et al. (2015), estimating VKT using straddling intervals is limited by the frequency of odometer readings. Where actual VKT may fluctuate significantly over the period of assessment, straddling methods represent variation in aggregate usage over short timescales. Figure 2.2 illustrates the concept behind straddling intervals.

Figure 2.2 Illustration of the concept of straddling intervals (reprinted from Wilson et al., 2013, p. 142)



The UK Department for Transport (2013) created a set of experimental statistics adapting this methodology to individual vehicles. It notes that, while it is possible to create data to estimate values of VKT, care should be taken as newer cars tend to have higher VKT in any given year than older cars. As with New Zealand, vehicle inspection rates in the UK are less frequent for newer cars. The report suggested using other resources such as the National Travel Survey to estimate mileage patterns for the identified data gaps. We show its methodology in (Equation 2.1:

$$VKT_{n+1} = (R_{n+1} - R_n) \times \frac{365.25}{\Delta T} \quad \text{(Equation 2.1)}$$

where:

VKT_{n+1} = estimated annual VKT associated with test $n+1$

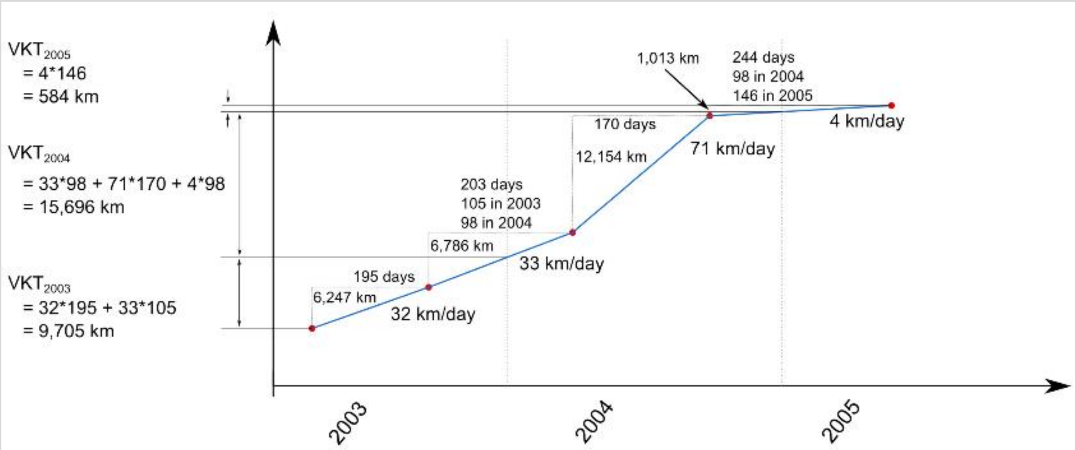
R_n = first odometer reading for the vehicle n in the category *i*

R_{n+1} = second odometer reading for the vehicle n in the category *i*

ΔT = number of days between the first and the second odometer readings (days).

Rendall et al. (2013) adopted this method for estimating geospatial distribution of VKT and fuel usage in New Zealand. Figure 2.3 illustrates the saddle point method for determining VKT over time for individual vehicles.

Figure 2.3 Example of VKT calculation using odometer data (reprinted from Rendall et al., 2013, p. 6)



Azevedo and Cardoso (2009) used a similar method for assessing VKT across vehicle types for the national fleet of Portugal, calculating annual VKT using odometer readings based on **Error! Reference source not found.**:

$$VKT_i = \left(\sum_{n=1}^{N} \frac{\frac{R_n^{T+\Delta T} - R_n^T}{\Delta T}}{N} \times F \right) \times Y \qquad \text{(Equation 2.2)}$$

where:

VKT_i = annual traffic volume for the vehicle category *i* (VKT)

RnT = first odometer reading for the vehicle n in the category *i* (km)

$RnT+\Delta T$ = second odometer reading for the vehicle n in the category *i* (km)

ΔT = number of days between the first and the second odometer readings (days)

N = number of inspected vehicles in the category *i*

F = total number of vehicles of the category *i* in the national vehicle fleet

Y = number of days in a year (365;366).

While the straddle point method provides a useful estimation of mileage rates between periods, it has known limitations of abnormal variations from short, irregular odometer readings. This makes analysis of dynamic, time-dependent changes in vehicle usage challenging, particularly for those occurring over rapid timescales. Additionally, the lack of information on individual mileage rates between the periods generates flawed outputs and missing seasonal trends in the data. Wilson et al. (2013) noted that we cannot know from odometer data how vehicle mileage is distributed between consecutive odometer readings. Furthermore, the distribution is likely to be non-uniform (tested in the second phase). They suggest a methodology using the straddle points available from many odometer readings from their dataset to generate a synthetic dataset with monthly temporal frequency for aggregated groups (areal units, demographic groups). They conclude that, at highly frequent timescales such as weekly and fortnightly, huge sample sizes would be needed to create a synthetic dataset close to the true variation in mileage rates.[4] Regardless, they argue that

[4] This report's large sample size using IDI will benefit from convergence to true values, which decreases any concern with inconsistent estimation.

combining the UK Department for Transport data with other data sources such as census data would be useful. We review subsequent studies using the dataset created by Cairns et al. (2014) in section 2.3.

2.1.1.2 Traffic counts

While traffic counts are not directly within the scope of this report, Jung et al. (2017) outline their use for VKT estimation and we briefly cover it here for comprehension. This multi-step process starts with adjusted 24-hour traffic counts, known as annual average daily traffic (AADT), gathered from traffic volume surveys. VKT is calculated for various road types by multiplying these traffic counts by the road length. For areas without direct measurements, VKT is estimated by subtracting the measured VKT from the total estimated VKT by taking the sum product of daily average VKT per vehicle by fleet size and area multiplied by the number of days in a year.

Azevedo and Cardoso (2009) provide a formula for determining annual VKT of the entire network using a sample of monitored road sections as per (Equation 2.3):

$$VKT_j = \left(\sum_{n=1}^{N}(AADT_n \times L_n)\right) \times \frac{L_{TOT\,j}}{\sum_{n=1}^{N} L_n} \times Y \qquad \text{(Equation 2.3)}$$

where:

VKT_j = annual traffic volume for the road class j

$AADT_n$ = AADT for the section n of the road class j (vehicles/day)

L_n = length of the road section n of class j (km)

$L_{TOT\,j}$ = total length of the class j sections in the whole road network (km)

N = number of road class j segments with available traffic counts

Y = number of days in a year (365;366).

Where target roads are road segments, traffic counts are not observed and traffic volume analysis and statistical methods are used to estimate traffic volume and VKT. The use of traffic counts to estimate VKT therefore has limitations as its counts are not representative of annual VKT in non-classified roads (Azevedo & Cardoso, 2009).

2.1.1.3 GPS tracking

GPS travel data is emerging as a reliable source for measuring VKT on roads. Nasri et al. (2019) used geocomputational methods to integrate GPS data with geographic information system (GIS) road networks and calculated VKT by road segment using 4 months of GPS data. The underlying methodology they adopt is like methods using traffic counts, effectively substituting traffic count data from traffic monitoring sites for GPS data. To estimate VKT for each road segment, the study multiplied the road segment length with traffic volume. Overall, the method was statistically viable, with trade-offs existing between computation time, sample size and accuracy. A notable limitation of this methodology is that access to vehicle GPS trajectory data is currently controlled by private companies, requiring government agencies to purchase the raw samples needed for analysis, which are usually costly, and extensive data cleansing is required given the high geographic granularity and frequency.

GPS logging can also be used as a means for data collection in household travel surveys. The Smartphone-based Individual Travel Survey System was tested as part of New Zealand's national HTS in 2015 (Safi et al., 2019). GPS logging devices continue to be used as part of the VKT measurement in the HTS (Ministry of Transport, 2022a).

2.1.2 Non-traffic measurement methods

2.1.2.1 Household/driver survey

The method involves sending periodic questionnaires to households with one or more cars. Respondents provide details such as yearly kilometres driven per vehicle. Large-scale household travel surveys are a common source for VKT analysis, with notable examples including the Netherlands National Travel Survey (Dieleman et al., 2002) and US and British national surveys (Zamir et al., 2014; Giuliano & Dargay, 2006; Polzin, 2006; Kweon & Kockelman, 2004). Zhao and Li (2021) employed the HTS for their VKT study in Wellington, New Zealand. Where national surveys are absent, smaller-scale studies fill the gap (Bansal et al., 2018; Hou et al., 2013; Majid et al., 2014). These surveys offer valuable data on VKT and its links to household features and travel reasons. However, they have limitations such as restricted sample sizes and durations, which miss infrequent long-distance travel patterns (Federal Highway Administration, 2023).

Household/driver surveys are commonly used as input for more-extensive VKT/transport models, which assume consistent driving patterns over a multi-year period and require infrequent recurring updates in driver numbers or household sizes. Aggregate VKT is calculated by multiplying the average annual mileage per household or licensed driver by the total number of households or drivers (Department for Transport, 2013).

2.1.2.2 Fuel sales

Methods for estimating annual VKT using fuel consumption rely on the assumption that national fuel sales records accurately reflect the fuel consumed in a specific year (Azevedo & Cardoso, 2009). Use of fuel sales as a means of estimating VKT is suggested to have first been used in the US in 1957 (Kumapley & Fricker, 1994). (Equation 2.4) shows the relationship between VKT, fuel consumption and fuel sales:

$$N \times d_{av} \times cons_{av} = FC <> FS \qquad \text{(Equation 2.4)}$$

where:

N = total number of vehicles in national fleet

d_{av} = average annual distance travelled by one vehicle of the national fleet (km/vehicle)

$N \times d_{av}$ = national annual VKT

$cons_{av}$ = average consumption of the nation vehicle fleet

FC = annual estimated volume of fuel consumption for the national vehicle fleet

FS = annual volume of fuel sales in the road sector.

The accuracy of fuel sales for estimating VKT largely depends on the precision of fuel sales and fleet efficiency data. Given the aggregated values it relies on, the method does not allow for disaggregation of VKT by vehicle attributes such as fuel type or engine capacity and fails to differentiate between demographic groups (but it could be useful for a robustness check of aggregate VKT measures at the macro level).

Table 2.1 summarises the methods of VKT estimation and their advantages and disadvantages, which is dependent on the intended purpose of analysis methods. Traffic measurement methods such as traffic counts and GPS tracking are best suited when spatial and temporal dimensions are required but do not allow for explicit linkages to demographic and vehicle attributes. Similar issues are present for fuel sales, which does not allow for demographic and vehicle disaggregation. Travel surveys can explicitly link VKT to demographics but can be costly to collect and are subject to the usual sampling errors. With the availability of administrative data in New Zealand in the IDI, odometer readings have the potential to capture VKT for the entire vehicle fleet with linkages to both demographics and vehicle attributes. Concentrating on user locations instead of vehicle routes and using odometer readings to assess travel distances, demographics and user locations is the most feasible and advantageous method for this analysis.

Table 2.1 VKT estimation methods, advantages and disadvantages

Method	Advantages	Disadvantages
Odometer readings	• Encompasses total distance travelled of entire vehicle fleet • Explicitly links individual vehicle attributes • Explicitly links individuals by vehicle ownership/registration • Regularly collects data as part of wider administrative operations	• Does not allow any association with geographical data regarding where these travelled distances were made • Irregular recording frequency • Possibility of reading/reporting errors, notation/transcription errors and odometer tampering • Vehicle drop-out (scrapped vehicles)
Traffic counts	• High spatial and temporal fidelity	• Prone to sampling and count errors • Not representative of non-sampled road • Subject to bias depending on count sites chosen • Cannot robustly disaggregate traffic by vehicle attributes (fuel type, engine capacity) • Cannot disaggregate between demographics
GPS tracking (telco)[5]	• High spatial and temporal fidelity	• Cannot robustly disaggregate traffic by vehicle attributes (fuel type, engine capacity) • Cannot disaggregate between demographics • Typically, private information must be sourced at a high cost
Household/ driver survey	• Travel explicitly linked to household/driver demographics • High level of detail and flexibility in collected data • Can capture purpose and destination of travel • Can capture non-private vehicle travel methods	• Subject to bias due to misreporting (either deliberate or accidental) • Subject to self-selection bias • Subject to sampling error • Assumes constant driving period over long periods (3–5 years) • High implementation costs
Fuel sales	• Estimates total area-wide VKT • Does not require travel distances	• Accuracy dependent on accuracy of retail fuel sales data and fleet fuel efficiency • Multiple sources of data needed for assumptions used in estimation • Cannot robustly disaggregate traffic by vehicle attributes (fuel type, engine capacity) • Cannot disaggregate between demographics

2.1.3 Available methods and models for the VKT analysis

Various models are used to estimate the relationships between VKT and household demographic features. A more granular data source improves the accuracy of these models. After establishing the models (and updating them), they can be used to evaluate different transport plans and investments.

VKT estimation models can be classified into four broad categories: econometric models, spatial regression models, agent-based modelling (ABM) and machine learning (ML) models. It is important to note that these models are not alternatives. ABMs are simulation models. Econometric models and SRMs are estimation techniques. Furthermore, SRMs are also econometric models. ML models combine estimation and simulation. We further describe these models below.

[5] GPS tracking could be for individuals or vehicles. Individual tracking can be further split into telco GPS and travel survey GPS, each with different pros and cons, which is beyond the scope of this report to further explore.

2.1.3.1 Econometric models

Econometric VKT models consider a wide range of factors, including household demographic features such as income, household size and vehicle ownership. Disaggregate models are typically estimated using regression analysis, which involves estimating the relationship between VKT and a set of independent variables, including household demographic features. The available VKT models usually use four groups of independent variables (Kasraian et al., 2022):

- Socio-economic conditions – demographics, tradition and culture, economy, industry and its structure, trade, tourism, climate.
- Land use – density and distribution of population (urban, rural), density and distribution of industry, density and distribution of tourism destinations, size and distribution of farming land.
- Transport system – available vehicle fleet, transport infrastructure, accessibility of transport, transport policy.
- Individual behaviour of population – mobility, length of trips, source and destination of trips, preferred means of transport.

While econometric models have been widely used for VKT estimation (Hasan & Şimşekoğlu, 2020; Kasraian et al., 2022; Wandani et al., 2018), they have a number of challenges. One challenge is that household travel behaviour can vary significantly within demographic groups. For example, there is a great deal of variation in VKT among households with the same income level. Another challenge is that household travel behaviour can change over time. For example, the rise of ride-hailing services and online shopping has led to changes in household travel behaviour in recent years.

Another challenge for econometric models is the assumption of independence of explanatory variables, known as multicollinearity. Multicollinearity and endogeneity (variables being correlated with the error terms) are critical concepts for the efficiency of estimations.[6] While land use and transport variables control for the spatial distribution of VKT, it is difficult to assume all variables are uncorrelated with the error term. One solution to this problem is using instrumental variables regression and two-stage least squares (2SLS) regression.

Models based on household demographic features are a useful tool for estimating VKT. However, it is important to be aware of the limitations of these models and address them carefully. As a result, it is important to use these models in conjunction with other data sources such as travel surveys and traffic count data to have the instruments required for further testing and addressing potential identification issues.

2.1.3.2 Spatial regression models

Spatial regression models are a type of statistical model that consider the spatial relationships across observations (Anselin, 1988). This can be useful for estimating VKT, as household travel behaviour can be influenced by a variety of spatial factors such as the proximity to jobs, schools and PT. Spatial modelling is not microdata regression modelling but instead an extension of modelling at the regional level. A typical regional data-based model would be the average annual VKT in region r as a function of:

- the share of the adult population of r aged less than 40
- the share of the adult population of r aged 65+
- the average income of households in r
- the population density of region r.

[6] Ridge regression can be very useful in dealing with multicollinearity.

Then the spatial characteristics should be added. Different types of spatial regression models can be used to estimate VKT. Some of the most common types include the following:

- **Spatial Lag Model (SLM):** The SLM is a method for controlling the spatial autocorrelation of a dependent variable when the dependent variable has a spatially dependent relationship. The SLM assumes that the VKT of a household is influenced by the VKT of its neighbouring households. This can be useful for modelling the diffusion of new travel technologies or the impact of changes in land use on household travel behaviour.
- **Spatial Error Model (SEM):** The SEM assumes that the error term in the VKT regression model is spatially correlated. This can be useful for modelling the impact of unobserved spatial factors such as traffic congestion or neighbourhood crime on household travel behaviour.
- **Spatial Durbin Model (SDM):** The SDM combines the SLM and SEM. This can be useful for modelling both the direct and indirect spatial effects of household travel behaviour.

Recent studies show spatial regression models offer several advantages over econometric models (Park & Ko, 2021; Rhee et al., 2016). Spatial regression models can improve their accuracy if the explanatory variables are spatially correlated and can identify spatial patterns in household travel behaviour, which is useful for transportation planning and policy analysis. However, spatial regression models also have some disadvantages. They are more complex and computationally intensive than traditional regression models and they require data on the spatial relationships between observations, which can be difficult to obtain accurately for many variables.

2.1.3.3 Agent-based modelling

ABM is a computational method that simulates the behaviour of agents in a system. Agents are autonomous entities that can interact with each other and with their environment. ABM is often used to study complex systems where the behaviour of individual agents can lead to emergent properties at the system level.

Once the agents have been assigned their attributes and preferences, ABM can be used to simulate their travel behaviour by using a variety of algorithms to model the decision-making process of the agents. For example, ABM could determine how agents choose their travel mode based on factors such as cost, travel time and convenience.

ABM can be used to estimate VKT based on demographic data by simulating the travel behaviour of individual agents. This can be done by assigning agents demographic attributes such as age, gender, household income and vehicle ownership. In ABM, transport users are simulated as individual agents. For each agent, daily travel plans have to be provided that describe the initial behaviour (when to end an activity and how to travel to the next activity location). The agents can adjust their initial behaviour by applying an evolutionary, iterative approach. In every iteration, the plans are executed (traffic flow simulation), the plans are evaluated (evaluation) and new plans are generated (learning) (Horni et al., 2016; Nagel & Flötteröd, 2012; Raney & Nagel, 2006).

Agents can be assigned travel preferences such as mode of transportation and trip destinations. Travellers are assigned to different utility units based on arrival, departure and stay at the activity's location. Staying more at an activity's location generates high utility while leaving early generates less utility. Travelling parameters include cost of using a transport mode per time or per distance (Hamadneh & Esztergar-Kiss, 2022). Staying a long time travelling means travellers lose utility as more money is spent on travelling. The score of the activity and the travel is determined using (Equation 2.5) (Charypar & Nagel, 2005):

$$V_{plan} = \sum_{i=0}^{n} \left(V_{activity,i} + V_{traveling,i,j}\right) \qquad \text{(Equation 2.5)}$$

where:

V_{plan} = utility of performing a selected plan

$V_{activity,i}$ = utility of performing the activity *i*

$V_{traveling,i,j}$ = (dis)utility derived from travelling to and from location of activity *i* by using transport mode *j*.

ABM can also model the interactions between the agents and their environment. For example, ABM could model how agents interact with traffic congestion and respond to changes in the transportation system. Once ABM has been calibrated and validated, it can estimate VKT for different scenarios. For example, ABM could be used to estimate VKT under different transportation policies or different future demographic conditions.

It is important to note that ABM is a complex tool that requires careful calibration and validation and that it is only as good as the data it is based on. Therefore, it is important to use high-quality demographic data when developing and using ABM to estimate VKT.

2.1.3.4 Machine learning

ML tends to create better predictive outcomes compared to more traditional regression models. ML algorithms use historical data, which makes them a good candidate for VKT models. ML algorithms can be used to predict VKT based on demographic data such as age, gender, income and household size.

Some common ML algorithms among researchers are support vector machines and decision trees (Narváez-Villa et al., 2021; van Dijk, 2018). A key benefit of using ML algorithms is that they can be updated with new data as it becomes available, which makes them more accurate over time. This could be done with traditional models too but is usually recognised as an advantage of ML algorithms because ML can better account for unexpected structural changes in relationships between variables.

However, some challenges are associated with using ML for VKT estimation. ML algorithms require a large amount of data to train, they can be complex and difficult to understand and they can be biased, depending on the data that they are trained on. Despite the challenges, ML has the potential to improve the way that VKT is estimated. ML algorithms can provide more accurate and up-to-date estimates of VKT, which can be used to improve transportation planning and policy.

2.1.4 Current modelling/allocating of private light VKT in New Zealand

Travel forecasting in New Zealand mainly focuses on quantifying VKT movements and their paths within the country. As such, little information is available on allocating private light VKT in New Zealand to demographic groups. One exception is the Ministry of Transport Household Travel Model. The model uses 2013 New Zealand census data and population projections from Stats NZ to estimate future population growth in specific demographic groups as a combination of region, location of residence, household type, income vehicle ownership (0, 1, 2 or 3+ vehicles) and age. These groups are then used to project vehicle ownership statistics. The second set of programs then uses results from HTS data to assign trip information such as distance, duration and number of trips (Ministry of Transport, 2017). Tailored surveys are undertaken in the development of transport models in New Zealand. This is evident in their use in the Wellington Transport Strategy Model, Waikato Regional Transportation Model, Auckland Strategic Planning Model and Auckland Regional Transport Model 3 (ART3) (Beca Carter Hollings & Ferner & Sinclair Knight Merz, 2004; Cornelis, 2015; Sinclair Knight Merz, 2008; Smith & Bevan, 2011).

2.1.4.1 Project Monty

Project Monty is a national-scale ABM to provide forecasts of travel behaviour at micro and macro levels. A description of Project Monty is provided in section 2.2.1.2.

2.1.4.2 Regional Land Transport Demand Model

The Regional Land Transport Demand Model (RLTDM) provides a hybrid approach to transport demand forecasting across New Zealand regions. The outputs of the model include deterministic and stochastic forecasts of a wide range of economic and transport series.

The hybrid modelling approach simplifies the relationship between transport demand and macroeconomic aggregates but combines top-down relationships with additional detail on behavioural parameters and often includes reduced (simplified) forms of the conventional regional transport models. Taking this approach, the RLTDM could be manipulated by people with differing degrees of modelling expertise.

The RLTDM's hybrid approach provides researchers and policy advisors with a useful tool for further investigation of the impact of the factors of demand (Torshizian & Fehling, 2024). The model evaluates transport demand scenarios looking out 30 years (Stephenson & Zheng, 2013), taking account of mega-trends in:

- population growth dynamics
- spatial demographic trends
- technology trends
- income and economic growth
- industrial composition
- policy and prices (environmental policy changes, fuel price escalation and volatility).

For the projections of VKT, the RLTDM considers the linkage between fleet information and travel behaviour. In summary the VKT projections are based on historical average vehicle kilometres per vehicle by vehicle type and age (Equation 2.6):

$$\Delta\left({}^{VKT_{i,j}}\!/_{VEH_{i,j}}\right) = \alpha_i \Delta COST_{i,j,t} + \beta_i \Delta INC_{i,j,t} \qquad \text{(Equation 2.6)}$$

where:

i = class of vehicle

j = technology and age combination within each class of vehicle

$\Delta COST$ = change in cost per kilometre for each vehicle

ΔINC = change in average household income in the case of light private vehicles (LPVs) and motorcycles and percent change in gross domestic product (GDP) per capita for other vehicle classes.

Each of these factors (income and GDP) depend on various socio-economic features, including age, household composition, household size and employment status. This is estimated using regression analysis with a constant and error term (Torshizian & Fehling, 2024).

The RLTDM provides projections of transport demand by region and can be used for scenario testing or measuring the uncertainty associated with future transport demand. The RLTDM can be used for scenario analysis of the structural impact on demand drivers such as fuel prices, industry activity, incomes and inter-regional migration. The RLTDM can be used to simulate household composition, VKT, fleet profile and mode choice.

The findings from the current project could be used to further increase the accuracy of the VKT estimates by disaggregating it for the identified variations across income groups. This will inform the parameter β_i in (Equation 2.6).

2.1.5 Factors of VKT demographics

The literature suggests that VKT is influenced by a multitude of factors such as economic development, population growth and household structure, demographic shifts, family income and employment status (Bastian et al., 2016; Grimal et al., 2013; Headicar, 2013; van der Waard et al., 2013). Similarly, a significant number of studies review the significance of built environment, identifying that VKT is affected by household/population density, job density, land-use composition, roading design, accessibility and distance to city centres (Ewing & Cervero, 2010).

Zhao and Li (2021) note that, while there is agreement that demographics, built environment and policy are factors of VKT, there is no consensus on the role each of these factors plays as determinants of VKT. They frame these factors as population structure changes, travel preferences and economic factors and assess the factors correlated with travel distance per trip to disaggregate travel elements into personal trips per day, car trip probability, car ownership and vehicle distance per trip for the Wellington region. Their results suggest a significant positive correlation with employment status and a mixed correlation for house accessibility and road length per person – sign dependent on age. Their results show a negative correlation with built environment determinants, including mixed land use, floor area and road density. The analysis shows how social characteristics, transport demand management and built environment relate to both mode choice and trip making but that these factors also vary by age. Subsequently, these factors affect the total VKT, highlighting the inter-relationship between factors, including demographics.

Table 2.2 shows the combined results of multiple models with making a trip, vehicle travel mode (excludes active mode), vehicle travel distance and car ownership as dependent variables across different models/groups based on age (five categories) and trip purpose (home-based work, home-based non-work or non-home-based). Variables vary in positive and negative correlations depending on age and trip purpose.

Table 2.2 Inter-related factors of travel behaviour and vehicle travel distance (adapted from Zhao & Li, 2021, p. 286)

Categories	Variables	Trip	Vehicle travel mode	Vehicle travel distance	Car ownership
Social characteristics	Household with children	±	±		–
	Household number of 4-wheel vehicles	+	±		
	Employment status	±		+	
	Household income		±		+
	Duration of a single trip	+			
Transportation demand management	Short-term on-street parking space number in local traffic zone		±		+
	Average commuting parking price at the end traffic zone		±		
	Commuting on-site parking space number in local traffic zone				–
	Average commuting on-street parking price in local traffic zone				+
	Number of bus stops in local traffic zone				–
	Number of rail stations in local traffic zone				–
	Frequency of bus		–		–
	Frequency of train		±		

Categories	Variables	Trip	Vehicle travel mode	Vehicle travel distance	Car ownership
Built environment characteristics	Distance to market	±	±		
	Job accessibility	±	±		–
	Service accessibility		±		
	House accessibility		±	±	
	Land area mixed use	–			–
	Mixed use floor area density	–		–	–
	Road length in local traffic zone		±		
	Road length per person in local traffic zone			±	
	Road length per area in local traffic zone			–	
Number of groups (models)		15	15	5	3
		Age* Trip types	Age* Trip types	Age	Trip types

* Signs indicate a certain variable has significant association with one of the four elements for at least one subgroup (age/trip type): + indicates positive association(s) in all significant groups; - indicates negative association(s) in all significant groups; ± indicates mixed effect among age or trip type subgroups.

2.1.6 New Zealand transport impact frameworks

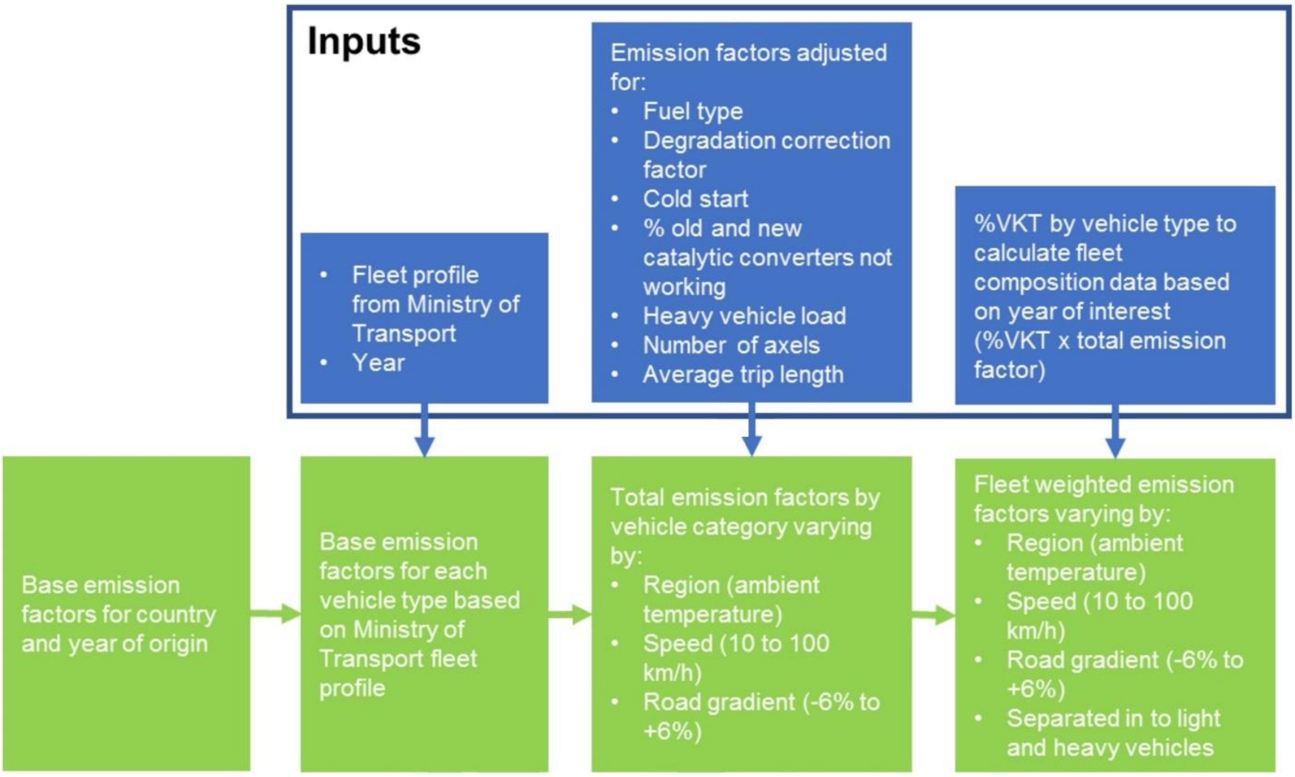
In this section, we provide an overview of the various modelling frameworks that NZTA and the Ministry of Transport use for evaluating VKT and emissions in New Zealand. For each model, we provide a description of the potential use of the outputs of this report to the model.

2.1.6.1 Vehicle Emissions Prediction Model (VEPM)

The VEPM was developed by NZTA and Auckland Council to predict emissions for New Zealand's vehicle fleet under common road and traffic conditions. The model includes emissions factors for the New Zealand fleet based on the different vehicle types and the VKT of each vehicle class. It includes default fleet profiles developed from national fleet data collected by the Ministry of Transport together with predictions of future fleet trends. Fleet-weighted emissions factors are calculated by multiplying the emissions factors in grams per kilometre (g/km) for each vehicle class by the proportion of kilometres travelled by that class for any given year. Updated annually, the model covers various pollutants such as carbon monoxide, hydrocarbons, nitrogen oxides, carbon dioxide (CO_2), PM_{10} and $PM_{2.5}$ as well as brake and tyre particulates. The model provides estimates suitable for assessing air quality and compiling regional emissions data (Metcalfe & Peeters, 2022).

The components of the VEPM shown in Figure 2.4 show how aggregated fleet profile by year and VKT by vehicle type are combined with emissions factors by vehicle category as part of determining fleet-weighted emissions factors in the VEPM. A disaggregated dataset of VKT for individual vehicles allows for determining emissions for each vehicle using the same methodology. Effectively, this is a substitution of the fleet profile and %VKT by vehicle type assumptions shown in Figure 2.4. It should be noted that, as the scope of our research does not determine vehicle travel paths, high-level assumptions will need to be made to account for roading factors such as speed and road gradients.

Figure 2.4 Calculation steps of fleet-weighted emissions factors in the VEPM (reprinted from Tonkin & Taylor, 2021, p. 3)



2.1.6.2 Vehicle Emissions Mapping Tool (VEMT) and National Vehicle Emission Dataset (NVED)

The VEMT is an automated system for calculating harmful air pollutants and greenhouse gases (GHGs) from motor vehicles on all public roads in New Zealand. Using a GIS framework, it maps emissions at highly granular road segment resolution.

The dataset used in the VEMT includes road centrelines and is based on traffic counts and road gradients derived from the Digital Elevation Model from Land Information New Zealand. These datasets are combined with roads and are divided into 50 m segments, each assigned a speed and gradient. Using this information, emissions factors for light and heavy vehicles are applied, derived from the VEPM table. It additionally incorporates the average air temperatures of territorial authorities as inputs into the VEPM. Emissions for each road section are then calculated in g/km per day.

The model determines emissions for all road sections in New Zealand, providing a spatial view of the emissions generated by motor vehicles (Hastings, 2016, 2018). The fleet profile projections used in the VEPM are determined by the Vehicle Fleet Model (VFM). The model determines emissions for all road sections in New Zealand, providing a spatial view of the emissions generated by motor vehicles. Further, the fleet profile projections used in the VEPM are determined by the VFM, which we describe in section 2.1.6.3.

Outputs from the VEMT are used to generate the annual NVED. The data inputs and a description of its use in the VEMT and NVED are shown in Table 2.3.

Table 2.3 Data inputs used in the VEMT (reprinted from Hastings, 2018, p. 6)

Parameter	Source	Database layer	Database field used	Date	Notes
Traffic Count	Core Logic	CoreLogic_RAMMCentreline	trafficVolume	2016	This attribute is provided from Core Logic's RAMM_ONRC dataset, but is brought directly through from the Transport Agency or TLAs own RAMM databases. It represents the average daily traffic count of a particular road section.
Fleet Profile	Core Logic	CoreLogic_RAMMCentreline	hvyVehicleVolume	2016	This attribute is provided from Core Logic's RAMM_ONRC dataset, but is brought directly through from the Transport Agency or TLAs own RAMM databases. It represents the average daily traffic count for heavy vehicles of a particular road section, and is used in conjunction with the trafficVolume attribute to calculate the light/heavy vehicle ratio.
Speed	Core Logic	CoreLogic_Centreline	routableSpeed	2016	This attribute is provided from Core Logic's Road Network dataset. It represents a realistic speed that a vehicle can traverse the road segment, and incorporates speed constraints such as surface type, intersections, roundabouts etc.
Gradient	LINZ	Nidem_25nztm.img and Sidem_25nztm.img	*derived*		This is a raster dataset that has an elevation value every 25 metres. The road centrelines are overlaid over this elevation surface to derive the gradient of a particular road section.
TLA Boundaries	Statistics NZ	TLABoundaries	Region	2015	These boundaries are used to divide the processing into TLA areas (ie regional council and unitary authorities).

The VEMT/NVED provides an alternative view of vehicle emissions in New Zealand, focusing on the location of where they are emitted based on the roads that individuals are travelling. A notable feature of the VEMT/NVED is that the calculation of emissions accounts for vehicle movement patterns and roading features such as travel speed and road gradients. Our report adds to the information available for vehicle emissions in New Zealand by providing an understanding of the geodemographics of who is contributing to vehicle emissions and their location of residence, while the VEMT/NVED provides insight on where emissions are occurring.

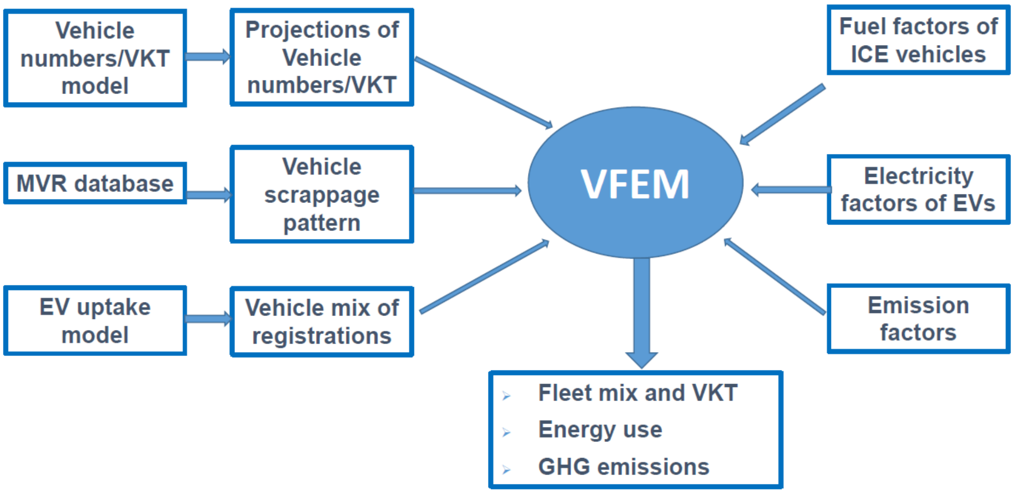
2.1.6.3 Vehicle Fleet Model (VFM)

The VFM forecasts the future make-up of the vehicle fleet as well as travel patterns, energy consumption and GHG emissions. The fleet profile projections of the VFM are used in the VEPM (Ministry of Transport, 2022c).

The model projects the future composition of the vehicle fleet up to 2055 using historical fleet data. It calculates each year's fleet by first determining the number of vehicles that will survive from the previous year based on recent scrappage rates. It then estimates how many new vehicles will be registered across different vehicle categories that match the VKT/Vehicle Numbers Model. Finally, the mix of new registrations is determined by fleet compositions, which include variables like age, fuel type and engine size. A separate model predicts the rate of future EV adoption.

The VFM also uses projected annual travel data from the VKT/Vehicle Numbers Model and distributes it across the fleet based on historical travel patterns. It then calculates the liquid fuel and electricity needs for each vehicle type in each year (Ministry of Transport, 2022c). Figure 2.5 illustrates the structure of the VFM modelling framework (now renamed from Vehicle Fleet Emissions Model).

