Modelled regional income timeseries

Alan Bentley

December 2021

Contents

1 Introduction

1.1 Data needs

Regional income series – split by regional council and territorial authority – are needed to understand households' and individuals' economic resources. For example, we can use these series alongside rentals for housing and house prices to provide indicators of housing affordability.

Regional timeseries of household income are not readily available as official statistics. This document describes our modelling approach, which combines several data sources, to estimate regional household income.

1.2 Modelling approach

Our modelling strategy is to use timely, but less comprehensive, indicators, calibrated to reliable, but infrequent, benchmarks available with a notable lag. This is a three stage approach:

1. Create **income benchmarks**: The Household Economic Survey (HES) is used to estimate *calibration and coverage factors* to apply to Census income estimates
2. **Interpolate** *between* income benchmarks: Linked employer-employee (LEED) tax data is used to estimate regional income timeseries between benchmarks (created in step 1)
3. **Extrapolate** *beyond* income benchmarks: LEED and other tax data is to used to extend the regional income timeseries to the latest period

The approach is flexible so can be used to estimate several aggregate statistics:

- Gross (before tax) and disposable (after tax) income
- For various income sharing units: personal, household, equivalised household (accounts for different household composition)
- By *tenure* (renters and all households)
- For different *periodicities* (annual, quarterly)

1.3 Data used

Four income data sources are used: two surveys and two variants of administrative tax data (see Table 1). Individually, each source has critical weaknesses. Collectively, we can borrow strength from each source to model timely, reliable, estimates.

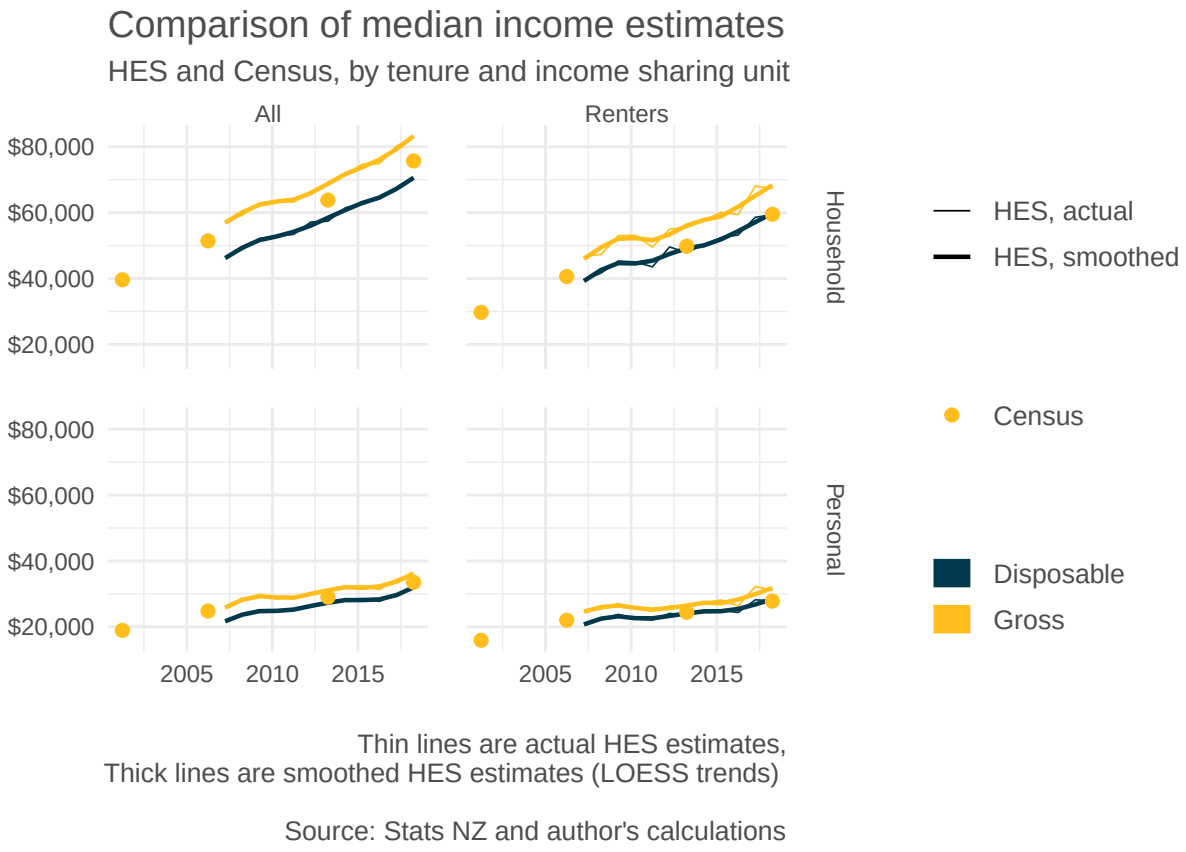
Table 1: Summary of data sources

Source	**Availability**	**Quality characteristics**
Household Economic Survey (**HES**)	Yearly, reporting lag of about 18 months	Comprehensive coverage of income sources (gross and disposable); limited regional coverage as sample survey
Census of population and dwellings	5-yearly, reporting lag of about 18 months	Comprehensive regional coverage; only gross (before tax) income - overall quality thought to be lower than HES
Tax data: Linked employer-employee data (**LEED**)	Quarterly, 4-5 quarter lag	Good regional coverage; *median gross earnings per job, based on employment location* (rather than home address)
Tax data: **Business employment data**	Quarterly, 2 month lag	Good regional coverage; *mean* gross earnings per job, based on *employer location* (rather than home address)

