# The income elasticity of household infrastructure expenditure

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). *The income elasticity of household infrastructure expenditure*. 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' research programme that Te Waihanga has undertaken. 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 **Stats NZ** for providing access to the data that was used for the analysis – see IDI disclaimer above.

# Cut to the chase

The Infrastructure Commission is undertaking a series of interrelated analyses to improve our understanding of household spending on infrastructure services. Our first key piece of analysis was to estimate the amounts that households in New Zealand spend on infrastructure services such as roading, passenger transport, energy, water and telecommunications. The results of that work were published in mid-2023 in a research insights report[1] titled 'How much do we pay for infrastructure?'

This supplementary research note extends that earlier analysis. It provides updated estimates of the income elasticity of household infrastructure spending. (An income elasticity is a measure of the extent to which demand for a product responds to changes in the customer's income. The higher the income elasticity of demand for a product, the more the level of demand changes in response to changes in income.)

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

That earlier report also found that both the dollar amount spent on infrastructure, and the share of income spent, varies significantly in relation to a household's income. Households with lower incomes spend smaller amounts in dollar terms, but a higher proportion of their after-tax income. Figure 1 shows those results for each after-tax income quintile.

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


While not its primary purpose, that earlier report also included a preliminary estimate of the income elasticity of household infrastructure spending. It concluded that on average, a 1% increase in household income is associated with a 0.24% increase in overall infrastructure spending.

[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.

The purpose of this supplementary research note is to assess the relationship between household income and infrastructure spending in more depth. It provides updated income elasticity estimates that we consider to be more robust.

## Summary of findings

We have calculated income elasticity estimates for a total of 12 interrelated models. This relatively large number of models served two purposes. First, it provided a form of sensitivity testing, giving us more confidence in the robustness of our elasticity estimates. Second, it allowed us to derive a wider range of elasticity estimates that can be used in a wider range of potential policy situations.

The 12 models differ in the following ways:

- First, half of the models calculate elasticities with respect to *total* expenditure on infrastructure, while the other half calculate them with respect to *variable* expenditure.
- Second, for each form of the model we ran one version that included dummy variables for each household's 'location type' (which we hoped might capture differences in commuting distances and modes) and one version without them.
- Third, for each form of the model we ran one version with just current after-tax income included as an explanatory variable, one version just with a proxy for permanent income included, and one with both forms of income included.

The income elasticities we calculated from each model are shown in Table 1, for the models with total expenditure as the dependent variable, and Table 2 for those with variable expenditure as the dependent variable.

*Table 1: Estimated income elasticities for total infrastructure spending*

| | Dependent variable | | | | | |
|---|---|---|---|---|---|---|
| | (Log) All expenditure | (Log) All expenditure | (Log) All expenditure | (Log) All expenditure | (Log) All expenditure | (Log) All expenditure |
| **Explanatory Variables** | (Model 1) | (Model 2) | (Model 3) | (Model 4) | (Model 5) | (Model 6) |
| (Log) Income after-tax | 0.278*** | 0.074*** | | 0.278*** | 0.074*** | |
| | (0.006) | (0.007) | | (0.006) | (0.007) | |
| (Log) Expenditure on non-capital items | | 0.373*** | 0.408*** | | 0.373*** | 0.410*** |
| | | (0.008) | (0.007) | | (0.008) | (0.007) |
| **Model fit and description** | | | | | | |
| $R^2$ | 0.4331 | 0.5480 | 0.5439 | 0.4320 | 0.5477 | 0.5435 |
| Includes control variables for household location type? | Yes | Yes | Yes | No | No | No |

Notes:

- The number of asterisks following each estimated coefficient shows the level of significance of each coefficient (in effect, how confident we can be that the true coefficient is different from zero). Three asterisks denote a 99% probability that the true coefficient is different to zero. In turn, two asterisks denote a 95% probability, and one denotes a 90% probability.
- The figures in brackets show the standard error of each coefficient.

*Table 2: Estimated income elasticities for variable infrastructure spending*

| | Dependent variable | | | | | |
|---|---|---|---|---|---|---|
| | (Log) Variable expenditure | (Log) Variable expenditure | (Log) Variable expenditure | (Log) Variable expenditure | (Log) Variable expenditure | (Log) Variable expenditure |
| **Variable** | (Model 7) | (Model 8) | (Model 9) | (Model 10) | (Model 11) | (Model 12) |
| (Log) Income after-tax | 0.356*** | 0.057*** | | 0.360*** | 0.058*** | |
| | (0.012) | (0.013) | | (0.012) | (0.013) | |
| (Log) Expenditure on non-capital items | | 0.551*** | 0.579*** | | 0.554*** | 0.582*** |
| | | (0.012) | (0.01) | | (0.012) | (0.010) |
| **Model fit and description** | | | | | | |
| $R^2$ | 0.3105 | 0.4150 | 0.4139 | 0.3072 | 0.4128 | 0.4118 |
| Includes control variables for household location type? | Yes | Yes | Yes | No | No | No |

As noted, these elasticities measure how sensitive infrastructure spending is to changes in income. The numbers preceding the asterisks in the tables above show the elasticity estimates for current after-tax income and expenditure on non-capital items, which we have included as a proxy for permanent income. Each estimate can be interpreted as meaning that 'a 1% increase in income is associated with an X% increase in infrastructure spending'. Elasticities less than 1 suggest that infrastructure spending will increase more slowly than income, whereas those higher than 1 suggest that spending will increase faster than income.

There are several aspects of the results shown in the tables above that we would draw attention to.

First, our estimated elasticities for current income range from 0.074 to 0.278 in relation to total expenditure on infrastructure, and from 0.074 to 0.360 in relation to variable expenditure. As all our estimated elasticities are less than 1, we can conclude that while higher incomes increase infrastructure expenditure, that effect is less than one-to-one. These estimates are slightly higher than the preliminary income elasticity estimate of 0.24 we reported in our earlier report 'How much do we pay for infrastructure?' (That model only included current income, so is best compared to model 1 from this research note, which has an elasticity of 0.278.)

Second, the estimated elasticities for *permanent* income are consistently higher than those for *current* income. The elasticity for current after-tax income in model 1, for example, is 0.278. Whereas the estimate for permanent income in model 3 is 0.408. This is in line with our a priori expectations. We would expect current income to be more volatile than permanent income, and for households to 'look through' those variations in current income to some degree.

Third, our estimated elasticities for both current and permanent income are lower in models where both types of income are included relative to models where only of them is included. By way of example, the elasticity for current income in model 1, which does not include permanent income, is 0.278. That drops to 0.074 in model 2 which includes both forms of income. However, the size of that drop is smaller for permanent income. Model 3, for example, which only includes permanent income, has an elasticity of 0.408 for permanent income. In contrast, model 2, which includes both current and permanent income, has an elasticity for permanent income of 0.373.

Fourth, our estimated elasticities relating to variable expenditure on infrastructure are higher than those relating to total expenditure. Model 7 for example, which relates to variable expenditure, has an elasticity for current income of 0.356. In contrast, model 1, which relates to total expenditure, has an elasticity for current income of 0.278.

Fifth, the model 'fit' ($R^2$ figure in the tables) is noticeably higher for the models including permanent income than it is for those only including current income (compare models 1 and 3, for example, which have $R^2$ figures of 0.4331 and 0.5439, respectively).

