SCARLATTI
Evaluation | Analytics | Insights

ALTA

NEW ZEALAND INFRASTRUCTURE COMMISSION
Te Waihanga

Prepared for:
Te Waihanga (New Zealand Infrastructure Commission)

Infrastructure workforce capacity baselining study

Technical introduction to modelling methodology and calibration

September 2023

Scarlatti.co.nz

About this document

The following document has been prepared for Te Waihanga as a part of the final deliverable of the *Infrastructure workforce capacity: Baselining study* (Te Waihanga project #30032). This report is delivered as a technical summary of the methodologies used to estimate the supply of and demand for the infrastructure workforce – refer to accompanying databases.

This report is split into the following sections:

1. **Overview of the research programme.** An introduction to the context in which this project was established and an overview of the description of services and project deliverables
2. **The infrastructure workforce definition.** Presentation of the infrastructure workforce definition and what is considered in and out of consideration
3. **Methodology: Modelling the supply of the infrastructure workforce.** An overview of the methodologies adopted to model the infrastructure workforce's current capacity/supply
4. **Methodology: Modelling the demand for the infrastructure workforce.** An overview of the methodologies adopted to model the current and future demand for the infrastructure workforce
5. **Reconciliation of supply and demand models.** A comparison of supply and demand model outputs and an overall estimate of workforce size
6. **Recommendations for refining methodology and inputs.** A list of recommendations for improvements that could be made to subsequent iterations of the models, and how they might be expected to improve the analysis.
7. **Appendices.** Additional supporting information.

Disclaimer

The methodologies presented in this report have been used to establish a 'point-in-time' estimate for the supply of and demand for the infrastructure workforce. There remains some uncertainty on several of the inputs used in the model which would benefit from refinement to improve the analysis outputs over time. Recommendations for where refinement would be most valuable is included at the conclusion of this document before the appendices.

Contents

Overview of the research programme

Context

Rautaki Hanganga o Aotearoa: The New Zealand Infrastructure Strategy 2022 – 2050 recommends that Te Waihanga provide industry and government with a long-term view on the likely resources required to plan, deliver and maintain infrastructure assets in New Zealand.

The New Zealand infrastructure industry is experiencing historic workforce shortages that are impacting the industry's ability to provide and maintain the infrastructure assets New Zealand needs. An important foundation piece of research is to define and quantify the current capacity of the workforce (i.e. current workforce supply) as a baseline and to characterise what is to be expected from the workforce in the future (i.e. current and future workforce demand). If the industry understands the current situation and what the future needs are, it will be well positioned to address the gap by reviewing and adapting skill requirements, education opportunities, and immigration policies.

Te Waihanga commissioned Scarlatti and Alta Consulting (we/us) to undertake this research programme. The aim was to build an evidence-based foundation for future decision-making about infrastructure workforce matters such as skills, education, and immigration policies. The objectives of this project were to:

- Holistically define the infrastructure workforce
- Estimate the current state supply for the infrastructure workforce
- Estimate the current and future demand for the infrastructure workforce.

Understanding workforce demand and supply

We proposed an approach to model both the supply and demand of the infrastructure workforce. Limitations in the available data mean both the supply and demand models are estimates of the size of the workforce, rather than perfect representations. The intention was to undertake a calibration exercise to reconcile the workforce estimates from the two models to present a 'single source of truth' estimate for the infrastructure workforce. Reconciliation of the models was therefore key to deriving a single estimate.

We reconcile the two models at the level of full-time equivalents (FTEs)[1] (or equivalently number of hours worked), as opposed to individuals. This means that the number of FTEs is the primary measure used to describe the size of the workforce, albeit that the virtual database created by the supply model consists of individuals.

Description of services

The following services aligned with the three objectives and were accepted by Te Waihanga:

1. **Defining the infrastructure workforce.** We developed a framework to define the infrastructure workforce drawing on a variety of publicly available datasets and expert knowledge (Alta Consulting, Construction Sector Accord, and Te Waihanga). A broad scope was used to define

[1] The calculation of FTE is an employee's scheduled hours divided by the employer's hours for a full-time work week. When an employer has a 40-hour work week, employees who are scheduled to work 40 hours per week are 1.0 FTE. Employees scheduled to work 20 hours per week are 0.5 FTE.

the workforce such that the output of this work gives users the ability to filter and query the workforce capacity model in ways that, in effect, allow them to narrow the scope to suit their specific purpose (i.e. define the infrastructure workforce relevant to them).

2. **Estimating the size and shape of the infrastructure workforce.** We estimated the size and shape (e.g. attributes – region, occupation, etc.) of the infrastructure workforce by considering both supply and demand side dynamics. We quantitatively estimated the supply and demand of the infrastructure workforce as follows:

   - The supply estimates were modelled using datasets accessed through Statistics New Zealand's Integrated Data Infrastructure (IDI) to estimate the real-world size and capacity of the infrastructure workforce as per the workforce definition framework. These were developed into a database of virtual workers with attributes matching real-world demographic distributions.

   - The demand estimates were modelled using a bottom-up approach that breaks down a project's value by the elements required to complete it, and the occupations required to undertake the elements. The approach used infrastructure pipeline data to estimate full-time equivalent (FTE) requirements for the infrastructure workforce according to the workforce definition framework.

   We then compared the two models against one another to understand where and why there were significant differences and used this to refine our assumptions in each model. A full discussion, along with guidance as to how the two models should be used of this is given in *Reconciliation of supply and demand models.*

3. **Engagement with internal and external stakeholders.** The supply and demand estimates were periodically reviewed and informed by engaging with stakeholders and end-user groups to ensure the outputs meet their needs. We engaged with the following groups:

   - A project reference group with representatives from industry organisations including Te Waihanga, Scarlatti, Alta Consulting, Waihanga Ara Rau and the Construction Sector Accord.

   - End-users and key partners to provide insights and judgement on the inputs used in modelling supply and demand estimates.

   - Other infrastructure stakeholders were engaged passively via newsletters and social media to be updated of on progress and the availability of project outputs.

Project deliverables

The project deliverables are presented in Table 1.

Name	Description	Date
Deliverable 1: Initial reference group workshop and targeted interviews	Confirm scope, timeframe, and deliverables for project, and identify any dependencies. Targeted interviews should achieve adequate coverage of key stakeholders across infra sectors.	23/12/2022
Deliverable 2: Draft report on defining the infrastructure workforce	Is informed by internal and external engagement. Provides a holistic view across all stages of infrastructure planning and delivery.	31/03/2023
Deliverable 3: Draft report and dataset on estimating the size and shape of the infrastructure workforce	Is informed by internal and external engagement. Uses a transparent and replicable approach.	31/05/2023
Deliverable 4: Final report on defining and estimating the size and shape of the infrastructure workforce	Incorporates feedback provided on draft reports.	31/07/2023
Deliverable 5: Final infrastructure workforce dataset and framework	Is provided in a non-proprietary form that can be used and applied by Te Waihanga in future.	31/07/2023

Table 1: Infrastructure workforce capacity: Baselining study project deliverables

Project reference group

To support and guide the project's progress, we established a reference group of members from Te Waihanga, Waihanga Ara Rau, and the Construction Sector Accord (Table 2). The reference group was engaged periodically to contribute to the development and review of the deliverables.

Name	Organisation
Peter Nunns	Te Waihanga
Matthew Keir	Te Waihanga
Hemant Passi	Te Waihanga
John Hemi	Te Waihanga
Mark Williams	Waihanga Ara Rau
Graham Burke	Construction Sector Accord

Table 2: Members of the project reference group (in addition to Scarlatti and Alta Consulting as the researchers)

The infrastructure workforce definition

Introduction

A framework was developed to define the infrastructure workforce. We drew on publicly available datasets, Te Waihanga's project pipeline database, the ANZSCO (Australia and New Zealand Standard Classification of Occupations) and ANZSIC (Australian and New Zealand Standard Industrial Classification) standards, as well as expert knowledge from the reference group. The definition is an important prerequisite to estimating the supply of and demand for the workforce, as it allows us to understand and justify what is included and excluded.

Infrastructure workforce definition

We start by defining the infrastructure workforce as those people who contribute labour directly to the planning, construction or asset management of horizontal and vertical infrastructure assets. We further characterise the workforce by considering the following data 'dimensions':

- Occupation
- Geographical area
- Project (or asset) sector – for example, an infrastructure project in water, transport or waste
- Stages of work
- Demographics
- Industry.

Each dimension was characterised in collaboration with the reference group and other industry stakeholders to establish mutually exclusive, collectively exhaustive lists to segment the workforce. We discuss the scope of each dimension in the following sections.

Occupation

To compare supply and demand side figures like-for-like, we must settle on a consistent single source of truth to determine which occupations are included in the workforce definition. To this end, we used the ANZSCO framework, which groups similar occupations in a hierarchical fashion. We made one amendment to this framework, introducing an 8-digit 'sub-occupation' tier below the standard 6-digit occupation tier. This allowed us to model certain occupations with greater precision where we felt that the standard ANZSCO code lacked the desired specificity.