2 Income benchmarks

The **first stage** uses survey data; namely, the Census of population and dwellings, and Household Economic Survey (HES). The Census has comprehensive regional coverage (for example, series are available for territorial authorities and Auckland local boards), but only gross (before tax) income is reported. Census income is collected as discrete, banded, amounts, so is less precise than income reported as dollars. The quality of HES income series is considered superior as detailed collection, by income source (in dollars), aids coverage of income sources and allows for the calculation of disposable (after tax) income.

Figure 1



Shown in Figure 1, gross income is a higher estimate using the HES compared with Census. We smoothed the HES timeseries using a LOESS (locally estimated scatterplot smoothing) regression to reduce HES sampling errors; HES 2018/19 data was not used due to a discontinuity in the source of income data. The HES series overlap with the the 2013 and 2018 Censuses (the income reference period is assumed comparable: Censuses are conducted in March; HES is conducted continuously, reported for each year to June). Calculating the average ratio of HES to Census gross income estimates produces gross income *calibration factors* (see Table 2 and Figure 2). These have been calculated for household and personal income, by tenure type (renters and all households).

Replacing HES gross income with disposable income we can calculate analogous *calibration and coverage factors* for disposable income.

		Income type	
Income unit	Tenure	Disposable	Gross
Household	All	0.92	1.09
Household	Renters	1.00	1.14
Personal	All	0.95	1.08
Personal	Renters	1.00	1.11

