NEW ZEALAND INFRASTRUCTURE COMMISSION
Te Waihanga

Drivers of household expenditure on infrastructure

An analysis of the factors that explain variations in household infrastructure spending

Research note

tewaihanga.govt.nz

New Zealand Infrastructure commission / Te Waihanga

Te Waihanga seeks to transform infrastructure for all New Zealanders. By doing so our goal is to lift the economic performance of Aotearoa and improve the wellbeing of all New Zealanders.

We are an autonomous Crown entity, listed under the Crown Entities Act 2004, with an independent board. We were established by the New Zealand Infrastructure Commission/Te Waihanga Act 2019 on 25 September 2019. Information on the Commission is available at www.tewaihanga.govt.nz/

How to cite this document

New Zealand Infrastructure Commission (2024). *Drivers of household expenditure on infrastructure: An analysis of the factors that explain variations in household infrastructure spending*. Wellington: New Zealand Infrastructure Commission / Te Waihanga.

Contact

Peter Nunns
Director, Economics
Email: peter.nunns@tewaihanga.govt.nz
Website: tewaihanga.govt.nz
LinkedIn: tewaihanga

Judy Kavanagh
Director, Inquiries
Email: judy.kavanagh@tewaihanga.govt.nz
Website: tewaihanga.govt.nz
LinkedIn: tewaihanga

ISBN 2024 978-1-7385911-8-3 (Online)

Acknowledgements

This work supports the 'What's fair? Providing and paying for infrastructure' project that Te Waihanga is undertaking. The research note was drafted by **Andrew Sweet** with support from **Peter Nunns**. We are grateful for the review and feedback provided by **Judy Kavanagh and Dave Heatley**.

We would also like to thank **Statistics New Zealand** for providing access to the data that was used for the analysis – see IDI disclaimer above.

Table of contents

Glossary

Access charge	An access charge is a cost that must be met in order to use a particular service. In the context of this paper, the cost of buying or renting a vehicle to use the road network can be seen as an access charge.
Categorical variable	A categorical variable is a variable that can only have one of a limited, and usually fixed, number of possible values. The variable 'household owns a private vehicle' used in this paper is a good example. The entry for each household for that variable can only be 'yes' or 'no'.
Dependent variable	A dependent variable in regression analysis is the main factor that one is trying to understand or predict. For this paper, the dependent variable is household expenditure on infrastructure services.
Dummy variable	Dummy variables are used in regression analysis to represent categorical data. If a categorical variable can only have one of four possible values, three dummy variables would be included in the regression analysis.
Explanatory or independent variable	An independent or explanatory variable in regression analysis is a factor that you include in the model to explain or predict changes in the dependent variable.
Fixed expenditure	Fixed infrastructure expenditure is spending that remains constant regardless of a household's level of infrastructure usage.
Ordinary least squares (OLS)	Ordinary least squares, or OLS, is a form of regression analysis.
Regression analysis	Regression analysis is a statistical method used to estimate the relationships between a dependent variable and one or more independent variables.
Statistical significance	A model (or coefficient) is statically significant if appropriate tests conclude that the result is highly unlikely to be explained solely by chance or random factors.
Total expenditure	Total expenditure is the sum of fixed and variable expenditure.
Variable expenditure	Variable infrastructure expenditure is spending that changes in proportion to the level of a household's infrastructure usage.

1. Summary

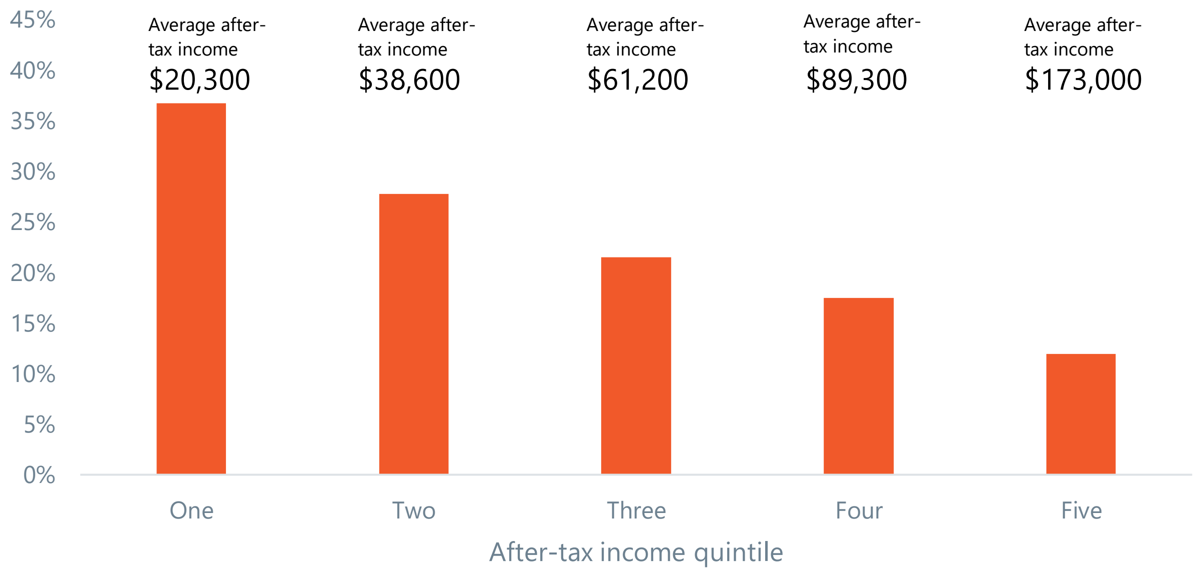
Infrastructure networks are vital to our quality of life, and the functioning of our economy. Safe transport, reliable electricity, extensive mobile phone coverage, fast internet services, and clean water underpin our modern standards of living.