Lastly, the inclusion of the household location type variables had little or no impact on our elasticity estimates. The co-efficient for current income in models 1 and 4, for example, are identical at the level of 3 decimal places.

In turn, there are several wider policy implications of these results that we would highlight.

Perhaps most importantly, it will be important for future users of these elasticity estimates to be clear in advance what 'type' of income elasticity is most appropriate for their intended use. The two key decisions will be whether to use a variable or total expenditure elasticity, and whether to use a current income or permanent income elasticity. Both decisions will have a meaningful impact on the size of elasticity, and therefore the results of any future policy analysis.

With that said, because permanent income appears to be a more important determinant of infrastructure expenditure than current income, we would suggest future users of these results use our permanent income elasticities unless they are clearly inappropriate for the specific use that is intended. Indeed, where sensible we would suggest the results from our models including *both* permanent and current income (models 5 and 11).

In terms of the impact of the inclusion of household location type variables, our results are inconclusive. We would therefore recommend *against* the use of the results from our models that include those household location type variables.

Lastly, we would note that these results could be used by policy advisors to forecast levels of future demand for infrastructure services well into the future, if combined with forecasts of population growth and GDP that Stats NZ and Treasury publish at regular intervals.

# Table of contents

# Introduction

The Infrastructure Commission is undertaking a series of interrelated analyses to improve our understanding of household spending on infrastructure services. Our first key piece of analysis was to estimate the amounts that households in New Zealand spend on infrastructure services such as roading, passenger transport, energy, water and telecommunications. The results of that work were published in mid-2023 in a research insights report[2] titled 'How much do we pay for infrastructure?'

This supplementary research note extends and improves on that earlier work. It provides updated estimates of the income elasticity of household infrastructure expenditure. An income elasticity is a measure of the extent to which demand for a product responds to changes in the customer's income. The higher the income elasticity of demand for a product, the more the level of demand changes in response to changes in income.

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

# Methodology

Our earlier report 'How much do we pay for infrastructure?' drew on data from Stats NZ's Household Economic Survey (HES). While the 'core' HES survey is run annually, detailed expenditure data is only collected every three years, and requires participating households to keep a careful diary of their spending over the course of a week. We used data from the five most recent 'waves' of that expenditure survey between 2006/07 and 2018/19. Slightly over 3,000 households, on average, were included in each survey wave.

## Nature of the data used

Broadly speaking, econometricians differentiate between three key 'types' of data. They are:

- Cross-sectional data: information on a number of different 'subjects' (such as individuals, households, or firms) at a single point in time. Data collected through a one-off survey of a selection of households in a city is a good example of cross-sectional data.
- Time series data: information on one subject at regular intervals over time. Data recording the number of staff employment by a firm over a 20-year period is a good example of time series data.
- Panel data: information on a number of subjects over multiple time periods. Data from the 'Dunedin Study', which collects information on the health, development and behaviour of the same group of around 1,000 people at regular intervals over the course of their lives, is a good example of panel data.

A key point to note about the data used for this analysis is that the selection of households in each 'wave' of the HES *changes* each time the survey is undertaken. Our five years of data therefore provides information on close to 16,000 different households across five years (rather than the same group of slightly over 3,000 households for each of those five years).

For that reason, the data we have used from the HES is *not* panel data. Rather we would refer to it as 'pooled, cross-sectional data'. (Pooled because it aggregates data from multiple surveys).

This characteristic of our dataset means we needed to adjust for 'group-specific' or fixed effects. By way of example, we know from our earlier work that the size and makeup of a household affects its level of infrastructure spending. Households without a working adult, for example, tend to spend less than those with a working adult. Similarly, households with dependent children spend more than households without children.

Because the households asked to participate in the expenditure module of the HES are selected randomly, the mix of households with different sizes and compositions will likely vary to some degree from survey wave to survey wave. We therefore needed to include 'controls' in our analysis to take account of this variation. In addition, we tested the need to include a dummy variable for each survey year, to allow for the possibility that there were additional 'between survey' effects that our controls were not adequately adjusting for.

The process we used to select those controls is discussed in more detail in the subsections below.

## Model development

For this supplementary analysis, we have considered a wider range of possible explanatory variables than was possible in the earlier 'How much do we pay for infrastructure?' report. We have also investigated a wider range of different possible model specifications.

We developed our statistical models through a multi-step process. First, we identified as many variables as possible in the HES and other linked IDI datasets that, a priori, seemed likely to have an effect on infrastructure spending. This became our variable 'longlist'.

Second, we assessed the explanatory power of each variable on the longlist by determining the simple correlation between it and infrastructure spending. This allowed us to discard any variables with little 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.

Third, we included all of the variables with meaningful explanatory power into an initial overall statistical model and assessed the overall explanatory power of that initial model.

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

Fifth, and noting that the primary objective of this analysis is to derive as accurate estimates as possible of the income elasticity of infrastructure spending, we developed conceptual 'causal models' setting out how we expect each variable to affect, or be affected by, infrastructure spending and after-tax income. This last step led us to exclude an additional set of variables that we felt were unlikely to be true 'confounders', which if included ran the risk of skewing our income elasticity estimates.

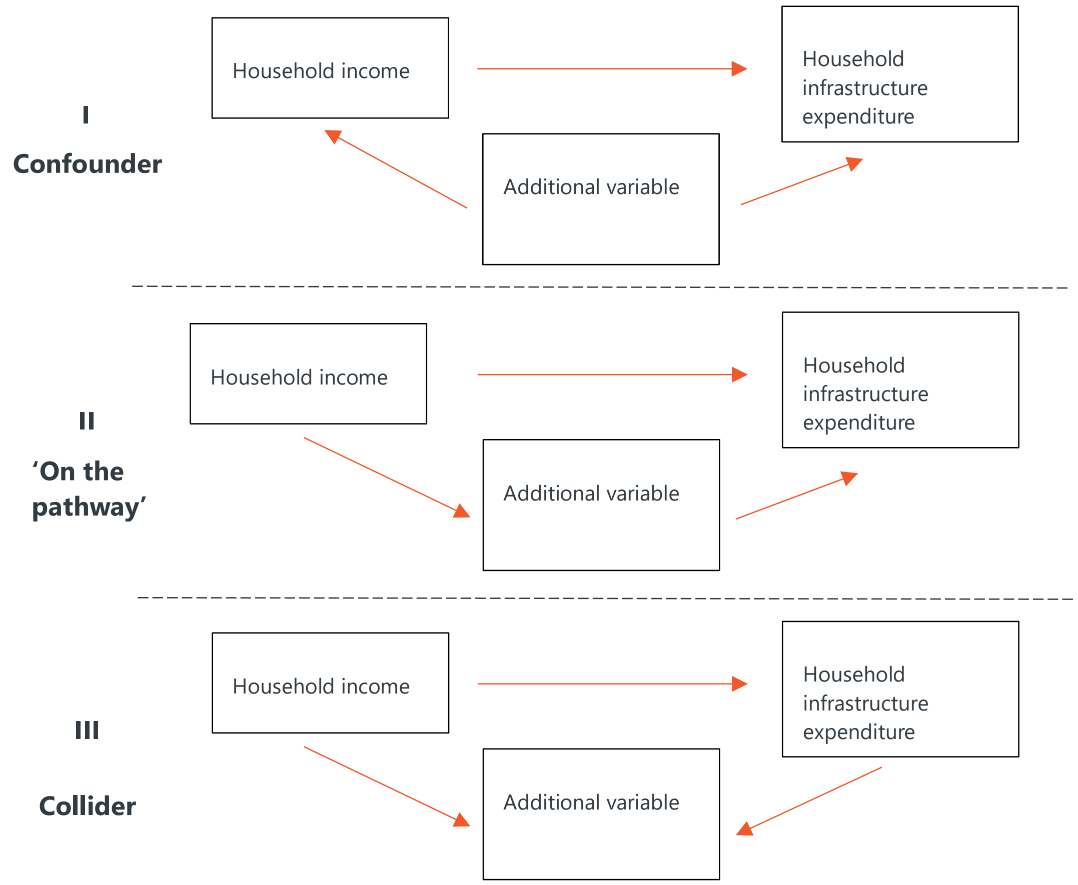
## Addressing potential confounding biases