Table 2: Census calibration and coverage factors

Figure 2

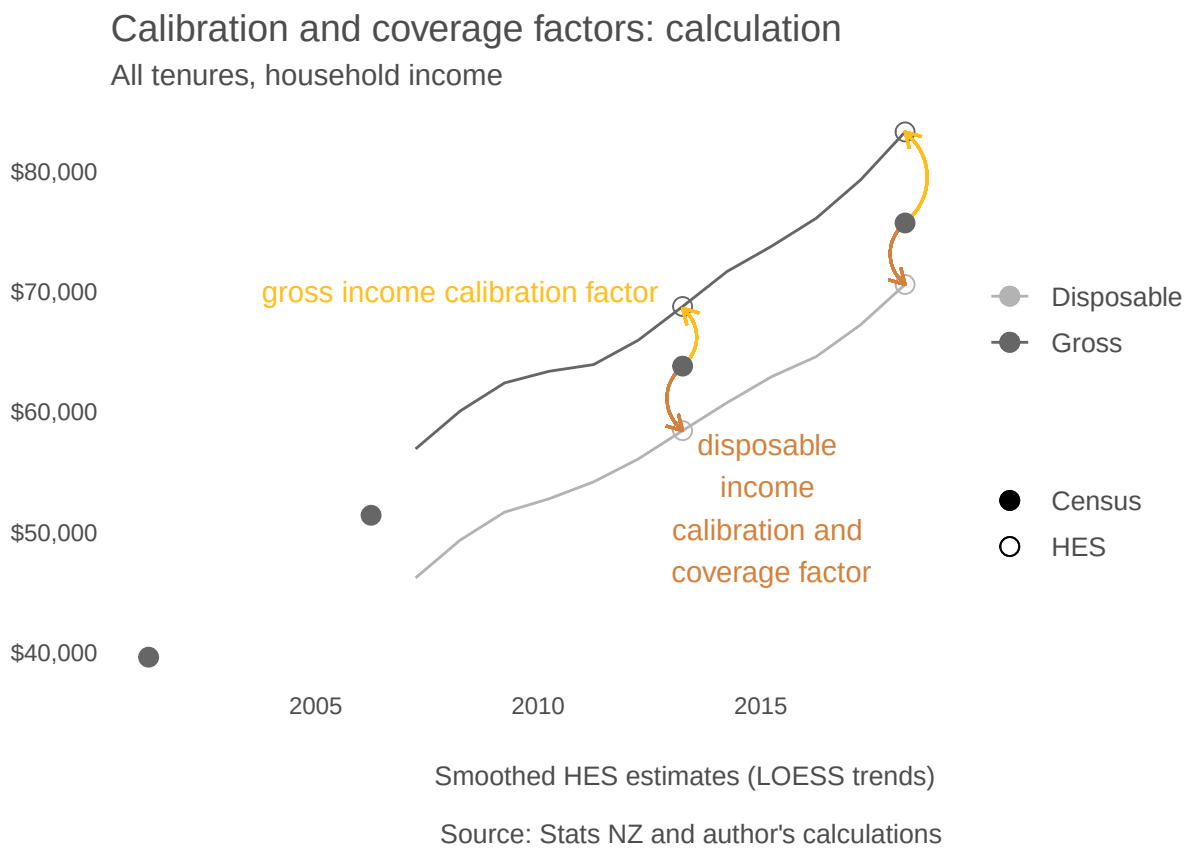
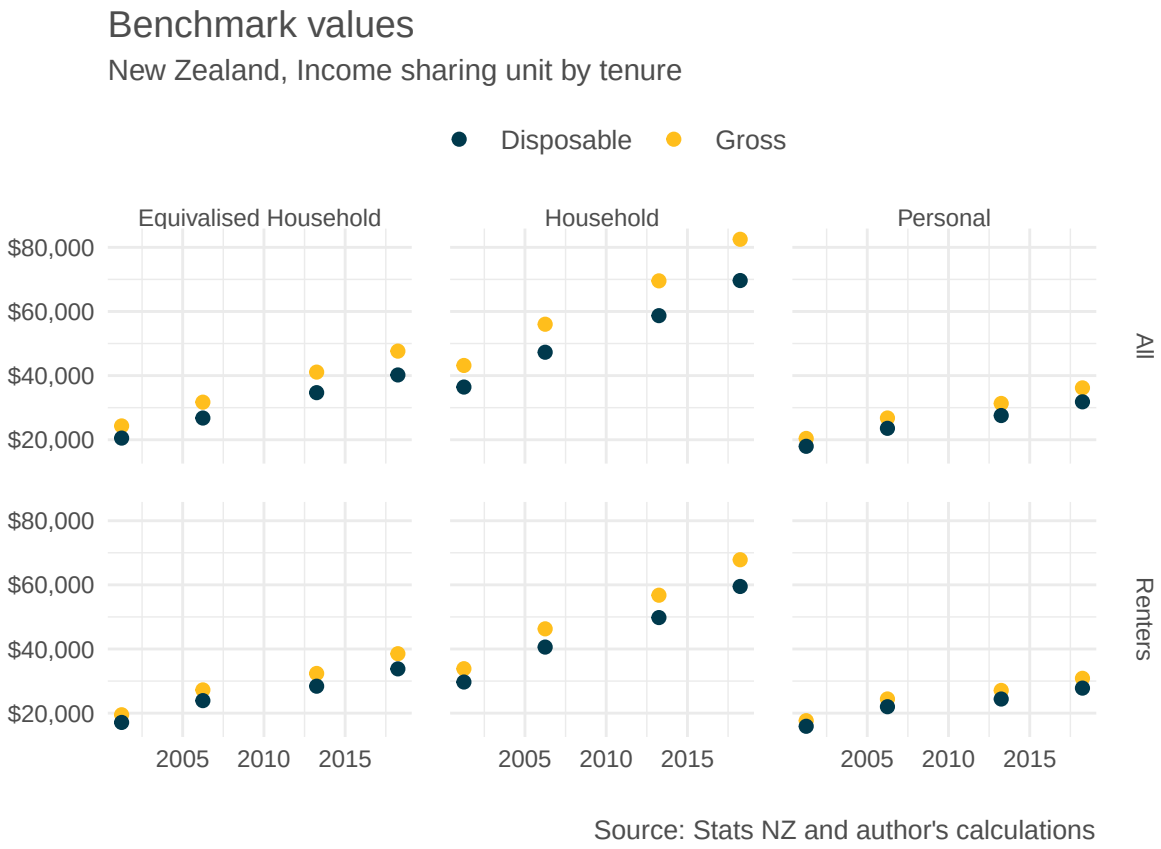


Figure 3

The resulting New Zealand level benchmark estimates, after applying the calibration and coverage factors, to Census data, is shown in Figure 3. The same factors were applied for each regional area of New Zealand.



3 Interpolation and extrapolation

3.1 Method explained

The **second stage** uses Linked employer-employee (LEED) tax data to **interpolate** *between* income benchmarks.

1. LEED median gross earnings per job are converted from *quarterly* to *annual* amounts, by aggregating over a rolling 4-quarter period.
2. *Benchmark factors* are calculated as the ratio of annualised LEED median earnings to income benchmarks (estimated from HES and Census in stage 1)
3. A timeseries of benchmark factors (between Census years) is created by linear interpolation
4. Applying these factors to the LEED earnings timeseries results in calibrated income timeseries