The Infrastructure Commission is preparing a series of reports that collectively aim to improve our understanding of household spending on infrastructure services. Our first report estimated the amounts that households in New Zealand spend on infrastructure services such as roading, passenger transport, energy, water and telecommunications. It was published as a research insights report 'How much do we pay for infrastructure?'[1] in mid-2023.

New Zealanders pay for their infrastructure use in a range of ways, including through council rates, central government taxes, fuel excise duties, and of course user charges. Taking all of those various funding sources into account, that first 'How much do we pay?' report found that on average, New Zealand households spent around 16% of their after-tax income on infrastructure services in 2018/19, which is equal to around $260 per week or slightly over $13,500 per year.

Importantly, that report also found considerable variation in household spending around that national average. Households with lower incomes, for example, were found to spend smaller amounts on infrastructure in dollar terms, but a higher proportion of their after-tax income. Figure 1 below shows the different levels of household spending for each after-tax income quintile.

Figure 1: Proportion of households' after-tax income spent on infrastructure services by after-tax income quintile (2006/07 to 2018/19)



In addition to income, that earlier report also found that:

- Households in cities spend less on infrastructure than those in rural areas or regional towns.
- Households without a working adult spend a significantly higher portion of their income on infrastructure than those with at least one working adult.

[1] New Zealand Infrastructure Commission (2023). *How much do we pay for infrastructure? Household expenditure on infrastructure services*. Wellington: New Zealand Infrastructure Commission / Te Waihanga.

- Households with one or more dependent children spent modestly more than households with no children.

This research note builds on, and extends, that earlier work. Its objectives are to:

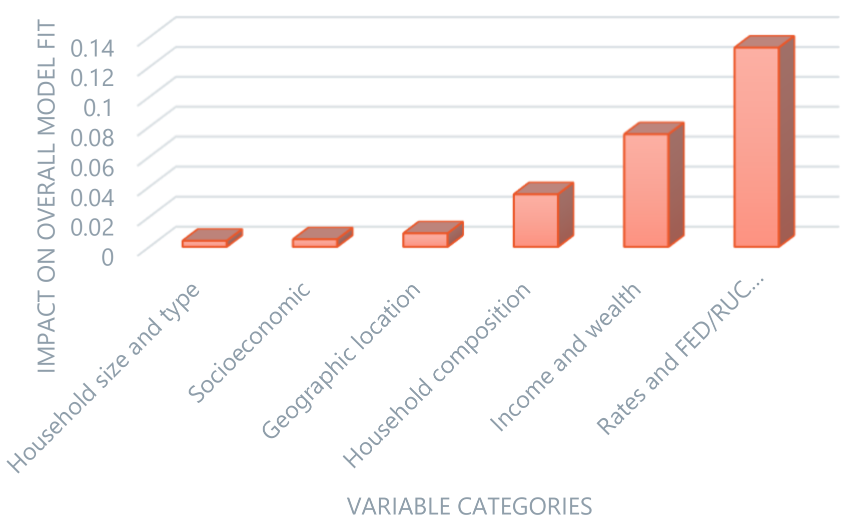
- identify as many additional factors as possible that influence levels of household infrastructure spending
- assess the size of the effect of each factor
- explain as much of the observable variation in household infrastructure spending as possible given the available data.

We used a multi-step process (described in the methodology section) to identify a set of between 30 and 36 variables[2] (including dummy variables) that look to have a statistically significant impact on levels of household infrastructure spending. That list of variables can be found in Table 3 in the body of this research note.

Using a set of four slightly different model specifications, we in turn found that those variables were able to explain between 64% and almost 70% of the observable variation in household infrastructure spending. While further work could always be done to look for an even greater number of relevant variables, we view an R-squared statistic of close to 0.7 as meaning that we have been able to identify most of the important factors that influence infrastructure spending. As such we see these results as being sufficiently strong to allow them to be used to inform policy making.

We then assessed the relative explanatory power of different groups of those variables by re-running our key model with and without each group. The results of that decomposition analysis are shown in Figure 3 below.

Figure 2: Contribution of different variable categories to the model's overall fit



[2] The variation in the number of variables across the different models we developed reflects the fact that the statistical significance of some variables dropped when we changed our model specification, resulting in their removal from that particular model.

As you can see, the group of variables with the largest explanatory power is that which influences the amounts each household pays in rates, fuel excise duty (FED) and road user charges (RUC). The variables included in that category are income after tax, expenditure on non-capital items, household expenditure on rates, and a dummy variable indicating whether the household owns at least one car.

Removing that group of variables reduced the overall explanatory power of our model from almost 70% to 56.6%; a difference of 13.3 percentage points. The importance of this group of variables is not surprising, as the variables in it directly determine the amounts each household is required to pay for their water and roading services, rather than indirectly estimating levels of infrastructure use, as the remaining groups of variables do.

The next most important group of variables relates to household income (both current and permanent income). Together those variables explain roughly 7.5 percentage points of the overall explanatory power of model 2. This again is not surprising, as we would expect both current and permanent income to be strongly related to each household's level of infrastructure consumption.

After that, the explanatory power of the remaining groups of variables is much smaller. Household composition variables – showing the number and type of household members – are the next most significant, explaining roughly 3.5% of the overall variation in spending. Geographic location variables are next, explaining 0.9%. Followed by socio-economic variables (0.5%) and variables relating to household size and type (0.4%).