The first four of the five steps outlined above are commonly used. However, step five – developing conceptual 'causal models' – is less common and warrants further explanation.

Again, our primary objective in undertaking this piece of analysis was to estimate the size of the impact of changes in household income on levels of household infrastructure spending. Where the goal is to estimate the impact (or 'coefficient') of a particular variable as accurately as possible, care is needed to avoid confounding biases.

To explain this, it is useful to start with a summary of the possible causal relationships each additional explanatory variable can have with our primary variable of interest (household income) and the outcome variable we are looking to better understand (household infrastructure spending). For any given additional variable there are three possible causal relationships as shown in Figure 2.

*Figure 2: Different possible causal relationships*




The first possible causal relationship is one where the additional variable impacts on both household income and infrastructure spending. That sort of variable is referred to as a 'confounder'. The second possible causal relationship is one where the additional variable is affected by household income, and in turn then affects infrastructure spending. A variable of that nature is referred to as being 'on the pathway' between the primary explanatory variable of interest and the dependent variable. Lastly, the additional variable can be impacted by household income and household expenditure, rather than having an impact on them. This third type of variable is referred to as a 'collider'.

Depending on which of these causal relationships is thought to exist for a variable, the consequences of including it in a regression model will differ. In the first example of a classical confounder, the accuracy of the income elasticity estimate will be improved by including the additional variable in the model.

In contrast, the additional variables in causal relationships II and III are not confounding factors and should not be included. Including such 'collider' and 'on the pathway' variables would either bias our income elasticity estimate, or make that estimate less efficient (i.e., subject to a larger confidence interval).[3]

There is no quantitative way of determining the direction of the causal relationship between different variables. Rather, researchers need to use their judgement to decide which ways the causal relationships are most likely to flow. Our hypothesised causal models are outlined in the section below.

[3] Readers wanting to understand the impact of these causal relationships in more detail can find more information here: How to control confounding effects by statistical analysis - PMC (nih.gov).

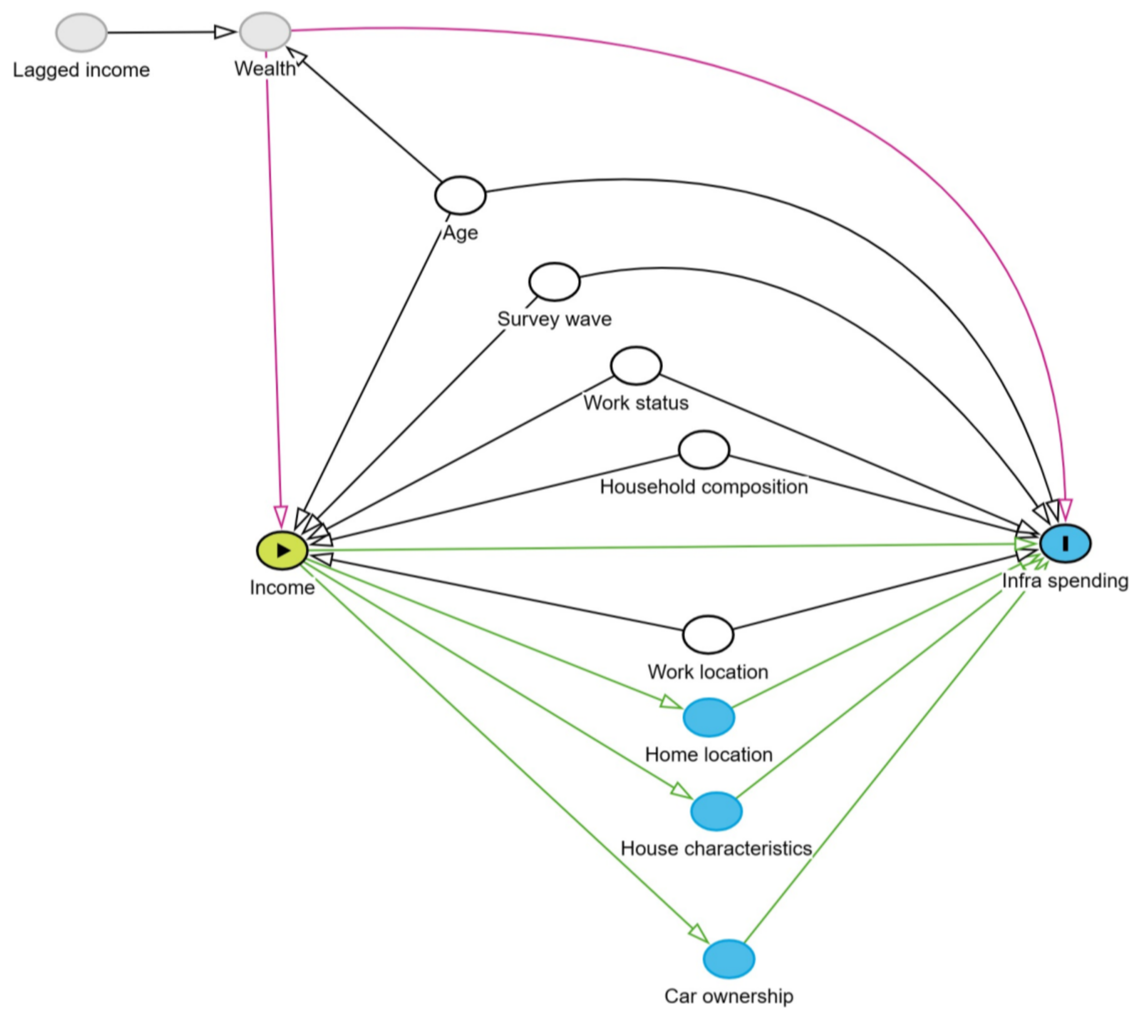
## Our hypothesised causal models

As the purpose of this analysis is to estimate the size of the impact of household income on household infrastructure expenditure, household income is our primary *explanatory* variable of interest, and household infrastructure spending is our *dependent* variable.

However, we would expect a range of other factors to also have an impact on household income and infrastructure spending. Key examples of those other factors are shown in the two causal models below.