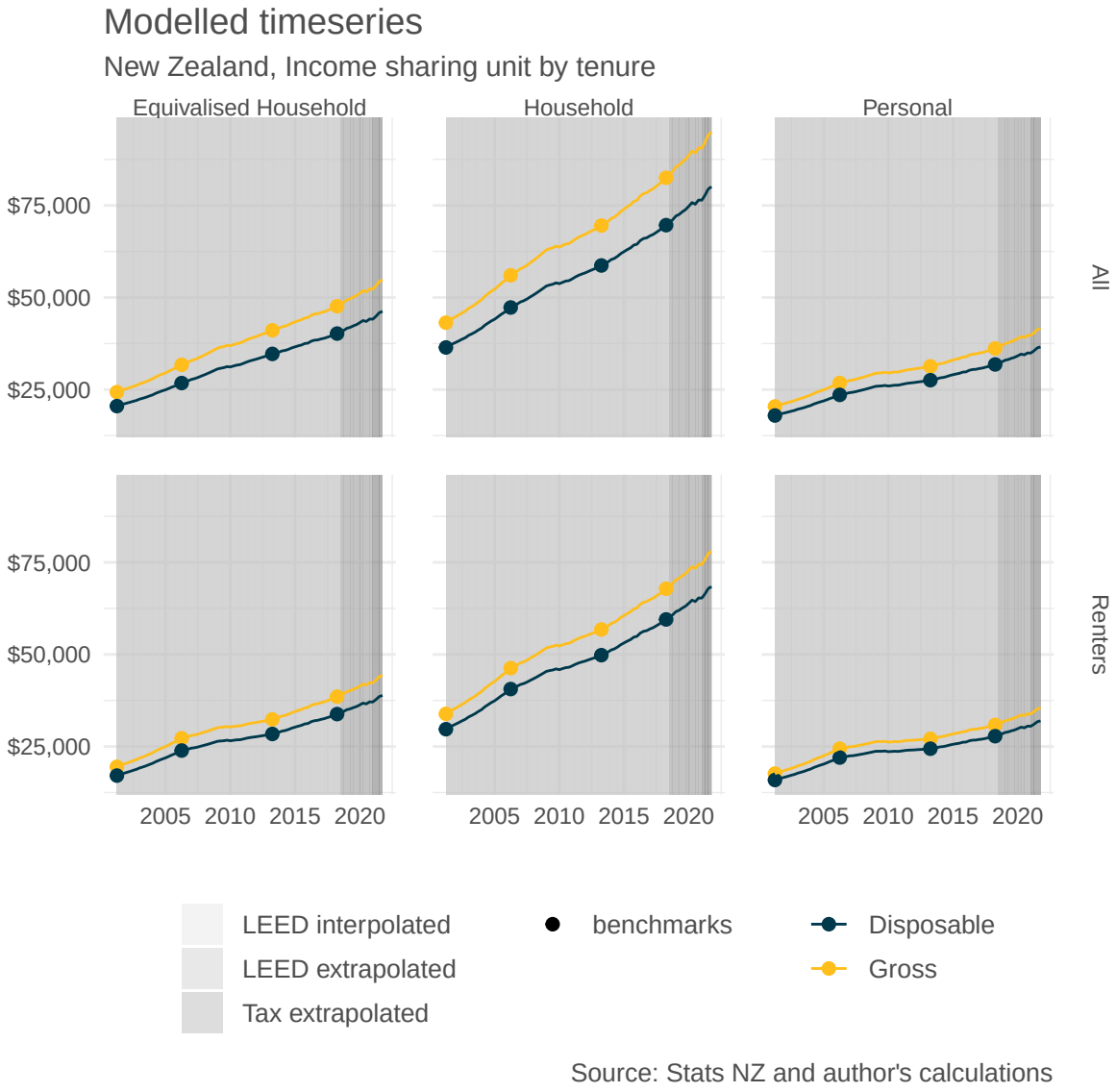
The **third stage** uses LEED and other tax data to extend the timeseries to the latest period.

1. LEED quarterly series are available with a 4-5 quarter lag. The latest available benchmark factor (from the time of the most recent Census) is applied to the LEED earnings series available beyond the benchmark.
2. This series is further extended to the latest period using annual changes (on a quarterly frequency) in another source of tax data, *Business employment data*, mean gross earnings are now used as median series are currently unavailable.