2. Methodology

2.1. Data sources

We extended the data set originally developed for our first report 'How much do we pay for infrastructure?' for this analysis. The data on infrastructure spending is drawn from the expenditure module of Stats NZ's Household Economic Survey (HES). Households participating in that component of the survey are required to keep a careful diary of their spending over the course of a week. The detailed expenditure data reported in that expenditure module is collected every three years. We used data from the five most recent 'waves' of that expenditure survey between 2006/07 and 2018/19.

In turn, we took data on other factors that we thought might affect infrastructure spending from a range of other linked data sets in the Integrated Data Infrastructure (IDI)[3].

2.2. Identification of relevant explanatory variables

As noted, our first objective was to identify as wide a range of factors as possible that appear to influence levels of household infrastructure spending. We did that through a multi-step process. First, we looked through a wide range of datasets contained in the IDI to identify variables that, a priori, seemed likely to have an effect on infrastructure spending. This became our variable longlist.

Second, we assessed the impact of each variable on the longlist in isolation, by calculating the simple correlation between it and infrastructure spending. That allowed us to discard any variables with limited or no explanatory power, and to determine an initial ranking of the remaining variables in terms of the strength of their relationship with infrastructure spending. The remaining variables became our shortlist.

Third, we then simultaneously included all of the variables from our shortlist into an initial statistical model and assessed its overall explanatory power using ordinary least squares (OLS) regression analysis.

Finally, we assessed the effect of dropping individual variables, or clusters of variables, from that initial model, on its overall explanatory power. Individual variables that were found to reduce the overall explanatory power or to have no statistical significance once included alongside the other variables from our long list were then dropped.

2.3. Structure of our models

We ran the process described above for four slightly different statistical models. All four models had the following high-level structure:

$$
\begin{aligned}
Log(infrastructure\ expenditure\ ) \\
&= \alpha + \beta 1 \log(after\ tax\ income\ and/or\ expenditure\ on\ non\ capital\ goods) \\
&+ \beta 2\ (other\ variables) + \epsilon
\end{aligned}
$$

Where:

α = constant
β1 = impact of income and/or wealth on infrastructure expenditure
β2 = impact of other explanatory variables on infrastructure expenditure

[3] Readers wanting more information on our use of the HES and additional IDI datasets should refer to Chapter 2 of that earlier paper.

However, the models differed on two key dimensions. The first dimension relates to our choice of dependent variable. Models 1-3 used *total* infrastructure expenditure as the dependent variable (the variable we are attempting to explain variations in), while model 4 used expenditure on infrastructure *as a percentage of after-tax income* as the dependent variable.

The second dimension relates to how we treated income in each model. Economists sometimes differentiate between:

- current income: the income earned by the members of a household in any given year; and
- permanent income: the long-term average income that the members of a household expect to earn over their lifetimes.

Proponents of the permanent income hypothesis argue that consumers will base their spending decisions on their expected long term average income (or 'permanent income'), rather than their actual income in any given year (which will sometimes fluctuate up and down between years). To allow for that possibility we have included current income and a 'proxy' for permanent income in our models in different combinations.

We say a 'proxy' for permanent income, because it is typically not possible to measure permanent income directly. Annual household expenditure on non-capital items is often used as a proxy for permanent income because it tends to be more stable over time than current income, and is thought to align more closely with a household's long-term income expectations.

In model 1 we only included current after-tax income. In model 2 we included both current income and a proxy for permanent income. In turn in model 3 we only included a proxy for permanent income.

This use of four slightly different models served two purposes. First, it provided a form of sensitivity testing, giving us more confidence in the robustness of our findings. Second, it allowed us to explore which measures of expenditure and income are best suited to our objectives.

A summary of the design of the four models is provided in Table 1 below.

Table 1: Summary of models

	Model 1	Model 2	Model 3	Model 4
Dependent variable	(Log of) Total expenditure on infrastructure ($)	(Log of) Total expenditure on infrastructure ($s)	(Log of) Total expenditure on infrastructure ($s)	Infrastructure expenditure as a percentage of after tax income (%)
Includes current income after tax?	Yes	Yes	No	Yes
Includes expenditure on non-capital items (as a proxy for wealth / permanent income)	No	Yes	Yes	Yes

2.4. Infrastructure spending vs infrastructure use

As noted, New Zealand households pay for their infrastructure use in a range of different ways, including through rates, general taxation, fixed charges and variable charges.

A key implication of this use of multiple funding streams is that for many types of infrastructure, the amount an individual is required to pay to use the service will only partly relate to their level of use of it. Take the funding of New Zealand's road network. For vehicles that are covered by road user charges (RUC), the amount each owner is required to pay is almost perfectly related to the distance they have travelled, or intend to travel. However, that is not true for the amount an individual is required to pay for roads via local government rates. For that funding stream it is home ownership, and the value of each home, that determines how much each household is required to pay, not the distance they have travelled.

While the funding arrangement for our roads offers perhaps the best example of funding sources that do not relate to levels of use, a degree of divergence exists for most infrastructure types. For example, even for electricity, which is provided by independent, profit motivated firms, users tend to be charged through a mix of fixed daily charges, and variable used-based charges.

This means that when undertaking analysis of this nature, it is critically important to be clear whether the overarching focus is on infrastructure spending, or infrastructure use. While both spending and use are of potential policy interest, in this research note our focus is on determining the factors that affect the amounts that households *spend* on infrastructure.

That focus on spending, not use, has influenced the choice of variables we have included in our models.

One of the key variables included in our models, for example, shows the amount each household spends on rates (or is estimated to contribute towards the owner's rates bill through their rent payments, for renting households). That variable is highly unlikely to have been found to have a statistically significant impact on infrastructure *use*, if that had been our focus.