Our first proposed causal model uses current after-tax income as the explanatory variable of interest and is shown in Figure 3. This model views wealth, age, the HES survey wave, household composition, the number of working adults in the household, and work location as influencing both income and infrastructure spending. As discussed, that makes them 'confounding variables', and we have therefore included them in our model. In contrast, home location, house characteristics and car ownership are seen as factors *on the pathway* between income and infrastructure spending. For the reasons discussed, those variables have therefore been excluded from our model.

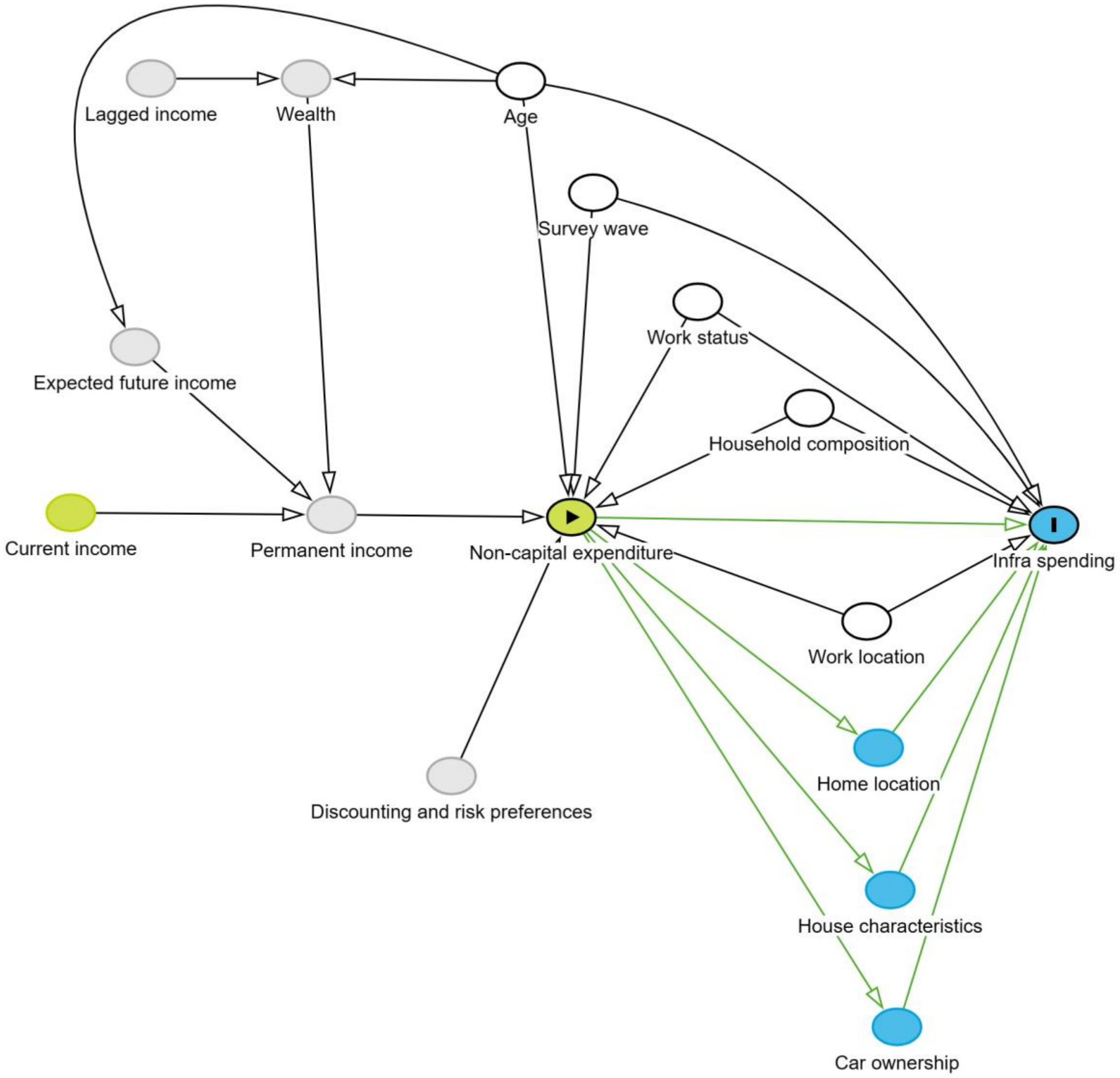
*Figure 3: Causal model with income after tax as explanatory variable of interest*




While this first causal model includes the effect of a household's wealth, we do not have access to suitable data on wealth and therefore cannot include it as a variable in our model. While unavoidable, the exclusion of household wealth is likely to bias our estimate of the relationship between income and infrastructure spending.

For that reason, we have developed a second causal model that expands on the relationship between wealth, current income, permanent income, and infrastructure spending. That second model is shown in Figure 4.

*Figure 4: Causal model with non-capital expenditure as explanatory variable of interest*




This second model posits that annual household expenditure on non-capital items is likely to be heavily influenced by current income and permanent income.[4] So long as wealth, expected future income and current income *only* affect infrastructure spending through this permanent income channel, we can use non-capital expenditure as a proxy for those unobservable variables.

We would note that neither of these conceptual models are perfect. For this reason, we have chosen to report our findings using both approaches.

[4] Permanent income is an economic concept that describes the average annual income an individual or household expects to receive in the long run. The permanent income hypothesis posits that people will spend money at a level that is consistent with their expected long-term average income, not their current after-tax income.

## More detail on permanent income

Permanent income is at best challenging to measure directly, and at worst impossible to measure. That is why the second model above includes each household's expenditure on non-capital items as a proxy measure for permanent income. The use of proxy measures is a well-established technique in econometrics, and expenditure on non-capital items is a widely used proxy for permanent income.

This approach is not without risk, however. As noted, the data in the expenditure module of the HES is collected by households keeping a detailed diary of their spending over the course of a week. Such expenditure surveys are complex to administer, so there is always a risk of some spending being missed or mis-categorised, resulting in a classical measurement error. A second source of error can arise from volatility in household's spending from week to week. Even if the reported 'snapshot' of expenditures is true for the week in question, the household's spending in that week may have been abnormally low or high, or atypical in nature.

Evidence from the literature suggests that where present, such errors are likely to lead to the impact of permanent income in the model in question being *underestimated*.[5]

We have no way of directly assessing how 'good' a proxy measure expenditure on non-capital items is in this instance. However, we have undertaken some high-level 'sense checks' to ensure that the relationship between expenditure on non-capital items and current income is broadly in line what we would expect.

We first looked at the strength of the relationship between current and permanent income. We found a Person's correlation of .6055 between the two variables,[6] which indicates a relatively strong, positive relationship between them. That is broadly what we would expect to see.

We then used OLS regression analysis to further investigate the relationship between the two variables using the following model:

$$Log(proxy\ for\ permanent\ income) = \alpha + \ \beta 1 \log(current\ after\ tax\ income) + \epsilon$$

Where: α = constant
β1 = coefficient of current after-tax income
ε = error term.

Our estimate of the coefficient for current after-tax income (β1) was 0.683, with an associated p-value of less than 0.001. This p-value confirms that the relationship between the two variables is statistically significant. In turn the coefficient β1 suggests that current income explains roughly 68% of the total observable variation in permanent income.

These results again are in line with what one would expect. In other words, we would expect current income to be statistically significant when used as a variable, and for it to explain much, but not all, of the variation in permanent income.