Figure 2.5 Vehicle Fleet Emissions Model (reprinted from Ministry of Transport, 2022c, p. 4)



2.1.6.4 Health and Air Pollution in New Zealand

Health and Air Pollution in New Zealand (HAPINZ 3.0) assesses the impact of air pollution on public health. The model calculates yearly average concentrations of $PM_{2.5}$ and nitrogen dioxide (NO_2) for different regions of New Zealand and includes health and social costs of human-caused air pollution from cars, domestic fires, windblown dust and industry. It also accounts for natural sources like sea spray and secondary particulate matter. Its outputs include measurable outcomes such as hospital admissions and premature deaths.

The HAPINZ data on premature deaths due to motor vehicle air pollution is available at the territorial local authority (TLA) level. HAPINZ measures emissions where they occur using data air quality monitoring sites across New Zealand for PM_{10} and $PM_{2.5}$ emissions and results from NVED for NO_2 emissions (Kuschel et al., 2022). Similar to the VEMT, the HAPINZ outputs focus on the location of emissions, while our project provides information on the residential area of vehicles contributing to emissions. Integrating the default health effect values determined as part of the HAPINZ 3.0 Health Effects Model with the VEPM and the VKT estimates from this report will provide information on health impacts from emissions both spatially and by vehicle type.

2.1.7 International practice on modelling/allocating private light VKT and use of household features

The allocation of private light VKT in transport models both in New Zealand and internationally predominantly relies on household travel surveys, complemented by calibration through supplementary datasets such as odometer readings and traffic volume counts (KPMG & ARUP, 2017; Ministry of Transport, 2022b; Stephenson, 2016).

The UK Department for Transport similarly uses its National Travel Survey for demographic travel information, including trip frequency, vehicle ownership and employment by household structure. The model follows a four-stage modelling approach of trip generation, trip distribution, modal split and highway assignment (Department for Transport, 2020).

The Southern California Association of Governments ABM also uses a household travel model as a basis for deriving household activity by demographic groups. It uses a multiple discrete-continuous extreme value model to jointly predict activity participation for all household individuals by activity purpose. These are then used as inputs in simulation modelling of activity generation and scheduling for activity-travel patterns by individuals over 24 hours (Bhat et al., 2012).

Reviewing the relationship between household income and transport use, the Bureau of Infrastructure and Transport Research Economics (2011) used three datasets for its analysis. This included a household travel survey, a household, income and labour survey to measure commuting time and a household expenditure survey to measure fuel expenditure.

Our review of the available literature on the use of VKT demographics for transport appraisal suggests that this is a current effort across different jurisdictions. According to the International Transport Forum (2022) roundtable discussions regarding the expansion of transport appraisal, the objectives of transport planning have expanded to meet non-traditional objectives such as long-run sustainability, urban liveability, decarbonisation, resilience and social equity. Cost-benefit analysis has been argued to have been inadequate in addressing this either through undervaluation or not being fully quantified or monetised. The discussions underscore the need for enhancements in cost-benefit analysis to account for non-traditional objectives effectively. While acknowledging its limitations, the International Transport Forum recommends supplementing aggregate cost-benefit analysis with purpose-specific measures. Results of broader appraisals that assess performance against a more comprehensive range of policy objectives cannot be summarised into a single metric. Methods should be chosen based on their capacity to address the specific issues at stake clearly. Also, the presentation of the relative performance of various options to policy makers should be transparent and concise. Various reports have identified the misalignment between the evaluation framework and the new planning paradigm (Litman, 2013, 2024; Transport Planning Society, 2020).

As discussed by the Transport Planning Society, the current transport appraisal system is inconsistent with the requirements of recent transport policies:

> *The current systems of transport appraisal, forecasts and modelling do not reflect current realities and priorities, notably decarbonising transport, support for disadvantaged people and communities and the promotion of active travel. The Government should conduct a fundamental reform of these systems and the business cases that result from them to ensure they support and deliver transport policy objectives.* (Transport Planning Society, 2020, p. 5)

Unsurprisingly, the international practice of adopting VKT demographics to address policy issues is limited.

2.2 IDI, data and transport models

2.2.1 Transport demand models

2.2.1.1 Regional Land Transport Demand Model (RLTDM)

Stephenson and Zheng (2013) developed a national long-term land transport demand model (NLTDM) and Stephenson (2016) then enhanced a national demand model for regional analysis (RLTDM). The RLTDM is intended to be used to construct quantitative long-term (30-year) regional transport planning scenarios.

The RLTDM illustrates the socio-demographic and transport demand projections in 12 regions and contains considerable socio-demographic detail to connect long-term transport demand to primitive drivers of demand such as numbers of people in a region and their age or industrial composition of a region. The RLTDM can be used to measure uncertainty associated with future transport demand given observable historical drivers of uncertainty in the economy.

Stephenson (2016) found that regional sensitivity to costs of travel reflects several underlying factors including population composition and relative incomes. People who live alone (often retired people) and solo-parent families tend to be more income-constrained than others and exhibit greater sensitivity to travel costs than other households. VKT of households in Northland are the most sensitive to changes in travel costs and people in Taranaki and Auckland are the least sensitive. In densely populated and urban regions, where speeds are relatively low, increased speeds (through new infrastructure or increasing speed limits on arterial routes) have a positive effect on the number of journeys, but in rural areas, where speeds are relatively high, an increase in speeds reduces the number of journeys.

Stephenson (2016) found that a 10% increase in the cost of driver travel per dollar of household income causes an average 0.2% reduction in driver trips, which is very low. When driver trips are reduced, most of the change is caught in increased passenger trips. The proportionate change in passenger trips (0.4%) is similar to the proportionate increase in PT (bus and train) use but passenger volumes are a much larger absolute share, accounting for 26% of journeys compared to 2% for PT.

Some key results from this study include:

- stronger mode substitution in urban areas such as Auckland and Wellington – where population density is higher and PT accessibility is higher
- overall sensitivity to changes in costs appear largest in regions with lower incomes
- mode choice in Wellington stands out as being remarkably unresponsive to changes in travel costs.

Stroombergen et al. (2018) used the RLTDM to assess the impact of a range of scenarios. Their analysis of the impact of a lower fuel price suggests an increase in both PT and private vehicles – because its income effect is larger than the substitution effect. For their analysis of the impact of higher density, they increased the population density of the Auckland and Wellington regions by 10%, over and above what occurs endogenously in the base case. They held population constant overall so that density in other regions decreases. Their results suggest that a 10% increase in density leads to the following:

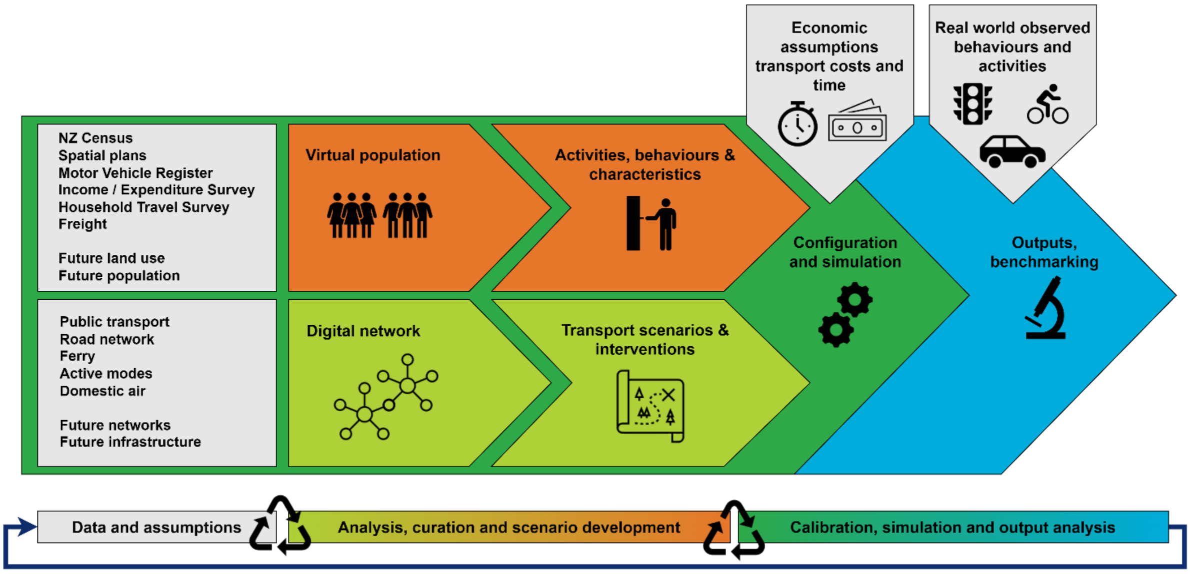
- An increase in PT use and active modes in Auckland and Wellington and, alongside that, an increase in VKT in the regions that experience a corresponding decline in density. At the national level, results show an increase in demand for travel by private vehicle by 2 billion km (or 3.2%) by 2050. The increase in VKT is dominated by Auckland and Wellington – by 2050, it is 18.3% above the base case, implying that total VKT declines in other regions by 4.1%.
- A 12.4% increase in national demand for PT (which is significantly more than VKT). The PT demand increases in Auckland and Wellington, where PT demand is 16.2%, implying a reduction in PT demand in other regions by 4.8%.

2.2.1.2 Project Monty

In partnership with ARUP, the Ministry of Transport has developed a national-scale ABM to provide forecasts of travel behaviour at micro and macro levels. For residential travel, Project Monty links travel behaviours from the HTS to New Zealand census data and roading network. It simulates travel choices made by each agent (virtual person) in undertaking their daily transport activities such as travel to work, school or shopping. The modelled choices are additionally affected by travel costs and travel time for different travel modes (ARUP, 2020).

By integrating individual data with household travel patterns and the roading network, Project Monty can simulate travel patterns at continuous national scale allowing for analysis of VKT by demographic groups and the estimation of emissions based on their travel patterns. Figure 2.6 illustrates the high-level data inputs and methodology used in Project Monty.

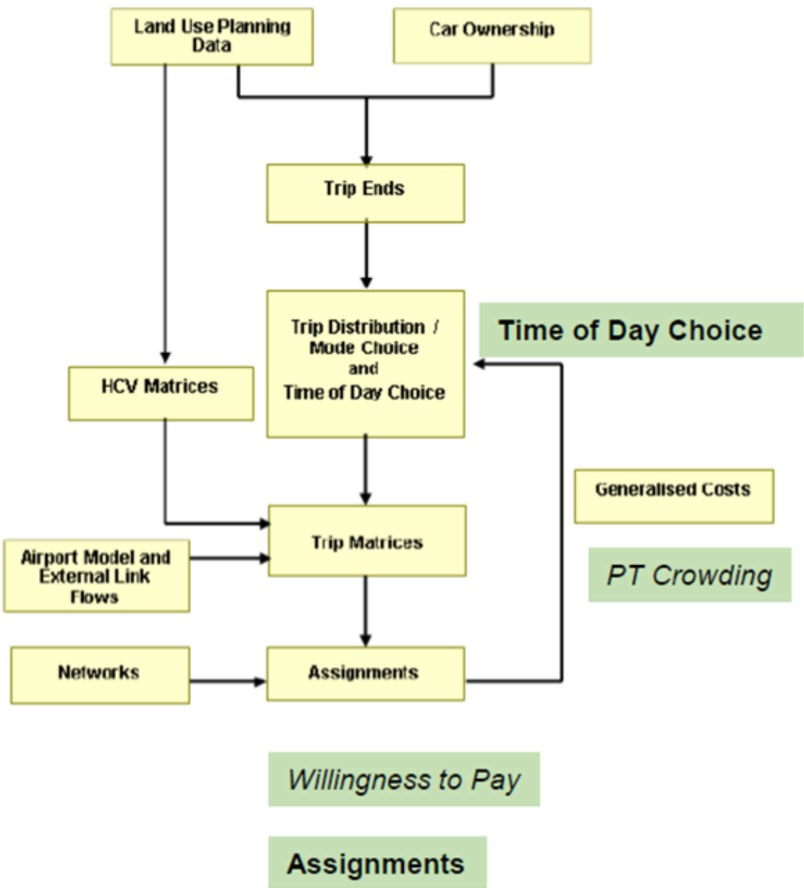
Figure 2.6 Schematic overview of Project Monty (adapted from Ministry of Transport, 2022b, p. 7)



2.2.1.3 Regional models

The Auckland Regional Transport Macro Strategic Model (MSM), previously named the Auckland Region Transport Model ART3, is a four-stage transport model used in Auckland to predict zone-to-zone trip generation. Figure 2.7 illustrates the model structure.

Figure 2.7 MSM distribution and mode choice model structure (reprinted from Valero, 2018, p. 6)



The Auckland MSM is a bottom-up model that considers multi-modal access (including PT, walking and cycling) and trip purposes (including home-based work, home-based education, home-based shop, employer business, home-based other and non-home based other). The model is part of the integrated transport and land-use models in Auckland, which includes the Auckland Strategic Planning model (ASP3.2) (Sinclair Knight Merz, 2008). The MSM uses separate distribution and mode choice models for each trip purpose and distributes trips by mode based on generalised travel costs for each trip purpose, time and destination attractor. MSM outputs provide counts of multi-modal trips between origin and destination transport zones over a 24-hour typical weekday period. Consequently, VKT derived from the MSM excludes weekends. Linkages between VKT and demographics using the MSM outputs are limited to transport unit level by comparing estimated travel distances to external demographic data such as the census. Emissions estimates can be determined using the MSM's outputs, incorporating VEPM emissions factors and distance-duration data from the MSM to calculate VKT and speed parameters (Torshizian et al., 2022).

2.2.1.4 Simplified models

A simplified transport strategy model (SM) was developed for Auckland, Wellington and Christchurch (Torshizian et al., 2025b). Unlike more complex, traditional transport models, the SM offers speed and flexibility, facilitating quick analyses and enabling the rapid evaluation of multiple scenarios. Its streamlined design focuses on key objectives rather than delving into detailed variables and market segments. The SM differs from conventional, more detailed transport models in several ways:

- **Speed and flexibility:** The SM's simplified approach enables quick analyses, allowing for the rapid testing of multiple scenarios and strategies.
- **Focus:** While detailed models often include multiple variables and market segments, the SM is more streamlined, concentrating on key objectives such as reducing light vehicle VKT and GHG emissions.
- **Level of detail:** The SM operates at a higher, more aggregated level, not diving into the minutiae often found in conventional models.

The SM is designed to complement existing, more detailed regional models and serve as a first-step tool for evaluating the effectiveness of potential transport strategies, providing directional guidance for more in-depth studies. Linkages between VKT and demographics using SM outputs are limited to transport unit level. Emissions estimates can be determined using the SM's outputs, incorporating VEPM emissions factors and distance, average speed outputs from the SM as inputs for VKT and speed parameters.

2.2.2 Sources of data

In New Zealand, observed VKT is estimated through two primary methods: odometer data from the MVR and road traffic counts from the RAMM database. Odometer readings are gathered at each vehicle's inspection (and change of ownership) and aggregated to offer VKT estimates by vehicle type and region. RAMM employs road traffic counts on state highways and local roads for its estimates. These counts are collected by NZTA and local authorities and used to model odometer readings by road type and by region. RAMM separates data for light and heavy vehicles but only on state highways, leaving a gap in information for local roads. Both datasets focus on producing VKT defined by where they are generated. In the case of odometer data, adjustments are made to infer regional VKT from the location where the vehicle was inspected, leading to potential inaccuracies from regions with high tourism counts (locations where vehicles are registered in one region and driven in another). Additionally, as odometer readings vary significantly in timing, adjustments are made to assign VKT to individual years (Ministry of Transport, 2019).[7]

[7] Rental cars are recorded as commercial vehicles. Uber cars are not necessarily registered as commercial vehicles.

McKibbin et al. (2022) determined a standardised approach for the distribution of national light vehicle VKT and GHGs across New Zealand tier 1 and 2 urban areas and reviewed the VKT data in New Zealand, including sources, continuity, granularity and coverage. A summary of their findings on key existing datasets relating to VKT and vehicle emissions in New Zealand is shown in Table 2.4.

Table 2.4 Key datasets and models on VKT and corresponding GHG emissions (adapted from McKibbin et al., 2022, p. 12)

Source	Dataset	Description	Continuity	Granularity	Coverage
NZTA	TMS	Records of AADT for continuous traffic count sites across the state highway network in New Zealand.	2012–2022	Links	Localised sites
	RAMM	Link level traffic counts observed (where there are count sites) and estimated for all roads across New Zealand. Annual VKT estimates at TLA/regional level.	2002–2020	TLAs	Nationwide
	Projected VKT	Quarterly VKT projection by vehicle type at national level. Estimated from econometric models by vehicle type. Input data includes GDP, employment, imports, exports and demographics information sources from Stats NZ. Includes RAMM and not HES.	2002–2050	Nation	Nationwide
	Emissions model	The VEPM provides estimates that are suitable for air quality and GHG assessments and regional emissions inventories.	2001–2050	N/A	N/A
	Emissions dataset	NVED, emissions estimates for all public roads, taking account of traffic count, fleet profile, speed and gradient.	2019 and 2020	Links	Nationwide
Ministry of Transport	Quarterly observed VKT	Quarterly VKT estimates by vehicle type at regional level. Estimated from WOF/COF testing station observed odometer reading. This is also the latest VKT estimates produced by Ministry of Transport.	2002 Q1 – 2021 Q2	11 regions	Nationwide
	Regional observed VKT	Annual VKT and vehicle number estimates by vehicle type at regional level. Estimation based on odometer readings from MVR and road use from RAMM data.	2012/13–2018/19	14 regions	Nationwide
	Projected VKT	Annual VKT and vehicle number estimates by vehicle type at regional level. Estimated from transport outlook household travel model, which provides projection of household travel in person kilometres by various modes.	5-year increments (2022/23–2057/58)	14 regions	Nationwide
	Emissions model	The VFM projects the make-up of future vehicle fleets and kilometres travelled, energy use and GHG emissions.	2001–2055	Nation	Nationwide
Local government transport models	Auckland, Christchurch, Wellington	Three regional transport models were available. The models provide projections of vehicle travel, including vehicle volume and speed predictions at a link level for light and heavy vehicles.	AKL 2018, 2038; CHC 2018, 2038; WLG 2013, 2036	Links	Urban centres
Stats NZ	Census	Census estimated and projected household by type and population by age group at statistical area 2 (SA2) level.	5-year increments (2018–2058)	SA2s	Nationwide

2.2.3 Measurement issues

For accurate measurement of VKT demographics, the minimal data requirement is to have accurate odometer readings, address information for the vehicle owner and socio-economic features of the owner's household when available in the IDI. However, there are challenges with linking these datasets.

2.2.3.1 Address information

Each vehicle should be attributed to an individual/household through the address information. The New Zealand census provides accurate addressing information but is only held every 5 years. Sourcing accurate addressing data for the years between censuses is challenging. Use of administrative data alleviates the issue of addressing collection between census years, with some limitations. The challenges of consolidating address data from various sources into a single reliable dataset that can be found in the IDI is noted and reported on by Stats NZ (2022). The IDI pools address data from at least nine government organisations, each with its own level of accuracy and update frequency. While the existing methodology focuses on using the most recent address as the most reliable, it also allows for data from less-trusted sources to be overridden by more credible ones (Anastasiadis et al., 2020). Table 2.5 shows the accuracy of using administrative data to derive individual address. Address information from Census 2013 is assumed to be a true and accurate source and is used as a baseline for other datasets providing address information.

Table 2.5 Percentage of individuals with address information matching Census 2013 by administrative source (adapted from Stats NZ, 2017, p. 34)[8]

Data source	Address ID	Meshblock	Area unit	TLAB	TA
IR	66	68	72	86	91
IR (excluding outlier dates)	67	69	73	86	91
NHI	75	77	80	90	93
PHO	73	75	79	88	93
MOE	66	68	73	87	91
ACC	67	69	73	86	90
MSD residential	66	68	72	84	89
MSD postal	27	29	36	61	68
Combined address	82	84	87	93	96
Any source[9]	86	89	91	95	97

In studies using administrative records and Census 2013 data, the address ID from the census is not present for 11% of individuals in any prior administrative records.[10] This limitation caps the individual-level accuracy of such data to a maximum of 89%. Anastasiadis et al. (2020) attempted to enhance the quality of address information in the IDI using alternative cleaning rules and validation methods against Census 2013. However, their methodology found no improvement over the existing address table in the IDI. As part of the

[8] As no historical back series for NZTA addressing is available in the IDI, no comparison could be made against Census 2013 addresses. However, comparisons with the Household Labour Force Survey suggested that the information may be higher quality than any other sources, particularly for age groups that are otherwise a difficult source (Stats NZ, 2022).

[9] Percent of total linked IDI census population (including those with no address).

[10] That is not necessarily because the administration data is wrong. For example, suppose a visit to a hospital in March 2012 yields an address for a person and the person changes address between March 2012 and March 2013. In that case, the Census 2013 address will be different but there are no errors in the measurements of the person's address.

study, the authors evaluated the accuracy of address notifications across various datasets in addition to Census 2013, including the General Social Survey (GSS), Household Economic Survey (HES) and Household Labour Force Survey (HLFS). We show the results of their address accuracy evaluation in Table 2.6. It is important to note that each source has a different population size and hence accuracies within each source, not the accuracy of the entire (might be assumed the same) population.

Table 2.6 Accuracy and age-range by administrative source (adapted from Anastasiadis et al., 2020)

Source	Age range	Number of notifications	Average notifications per person	Accuracy
ACC	Birth–100+	34,884,723	6.2	82.4%
HNZC	Birth–100+	35,279,364	57.1	90.5%
IR-applications	Birth–100+	32,832,552	3.5	78.1%
IR-timestamps	16–75	2,893,953	1.3	58.1%
MOE	Birth–60	1,344,450	1.0	81.2%
MOH-NHI	Birth–100+	30,941,499	4.1	84.7%
MOH-PHO	Birth–100+	24,988,050	4.2	83.5%
MSD-residence	Birth–100+	18,476,316	3.7	82.1%
MSD-postal	16–100+	1,968,099	1.6	55.1%
MSD-partner	17–70	490,749	2.1	79.8%
MSD-child	Birth–18	3,789,891	4.3	73.0%
NZTA	15–100+	8,741,064	1.8	

Investigating commuter patterns over time, Fabling and Maré (2020) constructed a commuter address table for workers in New Zealand, including workplace and residential locations. The dataset is restricted to workers only and can be accessed by approved researchers. They use an alternative prioritisation method to Stats NZ and make additional adjustments to compensate for potential errors due to changes in meshblock boundaries and coordinate-based matching. Their analysis suggests that adjacent meshblock is a more suitable matching variable for residential address data than the use of x-y coordinates that is sometimes employed for address linking in the IDI. Table 2.7 shows the accuracy of their address matching method.

Table 2.7 Comparison of Fabling and Maré prioritisation methods versus IDI (adapted from Fabling & Maré, 2020, p. 44)

Method	Geographic unit	Census night address	Elsewhere in New Zealand	No fixed abode	Total
Fabling and Maré (2020)	Meshblock	85.7%	70.4%	28.7%	85.4%
	TA	96.0%	87.6%	55.2%	95.7%
IDI	Meshblock	83.1%	69.9%	27.6%	82.8%
	TA	94.9%	87.5%	54.0%	94.7%
Gain over IDI Method	Meshblock	2.6%	0.5%	1.1%	2.6%
	TA	1.1%	0.1%	1.1%	1.1%

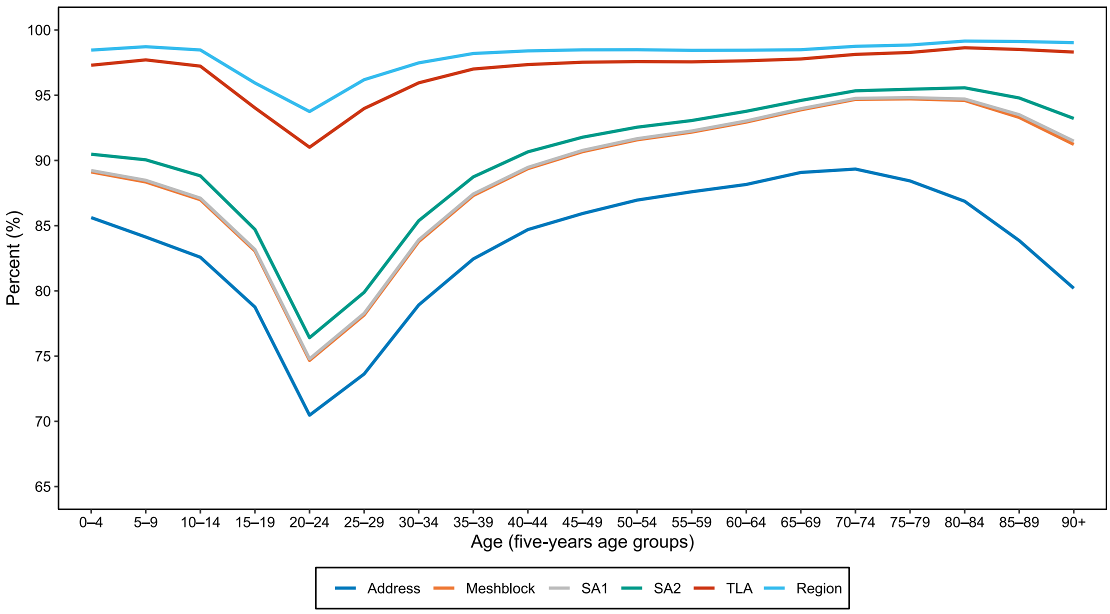
Another source of addressing information in the IDI is the administrative population census (APC). The dataset is part of a wider initiative by Stats NZ to generate census information using administrative data from the IDI. The current iteration covers 2006–2021 and includes 19 census variables (Stats NZ, 2022). In a comparison of usual residence addresses in Census 2018, the APC shows an 83% match with Census 2018 addresses and an 88% match at the meshblock level. The consistency of these addresses improves for broader geographic categories, reaching as high as 98% agreement for regional council areas. It uses a similar approach to the standard IDI process but excludes any data found in Census 2018. The higher levels of agreement in address matching are partly attributed to improvements made in March 2019 to the IDI's address-matching process (Stats NZ, 2022). The agreement of usual residence addresses in the APC (third iteration) and Census 2018 is shown in Table 2.8.

Table 2.8 Consistency of APC geography compared to Census 2018 (adapted from Stats NZ, 2022)

Geographic level	Agreement between APC in 2018 and Census 2018
Address	83%
Meshblock	88%
Statistical area 1	88%
Statistical area 2	89%
Territorial authority	97%
Regional council	98%

Figure 2.8 shows that the consistency of administrative address information is weakest amongst young adults, particularly for the 20–24 age group. This pattern is observed across all geographic scales, although it is less prominent in larger territorial authority and regional council areas (Stats NZ, 2022). For individuals in this age group, the address information from NZTA's data on driver licences and motor vehicle registrations could be a useful alternative data source and is noted as having high accuracy for young adults (Gath & Bycroft, 2018; Stats NZ, 2017). Due to a lack of historical back series, the NZTA datasets have not been used in the Stats NZ APC experimental dataset (Stats NZ, 2017). Most of the low consistency at ages 20–24 is due to intra-regional residential mobility (such as young people changing flats or enrolling in a tertiary institution). The higher the spatial level, the more consistent the coverage. If VKT calculations are aggregated to the regional level, inaccuracy of address data is not a problem.

Figure 2.8 Address consistency between APC in 2018 and Census 2018 by 5-year age groups (adapted from Stats NZ, 2022)



2.2.3.2 Household allocation

The allocation of individuals to households relies on the accuracy of address information. Previous attempts to determine household membership using administrative data (as opposed to relying on the census) had a 68.2% match with Census 2013 (Gath & Bycroft, 2018). While progress has been made in allocating individuals to households, issues with accuracy persist. The third iteration of the APC, which includes household information, has a 72% match with Census 2018. Agreement is higher for single persons and couples (with and without children) and significantly less accurate for single parents and more complex large households (Stats NZ, 2023b).

2.2.3.3 Odometer measurements

When using odometer data to estimate VKT, a range of potential errors can affect the accuracy of the analysis such as incorrect odometer readings at the time of vehicle testing, mechanical issues from odometer rollover (older vehicles with five-digit odometers reset to 0 after passing 99,999 VKT) and other data-entry issues such as duplicated records and excessive readings.

Known errors and potential amendments from the literature and discussions with NZTA and Ministry of Transport are identified Table 2.9.

Table 2.9 Known/potential issues with recorded odometer measurements and possible amendments

Issues	Potential amendments/comments
Zero mileages	Likely from duplicate records. Remove from dataset.
Excessive VKT	Likely misrecording of odometer readings. Remove from dataset. Kirley et al. (2022) and Rendall et al. (2013) remove readings above the 99th percentile to filter for excessive VKT intervals. Cairns et al. (2014) implement a filtering system based on a mileage-per-day threshold of 720 miles per day, akin to continuous driving at 30 mph for 24 hours.
Odometer rollover	Adjust and calculate correct number of kilometres driven. To be assessed before negative distances. More common for older vehicles with five-digit odometers.
Negative distance travelled between two tests	Check for odometer rollover issues. If unresolvable, remove from dataset.
Cheating odometers	Difficult to detect or to estimate significance. Minimum mileage requirements may alleviate some issues.
Change in vehicle use between two periods	Occurs from change in private to commercial vehicle or vice versa. We will attempt to identify such changes by investigating changes in vehicle registration data.
Absence of data for initial years of ownership	Simple solution is to assume average mileage rates for vehicles with the same attributes and household features. Statistical modelling solutions may be useful.
Absence of data for final years of ownership	Simple solution is to assume average mileage rates for vehicles with the same attributes and household features. Statistical modelling solutions may be useful.
Seasonality issues due to recording periods	No off-the-shelf solution currently exists. For annual estimates, this is likely to be an acceptable level of error. Methodologies presented by Cairns et al. (2014) may be useful.
Anomalous odometer readings	For example, odometer readings such as 99,999. Remove from dataset.
Outdated/incorrect addressing	Difficult to detect. To be investigated by comparing registration addressing with Census 2018 dataset in the IDI.

Source: Principal Economics stakeholder meetings with NZTA and Ministry of Transport; Task Force on Road Traffic, 2007; Department for Transport, 2013; Rendall et al., 2013; Cairns et al., 2014; Glensor, 2021; Kirley et al., 2022.

Cairns et al. (2014) investigated odometer readings, concluding that the UK vehicle testing data is relatively clean of errors and opting to omit all erroneous observations. Using a dataset of 150 million observations, they identified 2.77% records affected by errors, including missing first-year observations, odometer readings of 0, anomalous odometer readings (such as 99,999) and negative or unfeasibly high mileage rates.

The use of odometer readings has limitations for VKT estimation. As readings are snapshots in time that may not be specific to periods of interest, some estimation is required to allocate them to a precise time intervals. The issues worsen with longer recorded intervals such as using 2-year odometer readings to estimate annual VKT. Some discrepancy between VKT estimates and actual data is inevitable. The level of adjustment required depends on the significance of the issues encountered and the purpose of the analysis. Typical corrections involve adjusting based on broader traffic trends to account for seasonal and time-related variations (Cairns et al., 2014; Task Force on Road Traffic, 2007).

As shown in Table 2.10, the frequency of WOF inspections in New Zealand varies by vehicle age. This results in fewer odometer readings for newer vehicles. To address this gap, estimates are needed, particularly since newer vehicles generally have higher VKT than older ones (Department for Transport, 2013; Glensor, 2021; Task Force on Road Traffic, 2007; NZ Transport Agency Waka Kotahi, 2025).

Table 2.10 Length of WOF for light motor vehicles

Vehicle first registration anywhere in the world	How long WOF is issued for
New vehicle that’s never been registered	3 years
Less than 2 years ago	To the vehicle’s 3rd anniversary of when first registered
More than 2 years ago but less than 3 years ago	12 months
On or after 1 January 2000	12 months
Before 1 January 2000	6 months

Source: NZ Transport Agency Waka Kotahi, 2025.

An additional challenge with relying on odometer readings to estimate VKT is accounting for the last leg of a vehicle’s life. Prior to a vehicle being scrapped, it may still cover a significant distance but this final distance is not captured by the odometer records (no vehicle is scrapped with a final WOF check). Consequently, VKT estimates based on these readings may be systematically understated. Potential solutions for addressing this issue include assuming the vehicle maintains the same VKT from prior periods or assuming a diminished use proportional to survival rates (Glensor, 2021).

2.2.4 Addressing confidentiality issues

All output of data from the IDI follows strict confidentiality rules. These vary depending on the output type and underlying population or dataset used. Outputs relating to individuals and household information and a household’s contribution to a value magnitude cannot be estimated accurately and must be calculated from at least 20 observations. Unweighted count data must be randomly rounded to base 3. For counts smaller than 6, the data must be suppressed along with any values derived from it. Simulated outputs must all be clearly identified so as not to be mistaken for real data. Suppression rules follow the same rules, with underlying counts smaller than 6 requiring suppression. Depending on the level of output required, simulated and confidentialised counts at SA2 granularity is a feasible option for generating unit record data. In regard to synthetic data such as that generated using the synthpop R package, there are no specific rules published on confidentialising outputs. In our discussions with Stats NZ, any synthetic data method proposal will need to be assessed. The process will also depend on the method adopted to create the dataset. Further details are available in chapter 6.

2.3 Measurement of VKT profiles and correlations

For accurate measurement of VKT, this section provides a review of the available measures and the methodology for estimating VKT, and the impact of other factors such as COVID-19 are investigated. We provide a review of the methods for linking VKT outcomes with other outcomes such as equity, safety and emissions. The findings described in this section were particularly useful for the second phase of this project on the methods and range of explanatory variables considered for estimating VKT.

2.3.1 Analysis of VKT and estimation of the impact of socio-economic factors

Diao and Ferreira (2014) investigated the relationship between VMT and built environment (with demographic controls) using odometer readings of private passenger vehicles registered in the Boston Metropolitan Area. They used spatially detailed built environment data and adopted spatial-error modelling. Their results suggested statistically significant relationships between VMT, demographics and built-environment factors. They find that built environment has a greater impact on VMT per household than demographic factors. Built environment factors such as distance to non-work destinations, inaccessibility to transit and jobs and 'auto-dominance' (capturing high density of vehicle-based infrastructure) were associated with higher VMT per household, while walkability and connectivity were associated with lower VMT. Regarding demographic attributes, they used a range of variables, including home ownership, poverty level, education, income, ethnicity and income rate to account for wealth effects, household count and individuals enrolled at elementary school and population under 5 to capture households with children and population 65 and older to capture working status.

Aslanyan and Jiang (2021) investigated GHG emissions from production-based and consumption-based perspectives. For production-based estimations, they used emissions factors, road segments and speed limits and compared this against consumption-based estimates using VMT odometer-reading data. Their results showed significantly higher VKT estimates for production-based measures, which they attributed to potential errors from external traffic from outside of the Boston Metro area (VMT made by users not included in the odometer dataset).They also analysed factors of VMT such a built environment characteristics with demographic controls. Their results suggested that transport accessibility to destinations and road network density are associated with VMT as strongly as demographic factors and indicated a statistically significant inverse relationship between land-use measures of population, employment and road densities and VMT after controlling for the effects of household size, vehicle ownership and household income.

Zhao and Li (2021) used a bottom-up people-based approach to estimate daily VKT per capita within the Wellington region. They used travel data from the HTS and assigned VKT to the Wellington population based on demographic factors and used ordinary least squares (OLS) regression to assess the factors correlated with travel distance per trip to disaggregate travel elements into personal trips per day, car trips probability, car ownership and vehicle distance per trip. They then extrapolated these elements to the Wellington region by matching regression results to age-group compositions within transport analysis zones. Their results suggest a significant positive association for employment status and mixed correlation for house accessibility and road length per person – the sign is dependent on age. Their results showed negative correlation with built environment determinants, including mixed land use, floor area and road density.

Dieleman et al. (2002) used the Netherlands National Travel Survey to investigate the relationship between residential environment and personal characteristics on mode choice and distance travelled. Their results suggested two sets of factors are of equal importance. Individuals with higher incomes tend to own and use private cars more than those with lower incomes. Households with children are more inclined to use cars compared to one-person households. The travel purpose (work, shopping or leisure) affects both mode of travel and distance travelled.

Narváez-Villa et al. (2021) used ML models, including CART, random forest and gradient boosting, to investigate passenger vehicle travel patterns. Using performance metrics for optimisation, they identified age, engine size and tare weight as factors affecting VKT. VKT drops as a vehicle ages, with a significant decline observed between 4 and 6 years. Engine size is linked to higher usage and better vehicle condition. Higher tare weight correlates with increased annual VKT. There is a reduction in mobility as the driver ages.

Singh et al. (2018) used a simultaneous equations model to determine the impacts of socio-economic and demographic characteristics, built environment attributes and residential self-selection effects on weekday household VKT. Using household travel survey data for the New York region, they linked a multinominal probit model of residential (density) choice to a continuous linear regression associated to household VKT. Their results suggested that socio-economic and demographic characteristics explain 33% of VKT variance in household VKT. After adjusting for self-selection effects, built environment attributes explain 12% of the variance and residential self-selection effects contribute 11% to the VKT variance, leaving 44% of the variance unexplained. The authors suggested that the socio-economic and demographic characteristics have a larger influence on household VMT than either the built environment or residential self-selection.