3.2 Final income timeseries

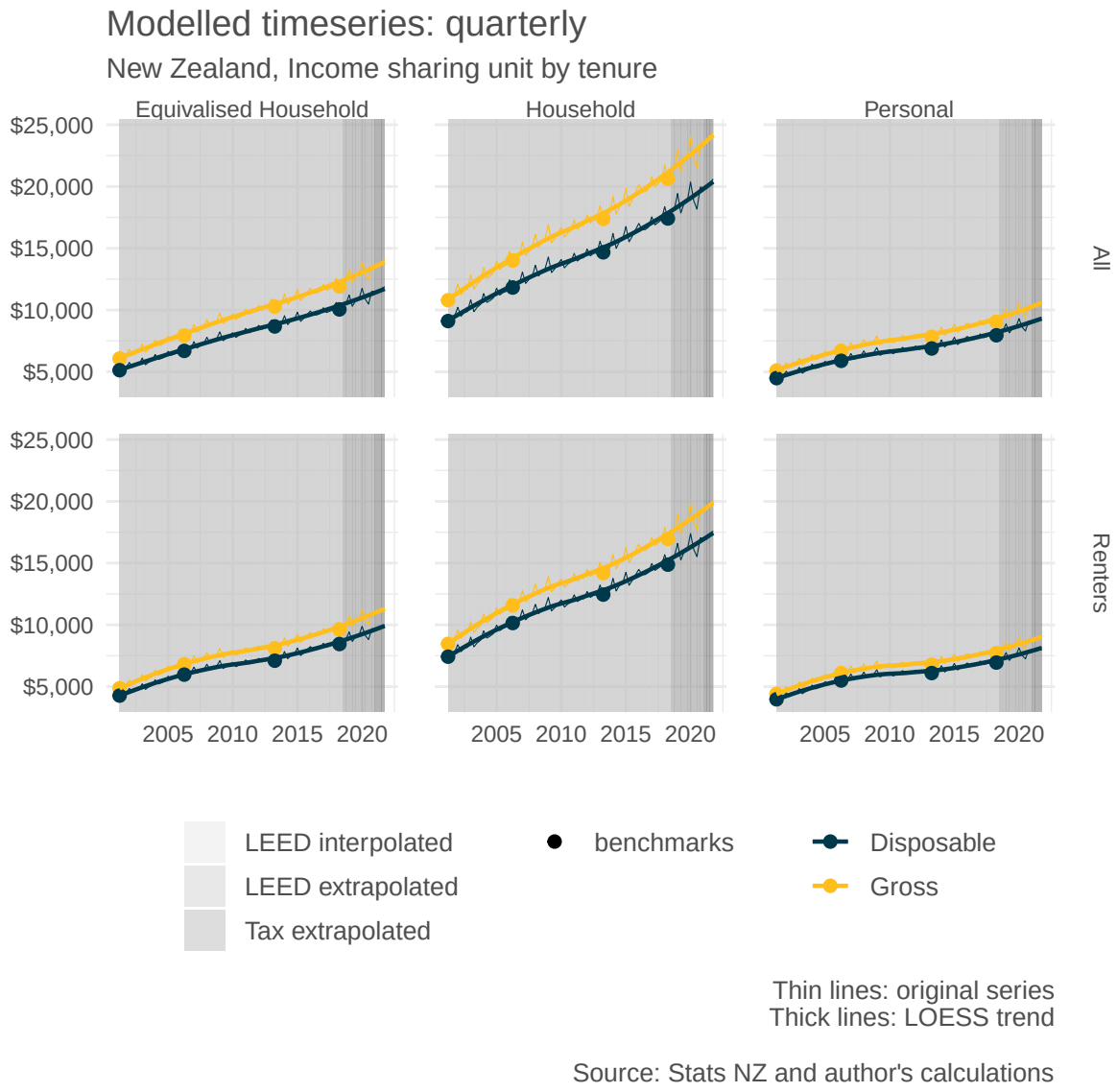
The resulting timeseries, for total New Zealand, are shown in Figure 4. The same methodology is applied for each regional area of New Zealand.

Figure 4



The method can also be used to create quarterly timeseries. LEED series now use the original quarterly values (not rolled up to annual amounts) and the Census-HES benchmarks are divided by four to estimate quarterly values. Shown in Figure 5, the series are more volatile, but changes in income growth may be apparent sooner using a quarterly income series.

Figure 5



Median annual household disposable income by regional council is shown in Figure 6, and by territorial authority in Figure 7, as an examples of regional timeseries.