We considered the in-scope infrastructure workforce occupations to be those that contribute directly to the planning, construction or asset management of horizontal and vertical infrastructure assets (i.e. those that are costed directly to a project's budget). The following types of occupations were not considered in scope for the infrastructure workforce:

- **Support functions.** Occupations that act as support functions in infrastructure businesses (e.g. marketers, human resource managers, accountants, etc) were not considered in scope for the infrastructure workforce due to their transferrable skills across sectors and because their salaries/wages are covered by the margins in a project rather than directly costed.

- **Operators of infrastructure assets**. The operators of infrastructure assets (e.g. train drivers) were not considered in scope for the infrastructure workforce except insofar as they are involved in maintenance or renewal activities. The occupations should be seen as a third-party user of infrastructure rather than as part of the workforce involved in the asset lifecycle.
- **Offsite manufacturers**. The occupations of people involved in the manufacturing of products offsite were not considered in scope for the infrastructure workforce. The approach to model the demand for the infrastructure workforce first breaks down a project's value by the costs attributed to labour, materials, and margins. The products manufactured offsite are brought onsite as a material cost rather than a labour cost.

See Appendix 1: Occupation and workforce size for a full list of in-scope occupations including ANZSCO code and workforce size.

Geographical area

We used New Zealand's 16 regional and unitary council[2] boundaries as the basis to define the geographical area of the infrastructure workforce. In cases where the data used to model the supply (IDI) and demand (project pipelines) specified the location of a project with more precision (e.g. district council or territorial authority boundary), the more detailed location was used.

It is important to note that work on a project may be done in a different region to that where the project is physically located. This is especially true for professional services workers who predominantly live and work in the major New Zealand cities. Unless corrected for, this would appear to create an oversupply of these workers in Auckland, Waikato, Wellington and Canterbury regions, and an undersupply in all others.

We considered correcting for this by applying an adjustment to the 'region' characteristic of the demand model's labour hours output, based on the nature of each occupation. However, this adjustment came with its own limitations and would not have significantly improved the model's accuracy. As this project focusses on the workforce at a national level, this was not a priority. This could be an area for future work.

Project sectors

Projects in different infrastructure sectors have significantly different workforce requirements. To accurately model the supply and demand for the infrastructure workforce we must specify model parameters for projects in each sector accordingly. This necessitated a consistent, precise definition of the sectors, and which projects belong to which sector.[3]

Te Waihanga's infrastructure project pipeline was used as the primary dataset to estimate the demand for the infrastructure workforce (see page 26). We therefore chose to use the same classification system to classify the project sectors (Table 3). See Appendix 2: Project sectors for a full description of this classification.

---

[2] Northland, Auckland, Waikato, Bay of Plenty, Gisborne, Taranaki, Hawkes Bay, Manawatu-Whanganui, Wellington, Tasman, Nelson, Marlborough, West Coast, Canterbury, Otago and Southland.

[3] We note here a subtle but important distinction between a project's sector and its owner. For example, the 'Let's get Wellington Moving' project falls under the transport *sub-sector*, but its *owner* is defined as local government.

Sector	Sub-sector	Examples of project types
Horizontal infrastructure	Water sector	Irrigation, storm water, and wastewater
Horizontal infrastructure	Waste sector	Carbon management, collection, and landfill
Horizontal infrastructure	Transport sector	Airport, rail, and road
Vertical infrastructure	Social sector	Health, housing, and justice
Horizontal infrastructure	Energy sector	Electricity, gas, and liquid fuels
Vertical infrastructure	Education and research sector	School, tertiary, and science and research
Vertical infrastructure	Community sector	Community, recreation, and venue
Horizontal infrastructure	Communications sector	Data centre, switching, and transmission
Vertical infrastructure	Commercial	Commercial, logistics, and manufacturing

Table 3: Classification of project sectors as defined by Te Waihanga's infrastructure project pipeline

We used the columns in Table 3 to establish a hierarchical classification of projects and respective work requirements. Demand-side modelling was carried out at the level of *project sub-sector* and *project type.*

Stages of work

The brief for this project stipulated a holistic approach to defining the workforce which includes the three stages of asset lifecycle:

1. The **planning and design** of an infrastructure asset
2. The **construction** of the asset
3. The ongoing **maintenance** of the asset.

These sequential stages have significantly different characteristics and are susceptible to labour market dynamics in different ways. In particular, the distribution of occupations varies significantly between stages of an infrastructure project. We broke down the three stages into their constituent elements, and estimated the distribution of labour for each element. Each stage is discussed in more detail below, and a full treatment of the assumptions and calibration processes used is given in the *Methodology: Modelling the demand for the infrastructure workforce* section.

The timelines in the figures that follow span from the beginning of planning to the time of delivery of the asset – the maintenance stage is discussed separately. The development of the timelines was informed by our discussions with industry experts and Alta Consulting. Accordingly, we identified that the elements of the planning and design stage take place over the first half of the lifetime of a project, while the construction stage occurs during the second half. The experts we consulted noted that while not all projects conform to this breakdown, this is nevertheless a general and realistic approximation of the distribution of work across the project pipeline.

Planning and design

The planning and design stage consists of three elements: planning, property, and procurement[4], consenting, and design. These elements continue throughout a project's lifetime (up to asset delivery) to varying extents of intensity. Figure 1 shows the phasing of planning and design elements over the lifetime of a project. Note that although the weightings of elements differ between project types, the phasing is assumed to be the same for all.