Chatterton et al. (2015) used estimated VKT (by Cairns et al., 2014) available from national odometer readings in the UK to assess variations in car type, size, usage and emissions and their relationships with socio-economic characteristics. They used the location of vehicle testing stations as a proxy for vehicle location and geospatially linked VKT and vehicle attributes to household characteristics within postcode areas using UK Census 2011 and linked VKT and vehicle attributes to national emissions data. They identified strong positive correlations between the percentage of people driving to work and VKT per vehicle and strong negative correlations with population density and people with no access to a car. The results show moderate positive correlations between average household age, household income and social grade (employed, intermediate managerial, administrative, professional). The study identified no correlation between VKT per vehicle and population of the postcode area. The dataset omits vehicles under 3 years (due to lack of vehicle testing requirements and no odometer readings recorded), suggesting a potential bias due to data limitations. We show the correlation between VKT and socio-demographics by postcode area in Figure 2.9. They additionally tested the correlation between vehicle parameters and emissions per vehicle by matching emissions factors (Barnes & Baily, 2014) with vehicle attributes (fuel type, engine size and date of registration). We show the correlation plots between CO_2 emissions and vehicle attributes in Figure 2.10.

Figure 2.9 Relationship between average km driven per vehicle per year and average household socio-demographic characteristics by UK postcode area (reprinted from Chatterton et al. 2015, p. 11)

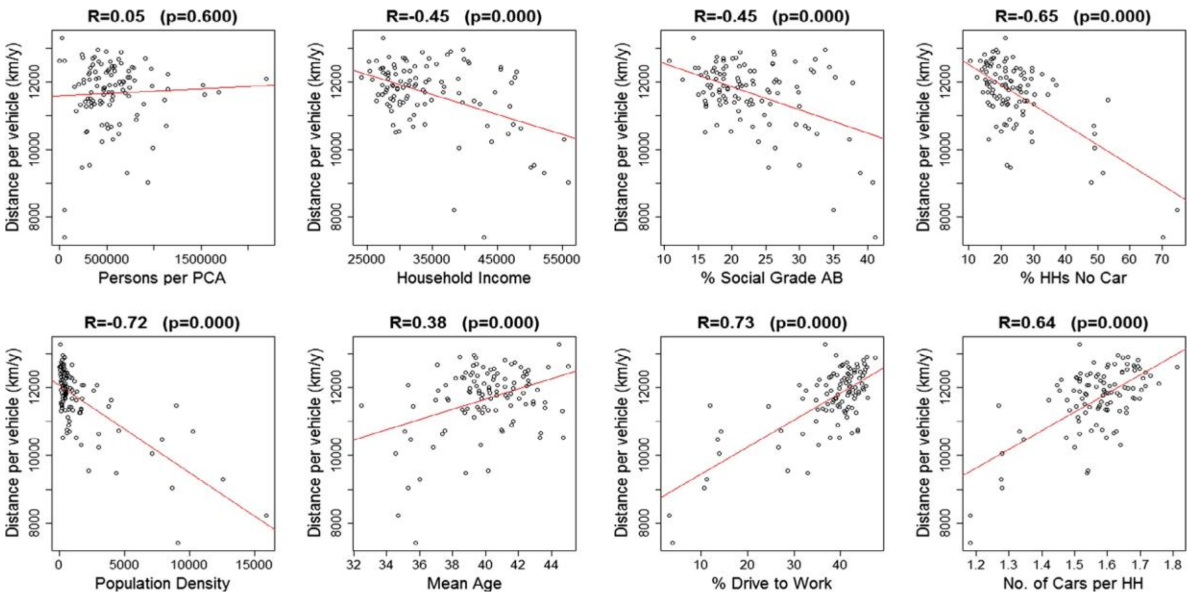
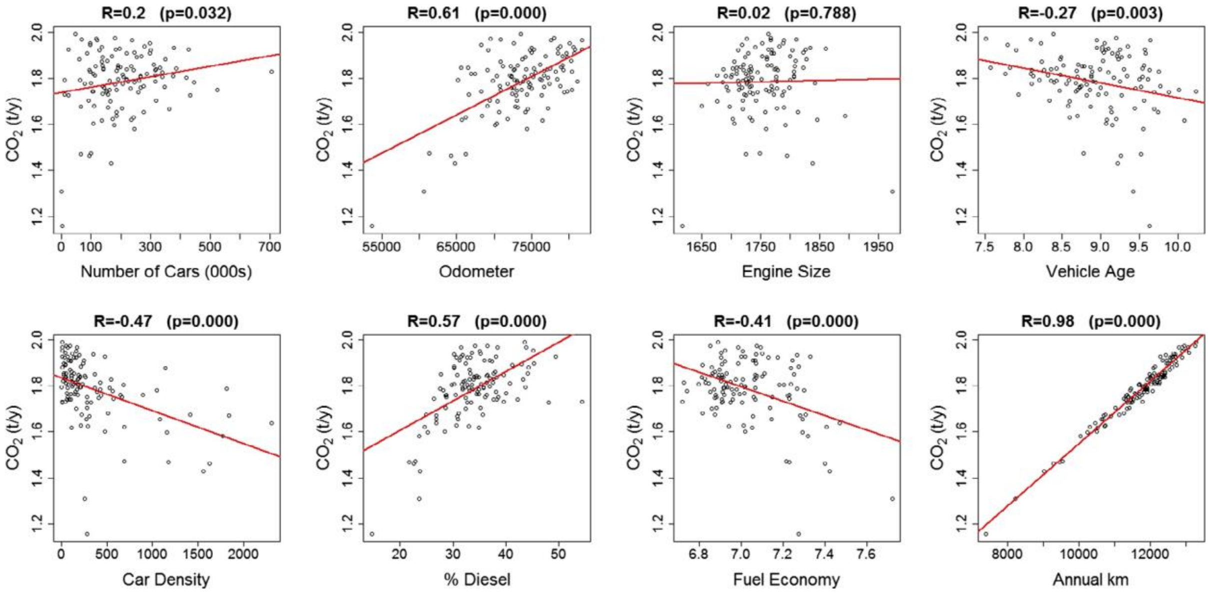
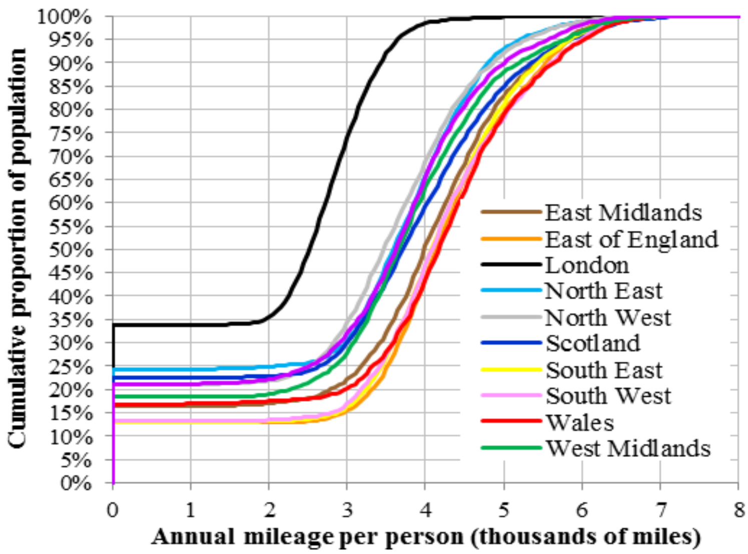


Figure 2.10 Relationship between mean key vehicle parameters and mean CO_2 tonnes per year per vehicle by UK postcode area (reprinted from Chatterton et al. 2015, p. 9)



Ball et al. (2016) build on the work by Cairns et al. (2014) using VKT derived from odometer measurements and link this dataset to the UK census at Lower-layer Super Output Areas (LOSAs)[11] using vehicle registration details. Taking the mean of estimated VKT located within each LOSA, they derive plausible differentiations of mean VKT per person between LOSAs. They note that, when comparing metrics other than the mean such as proportion of vehicles driving more than 12,000 miles per year (the proportion of high VKT households in an area), the variation across areas is significantly higher. The distribution of high VKT households for individual areas is more spread out rather than clustered to a consistent level of high VKT. We show regional distributions of LOSA average household VKT results in Figure 2.11.

Figure 2.11 Annual mileage distribution per person in Great Britain 2011 (reprinted from Ball et al., 2016, p. 8)



We summarise a range of studies on VKT demographics in Table 2.11.

[11] This delineates the UK into 41,729 geographic units, comprising of typically 200–500 households per unit in Scotland and 500–800 households per unit in other areas.

Table 2.11 Methods and factors of VKT demographics

Authors	Study area	Estimation method	Sample size	Dependent variable	Independent variables
Dieleman et al., 2002	Netherlands	OLS	73,702	VKT (work and shopping trips per person)	• Car ownership *** • Income *** • Household type *** • Education *** • Residential environment *** • City size ***
Kockelman, 1997	San Francisco, US	OLS	8,050	VMT	• Household size ** • Car ownership ** • Income per household member ** • General mixed use **
Sun et al., 1998	Portland, US	OLS	300	VMT	• Household size * • Income * • Number of vehicles * • Number of activities * • Land use ** • Accessibility (household to job) * • Accessibility (job to household) *
Diao & Ferreira, 2014	Boston Metro Area, US	Regression – Spatial Error Model	53,188	VMT per household	• Distance to non-work destinations ** • Connectivity ** • Inaccessibility to transit and jobs ** • Auto-dominance ** • Walkability ** • Wealth ** • Children * • Working status **
Aslanyan & Jiang, 2021	Massachusetts, US	OLS FE GS2SLS MLGMM SEM SLM SELM	1,454	VMT	• Per capita income ** • Average household size *** • Population over 65 years old ** • Employment status • Average number of vehicles *** • Average passenger vehicle age *** • Average fuel economy *** • Commute mode *** • Commute travel time *** • Population density *** • Road network density *** • Employment density *** • Accessibility * • Distance to CBD • Weighted VMT *** • Metro area **
Zhao & Li, 2021	Wellington, New Zealand	Joint simultaneous equation	6,913	VKT	• Employment status *** • Road length per person *** • Road length per area *** • House accessibility *** • Mixed-use floor area density ***
Stroombergen et al., 2018	New Zealand	Simulation using the RLTDM	N/A	VKT	• Population growth dynamics • Spatial demographic trends • Technology trends • Income and economic growth • Industrial composition • Prices

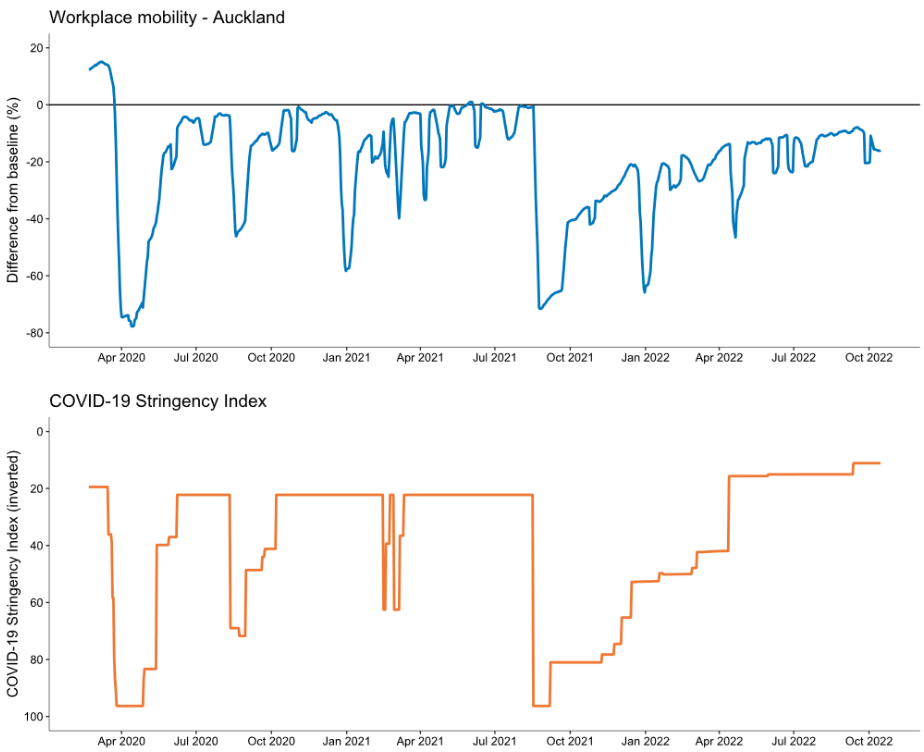
Source: Principal Economics based on the studies listed.

Our review suggests that most explanatory variables are statistically significant, but the causal impact is indeterminate – VKT can determine household variables and petrol prices. This is unsurprising given that each travel purpose is driven by a range of factors, including the socio-economic features of households, the economic conditions, the price of fuel and the various components of the generalised cost function of LPVs. All these factors determine the need for travel, and the generalised cost elasticities indicate the change in VKT (and use of other modes). For our report, it is important to note that most of the considered explanatory variables in the literature are highly correlated with socio-economic features such as age, income and household composition as well as the features of the vehicles and the overall economic conditions.

2.3.2 Methods available for analysing the impact of other factors

Assessing VKT between 2020 and 2023 poses significant challenges. The impact of the COVID-19 pandemic disrupted usual travel patterns with impacts from government lockdowns, shifts from public to private transport and remote working. In assessing the impact of half-price PT, Principal Economics (2022) isolate the fluctuations from the impacts of COVID-19 and those driven by half-price PT by constructing a COVID-19 exposure index. The index is extrapolated using the results of Das et al. (2021), which provides information on the likelihood of modal shift from PT to private vehicles as a result of COVID-19. Indices such as the Government Stringency Index (Mathieu et al., 2020) offer potential solutions for controlling for various impacts of COVID-19 on travel behaviour. Figure 2.12 shows workplace mobility in Auckland as a difference from pre-COVID-19 baseline and stringency of government policies in response to COVID-19 in New Zealand. While it shows some recovery to pre-COVID-19 levels, the impact of the pandemic on potentially longer-term trends such as working from home is not yet fully understood (Currie et al., 2021).

Figure 2.12 Workplace mobility and COVID-19 policy responses[12]



Source: Google Mobility Report, https://ourworldindata.org/coronavirus.

[12] Baseline covers 3 January – 6 February 2020. For illustrative purposes, we show weekly average workplace mobility.

2.3.3 Methods available for linking VKT outcomes with other outcomes

2.3.3.1 Equity

In New Zealand, a typical measure used for capturing the distributional outcomes of transport investments is the New Zealand Deprivation Index (NZDep), which is an area-based measure of socio-economic deprivation (Atkinson et al., 2019). NZDep has been used in a variety of studies concerned with distributional effects across New Zealand communities (Curl et al., 2020). Studies use both the index and score measures of deprivation. The distribution and use of the score and index measures have important differences. Torshizian (2017) provides further details on the advantages and disadvantages of using each measure. The New Zealand Index of Multiple Deprivation (NZIMD) is another area-based deprivation measure. It differs from the NZDep in that it consists of seven domains of deprivation as opposed to a single index and score. The NZIMD includes indices related to the domains of employment, income, crime, housing, health, education and access (Exeter et al., 2017).

Torshizian et al. (2022) reviewed the current Monetised Benefits and Costs Manual approach to distributional analysis, providing an overview of considerations for transport equity as well as a method for accessing transport equity impacts geospatially and their effects on different population groups. Their approach follows the NZTA's Monetised Benefits and Costs Manual mobility-focused approach.

2.3.3.2 Safety

In New Zealand, three main safety rating systems are in place: Australasian New Car Assessment Program (ANCAP), Used Car Safety Ratings (UCSR) and the newly added Vehicle Safety Risk Ratings (VSRR). ANCAP focuses on new vehicles and assesses them based on various crash tests, adjusting their criteria as safety technology evolves. UCSR ratings are derived from real-world crash data but does not consider crash avoidance features. VSRR is used for vehicles without ANCAP or UCSR ratings, providing an estimate based on similar vehicles. The mix of safety ratings is due to the composition of the New Zealand vehicle fleet being comprised of new and imported vehicles in the country. These ratings have been collated together into a single database and linked to vehicles by plate number by NZTA on the Rightcar website.[13]

2.3.3.3 Emissions

As discussed in section 2.1.6.1, the VEPM is used to predict emissions for New Zealand's vehicle fleet under common road and traffic conditions (Metcalfe & Peeters, 2023). The model includes emissions factors for the New Zealand fleet based on the different vehicle types and vehicle class. These factors can be used to provide estimates of vehicle emissions by multiplying the emissions factors, which are denoted in g/km.

The VEPM primarily focuses on predicting emissions for the entire vehicle fleet, resulting in high-level emissions factors. Specifically, when considering the private light vehicle fleet, the vehicles are categorised solely based on their motive power, without further disaggregation.

Nevertheless, as fleet emissions factors are determined through a bottom-up approach, the model has incorporated more detailed disaggregation of emissions factors for different vehicle types within its internal structure. This involves making assumptions about emissions standards based on factors such as vehicle category, fuel type, year of manufacture and country of origin. Emissions factors are then calculated for each unique combination of vehicle type/emissions standard. Fleet ratios and VKT are determined using motor vehicle registration data, with these ratios informing the calculation of the aggregated emissions factors. This provides the high-level emissions factors in the VEPM.

[13] https://rightcar.govt.nz/

It is therefore possible to use the raw emissions factors for the disaggregated vehicle types to estimate emissions for individual vehicles by assuming the same linkages between emissions standards and vehicle attributes made in the VEPM. This can be undertaken using the MVR dataset present in the IDI, which holds the same information on vehicle category, fuel type, year of manufacture and country of origin that are used to determine fleet-weighted VEPM emissions factors.

2.3.4 Approaches for linking VKT profiles to projection from simulation models

Project Monty and the RLTDM are the simulation models providing projections of VKT profiles. As usual, the simulation models are established based on stylised relationships among economic, demographic and transport behaviour factors. Both models use the household travel survey as a base input to VKT for different demographic groups and use external VKT measurements such as traffic count data for regions for allocation/calibration of VKT within areas.

The significant difference between this report and Project Monty (and the RLTDM) is that those models are calibrated by the HTS whereas, in this project, the admin data of registered owners of light motor vehicles are the spine of the data, linked in the IDI with their household characteristics and WOF odometer readings of their vehicles.

3 Available data and methods for VKT demographics

This chapter provides information about the available data. It includes a description of the identified issues with different data sources and our approach to addressing them. We also provide our approach to potential methods for addressing measurement errors, which will be further tested and implemented in the next chapter.

Based on our literature review, most of the explanatory variables of VKT considered in the literature are highly correlated with socio-economic features such as age, income and household composition as well as the features of the vehicles and the overall economic conditions. Hence, in this chapter, we are primarily focused on the feasibility of disaggregating VKT by those features.

The data available from the IDI includes odometer readings, vehicle attributes, individual demographics, geographic location and household features. In addition to the inherent data manipulation required for the IDI data, we addressed various critical issues:

- To address issues of irregular odometer observations for vehicles, we have adopted methods from Wilson et al. (2015). We assume a snapshot of daily VKT as of 30 June for each year as constant over the year to evaluate the variance of VKT between demographic groups, geographic regions and income quintiles. We expanded this analysis to estimate temporal values for a greater range of demographic groups, which are used for econometric analysis in chapter 5.
- On the construction of household tables, the most reliable source of household information is Census 2018. Additionally, an experimental linkage between individuals and household units was available to researchers during the writing of this report as part of the October 2023 IDI refresh. This dataset is currently the most comprehensive information available for household analysis over inter-census periods, albeit there are known issues with accuracy, particularly for larger, more complex household compositions such as multi-family households and households of unrelated people (Stats NZ, 2023a). Given these sources of information on household membership, we use Census 2018 for the linkages of individuals to households, which is consistent with earlier studies such as Principal Economics (2022). Furthermore, to improve statistical robustness, a timeseries dataset of the MVR is loaded to the IDI to provide a comprehensive estimate of odometer readings captured during Census 2018. This methodology could be further updated when Census 2023 is added to the IDI.
- To link registered vehicle ownership periods to individuals, we have relied on ownership data available in the MVR. Our initial strategy involved using Inland Revenue (IR) identifiers and historical IDI individual identifier records to link individuals across MVR snapshots. Since the release of the October IDI refreshes, we have updated our methodology to use the official IDI identifier linkages over time. We have used date information available in the MVR as indicators for periods of vehicle ownership. As individuals may have multiple vehicles, we have derived provisional VKT estimates for both vehicles and individuals over time. Overall, our investigation of data suggests that the coverage of both demographic and granular geographic units is sufficient for robust analysis.
- To manage the size and complexity of data, we have used a variety of techniques to improve the efficiency of our analysis. This includes leveraging the use of both SQL and R dependent on the speed of compute and memory management. Additionally, we have developed fast algorithms for detecting erroneous odometer readings. Given the processing time required for estimating annual VKT using all odometer readings, we have opted for taking a single snapshot in time for this analysis phase.

For joining safety standards to vehicles in the second phase of the analysis, we will use the granular vehicle information available in the MVR. This will be accomplished using ad hoc data loading depending on the level of granularity of safety standards provided (at individual vehicle level).

3.1 Criteria for choice of data

Initially, we suggest that the data included in the analysis needs to be available to researchers (in the IDI), to be useful to the end user (to inform policy and model as discussed above) and to provide frequent updates. While census data provides more accurate addressing information, the low frequency of the census (available every 5 years) limits the usefulness of census data to this report. Hence, we considered other sources of data, including the APC, address information in the IDI and the address data created by Fabling and Maré (2020). We note that, as the IDI sources data from a range of data providers, there is a high variability in the timeliness of data in the IDI. While data providers provide raw data to Stats NZ on a regular basis, this is according to their own schedule (Welsh, 2023).

3.2 Datasets

3.2.1 Odometer dataset

Due to the absence of temporal odometer data in the main IDI MVR dataset (which provides a snapshot of the current fleet profile), we use a method of stacking IDI refreshes (or snapshots) and calculating VKT by analysing the difference in odometer readings between refreshes. We list the IDI refreshes we have extracted odometer readings from in **Error! Reference source not found.**.

Table 3.1 IDI refreshes[14]

IDI refreshes	Database name
April 2018	IDI_Clean_20180420
October 2020[15]	IDI_Clean_2020020
March 2021	IDI_Clean_2020320
June 2021	IDI_Clean_2020620
October 2021	IDI_Clean_2021020
March 2022	IDI_Clean_202203
June 2022	IDI_Clean_202206
October 2022	IDI_Clean_202210
March 2023	IDI_Clean_202303
June 2023	IDI_Clean_202306
October 2023	IDI_Clean_202310
March 2024	IDI_Clean_202403
June 2024	IDI_Clean_202406

Source: Stats NZ.

[14] As user identifiers change with each IDI refresh, we had been using IR-to-NZTA identifiers to link users across IDI refreshes, while also accounting for revised NZTA identifiers refreshes corrected by Stats NZ. In the October 2023 refresh, Stats NZ introduced a new IDI dataset that provides linkages for user identifiers across refreshes. We have since updated our methodology to use this dataset.

[15] This database has since been removed from the IDI as part of regular maintenance. We have extracted vehicle attributes, odometer readings and the corresponding individual identifiers from this refresh and stored these in the IDI project folder and project database.

By 'stacking' IDI refreshes, we can assess the variations in odometer records across each refresh, allowing us to gather VKT data for each vehicle and its registered owner over time. We use a total of eight snapshots for the IDI MVR dataset, covering the period March 2022 to June 2024.

3.2.2 Demographic datasets

While the census data has low frequency, it provides the most comprehensive and robust source of demographic data available at the point in time when the census is conducted. At present, it is the most reliable source for determining the households to which individuals belong. The demographic variables that are largely restricted to using census data include number of children, household composition and highest qualification in households. As of writing this report, Census 2023 has not yet been included in the IDI. The census is often used as benchmark for testing the validity of other administrative data and methods within the IDI (Anastasiadis et al., 2020; Fabling & Maré, 2020; Gath & Bycroft, 2018; Stats NZ, 2022). A notable limitation of the census is its lack of granularity on its information on income, which is limited to the classification bins outlined in census questions.

The IDI provides core data using administrative and census data to provide the best estimate of socio-demographic information. In terms of the age of individuals, the data is derived from the Department of Internal Affairs birth register, Ministry of Business, Innovation and Employment border movements and visas, and census and IR data, providing almost complete coverage of information on age, sex and ethnicity for all individuals in the IDI. Personal income is also available as part of the IDI core data derived from data from IR, Ministry of Social Development (MSD) and Working for Families. A known limitation of the use of administrative data for deriving income is the lack of information on income from other superannuation, pensions or annuities (Stats NZ, 2022).[16] These datasets are updated with each IDI refresh and allow for the creation of timeseries datasets on individual socio-demographic variables on an annual basis.

A recent addition to the IDI is the development of the APC, which includes a range of socio-economic and demographic variables. In the most recent iteration (October 2023), household variables have also been produced, albeit with significant limitations – with a 75% exact household membership match rate with Census 2018 (Stats NZ, 2023a).

The APC differs from census data in that it is derived solely from administrative data. The APC provides valuable information for this project on many socio-economic variables that have been derived from a range of IDI datasets over time. This includes variables such as personal income, highest qualification attainment, employment status, industry of employment and Māori descent. Being a complete dataset of vehicle registrations, matching the MVR to the APC enables determining vehicle ownership for the population.

3.2.3 Addressing datasets

Between New Zealand census periods, which happen every 5 years, it is difficult to accurately connect vehicles to individuals or households using address data. In addition to methodologies for deriving alternative addressing datasets created by external parties, multiple datasets of collated addressing data derived using administrative data exist in the IDI (Anastasiadis et al., 2020; Fabling & Maré, 2020; Stats NZ, 2023b).

We identify three addressing datasets that are likely to have the best accuracy and coverage of the MVR:

- The IDI core data is the standard dataset provided to all IDI researchers, which pools address data from at least nine government organisations, each with its own level of accuracy and update frequency.

[16] We note that Census 2018 had an 81.2% response rate for total personal income (Welsh, 2023) – lower than the coverage we determined using core data when comparing against individuals in our estimated VKT dataset.

- The APC adopts a largely similar methodology with a slightly different ruleset and exclusively uses the administration dataset, excluding census data.
- The methodology proposed by Fabling and Maré (2020) includes pooled MVR addressing data from prior IDI snapshots. We note that they derive addressing for employees only.

3.3 IDI and feasibility of deriving the outputs required

As part of the feasibility study of this report, we assessed each dataset for missing values and erroneous observations. We first undertake the preliminary data matching and then further refine and cleanse the data.

3.3.1 Motor Vehicle Register

We estimate annual VKT for vehicles using the methodology used by the Department for Transport (2013) for creating its experimental statistics for individual vehicles (see section 2.1.1.1). As the MVR dataset in the IDI only covers vehicles currently on the road, we use all available snapshots or IDI refreshes and collate current and historical datasets together to create a timeseries of recorded MVR records.

While the primary objective of this phase of the study is to determine the feasibility of determining the relationship between VKT, geography, demographics and vehicle characteristics, we have created an initial dataset of VKT estimates using odometer readings. This has been undertaken to test the matching rate of relevant observations against available demographic variables.

In investigating the MVR datasets, we find duplicate observations across different IDI refreshes such as vehicles with multiple odometer readings on the same day. This is due to updates and amendments prior to odometer recordings. Where these duplicates exist, we keep only observations found in the most recent IDI refreshes, assuming that these represent the most accurate readings available.[17]

Additionally, we use only odometer readings from vehicle inspections with a passing grade, which further mitigates issues with multiple readings on the same date.[18]

We also find that the allocation of vehicles to individual identifiers in MVR datasets in the IDI varies over each IDI refresh. To account for these variations, we defer vehicle to individual linkages post-cleaning of the VKT dataset. We adopt a similar approach for linkages with vehicle attributes. In our initial dataset, we include only vehicle ID, industry class, inspection date and odometer reading.

We list the methods we used for data cleaning in Table 3.2.

Table 3.2 Missing data/errors

Issues	Potential amendments/comments
Zero odometer readings/VKT	Removed from dataset.
Excessive VKT	We use a generous limit of 438,000 km/yr to define excessive VKT. This represents a vehicle that is driven at 50 km/h, 24 hours a day, every day for a year. This aligns with the methodology used by Cairns et al. (2014).
Odometer rollover	See below.

[17] An alternative method for determining the 'best' reading is provided by Wilson et al. (2013).

[18] Vehicles that fail inspection should be kept off the road between tests and are likely to have little to no VKT, which may affect the analysis. This approach aligns with that employed by the Department for Transport (2013).

Issues	Potential amendments/comments
Negative distance travelled between two tests	We develop a fast algorithm for determining the odometer readings based on sequential observations. When a negative VKT is derived, the algorithm tests the relative position of the related odometer reading and the sequence of odometer readings for the vehicle and compares against odometer readings one to two periods ahead and prior depending on its position in the set of readings available for that vehicle. This is used to identify the position of the erroneous odometer reading and flag it for removal. VKT is then recalculated for the new set of odometer readings and compared against prior estimates. Subsequently, this two-stage analysis allows for the identification of odometer rollover or odometer replacement where a negative VKT is determine post-removal. In these cases, we drop the estimated VKT for that odometer span but keep all other estimated VKT spans following odometer rollover/replacement.
Cheating odometers	Depending on how odometers are cheated, a proportion will be captured and removed from our estimates when checking for excessive VKT estimates and negative distances between motor vehicle tests. In more-complex cases where cheating is undertaken periodically while the vehicle is in use, we have deemed these too difficult to detect.
Change in vehicle use between two periods[19]	Given this report's focus on private vehicle use, we remove all odometer observations where a change in vehicle use is recorded. Additionally, we remove all VKT recordings prior to and following this change as this represents periods where a vehicle may have been used for non-private purposes.
Absence of data for initial years of ownership	No action has been undertaken to amend this information at this phase. We separate these values out from our analysis and identify the proportion of missing VKT data across different demographic groups and vehicle ages.
Absence of data for final years of ownership	No action has been undertaken to amend this information at this phase. We separate these values out from our analysis and identify the proportion of missing VKT data across different demographic groups and vehicle ages.
Seasonality issues due to recording periods	We ignore issues of seasonality, assuming that snapshots allow for an acceptable level of error. As with Cairns et al. (2014), further refinement could be made by taking a greater number of snapshots (quarterly or monthly) to estimate annual VKT.
Anomalous odometer readings	To account for VKT relying on an accurate sequencing of observations (deriving VKT relies on having accurate pairs of observations), we employ an iterative process to flag anomalous observations. We then test different combinations, removing either the flagged observation, the preceding one or the following one and removing the most likely erroneous observation.

Source: Principal Economics analysis.

As we are using snapshots of the MVR as opposed to a timeseries dataset, there are a number of missing odometer recordings where a corresponding odometer reading for the same vehicle is not recorded. This is a likely consistency issue since our snapshot range consists of a single snapshot in 2018, then there is gap of 1 year and the next snapshot is available for the period 2020–2023.

We show the count of odometer readings in our pooled dataset at their year of inspection in Table 3.3. As additional IDI refreshes are added, the availability of observations will increase, assuming the dataset is updated regularly.

Alternatively, loading a timeseries dataset of the MVR to the IDI would provide a complete range of odometer observations for all years. In our regression analysis (in chapter 6), we test the change in coefficients by considering different years.

[19] We highlight the importance of correcting for erroneous recordings prior to the removal of negative VKT intervals in Appendix A.

Table 3.3 Number of odometer readings by year of inspection

Year	Number of records
Prior to 2016	597,291
2016	381,120
2017	3,060,363
2018	624,972
2019	1,674,174
2020	3,794,370
2021	3,875,478
2022	3,886,164
2023	2,126,439
No date	152,880

3.3.2 Recording errors and odometer data availability

Table 3.4 shows the count of observations collected, the number of records/intervals affected by different errors and omissions and their proportion of total records/intervals.[20] We defined records as individual odometer testing records and intervals as the inter-period spans between testing records that VKT is estimated from. At this stage of the data cleaning, we filter the pooled MVR dataset for observations that should be omitted from this analysis. This includes common data issues such as missing odometer reading information and vehicles out of scope, including non-private vehicles and/or commercial vehicles.

Table 3.4 Accuracy and usefulness of MVR data

Reason for omission	Records affected	% of records
Missing odometer reading	160,848	0.8%
Missing inspection data	152,880	0.8%
Zero odometer reading[21]	2,199,123	10.9%
Failed inspection	737,511	3.7%
Non-private vehicle[22]	1,477,734	7.3%
Non-LPV classification[23]	5,182,350	25.7%
Total omissions	5,970,555	29.6%
Total observations	20,173,188	100.0%

Source: Principal Economics analysis.
Note: As observations can have multiple reasons for omission, the percentages across all reasons do not sum to the total.

[20] This excludes duplicate observations from earlier IDI refreshes that were the most recent recording, assuming that, where variation occurs, these are corrections that have been made since the initial recording.

[21] The majority of zero odometer readings reflect trailers and caravans that require a WOF but do not have related recorded odometer readings. This overlaps with non-LPV classifications.

[22] For vehicles that change between private and commercial uses, we omit both the observation where it is recorded as a non-private vehicle and the observation following to remove the period for which it operated as a commercial vehicle while keeping observations where it operated as a private vehicle.

[23] We use the Ministry of Transport definition of light passenger vehicles as a passenger car or van up to 3,500 kg.

Out of the extensive pooled MVR dataset comprising over 20 million observations, we omit just under 30% of all MVR observations available. The primary reason for omitting a large proportion of observations is that they do not relate to LPVs (non-LPV classification). Failed inspections are also omitted, assuming that vehicles are retested later.

After omitting these observations, we determine the number of single observations. These consist of single odometer readings where VKT cannot be estimated due to a lack of data. This largely consists of newer vehicles that have only had their first WOF check. We show the count of single odometer observation records and the remaining observations available for VKT estimation in Table 3.5.

Table 3.5 Number of odometer readings available for VKT estimation

Type of issue	Number of records
Total remaining observations	14,202,633
Single observations	840,408
Observations available for VKT estimation	13,362,225

Source: Principal Economics analysis.

As discussed in section 2.2.3.3, the requirements for WOF inspection frequency varies by vehicle age, with newer vehicles requiring fewer inspections than their older equivalents. This means that, for newer vehicles, a second odometer observation is not available to determine VKT. We noted the methods for approximating VKT in section 2.2.3.3. These methods will be used in the next chapter for addressing the missing value issue. We provide more details on the distribution of single observations in Appendix A.

Given the potential for bias from these single odometer observations, we further disaggregated the distribution of single odometer observations by inspection year, age and vehicle age. It is important to note that the distribution is more important than the value of VKT (at level) for the purpose of this phase of the project. This is because we could possibly calibrate the VKT values, but unbiased distribution is critical for VKT profiles. If we identified issues with the distribution, we would use the available identification methods to address the problem in the second phase of this project (discussed in the next chapter).

3.3.3 VKT estimation and odometer interval errors

Estimating VKT relies on the accuracy of two sequential odometer readings. This presents some challenges in its determination. Wilson et al. (2015) highlight the issues with erroneous observations when used with spanning intervals. In situations where a 'bad' odometer reading exists, the calculation of VKT from 'good' odometer readings on either side of the bad reading is impacted such that an incorrect mileage rate is determined. Removal of the erroneous observation is likely to provide the correct VKT over the assessment period. We develop a fast algorithm to identify and remove erroneous odometer readings when a negative VKT reading is identified (see Appendix B).

For the available observations of over 13 million, each must be paired to a second observation to determine VKT for that vehicle over two points in time. There were a number of errors in the data during this process, including the presence of negative VKT estimates and duplicated dates. We follow the data cleaning process outlined in Table 3.2 and additionally remove observation spans over 4 years (Table 3.6). This is to keep vehicles that are required to be tested at least once over a 3-year period while omitting vehicles that may have been off the road for an extended period.

Table 3.6 Number of odometer pairs available for VKT estimation

Type of issue	Number of records
Single observation	1,104
Negative VKT	321,456
Duplicated dates	15
Excessive VKT	2,694
4+ years interval	192,921
First odometer reading	3,312,288
Total VKT intervals	13,362,225
Remaining VKT intervals	10,631,196

Source: Principal Economics analysis.
Note: There are various reasons for the omission of each VKT interval so subtracting the sum of all issues from total VKT intervals will not result in remaining VKT intervals.

With over 10 million odometer-pairs available, we use methods outlined in section 2.1.1.1 to calculate VKT, taking a snapshot for each year at 30 June 30 and determining the average daily VKT for each vehicle (and registered owner) and multiplying this value by number of days in a year.[24] As a proof of concept, we choose a single date for each year ranging from 2018–2023. This could be further extrapolated to a daily granularity to provide an accurate estimation of annual VKT by estimating the VKT for each day and taking the sum of estimations. As the focus of the phase is to determine the potential for linking VKT to demographics, we have opted for a single point in time for each year.[25]

We calculated VKT for each period using annual VKT as outlined in **Error! Reference source not found.**:

$$VKT_{n+1} = (R_{n+1} - R_n) \times \frac{365.25}{\Delta T} \qquad \text{(Equation 3.1)}$$

where:

VKT_n = estimated annual VKT associated with test $n+1$

R_n = first odometer reading for the vehicle n in the category *i*

R_{n+1} = second odometer reading for the vehicle n in the category *i*

ΔT = number of days between the first and the second odometer readings (days).

To estimate annual VKT for individuals, we take the sum of estimated daily VKT for each vehicle registered to the individual and multiply this by days in a year using (Equation 3.2) to account for individuals who own and drive multiple cars:[26]

$$VKT_{in} = \sum \frac{(R_n - R_{n-1})}{\Delta T} \times 365 \qquad \text{(Equation 3.2)}$$

[24] For consistency with the APC dataset found in the IDI, we use 30 June for each year (McKibbin et al., 2022).

[25] The Ministry of Transport has its methodologies for determining annual VKT. We use the dataset we have created solely as a proof of concept.