Figure 6

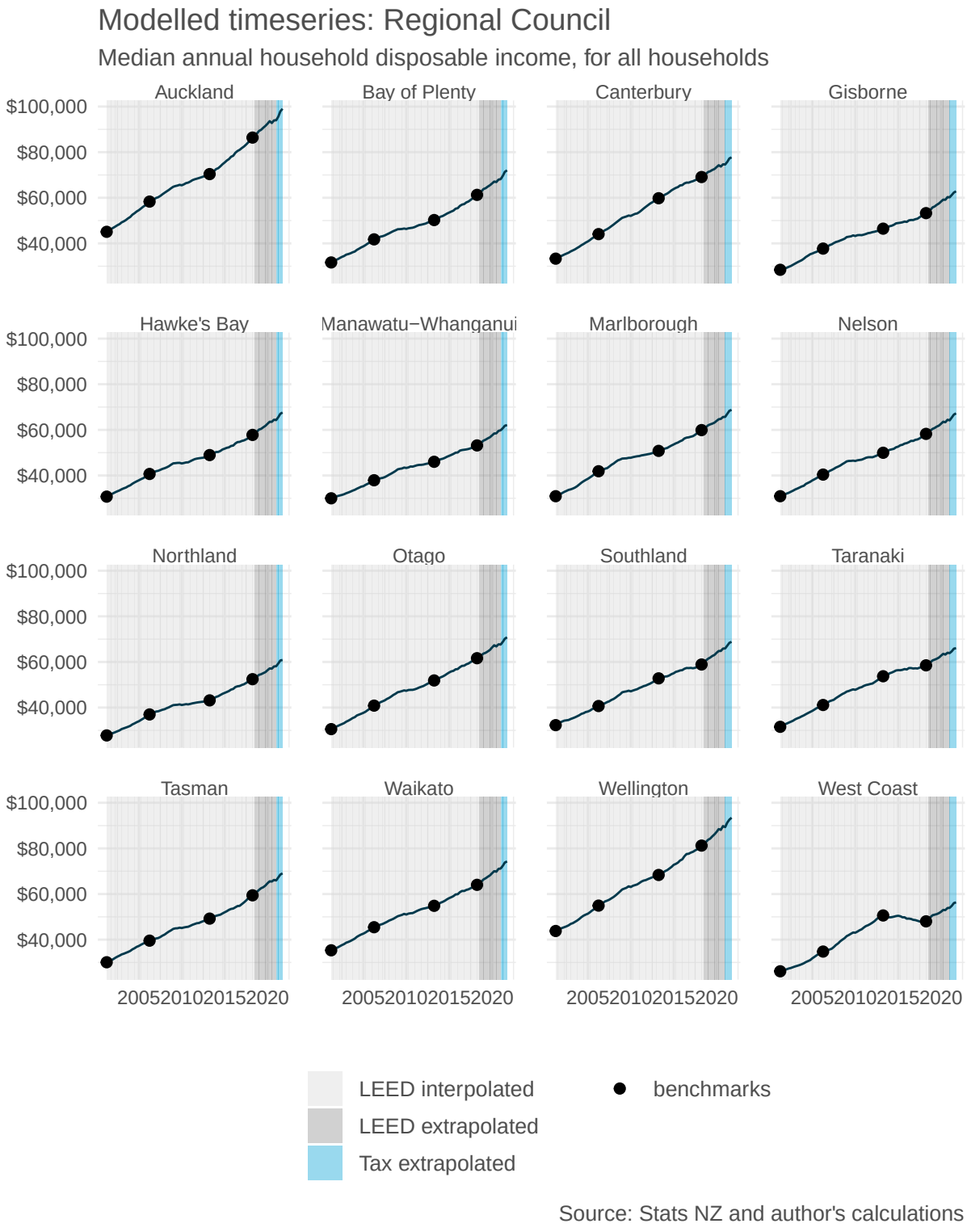
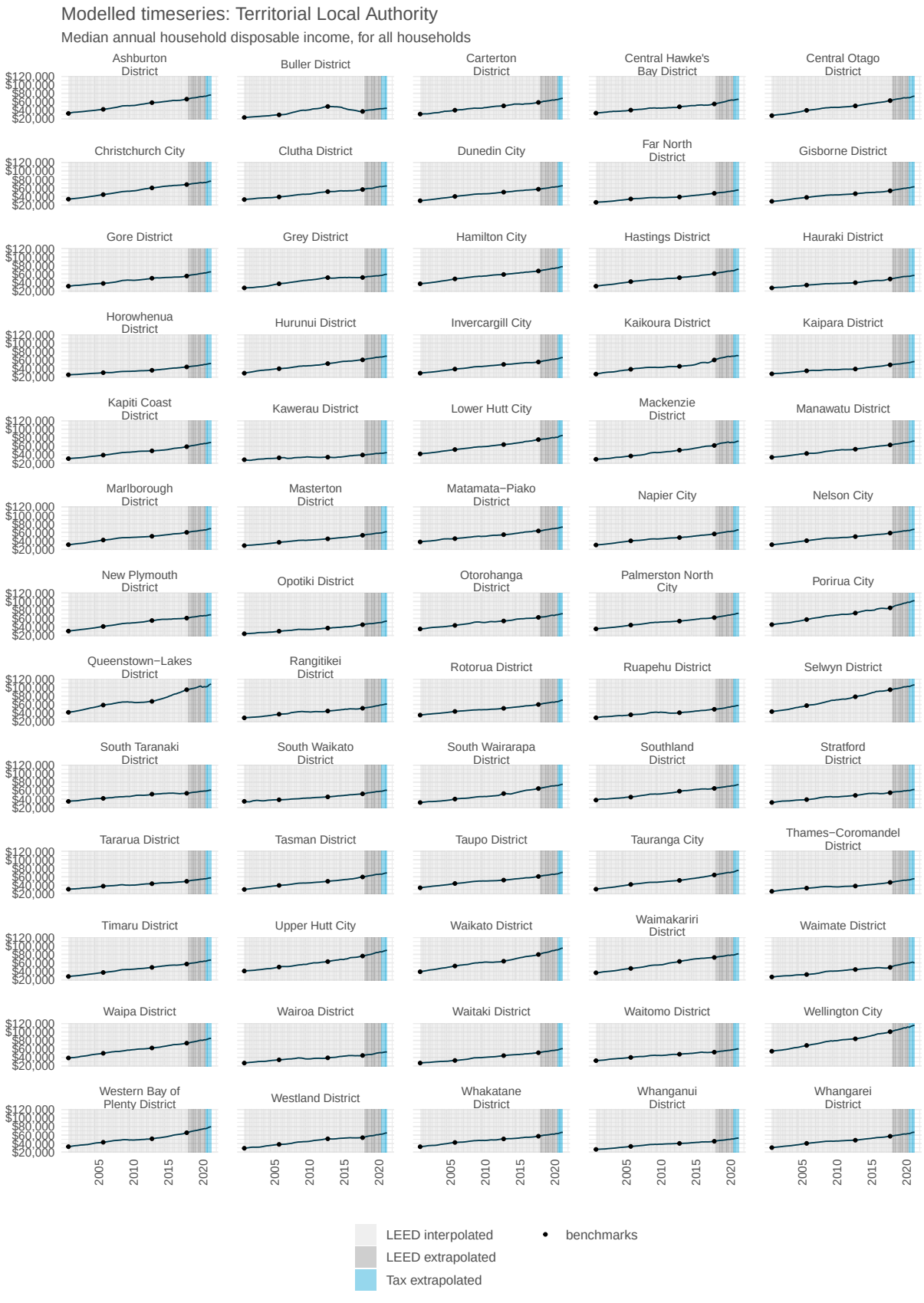