3. Results

As noted, the purpose of the research presented in this research note is to:

- identify as many of the factors that drive differences in household infrastructure spending as possible
- assess the size of the effect of each factor
- explain as much of the total observable variation in household infrastructure spending as possible given the available data.

Our results in each of those areas is discussed in the subsections below.

3.1. Overall level of explanatory power achieved

Table 2 below shows the maximum R-squared value we were able to achieve for each model. The R-squared statistic for a model indicates how much of the overall level of observable variation in infrastructure spending was able to be explained. The statistic can range from 0-1. A result of 0 means that the model is unable to explain any of the variability of the dependent variable, whereas a value of 1 means that the model explains all of that variability. The R-squared statistic for model 2 of 0.6989, for example, means that that model was able to explain close to 70% of the observable variation in household infrastructure spending.

Table 2: Overall explanatory power of each model

	Model 1	Model 2	Model 3	Model 4
R-squared value (overall explanatory power)	0.6428	0.6989	0.6978	0.6758

As can be seen from the table, the maximum explanatory power we were able to achieve in each of our four models ranged from 64% to almost 70%.

The two models with the highest level of fit are models 2 and 3. Those models both use total expenditure in dollar terms (not as a proportion of income) as the dependent variable, and include expenditure on non-capital items as an explanatory variable. We view these as the best two models, and would recommend that future users of our results use the parameters from them, rather than models 1 or 4.

While further work could usefully be undertaken to attempt to find additional variables that increase this overall explanatory power further, we view an R-squared statistic of 0.7 as a credible and useful start. In our view it is certainly sufficiently strong to allow these results to be used to inform policy making.

3.2. List of variables that were found to influence infrastructure spending

As discussed in the methodology section, we used a multi-step process to identify as many variables as possible that influence levels of household infrastructure spending. Through that process we derived a set of between 30 and 36 variables (including dummy variables) to include in each model.

A list of the variables we have included in our models as a result of that process is provided in Table 3 below. The relevant coefficient from model 2 is also shown, along with an indicator showing whether the variable was included in all of the four models, or only a subset.

Table 3: List of variables included in our models

Variable description	Coefficient (from model 2)	Included in all models?
(Log) Income after-tax	0.027	N
(Log) Expenditure non-capital after-tax	0.258	N
Household owns private vehicle (Y/N)	0.623	Y
(Log) Expenditure on rates	0.163	Y
Number of non-working adults in household	0.135	Y
Number of working adults in household	0.15	Y
Number of bedrooms in dwelling	0.029	Y
Household is renting (Y/N)	-0.032	Y
Number of dependent children in household	0.01	Y
X Coordinate of household location	-0.030	Y
Y Coordinate of household location	0.019	Y
Household location 'type': rural area	0.038	N
Household location 'type': Low density part of metropolitan area	0.053	Y
Household location 'type': Medium to low density part of metropolitan area	0.028	N
Household location 'type': Medium density part of metropolitan area	0.025	N
Household location 'type': Medium to high density part of metropolitan area	0.027	N
Household located in Far North District	-0.207	Y
Household located in Whangarei District	-0.146	Y
Household located in Kaipara District	-0.205	Y
Household located in Hauraki District	-0.099	N
Household located in Waikato District	-0.113	Y
Household located in Hamilton City	-0.103	Y
Household located in Tauranga City	-0.07	Y
Household located in Whanganui District	-0.145	Y
Household located in Manawatu District	-0.134	Y
No reported difficulty paying bills	-0.043	Y

Late paying bills more than once across multiple bill types (utilities, car, and rent/mortgage)	-0.061	N
Household not crowded (Y/N)	-	N
Dwelling 'non-private' (Y/N)	-0.045	Y
Household located in a meshblock with deprivation index of 2	-0.09	Y
Household located in a meshblock with deprivation index of 3	-0.082	Y
Household located in a meshblock with deprivation index of 4	-0.066	Y
Household located in a meshblock with deprivation index of 5	-0.089	Y
Household located in a meshblock with deprivation index of 6	-0.071	Y
Household located in a meshblock with deprivation index of 7	-0.069	Y
Household located in a meshblock with deprivation index of 8	-0.074	Y
Household located in a meshblock with deprivation index of 9	-0.086	Y
Household located in a meshblock with deprivation index of 10	-0.096	Y

In turn, a list of the variables we investigated but ultimately excluded from all of our models is provided in the technical notes.

3.3. Interpretation

It should be emphasised that we converted all financial variables in our models (i.e. those expressed in absolute dollar terms, such as after tax income) into logs before carrying out our analysis. That is a common approach to take when using financial data, as financial variables tend to increase exponentially over time, rather than linearly. All other non-financial variables, such as the number of working adults in a household, were left in their original form.

While a common approach to take, care is needed to interpret the results of models such as these which include a mix of logged and un-logged independent variables.

For the variables that have been logged, such as income after tax, the coefficients shown in the table above can be interpreted as meaning that *'a 1% increase in that variable is estimated to result in an x% increase in the dependent variable'* (expenditure on infrastructure services in this instance).

However, the coefficients of the variables that have not been logged need to be interpreted slightly differently. For those variables the coefficient shown in the table above should be interpreted as meaning that *'a one unit increase in the variable is estimated to increase expenditure on infrastructure by 100 × x%'*.

3.4. Decomposition analysis

We then assessed the relative explanatory power of different groups of variables by re-running model 2 with and without each group. The results of that decomposition analysis are shown in Table 4 below.