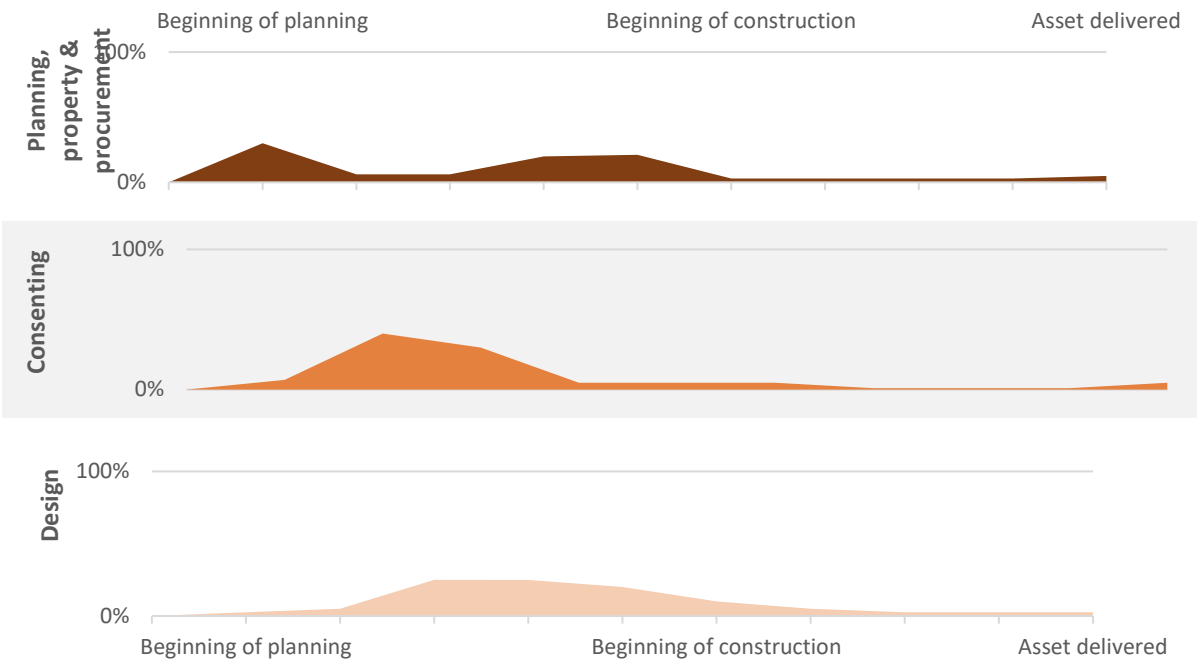


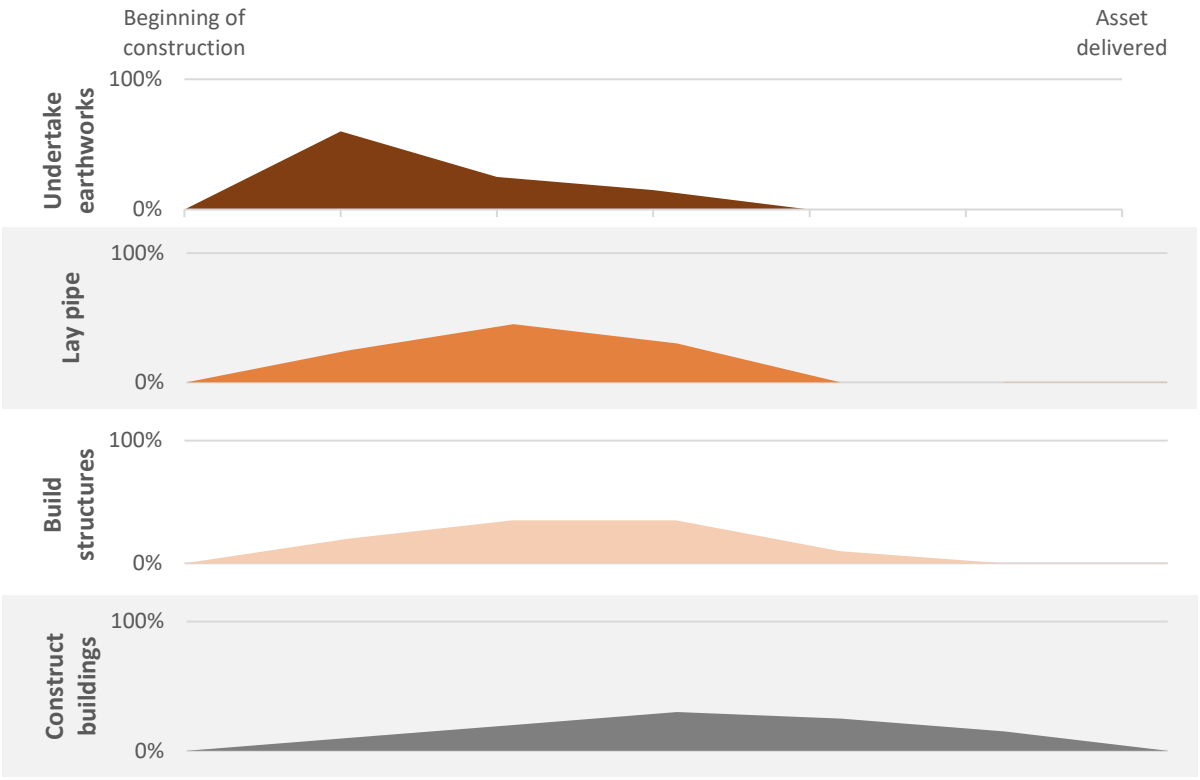
Figure 1: Phasing of planning and design elements over the lifetime of a project (note the difference in timescale compared to Figure 2 below)

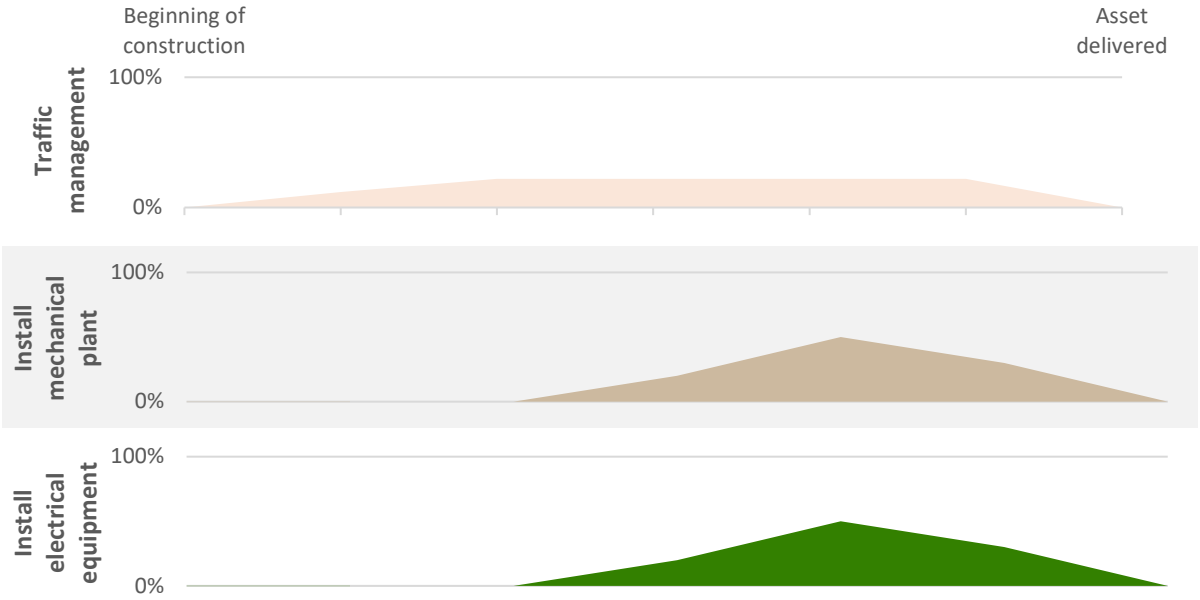
Construction

The construction stage consists of nine elements: undertaking earthworks, laying pipe, building structures, constructing buildings, traffic management, installing mechanical plant, installing electrical equipment, and placing pavement. Not all elements will be relevant to all project types. For instance, a roading project would not be expected to include a 'construct buildings' element.

Figure 2 shows the phasing of construction elements over the lifetime of a project.

[4] This element accounts for the labour involved in business case development, however it should be noted that some business case development activities also overlap with the consenting and design elements.





SCARLATTI ALTA

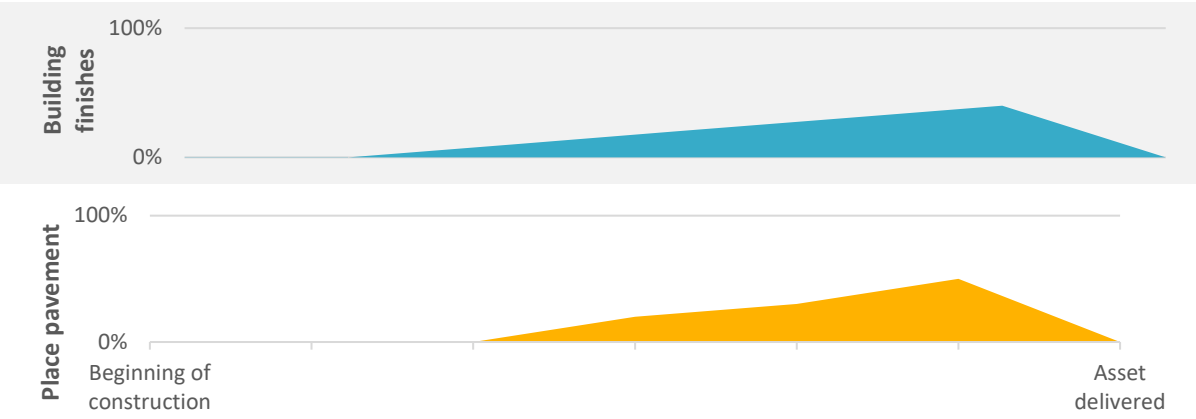


Figure 2: Phasing of construction elements over the lifetime of a project (note the difference in timescale compared to Figure 1 above)

Maintenance

Unlike the planning and design and construction stages, the maintenance stage has no well-defined timeframe. We assume an asset gets one quarter worth of routine maintenance each year and one quarter worth of moderately intense maintenance every five years. The amount spent in each of these maintenance periods is defined as a percentage of the project's initial cost, varying depending on the project's sector (see

Asset type/subsector	Total asset value ($)	Actual ratio of maintenance / renewal spend to asset value	Estimated ratio of maintenance / renewal spend to asset value, based on depreciation trends	Ratio of labour cost to total project cost	Adjustment for maintenance / renewal projects in NIP	Average annual maintenance / renewal labour cost as % of asset value	Notes/sources
Water supply	13,721,287,059	NA	1.40%	45%	25,905,965	0.44%	National Accounts average CFK/NKS ratio for local government water,

							sewerage, drainage, and waste services for 2013-2022 period, marked down by 40% to account for under-renewal trend observed in Council LTPS
Stormwater	25,418,960,599	NA	1.40%	45%	12,057,357	0.58%	National Accounts average CFK/NKS ratio for local government water, sewerage, drainage, and waste services for 2013-2022 period, marked down 40% to account for under-renewal trend observed in Council LTPS
Wastewater	33,857,306,326	NA	1.40%	45%	29,742,644	0.54%	National Accounts average CFK/NKS ratio for local government water, sewerage, drainage, and waste services for 2013-2022 period, marked down 40% to account for under-renewal trend observed in Council LTPS
Road	170,855,574,604	1.40%		39%	84,992,812	0.50%	Waka Kotahi actual local road and state highway maintenance spending
Port	13,000,000	NA	2.00%	39%		0.78%	No relevant data on either depreciation or maintenance/renewal expenditure. Midpoint of roads and electricity used.
Airport	9,021,900,000	NA	2.80%	50%	9,965,538	1.29%	No relevant data on either depreciation or maintenance/renewal expenditure. Midpoint of roads

							and education and research used.
Commercial	343,000,000	NA	2.50%	50%	80,669	1.23%	No relevant data, used mean of other sectors
Waste	218,977,161	NA	4.50%	45%	14,078	2.02%	National Accounts average CFK/NKS ratio for commercial/private water, sewerage, drainage, and waste services for 2013-2022 period, marked down by ~15% to account for under-renewal trend observed in Council LTPS
Electricity	47,166,000,000	2.60%		44%	16,229,499	1.12%	Actual maintenance spend from infrastructure network capital stock and investment baseline data
Gas	1,064,000,000	1.00%		36%	657,390	0.30%	Actual maintenance spend from infrastructure network capital stock and investment baseline data
Rail	10,332,000,000	NA	1.40%	46%	31,117,435	0.34%	National Accounts data includes vehicles and above-track infrastructure, resulting in a higher depreciation ratio. Value for roads used as a proxy.
Social	67,454,368,202	NA	3.40%	49%	156,201,846	1.43%	National Accounts average CFK/NKS ratio for central government hospitals for 2013-2022 period, marked down by ~15% to account for under-renewal