These 'sense checks' leads us to conclude that the use of expenditure on non-capital items as a proxy for permanent income is valid in this instance.

[5] See for example, John Gibson, Bonggeun Kim, 'How reliable are household expenditures as a proxy for permanent income? Implications for the income–nutrition relationship', *Economics Letters*, Volume 118, Issue 1, 2013, Pages 23-25.

[6] The Pearson's correlation coefficient is a statistical measure that quantifies the linear relationship between two continuous variables. It ranges from 1 (for a perfect positive correlation) to -1 (for a perfect negative correlation). A result of 0 indicates no linear relationship.

## Commuting patterns

One would expect, a priori, that the physical distance between a house, and the sites where the household's members work, would have an effect on transport expenditure. For that reason, we investigated the possibility of including data on workplace location in our analysis.

We ultimately concluded that the available data was not sufficiently robust for use in that way. The longitudinal business database (LBD) – a 'sister' database to the IDI that contains anonymised information about firms rather than people – does contain detailed employment data, including employer addresses. This data can be linked to the IDI through the Linked Employer-Employee Dataset (LEED).

However, there are several practical challenges in attempting to use LBD data to identify work locations. A key challenge relates to firms that operate from more than one physical site, such as supermarket chains. Where that is the case, the address data contained in the LBD is often for the firm's national or regional head office, not the actual site where each employee works. A second challenge arises from the fact that the LBD does not capture the extent to which individual employees work from home, or travel as part of their job. These practical challenges make it difficult to identify work locations with any accuracy.

We have therefore trialled a simpler approach in this analysis, looking at the 'type' of location that each household is situated in (referred to as 'household location type' variables for the rest of this research note). Combining the meshblock data from the HES with data sourced externally, we have been able to categorise all households as being situated 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 locations in this way, we are hypothesising that there are commonalities in the distance that people commute to work, as well as how they commute, in each type of area. Residents in rural areas, for example, may have longer commutes on average, and be less likely to be able to use public transport.

We cannot be certain, however, these location type variables will in practice capture differences in commuting distances and practices. For that reason, we have run all of our results both with and without their being included as explanatory variables.

## Variable or total expenditure

Our earlier report 'How much do we pay for infrastructure?' differentiated between infrastructure spending that is fixed and spending that is variable. Fixed expenditure refers to spending that does not change in line with changes in the level of a household's infrastructure use. The cost of licencing a car is a good example of a fixed cost, as people that drive their car a lot pay the same licence fee as light users. In turn, variable expenditure is that which *does* change in line with use. Expenditure on petrol is a good example of a variable cost, as people that drive their cars more will need to spend more on fuel.

We have assessed which explanatory variables are relevant, and estimated income elasticities, separately for both variable and total infrastructure expenditure. This should allow future users of the results of this work to select an income elasticity estimate that is most appropriate for their intended purpose.

## Choice of models

We estimated the income elasticity of infrastructure spending for a selection of 12 slightly different models. The different model specifications were designed to ensure that the estimates we have derived are useful in as wide a range of applications as possible, and to give us more confidence in the robustness of our estimates.

Half of our models investigated the relationship between income and *total* household infrastructure spending, while the other half looked at the relationship just with the *variable* component of infrastructure spending.

In turn, for both total and variable expenditure we treated household income in our model in three different ways:

- only including current after-tax income as an explanatory variable
- including both current income and expenditure on non-capital items (as a proxy for permanent income)
- only including a proxy for permanent income.

Lastly, we ran all of our models both with and without the household 'location type' variables discussed earlier in the 'commuting patterns' section. To recap, by including household location types we are hypothesising that there are commonalities in the distance that people commute to work, as well as how they commute, in each type of area. However, we cannot be certain these household location type variables will have the effect we are wanting. We have therefore run all of our results both with and without their inclusion.

Collectively these slightly different approaches led us to estimate 12 different models.

## Explanatory variables included in our models

As noted, the primary purpose of this analysis is to estimate the income elasticity of household infrastructure spending. In mathematical terms, the models we have used to calculate our elasticity estimates all take the form:

$$Log(infrastructure\ expenditure\ ) = \alpha + \ \beta 1 \log(after\ tax\ income) + \beta 2\ (other\ variables) + \epsilon$$

Where:

α = constant
β1 = income elasticity of infrastructure expenditure
β2 = coefficients for other explanatory variables
ε = error term.

Using the approach described in the methodology section, we derived a collection of six sets of explanatory variables to include in our models in addition to current and permanent income. A list and brief description of those six sets of variables is provided in Table 3.

Three of those sets of variables were not included in the analysis undertaken as part of the earlier work reported in the report 'How much do we pay for infrastructure? They are marked as 'new variable' in the comment column of the table.

*Table 3: List of variables included in our 12 models*

| *Dwelling / property characteristics* | **Source** | **Comment** |
|---|---|---|
| Area 'type' household is located in (i.e., rural, vs large regional centre, vs low-density area in metropolitan city) | Meshblock ID from HES combined with external data | See section on 'work location data' for more information. |

| *Occupant characteristics* | **Source** | **Comment** |
|---|---|---|
| No. of people in household | HES | |
| No. of working adults in household | HES | |
| No. of dependent children in household | HES | |
| Age of primary income earner in household | Derived from HES data | New variable |
| Log of after-tax household income | Derived from HES data | |
| Log of expenditure on non-capital items | Derived from HES data | New variable. Included as a proxy for permanent income. |

| *Other* | **Source** | **Comment** |
|---|---|---|
| HES survey year | HES | New variable. While all data has been inflation adjusted, the survey year remains important given steadily rising incomes. |

The variables that we assessed, but ultimately chose not to include in any of our models are listed in the notes. Table 6 in the notes section lists the variables we excluded to avoid confounding biases. In turn, *Table 7* lists the variables that we concluded do not add any additional explanatory power.

# Results

The income elasticities we derived for each of these twelve models are shown in Table 4 (for total expenditure), and Table 5**Error! Reference source not found.** on the following page[7] (for variable expenditure). It should be noted that we have transformed all financial variables (including both types of income and expenditure on infrastructure) into logs before including them. Income elasticities for models of this nature can be interpreted as meaning that 'a 1% increase in income is estimated to result in an *x*% increase in expenditure on the good or service in question', where x is the estimated coefficient.

As can be seen from the tables, our income elasticity estimates for current after-tax income range from 0.074 to 0.278 for total expenditure on infrastructure, and from 0.057 to 0.360 for variable expenditure. In turn our elasticity estimates for permanent income range from 0.373 to 0.41 for total expenditure and 0.551 to 0.582 for variable expenditure.