Table 4: Impact on explanatory power of dropping different categories of variables

Categories of variables removed	Overall R^2 Value	Difference in R^2 from baseline
Core model including all variables (our baseline)	0.699	-
Income and wealth and rates and FED/RUC variables removed	0.429	-0.270
Rates and FED/RUC payments removed	0.566	-0.133
Income and wealth variables removed	0.624	-0.075
Household composition variables removed	0.664	-0.035
Socioeconomic and geographic variables removed	0.689	-0.010
Geographic location variables removed	0.690	-0.009
Socioeconomic variables removed	0.694	-0.005
Household size and type variables removed	0.695	-0.004

These findings show that the group of variables with the largest explanatory power is the group that influences the amounts each household pays in rates, fuel excise duty (FED) and road user charges (RUC). The variables included in that category are income after tax, expenditure on non-capital items, household expenditure on rates, and a dummy variable indicating whether the household owns at least one car.

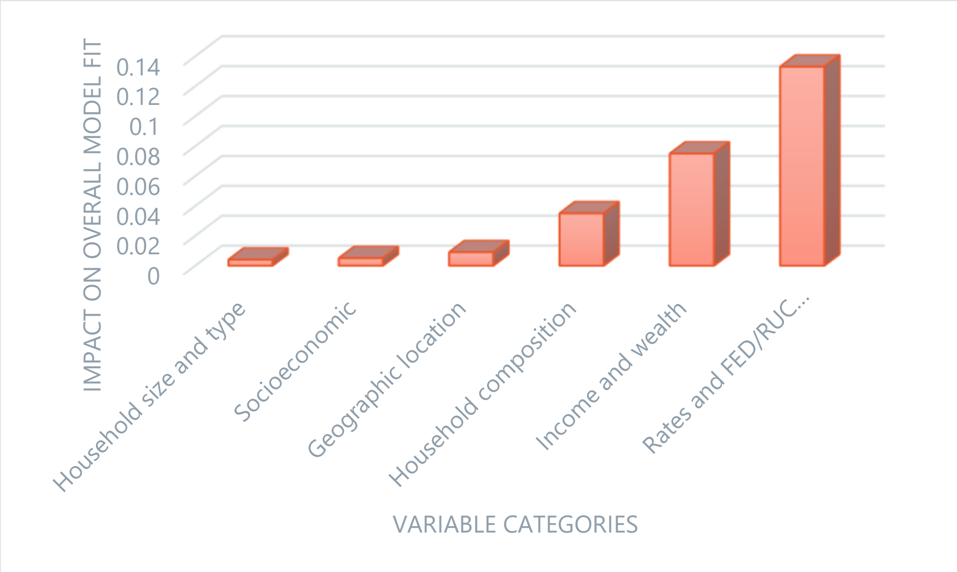
Removing that group of variables reduced the overall explanatory power of our model from almost 70% to 56.6%; a difference of 13.3 percentage points. The importance of this group of variables is not surprising, as the variables in it directly determine the amounts each household is required to pay for their water and roading services, rather than indirectly estimating levels of infrastructure use, as the remaining groups of variables do.

The next most important group of variables relates to household income (both current and permanent income). They explain roughly 7.5 percentage points of the overall explanatory power of model 2. This again is not surprising, as we would expect both current and permanent income to be strongly related to each household's level of infrastructure consumption.

After that, the explanatory power of the remaining groups of variables is much smaller. Household composition variables – showing the number and type of household members – are the next most significant, explaining roughly 3.5% of the overall variation in spending. Geographic location variables are next, explaining 0.9%. Followed by socio-economic variables (0.5%) and variables relating to household size and type (0.4%).

These relative contributions to the overall explanatory power of our model are shown diagrammatically in Figure 3 below.

Figure 3: Contribution of different variable categories to the model's overall fit



3.5. Discussion

In these last subsections we then discuss our findings for each of these groups of variables in more detail.

3.5.1. Income and wealth variables

We know from earlier work that household income is an important predictor of household infrastructure spending. However, that earlier work did not look at the question of whether it is current after-tax income that has the biggest impact on infrastructure spending, or permanent income.

This supplementary analysis confirmed the earlier result that higher income households spend more in dollar terms on infrastructure, but less as a proportion of their overall after-tax income. However, it extended that original analysis by looking at the impact of including a proxy for permanent income instead of, or in addition to the household's current after tax income.

We found that our models' overall explanatory power was increased by including permanent income in the model. Indeed, when both income variables are included, the impact of permanent income (at 0.258 for model 2) is considerably higher than that of current income (0.027 in model 2).

Another way of considering the relative impact of current and permanent income is to compare the overall explanatory power of model 2 (that includes both current after-tax income and permanent income) and model 3 (which only includes permanent income). The overall explanatory power of model 2 (at 69.89%) is only slightly higher than that of model 3 (at 69.78%).

These two results lead us to conclude that a household's expected long term average future income (or permanent income) is a considerably more important predictor of infrastructure spending than current income. In other words, the permanent income hypothesis appears to apply in this instance.

3.5.2. Household composition variables

Several of the variables included in our models relate to the number and nature of the members in each household[4]. We found that the greater the number of occupants in a household, the higher the infrastructure spending. The type of household member mattered too. Additional working adults have a (modestly) larger effect on infrastructure expenditure than additional non-working ones. Similarly, dependent children have a smaller effect on infrastructure spending than adults.

These findings are largely intuitive.

3.5.3. Dwelling size and type variables

We also found that the size and nature of the dwelling a household lives in impacts on their infrastructure expenditure.

A higher number of bedrooms (the best proxy we could find for overall dwelling size) leads to higher levels of expenditure. That makes intuitive sense, as bigger houses are more expensive to run.