[26] This makes a potentially unlikely assumption that owners of vehicles spread their usage evenly across their year. However, this is prone to overcounting VKT in situations where vehicles are only used and owned for a proportion of the year. To mitigate this, we suggest exploring the possibility of increasing VKT snapshots to daily granularity in the second phase of this project.

where:

VKT_{in} = estimated annual VKT associated with individual i over the period n

R_{n-1} = prior odometer reading for the vehicle at period n-1

R_n = odometer reading for the vehicle at time n

ΔT = number of days between the first and the second odometer readings (days).

3.3.4 Matching VKT intervals with demographic features

For the remaining VKT intervals, we join demographic data from the IDI core dataset and APC data and test their match rate – the proportion of demographic records that can be matched with estimated VKT (Table 3.7).

Table 3.7 IDI core data and APC match rates[27]

IDI core data	2016	2017	2018	2019	2020	2021	2022
Age	99.7%	99.5%	99.8%	100.0%	100.0%	100.0%	100.0%
Sex	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Ethnicity	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Personal income	88.5%	88.4%	90.1%	92.5%	93.2%	93.1%	91.7%
APC[28]	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Personal income	94.5%	93.3%	94.4%	95.7%	96.6%	96.4%	94.5%
Employment indicator	95.8%	94.6%	95.4%	96.5%	97.3%	97.0%	95.4%
Employment status	71.5%	70.7%	70.2%	71.1%	71.5%	71.7%	72.0%
Highest qualification	88.5%	86.4%	86.7%	87.3%	87.0%	87.0%	85.6%

Source: Principal Economics analysis.

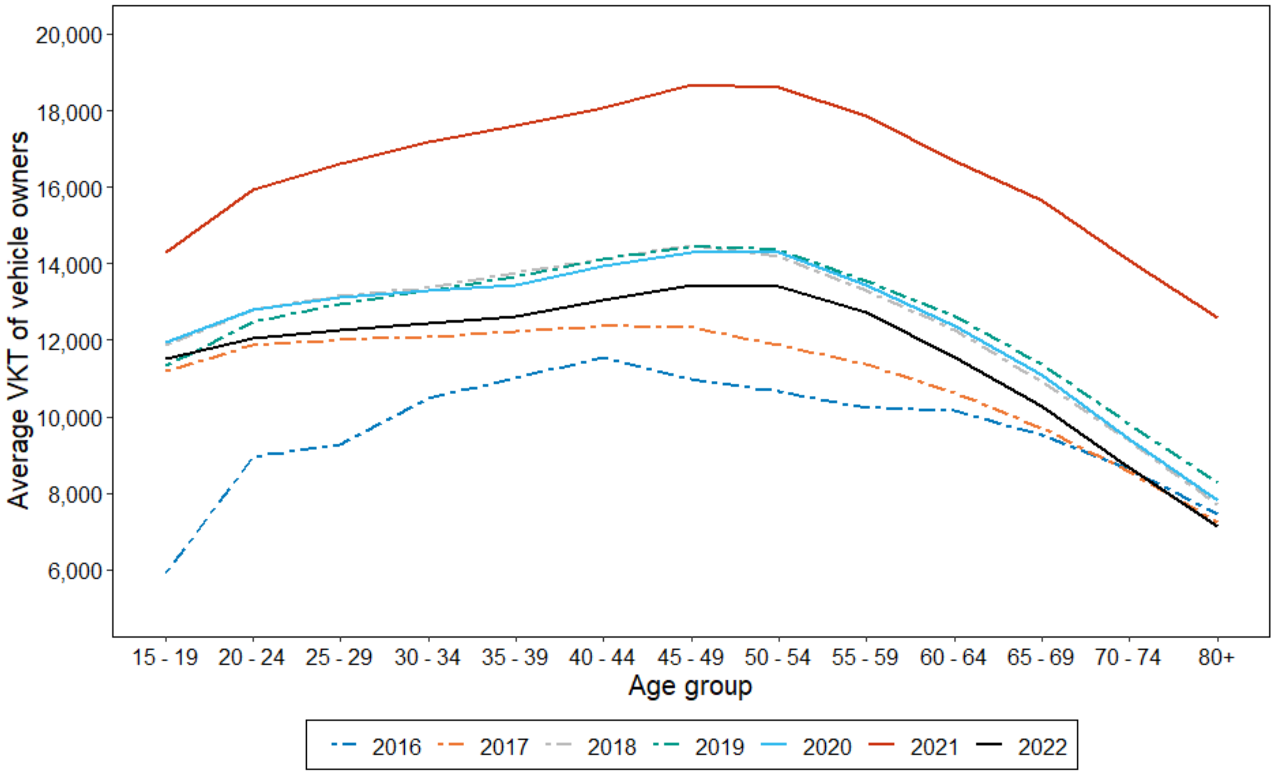
Our results show a strong match rate when using IDI core data with near perfect coverage for age groups (less than 0.1% unmatched for the 2019–2023 period). The match rates improved for income quintiles when using the APC as opposed to IDI core data. This is due to slightly different methods for determining personal incomes in the two datasets, where the APC uses a wider range of data sources including income from benefits and New Zealand Superannuation, investment income, child support payments and income tax credits (Stats NZ, 2023b).

We show preliminary outputs for average VKT by age groups in Figure 3.1 and Table 3.8. Note this does not include any adjustments for single odometer observations or vehicle scrappages. The average VKT values will marginally change as our methodology is further developed in the next chapter. We will further explore the higher VKT observed in 2021 below.

[27] Currently, the APC does not cover 2023.

[28] The APC also includes age and sex observations. As these have lower match rates than the IDI core data, we have not included them in this table.

Figure 3.1 Average VKT of vehicle owners by age group



Source: Principal Economics analysis.

Table 3.8 Average VKT of vehicle owners by age group[29]

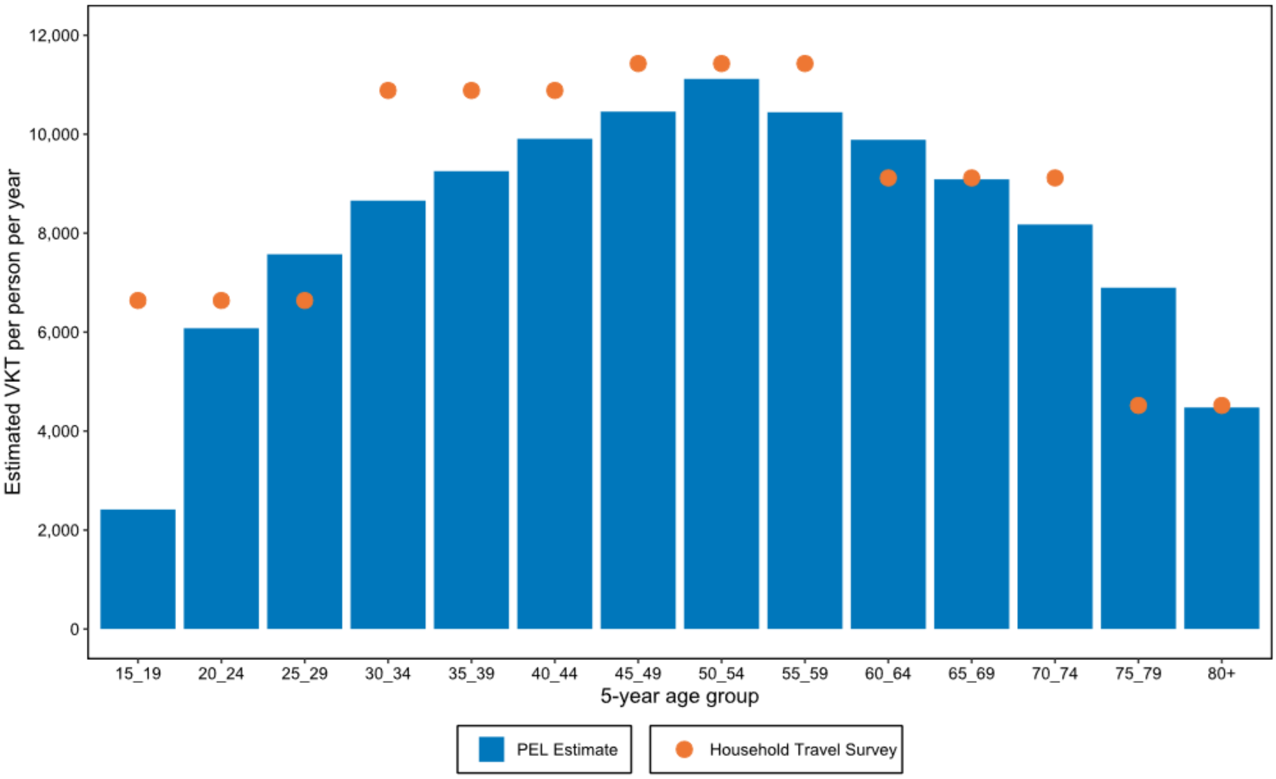
Age group	2016	2017	2018	2019	2020	2021	2022
15–19	5,921	11,205	11,883	11,337	11,944	14,307	11,501
20–24	8,941	11,873	12,804	12,480	12,807	15,942	12,036
25–29	9,260	12,029	13,158	12,939	13,107	16,599	12,275
30–34	10,474	12,073	13,383	13,287	13,290	17,183	12,443
35–39	11,010	12,219	13,766	13,643	13,450	17,620	12,605
40–44	11,565	12,388	14,134	14,126	13,957	18,063	13,038
45–49	10,986	12,337	14,463	14,450	14,284	18,694	13,439
50–54	10,667	11,869	14,187	14,358	14,283	18,618	13,422
55–59	10,220	11,375	13,314	13,565	13,444	17,862	12,712
60–64	10,181	10,616	12,271	12,606	12,366	16,687	11,558
65–69	9,537	9,696	10,913	11,367	11,075	15,657	10,282
70–79	8,638	8,544	9,374	9,816	9,404	14,068	8,677
80+	7,452	7,245	7,705	8,263	7,797	12,590	7,147

Source: Principal Economics analysis.

29 For analysis, we have grouped vehicle owners by age. The data available includes all ages as nominal values.

We compare our average VKT estimates against those from the HTS 2019–2022. As our values encompass vehicle owners only, we use our 2021 estimate of total VKT, which we have determined as the year for which we have the most comprehensive data (the least number of single odometer observations) and divide by the estimated population count in 2021.[30] As shown in Figure 3.2, the estimated VKT by age group is highly aligned with the pattern of the HTS. Additionally, we note the age brackets of the HTS are more aggregated and differ slightly from the age groups we have shown in Figure 3.2 to allow for high-level comparison.[31]

Figure 3.2 Estimated VKT by age group compared to the HTS



Source: Principal Economics analysis, Ministry of Transport.
Note: Numbers are adjusted to population as opposed to VKT for vehicle owners only. This is to match the public values reported for the HTS by the Ministry of Transport.

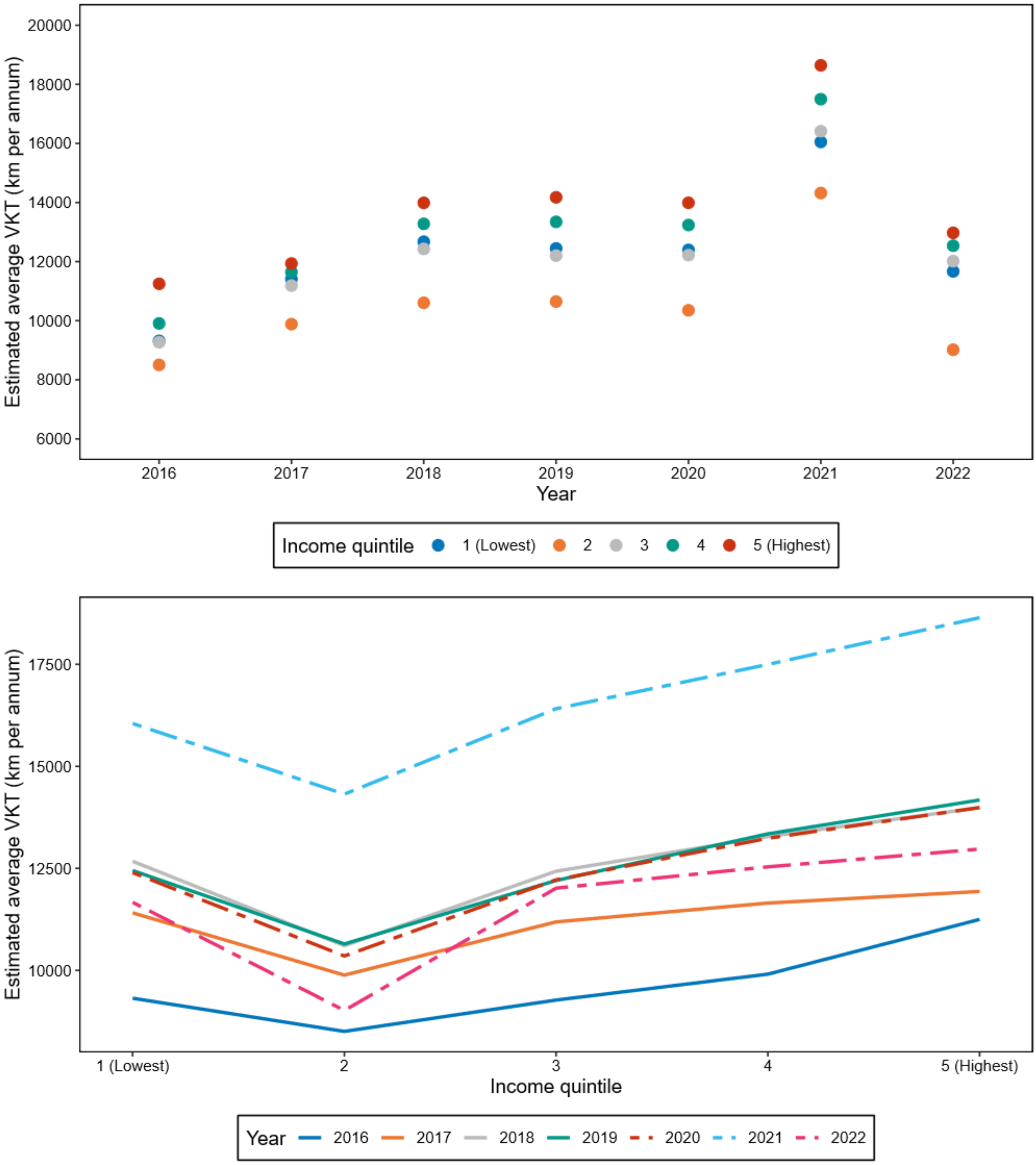
Figure 3.3 shows the average VKT of vehicle owners across different income quintiles. As expected, the overall pattern of VKT over time is similar across income quintiles. The pattern over time is less important for the purpose of this phase of the project than the distribution across income quintiles. Overall, the relative average VKT of income quintiles is almost fixed over time, except for a slight increase in income group 3's average VKT and a slight decrease income quintile 5's average VKT for the years after 2020. We investigate this more accurately in chapter 4 using a regression analysis after controlling for the impact of other factors to provide more information on the proportion of the variation in average VKT across income quintiles that is attributable to a trend versus the proportion attributable to a wider range of control variables. We provide further details on average VKT by year and income in Appendix C.

---

[30] Inclusion of future IDI MVR snapshots will reduce the number of single odometer observations for years past, improving the accuracy of VKT estimates. This could similarly be achieved by the inclusion of a timeseries dataset of the MVR in the IDI.

[31] For driver statistics, these include 0–15 years, 31–45 years, 46–60 years, 61–75 years and 76 years and over.

Figure 3.3 Average VKT of vehicle owners over time and by income quintile[32,33]



Source: Principal Economics analysis.

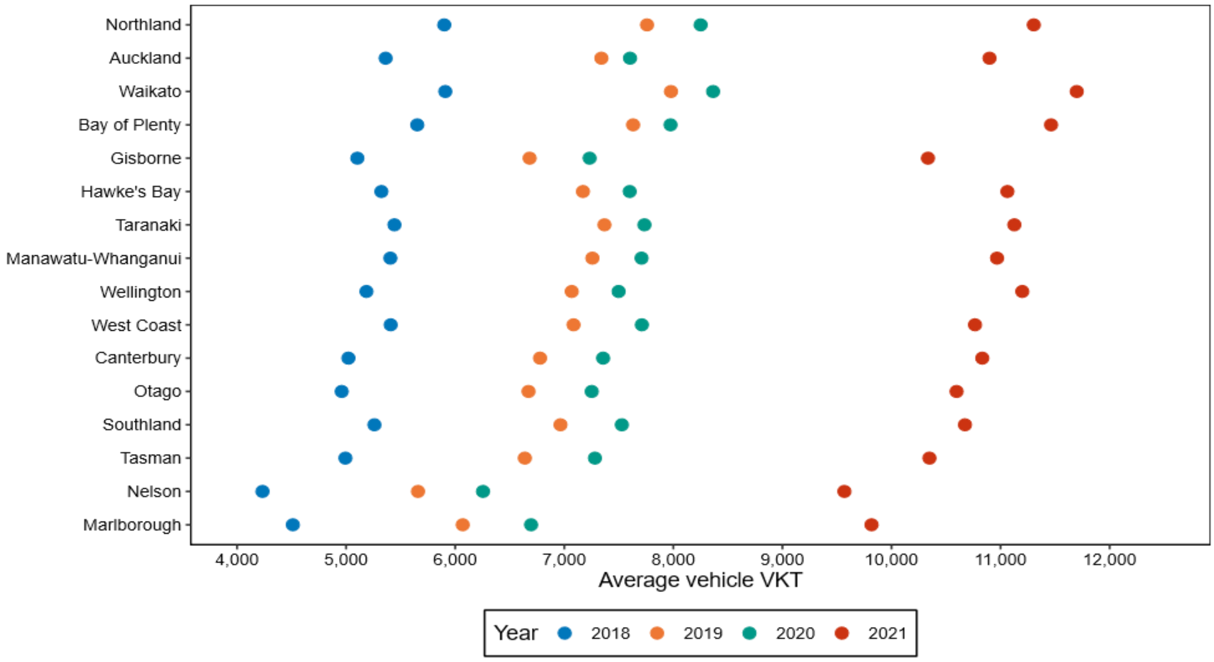
---

32 We have aggregated income into quintiles for analysis. Incomes available in the IDI are in nominal dollar values.

33 As the current dataset is based on a single 30 June saddle point, significant national variances in VKT can impact results. We suspect that the higher average VKT estimated in 2021 is due to the impact of COVID-19 on individual travel patterns and the corresponding odometer readings before and after 30 June 2021. This will be mitigated by taking saddle points for each day of the year in the second phase of this project alongside other adjustments.

We show regional average VKT by vehicle owners in Figure 3.4 and Table 3.9. This has been aggregated based on linkages between VKT, MVR and demographic datasets (greater disaggregation to individual addressing and meshblock granularity for this data is available).

Figure 3.4 Average VKT of vehicle owners by year



Source: Principal Economics analysis.

Table 3.9 Average VKT of vehicle owners by region

Region	2016	2017	2018	2019	2020	2021	2022
Northland	9,828	12,242	13,681	13,642	13,538	17,018	12,573
Auckland	10,178	11,239	12,702	12,694	12,409	16,301	11,414
Waikato	10,501	12,443	13,923	13,913	13,770	17,642	12,903
Bay of Plenty	10,705	11,792	13,194	13,189	12,969	17,091	12,023
Gisborne	9,660	10,334	11,511	11,569	11,670	15,200	10,809
Hawke's Bay	10,079	11,226	12,515	12,477	12,504	16,721	11,546
Taranaki	9,689	11,451	12,908	12,913	12,708	16,946	11,902
Manawatū-Whanganui	9,484	11,413	12,761	12,737	12,662	16,786	11,807
Wellington	10,470	10,971	12,220	12,254	12,123	16,833	11,384
Tasman	9,723	10,759	12,098	12,119	12,013	16,522	11,451
Nelson	8,448	9,101	9,861	9,845	9,975	14,480	9,196
Marlborough	8,505	9,647	10,835	10,804	10,982	15,051	10,215
West Coast	8,681	11,202	12,665	12,544	12,579	16,593	11,904
Canterbury	9,097	10,696	12,101	12,119	12,211	16,942	11,696
Otago	9,475	10,702	11,883	11,978	11,984	16,452	11,530
Southland	9,434	11,353	12,777	12,789	12,788	16,899	12,164

Source: Principal Economics analysis.

The results show a consistent distribution of average VKT particularly for periods less affected by issues from vehicles without odometer pairs (2018–2020). The higher VKT levels in 2021 are due to the greater coverage (higher number of odometer readings) for this year compared to the other years (including vehicles with the 3-year WOF inspection). As shown, the regional patterns are almost the same across the years. This is be further investigated in chapter 6, after controlling for year effects in our regression analysis.

3.4 Summary of the first phase of the project

The purpose of the research is to better understand the motor vehicle profile of households across New Zealand. The objective of the first phase of this project was to identify the usefulness and feasibility of the research project. This section provides a summary and conclusion for the phase of this project.

VKT demographics are critical for informing policy and accuracy of any transport policy analysis. In addition to informing equitable transition to a resilient transport system, VKT profiles will help to improve effectiveness of pricing (and mode shift) policies – identified as critical for estimating generalised costs between modes (Torshizian et al., 2025b). The findings of this report will be critical for any assessment of the effectiveness of pricing policies and the other drivers of VKT. Technically, we identified a critical role for VKT demographics and various modelling frameworks used for informing policy decisions. To this end, we identified the usefulness of this research project to inform various policy topics and the analysis of VKT initiatives.

In chapter 2, we investigated the literature and identified socio-economic features as the most common factors considered for the estimation of VKT, together with features of vehicles and the economic and the spatial characteristics, which are more important for analysis of the level of VKT – noting that the primary focus of this phase of the project was the relationship of VKT with socio-economic features. In this chapter, we investigated the available data sources and described our methodology for linking the large datasets available from the IDI to link odometer readings with locational, social and economic features of households. After an extensive data cleansing process, which was informed by previous attempts to cleanse similar datasets locally and internationally and is described in detail, we derived a reliable large dataset containing information about VKT and household features.

We further investigated the distribution of VKT across socio-economic features and identified that **the available data provides reliable information about the distribution of VKT.** Our investigation of data suggested that the VKT distribution is available for the important factors of VKT, including income, age, sex, region, ethnicity and type of vehicle.

In terms of income, our descriptive analysis indicates that VKT increases with higher income, with the exception of the lowest income group, which has a similar VKT profile to that of the third income quintile. The availability of income data (nominal values) that can be linked to VKT estimates is high, with 2017 having the lowest number of linkages at 93.3%. Ethnicity, age and sex near complete coverage, with VKT estimates of over 99% match rates to individuals and vehicles across all years.

Similarly, as vehicle type is a part of the MVR dataset, there is near complete coverage of vehicle types linked to VKT, their owners and demographic attributes. Vehicle information includes fuel type, engine size, first registration date, gross vehicle mass, industry class, country of origin, make, model and vehicle age. Given the granularity of vehicle information available, it is possible to derive emissions using the same assumptions adopted in the VEPM for deriving aggregated fleet-weighted emissions factors. This means emissions could be estimated for individual vehicles based on their related attributes and estimated VKT.

We conclude that further statistical analysis of patterns of VKT is feasible and the cleansed data can be further linked with various data sources to provide information about a wider range of factors. As described, we compared the VKT distribution across age groups and income quintiles with comparable sources of information and identified similar patterns. We also suggested methodologies for addressing potential measurement errors in the regression analysis of VKT. In the second phase of this project (discussed in the following chapters), this combination of personal, household and location information has been used to create a statistical model that identifies the main characteristics of people and households and their vehicles associated with low and high VKT.

3.4.1 The value add for policy

To ensure our methodology best meets the requirements of this project, this section provides a summary of the valuable insights that could be gained from the findings and application of the identified data (and models). These insights will serve as a foundation for shaping the specific details of our methodology as outlined in the next chapter.

By considering demographic factors, the outputs of the study aim to enhance transport fleet projections, leading to more accuracy with VKT projections and analyses of VKT-related policies. The following topics are potential insights and developments from the outputs of this research project:

- Providing a deeper understanding of how demographic factors influence VKT and fleet composition, which has applications in improving projections of fleet compositions, emissions and transport patterns.
- By refining the composition of individual vehicle characteristics, this research can provide more precise estimations of emissions in the VEPM, contributing to more accurate emissions modelling.
- Improvements of the VEPM include greater disaggregation of fleet composition and potential for alternative emissions modelling based on changes in demographic factors to complement the information available on fleet composition.
- This analysis provides a nuanced understanding of the emissions characteristic of different geographic areas. Comparison of emissions related to specific communities and geographic regions will assist in guiding targeted interventions and policy development – areas where improvements in PT coverage may affect VKT and emissions.
- By examining the variations between emissions patterns in residential areas and those generated from transport paths, the research provides insights into the different sources and distribution of emissions. This will assist in identifying areas where emissions reduction strategies can be most effective.
- Insight into household vehicle choice, level of safety and actual recorded vehicle crash incidents.
- By establishing linkages between demographic factors, fleet composition and VKT, the research supports the validation of ABM techniques such as Project Monty.

We suggest this project will have significant use in its connection with other existing transport and emissions models. Outputs from this modelling can substitute traffic volumes and fleet composition data in the VEMT, alongside improvements in the VEPM, allowing for detailed granular projections of motor vehicle emissions by road segments. To further investigate mode shift, the RLTDM could be used. This provides a comprehensive analytical package for transport and emissions policy assessment.

In a similar scoping study, Cairns et al. (2014) suggest a number of spatially based datasets that could be integrated for analysis beyond those already mentioned. These include indices of deprivation, accessibility indicators and annual energy use. Research undertaken by Rendall et al. (2013) similarly linked New Zealand odometer data to geographic units (area unit granularity) to evaluate fuel use by households and energy consumption for planning purposes.

In our assessment of emissions policy levers for GHG reductions, we provide estimated elasticities of public and private transport for both geographic and demographic attributes (Principal Economics, 2022). Combining these elasticities with VKT allows for the highly granular assessment of pricing interventions and their impact on travel patterns and spatially in terms of who is affected and the demographic groups impacted. This additionally allows for highly granular analysis of equity impact from various VKT policies.

4 VKT and household features

As shown by our literature review, the features of households play an important role in explaining VKT. Hence, any estimation of VKT must allow for household features such as age, household composition and household income. This chapter provides a detailed description of the variation in VKT by the features of households. The number of cross-tabulations that could improve our understanding of VKT patterns is unlimited. Hence, in this chapter, we mainly focus on reporting the correlations between different features and VKT patterns and leave further investigation of cross-tabulations to the next chapter's econometric analysis. There is a very extensive range of data and information available relevant to the topic of this project – our intention is to provide the most relevant. The modelling frameworks used in the next chapters cover a wider range of information and it is neither useful nor feasible for this chapter to provide that.

This chapter starts with a summary of the most relevant information to patterns of VKT and then presents our findings from our granular data. We also provide a summary of our observations from the data.

4.1 A summary of relevant regional patterns of VKT

In our analysis of descriptive statistics of VKT in this chapter, we demonstrate the usefulness of having a better understanding of the regional patterns of VKT and PT for explaining the observed patterns. This section provides a brief summary. It is important to note that the source of the data used in this section is more aggregated information available outside of the IDI – the data source is different from the data presented in the next sections.

Torshizian et al. (2025b) identified policy levers for reducing VKT in Auckland, Wellington and Christchurch and measured their impacts. That study provided an extensive description of VKT patterns across the three large urban areas. This includes demographic and road transport features, variation in mode choice with changes in the features of trip, VKT status, determinants of PT choice and factors affecting PT use across the three regions. A few important findings that are relevant to this project are as follows:

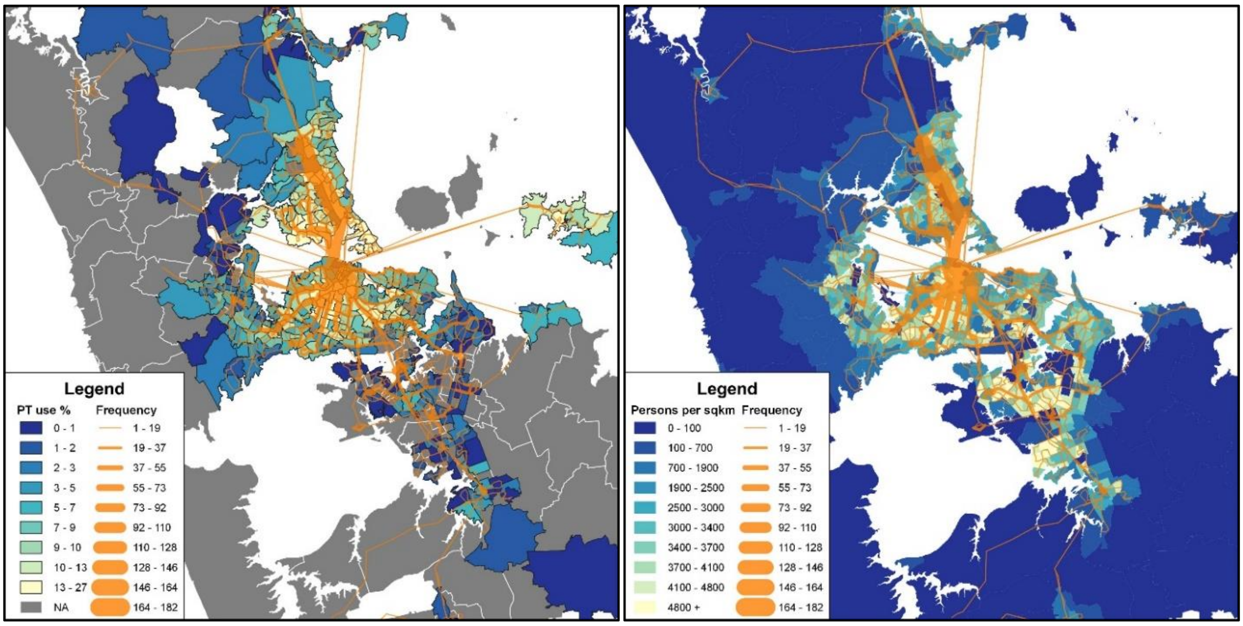
- New Zealand has the fourth-highest ratio of passenger cars per capita in the world with 645 cars per 1,000 inhabitants in 2018 (based on the OECD's 2023 passenger cars dataset).
- Auckland and Christchurch have lower PT trips per capita compared to cities of similar density and for some Australian cities of lower density.
- Canterbury has the longest road network of the three regions, while Wellington has the shortest. As the most populous region, Auckland emits three times the direct CO_2 emissions at the regional level compared to the Canterbury region. The Wellington region emits the least CO_2 of the three regions despite having a higher population density than Canterbury.
- Personal travel is responsible for a large portion of VKT at different times of the day. This may suggest that the policies affecting personal travel may have greater impact on VKT than policies targeting other types of travel.
- Higher-income households in New Zealand tend to have greater access to private vehicles. Across all deprivation deciles, a private vehicle is the predominant transport mode for workplace commutes by a significant margin. Other transport modes such as PT and active transport (walking and cycling) account for less than 20% of commutes across all regions and deprivation deciles.
- PT use for workplace commutes in the Auckland region is most common for deciles 5–6, with notably less use in decile 10 (most-deprived) areas. In the Wellington region, PT use is more prevalent in low-decile (less-deprived) areas and less used in high-decile (most-deprived) areas. In the Canterbury region, PT use is more common in high-decile (most-deprived) areas.

- Across all deprivation levels, individuals who work in the city centre are significantly more likely to travel using PT. Despite the city centre typically being the most connected area in terms of PT, most individuals across the Auckland, Wellington and Canterbury regions work outside of the city centre. In high-deprivation areas within the Auckland region (deciles 9–10), a lower proportion of workers (7–14%) commute to the city centre compared to areas with lower deprivation (deciles 1–8), where a higher proportion of workers commute to the city centre (17–21%). This observation partly explains the lower use of PT by lower-income groups in Auckland.
- For individuals who reported that using PT was difficult or very difficult, among lower-income households, the main reason is due to a lack of available PT, while high-income households reported that PT is too far away (Torshizian et al., 2025b, p. 129).

An important finding of the Torshizian et al. report is that PT coverage has significant implication for VKT patterns.[34] Figures 4.1–4.3 show the PT network stop-to-stop service frequency for an average Tuesday morning (7–10am) and the use of PT for commuting to work. We aggregate stops within close proximity as single interchanges and determine the total number of services travelling from stop to stop, aggregating all services regardless of route. PT use is based on the proportion of PT use for commuting to work at SA2 granularity using Census 2018 data.

Figure 4.1 shows that areas in Auckland with a high PT frequency tend to have a high proportion of PT commuters (most notable in the inland North Shore and central areas). Similarly, areas with ferry services (including Waiheke Island, Devonport, Bayswater, Birkenhead and Beach Haven) also have high PT use.

Figure 4.1 Auckland PT network and use/persons per square kilometre

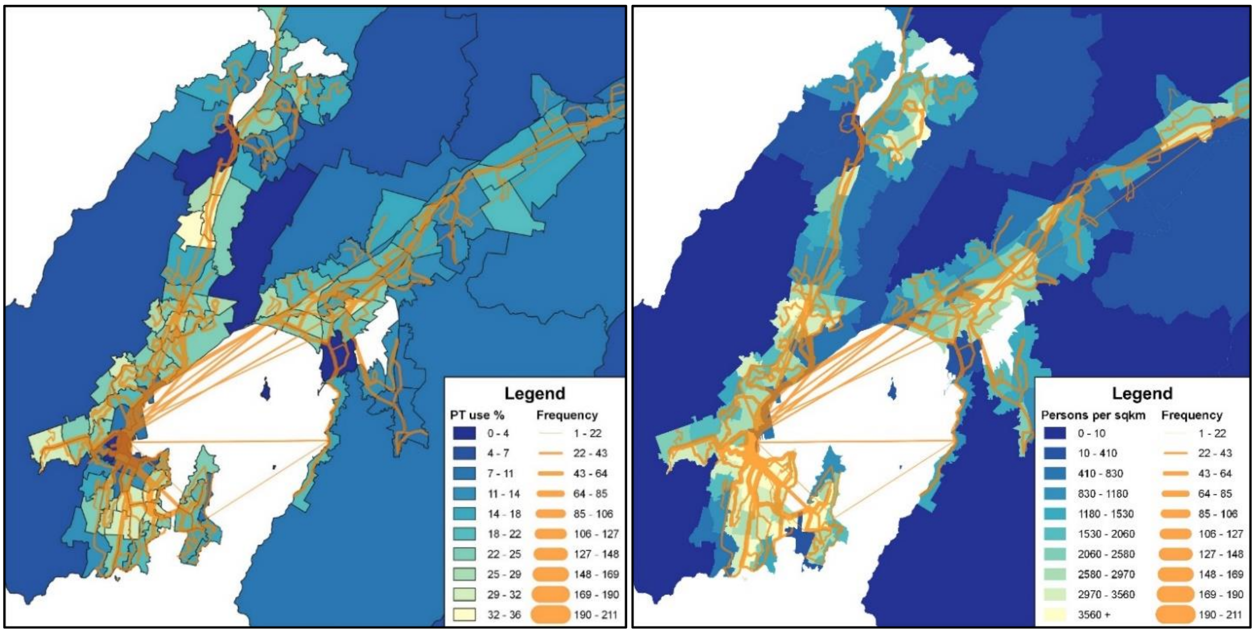


Source: Torshizian et al., 2025b.
Note: Route frequencies reflect the number of services running from 7–10am on Tuesday 4 October 2022 using General Transit Feed Specification (GTFS) data published by Auckland Transport for 29 September 2022.

Figure 4.2 shows more consistent PT frequency in Wellington. A smaller proportion of the population use PT in the CBD due to a high proportion of commuters using active transport modes (walking and cycling).

[34] In addition to the descriptive statistics, the statistical significance of the findings was tested and its implication for effectiveness of VKT reduction policies assessed.

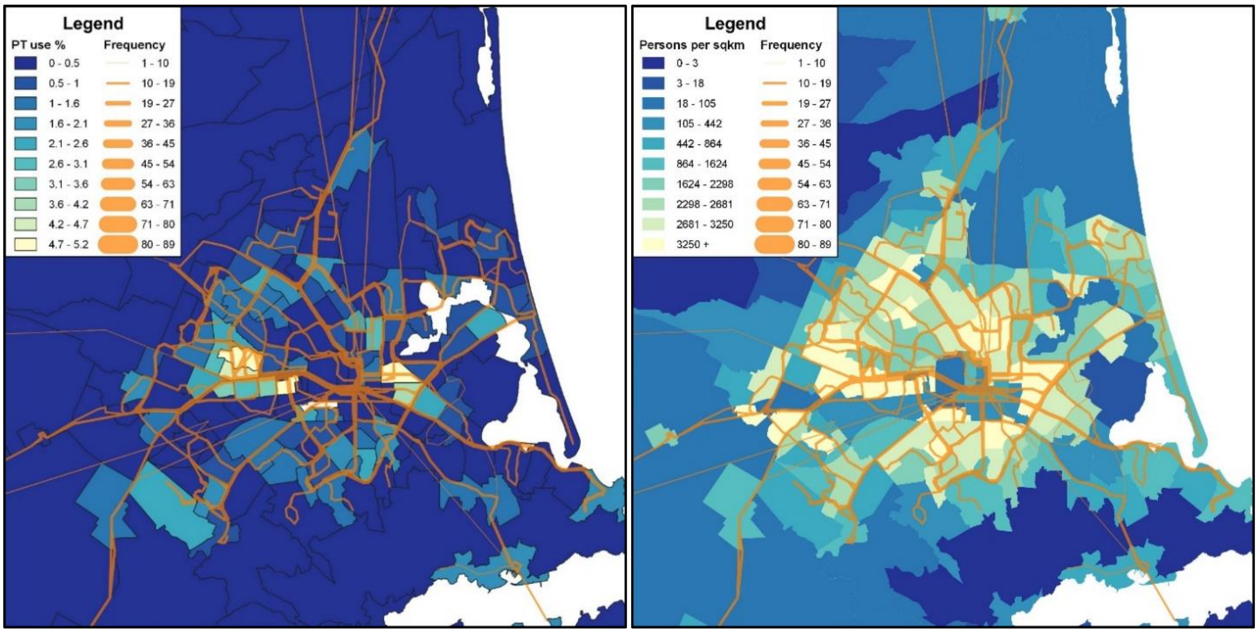
Figure 4.2 Wellington PT network and use



Source: Torshizian et al., 2025b.
Note: Route frequencies reflect the number of services running from 7–10am on Tuesday 10 May 2021 using GTFS data published by Metlink for 3 May 2021.