*Table 4: Estimated income elasticities for total infrastructure spending*

| | Dependent variable | | | | | |
|---|---|---|---|---|---|---|
| | (Log) All expenditure | (Log) All expenditure | (Log) All expenditure | (Log) All expenditure | (Log) All expenditure | (Log) All expenditure |
| Explanatory Variables | (Model 1) | (Model 2) | (Model 3) | (Model 4) | (Model 5) | (Model 6) |
| (Log) Income after-tax | 0.278*** | 0.074*** | | 0.278*** | 0.074*** | |
| | (0.006) | (0.007) | | (0.006) | (0.007) | |
| (Log) Expenditure on non-capital items | | 0.373*** | 0.408*** | | 0.373*** | 0.410*** |
| | | (0.008) | (0.007) | | (0.008) | (0.007) |
| **Model fit and description** | | | | | | |
| $R^2$ | 0.4331 | 0.5480 | 0.5439 | 0.4320 | 0.5477 | 0.5435 |
| Includes control variables for household location type? | Yes | Yes | Yes | No | No | No |

Notes:

- The number of asterisks following each estimated coefficient shows the level of significance of each coefficient (in effect, how confident we can be that the true coefficient is different from zero). Three asterisks denote a 99% probability that the true coefficient is different to zero. In turn, two asterisks denote a 95% probability, and one denotes a 90% probability.
- The figures in brackets show the standard error of each coefficient.

[7] The full regression analysis results for the twelve models are provided in the notes.

*Table 5: Estimated income elasticities for variable infrastructure spending*

| | Dependent variable | | | | | |
|---|---|---|---|---|---|---|
| | (Log) Variable expenditure | (Log) Variable expenditure | (Log) Variable expenditure | (Log) Variable expenditure | (Log) Variable expenditure | (Log) Variable expenditure |
| **Variable** | (Model 7) | (Model 8) | (Model 9) | (Model 10) | (Model 11) | (Model 12) |
| (Log) Income after-tax | 0.356*** | 0.057*** | | 0.360*** | 0.058*** | |
| | (0.012) | (0.013) | | (0.012) | (0.013) | |
| (Log) Expenditure on non-capital items | | 0.551*** | 0.579*** | | 0.554*** | 0.582*** |
| | | (0.012) | (0.01) | | (0.012) | (0.010) |
| **Model fit and description** | | | | | | |
| $R^2$ | 0.3105 | 0.4150 | 0.4139 | 0.3072 | 0.4128 | 0.4118 |
| Includes control variables for household location type? | Yes | Yes | Yes | No | No | No |

## Discussion

### Interpreting and using income elasticities

As noted, these elasticities measure how sensitive infrastructure spending is to income. An elasticity of 1 indicates that a 1% increase in income is associated with a 1% increase in infrastructure spending. Smaller elasticities suggest that infrastructure spending increases more slowly than income. An elasticity of 0.2, for example, indicates that a 1% increase in income is associated with only a 0.2% increase in spending.

Income elasticities can help us understand current and future demand for infrastructure. If infrastructure spending and use is highly responsive to incomes, we might expect per capita infrastructure use to be higher in higher income parts of the country. This might affect the number of different types of infrastructure needed in different places. Similarly, if infrastructure spending and use tends to rise as incomes increase, then future increases in per-capita incomes would be expected to result in increased demand for infrastructure even if the population is not growing.

We would emphasise that the elasticities reported in this research note are *averages* for the overall New Zealand population. We would expect income elasticities to vary between different sub-populations. Retired people for example may have more ability to shift their travel to avoid peak periods. If so, their income elasticity for transport spending would be lower than that of the population as a whole. The dataset we have used for this analysis does not include a sufficient number of households to allow us to estimate such sub-population specific elasticities. However, future users of these results should remember that different parts of the New Zealand population are likely to have different responses to those implied by the average elasticities reported above.

## Comparison to the Commission's earlier preliminary estimates

These elasticity estimates are slightly higher than those reported in our earlier report 'How much do we pay for infrastructure?' That earlier report concluded that on average, a 1% increase in household income is associated with a 0.24% increase in overall infrastructure spending. That model did not include a proxy for permanent income, so is best compared to the results of models 1 and 4 from this analysis, which both returned an elasticity of 0.278.

We consider the updated estimates reported in this research note to be superior to those earlier results as we have been able to include a greater range of appropriate controls in our models. These updated results should therefore be used instead of the earlier ones.

## Comparison to international estimates

Despite being slightly higher than our earlier, preliminary estimates, our updated income elasticity estimates remain at the lower end of those reported internationally. A literature review undertaken by the UK National Infrastructure Commission in 2018 found that income elasticity estimates mostly range between 0.2 and 0.9, depending upon the type of infrastructure service being considered, and the location. The two key exceptions to that overall finding are for land transport, where some estimates were slightly above 1, and air travel where their reported estimates ranged from 1.3 to 2.4.

As noted in our earlier 'How much do we pay for infrastructure?' report, the fact that our estimates are at the lower end of similar international estimates likely reflects our use of cross-sectional household data. Meta-analyses of income elasticities of demand for infrastructure services consistently find that cross-sectional methods, which compare different households or individuals at a point in time, result in lower estimates than methods that analyse how household or individual demand changes over time.

## Total expenditure vs variable expenditure

Our income elasticity estimates relating to total expenditure are mostly lower than those relating to variable expenditure. For example, the estimated elasticity for current income in model 1 is 0.278, whereas the equivalent estimate in model 7 (which has variable expenditure as the dependent variable but includes the same explanatory variables) is 0.356. Similar differences can be seen between the elasticity estimates for permanent income relating to total and variable infrastructure expenditure (such as between models 2 and 8).

These higher income elasticity estimates for the models explaining variable expenditure are consistent with our a priori expectations. Variable expenditure is, by definition, spending that varies in relation to the quantity of use of a good or service. We already know that higher-income households spend more on infrastructure in absolute terms than lower-income households, so would expect the portion of infrastructure spending that varies in relation to use to be more sensitive to changes in income than the portion that is unrelated to levels of use.

## Permanent and current income

As noted, our estimates of the elasticity for current income drop significantly if a proxy for permanent income is also included in the model. However, both permanent and current income retain significant explanatory power in all of our models, regardless of whether they are included on their own or together. This subsection discusses current and permanent income in more detail and suggests how these results should be interpreted.

To start, we would note that the overall level of 'fit' of our models is noticeably higher when permanent income is included as an explanatory variable. Model 1 for example, which does not include a proxy for permanent income, has an R-squared of 0.433. That compares to an R-squared of 0.548 for model 2, which includes both current and permanent income as explanatory variables. In lay terms that means that model 1 was able to explain only 43% of the observable variation in total household infrastructure spending, whereas model 2 explains 55%. Similar differences in the R-squared figure can be seen when comparing models 4, 5 and 6, models 7, 8 and 9, and models 10, 11 and 12.

In contrast, there are only relatively minor differences in overall fit between the models that include both current and permanent income as explanatory variables, and those that only include permanent income. For example, the overall fit of model 2 (which includes both current and permanent income) is 55%. That is only slightly higher than the fit of model 3, at 54%, which only includes permanent income. Similarly small differences in the coefficient for permanent income can be seen when comparing models 5 and 6, models 8 and 9, and models 11 and 12.

Turning to our estimated elasticities, and referring again to the tables on pages 18 and 19, the estimated elasticity for current income drops significantly if permanent income is also included in a model. For example, the coefficient for current income in model 1, which excludes permanent income, is 0.278. That drops to 0.074 in model 2, which includes both current and permanent income. This effect is much smaller for the coefficients relating to permanent income, however. The coefficient for permanent income in model 3, for example, is 0.409. That drops to 0.373 when current income is included in the model as well.

This result, that the estimated elasticity for current income drops significantly if permanent income is also included in a model, is even more pronounced in the models explaining *variable* infrastructure expenditure. In model 7 for example, which excludes permanent income, our estimated elasticity for current income is 0.356. That drops to 0.057 in model 8, which includes both current and permanent income.