Community	20,153,012,989	NA	3.80%	49%	73,662,340	1.50%	Midpoint of social and education/research
Education and Research	15,350,000,000	NA	4.20%	49%	15,829,302	1.95%	National Accounts average CFK/NKS ratio for central government preschool, school education and tertiary education for 2013-2022 period, marked down by ~15% to account for under-renewal
Other	6,000,000	NA	2.50%	49%		1.23%	Unsure what this category includes – use whatever seems most relevant.
Communications	9,521,399,756	NA	7.00%	44%	388,202	3.11%	National Accounts data for telecommunications services suggests a depreciation ratio of 9.5%, reflecting greater role of software and shorter cycles of technological change. A lower figure has been used to better align with the IRD's range of asset life estimates for fixed telecommunications assets.
Liquid fuels	607,260,188	NA	2.60%	46%	0	1.20%	Use figure for electricity sector
Irrigation	607,260,188	NA	2.00%	44%	0	0.88%	Use figure for water supply
Protection	1,214,520,377	NA	2.00%	44%	3,803,540	0.57%	Use figure for water supply

in Appendix 7: Breakdown of existing asset register by asset type

). This continues for 20 years after the project finishes. A more detailed discussion of the validity of these assumptions is given in the calibration section of the demand model methodology chapter.

For simplicity, we assume (based on the expert judgement of Alta Consulting) that the occupations involved in maintenance work are the same as for the equivalent elements of the construction stage. Therefore, we model the maintenance stage as a repetition of the superficial elements of the construction stage (constructing buildings, installing mechanical and electrical equipment, and building finishes). The distribution of work across these elements matches the distribution used during the construction stage for that project type and sub-sector. A full treatment of these assumptions are given in the *Methodology: Modelling the demand for the infrastructure workforce* section.

Demographics

The above sections describe the dimensions we have used to *define* the workforce. We consider other characteristics of the workforce in this section, drawing the distinction that these do not form part of the criteria for defining who is part of the workforce. We used IDI data to explore the distributions of the following characteristics (see Appendix 3: IDI queries):

- Age
- Gender
- Ethnicity
- Tenure in current occupation
- Highest qualification
- Visa status
- Size of employing firm
- Industry of employing firm.

Our virtual workforce was constructed such that:

- Each virtual worker had a value for each of these attributes; and
- The virtual workforce as a whole is representative of the wider New Zealand population distribution.

The virtual workforce produced in this work may be used in the future to better understand the segmentation of the workforce, for example, to give insight into the state of training, diversity, or immigration.

Industry

We note here an alteration of our planned approach. Initially, we intended to use industry (defined by ANZSIC codes) as a filtering dimension. That is, to only consider in-scope those individuals whose employing firm belonged to an industry considered 'in-scope'. However, we found that this approach was flawed for two reasons. Firstly, it did not help to segment the supply-side workforce in a sensible way for comparison with demand. This is because firms in some industries (e.g., electrical services) supply labour to both infrastructure and residential construction, as well as serving non-construction

businesses such as manufacturing firms. Therefore, the boundaries of the *infrastructure* workforce do not align with the boundaries of industries. Secondly, we found that many relevant occupations had a long tail of individuals in seemingly irrelevant industries (e.g., plumbers working for religious services firms). By excluding these individuals, we were systematically underestimating the size of the available workforce. The magnitude of this effect varied greatly between occupations, so applying a simple correction factor was not an option. We therefore decided that it would be best to retain industry as an observed characteristic, but not use it to set the boundaries of the workforce itself.

With industry no longer used as a filter, another mechanism was required to restrict the workforce to those relevant to infrastructure. We approached industry experts to understand the breakdown of workers between the vertical infrastructure, horizontal infrastructure, and residential construction sectors. We also developed a measure called *specificity*, which described the propensity for individuals in a given occupation to work in more than one sector simultaneously (e.g., an architectural draftsperson has *high* specificity, as their skills tend only to be applicable to the sector they've trained for, while a general labourer has *low* specificity as they can easily supply labour to whichever sector requires it).

In addition to the idea that a single worker could split their time between sectors, we recognised that some individuals only work part-time. This is accounted for in the supply model by making reference to household labour force survey data regarding usual hours worked per week across the relevant industries and occupations. Both factors contributed to the outcome that the number of *worker*s reported in our outputs tended to be significantly higher than the number of *FTEs.*

Methodology: Modelling the supply of the infrastructure workforce

Introduction

On the supply side, we used datasets accessed through the IDI to estimate the real-world size and capacity of the workforce under the framework. These were developed into a database of virtual workers characterising the entire infrastructure workforce.

IRD and census data (March 2018) were used to explore the demographic characteristics of the workforce. The virtual individuals in the output database were assigned attributes according to the real-world New Zealand demographic distributions.

Relevant workforce dimensions

The supply of the infrastructure workforce (i.e. each virtual worker) has been characterised with respect to the following workforce dimensions:

- The **occupation** of the virtual worker
- The **industry** the virtual worker is employed in
- The **sector(s)** (i.e., vertical, horizontal, residential, or other) the virtual worker is employed in
- The **region (i.e., geographical area)** the virtual worker is situated
- The **demographics** of the virtual worker (e.g., age, gender, ethnicity, etc.).

Model framework and methodology

The framework to model and estimate the supply of the infrastructure workforce involved the following high-level stages:

- Stage 1: Identify relevance to workforce definition
- Stage 2: Retrieve data from IDI
- Stage 3: Generate characteristic distribution from IDI data
- Stage 4: Assign employees labour hours for each sector.

Stage 1: Identify relevance to workforce definition

We began by defining the boundaries of the workforce. We considered only those individuals whose occupation at the time of the 2018 census was on the list of in-scope occupations (Appendix 1: Occupation and workforce size). Industry, geographical area, and demographics were not used to restrict the population but were all deemed important information to record – see below. Note: for the remainder of this document, we refer only to the horizontal and vertical sectors when describing 'total (infrastructure) workforce', however the output database also contains entries for FTEs in the residential and 'other' sectors.

Key inputs and assumptions to stage 1

The key input to stage 1 was the workforce definition described in the previous section. This was developed via an iterative feedback process between the project reference group (Table 2).

Stage 2: Retrieve data from IDI

We identified a 'base population' from 2018 census data. This consisted of individuals who:

- Worked in a relevant occupation
- Earned at least 30 hours of minimum wage equivalent per week[5]
- Worked for at least 1 month that year in their relevant occupation.

We then queried this base population to find the joint distributions of our ten attributes (occupation, industry, age, gender, region, ethnicity, visa status, firm size, highest qualification, and tenure). Noting that an exhaustive search of all possible attribute triples would require 120 queries, we then identified pairs and triples of attributes we believed to have nontrivial correlations. For example, occupation is strongly correlated with an individual's industry, ethnicity is strongly correlated with visa status, but age is not strongly correlated with ethnicity for working age people[6].

IDI data is subject to strict confidentiality rules. Outputs will have rows suppressed where the total number of individuals is sufficiently small ($n < 6$), therefore attributes were only queried two or three at a time.

Finally, we queried the household labour force survey dataset for 2018 to understand the distribution of hours worked per week across the relevant occupations. As the household labour force survey is representative, the population definition does not apply in the same way as the census-based queries. The hours-worked distribution is applied as a type of post-processing step to the virtual workforce (step 4).

Key inputs and assumptions to stage 2

The key inputs to this stage were the data tables accessed through the IDI including 2018 census data, IRD records, and HLFS survey responses. See Appendix 3: IDI queries for a full description of the queries used in this stage.

Stage 3: Generate characteristic distribution from IDI data

We used outputs from IDI queries to estimate the overall attributes of the infrastructure workforce, including how the workforce breaks down by occupation, industry, sector, region, and demographic characteristics. To do so, we used a statistical technique to cross-reference between separate IDI queries. We considered a 10-dimensional 'characteristic space' consisting of all possible combinations of values for each characteristic, and completed this using the *iterative proportional fitting procedure (IPFP).* A simplified explanation of the algorithm is as follows:

---

[5] We expand on this point later to account for part-time workers; the 30-hour per week threshold is simply used to establish a consistent baseline and minimize the impact of data errors.