Figure 4.3 shows low PT use in Christchurch and lower frequency compared to Auckland and Wellington. Areas with higher PT use include areas surround Canterbury University and the Linwood Bus Hub.

Figure 4.3 Christchurch PT network and use

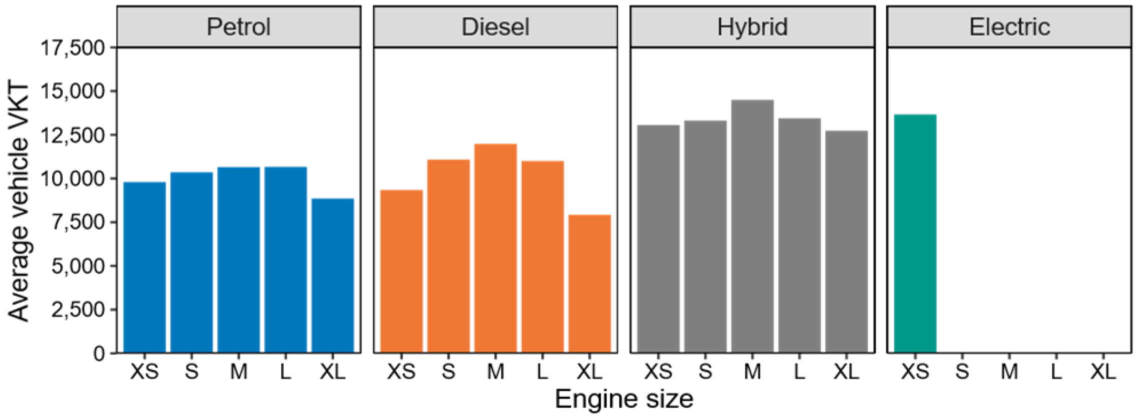


Source: Torshizian et al., 2025b.
Note: Route frequencies reflect the number of services running from 7am–10am on Tuesday 18 October 2021 using GTFS data published by Metlink for 17 October 2021.

4.2 Odometer readings are higher for electric and newer vehicles

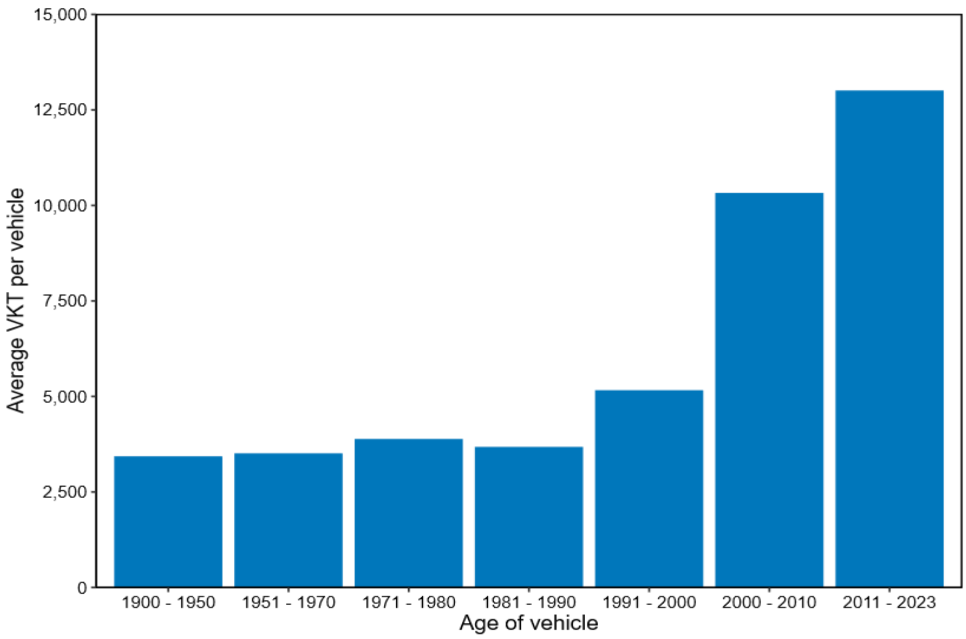
There is a consensus about the important role of vehicle type on VKT. As show in Figure 4.4, average VKT is higher for electric and hybrid vehicles than the other types. This may suggest that EV users maximise use of their vehicle due to its lower (per km) variable cost. Across larger vehicle classes, hybrid still has the highest average VKT. An important takeaway for emissions reduction policies is that a policy objective to increase EV uptake could have opposing impacts to a potential VKT reduction policy (unless these policies will be associated with alternative policies/incentives). As expected, the average VKT is lower for older vehicles (Figure 4.5). There is generally a positive correlation between emissions star rating and VKT (Figure 4.6).[35]

Figure 4.4 VKT per vehicle by type for 2021



Source: Principal Economics analysis.

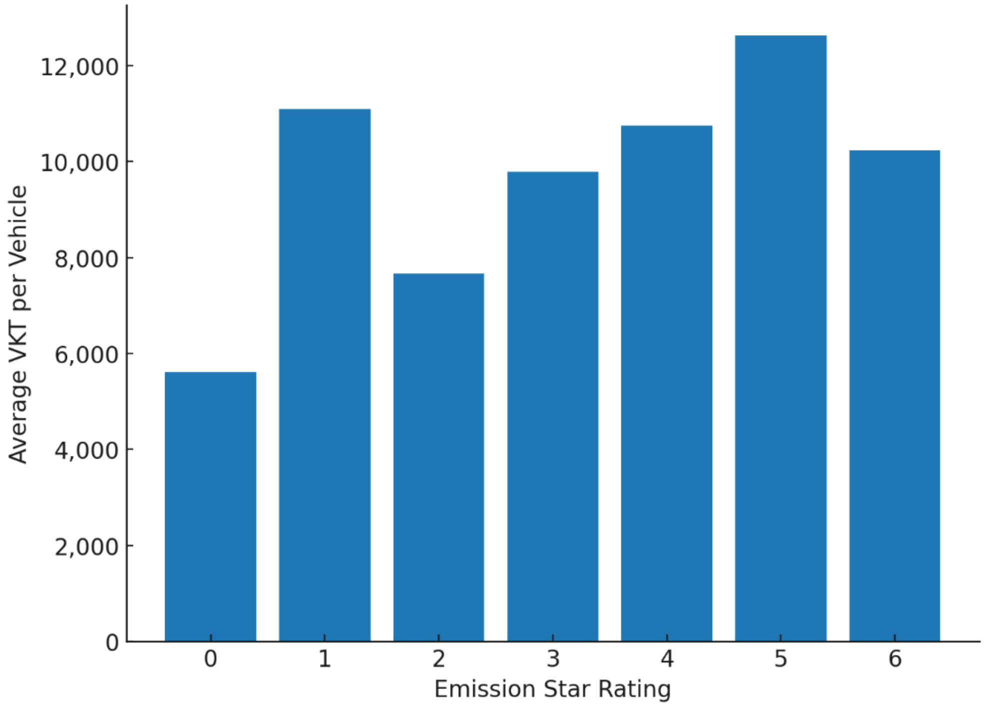
Figure 4.5 Average VKT the age of vehicle for 2021



Source: Principal Economics analysis.

[35] A rating (1–6 stars) is applied. EVs score 6 stars and other vehicles get fewer stars, depending on the health impacts of their emissions.

Figure 4.6 Average VKT by fuel efficiency of vehicles for 2021



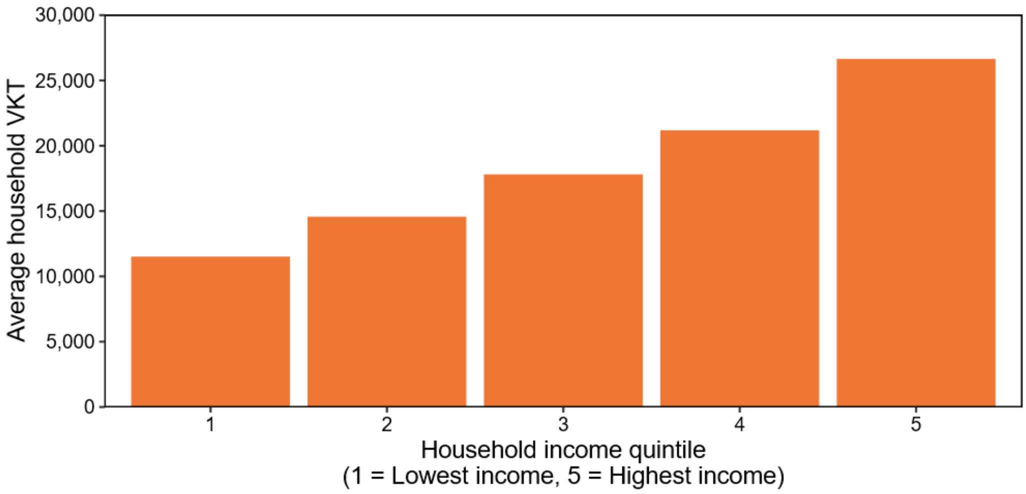
Source: Principal Economics analysis.

4.3 Variations in VKT by population demographics

There is a positive relationship between VKT and income, household size, youth and having dependent children.

Figure 4.7 shows the level of VKT per household for each household income quintile. Accordingly, there is a positive correlation between income and VKT.

Figure 4.7 VKT per vehicle by household income quintile

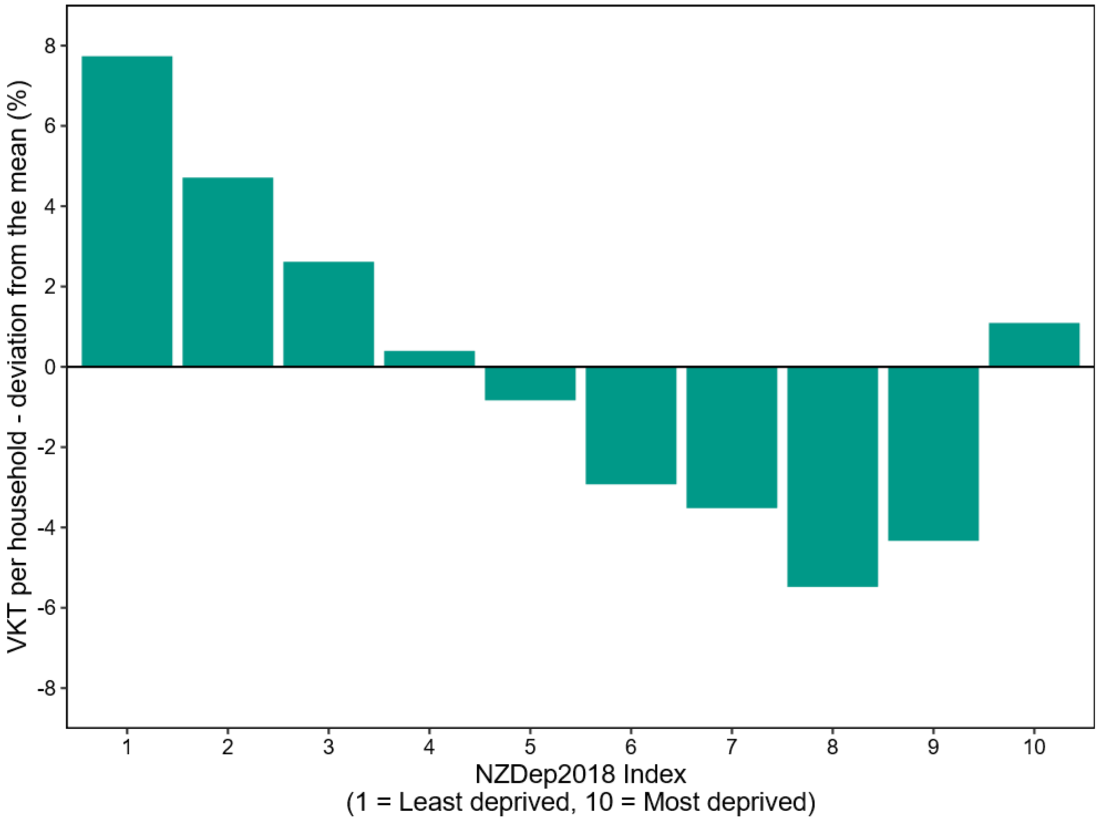


Source: Principal Economics analysis.

This is particularly important for policies targeting VKT. In Auckland, for example, we observe that high-income groups are the main users of PT. Our observation that the high-income group also has higher VKT suggests an overall higher use of the PT system by the high-income group due to better alignment of their origin and destination with PT routes. Also, Torshizian et al. (2025b) identified a lower fuel price sensitivity for the higher-income group. The mix of these factors has important implications for any VKT reduction policy. We will need to further test the variation in our observation of VKT by household income by regions.

The average VKT per household across NZDep[36] groups is 18,947 km. Figure 4.8 shows the deviation of VKT per household from the mean for different NZDep groups. Accordingly, the least-deprived groups have higher VKT per household, with the exception of deprivation group 10 (most-deprived), which has a higher VKT per household than the average.

Figure 4.8 VKT per household deviation from the mean by NZDep2018

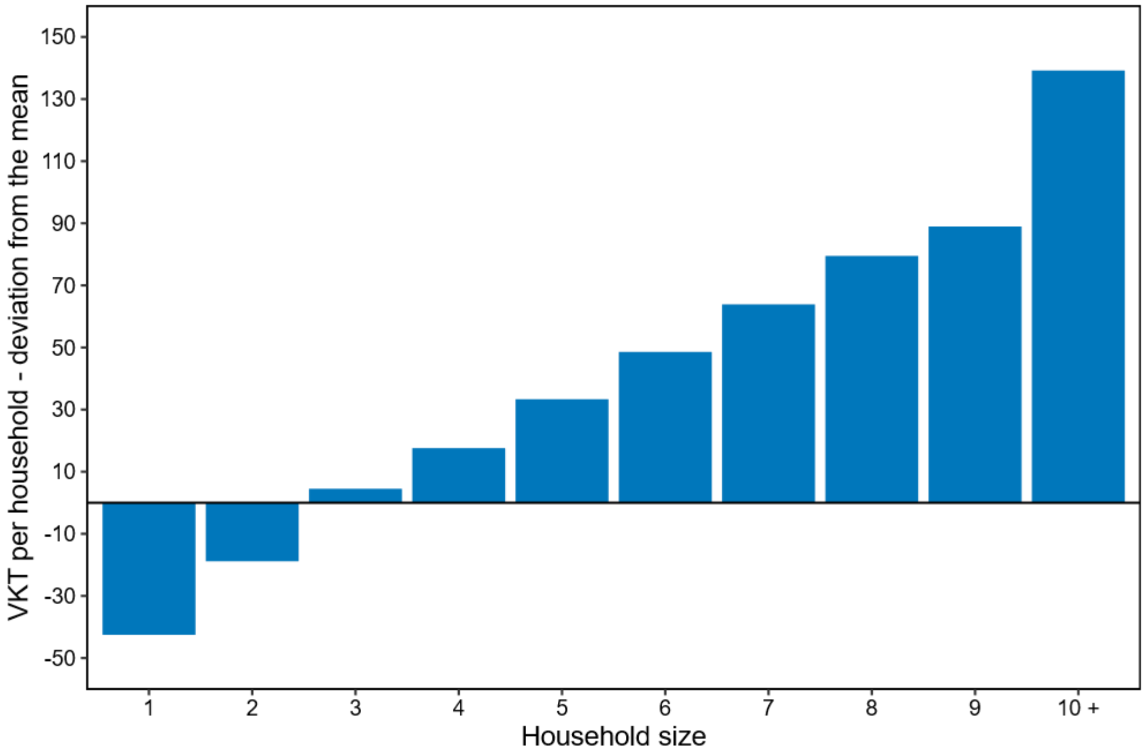


Source: Principal Economics analysis.

The variation in VKT by household size is shown in Figure 4.9. Accordingly, households with one or two people have a lower VKT that the mean. Overall, there is a positive correlation between household size and VKT. In the next chapter, we further investigate this relationship after controlling for confounding factors such as income, age and car ownership.

[36] The NZDep2018 ordinal scale ranges from 1 to 10, where 1 represents the areas with the least-deprived scores and 10 the areas with the most-deprived scores (Atkinson et al., 2020, p. 8)

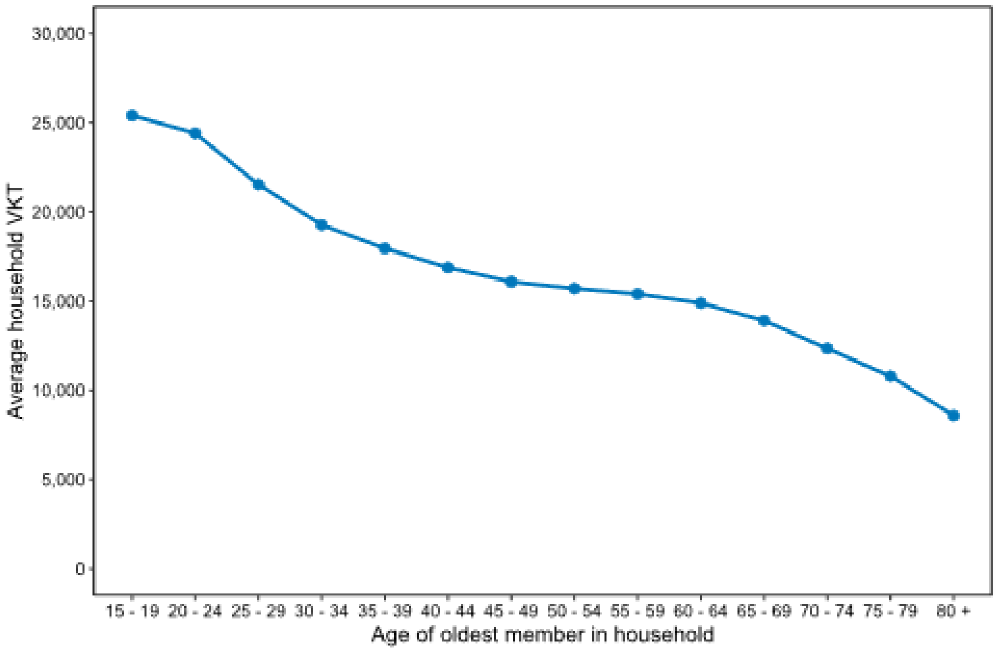
Figure 4.9 VKT per household deviation from the mean by household size



Source: Principal Economics analysis.

Figure 4.10 shows a negative relationship between average household VKT and the age of the oldest member in household. Figure 4.11 shows that the VKT per household increases with the number of dependent children.

Figure 4.10 VKT per household by oldest member in household



Source: Principal Economics analysis.

Figure 4.11 VKT per household by number of dependent children

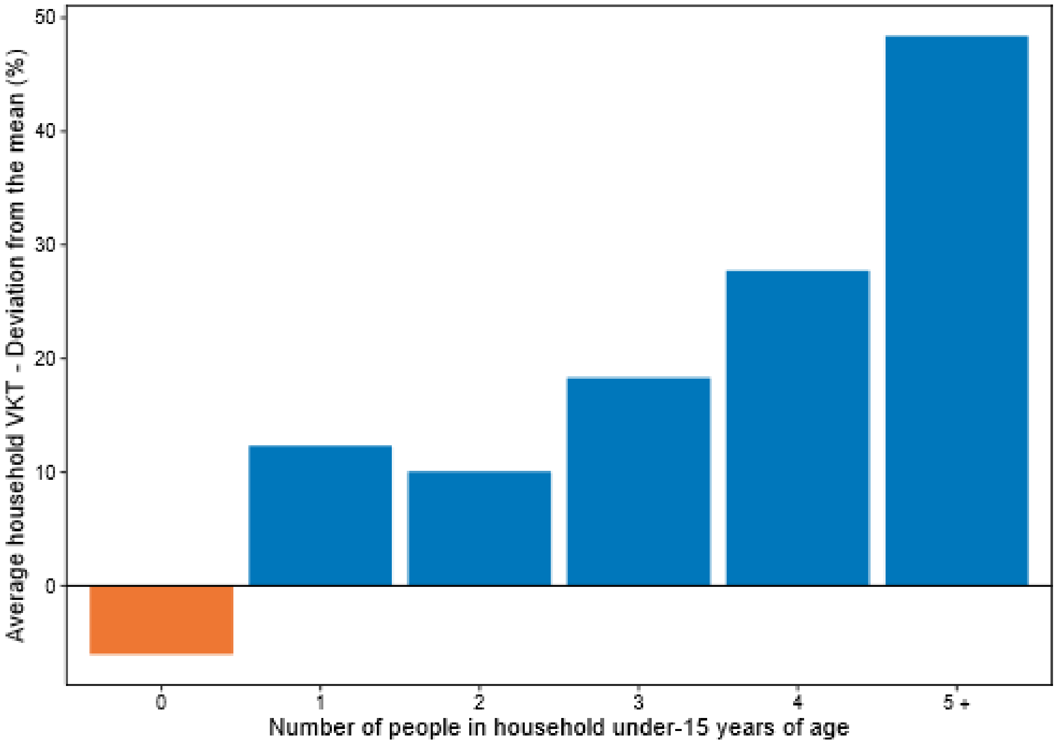
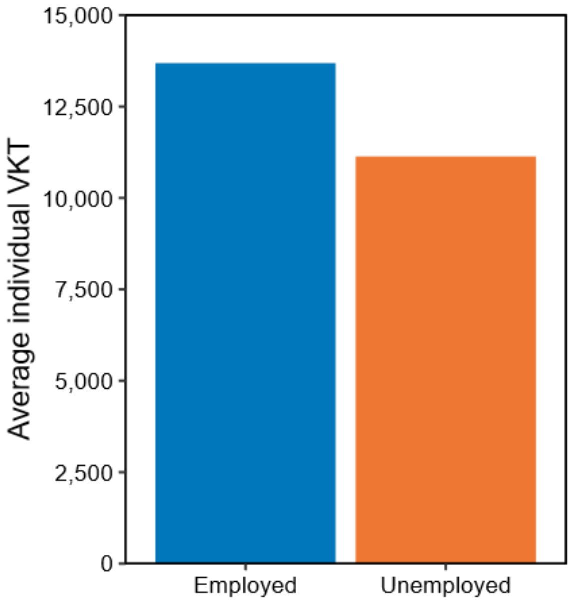


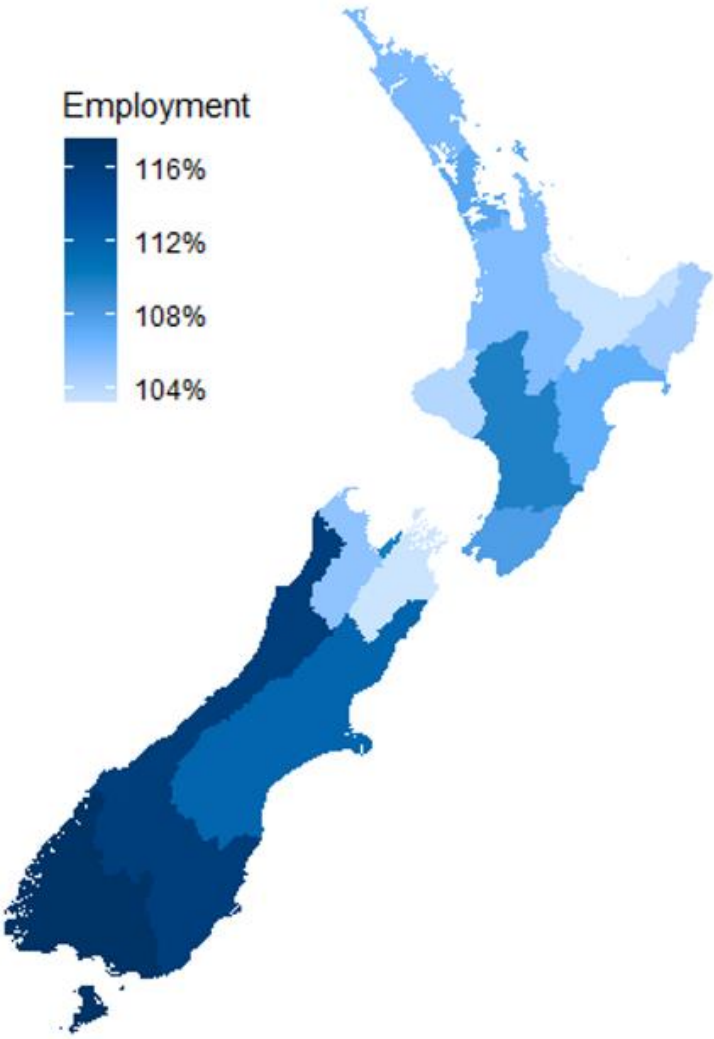
Figure 4.12 shows the average individual VKT by employment status. Accordingly, employed individuals report higher VKT. We further investigated the variations across regions in Figure 4.13, which shows the ratio of VKT for employed compared to unemployed across regions. Accordingly, employed people have higher VKT than unemployed people, particularly in the South Island. This could be due to the regional features or the features of industries in the South Island.

Figure 4.12 VKT and employment status



Source: Principal Economics analysis.

Figure 4.13 The ratio of VKT for employed to unemployed across regions



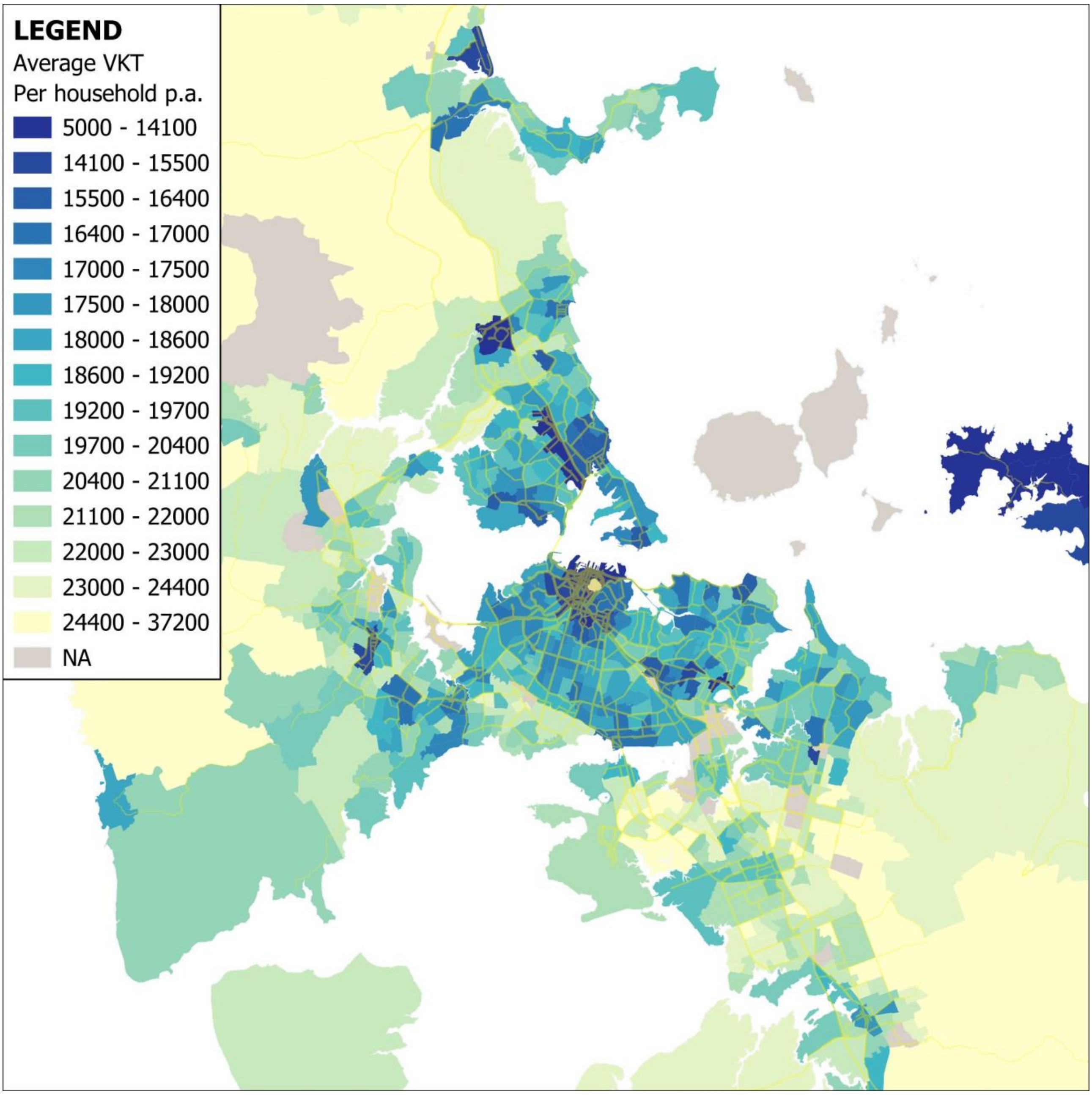
Source: Principal Economics analysis.

4.4 Spatial patterns of VKT indicate close correlation with PT coverage

Figure 4.14 shows the distribution of VKT per household across Auckland suburbs. A comparison between this map and the PT network coverage in Auckland[37] shows that the areas with lower VKT have a high PT coverage. As expected, the more distant and rural areas have the highest VKT per household in Auckland. The spatial patterns of VKT per household for Wellington and Christchurch are shown in Figure 4.15 and Figure 4.16. Despite the seemingly less-clustered distribution of VKT across Christchurch suburbs, we suggest that PT coverage could potentially explain the observed patterns.

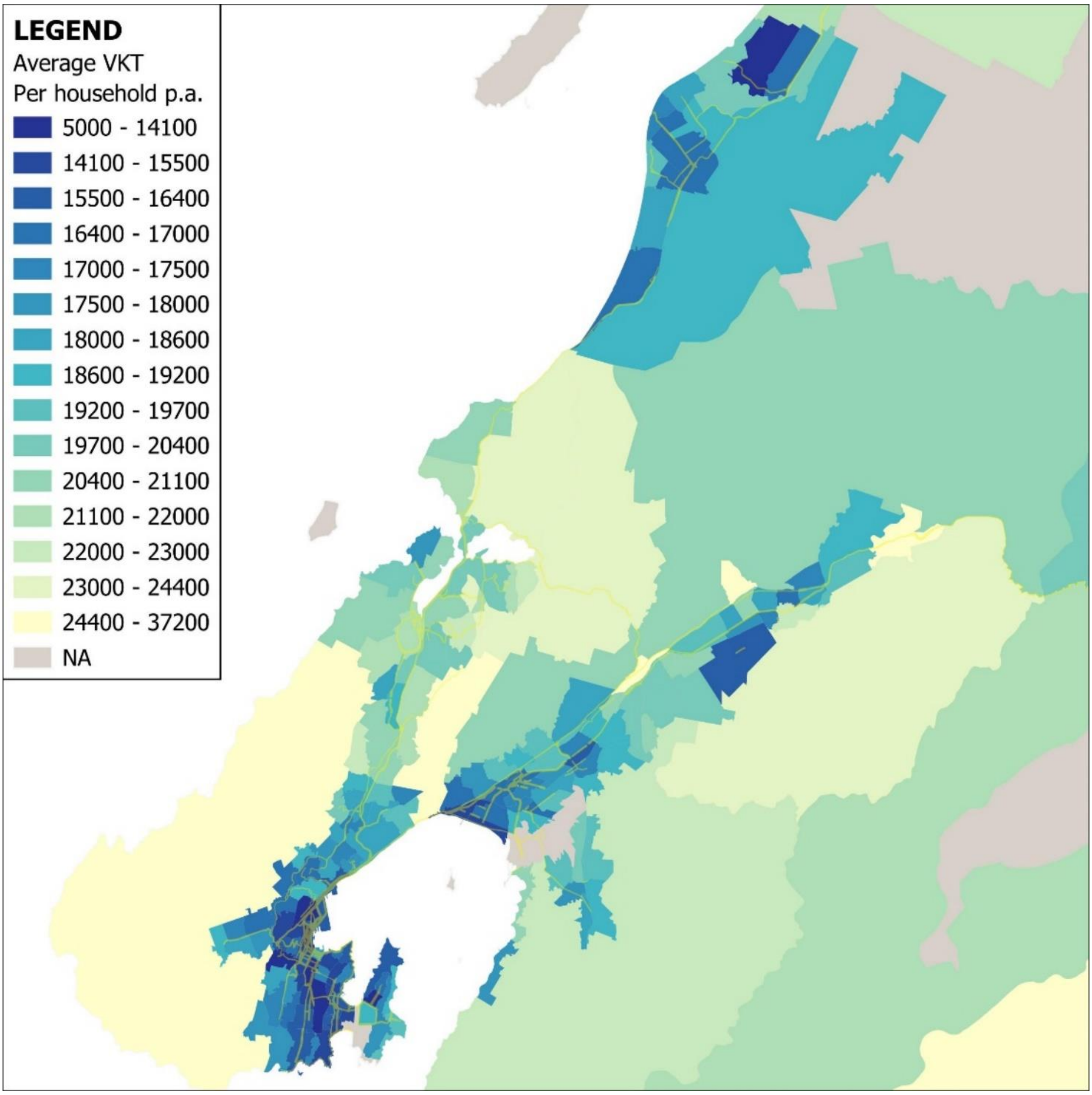
[37] For the PT network, refer to Torshizian et al., 2025b, pp. 122–123.

Figure 4.14 Average VKT per household across Auckland suburbs



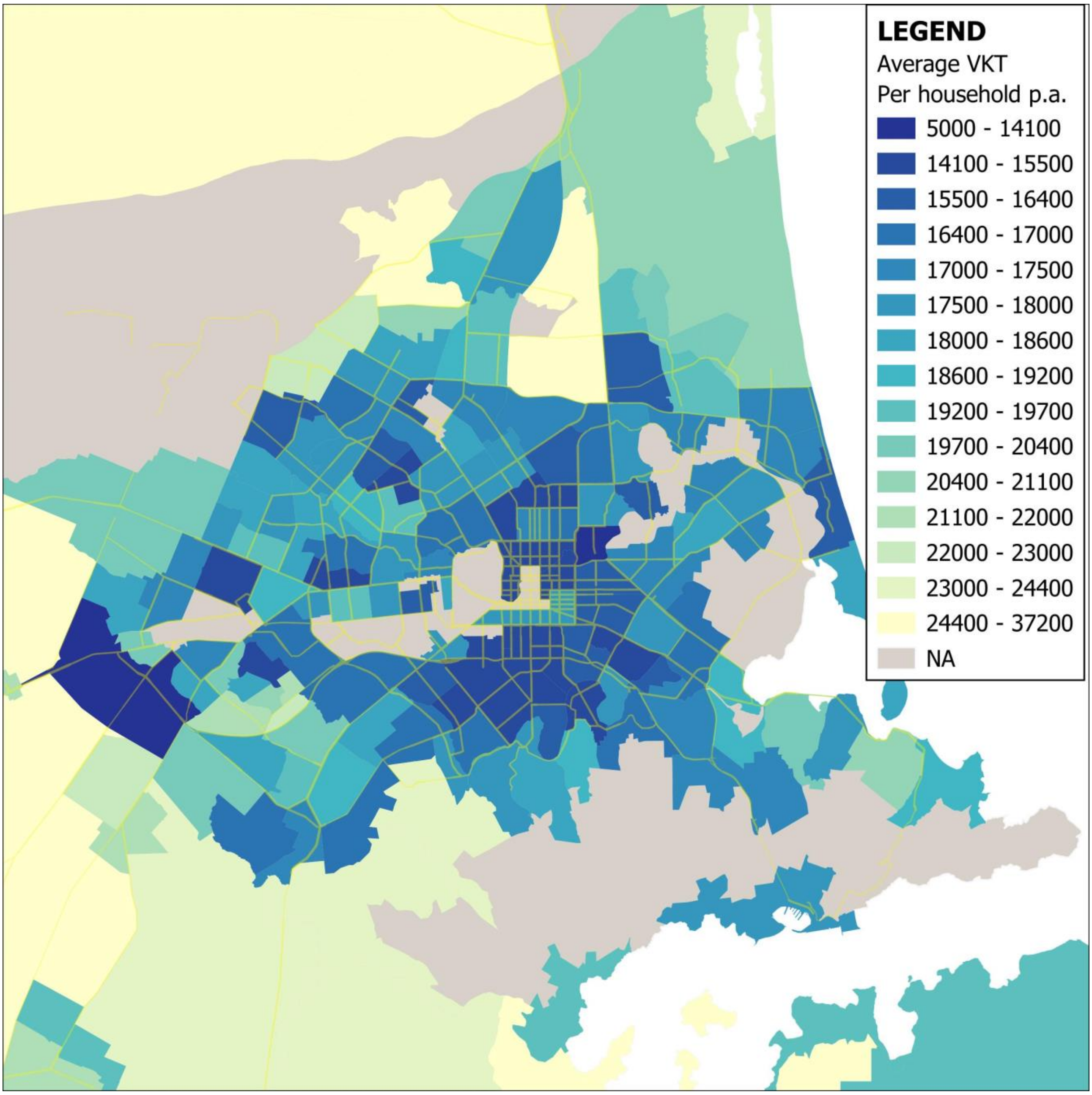
Source: Principal Economics analysis.