In turn, we found that occupants of 'non-private' dwellings spend (slightly) less on infrastructure. Non-private dwellings are those where some or all of the living spaces are open to members of the public. This includes institutions like retirement villages, boarding houses, and motels. The negative relationship between living in a non-private dwelling and infrastructure expenditure has several possible explanations. One is that some infrastructure costs (such as heating and lighting) are paid for by the managers of the institution, and recouped through each occupant's rental payment. A second explanation is that the occupants of such dwellings are experiencing slightly higher levels of disadvantage than their incomes suggest, and that as with income, this higher level of disadvantage leads to lower levels of spending. Lastly, a third possibility is that the private living spaces are smaller than in private dwellings, and therefore that this variable is picking up an additional dwelling size effect. A clear interpretation of this result can only be achieved through further analysis.

We also found that households in uncrowded dwellings had slightly lower expenditure levels than those in crowded dwellings. A possible explanation of this result is that crowded dwellings occur mostly in high-cost housing markets, and that this variable is picking up the additional level of disadvantage that comes from facing high housing costs on a relatively low income. However, further work is again required to reach a clear interpretation of this result.

3.5.4. Other socioeconomic variables

Several additional socioeconomic variables remained significant in our models after the inclusion of income and wealth.

Renting households spend less on infrastructure than households that own their dwelling, even once the impact of lower incomes is taken into account. Similarly, the deprivation index of the meshblock a household is located in also affects infrastructure expenditure, with households located in more deprived areas spending less on infrastructure.

That is intuitively what one might expect, given the broader relationship between income and infrastructure expenditure. However, these deprivation dummy variables are picking up an additional effect over and above the impact of lower incomes which is not easy to explain. It could be that the most deprived areas have less good access to public transport and/or services such as supermarkets. However, this hypothesis does not fully explain why the negative relationship holds across all deciles, rather than just applying to the most deprived ones. By way of example, it does not explain why

[4] They are: the number of non-working adults in a household; the number of working adults in a household; and the number of dependent children in a household.

households located in the second least deprived decile of meshblocks (decile 2) would spend less than those in the least deprived areas (decile 1). This finding therefore requires further investigation.

A self-reported difficulty in paying bills also showed up as significant in our model. However, the results are hard to interpret. We created a categorical variable for this measure with 7 possible entries, ranging from no difficulty paying any type of bill, through to multiple instances of difficulty paying multiple types of bill (including utilities, car-related and housing-related bills). Only the two dummies at either end of that spectrum showed up as significant in our model, which is not surprising. What is surprising, however, is that the sign of the co-efficient of both of those dummies was negative. We have no credible explanation for this result and would recommend that it be treated with caution.

3.5.5. Factors affecting rates and road user charges / fuel excise duty

Again, the objective of this analysis is to predict infrastructure *expenditure*. As noted, while the two variables are related, higher levels of infrastructure spending will often only be partly correlated with higher levels of infrastructure use.

Where that is the case, the variables showing up as significant in our models will sometimes relate to the *design* of the underlying funding mechanisms. That is the case for the two variables discussed in this sub-section; a dummy variable identifying whether or not the household owns at least one car, and the dollar value of each household's expenditure on rates.[5]

We would expect these two variables to be significant in explaining infrastructure expenditure. It is only car owners that pay vehicle registration and licencing fees, and mostly only car owners that buy petrol (the price of which includes the fuel excise duty used to fund the maintenance and improvement of the roading network). Similarly, those households that pay the most in rates will be meeting a higher proportion of the cost of water services and local roads, again regardless of their level of use of those services.

The fact that both of these variables were found to be highly significant in our models is therefore not surprising. But care must be taken when interpreting these findings to be clear that their impact stems from the way that our roads are funded, rather than through an impact on the levels of household use of our roading system.

3.5.6. Location variables

We found several variables relating to the geographical location of each household to be statistically significant in our models.

Specific TLA each household is located in

We tested dummy variables for each TLA in New Zealand that a household can be located in. Only 9 TLAs showed up as having a significant impact on infrastructure spending: The Far North District; Whangarei District; Kaipara District; Hauraki District; Waikato District; Hamilton City; Tauranga City; Whanganui District and Manawatu District. All of these TLAs had a negative co-efficient. In other words, being located in one of TLAs lead to households spending less on infrastructure than equivalent households elsewhere.

We do not have a clear explanation for lower expenditure in these particular councils. The Far North and Kaipara districts both receive relatively high rates of financial assistance for local roads from NZTA, which may provide part of the answer. However, there are other TLAs that receive at least as high or higher rates of assistance which did not show up as significant. While all of these TLAs are in the North

[5] We calculated an implied level of rates for households in rental accommodation, as over the longer term we would expect housing related costs such as rates and home maintenance to be passed on to renters rather than be born by the owner.

Island, this finding is unlikely to be climate related as Auckland Council did not show up as significant. Further investigation is required to better understand this finding.

Household location type

We have also categorised each dwelling in our data set as being located in one of the following 'types' of area:

- rural
- small regional centre
- medium regional centre
- large regional centre
- metropolitan area with low population density
- metropolitan area with moderately low population density
- metropolitan area with medium
- metropolitan area with moderately high population density
- metropolitan area with high population density

By including household location types in this way, we are hypothesising that there are common factors that influence infrastructure spending in each area. Public transport, for example, is mostly only available in New Zealand's larger towns and cities. So, we would expect households in rural areas or small regional centres to spend less of passenger transport. Conversely, people living in those areas may need to drive further when travelling to work or school, so may spend more on private vehicle transport.