Modelled timeseries: Regional Council
Median annual household disposable income, for all households
Auckland
Bay of Plenty
Canterbury
Gisborne
Hawke's Bay
Manawatu−Whanganui
Marlborough
Nelson
Northland
Otago
Southland
Taranaki
Tasman
Waikato
Wellington
West Coast
$100,000
$80,000
$60,000
$40,000
2005201020152020
LEED interpolated
LEED extrapolated
Tax extrapolated
benchmarks
Source: Stats NZ and author's calculations

Figure 7



Source: Stats NZ and author's calculations

Appendix

Figure A1: Comparison of modelled timeseries with HES

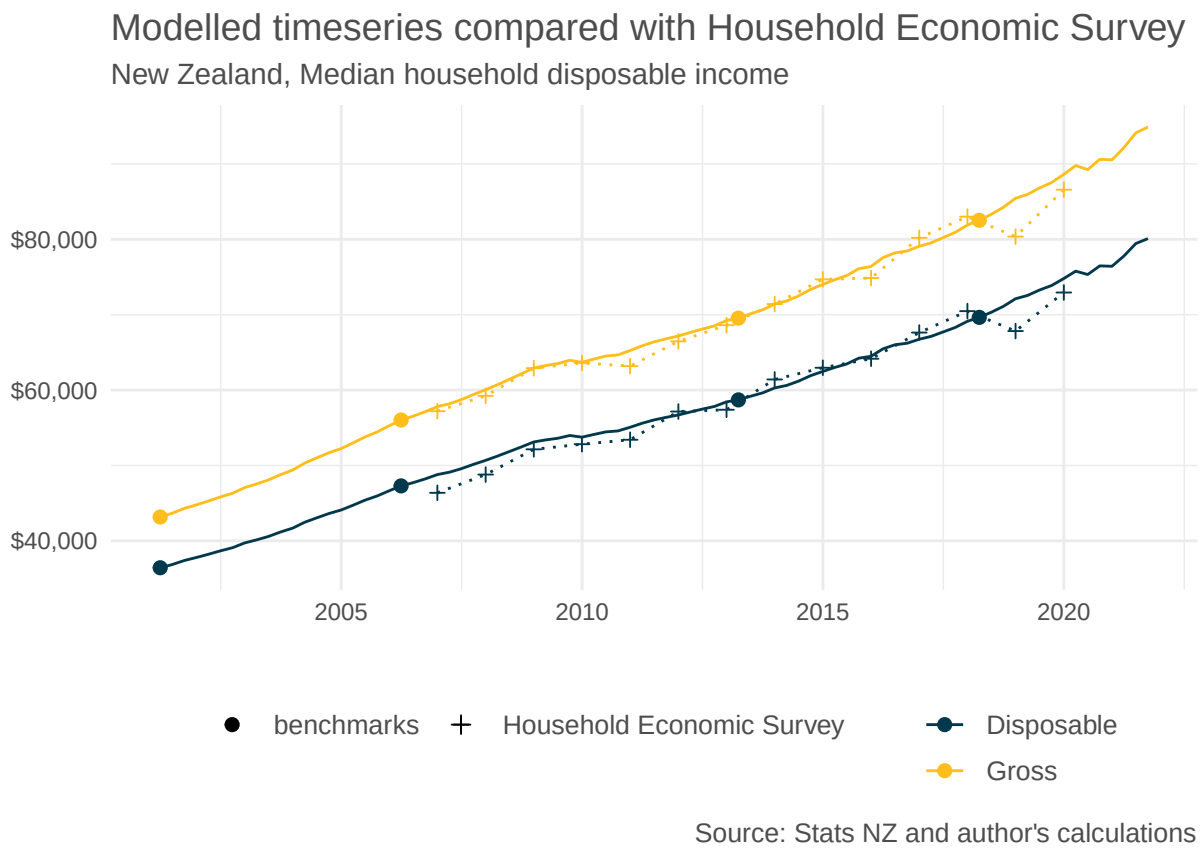
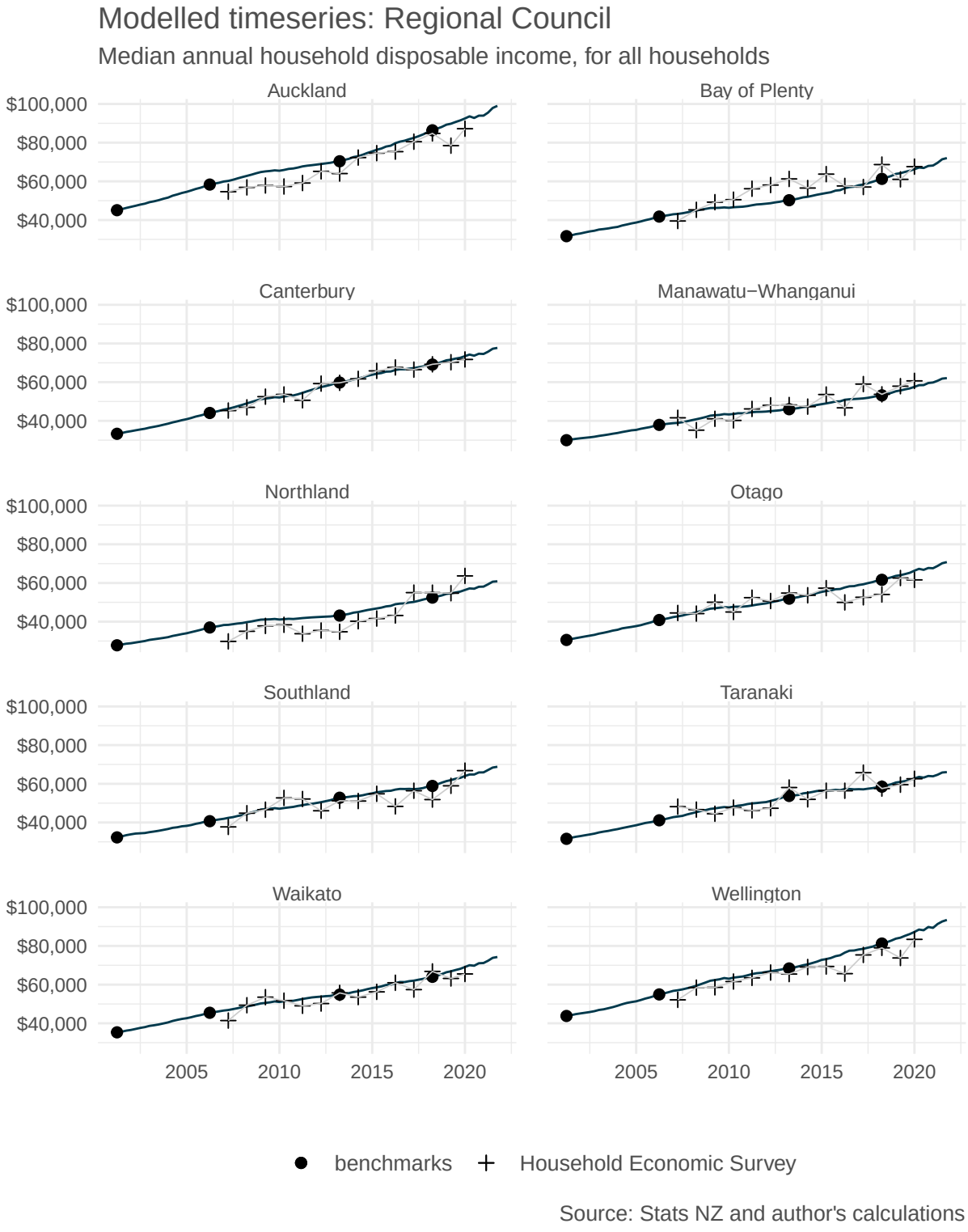


Figure A2: Regional comparison of modelled timeseries with HES



Links to input data sources

Household Economic Survey (HES)

Years ending June 2007–2019 https://catalogue.data.govt.nz/dataset/hes-gross-and-disposable-income-timeseries-2006-07-2018-19

Years ending June 2019–2020 https://www.stats.govt.nz/information-releases/household-income-and-housing-cost-statistics-year-ended-june-2020

Censuses of Population and Dwellings (Census)

https://catalogue.data.govt.nz/dataset/census-household-and-personal-income-series

Linked Employer-Employee Data (LEED)

https://www.stats.govt.nz/information-releases/linked-employer-employee-data-september-2020-quarter-nz-stat-tables

Business Employment Data

https://www.stats.govt.nz/information-releases/employment-indicators-september-2021