Figure 4.15 Average VKT per household across Wellington suburbs



Source: Principal Economics analysis.

Figure 4.16 Average VKT per household across Christchurch suburbs

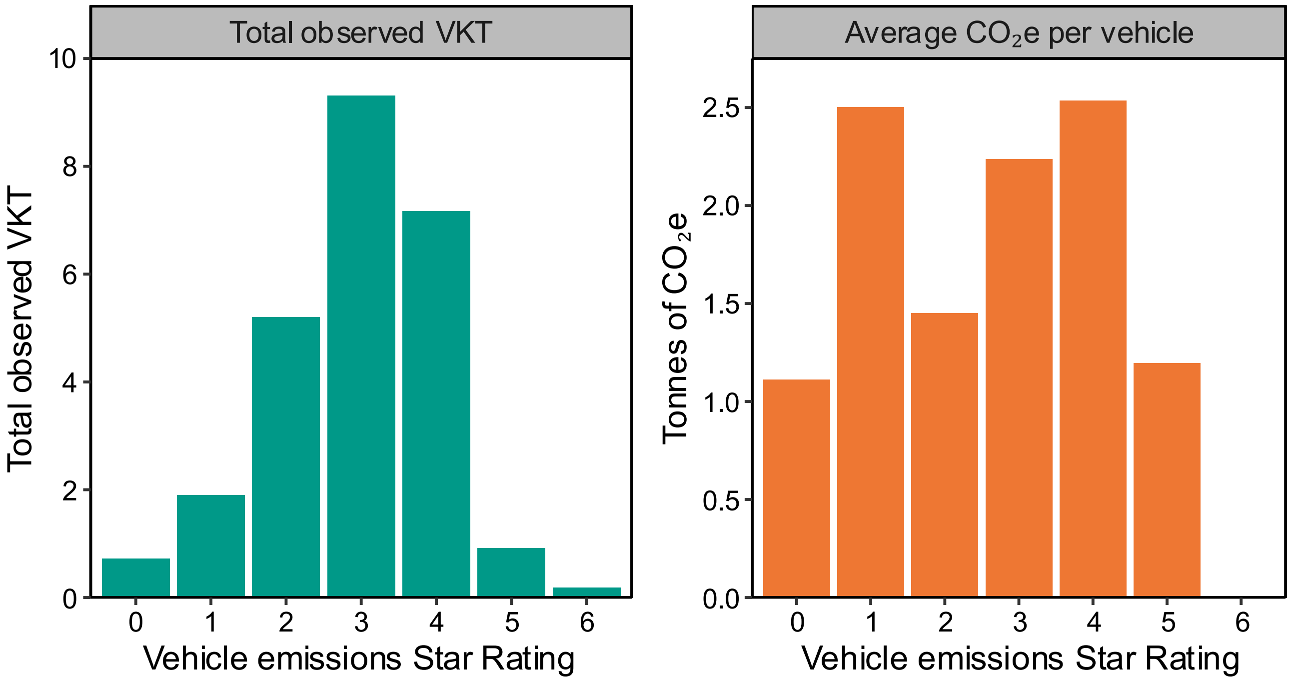


Source: Principal Economics analysis.

4.5 Demographics of VKT and GHG emissions

Figure 4.17 compares the fleet's emissions profiles (star rating) with total observed VKT and average CO_2e per vehicle. The total VKT shows a higher VKT for average emissions star vehicles. We do not observe a clear pattern for tonnes of CO_2e. Hence, we suggest that we need to further control for other factors to derive a useful inference. Compared to Figure 4.6, which shows that the average VKT of 1-star emissions vehicles is high, their total VKT is relatively lower than other higher star ratings.

Figure 4.17 VKT and emissions star rating for 2021

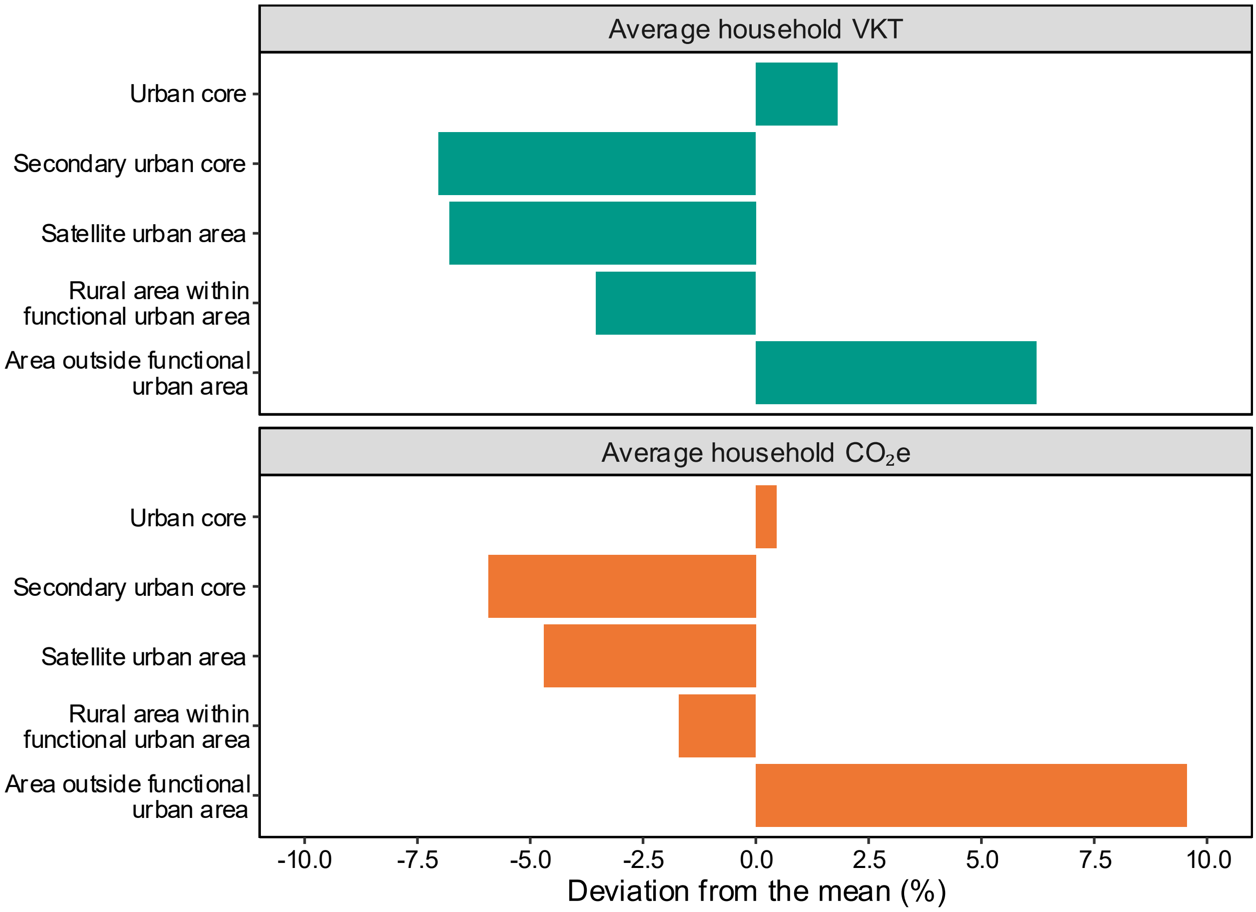


Source: Principal Economics analysis.

Figure 4.18 shows the deviation from the mean VKT and CO_2e per household for different urban forms. The areas categorised as urban core indicate slightly higher average VKT compared to other urban forms. As expected, the suburbs outside the urban area have the highest average VKT and CO_2e per household.

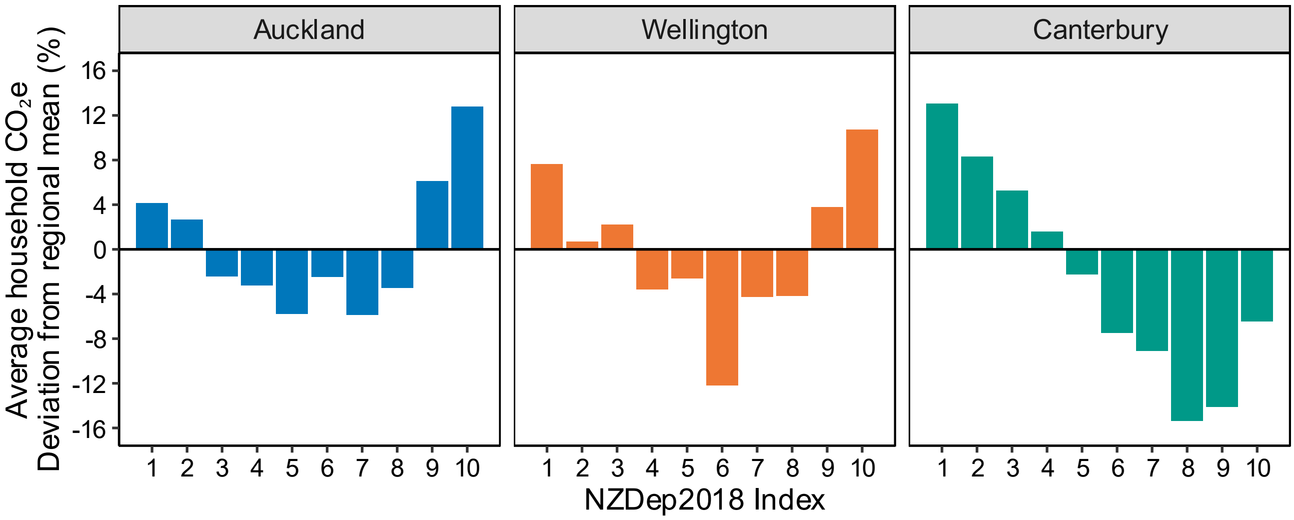
To further investigate the pattern of CO_2e and VKT, Figure 4.19 presents the deviation from the mean for the average household CO_2e by NZDep groups. While the most-deprived areas indicate high emissions, there are variations in the emissions observed for the least-deprived areas across the regions. Given that we are presenting descriptive statistics, it is difficult to reach causal inference at this stage, but we further investigate the observed pattern in the next chapter using econometric analysis.

Figure 4.18 VKT and urban form



Source: Principal Economics analysis.

Figure 4.19 Average household CO_2e (deviation from the mean) by NZDep



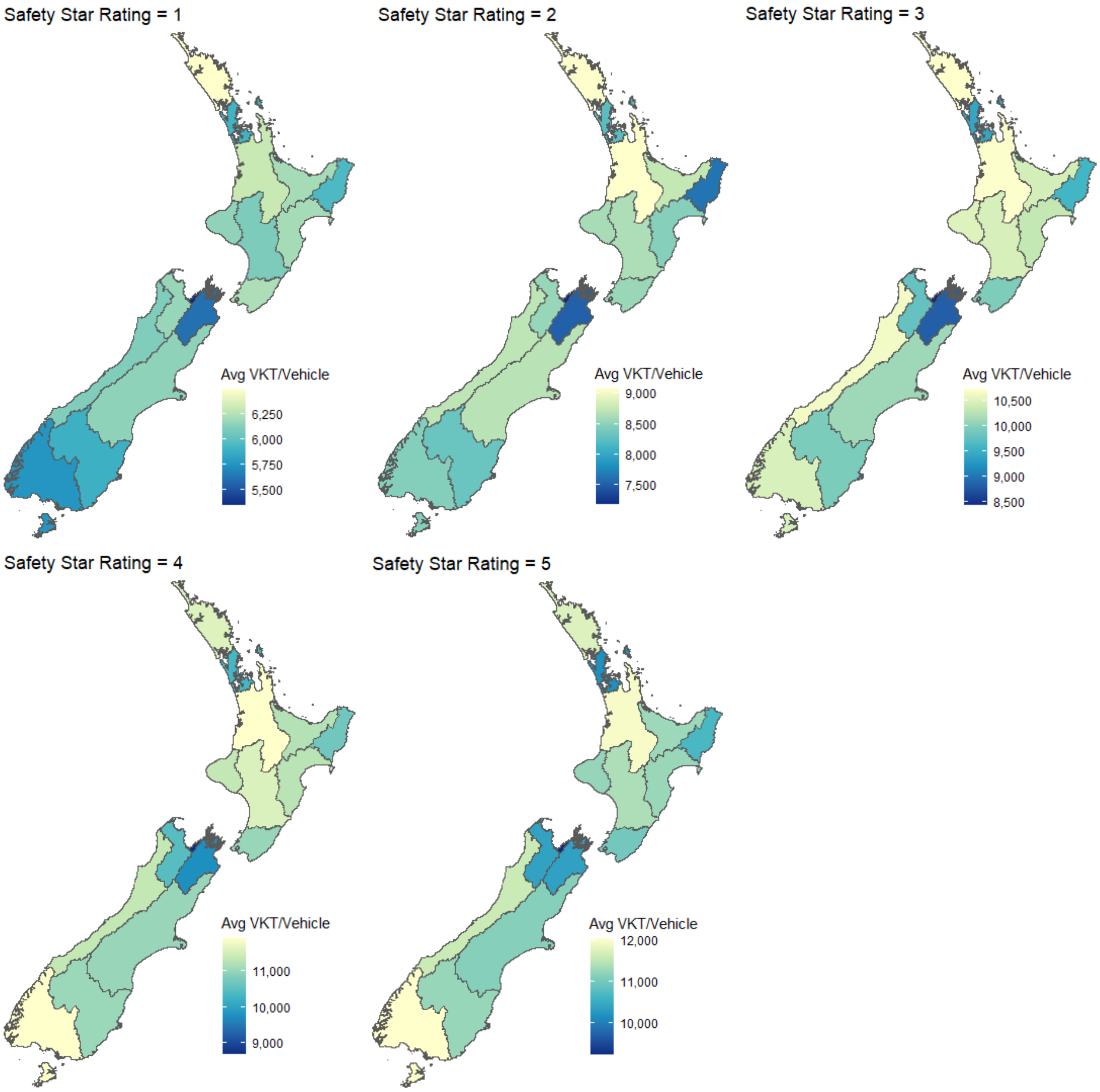
Source: Principal Economics analysis.

4.6 Regional safety profile of the fleet

NZTA's safety ratings provide information about a vehicle's likely safety in a crash. A higher star rating indicates better safety.

Figure 4.20 shows the regional distribution of VKT per vehicle for each safety rating. While an illustration of safety rating alone would show that larger urban areas have the safest vehicles, it is interesting to see that Auckland still has a relatively high VKT per vehicle for low-safety star vehicles.

Figure 4.20 Regional distribution of safety ratings

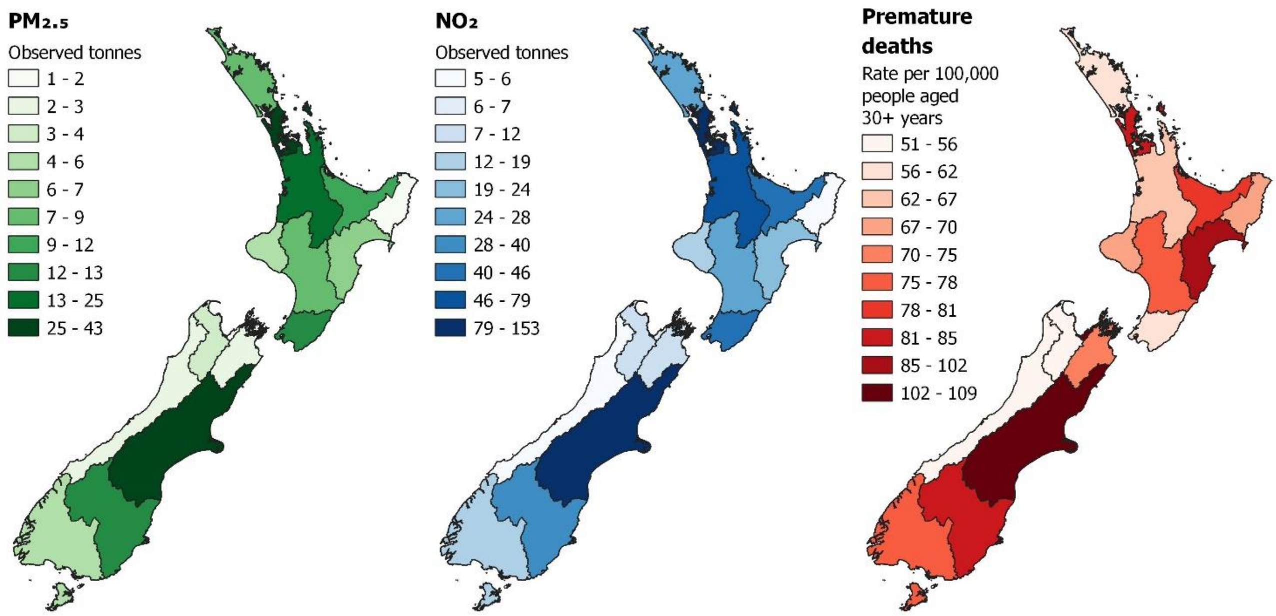


Source: Principal Economics analysis.

4.7 Air quality health statistics and spatial distribution of fleet

A technical requirement of this project was to derive calculations of VKT emissions. Figure 4.21 shows the spatial pattern of air quality, including $PM_{2.5}$, NO_2 and premature death across regions.

Figure 4.21 Air quality by region



Source: Principal Economics analysis.

Summary of the chapter

This chapter describes the patterns of VKT for vehicle features and household demographics and identifies its spatial and air quality co-variants. The main findings are as follows:

- Higher income, younger age and more dependent children are positively correlated with VKT.
- Employment is positively correlated with VKT.
- VKT is higher for households located in the urban core and in areas outside of the urban area. This pattern might be explained by demographic differences, which are investigated in the next chapter.
- From the spatial pattern of VKT across suburbs of large urban areas, we infer negative correlation with PT coverage.
- VKT is higher for electric and hybrid vehicles and (on average) decreases with the age of vehicle.
- Total VKT is higher for average emissions star vehicles.
- Auckland has relatively high VKT per vehicle for low-safety star vehicles.

We further investigate these correlations in the next chapter using econometric analysis.

5 Data, statistical model and information for end use

As described in the previous chapter, there are many factors associated with VKT. The key objectives of this chapter are to establish a statistical model for modelling VKT with a focus on demographic factors and to test the partial correlation between those factors, keeping other factors constant, and VKT.

5.1 Our stepped regression analysis provides an understanding of VKT demographics and confounding factors

To analyse VKT, we rely on the features of the population, suburbs, regions and vehicles. Since our objective is to test the correlation between demographic factors and VKT, we first estimate the impact of demographics on VKT based on (Equation 5.1):

$$vpd_i = \theta + \beta_j . d_{ji} + \rho . y_i + \varepsilon_i \qquad \text{(Equation 5.1)}$$

where:

vpd_i = VKT per day for the specific reading period

θ = intercept

d_{ji} = a vector of demographic features, including sex, age, income, household composition and education

y_i = year dummies

β_j = parameter vector estimating correlation between d_{ji} and vpd_i

ρ = parameter estimate capturing the trend impact

ε_i = the random error term.

Given the skewed distribution of VKT per day, we use a Poisson model for the estimation and check the consistency of the results with an OLS regression with log-transformed output – log(vpd). This is because, with our large sample size, Poisson regression is computationally intensive. If the results of the OLS, particularly the estimated standard errors, do not differ significantly from the Poisson model, we will use OLS.

Then we estimate the following regression models to understand the overlapping effects of various features:

- The first model only includes the demographic variables (d_{ji}). This is because the primary focus of our work is on demographics of VKT and we are interested to see how the impacts of demographics may change with the inclusion of other features.
- The second model adds the regional control variables, including region dummies, population density (at the SA2 level), time measures for travel to the nearest town centre using car, PT, walking and cycling. We will compare the outputs of this model with the first model to see if the regional features are the driver of any change in correlation between demographic features and VKT.[38]
- The third model includes vehicle features such as vehicle year, body type, engine technology and engine size. Given that vehicle features are expected to be highly correlated with VKT, we separated them from demographics and other features in this model to see their impact in isolation.

[38] We acknowledge that, due to their macroeconomic nature, the inclusion of regional features will be associated with an increase in within-group error for these variables. Hence, the results need to be interpreted carefully.

- The fourth model adds the regional features to the third regression. This is informative for understanding how regional features affect the correlation between vehicle features and VKT versus the wider range of factors considered in the next model.
- The fifth model includes all demographic, regional and vehicle features. This presents a full set of results, but given the importance of PT coverage and recent policy emphasis on EV adoption, we continue the analysis by including those factors in the seventh and eighth model respectively.
- In addition, the sixth model includes the same variable list as the fifth model but limits the observation year to 2021. This provides information about how the coefficients change in the year that we have the most complete dataset, but we note that it is a year affected by COVID-19 measures.

The same methodology is used for car ownership regression analysis with the outcome being a dummy variable of having a car (1) or not (0). Then we model the impact of various features on the uptake of EVs. This is beyond the scope of the current project, but given the granularity of our dataset, we decided to provide preliminary results on the factors of EV uptake. EV uptake is defined for the households that did not have an EV before and then reported having an EV. For this, we use the same set of explanatory variables, but the output is a dummy variable of EV uptake or not. For the estimation, we used a logistic regression. This regression analysis will only focus on a few correlates of EV adoption. While limited, the results provide the most disaggregated analysis of EV adoption that we are aware of and shows the value of the constructed dataset for further investigation of the factors of EV adoption and other policies in the future.

5.2 Estimation results suggest a relatively stable correlation between demographics and VKT but no significant explanation for vehicle ownership

The extensive results table is provided in Table D.1 in Appendix D. The best goodness of fit is for the VKT model that includes all variables. In comparison between the model that only includes demographics and the one that only includes vehicle variables, vehicle features provide a better explanation for VKT (based on the goodness of fit using the adjusted R-squared measure). After adding all variables, the results are presented in column 5. The regression analysis using the sampling weights is computationally burdensome. Hence, we tested if the results changed by limiting the sample population to 2021, our most comprehensive dataset for a year. The sign and statistical significance of the key variables are unchanged, and hence, the 2021 sample could be used for future analysis. In this table, we keep using the full sample, but the consistency of the 2021 results with the full sample results indicates that our results are not biased due to dropped observations in other years. Then we tested the impact of proximity to the EV charging station in column 7, and in the last column, we tested the change in coefficients by considering PT coverage.

A comparison between models 2 and 5 suggest that the correlation with being female becomes negative once the features of vehicles considered. This is a simple indication that the vehicle features are endogenous to the purpose of use and have a significant impact on VKT. The other demographic factors show a relatively persistent impact across all regressions.

The addition of (the logarithm of) distance to the EV charging station (log_ev) in column 6 does not change the sign and statistical significance of the demographic variables. However, the coefficient of EV fuel technology changes from positive to zero (it becomes statistically insignificant). This change indicates the importance of proximity to EV charging stations in the estimated higher VKT of EV vehicles compared to petrol cars. Our results suggest that the impact of proximity to an EV charger on VKT is positive, and this correlation decreases with (population) density.

The addition of PT coverage variables[39] in column 7 does not change demographic variables' sign and statistical significance. However, the sign of the correlation for EV vehicles changes back to positive. This implies that, if PT coverage was fixed across suburbs, an EV vehicle would be travelling more than a petrol vehicle, which reflects the mode shift between petrol cars and PT. The sign and statistical significance of PT coverage variables are negative as expected – improved PT coverage is associated with lower VKT.[40]

Further consideration of the interaction terms provided in the table yields a range of interesting results. For example, increased household income is associated with higher VKT for all fuel types.[41] The interaction of vehicle size and fuel type is also reported.

Our findings for the correlation of VKT with demographic and regional factors using the results of the most comprehensive equation (column 7) are as follows:

- Being female is associated with lower VKT.
- Europeans and Asians travel less (-ve) and Māori and Pacific people travel more (+ve).
- VKT decreases with household income but at a decreasing rate.
- VKT decreases with age.
- VKT is higher for households with more children (# under 15) and working-age adults (and these relationships are also decreasing).
- In general, educated people have higher VKT.
- An increase in the number of cars is associated with lower VKT (one additional car decreases VKT by 6%), and this relationship is decreasing
- As shown, as we increased the number of explanatory variables, the initial negative impacts in Otago, Nelson and Marlborough changed to positive. Based on the regression results in column 6, compared to Auckland, VKT is lower in most tier 3 urban environments (Gisborne, West Coast and Tasman). The highest VKT due to regional features is for Waikato, Bay of Plenty, Canterbury and Hawke's Bay.
- While population density initially has a negative correlation with VKT, the positive correlations in models 7 and 8 are driven by the interaction between distance to the EV charger as a measure of proximity. Hence, the lower VKT in the dense areas is driven by access to amenities and facilities.[42] We only report this as an observation, which requires further testing.
- As expected, VKT is positively correlated with travel time to the nearest town centre using PT, walking and cycling and negatively correlated with travel time using a car.

A comparison between our regression results and the descriptive statistics suggests the following:

- While descriptive statistics show a positive correlation between VKT and income, this correlation changes to negative after controlling for other demographic factors.
- While employment is positively correlated with VKT, after controlling for other demographic factors, the correlation is negative. The self-employed relationship remains positive, which might be due to some personal use actually being work use.

[39] Availability of PT with 15-minute frequency, availability of PT with 30-minute frequency, distance to bus stop (metres).

[40] Due to lack of information on PT coverage in a few regions, the number of observations in the last column decreases by 2,644,743. Our comparison of column 7 results with a regression of column 7 of the reduced sample population suggests that our interpretation of results is robust to this sample drop.

[41] There is a weak exception of diesel hybrid and petrol hybrid, which decreases with household income and could be due to the lower use of these vehicles as income increases.

[42] Given its implication for land use and transport policy, we suggest further exploring this correlation in a future study.

- VKT is higher for households located in the urban core and in areas outside of the urban area, but this relationship changes after we control for local features, particularly PT coverage and access to amenities.
- VKT has a significant negative correlation with PT coverage.

Our findings for the correlation of VKT with vehicle factors using the results of the most comprehensive equation (column 8) are as follows:

- Newer cars travel more, and the relationship is (non-linear and) decreasing.
- Station wagons have higher VKT than other body types, and convertibles have the lowest VKT.
- EVs travel more than petrol cars (with the caveats highlighted above).
- While diesel vehicles seem to travel more than petrol cars initially, the correlation becomes negative after controlling for demographic features.
- Larger cars travel more than other vehicle sizes.

An important finding is about the change in VKT over time (captured by the year dummies). As shown, the captured trend changes significantly as we add demographic, vehicle features, regional features and other factors to the equation. For example, the first equation, controlling only for demographics, shows that general VKT is lower in the COVID-19 years (2021 and 2022) compared to 2020 and earlier. However, after controlling for other factors, the trend changes to positive. Torshizian et al. (2025b) developed a COVID-19 index based on individual features to control for that impact. Hence, we suggest that VKT forecasts must consider the role of demographics and vehicle and regional features. A helpful approach is using a socio-economic modelling framework such as the RLTDM and modelling the VKT trend and cycle separately – a strategy developed for cost escalation modelling by Principal Economics (2024). The uncertainty with forecast could be further incorporated into modelling using a stochastic approach, similar to the RLTDM, and scenario testing.

Table 5.1 shows the summary results (RLT) and the findings from the literature (LR) for the VKT, vehicle ownership and EV uptake analyses. For details about the literature review for car ownership and EV uptake, refer to Appendix E. While we identify significant impacts on car ownership, the low goodness of fit suggests that a broader range of factors may affect the car ownership decision. This finding helps reduce car ownership policy decisions. We also tested the correlation between these factors and vehicle emissions, excluding VKT. The results suggested a very low goodness of fit (at around 4%). This finding may indicate that users do not consider vehicle emissions in their vehicle ownership decisions.

Table 5.1 Summary results from Appendix E

Categories	Variables	(1)	(2)	(3)	(4)	(5)	(6)
		VKT		Ownership		EV uptake	
		LR	RLT	LR	RLT	LR	RLT
Demographics	Sex (being male)		- →+		-	±	-
	Household with children	+	+	–	+	±	+
	Household size	+	+	+	+		+
	Household number of working age	+	+		-		0
	Education	+	+		+	+	+
	Being employed	+	+→-		-	+	0
	Being self-employed		+		+		0
	Household income	±	+→-	+	+	±	+→-

Categories	Variables	(1)	(2)	(3)	(4)	(5)	(6)
		VKT		Ownership		EV uptake	
		LR	RLT	LR	RLT	LR	RLT
	Age		-		+		
	Ethnicity (being Māori)		+		-		
Vehicle features	Household number of cars	+	-			±	-
	Vehicle age		-				
	Body type		+ station wagons				
	Electric vehicles		+				
	Diesel vehicles		+→-				
	Vehicle size		+	−			
Built environment characteristics	Population density	-	-		-		-
	Distance to market	+	+				
	Distance to town centre by car	+	+		-	-	-
	Distance to town centre by PT	±	+	−	+	±	0
	Distance to town centre by walking	+	+		+	±	+
	Distance to town centre by biking	±	+		+	±	-
	Land area mixed use	-	+ → -	−	+ → -		
	Distance to EV charger		+		-	-	-
	PT coverage		-		-		

Source: Principal Economics.

5.3 Improved proximity to EV chargers by 344 m is correlated with 0.07% increase in EVs

For EV uptake, a simple log-log regression was used to identify the impact of a percentage change in proximity to EV charging infrastructure on the likelihood of EVs across the sample population. The output is a dummy variable of a household having an EV (1) or not (0). We used logistic regression analysis at the granular household level. For this, we used the sampling weights with pooled cross-section, which, over the timeframe of 2017–2023, provided us with around 19 million weighted sample population. The results show the following (Table 5.2):

- A 1% improved proximity to an EV charger (344 m) increases the likelihood of EV ownership by 0.07%.
- This relationship decreases with density (and income, which is not in the table, but we tested separately). Being in a dense area is associated with a lower likelihood of EV ownership.
- The marginal impact of log-dens is (approximately) -0.05. This implies that a 1% higher population per square kilometre (equal to 19.1 people) is associated with a 0.05% decrease in the likelihood of EV ownership.

The metropolitan area is most likely to adopt EV across functional urban areas (FUAs). A wide range of questions remain about the role of omitted variables, the interaction with other predictors and the causality of this impact. However, this simple regression presents the best available estimate of the impact of proximity to EV infrastructure using a granular household-level analysis, which was not available before. Further, it illustrates the value of the established VKT dataset for this project.

Table 5.2 Regression of the factors of EV uptake

Variable	Coefficient
Log distance to EV charger (log_ev)	0.078**
	(0.002)
log_ev interaction with log_dens	-0.0093***
	(0.001)
log_dens	-0.047**
	(0.002)
FUA (base: Metropolitan)	
Outside of FUA	0.007
	(0.002)
Small centre	-0.023***
	(0.002)
Medium centre	-0.057***
	(0.001)
Large centre	0.023**
	(0.001)
_cons	-1598.3***
R squared	0.287
Number of observations	11,086,350

Source: Principal Economics.
Note: Significance codes: *** p<0.01, ** p<0.05, * p<0.1.

5.4 Impacts on high and low VKT are similar to general VKT

We also tested the impact of demographics on high and low VKT. This was based on a logistic regression of a dummy output variable on demographic characteristics. The output was defined at the regional level for each year. We tested the highest 10th (90th), 20th (80th) and 30th (70th) for the high (low) dummy variables. Results show that the sign and statistical significance of the demographic factors are like their impacts on general VKT. We did not further control for the effects of regions and vehicles in this analysis due to the computational limits using logistic regression – given the large size of our dataset, it is possible to simply use OLS, but we did not pursue that further.

6 We used synthpop to release IDI data for improved access to VKT data

We used the synthpop R package to generate a synthetic unit-record file of annual VKT for individuals and their main vehicle relationship that preserves the statistical properties of our original. We use the synthpop implementation of classification and regression tree (CART) modelling to generate the synthetic data to closely mimic the relationships and distributions found in the original dataset. The CART modelling technique creates decision trees that split the data into smaller, homogeneous groups based on the values of predictor variables, handling both numeric and categorical data to replicate the relationships and distributions in the original dataset (Nowok et al., 2016).

To prepare the confidential dataset for synthesis, we first subset the data to variables relevant to our research goals. This step aligns outcomes with our objectives and makes the dataset more manageable, reducing the time required for synthesis. Numeric variables, including household income, vehicle year and VKT, were winsorised[43] prior to synthesis to improve outputs and maintain confidentiality requirements. Aggregation of variables, including household income and age groups into interval brackets, was done post-synthesis. We show variables included in the synthesised dataset in Table 6.1. To reduce processing time, we segment the dataset into regional groups before applying the synthpop CART modelling and data synthesis to each segment.

Table 6.1 Variables in the synthesised dataset

Variable	Description of variable
year	Year of observation
sex	Sex of individuals (1 for male, 0 for female)
age	Age of individuals (5-year intervals)
u15_nbr	Number of usual residents aged under 15 in household
o15_nbr	Number of usual residents aged over 15 in household
hhld_size	Household size
hhd_inc	Household income[44]
region_code	Region code
hst_qual	Highest qualification – NZQCF qualification level
eth1	Ethnic group 1 (European, 1 if true)
eth2	Ethnic group 2 (Māori, 1 if true)
eth3	Ethnic group 3 (Pacific Peoples, 1 if true)
eth4	Ethnic group 4 (Asian, 1 if true)
vkt_span	Annual VKT-specific individual and vehicle
vehyear	Year of vehicle manufacture
bodytype	Vehicle body type (hatchback, SUV, convertible, etc.)
veh_ccpwr	Vehicle engine size (XS, S, M, L, XL)
veh_tech	Vehicle fuel type

Source: Principal Economics.

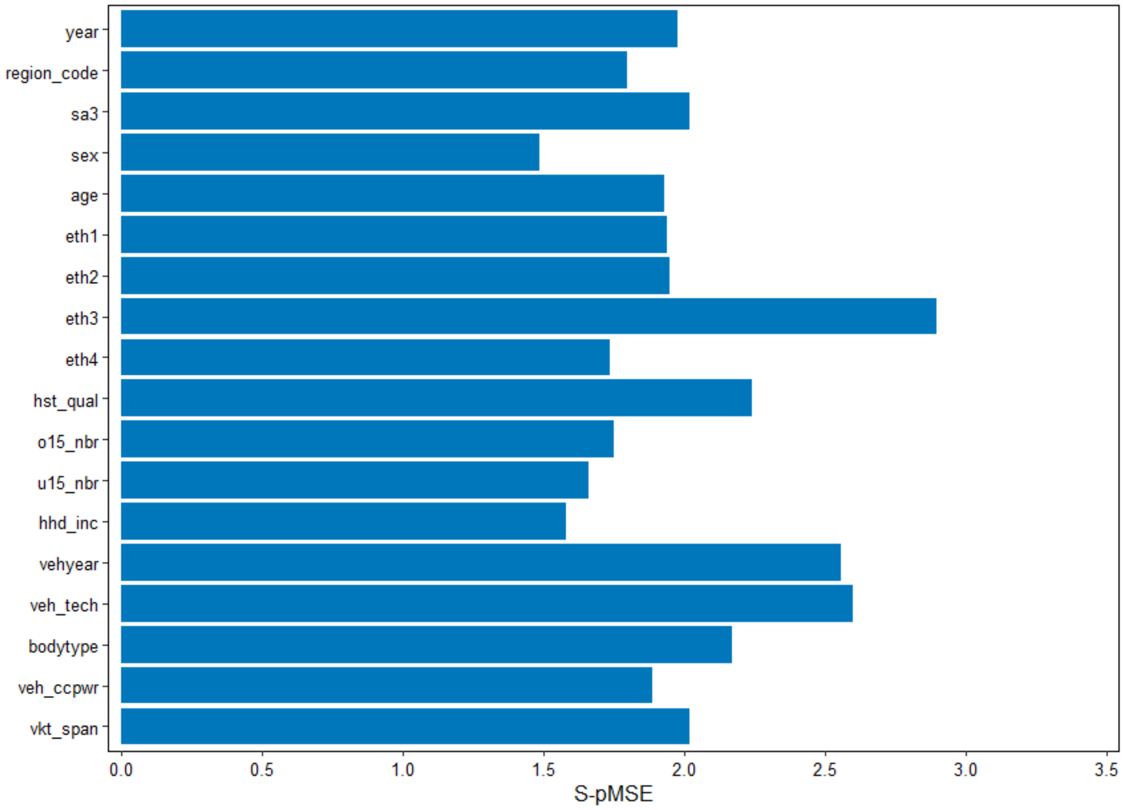
[43] Extreme outliers replaced with the smallest and largest values of a dataset.

[44] Grouped to income brackets from -$5,000 to $500,000 and over in $5,000 intervals.

To verify the validity of the synthetic data, we used the propensity score estimate (pMSE) ratio and standardised pMSE, which adjusted the statistic based on its expected value and standard deviation under the null. These measures test whether the model used to generate the synthetic data accurately reflects the original data (Snoke et al., 2018). The ideal utility ratio target is 1.0, indicating that the synthetic data perfectly matches the statistical properties of the actual data. However, as suggested by Raab et al. (2021), a good rule of thumb is to aim for utility ratios below 10, which provides a practical benchmark for ensuring the data is useful and valid without perfectly matching the actual data.