We draw several conclusions from these results. First, and perhaps most importantly, permanent income looks to be a more important determinant of infrastructure spending than current income. Users of this analysis should draw on the elasticity estimates from our models that include permanent income where possible. However, current after-tax income still provides some additional explanatory power in addition to that provided by permanent income, so the models which include both types of income would appear to be the most useful.

These conclusions highlight the importance of future users of these elasticity estimates being clear in advance what ‘type’ of income elasticity is most appropriate for their intended use. The two key decisions will be whether to use a variable or total expenditure elasticity, and whether to use a current income or permanent income elasticity. Both decisions will have a meaningful impact on the size of elasticity, and therefore the results of any future analysis.

## Household location types

Our income elasticity estimates vary only slightly between the models that include household location types and those that exclude them. To recap, these household location types were included in the hoped that they would pick up differences in commuting patterns. Referring again to the tables on pages 18 and 19, our income elasticity estimates relating to total infrastructure expenditure do not vary at all (to the level of 3 decimal places) between the models that include household location types and those that don’t. In turn the elasticity estimates relating to variable infrastructure expenditure differ only slightly between the models that include household location types and those that don’t. Our elasticity estimate for current income, for example, varies only slightly between models 7 and 10 (where the elasticity estimates are 0.356 and 0.360, respectively).

This leads us to conclude that these household location type variables are too 'coarse' a measure to capture meaningful differences in commuting distances, and modes of travel. We therefore recommend that the results of the models that include them (models 1-3 and 7-9) not be used.

## Using these results to help predict New Zealand's long-term infrastructure needs

Finally, we would note that these results could be used to forecast levels of future demand for infrastructure services.

Changes in New Zealand's overall population and increases in its per capita incomes are likely to be the two biggest drivers of changes in the demand for infrastructure services in the future. Combining the results from this research note with existing long-term economic and population growth projections prepared by Stats NZ and Treasury would allow more robust long-term forecasts of the future demand for infrastructure services to be developed.

In doing so, however, it will be important to distinguish between increased demand for the same quality of service, and demand for higher quality services. Rising incomes will likely increase the demand for services of the existing level of quality *and* the demand for higher quality services. In contrast, population growth will most likely only increase the demand for the same quality of service.

# Technical notes – List of excluded variables

The tables in these notes list the variables which have been excluded from our models. Table 6 lists those variables removed due to concerns around confounding biases. *Table 7* then lists those variables that were found to have little or no explanatory power when combined in models with other more important variables.

*Table 6: Variables excluded due to concerns around confounding biases*

| **Variable** |
|---|
| ***Dwelling characteristics*** |
| Number of bedrooms |
| Household type (dwelling code) |
| Expenditure on rates |
| X & Y location coordinates of dwelling |
| ***Occupant characteristics*** |
| Household composition (whether one or more earners in household, and whether one or more dependent children) |
| Renting? (Y/N) |
| Own a car? (Y/N) |
| Reported difficulty in paying bills |
| Number of adults in household not working |
| ***Other socioeconomic variables*** |
| Deprivation decile of meshblock household lives in |
| Degree of overcrowding of dwelling |

*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 (meshblock) |
| Location of property (SA1 and SA2) |
| TLA property is located in |
| Regional council property is located in |
| SA2 population density |
| Dwelling code[8] |
| ***Occupant characteristics*** |
| Number of families in dwelling |
| Highest qualification of primary earner |
| Expenditure on rent |
| Household income class (breakdown of income quintile 1) |
| $ amount of outstanding mortgage |
| Total expenditure on housing related expenses |
| Household tenure code[9] |
| Employment status of adults in household |
| DEP-17 material wellbeing score of family |

[8] Denotes whether the dwelling is private, or public (such as a boarding house or motel).
[9] Denotes whether the dwelling is fully or partly owned by the individual, a family trust, or some other entity.

# Technical notes – Full regression results

In these notes we present our full regression results. The first two rows of the following tables show our estimated expenditure elasticities with respect to current income and permanent income respectively. This is done by taking the natural logarithms of household expenditure on infrastructure, household income after tax (current income) and household expenditure on non-capital items (a proxy for permanent income).

Expenditure elasticities can be interpreted as 'a 1% increase in income is estimated to result in an *x*% increase in expenditure. The expenditure elasticities are calculated using ordinary least squares (OLS) regression.

The elasticity regressions can be defined by the following equation,

$$\begin{aligned} &log(total\ expenditure\ on\ infrastructure\ ) \\ &\qquad = \alpha + \log(after\ tax\ income)'\beta_1 + \log(\text{expenditure on non} - \text{capital items})\beta_2 + Z'\delta \end{aligned}$$

Where total expenditure on infrastructure is the dependent variable, $\alpha$ is the estimated constant term, $\beta_1$ is the estimated effect of after-tax income on the dependent variable, $\beta_2$ is estimated effect of permanent income on the dependent variable, $Z$ is a vector of controls and $\delta$ is the vector of coefficients on the controls.

Only the financial variables included in our models were transformed into logs. For all other variables the interpretation of the coefficients is somewhat different. The $\delta$ values are referred to as semi-elasticities. These can be interpreted as 'a one unit increase in the relevant $Z$ control is estimated to increase expenditure on infrastructure by 100 × α percent'.

*Table 8: OLS, expenditure elasticities with respect to income for total expenditure on infrastructure services. Models 1-3*

| Dependent variable | Log (all expenditure) | Log (all expenditure) | Log (all expenditure) |
|---|---|---|---|
| | (Model 1) | (Model 2) | (Model 3) |
| Log (income after-tax) | 0.278*** | 0.074*** | |
| | (0.006) | (0.007) | |
| Log (expenditure non-capital after-tax) | | 0.373*** | 0.408*** |
| | | (0.008) | (0.007) |
| Number of people in household | 0.182*** | 0.153*** | 0.156*** |
| | (0.005) | (0.005) | (0.005) |

Note: Standard errors in parentheses, *** = p<0.01, ** = p<0.05, * = p<0.1.

*Table 8 continued*

| | | | |
|---|---|---|---|
| Number of dependent children in household | -0.133*** | -0.123*** | -0.126*** |
| | (0.006) | (0.006) | (0.006) |
| Number of working adults in household | 0.042*** | 0.022*** | 0.042*** |
| | (0.005) | (0.005) | (0.004) |
| Age of household primary income earner | 0.001*** | 0.002*** | 0.002*** |
| | (0.0002) | (0.0002) | (0.0002) |
| 2006/07 HES Survey | 0.016* | | |
| | (0.009) | | |
| 2009/10 HES Survey | 0.032*** | 0.02*** | |
| | (0.008) | (0.007) | |
| 2012/13 HES Survey | 0.087*** | 0.088*** | 0.083*** |
| | (0.008) | (0.007) | (0.007) |
| Large regional centre | -0.03*** | -0.022*** | -0.022*** |
| | (0.009) | (0.007) | (0.007) |
| Medium regional centre | -0.032*** | | |
| | (0.012) | | |
| Small regional centre | | | |
| | | | |
| Rural area | 0.029** | 0.061*** | 0.06*** |
| | (0.012) | (0.01) | (0.01) |
| Metro and PDQ 1 | 0.073*** | 0.041*** | 0.046*** |
| | (0.015) | (0.013) | (0.013) |
| Metro and PDQ 2 | 0.023** | | 0.02** |
| | (0.011) | | (0.01) |
| Metro and PDQ 3 | 0.018* | | |
| | (0.01) | | |
| Metro and PDQ 4 | | | |
| | | | |
| Metro and PDQ 5 | | | |
| | | | |
| Constant | 5.771*** | 4.048*** | 4.431*** |
| | (0.069) | (0.073) | (0.07) |
| $R^2$ | 0.4331 | 0.5480 | 0.5439 |
| Observations | 15,726 | 15,708 | 15,708 |

Note: Standard errors in parentheses, *** = p<0.01, ** = p<0.05, * = p<0.1.