Dummy variables for each of these 'location types' were also included in the analysis reported in our earlier research insights report 'How much do we pay for infrastructure'. Unhelpfully, the results reported in this research note are quite different to those from the earlier report. Different location types were found to be significant across the two reports, and in some instances where a specific location type was found to be significant in both papers, the signs of the coefficients are different. We cannot explain these differences and would suggest that considerable caution is exercised when interpreting these particular results.

X and Y coordinates of household location

The most precise household location variables available in the IDI are the X and Y coordinates of the centre of the meshblock that each dwelling is located in. These X and Y coordinate variables express location as the number of meters a meshblock is away from a reference point to the south-west of New Zealand off Antarctica. The Y coordinate shows the number of metres north of that point and the X coordinate shows the number of metres to the east.

We have included these X and Y location coordinates along with some territorial authority areas as both sets of location variable show up as significant in our models, and appear to be picking up different effects.

Both sets of coordinates were found to be highly significant in our models. However, the signs of the coefficients are surprising. Our hypothesis in including these two variables was that both coordinates would pick up the effects of the colder climate in the South Island. (Because the North Island mostly lies to the west of the South Island the X and Y variables will both differentiate between the two islands to some degree). However, we found that infrastructure expenditure increases slightly the further north a household is located, and decreases slightly the further east it is.

We are unclear how to interpret this finding. One possibility we investigated was that these coordinate variables were picking up areas of high or low income that weren't captured by our other geographic location variables. We tested that by using different polynomials of the coordinate variables as our explanatory variables. If we were picking up particular geographic areas of high or low expenditure we would have expected that converting those variables into polynomials would result in increases in their statistical power. However, their statistical power was unchanged at the level of 3 decimal places. The interpretation of these two variables in our models therefore also requires further analysis.

Technical notes: Full regression results

Table 5: OLS models explaining variations in total infrastructure expenditure.

	(Model 1)	(Model 2)	(Model 3)
	Dependant variable		
Explanatory variables	(Log) All expenditure	(Log) All expenditure	(Log) All expenditure
(Log) Income after-tax	0.144***	0.027***	
	(0.007)	(0.007)	
(Log) Expenditure non-capital after-tax		0.258***	0.268***
		(0.008)	(0.007)
Household owns private vehicle (Y/N)	0.696***	0.623***	0.624***
	(0.014)	(0.013)	(0.013)
(Log) Expenditure on rates	0.198***	0.163***	0.165***
	(0.019)	(0.016)	(0.016)
Number of non-working adults in household	0.155***	0.135***	0.139***
	(0.005)	(0.005)	(0.005)
Number of working adults in household	0.191***	0.15***	0.161***
	(0.006)	(0.005)	(0.005)
Number of bedrooms in dwelling	0.037***	0.029***	0.029***
	(0.005)	(0.004)	(0.004)
Household is renting (Y/N)	-0.019**	-0.032***	-0.036***
	(0.008)	(0.007)	(0.007)
Number of dependent children in household	0.029***	0.01***	0.012***
	(0.003)	(0.003)	(0.003)
X Coordinate of household location	-0.029***	*-0.030****	-0.031***
	(0.003)	(0.*003*)	(0.003)
Y Coordinate of household location	0.019***	*0.019****	0.019***
	(0.002)	(0.*002*)	(0.001)
Household located in rural area		0.038***	
		(0.012)	
Household located in Metro and PDQ 1 (low density) area	0.065***	0.053***	0.037**
	(0.016)	(0.015)	(0.015)
Household located in Metro and PDQ 2 (medium to low density) area		0.028***	
		(0.011)	

(Table 5 continued)

	(Model 1)	(Model 2)	(Model 3)
		Dependant variable	
Explanatory variables	(Log) All expenditure	(Log) All expenditure	(Log) All expenditure
Household located in Metro and PDQ 3 (medium density) area	0.024**	0.025***	
	(0.01)	(0.01)	
Household located in Metro and PDQ 4 (medium to high density) area	0.029***	0.027***	
	(0.009)	(0.009)	
Household located in Far North District	-0.183***	-0.207***	-0.189***
	(0.045)	(0.041)	(0.04)
Household located in Whangarei District	-0.162***	-0.146***	-0.154***
	(0.027)	(0.025)	(0.025)
Household located in Kaipara District	-0.172***	-0.205***	-0.182***
	(0.045)	(0.04)	(0.039)
Household located in Hauraki District	-0.136***	-0.099**	
	(0.05)	(0.047)	
Household located in Waikato District	-0.117***	-0.113***	-0.106***
	(0.032)	(0.029)	(0.029)
Household located in Hamilton City	-0.1***	-0.103***	-0.094***
	(0.019)	(0.018)	(0.017)
Household located in Tauranga City	-0.057***	-0.07***	-0.056***
	(0.018)	(0.018)	(0.017)
Household located in Whanganui District	-0.133***	-0.145***	-0.156***
	(0.024)	(0.023)	(0.023)
Household located in Manawatu District	-0.172***	-0.134***	-0.129***
	(0.046)	(0.044)	(0.044)
No reported difficulty paying bills	-0.032***	-0.043***	-0.041***
	(0.01)	(0.009)	(0.009)
Late paying bills more than once across multiple bill types (utilities, car, and rent/mortgage)	-0.086***	*-0.061***	-0.065**
	(0.032)	*(0.029)*	(0.029)
Household <u>not</u> crowded (Y/N)	-0.184***		-0.103**
	(0.05)		(0.042)
Dwelling 'non-private' (Y/N)	-0.04***	*-0.045****	-0.046***
	(0.009)	*(0.009)*	(0.009)