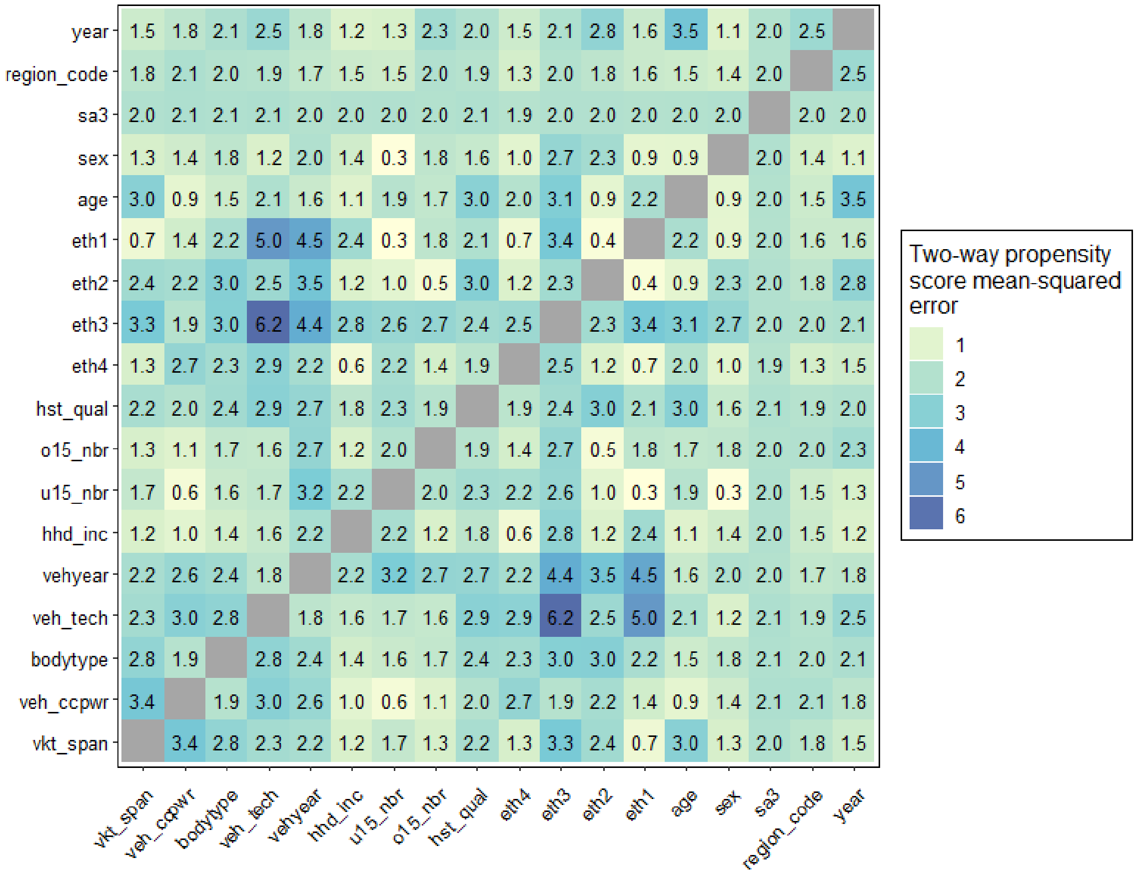
A more practical and effective approach is to evaluate the agreement between the synthetic and original data, suggested by analysing low-order margins, beginning with one-way marginals and then progressing to two-way and possibly three-way marginals. Figure 6.1 shows the results of the S_pMSE utility measure for one-way marginals are being below 3, indicating high utility for analysis.

Figure 6.1 S-pMSE for one-way marginals



Additionally, we determine the two-way marginals as shown in Figure 6.2, with the highest two-way pMSE ratio being 6.2 between the variables eth3 – Pacific Peoples and veh_tech – fuel type well below the recommend score in the literature. This suggests that the synthetic data is likely to reasonably replicate the relationships between variables in the actual data.

Figure 6.2 Two-way pMSE ratios



Using the extracted data, we established a publicly accessible VKT demographics dashboard, available at https://principaleconomics.com/vkt-dashboard/.

7 Conclusion, limitation and future research

This research project identified sources of data and available methods for establishing a granular dataset of VKT using household features, vehicle features and regional context. Then we applied the identified methods to address the complexities of merging the required datasets using the IDI. After that, we generated descriptive statistics and applied a range of regression analyses to investigate the role of demographics, vehicle features and regional context on VKT. In the end, we developed a VKT dataset that is publicly available for future research. Our VKT demographics dashboard extracted from the dataset is available at https://principaleconomics.com/vkt-dashboard/.

For demographic factors, we investigated the VKT impact of sex (being male), household with children, household size, household number of working age, education, being employed, being self-employed, household income, age and ethnicity.

Our identified key demographic factors are as follows:

- All demographic factors explained VKT (statistically) significantly.
- Being female is associated with lower VKT.
- Europeans and Asians travel less (-ve) and Māori and Pacific people travel more (+ve).
- VKT decreases with household income but at a decreasing rate.
- VKT decreases with age.
- VKT is higher for households with more children (# under 15) and working-age adults (and these relationships are also decreasing).
- In general, educated people have higher VKT.
- An increase in the number of cars is associated with lower VKT (one additional car decreases VKT by 6%), and this relationship is decreasing.
- As shown, as we increased the number of explanatory variables, the initial negative impacts in Otago, Nelson and Marlborough changed to positive. Based on the regression results in column 6 of Table D.1 in Appendix D, compared to Auckland, VKT is lower in most tier 3 urban environments (Gisborne, West Coast and Tasman). The highest VKT due to regional features is for Waikato, Bay of Plenty, Canterbury and Hawke's Bay.
- Population density has a negative correlation with VKT. The lower VKT in the dense areas is driven by access to amenities and facilities.
- VKT is positively correlated with travel time to the nearest town centre using PT, walking and cycling and negatively correlated with travel time using a car.

In addition to interesting findings on the correlates of VKT, a comparison between our regression results and the descriptive statistics suggests the following:

- While descriptive statistics show a positive correlation between VKT and income, this correlation changes to negative after controlling for other demographic factors.
- While employment is positively correlated with VKT, after controlling for other demographic factors, the correlation is negative.
- VKT is higher for households located in the urban core and in areas outside of the urban area, but this relationship changes after we control for local features, particularly PT coverage and access to amenities.
- VKT has a significant negative correlation with PT coverage.

Our comparison between the explanatory power of demographics, vehicle and spatial features and the features of public and road transport networks suggests that vehicle features explain VKT most significantly.

Our results suggest that demographic features and other factors are important to VKT. The question is, what is the purpose of VKT analysis? If the purpose is monitoring VKT and a small set of variables is desired, our results suggest that fleet composition provides a better prediction of VKT. If the purpose is to model VKT in detail, we identify significant correlations with demographic factors, vehicle and spatial features and the features of public and road transport networks. If the purpose is to assess policy initiatives (ex-post or ex-ante, all factors matter, as well as further interaction of policy-targeted features of VKT with other factors.

We also explored the impact of demographics and other factors on car ownership

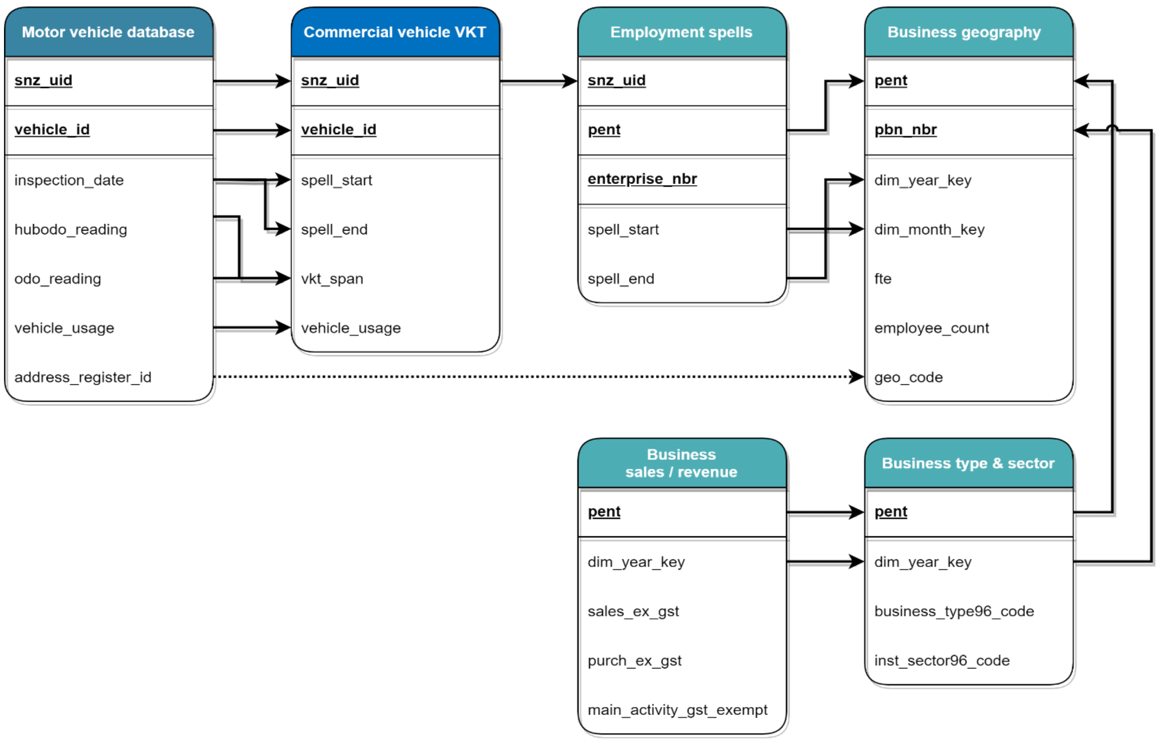
Our results suggested that, while important, demographic factors do not explain car ownership significantly. The low goodness of fit of the demographic factors in the vehicle ownership equation suggests that a broader range of factors may affect the car ownership decision. This finding helps reduce car ownership policy decisions. We also tested the correlation between these factors and vehicle emissions, excluding VKT. The results suggested a very low goodness of fit (at around 4%). This finding indicates that users do not consider vehicle emissions in their vehicle ownership decisions.

Opportunities for future research

The established data and methods in this analysis provide opportunities for future research of the patterns of VKT, fleet composition and household behaviour. A few of the essential topics are listed below:

- Using a similar approach to establish a VKT dataset for freight will provide an invaluable data source for freight policy. Figure 7.1 illustrates how the linkage between the Stats NZ's Longitudinal Business Database (LBD) and other datasets could be established.

Figure 7.1 Suggested use of LBD to establish freight VKT



Source: Principal Economics.

- Further analysis of the endogeneity of the factors of VKT. This report used a large dataset to establish the correlations between VKT and a wide range of factors. The large size of the dataset and the persistence of the parameters across regression models indicate the robustness of the established correlations. Further analysis will be required to understand the impact of each identified factor. For example, a further assessment of the correlation between proximity to EV charging stations on VKT suggested that there is a significant role for density, household income and location features. The established dataset provides the information needed for future research on different aspects of VKT.
- A granular analysis of own and cross-VKT price elasticities is unavailable in New Zealand. This would be particularly useful for understanding the impact of pricing policies such as tolling.
- Further assessment of the drivers of EV uptake. This report provided initial results to inform EV uptake policies. A future report should focus on specific hypotheses and consider various policy levers.
- Our discussions with stakeholders suggest that the EV charging station location dataset available from NZTA needs to be updated. This requires further consideration in a future research project.
- Further analysis of the correlates of vehicle ownership will be helpful to complement the VKT analysis.
- Improving forecasts of VKT and fleet composition by linking the established dataset with an advanced socio-economic model such as the RLTDM.
- The current NZTA data on the location of EV charging stations is outdated and does not include all the charging stations. While our large sample size improves the accuracy of our estimates, if the unobserved EV charging stations are systematically located, our results will be inconsistent. We suggest further exploring this issue in the future.

Overall, this research project significantly improves our knowledge base of VKT demographics and data availability. Further exploration of a wide range of policy and technical implications is recommended.

References

Anastasiadis, S., Chokkanathapuram, A., & Wright, C. (2020). *Constructing households from linked administrative data: An attempt to improve address information in the IDI*. Social Wellbeing Agency | Toi Hau Tāngata. https://www.sia.govt.nz/assets/Document-Library/Constructing-households-from-linked-admin-data.pdf

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Springer Science & Business Media.

ARUP. (2020). *Building an Agent Based Model for New Zealand: Stage 2 report*. Ministry of Transport. https://www.transport.govt.nz/assets/Uploads/EvaluationWorkProgramme_Part1.pdf

Aslanyan, T., & Jiang, S. (2021). Examining passenger vehicle miles traveled and carbon emissions in the Boston metropolitan area. In S. C. M. Geertman, C. Pettit, R. Goodspeed & A. Staffans (Eds.), *Urban informatics and future cities* (pp. 319–340). Springer International Publishing. https://doi.org/10.1007/978-3-030-76059-5_17

Atkinson, J., Salmond, C., & Crampton, P. (2020). *NZDep2018 Index of Deprivation*. University of Otago. https://www.otago.ac.nz/__data/assets/pdf_file/0020/326711/nzdep2018-index-of-deprivation-research-report-final-dec-2020-823833.pdf

Axsen, J., & Kurani, K. S. (2013). Hybrid, plug-in hybrid, or electric – What do car buyers want? *Energy Policy*, *61*, 532–543. https://doi.org/10.1016/j.enpol.2013.05.122

Azevedo, C. L., & Cardoso, J. (2009, June 3–5). *Estimation of annual traffic volumes: A model for Portugal* (Paper presentation). Young Researchers Seminar, Torino, Italy. https://www.ectri.org/wp-content/uploads/2019/01/Azevedo_C_Cardoso_J_Session1_Traffic1.pdf

Ball, S. D., Emmerson, P., Anable, J., Cairns, S., Wilson, R. E., & Chatterton, T. (2016). *Understanding variation in car use: Exploration of statistical metrics at differing spatial scales using data from every private car registered in Great Britain*. UK Engineering and Physical Sciences Research Council.

Bansal, P., Kockelman, K. M., Schievelbein, W., & Schauer-West, S. (2018). Indian vehicle ownership and travel behavior: A case study of Bengaluru, Delhi and Kolkata. *Research in Transportation Economics*, *71*, 2–8. https://doi.org/10.1016/j.retrec.2018.07.025

Barnes, J., & Bailey, R. (2014). *Quantitative and qualitative assessment of the South Yorkshire ECO Stars fleet recognition scheme*. Barnsley Metropolitan Borough Council. https://uwe-repository.worktribe.com/output/817590

Bastian, A., Börjesson, M., & Eliasson, J. (2016). Explaining “peak car” with economic variables. *Transportation Research Part A: Policy and Practice*, *88*, 236–250. https://doi.org/10.1016/j.tra.2016.04.005

Beca Carter Hollings & Ferner & Sinclair Knight Merz. (2004). *Wellington Transport Strategy Model: Validation report*. Greater Wellington Regional Council. https://wellingtontransportanalytics.co.nz/wp-content/uploads/2022/03/WTSM-2001-TN221-ValidationReport.pdf

Bhat, C. R., Goulias, K. G., Pendyala, R. M., Paleti, R., Sidharthan, R., Schmitt, L., & Hu, H. (2012, January 22–26). *A household-level activity pattern generation model for the Simulator of Activities, Greenhouse Emissions, Networks, and Travel (SimAGENT) system in Southern California* (Paper presentation). Transportation Research Board 91st Annual Meeting, Washington, DC. https://scag.ca.gov/sites/default/files/old/file-attachments/intrahouseholdinteractions_trb_1aug11.pdf?1605574685=

Bureau of Infrastructure, Transport and Regional Economics. (2011). *Road vehicle kilometres travelled: Estimation from state and territory fuel sales*. https://www.bitre.gov.au/sites/default/files/report_124.pdf

Cairns, S., Wilson, R., Chatterton, T., Anable, J., Notley, S., & McLeod, F. (2014). *Using MOT test data to analyse travel behaviour change – scoping report*. Transport Research Laboratory. https://www.trl.co.uk/uploads/trl/documents/PPR578%20FINAL.pdf

Charypar, D., & Nagel, K. (2005). Generating complete all-day activity plans with genetic algorithms. *Transportation*, *32*(4), 369–397. https://doi.org/10.1007/s11116-004-8287-y

Chatterton, T., Barnes, J., Wilson, R. E., Anable, J., & Cairns, S. (2015). Use of a novel dataset to explore spatial and social variations in car type, size, usage and emissions. *Transportation Research Part D: Transport and Environment*, *39*, 151–164. https://doi.org/10.1016/j.trd.2015.06.003

Cornelis, G. (2015). *2013 WTSM Update: Technical Note 5: Model input parameters*. Greater Wellington Regional Council. https://www.gw.govt.nz/assets/Documents/2021/10/TN5-Model-Input-Parameters.pdf

Curl, A., Watkins, A., McKerchar, C., Exeter, D., & Macmillan, A. (2020). *Social impact assessment of mode shift* (Research report 666). Waka Kotahi NZ Transport Agency. https://www.nzta.govt.nz/assets/resources/research/reports/666/666-Social-impact-assessment-of-mode-shift.pdf

Currie, G., Jain, T., & Aston, L. (2021). Evidence of a post-COVID change in travel behaviour – Self-reported expectations of commuting in Melbourne. *Transportation Research Part A: Policy and Practice*, *153*, 218–234. https://doi.org/10.1016/j.tra.2021.09.009

Das, S., Boruah, A., Banerjee, A., Raoniar, R., Nama, S., & Maurya, A. K. (2021). Impact of COVID-19: A radical modal shift from public to private transport mode. *Transport Policy*, *109*, 1–11. https://doi.org/10.1016/j.tranpol.2021.05.005

Department for Transport. (2013). *Experimental statistics: Analysis of vehicle odometer readings recorded at MOT tests*. https://assets.publishing.service.gov.uk/media/5a7c729c40f0b626628ac212/experimental-statistics-mot-data.pdf

Department for Transport. (2020). *National Transport Model – analytical review*. https://assets.publishing.service.gov.uk/media/5fdc9ff3e90e0745298be3cd/DfT-National-Transport-Model-Analytical-Review-accessible.pdf

Diao, M. I., & Ferreira, J. (2014). Vehicle miles traveled and the built environment: Evidence from vehicle safety inspection data. *Environment and Planning A: Economy and Space*, *46*(12), 2991–3009. https://doi.org/10.1068/a140039p

Dieleman, F. M., Dijst, M., & Burghouwt, G. (2002). Urban form and travel behaviour: Micro-level household attributes and residential context. *Urban Studies*, *39*(3), 507–527. https://doi.org/10.1080/00420980220112

Exeter, D. J., Zhao, J., Crengle, S., Lee, A., & Browne, M. (2017). The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health research in Aotearoa, New Zealand. *PLoS ONE*, 12(8), e0181260. https://doi.org/10.1371/journal.pone.0181260

Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Fabling, R., & Maré, D. C. (2020). *Measuring commute patterns over time: Using administrative data to identify where employees live and work* (Motu working paper 20-05). Motu Economic and Public Policy Research. https://motu-www.motu.org.nz/wpapers/20_05.pdf

Federal Highway Administration. (2023). *Travel survey state of the practice*. https://nhts.ornl.gov/assets/NextGen%20NHTS_State%20of%20Practice_032423.pdf

Gath, M., & Bycroft, C. (2018). *Potential for linked admin data to provide household and family information*. Stats NZ. https://www.stats.govt.nz/assets/Uploads/Reports/The-potential-for-linked-administrative-data-to-provide-household-and-family-information/the-potential-linked-admin-data-household-family-information.pdf

Giuliano, G., & Dargay, J. (2006). Car ownership, travel and land use: A comparison of the US and Great Britain. *Transportation Research Part A: Policy and Practice*, *40*(2), 106–124. https://doi.org/10.1016/j.tra.2005.03.002

Glensor, K. (2021, June 3). *NZ MoT VKT and fleet estimation* (Presentation). Innovations in Applied Data Symposium, Wellington. https://terourou.org/symposium/speakers/kain_glensor.pptx

Grimal, R., Collet, R., & Madre, J.-L. (2013). Is the stagnation of individual car travel a general phenomenon in France? A time-series analysis by zone of residence and standard of living. *Transport Reviews*, *33*(3), 291–309. https://doi.org/10.1080/01441647.2013.801930

Hamadneh, J., & Esztergar-Kiss, D. (2022). Evaluation of the impacts of autonomous vehicles on the mobility of user groups by using agent-based simulation. *Transport*, *37*(1), 1–16. https://doi.org/10.3846/transport.2022.16322

Hardman, S., & Tal, G. (2021). Understanding discontinuance among California's electric vehicle owners. *Nature Energy*, *6*(5), 538–545. https://doi.org/10.1038/s41560-021-00814-9

Hasan, S., & Şimşekoğlu, Ö. (2020). The role of psychological factors on vehicle kilometer travelled (VKT) for battery electric vehicle (BEV) users. *Research in Transportation Economics*, *82*, 100880. https://doi.org/10.1016/j.retrec.2020.100880

Headicar, P. (2013). The changing spatial distribution of the population in England: Its nature and significance for 'Peak Car'. *Transport Reviews*, *33*(3), 310–324. https://doi.org/10.1080/01441647.2013.802751

Hastings, K. (2016). *Prototype development for four local authorities*. Waka Kotahi NZ Transport Agency. https://www.nzta.govt.nz/assets/Highways-Information-Portal/Technical-disciplines/Air-quality/Planning-and-assessment/Vehicle-emissions-mapping-tool/Local-Authority-Vehicle-Emission-GIS-Mapping-Prototype-Development-for-Four-Local-Authorities.pdf

Hastings, K. (2018). *National Vehicle Emission Dataset 2016*. Waka Kotahi NZ Transport Agency. https://www.nzta.govt.nz/assets/Highways-Information-Portal/Technical-disciplines/Air-quality/Planning-and-assessment/Vehicle-emissions-mapping-tool/National-Vehicle-Emission-GIS-Mapping-National-Vehicle-Emission-Dataset-2016-Final-Revised-Sep-2018.pdf

Holtzclaw, J. (1994). *Using residential patterns and transit to decrease auto dependence and costs*. Natural Resources Defense Council. https://www.nrdc.org/sites/default/files/sma_09121401a.pdf

Holtzclaw, J., Clear, R., Dittmar, H., Goldstein, D., & Haas, P. (2002). Location efficiency: Neighborhood and socio-economic characteristics determine auto ownership and use – Studies in Chicago, Los Angeles and San Francisco. *Transportation Planning and Technology*, *25*(1), 1–27. https://doi.org/10.1080/03081060290032033

Horni, A., Nagel, K., & Axhausen, K. W. (2016). *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press. https://doi.org/10.5334/baw

Hou, C., Wang, H., & Ouyang, M. (2013). Survey of daily vehicle travel distance and impact factors in Beijing. *IFAC Proceedings Volumes*, *46*(21), 35–40. https://doi.org/10.3182/20130904-4-JP-2042.00002

International Transport Forum. (2022). *Broadening transport appraisal: Summary and conclusions* (Roundtable report 188). OECD Publishing. https://www.itf-oecd.org/sites/default/files/docs/broadening-transport-appraisal.pdf

Javid, R. J., & Nejat, A. (2017). A comprehensive model of regional electric vehicle adoption and penetration. *Transport Policy*, *54*, 30–42. https://doi.org/10.1016/j.tranpol.2016.11.003

Jung, S., Kim, J., Kim, J., Hong, D., & Park, D. (2017). An estimation of vehicle kilometer traveled and on-road emissions using the traffic volume and travel speed on road links in Incheon City. *Journal of Environmental Sciences*, *54*, 90–100. https://doi.org/10.1016/j.jes.2015.12.040

Kasraian, D., Raghav, S., Yusuf, B., & Miller, E. J. (2022). A longitudinal analysis of travel demand and its determinants in the Greater Toronto-Hamilton Area. *Environment and Planning B: Urban Analytics and City Science*, *49*(8), 2230–2249. https://doi.org/10.1177/23998083221082109

Kirley, B., Wang, Y., Foss, R., Harrell, S., & Goodwin, A. (2022). *Estimating motorcycle miles traveled from state vehicle inspection records*. Department of Transportation. https://rosap.ntl.bts.gov/view/dot/62082

Kockelman, K. M. (1997). Travel behavior as function of accessibility, land use mixing, and land use balance: Evidence from San Francisco Bay Area. *Transportation Research Record: Journal of the Transportation Research Board*, *1607*(1), 116–125. https://doi.org/10.3141/1607-16

KPMG & ARUP. (2017). *Melbourne Activity-Based Model calibration and validation report*. Infrastructure Victoria.

Kumapley, R., & Fricker, J. (1994). *Estimating statewide vehicle miles traveled in Indiana*. Purdue University. https://doi.org/10.5703/1288284314211

Kuschel, G., Metcalfe, J., Sridhar, S., Davy, P., Hastings, K., Mason, K., Denne, T., Woodward, A., Hales, S., Atkinson, J., Berentson-Shaw, J., & Bell, S. (2022). *Health and air pollution in New Zealand 2016 (HAPINZ 3.0): Volume 2: Detailed methodology*. Ministry for the Environment, Ministry of Health, Te Manatū Waka Ministry of Transport & Waka Kotahi NZ Transport Agency. https://environment.govt.nz/publications/health-and-air-pollution-in-new-zealand-2016-detailed-methodology/

Kweon, Y.-J., & Kockelman, K. M. (2004, January 11–15). *Nonparametric regression estimation of household VMT* (Paper presentation). 83rd Annual Meeting of the Transportation Research Board, Washington, DC. https://www.caee.utexas.edu/prof/kockelman/public_html/TRB04VMT.pdf

Litman, T. (2013). The new transportation planning paradigm. *ITE Journal*, *83*(6), 20–28. https://www.vtpi.org/paradigm.pdf

Litman, T. (2024). *Evaluating accessibility for transport planning: Measuring people's ability to reach desired services and activities*. Victoria Transport Policy Institute. https://www.vtpi.org/access.pdf

Majid, M. R., Nordin, A. N., & Medugu, I. N. (2014). Influence of housing development designs on household vehicle miles traveled: A case of Iskandar Malaysia. *Transportation Research Part D: Transport and Environment*, *33*, 63–73. https://doi.org/10.1016/j.trd.2014.09.001

Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., & Roser, M. (2020). *COVID-19 pandemic*. https://ourworldindata.org/coronavirus

McKibbin, B., Sewell, M., & Li, V. (2022). *VKT and GHG emissions baseline report* (Research note 008). Waka Kotahi NZ Transport Agency. https://www.nzta.govt.nz/assets/resources/research/research-notes/008/008-VKT-and-GHG-emissions-baseline-report.pdf

Metcalfe, J., & Peeters, S. (2022). *Vehicle Emissions Prediction Model: VEPM 6.3 update technical report*. Waka Kotahi NZ Transport Agency. https://www.nzta.govt.nz/assets/Highways-Information-Portal/Technical-disciplines/Air-quality/Planning-and-assessment/Vehicle-emissions-prediction-model/VEPM-6.3-technical-report-2022.pdf

Metcalfe, J., & Peeters, S. (2023). *Vehicle Emissions Prediction Model: VEPM 7.0 technical report*. Waka Kotahi NZ Transport Agency. https://www.nzta.govt.nz/assets/Highways-Information-Portal/Technical-disciplines/Air-quality/Planning-and-assessment/Vehicle-emissions-prediction-model/VEPM-7.0-technical-report-FINAL.pdf

Ministry of Transport. (2017). *New Zealand Transport Outlook: Household Travel Model SAS programmes*. https://www.transport.govt.nz/assets/Uploads/Data/Transport-outlook-updated/Household-Travel-Model-SAS-Version-2-Documentation-20171127.pdf

Ministry of Transport. (2019). *Background discussion of regional vehicle-kilometres travelled (VKT) data in New Zealand*. https://www.transport.govt.nz/assets/Uploads/Discussion/Background-Discussion-of-VKT-Data-20190716.pdf

Ministry of Transport. (2022a). *New Zealand Household Travel Survey 2018–2020: Methodology report*. https://www.transport.govt.nz/assets/Uploads/Research/NZHTS_Methodology_Report_2018_2020-v2.pdf

Ministry of Transport. (2022b). *Project Monty*.

Ministry of Transport. (2022c). *Vehicle Fleet Emissions Model (Version 3.2)*. https://www.transport.govt.nz/assets/Uploads/Data/Transport-outlook-updated/Vehicle-Fleet-Emissions-Model-Documentation-20220608.pdf

Nagel, K., & Flötteröd, G. (2012). Agent-based traffic assignment: Going from trips to behavioural travelers. In R. Pendyala & C. Bhat (Eds.), *Travel behaviour research in an evolving world: Selected papers from the 12th International Conference on Travel Behaviour Research* (pp. 261–294). Lulu.

Narváez-Villa, P., Arenas-Ramírez, B., Mira, J., & Aparicio-Izquierdo, F. (2021). Analysis and prediction of vehicle kilometers traveled: A case study in Spain. *International Journal of Environmental Research and Public Health*, *18*(16), 8327. https://doi.org/10.3390/ijerph18168327

Nasri, A., Zhang, L., Fan, J., Stewart, K., Younes, H., Fu, C., & Jessberger, S. (2019). Advanced vehicle miles traveled estimation methods for non-federal aid system roadways using GPS vehicle trajectory data and statistical power analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(11), 296–308. https://doi.org/10.1177/0361198119850790

Nowok, B., Raab, G. M., & Dibben, C. (2016). synthpop: Bespoke creation of synthetic data in R. *Journal of Statistical Software*, *74*(11). https://doi.org/10.18637/jss.v074.i11

NZ Transport Agency Waka Kotahi. (2025). *Warrant of fitness.* http://nzta.govt.nz/vehicles/warrants-and-certificates/warrant-of-fitness

Ostermeijer, F., Koster, H., & van Ommeren, J. (2019). Residential parking costs and car ownership: Implications for parking policy and automated vehicles. *Regional Science and Urban Economics*, *77*, 276–288. https://doi.org/10.1016/j.regsciurbeco.2019.05.005

Park, S., & Ko, D. (2021). Spatial regression modeling approach for assessing the spatial variation of air pollutants. *Atmosphere*, *12*(6), 785. https://doi.org/10.3390/atmos12060785

Polzin, S. E. (2006). *The case for moderate growth in vehicle miles of travel: A critical juncture in U.S. travel behavior trends*. Department of Transportation. https://www.cutr.usf.edu/oldpubs/The%20Case%20for%20Moderate%20Growth%20in%20VMT-%202006%20Final.pdf

Potoglou, D., & Kanaroglou, P. S. (2007). Household demand and willingness to pay for clean vehicles. *Transportation Research Part D: Transport and Environment*, *12*(4), 264–274. https://doi.org/10.1016/j.trd.2007.03.001

Potoglou, D., & Kanaroglou, P. S. (2008). Modelling car ownership in urban areas: A case study of Hamilton, Canada. *Journal of Transport Geography*, *16*(1), 42–54. https://doi.org/10.1016/j.jtrangeo.2007.01.006

Principal Economics. (2022). *Emission reduction initiatives: A review of policy levers for reducing transport GHG emissions*. https://www.principaleconomics.com/emission_reduction_initiatives/index.html

Principal Economics. (2024). *Sector-specific input price escalation forecasts for electricity distribution businesses (EDBs)*. https://principaleconomics.com/reports/sector-specific-input-price-escalation-forecasts-for-electricity-distribution-businesses-edbs/

Raab, G., Nowok, B., & Dibben, C. (2021, December 1–3). *Assessing, visualizing and improving the utility of synthetic data* (Presentation). Expert Meeting on Statistical Data Confidentiality, Poland. https://unece.org/sites/default/files/2021-12/SDC2021_Day2_Raab_AD.pdf

Raney, B., & Nagel, K. (2006). An improved framework for large-scale multi-agent simulations of travel behaviour. In B. Jourquin, P. Rietveld & K. Westin (Eds.), *Towards better performing transport networks*. Routledge (pp. 305–347). https://doi.org/10.4324/9780203965573

Rendall, S., Page, S., & Krumdieck, S. (2013). *Local Area Transport Energy Evaluation (LATEE): New Zealand warrant of fitness data and VKT analysis mapped to census unit areas method description and validation*. University of Canterbury. https://ir.canterbury.ac.nz/items/7d5155ea-c34a-4fbd-a96c-2d8c7fd156df

Rhee, K.-A., Kim, J.-K., Lee, Y., & Ulfarsson, G. F. (2016). Spatial regression analysis of traffic crashes in Seoul. *Accident Analysis & Prevention*, *91*, 190–199. https://doi.org/10.1016/j.aap.2016.02.023

Sabouri, S., Tian, G., Ewing, R., Park, K., & Greene, W. (2021). The built environment and vehicle ownership modeling: Evidence from 32 diverse regions in the U.S. *Journal of Transport Geography*, *93*, 103073. https://doi.org/10.1016/j.jtrangeo.2021.103073

Safi, H., Assemi, B., Mesbah, M., Ferreira, L., & Hickman, M. (2019). Design and implementation of a Smartphone-Based Travel Survey. *Transportation Research Record: Journal of the Transportation Research Board*, *2526*(1), 99–107. https://doi.org/10.3141/2526-11

Sinclair Knight Merz. (2008). *Auckland Transport Models Project (ATM2): ART3 Distribution and mode choice report*.

Singh, A. C., Astroza, S., Garikapati, V. M., Pendyala, R. M., Bhat, C. R., & Mokhtarian, P. L. (2018). Quantifying the relative contribution of factors to household vehicle miles of travel. *Transportation Research Part D: Transport and Environment*, *63*, 23–36. https://doi.org/10.1016/j.trd.2018.04.004

Smith, G., & Bevan, J. (2011, March 27–30). *The Waikato Regional Transportation Model* (Paper presentation). IPENZ Transportation Group Conference, Auckland. http://transportationgroup.nz/wp-content/uploads/conference/2011/Papers/Smith_Grant%3B%20Bevan_James_The%20Waikato%20Regional%20Transportation%20Model.pdf

Snoke, J., Raab, G. M., Nowok, B., Dibben, C., & Slavkovic, A. (2018). General and specific utility measures for synthetic data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *181*(3), 663–688. https://doi.org/10.1111/rssa.12358

Song, J., & Wang, R. (2017). Measuring the spatial dimension of automobile ownership and its associations with household characteristics and land use patterns: A case study in three counties, South Florida (USA). *Sustainability*, *9*(4), 558. https://doi.org/10.3390/su9040558

Stats NZ. (2017). *Experimental population estimates from linked administrative data: 2017 release*. https://www.stats.govt.nz/research/experimental-population-estimates-from-linked-administrative-data-2017-release/

Stats NZ. (2022). *Experimental administrative population census: Data sources, methods, and quality (second iteration)*. https://www.stats.govt.nz/research/experimental-administrative-population-census-data-sources-methods-and-quality-second-iteration/

Stats NZ. (2023a). *Experimental administrative population census (third iteration): Data sources, methods, and quality for household information*. https://www.stats.govt.nz/research/experimental-administrative-population-census-third-iteration-data-sources-methods-and-quality-for-household-information/

Stats NZ. (2023b). *Experimental administrative population census (third iteration): Information by variable*. https://www.stats.govt.nz/research/experimental-administrative-population-census-third-iteration-information-by-variable/

Stephenson, J. (2016). *Regionalisation of the National Land Transport Demand Model* (Research report 586). NZ Transport Agency. https://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* (Research report 520). NZ Transport Agency. https://www.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* (Research report 646). NZ Transport Agency. https://www.nzta.govt.nz/assets/resources/research/reports/646/646-Impacts-of-socio-demographic-changes-on-the-NZ-land-transport-system.pdf

Sun, X., Wilmot, C., & Kasturi, T. (1998). Household travel, household characteristics, and land use: An empirical study from the 1994 Portland Activity-Based Travel Survey. *Transportation Research Record: Journal of the Transportation Research Board*, *1617*(1), 10–17. https://doi.org/10.3141/1617-02

Task Force on Road Traffic. (2007). *Handbook on statistics on road traffic: Methodology and experience*. United Nations Economic Commission for Europe Transport Division. https://unece.org/DAM/trans/doc/2007/wp6/handbook_final.pdf

Tonkin & Taylor Ltd. (2021). *Vehicle Emissions Mapping Tool: Summary report*. Waka Kotahi NZ Transport Agency. https://www.ehinz.ac.nz/assets/HAPINZ3/TT_VEMT-draft-summary-report-v2.pdf

Torshizian, E. (2017). *Effects of crowding, density and deprivation on residential satisfaction* [Doctoral dissertation]. University of Auckland. https://researchspace.auckland.ac.nz/server/api/core/bitstreams/44dac1c5-02e6-43bf-aa00-ced2ad9ab7b9/content

Torshizian, E., & Fehling, A. (2024). *Regional Land Transport Demand Model: Technical notes for practitioners* (Research note 012). NZ Transport Agency Waka Kotahi. https://www.nzta.govt.nz/assets/resources/research/research-notes/012/012-regional-land-transport-demand-model-technical-notes-for-practitioners.pdf

Torshizian, E., & Isack, E. (2020). *Fair fare for public transport*. Auckland Transport.

Torshizian, E., & Meade, R. (2020). *Estimating demand for competition analysis: A statistical exploration, and some possible applications*. Productivity Hub. https://www.nzier.org.nz/publications/estimating-demand-for-competition-analysis-a-statistical-exploration-and-some-possible-applications

Torshizian, E., Byett, A., Isack, E., Fehling, A., & Maralani, M. (2022). *Incorporating distributional impacts (equity) in the cost-benefit appraisal framework* (Research report 700). Waka Kotahi NZ Transport Agency. https://www.nzta.govt.nz/assets/resources/research/reports/700/700-incorporating-distributional-impacts-equity-in-the-costbenefit-appraisal.pdf

Torshizian, E., Byett, A., Litman, T., Darlington, J., & Isack, E. (2025a). *Transport emission reduction initiatives* (Research report). NZ Transport Agency Waka Kotahi.