*Table 9: OLS, expenditure elasticities with respect to income for total expenditure on infrastructure services. Models 4-7*

| Dependent variable | Log (all expenditure) | Log (all expenditure) | Log (all expenditure) |
|---|---|---|---|
| | (Model 4) | (Model 5) | (Model 6) |
| Log (income after-tax) | 0.278*** | 0.074*** | |
| | (0.006) | (0.007) | |
| Log (expenditure non-capital after-tax) | | 0.373*** | 0.41*** |
| | | (0.008) | (0.007) |
| Number of people in household | 0.181*** | 0.152*** | 0.156*** |
| | (0.005) | (0.005) | (0.005) |
| Number of dependent children in household | -0.132*** | -0.123*** | -0.126*** |
| | (0.006) | (0.006) | (0.006) |
| Number of working adults in household | 0.043*** | 0.023*** | 0.043*** |
| | (0.005) | (0.005) | (0.004) |
| Age of household primary income earner | 0.001*** | 0.002*** | 0.002*** |
| | (0.0002) | (0.0002) | (0.0002) |
| 2006/07 HES Survey | 0.016* | | |
| | (0.009) | | |
| 2009/10 HES Survey | 0.032*** | 0.02*** | |
| | (0.009) | (0.007) | |
| 2012/13 HES Survey | 0.087*** | 0.088*** | 0.083*** |
| | (0.008) | (0.007) | (0.007) |
| Constant | 5.77*** | 4.04*** | 4.423*** |
| | (0.069) | (0.073) | (0.07) |
| $R^2$ | 0.4320 | 0.5477 | 0.5435 |
| Observations | 15,726 | 15,708 | 15,708 |

Note: Standard errors in parentheses, *** = p<0.01, ** = p<0.05, * = p<0.1.

*Table 10: OLS, expenditure elasticities with respect to income for variable expenditure on infrastructure services. Models 7-9*

| Dependent variable | Log (variable expenditure) | Log (variable expenditure) | Log (variable expenditure) |
|---|---|---|---|
| | (Model 7) | (Model 8) | (Model 9) |
| Log (income after-tax) | 0.356*** | 0.057*** | |
| | (0.012) | (0.013) | |
| Log (expenditure non-capital after-tax) | | 0.551*** | 0.579*** |
| | | (0.012) | (0.01) |
| Number of people in household | 0.168*** | 0.124*** | 0.125*** |
| | (0.009) | (0.008) | (0.008) |
| Number of dependent children in household | -0.093*** | -0.077*** | -0.075*** |
| | (0.011) | (0.01) | (0.009) |
| Number of working adults in household | 0.078*** | 0.05*** | 0.07*** |
| | (0.009) | (0.009) | (0.007) |
| Age of household primary income earner | -0.002*** | -0.001* | |
| | (0.0004) | (0.0004) | |
| 2006/07 HES Survey | 0.126*** | 0.105*** | 0.096*** |
| | (0.015) | (0.014) | (0.014) |
| 2009/10 HES Survey | 0.126*** | 0.111*** | 0.105*** |
| | (0.014) | (0.013) | (0.013) |
| 2012/13 HES Survey | 0.138*** | 0.141*** | 0.137*** |
| | (0.014) | (0.013) | (0.013) |
| Large regional centre | 0.075*** | 0.086*** | 0.083*** |
| | (0.017) | (0.016) | (0.016) |
| Medium regional centre | 0.058*** | 0.103*** | 0.098*** |
| | (0.022) | (0.02) | (0.02) |
| Small regional centre | 0.109*** | 0.126*** | 0.12*** |
| | (0.029) | (0.026) | (0.026) |

*Table 10 continued*

| | | | |
|---|---|---|---|
| Rural area | 0.184*** | 0.232*** | 0.227*** |
| | (0.022) | (0.02) | (0.02) |
| Metro and PDQ 1 | 0.204*** | 0.159*** | 0.158*** |
| | (0.026) | (0.024) | (0.024) |
| Metro and PDQ 2 | 0.103*** | 0.093*** | 0.093*** |
| | (0.021) | (0.019) | (0.019) |
| Metro and PDQ 3 | 0.091*** | 0.073*** | 0.073*** |
| | (0.019) | (0.018) | (0.017) |
| Metro and PDQ 4 | 0.057*** | 0.036** | 0.036** |
| | (0.018) | (0.016) | (0.016) |
| Metro and PDQ 5 | | | |
| Constant | 4.06*** | 1.489*** | 1.744*** |
| | (0.124) | (0.129) | (0.106) |
| $R^2$ | 0.3105 | 0.4150 | 0.4139 |
| Observations | 15,639 | 15,618 | 15,618 |

Note: Standard errors in parentheses, *** = p<0.01, ** = p<0.05, * = p<0.1.

*Table 11: OLS, expenditure elasticities with respect to income for variable expenditure on infrastructure services. Models 10-12*

| Dependent variable | Log (variable expenditure) | Log (variable expenditure) | Log (variable expenditure) |
|---|---|---|---|
| | (Model 10) | (Model 11) | (Model 12) |
| Log (income after-tax) | 0.36*** | 0.058*** | |
| | (0.012) | (0.013) | |
| Log (expenditure non-capital after-tax) | | 0.554*** | 0.583*** |
| | | (0.012) | (0.01) |
| Number of people in household | 0.164*** | 0.12*** | 0.123*** |
| | (0.009) | (0.008) | (0.008) |
| Number of dependent children in household | -0.088*** | -0.071*** | -0.074*** |
| | (0.011) | (0.009) | (0.009) |
| Number of working adults in household | 0.08*** | 0.056*** | 0.07*** |
| | (0.009) | (0.008) | (0.007) |
| Age of household primary income earner | -0.002*** | | |
| | (0.0004) | | |
| 2006/07 HES Survey | 0.127*** | 0.106*** | 0.096*** |
| | (0.015) | (0.014) | (0.014) |
| 2009/10 HES Survey | 0.127*** | 0.112*** | 0.105*** |
| | (0.014) | (0.013) | (0.013) |
| 2012/13 HES Survey | 0.137*** | 0.14*** | 0.136*** |
| | (0.014) | (0.013) | (0.013) |
| Constant | 4.078*** | 1.455*** | 1.764*** |
| | (0.123) | (0.128) | (0.107) |
| $R^2$ | 0.3072 | 0.4128 | 0.4118 |
| Observations | 15,639 | 15,618 | 15,618 |

Note: Standard errors in parentheses, *** = p<0.01, ** = p<0.05, * = p<0.1.