[6] At the population level, the median age of New Zealanders in the European ethnic group (41.4 years) is significantly higher than that of the Māori (25.4 years) and Pacific (23.4 years) groups. However, this difference is primarily driven by higher proportions of children and lower proportions of elderly people in these groups, which are not relevant to the demographics of working age people as they pertain to this report.

1. Initialise an empty matrix whose dimensions match the dimensions of our characteristic space
2. Set random initial values for the cells
3. Use IDI data to set desired row and column totals across all dimensions
4. For each cell, calculate an adjustment factor according to the respective row and column totals
5. Adjust the value in the cell by multiplying it with the calculated adjustment factor
6. Repeat steps 4 and 5 for all cells in the table
7. Check if the table now satisfies the desired constraints
8. If the constraints are not met, repeat the process from step 4 until the table converges to a solution that satisfies the constraints or until a specified stopping condition is reached.

The resulting matrix can be thought of as a probability distribution over every unique combination of attributes. Specifically, it is the maximum-likelihood probability distribution for the observed characteristics in our IDI query population. While at this stage this model does not provide the ability to conduct more detailed analyses (e.g., gender/ethnicity pay gaps), these attributes could be added to enable this in the future. This would require a subsequent IDI project.

Key inputs and assumptions to stage 3

A full treatment of the validation of the IPFP model is given in Appendix 4: Detailed analysis of IPFP. The key takeaway from this is that we believe the model faithfully preserves correlations between dimensions even when they are not explicitly supplied as inputs, and that correlation to real data is very high for combinations of up to three dimensions. Filtering by more than three dimensions is possible, but caution should be exercised in using any output at this level. Note this is also generally the level at which IDI queries start to face widespread suppression issues. In practice, we do not expect that this will be a significant limitation as it is difficult to imagine a use case in which more than 3 filters are required. In addition to the number of dimensions, it is important to consider the size of the cells produced by filtering – correlation to the real data decreases with mean attribute bucket size (see *Limitations of model framework and methodology*).

Stage 4: Assign employees labour hours for each sector

We created a set of 'virtual workers' by drawing (with replacement) samples from the characteristic distribution created in stage 3, until the number of virtual workers was equal to the total of individuals across all relevant occupations[7].

For each virtual worker, we did the following:

- Sample the HLFS hours worked per week distribution for that worker's occupation (we were limited to assuming that hours worked per week was dependent only on occupation due to the small sample size of HLFS preventing more detailed cross-tabulation). Divide this by 40 to get the worker's total FTE number.

[7] As this is a stochastic process, subsequent runs of the supply model will not produce identical estimates. However, we are confident that the magnitude of error introduced by this randomness is small compared to the other sources of error introduced by our assumptions. See *Appendix 4: Detailed analysis of IPFP* for a further discussion of this.

- Randomly choose a *primary* sector, weighted according to the sector distribution for that occupation (see Appendix 1: Occupation and workforce size). If the worker's occupation has *high* specificity, assign all the worker's total hours to this sector. If the worker's occupation has *medium* or *low* specificity, for each remaining sector do the following:
- Generate a random number between 0 and 1. Halve this number if the specificity is *low.*
- If this number is smaller than the proportion of workers in that sector, assign 25% of the worker's hours to that sector, otherwise do nothing.
- Assign the worker's remaining hours to their primary sector.

Worked example:

ANZSCO code	Occupation	Degree of sector specificity	Breakdown of an occupation by sector:			
			Vertical	Horizontal	Residential	Other
331212	Carpenter	Moderate	22%	8%	67%	3%

Excerpt from Appendix 1: Occupation and workforce size

Taking a virtual worker whose occupation is carpenter, we sampled the HLFS hours worked distribution for carpenters and returned a value of 40 hours per week, or 1.0 FTE. We then made a weighted random choice of the four sectors to determine that the carpenter's primary sector was *residential.* For each of vertical, horizontal, and other, we generated a random number between 0 and 1, and compared that to the sector breakdown value (see Excerpt from Appendix 1: Occupation and workforce size), e.g.,

- $rand(0,1) \rightarrow 0.13 < p_{vertical} = 0.22$. Assign 0.25 FTE to the *vertical* sector
- $rand(0,1) \rightarrow 0.76 > p_{horizontal} = 0.08$. Assign 0 FTE to the *horizontal* sector
- $rand(0,1) \rightarrow 0.44 > p_{other} = 0.03$. Assign 0 FTE to the *other* sector.

The output of this process in the virtual workforce was a carpenter who works full-time, 0.75 FTE in the residential sector and 0.25 in the horizontal sector (see row 2 in Table 4). These figures should be treated as an average over time, i.e., we can understand this distribution as the virtual carpenter working on a mix of approximately three quarters residential homes, and one quarter hospitals, schools, etc. (rather than specifically 30 hours on houses and 10 hours on hospitals every week).

Key inputs and assumptions to stage 4

The key input to this stage was the expert estimation of occupation sector splits, see *Appendix 1: Occupation and workforce size*. Collection methodology: send out the spreadsheet to experts to enter their estimates, calculate the average. Consult with industry stakeholders where confidence is low, or where estimates differ by more than 25%.

The assignment to sectors was done with independent random draws in order to preserve the sector distribution for each occupation at the population level. We validated that the distribution of FTEs between sectors in the modelled outputs was faithful to the distribution defined by the occupation sector map, with a root-mean-square error of 2 percentage points.

Limitations of model framework and methodology

The limitations of the above framework and methodology to model the supply estimates include:

- **Currency.** The only reliable data source for count of workers by occupation is the census, which occurs only every five years. The supply model represents a point-in-time estimate as of the 2018 census. For consistency, other data sources used in the supply model (IRD records, HLFS data) were also captured at March 2018.
- **Reduced accuracy at high granularity.** The modelled supply outputs lose accuracy the more granular the outputs are viewed. Based on the analysis in Appendix 4: Detailed analysis of IPFP, we recommend that caution should be used with outputs using more than 3 dimensions, or where the mean attribute bucket size is less than 50.
- **Workforce definition.** The mapping of occupations to sectors (i.e., horizontal, vertical, residential, and other) is highly dependent on what is considered in-scope for each sector; different sector definitions could produce very different results.
- **Sector specificity.** The concept of sector specificity, as defined in this project, is not well-studied and therefore difficult to validate other than by expert opinion. The algorithm for assigning workers' time to multiple sectors is a simple heuristic only, prioritising maintaining correct sector proportions more so than accurately representing the way workers split their time.

Structure of outputs

The output was a database of virtual workers. The number of rows was equal to the size of the population in the IDI queries. Each row had an entry for all 10 of the observed characteristics and a full-time equivalent (hours worked per week / 40) value for each sector.

ID	Occupation	Region	Industry	FTE split			
				Vertical	Horizontal	Residential	Other
1	Architect	AKL	Architecture	0.75	0.25	0	0
2	Carpenter	WLG	Residential building	0.25	0	0.75	0
3	Electrician	CAN	Building completion	1	0	0	0
…	…	…	…	…	…		
142,775	Carpenter	WAI	Residential building	1	0	0	0

**For each virtual worker, we assigned them a range of demographics (e.g. age, gender, ethnicity, tenure, highest qualification, etc) which sit to the right of this table that are modelled using data extracted from the IDI to ensure the virtual population is representative of the New Zealand population.*

Table 4: Structure of virtual workforce database

Summary of outputs

The total size of the virtual infrastructure workforce is 107,551 FTEs (157,003 workers).[8] Note that this number represents the supply-side estimate only; a treatment of how this number reconciles to the demand model estimates is given later in its own section. We also expect this number to change as more data is made available, particularly on the release of data from the 2023 census. The database of virtual workers is rich in detail and includes a range of attributes for each virtual worker as illustrated in the previous section. The attributes are available for the end-user to characterise the workforce in a range of different ways depending on what's important to them. Below, we have included a few analyses that describe the modelled workforce by sector, region and occupation.

Virtual workforce by sector

A breakdown of the 107,551 FTEs (157,003 workers) by the sector they are modelled to be working in is shown in Figure 3.