(Table 5 continued)

	(Model 1)	(Model 2)	(Model 3)
		Dependant variable	
Explanatory variables	(Log) All expenditure	(Log) All expenditure	(Log) All expenditure
Deprivation index for meshblock household is located in: Decile 2 (2nd least deprived)	-0.053***	*-0.09****	-0.081***
	(0.015)	*(0.013)*	(0.013)
Deprivation index for meshblock household is located in: Decile 3 (3rd least deprived)	-0.059***	*-0.082****	-0.076***
	(0.013)	*(0.012)*	(0.012)
Deprivation index for meshblock household is located in: Decile 4	-0.048***	*-0.066****	-0.059***
	(0.014)	*(0.013)*	(0.013)
Deprivation index for meshblock household is located in: Decile 5	-0.074***	-0.089***	-0.086***
	(0.013)	(0.012)	(0.012)
Deprivation index for meshblock household is located in: Decile 6	-0.072***	-0.071***	-0.068***
	(0.013)	(0.012)	(0.012)
Deprivation index for meshblock household is located in: Decile 7	-0.068***	-0.069***	-0.068***
	(0.014)	(0.012)	(0.012)
Deprivation index for meshblock household is located in: Decile 8	-0.074***	-0.074***	-0.071***
	(0.013)	(0.012)	(0.012)
Deprivation index for meshblock household is located in: Decile 9	-0.107***	-0.086***	-0.088***
	(0.013)	(0.012)	(0.012)
Deprivation index for meshblock household is located in: Decile 10	-0.113***	-0.096***	-0.099***
	(0.012)	(0.011)	(0.011)
Constant	4.632***	3.629***	3.82***
	(0.119)	(0.107)	(0.108)
R2	0.6428	0.6989	0.6978
Observations	9,936	9,924	9,924

Table 6: OLS models explaining variations in infrastructure expenditure as a proportion of income

	(Model 4)
	Dependent variable
Explanatory variables	All expenditure as % of after tax income
(Log) Income after-tax	-0.22*** (0.003)
(Log) Expenditure non-capital after-tax	0.065*** (0.002)
Household owns private vehicle (Y/N)	0.108*** (0.003)
(Log) Expenditure on rates	0.035*** (0.004)
Number of non-working adults in household	0.027*** (0.001)
Number of working adults in household	0.03*** (0.002)
Number of bedrooms in dwelling	0.009*** (0.001)
Household is renting (Y/N)	-0.011*** (0.002)
Number of dependent children in household	
X Coordinate of household location	-0.007*** (0.001)
Y Coordinate of household location	0.005*** (0.000)
Household located in rural area	0.013*** (0.003)
Household located in Metro and PDQ 1 (low density) area	0.015*** (0.004)
Household located in Metro and PDQ 2 (medium to low density) area	0.007** (0.003)
Household located in Metro and PDQ 3 (medium density) area	
Household located in Metro and PDQ 4 (medium to high density) area	0.008*** (0.002)
Household located in Far North District	-0.031** (0.014)
Household located in Whangarei District	-0.036*** (0.007)
Household located in Kaipara District	-0.043*** (0.014)

(Table 6 continued)

Explanatory variables	(Model 4) Dependent variable All expenditure as % of after tax income
Household located in Hauraki District	
Household located in Waikato District	-0.028*** (0.007)
Household located in Hamilton City	-0.023*** (0.005)
Household located in Tauranga City	-0.018*** (0.005)
Household located in Whanganui District	-0.035*** (0.006)
Household located in Manawatu District	-0.023*** (0.008)
No reported difficulty paying bills	-0.008*** (0.003)
Late paying bills more than once across multiple bill types (utilities, car, and rent/mortgage)	
Household not crowded (Y/N)	
Dwelling 'non-private' (Y/N)	-0.01*** (0.002)
Deprivation index for meshblock household is located in: Decile 2 (2nd to least deprived)	-0.01*** (0.004)
Deprivation index for meshblock household is located in: Decile 3 (3rd to least deprived)	-0.009*** (0.003)
Deprivation index for meshblock household is located in: Decile 4	-0.009*** (0.003)
Deprivation index for meshblock household is located in: Decile 5	-0.019*** (0.003)
Deprivation index for meshblock household is located in: Decile 6	-0.014*** (0.003)
Deprivation index for meshblock household is located in: Decile 7	-0.012*** (0.003)
Deprivation index for meshblock household is located in: Decile 8	-0.014*** (0.003)
Deprivation index for meshblock household is located in: Decile 9	-0.021*** (0.003)

(*Table 6 continued*)

	(Model 4)
	Dependent variable
Explanatory variables	All expenditure as % of after tax income
Deprivation index for meshblock household is located in: Decile 10	-0.026***
	(0.003)
Constant	1.354***
	(0.03)
R^2	0.6758
Observations	9,924

Technical notes: Variables excluded from our models

The table below lists those variables that we investigated as part of our analysis, but ultimately excluded from all models given their lack of explanatory power when combined with other more powerful variables.

Table 7: Variables excluded due to lack of significance once combined with other more powerful explanatory variables

Variable
Dwelling characteristics
Number of storeys
Capital value of dwelling
Capital value of underlying land
Location of property (SA1 ID and SA2 ID)
Regional council property is located in
SA2 population density
Occupant characteristics
Number of families in dwelling
Highest qualification of primary earner
Expenditure on rent
Household income class
$ amount of outstanding mortgage
Total expenditure on housing related expenses
Household tenure code[6]
DEP-17 material wellbeing score of household

[6] Denotes whether the dwelling is fully or partly owned by one or more of the occupants, a family trust, or some other entity.