Torshizian, E., Roberts, P., Isack, E., Wallis, I., Byett, A., Litman, T., Fehling, A., Maralani, M., & Darlington, J. (2025b). *Interventions to reduce land transport greenhouse gas emissions* (Research report 714). NZ Transport Agency Waka Kotahi.

Transport Planning Society. (2020). *State of the nations: Transport planning for a sustainable future*. https://tps.org.uk/public/downloads/gUZd5/TPS%20State%20of%20the%20Nations%202020%20-%20FINAL.pdf

Valero, J. (2018, September 17–18). *Auckland Model Refresh 2018* (Presentation). NZMUGS Conference, Auckland. https://www.transportationgroup.nz/wp-content/uploads/2-1-2-AFC_ModelRefresh_NZMUGS_Sept2018.pptx-1.pdf

van der Waard, J., Jorritsma, P., & Immers, B. (2013). New drivers in mobility: What moves the Dutch in 2012? *Transport Reviews*, *33*(3), 343–359. https://doi.org/10.1080/01441647.2013.801046

van Dijk, J. (2018). Identifying activity-travel points from GPS-data with multiple moving windows. *Computers, Environment and Urban Systems*, *70*, 84–101. https://doi.org/10.1016/j.compenvurbsys.2018.02.004

Wandani, F. P., Siti, M., Yamamoto, M., & Yoshida, Y. (2018). Spatial econometric analysis of automobile and motorcycle traffic on Indonesian national roads and its socio-economic determinants: Is it local or beyond city boundaries? *IATSS Research*, *42*(2), 76–85. https://doi.org/10.1016/j.iatssr.2017.07.001

Welsh, I. (2023). *Estimating income from linked admin data: Impact of new sources*. Stats NZ. https://www.stats.govt.nz/assets/Research/Estimating-income-from-linked-admin-data-Impact-of-new-sources-/Estimating-income-from-linked-admin-data-impact-of-new-sources.pdf

Wen, L., Sheng, M. S., Sharp, B., Meng, T., Du, B., Yi, M., Suomalainen, K., & Gkritza, K. (2023). Exploration of the nexus between solar potential and electric vehicle uptake: A case study of Auckland, New Zealand. *Energy Policy*, *173*, 113406. https://doi.org/10.1016/j.enpol.2022.113406

Wilson, R. E., Anable, J., Cairns, S., Chatterton, T., & Turnbull, O. (2015, March 25). *Techniques for inferring mileage from the Department for Transport's MOT data set* (Presentation). Centre for Transport Study Seminar, Imperial College London. https://www.abdn.ac.uk/media/site/engineering/documents/WilsonImperialSeminar2015.pdf

Wilson, R. E., Anable, J., Cairns, S., Chatterton, T., Notley, S., & Lees-Miller, J. D. (2013). On the estimation of temporal mileage rates. *Procedia – Social and Behavioral Sciences*, *80*, 139–156. https://doi.org/10.1016/j.sbspro.2013.05.010

Wilson, R. E., Cairns, S., Notley, S., Anable, J., Chatterton, T., & McLeod, F. (2012). Techniques for the inference of mileage rates from MOT data. *Transportation Planning and Technology*, *36*(1), 130–143. https://doi.org/10.1080/03081060.2012.745768

Wu, L., Yuan, X., Yin, C., Yang, M., & Ouyang, H. (2023). Car ownership behavior model considering nonlinear impacts of multi-scale built environment characteristics. *Sustainability*, *15*(12), 9658. https://doi.org/10.3390/su15129658

Zamir, K. R., Nasri, A., Baghaei, B., Mahapatra, S., & Zhang, L. (2014). Effects of transit-oriented development on trip generation, distribution, and mode share in Washington, D.C., and Baltimore, Maryland. *Transportation Research Record: Journal of the Transportation Research Board*, *2413*(1), 45–53. https://doi.org/10.3141/2413-05

Zhao, P., & Li, P. (2021). Rethinking the determinants of vehicle kilometers traveled (VKT) in an auto-dependent city: Transport policies, socioeconomic factors and the built environment. *Transportation Planning and Technology*, *44*(3), 273–302. https://doi.org/10.1080/03081060.2021.1883228

Appendix A: Single observations

The distribution of single observations is shown in **Error! Reference source not found.**. Given the dates of IDI snapshots, it is not surprising that we find a significant proportion of single observations for 2016–2018. These are likely to reflect the lack of odometer readings conducted prior to and between the date of IDI snapshots currently available for analysis. Additionally, we find that, for more periods where we have greater data covered, the presence of single observations increases over time, likely reflecting the 3-year WOF requirements for newer vehicles (we begin to capture more vehicles that have only had their first odometer reading).

Table A.1 Single odometer readings by year

Year	Single observations	Total observations	% of single observations
2016	58,242	190,662	30.5%
2017	322,173	2,297,046	14.0%
2018	41,853	382,992	10.9%
2019	13,425	1,198,203	1.1%
2020	39,777	2,802,549	1.4%
2021	77,481	2,774,091	2.8%
2022	136,578	2,792,409	4.9%
2023	103,299	1,565,082	6.6%

Source: Principal Economics analysis.
Note: We omit odometer readings that were conducted on dates prior to 2016.

The impact of limited historical data needed to determine VKT for earlier years and the presence of limited odometer readings for newer vehicles can be seen when comparing the proportion of single observations by vehicle age and year. This is shown in Figure A.1.

Figure A.1 Single odometer observations by year and vehicle age

Vehicle age \ Year	2016	2017	2018	2019	2020	2021	2022	2023
Years 0 - 2	2.5%	2.0%	1.5%	2.6%	28.6%	77.8%	86.7%	90.3%
Years 3 - 6	6.5%	2.6%	2.8%	0.3%	0.4%	0.3%	5.6%	9.4%
Years 7 - 10	18.4%	4.7%	4.4%	0.7%	0.6%	0.5%	5.5%	10.2%
Years 11 - 14	35.3%	9.1%	6.6%	1.0%	0.8%	0.5%	2.6%	6.4%
Years 15 - 18	53.8%	20.3%	13.2%	0.8%	0.6%	0.2%	0.6%	0.6%
Years 19 - 21	65.8%	32.0%	21.2%	1.8%	0.5%	0.2%	0.4%	0.3%
Years 22 - 25	57.9%	35.6%	22.9%	2.9%	0.5%	0.1%	0.3%	0.5%
Years 26 - 29	46.2%	32.2%	20.3%	2.9%	0.5%	0.1%	0.1%	0.3%
Years 30 - 33	32.1%	22.9%	15.0%	3.1%	0.6%	0.2%	0.2%	0.4%
Years 34 - 37	17.0%	10.9%	9.8%	3.0%	0.8%	0.3%	0.4%	0.6%
Years 40 +	12.7%	3.9%	3.9%	1.5%	0.5%	0.3%	0.8%	1.6%
NA	6.4%	6.5%	2.7%	1.0%	0.5%	0.5%	0.8%	2.9%

Source: Principal Economics.

Figure A.2 shows the distribution of single observations by year and individual age.

Figure A.2 Single odometer observations by year and individual age

Individual age	2016	2017	2018	2019	2020	2021	2022	2023
Aged 15 - 19	60.9%	33.6%	20.5%	2.7%	1.7%	1.3%	4.0%	5.8%
Aged 20 - 24	52.4%	25.5%	17.9%	2.3%	1.7%	1.6%	4.5%	6.7%
Aged 25 - 29	42.9%	19.6%	15.4%	1.7%	1.5%	2.0%	5.1%	8.4%
Aged 30 - 34	35.6%	15.4%	12.1%	1.5%	1.5%	2.6%	5.9%	9.5%
Aged 35 - 39	32.6%	13.7%	10.5%	1.3%	1.5%	2.8%	6.2%	9.5%
Aged 40 - 44	29.9%	13.2%	10.4%	1.2%	1.4%	2.9%	5.8%	8.6%
Aged 45 - 49	27.8%	13.1%	10.3%	1.0%	1.3%	2.7%	5.1%	6.9%
Aged 50 - 54	25.5%	12.5%	10.2%	0.9%	1.3%	2.7%	4.4%	5.6%
Aged 55 - 59	22.0%	11.2%	9.2%	0.8%	1.2%	2.8%	4.3%	4.9%
Aged 60 - 64	17.6%	9.7%	7.9%	0.7%	1.3%	3.2%	4.3%	4.8%
Aged 65 - 69	11.9%	7.7%	6.7%	0.6%	1.6%	3.9%	5.1%	5.4%
Aged 70 - 74	10.5%	6.7%	5.5%	0.5%	1.6%	4.1%	4.9%	5.1%
Aged 75 - 79	8.9%	6.3%	5.4%	0.4%	1.5%	3.7%	4.6%	4.8%
Aged 80+	9.8%	6.1%	5.9%	0.3%	1.0%	2.4%	3.1%	3.2%
NA	49.2%	24.9%	20.9%	1.5%	1.0%	0.6%	0.8%	7.8%

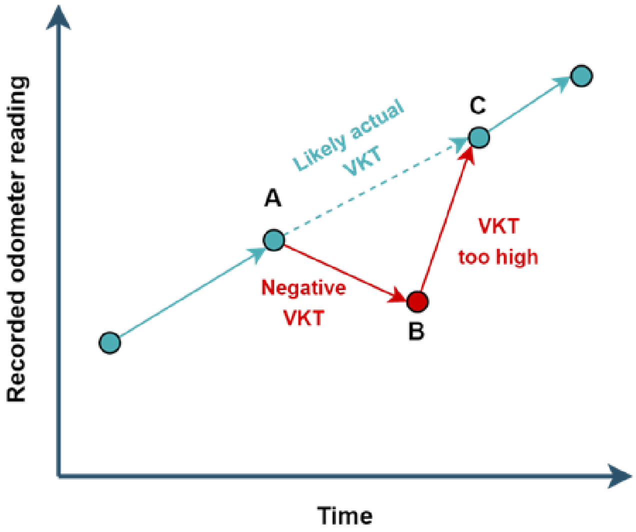
Source: Principal Economics.

Appendix B: Further details on erroneous odometer readings and VKT intervals

Wilson et al. (2015) highlight the issues with erroneous observations when used with spanning intervals. In situations where a 'bad' odometer reading exists, the calculation of VKT from 'good' odometer readings on either side of the bad reading is impacted such that an incorrect mileage rate is determined. Removal of the erroneous observation is likely to provide the correct VKT over the assessment period.

Figure B.1 shows how an incorrect odometer reading (point B) can lead to incorrect VKT for periods A-B and B-C. Removing reading B and recalculating estimated VKT using only points A-C provides a more robust VKT estimation.

Figure B.1 Negative VKT and erroneous readings (adapted from Wilson et al., 2015, p. 20)

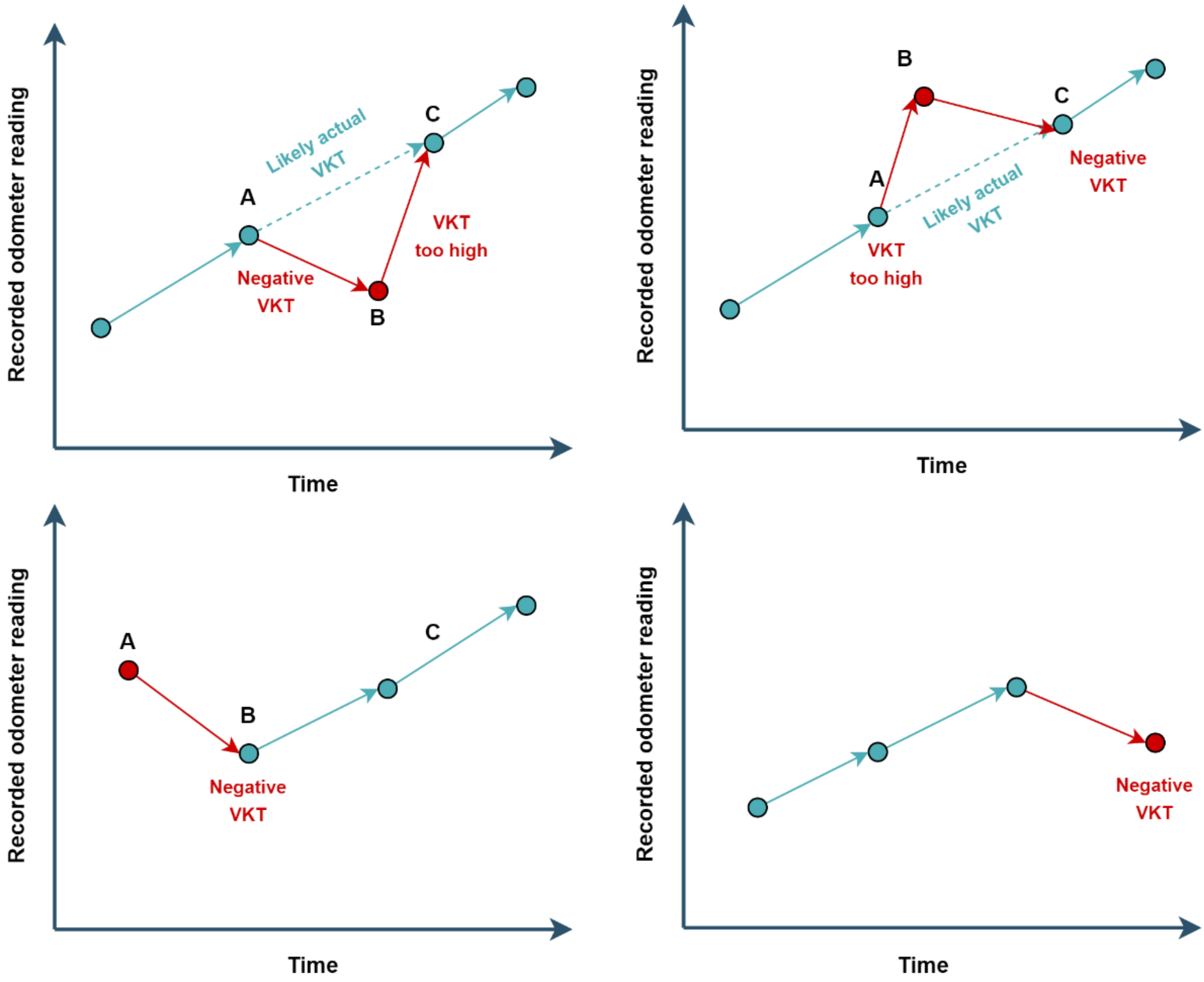


Wilson et al. (2015) and Glensor (2021) note other variations of this error when assessing odometer data. We expand on the variation of negative odometer readings and their corresponding erroneous readings.

As shown in Figure B.2, we develop a fast algorithm for determining the odometer readings based on sequential observations. When a negative VKT is derived, the algorithm tests the relative position of the related odometer reading and the sequence of odometer readings for the vehicle and compares against odometer readings one to two periods ahead and prior, depending on its position in the set of readings available for that vehicle. This is used to identify the position of the erroneous odometer reading and flags it for removal. VKT is then recalculated for the new set of odometer readings and compared against prior estimates. Subsequently, this two-stage analysis allows for the identification of odometer rollover or odometer replacement where a negative VKT is determined post-removal. In these cases, we drop the estimated VKT for that odometer span but keep all other estimated VKT spans following odometer rollover/replacement.

Zero odometer readings are removed prior to checking for negative VKT. Remaining excessive VKT intervals are removed following this algorithm.

Figure B.2 Variation of negative VKT and corresponding erroneous readings (adapted from Wilson et al., 2015, p. 20)



Appendix C: Further details on preliminary VKT estimates

Table C.1 Average VKT of vehicle owners by income quintile

Income quintile	2016	2017	2018	2019	2020	2021	2022
1 (Lowest)	9,318	11,411	12,674	12,448	12,397	16,051	11,669
2	8,505	9,884	10,606	10,648	10,353	14,320	9,018
3	9,273	11,189	12,431	12,203	12,220	16,413	12,015
4	9,908	11,650	13,278	13,346	13,238	17,498	12,538
5 (Highest)	11,249	11,935	13,989	14,176	13,991	18,643	12,973

Appendix D: Estimation results

Table D.1 Regression results of the VKT factors

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model: (OLS)	Demo	(1) + reg	Veh	(3) + reg	All	All 21	(5) + EV	(7) + PT
Sex (Male)								
Female	0.10***	0.09***			-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
Not specified	-0.2***	-0.21***			-0.22***	-0.30**	-0.23***	-0.27***
	(0.07)	(0.07)			(0.07)	(0.14)	(0.07)	(0.07)
Ethnicity								
European	-0.1***	-0.09***			-0.07***	-0.08***	-0.07***	-0.08***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
Māori	0.09***	0.08***			0.10***	0.11***	0.10***	0.10***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
Pacific	0.12***	0.12***			0.12***	0.12***	0.12***	0.12***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
Asian	-0.14***	-0.13***			-0.19***	-0.20***	-0.19***	-0.19***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
MELAA	-0.02***	-0.01***			-0.02***	-0.01***	-0.02***	-0.02***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
Other	-0.03***	-0.03***			-0.01***	-0.02***	-0.01***	-0.01***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
Employed	-0.04***	-0.04***			-0.03***	-0.03***	-0.03***	-0.02***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
Self-employed	0.10***	0.10***			0.08***	0.09***	0.08***	0.08***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
HH income	-0.17***	-0.15***			-0.24***	-0.18***	-0.24***	-0.22***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
HH income sq.	0.00***	0.00***			0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
Age								
15_19	-0.04***	-0.05***			0.06***	0.06***	0.06***	0.05***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
20_24	0.02***	0.03***			0.10***	0.11***	0.10***	0.10***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
25_29	0.03***	0.03***			0.06***	0.07***	0.06***	0.07***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
35_39	-0.04***	-0.04***			-0.05***	-0.06***	-0.05***	-0.06***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
40_44	-0.06***	-0.06***			-0.07***	-0.08***	-0.07***	-0.07***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
45_49	-0.08***	-0.08***			-0.08***	-0.09***	-0.08***	-0.08***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
50_54	-0.11***	-0.12***			-0.10***	-0.11***	-0.10***	-0.10***

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model: (OLS)	Demo	(1) + reg	Veh	(3) + reg	All	All 21	(5) + EV	(7) + PT
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
55_59	-0.16***	-0.17***			-0.14***	-0.16***	-0.14***	-0.14***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
60_64	-0.22***	-0.22***			-0.20***	-0.21***	-0.20***	-0.20***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
65_69	-0.26***	-0.26***			-0.27***	-0.27***	-0.27***	-0.26***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
70_74	-0.35***	-0.35***			-0.38***	-0.39***	-0.38***	-0.38***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
75_79	-0.47***	-0.47***			-0.51***	-0.53***	-0.51***	-0.52***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
80+	-0.76***	-0.76***			-0.79***	-0.81***	-0.79***	-0.79***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
# under 15	0.05***	0.04***			0.02***	0.02***	0.02***	0.02***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
# working age	0.03***	0.02***			0.03***	0.03***	0.03***	0.03***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
# under 15 x # working age	-0.00***	-0.00***			-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
Over 65	-0.01***	-0.02***			-0.02***	-0.02***	-0.02***	-0.02***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
NZQCF level (no certificate)								
1 Certificate - Basic knowledge	0.03***	0.03***			-0.01***	-0.01***	-0.01***	-0.01***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
2 Certificate - Operational	0.06***	0.06***			0.02***	0.02***	0.02***	0.01***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
3 Certificate - Knowledge and operation	0.08***	0.09***			0.03***	0.04***	0.04***	0.03***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
4 Certificate - Limited supervision	0.02***	0.02***			0.00***	0.00	0.00***	0.00***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
5 Diploma	0.09***	0.09***			0.04***	0.04***	0.04***	0.04***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
6 Diploma	0.11***	0.12***			0.03***	0.03***	0.03***	0.03***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
7 Graduate	0.14***	0.16***			0.04***	0.04***	0.04***	0.04***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
8 Bachelor and post graduate	0.16***	0.17***			0.04***	0.04***	0.04***	0.04***
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
9 Master’s degree	0.15***	0.17***			0.05***	0.04***	0.05***	0.04***

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model: (OLS)	**Demo**	**(1) + reg**	**Veh**	**(3) + reg**	**All**	**All 21**	**(5) + EV**	**(7) + PT**
	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
10 Doctoral degree	0.13***	0.16***			0.02***	0.00	0.02***	0.01***
	(0.00)	(0.00)			(0.00)	(0.01)	(0.00)	(0.00)
Year (2016)								
2017	0.08***	0.08***	0.22***	0.22***	0.23***		0.23***	0.22***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)
2018	0.08***	0.08***	0.21***	0.21***	0.23***		0.23***	0.22***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)
2019	0.08***	0.08***	0.18***	0.18***	0.21***		0.21***	0.20***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)
2020	0.07***	0.07***	0.14***	0.14***	0.16***		0.16***	0.15***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)
2021	-0.01**	-0.01**	0.04***	0.04***	0.06***		0.06***	0.05***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)
2022	-0.01***	-0.01***	-0.01*	-0.01*	0.02***		0.02***	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)
Regions (Auckland)								
Northland		0.03***	0.02***	0.06***	0.04***	0.05***	0.05***	0.05***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Waikato		0.11***	0.09***	0.12***	0.11***	0.12***	0.11***	0.11***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Bay of Plenty		0.12***	0.07***	0.09***	0.11***	0.13***	0.10***	0.10***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Gisborne		-0.06***	-0.06***	-0.00	-0.04***	-0.02***	-0.05***	0.01**
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Hawke's Bay		0.06***	0.04***	0.08***	0.09***	0.12***	0.08***	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Taranaki		0.07***	0.04***	0.06***	0.07***	0.09***	0.05***	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Manawatū-Whanganui		0.07***	0.03***	0.09***	0.09***	0.12***	0.09***	0.15***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Wellington		0.03***	0.04***	0.02***	0.03***	0.06***	0.04***	0.03***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
West Coast		-0.16***	-0.08***	-0.13***	-0.11***	-0.06***	-0.09***	
		(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	
Canterbury		0.02***	0.01***	0.05***	0.08***	0.12***	0.09***	0.09***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Otago		-0.06***	-0.02***	-0.01***	0.01***	0.05***	0.01***	0.02***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Southland		-0.07***	-0.02***	0.01***	0.02***	0.06***	0.02***	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	
Tasman		-0.11***	-0.05***	-0.07***	-0.02***	0.01**	-0.02***	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	
Nelson		-0.07***	-0.05***	-0.02***	0.01***	0.06***	0.01***	

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model: (OLS)	**Demo**	**(1) + reg**	**Veh**	**(3) + reg**	**All**	**All 21**	**(5) + EV**	**(7) + PT**
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	
Marlborough		-0.02***	-0.14***	0.00	0.04***	0.07***	0.03***	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	
Rural urban (Metropolitan)								
Outside of FUA		-0.03***	0.11***	-0.04***	-0.01***	-0.00	0.01***	0.00**
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Small centre		-0.08***	0.01***	-0.12***	-0.06***	-0.06***	-0.03***	-0.02***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Medium centre		-0.07***	-0.01***	-0.09***	-0.05***	-0.06***	-0.05***	-0.06***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Large centre		-0.01***	-0.02***	-0.02***	0.00***	-0.00	0.00***	0.02***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log population density		-0.02***		-0.01***	-0.02***	-0.03***	0.02***	0.04***
		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log driving distance		-0.08***		-0.12***	-0.09***	-0.09***	-0.09***	-0.08***
		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log PT distance		0.01***		-0.00*	0.01***	0.01***	0.00***	0.01***
		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log walking distance		0.03***		0.06***	0.03***	0.03***	0.03***	0.02***
		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log biking distance		0.08***		0.08***	0.08***	0.08***	0.08***	0.10***
		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Vehicle year			2.00***	2.01***	1.43***	1.29***	1.43***	1.55***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
vehyear # vehyear			-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Vehicle body (Station Wagon)								
Convertible			-0.95***	-0.95***	-0.84***	-0.83***	-0.84***	-0.84***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Hatchback			-0.17***	-0.16***	-0.13***	-0.13***	-0.13***	-0.14***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Heavy van			-0.23***	-0.23***	-0.15***	-0.08	-0.15***	-0.17***
			(0.04)	(0.04)	(0.04)	(0.09)	(0.04)	(0.04)
Light van			-0.09***	-0.09***	-0.04***	-0.04***	-0.04***	-0.04***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Motorcycle			-1.79***	-1.80***	-1.79***	-1.88***	-1.80***	-1.84***
			(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)
Saloon			-0.19***	-0.19***	-0.17***	-0.18***	-0.17***	-0.17***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Sports car			-0.90***	-0.89***	-0.83***	-0.83***	-0.83***	-0.84***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Utility			-0.33***	-0.33***	-0.30***	-0.30***	-0.30***	-0.30***
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Fuel type (Petrol)								

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model: (OLS)	Demo	(1) + reg	Veh	(3) + reg	All	All 21	(5) + EV	(7) + PT
Diesel			0.00***	0.00***	-0.04***	-0.03***	0.02***	-0.03***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Diesel hybrid			0.02	0.02	0.05	0.10	0.20	0.06
			(0.06)	(0.06)	(0.20)	(0.36)	(0.55)	(0.20)
Electric			0.12***	0.13***	0.06***	0.06***	-0.04	0.07***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.03)	(0.00)
Electric Petrol			-0.06***	-0.03	-1.47***	-1.48*	-1.59***	-1.45***
			(0.02)	(0.02)	(0.57)	(0.87)	(0.60)	(0.56)
Petrol Hybrid			0.38***	0.39***	0.41***	0.42***	0.57***	0.43***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)
Plugin			0.14***	0.14***	0.01	-0.04	0.12**	-0.00
			(0.01)	(0.01)	(0.03)	(0.06)	(0.06)	(0.03)
Vehicle size (Large)								
Medium			-0.05***	-0.05***	-0.03***	-0.03***	-0.03***	-0.02***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Small			-0.10***	-0.10***	-0.05***	-0.05***	-0.05***	-0.04***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Extra-Large			-0.21***	-0.22***	-0.23***	-0.24***	-0.23***	-0.23***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Extra-Small			-0.17***	-0.17***	-0.11***	-0.12***	-0.11***	-0.10***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
# Cars			-0.02***	-0.02***	-0.05***	-0.10***	-0.05***	-0.06***
			(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
# Cars sq.			-0.00	-0.00	0.02***	0.03***	0.02***	0.02***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Interaction: Size and Fuel type								
M # Diesel					0.03***	0.02***	0.03***	0.03***
					(0.00)	(0.01)	(0.00)	(0.00)
M # Diesel hybrid					0.12	0.10	0.12	0.12
					(0.21)	(0.38)	(0.21)	(0.21)
M # Electric					-0.24	-0.16	-0.26	-0.28
					(0.18)	(0.35)	(0.18)	(0.18)
M # Electric Petrol					1.84***	1.70	1.84***	1.82***
					(0.67)	(1.23)	(0.67)	(0.67)
M # Petrol Hybrid					0.04***	0.05***	0.04***	0.04***
					(0.01)	(0.01)	(0.01)	(0.01)
M # Plugin					0.17***	0.23***	0.17***	0.18***
					(0.03)	(0.06)	(0.03)	(0.03)
S # Diesel					0.04***	0.02**	0.04***	0.02***
					(0.01)	(0.01)	(0.01)	(0.01)
S # Electric					-0.36***	-0.31	-0.36***	-0.30**
					(0.13)	(0.27)	(0.13)	(0.14)
S # Electric Petrol					2.23**	0	2.27**	2.25**
					(0.98)	(.)	(0.98)	(0.97)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model: (OLS)	**Demo**	**(1) + reg**	**Veh**	**(3) + reg**	**All**	**All 21**	**(5) + EV**	**(7) + PT**
S # Petrol Hybrid					-0.10***	-0.12***	-0.10***	-0.11***
					(0.01)	(0.01)	(0.01)	(0.01)
S # Plugin					-0.11***	-0.07	-0.11***	-0.10***
					(0.03)	(0.06)	(0.03)	(0.03)
XL # Diesel					0.17***	0.17***	0.17***	0.17***
					(0.00)	(0.01)	(0.00)	(0.00)
XL # Diesel hybrid					0.77**	0.99	0.78**	0.78**
					(0.39)	(0.74)	(0.39)	(0.38)
XL # Electric					0.38***	0.49*	0.38***	0.37***
					(0.14)	(0.26)	(0.14)	(0.14)
XL # Electric Petrol					1.76***	1.96*	1.73***	0
					(0.63)	(1.06)	(0.63)	(.)
XL # Petrol Hybrid					-0.01	-0.00	-0.01	-0.03***
					(0.01)	(0.02)	(0.01)	(0.01)
XL # Plugin					-0.08	-0.21	-0.08	0.01
					(0.10)	(0.19)	(0.10)	(0.10)
XS # Diesel					0.06***	0.03	0.06***	0.05**
					(0.02)	(0.05)	(0.02)	(0.02)
XS # Electric Petrol					1.47***	1.51*	1.48***	1.44**
					(0.57)	(0.87)	(0.57)	(0.56)
XS # Petrol Hybrid					-0.08***	-0.09***	-0.08***	-0.11***
					(0.01)	(0.01)	(0.01)	(0.01)
XS # Plugin					-0.00	0.05	-0.00	0.01
					(0.04)	(0.08)	(0.04)	(0.04)
Interaction: Fuel type and household income								
Diesel # hhd_inc					0.21***	0.17***	0.21***	0.20***
					(0.00)	(0.01)	(0.00)	(0.00)
Diesel hybrid # hhd_inc					-0.73*	-1.26	-0.74*	-0.75*
					(0.41)	(0.83)	(0.41)	(0.41)
Electric # hhd_inc					0.10***	0.06***	0.10***	0.09***
					(0.01)	(0.02)	(0.01)	(0.01)
Electric Petrol # hhd_inc					0.09	0.08	0.09	0.09
					(0.06)	(0.08)	(0.06)	(0.06)
Petrol Hybrid # hhd_inc					-0.06***	-0.04***	-0.06***	-0.08***
					(0.01)	(0.01)	(0.01)	(0.01)
Plugin # hhd_inc					0.11***	0.14***	0.11***	0.12***
					(0.03)	(0.04)	(0.03)	(0.03)
Log distance to EV charger							0.06***	0.08***
							(0.00)	(0.00)
Interaction: Fuel type and Log distance to EV charger								
Diesel # log_ev							-0.01***	
							(0.00)	
Diesel hybrid # log_ev							-0.02	
							(0.07)	

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model: (OLS)	**Demo**	**(1) + reg**	**Veh**	**(3) + reg**	**All**	**All 21**	**(5) + EV**	**(7) + PT**
Electric # log_ev							0.01***	
							(0.00)	
Electric Petrol # log_ev							0.02	
							(0.03)	
Petrol Hybrid # log_ev							-0.02***	
							(0.00)	
Plugin # log_ev							-0.02**	
							(0.01)	
log_ev # log_dens							-0.01***	-0.01***
							(0.00)	(0.00)
Log distance to PT station								-0.01***
								(0.00)
PT 15 minutes frequency								-0.03***
								(0.00)
PT 30 minutes frequency								-0.02***
								(0.00)
Constant	2.87***	2.71***	-2040.90***	-2048.30***	-1473.10***	-1330.00***	-1473.10***	-1591.10***
	(0.00)	(0.01)	(4.30)	(4.30)	(4.20)	(9.12)	(4.20)	(4.84)
Observations	13731096	13731096	13731096	13731096	13731096	3232260	13731096	11086353
Adj R sq	0.08	0.08	0.22	0.23	0.29	0.27	0.29	0.29

Source: Principal Economics.
Note: Significance codes: *** p<0.01, ** p<0.05, * p<0.1.

Appendix E: Review of the car ownership and EV uptake literature

Table E.1 Methods and factors of EV uptake

Authors	Study area	Estimation method	Sample size	Dependent variable	Independent variables
Axsen & Kurani, 2013	San Diego County, California, US	Multi-part, multi-mode survey	3,053,793 (2009); 508 (2011)	EV adoption	+ Sex (being male) + Income + Education
Hardman & Tal, 2021	California, US	Binary logistic regression model	1,727	Battery EV discontinuance	+ Number of vehicles in the household *** + Convenience of charging ** + Level 2/level 1 ** (home charging categories)
Wen et al., 2023	Auckland, New Zealand	Spatial negative binomial model[45]	143	The intensity of the rate parameter	+ Solar photovoltaics * + Total EV in 2018 *** - Distance to CBD * - Drive alone *** + Two vehicles *** + High qualification *** - Being male (%)
Potoglou & Kanaroglou, 2007	Hamilton, Canada	Nested multinomial logit	3,856	The choice of vehicle type	+ No purchase tax*** + Access to high occupancy lanes * + Pollution level: 10% of a present-day average car (AF and hybrid) ** + Hybrid x university degree *** + Hybrid x household income ** - Price x household income: medium *** - Price x household income: high *** - Fuel availability: 10% of existing gas stations *** - Fuel availability: 25% of existing gas stations *** - Alternative fuelled vehicle x age of respondent >44 years *** - Alternative fuelled vehicle x long-distance commuters (>7 km) per household x FA at 10% **
Javid & Nejat, 2017	58 counties in California and Delaware Valley, US	Multiple logistic regression analysis	16,348	Vehicle choice – plug-in electric vehicles (PEVs) and conventional cars (non-PEVs)	+ Car sharing *** + Income *** + Maximum education *** + Charging station per capita *** + Gas price ***

[45] For simplicity, only show whether the independent variables are significant or not (indicated by *) when band = 75 km in the table.

Table E.2 Methods and factors of car ownership

Authors	Study area	Estimation method	Sample size	Dependent variable	Independent variables
Potoglou & Kanaroglou, 2008	Hamilton, Canada	MNL	774 households	Number of vehicles	One vehicle:[46] + Type of dwelling: single-family house *** - Number of persons employed part-time *** + Number of individuals with driver licence/household size *** + Household income: medium[47] ** + Number of individuals working at distance >6 km ** Two vehicles: + Type of dwelling: single-family house *** - Number of persons employed part-time *** + Number of full-time workers ** + Number of individuals with driver licence/household size *** + Household type: couple (with or without children) *** + Household type: extended family or unattached individuals*** - Mixed density index * - Land-use entropy index (EI500) * + Number of individuals working at distance >6 km *** Three or more vehicles: + Type of dwelling: single-family house *** + Number of full-time workers *** + Number of individuals with driver license/household size *** + Household type: extended family or unattached individuals *** - Mixed density index * - Land-use entropy index (EI500) * - Number of bus stops within 500 m from dwelling * + Number of individuals working at distance >6 km ***
Ostermeijer et al., 2019	Amsterdam, Rotterdam, the Hague and Utrecht, Netherlands	Two-step estimation methodology (multinomial logit model, average marginal effects)	98,659	Step 1: property sale price Step 2: household car ownership level	Step 1: + Presence of private parking *** + Parking type *** Property characteristics: + Size of the property + Parcel size + Number of rooms/bathrooms

[46] * for coefficients where the absolute value of the t-statistic is 2.0 or greater but less than 2.58; ** for coefficients where the absolute value of the t-statistic is 2.58 or greater but less than 3.29; *** for coefficients where the absolute value of the t-statistic is 3.29 or greater.

[47] CAN$30,000–$80,000.

Authors	Study area	Estimation method	Sample size	Dependent variable	Independent variables
					+ Amenities and features + Property condition Age and type of property: ± Construction year (depending on location) ± House type (different types have different price levels) Locational characteristics: - Distance to metropolitan centre - Distance to train station/highway ramp Step 2: - Estimated parking costs Household characteristics: + Household income + Household size + Number of working-age individuals + Presence of children + Education level + Employment status (including being self-employed) Built environment characteristic: - Population density + Distance to city centre + Distance to market/ town centre by car - PT availability and coverage - Land area mixed use Heterogeneous effects: By dwelling type: 1. Households living in flats - Parking costs effect *** 2. Households living in houses - Parking costs effect ** By tenure status, i.e. renter: - Parking costs effect * By region: 1. Amsterdam - Parking costs effect 2. Other municipalities - Parking costs effect ***
Song & Wang, 2017	Broward, Palm Beach, and Miami-	Poisson regression analysis	3,980	Number of vehicles per household	+ Number of drivers ** + Number of workers ** + Household income level *

Authors	Study area	Estimation method	Sample size	Dependent variable	Independent variables
	Dade in southern Florida, US				- Housing tenure ** - Proximity to schools ** - Net house density **
Sabouri et al., 2021	32 diverse regions in the US	Count regression models (quasi-Poisson and Poisson)	91,959 households	Actual number of vehicles owned by household	Socio-demographic characteristics: + Household size *** + Number of employed persons in household *** + Households income *** Built environment: - Density *** - Diversity (land use entropy) *** - Design (intersection density and percentage of four-way intersections) *** - Transit (stop density) *** - Destination accessibility (jobs within 10 and 30 minutes' travel time by auto and 30 minutes by transit) ***[48]
Wu et al., 2023	116 cities and 334 communities, China	Gradient-boosting decision tree (GBDT)[49]	17,186	Car ownership behaviour	Individual family attributes: + Household income + Family size Neighbourhood-scale built-environment characteristics: - Degree of mixed land use - Distance to bus station ± Distance to city centre Neighbourhood population density[50] City-scale built-environment characteristics: + Number of buses per 10,000 people Urban population density + Road area per capita

[48] The only exception occurs in the Poisson model, where the destination accessibility within 10 minutes is not statistically significant.

[49] Significance levels are not usually computed in GBDT.

[50] One factor with significant impacts on residents' car ownership, but the specific nature of its relationship (positive or negative) to car ownership is not detailed in the study.