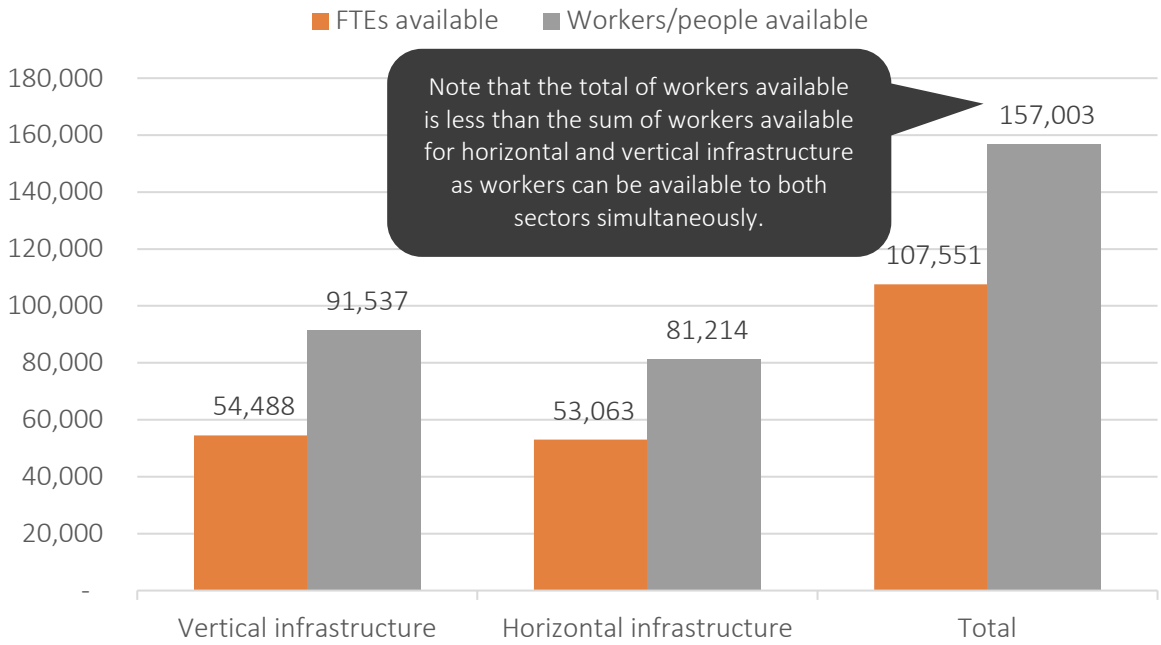


Figure 3: Total count of FTEs and workers available by sector

Virtual workforce by region

A breakdown of the 107,551 FTEs by region is illustrated in Figure 4.

[8] Note: This estimate is based on a particular run of the model described above. Re-running the model with identical input data results in small variations (<1%) to the total workforce size. As a result, reported estimates may differ slightly from this figure.

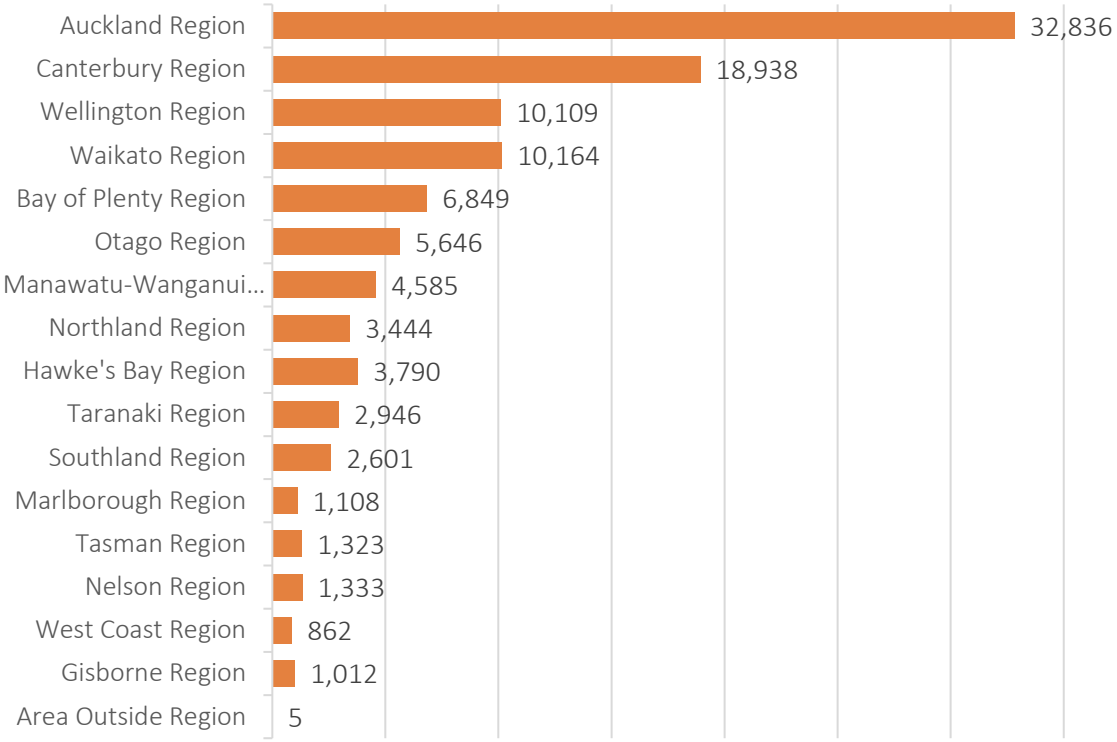


Figure 4: Total count of FTEs by region

Virtual workforce by occupation

A breakdown of the 107,551 FTEs by occupation is presented in Figure 5 . Only the top ten occupations are shown, which represents approximately 50% of the total modelled workforce.

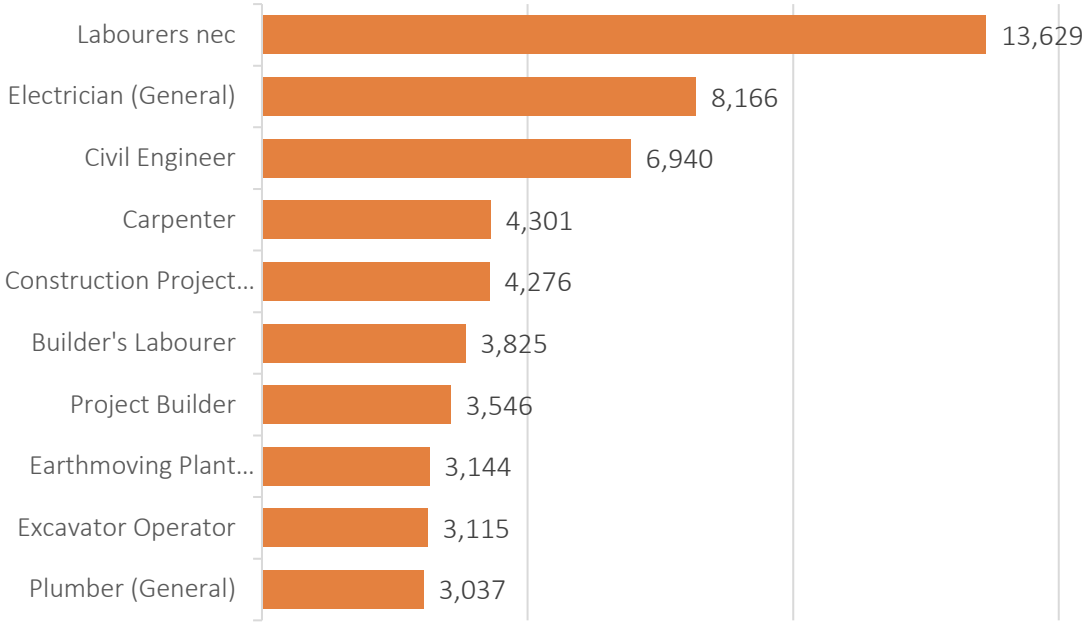


Figure 5: Total count of FTEs by top ten occupations

Calibration of supply model outputs

Note that the joint calibration of the supply model and demand model is discussed separately later in the report; this section refers only to the calibration of the supply model in isolation. There are two key calibration steps in this section; the validation of the IPFP model, which has been described in detail in *Appendix 4: Detailed analysis of IPFP*, and the use of expert judgement in the creation of the occupation sector map.

To increase confidence in these estimates, we asked three members of the reference group (John Hemi and Graham Burke), with support from Alta consultants, to review and update the proportions. These reviews were conducted independently, to minimise the impact of anchoring bias. We then took the average of the three estimates for each occupation, and highlighted instances where the difference in reported proportion was greater than 25%. There were ten such occupations, although six of these had fewer than 250 total individuals, and therefore were not expected to have significant impact on the outcome of the supply model. The four largest of these misaligned occupations (earthmoving plant operator, telecommunications technician, interior designer, and telecommunications engineer) were subsequently reviewed with industry stakeholders.

Methodology: Modelling the demand for the infrastructure workforce

Introduction

We used Te Waihanga's project pipeline data to measure the workforce demand in the infrastructure sector. This database holds a record of past, present and future infrastructure projects in New Zealand. For each project, where available, the database includes details of project sector, asset type, duration, and value. Using this database, we took a bottom-up project approach to estimate FTEs across the workforce which we discuss further in the section that follows Figure 6.

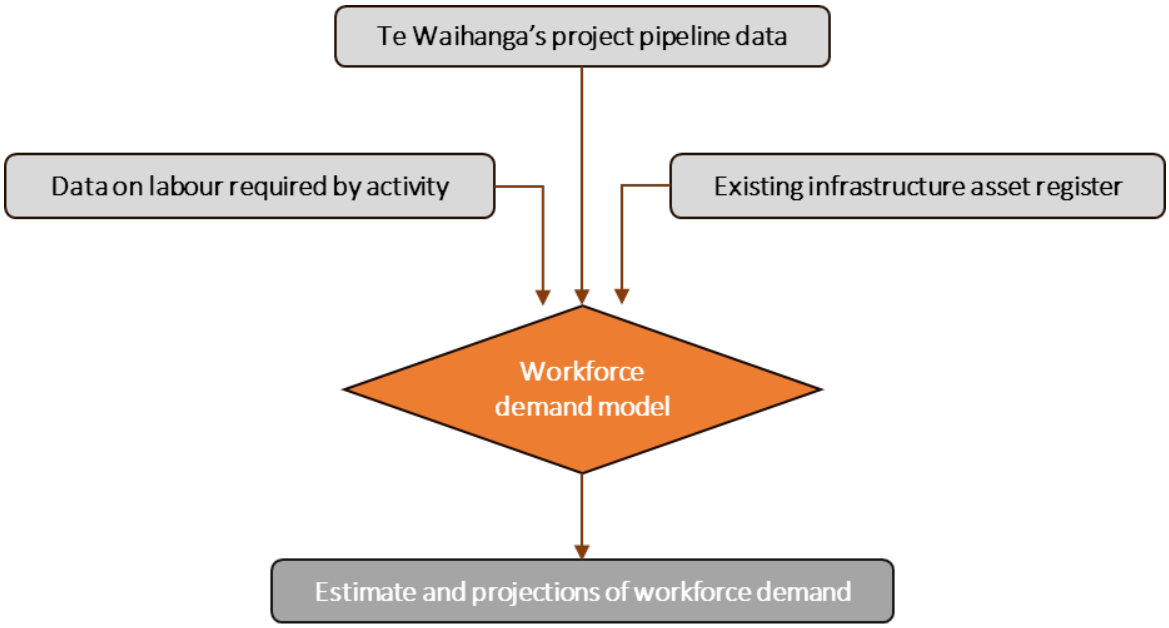


Figure 6: Infrastructure workforce demand model

Relevant workforce dimensions

The demand for the infrastructure workforce was characterised with respect to the following workforce dimensions:

- The **occupations** required to undertake activities in present and future projects
- The **project sector and sub-sector** where workers will be required
- The **region (i.e., geographical area)** where workers will be required
- The **project stage** which workers will be required for.

Model framework and methodology

The framework to model and estimate the demand for the infrastructure workforce involved the following high-level stages:

- Stage 1: Collect demand model inputs.
- Stage 2: Develop existing asset register.

- Stage 3: Process pipeline and existing asset data through demand model.

Stage 1: Collect demand model inputs

To predict the labour requirements for an infrastructure project, we developed the following estimates:

- The proportion of a project's value attributed to direct labour: we broke down the total project value into labour cost across planning and design, construction, and maintenance stages.
- The elements relevant to different project types/sub-sectors: we split each stage into elements, and weighted contribution of elements to total project value according to type (e.g., the majority of a stormwater project consists of laying pipe, the majority of a roading project is comprised of earthworks and traffic management). See Appendix 5: Demand model input tables.
- The occupations required to complete each of the elements and to what extent: we broke down each element into the occupations that contribute to it (e.g., traffic management consists of traffic management workers and general labourers).
- When the different elements are expected to be completed during a project's lifetime: we distributed elements across the lifetime of the project.
- The average wage rates for each relevant occupation: we converted dollar-value work requirements to labour hour requirements using an estimated charge-out rate for each occupation. The charge-out rates used have been previously collected QV's CostBuilder tool.

We also attached a *confidence* value to each project type in the project-element breakdown, and each occupation in the element-occupation breakdown. These were scored on a rubric scale with 4 levels:

- Placeholder
- Value estimated with expert knowledge
- Value modelled with reference to reliable dataset
- Value sourced directly from reliable dataset.

When the demand model was run, the minimum confidence value for project type and occupation was retained, meaning every output row in the database contained an indicative 'minimum confidence level'.

See Appendix 5: Demand model input tables for (abridged) versions of these input tables. We also present a brief worked example for a road project in Appendix 6: Example project calculation.

Key inputs and assumptions to stage 1

Inputs were collected via discussion with Alta Consulting. Estimates were calibrated with reference to balance sheets of real-world projects where possible. Proportion of the project's value as labour cost was the most important input as it set the overall magnitude of FTEs required for project work.

Stage 2: Develop existing asset register

The infrastructure workforce is not solely occupied by new project work. Existing infrastructure assets require maintenance, and the labour requirements for this represent a significant portion of the infrastructure workforce's capacity. In order to capture this segment of the workforce in our demand

model, we developed a register of existing infrastructure assets. This was built using the following process:

- Identify major infrastructure owners (e.g., central government, regional councils, state owned enterprises)
- Map these organisations to the project sectors for which they are likely to operate significant infrastructure assets (e.g., Waka Kotahi owns many roads but no hospitals)
- For each organisation-sector pair identified in the previous step, research annual statements (or equivalent publications as available) to determine total value of asset ownership
- Where these data are available only at the national level, apply per-capita scaling to break these down across the 16 regional and unitary council areas.

The total size of the existing asset register used to model the maintenance demand for these existing infrastructure assets was ~ $440B (see Appendix 7: Breakdown of existing asset register by asset type). We sought to triangulate this estimate with other sources, such as Statistics NZ National Accounts although differences in valuation and accounting practices meant this calibration is only approximate – see *Limitations of model and framework* below for more details.

Key inputs and assumptions to stage 2

We assert that major refits and renewals appear in the Te Waihanga pipeline as standalone projects, supported by the fact that just under 10% (3,032 out of 33,979) of project descriptions contain one of the keywords 'renewal', 'maintenance', 'refit' or 'LoSI[9]', representing approximately 8% of the total pipeline value. These will therefore be modelled in full by the regular demand model, including their own planning and construction phases. The *maintenance* stage therefore only refers to regular wear-and-tear remediation and general upkeep, and therefore represents a very small (~ 1% per annum – see

Asset type/subsector	Total asset value ($)	Actual ratio of maintenance / renewal spend to asset value	Estimated ratio of maintenance / renewal spend to asset value, based on depreciation trends	Ratio of labour cost to total project cost	Adjustment for maintenance / renewal projects in NIP	Average annual maintenance / renewal labour cost as % of asset value	Notes/sources

[9] Level-of-service improvement

Water supply	13,721,287,059	NA	1.40%	45%	25,905,965	0.44%	National Accounts average CFK/NKS ratio for local government water, sewerage, drainage, and waste services for 2013-2022 period, marked down by 40% to account for under-renewal trend observed in Council LTPS
Stormwater	25,418,960,599	NA	1.40%	45%	12,057,357	0.58%	National Accounts average CFK/NKS ratio for local government water, sewerage, drainage, and waste services for 2013-2022 period, marked down 40% to account for under-renewal trend observed in Council LTPS
Wastewater	33,857,306,326	NA	1.40%	45%	29,742,644	0.54%	National Accounts average CFK/NKS ratio for local government water, sewerage, drainage, and waste services for 2013-2022 period, marked down 40% to account for under-renewal trend observed in Council LTPS
Road	170,855,574,604	1.40%		39%	84,992,812	0.50%	Waka Kotahi actual local road and state highway maintenance spending
Port	13,000,000	NA	2.00%	39%		0.78%	No relevant data on either depreciation or maintenance/renewal expenditure. Midpoint of roads and electricity used.

Airport	9,021,900,000	NA	2.80%	50%	9,965,538	1.29%	No relevant data on either depreciation or maintenance/renewal expenditure. Midpoint of roads and education and research used.
Commercial	343,000,000	NA	2.50%	50%	80,669	1.23%	No relevant data, used mean of other sectors
Waste	218,977,161	NA	4.50%	45%	14,078	2.02%	National Accounts average CFK/NKS ratio for commercial/private water, sewerage, drainage, and waste services for 2013-2022 period, marked down by ~15% to account for under-renewal trend observed in Council LTPS
Electricity	47,166,000,000	2.60%		44%	16,229,499	1.12%	Actual maintenance spend from infrastructure network capital stock and investment baseline data
Gas	1,064,000,000	1.00%		36%	657,390	0.30%	Actual maintenance spend from infrastructure network capital stock and investment baseline data
Rail	10,332,000,000	NA	1.40%	46%	31,117,435	0.34%	National Accounts data includes vehicles and above-track infrastructure, resulting in a higher depreciation ratio. Value for roads used as a proxy.
Social	67,454,368,202	NA	3.40%	49%	156,201,846	1.43%	National Accounts average CFK/NKS ratio for central government hospitals for 2013-2022 period, marked

							down by ~15% to account for under-renewal
Community	20,153,012,989	NA	3.80%	49%	73,662,340	1.50%	Midpoint of social and education/research
Education and Research	15,350,000,000	NA	4.20%	49%	15,829,302	1.95%	National Accounts average CFK/NKS ratio for central government preschool, school education and tertiary education for 2013-2022 period, marked down by ~15% to account for under-renewal
Other	6,000,000	NA	2.50%	49%		1.23%	Unsure what this category includes – use whatever seems most relevant.
Communications	9,521,399,756	NA	7.00%	44%	388,202	3.11%	National Accounts data for telecommunications services suggests a depreciation ratio of 9.5%, reflecting greater role of software and shorter cycles of technological change. A lower figure has been used to better align with the IRD's range of asset life estimates for fixed telecommunications assets.
Liquid fuels	607,260,188	NA	2.60%	46%	0	1.20%	Use figure for electricity sector
Irrigation	607,260,188	NA	2.00%	44%	0	0.88%	Use figure for water supply

Protection	1,214,520,377	NA	2.00%	44%	3,803,540	0.57%	Use figure for water supply

for a full breakdown) marginal spend with respect to the project's initial value.

Stage 3: Process pipeline and existing asset data through demand model

We prepared project data from the Te Waihanga pipeline to be run through the demand model using the following steps:

- Remove projects which have duplicate IDs, missing information, or are out-of-date (most recent update before 2023Q1)
- Retrieve the project's sector, subsector, and type from a list of gpt-powered classifications[10],[11]
- Determine the project's start, end, and duration. The Te Waihanga pipeline has eight date fields (business case start and end, procurement start and end, construction start and end, and total project start and end). The fields that are completed varies across projects. As a data cleaning heuristic, we simply took the earliest and latest date across all eight fields to be the project's total date range and assumed that construction began at the exact midpoint of this. This estimate was based on expert judgement from Alta Consulting, and an analysis of the subset of projects which did have complete date information - we found that for the majority of projects, the construction start date fell between 40 and 70% of the way through the project's lifetime. Refinements to this filtering process to exclude projects with nonsensical date range may be appropriate, although the magnitude of error introduced by erroneous inclusions is not expected to be large. Additionally, if data coverage can be improved enough that the distinction between planning and construction phases can be reliably determined from the pipeline, it would be desirable to supersede the assumption that construction begins at the midpoint and instead model planning and construction date ranges explicitly.

As the maintenance model for the Te Waihanga pipeline is set up to run for 20 years from the completion date of the project, we needed to assign completion dates to the projects in the infrastructure asset register in order to run the maintenance model for them. Since the asset register represents an aggregation of all assets for a given type or subsector, we could not assign a single completion date to these projects. We therefore assign a completion date of 2020Q1 to all items in the asset register, so that their maintenance will be included in all near-term outputs.

The demand model produces labour hour requirements by occupation for both the pipeline and the asset register separately, applying only the maintenance stage to the existing assets. The outputs of these were then merged, and grouped by occupation, region, quarter, project owner, project type/sub-sector, and project stage.

[10] GPT3.5, via the OpenAI API, was used to assign sub-sector and type to the roughly 4000 projects in the 2023Q1 TW pipeline update, based on the project's owner, name, and description. This was benchmarked against a sample of 400 projects which had been manually classified by Te Waihanga personnel and found concordance of ~ 90%. Where invalid types were generated by the GPT classification, the sub-sector was used instead.

[11] At the time of writing, Te Waihanga are working to develop more coding for this dataset.

Key inputs and assumptions to stage 3

Strictly speaking, the model only produces estimates of required *labour hours*. In order to convert these values into *employee* requirements, we applied a conversion factor. We made the following assumptions:

- A full-time employee works 52 40-hour weeks for a total of 2080 hours per year
- Receives 160 hours of annual leave, 80 hours of sick leave, and ~ 50 hours of paid breaks, bringing the total down to ~ 1800 hours per year
- Spends roughly 10% of the remaining time on non-project matters (travel, business administration, etc.).

Since we based the demand model calculation on direct hours only, the outcome was a figure of 1600 labour hours per year per FTE. This figure is generally consistent with industry estimates of billable hours per year[12].

Although we prefer to report on FTEs as the primary metric for workforce size, we do include a 'workers' column in the database, for the sake of comparison to supply-side estimates. This is calculated by multiplying the FTE value by the average number of hours per week worked for that occupation group (ANZSCO 2-digit level) from HLFS data, divided by 40. This is consistent with the adjustment used in the supply model.

Limitations of model framework and methodology

Current and future coverage of Te Waihanga's project pipeline

Current coverage

The National Infrastructure Pipeline is intended to provide exhaustive coverage of current infrastructure works; however, it faces limitations in that it relies on infrastructure service providers to supply project data, and this process is not perfect. It is our understanding that gaps in coverage are known to exist. These depend on both the geographical location of projects and the organisations responsible for them. In general, data coverage is better for public sector infrastructure (which is primarily in the horizontal sector) than private (primarily vertical sector). This suggests that the demand model estimate of the vertical sector workforce is more likely to be an underestimate than for the horizontal sector.

Future coverage

Te Waihanga's pipeline of current and future construction and infrastructure projects is comparatively rich in data for the projects over the next few years, however, there is a lack of data available currently for construction and infrastructure projects occurring more than two years from now. The effect of this on the worker demand modelled through this work is that there is a drop-off in modelled workforce demand approximately two years from now (Figure 7). As Te Waihanga updates its pipeline with information collected from organisations across the industry on future projects, the modelled demand will also update to reflect the improved clarity on the future of the pipeline.

[12] See, for example, https://www.trinityp3.com/fee-models/how-many-billable-hours-per-year/

Demand estimates for vertical construction

Since Alta Consulting is primarily involved in horizontal infrastructure, confidence is lower in the values provided for projects in the vertical sector. We tried to address this via case-study calibration but note this as an area for potential improvement.

The model assumes linearity

The model assumes linearity; that is, a billion-dollar project requires the same number of workers as a hundred different 10-million-dollar projects. This is assumption is unlikely to be perfect, as economies of scale may mean that larger projects require proportionally fewer professional and technical staff and may have higher proportions of material costs.

However, a non-linear scaling model presents its own limitations. It is desirable for the sake of tractability that the model be scale-independent, since the level of aggregation at which projects are entered into the pipeline is highly variable. A series of roading improvements could be entered as one main project, or several smaller sub-projects. Having the model produce different estimates in these cases would lead to unwanted complications and make drawing comparisons between projects difficult. In the case of the asset register, only national- and sector-level figures can be recorded in many cases – so these totals are orders of magnitude larger than the individual assets they represent. A non-linear and therefore scale-dependent model would require estimating not only the total value, but the number and cost distribution of assets in every category. We therefore assumed linearity, calibrating the model to the mean-valued project for each sector.

Data inputs for estimating the size of the existing asset register

Although we found several high-level estimates of the total value of capital stock, differences in valuation and accounting practices make calibration difficult. These estimates range from ~ $180B (based on the National Accounts) to ~ $834B (based on an IMF report). Our total valuation of ~ $440B sits in the middle of this range. We note the following specific limitations:

- Some sectors (e.g., social housing via Kainga Ora) may include significant amounts of land in their valuation. This distorts the value of maintenance work required. The magnitude of this error is likely to vary significantly between sectors.
- Roading – many local roads are jointly maintained by regional/district councils and Waka Kotahi, which could lead to double-counting. In total, roading accounts for approximately one third of the value of the asset register, c.f. a quarter of the National Accounts' valuation.
- Some sectors have poor data availability, either due to lack of clarity regarding asset ownership, or organisations failing to publish clear property, plant and equipment (PPE) figures. We note this as a limitation but expect the impact of this to be relatively minor, as this shortfall generally occurs in smaller sectors.

Structure of outputs

The output from the demand modelling is a database of occupations required by quarter (Table 5).

Project sector	Stage	Region	Quarter	Occupation	Labour hours	FTEs	Workers

Roading	Construction	Auckland	2023Q4	Paving and Surfacing Labourer	3141.59	7.85	9.42

Table 5: Example of the structure of demand model output

Summary of outputs

Unlike the outputs from the supply model, which are estimates of the current state supply, the demand model outputs estimate demand over a period of time as informed by Te Waihanga's demand pipeline. Below, we have included analyses that describe the modelled demand by sector over time, and then by region and occupation for the current quarter workforce demand (2023Q3). We note that these outputs are still a work in progress and should be expected to change as refinements are made to the methodologies and inputs, as well as improvements in the coverage of the Te Waihanga pipeline.

Workforce demand by sector

Modelled workforce demand for the current quarter sits at 103,954 FTEs, with horizontal infrastructure representing the majority of these (59,022 FTEs in 2023 Q3 compared to 44,932 in the vertical sector – see Figure 7). This is a point-in-time estimate based on the 2023 Q1 update to the National Infrastructure Pipeline. Note that the drop-off in worker demand over time is a reflection of the lack of data available for projects starting more than two years or so from the present (see methodology limitations above).

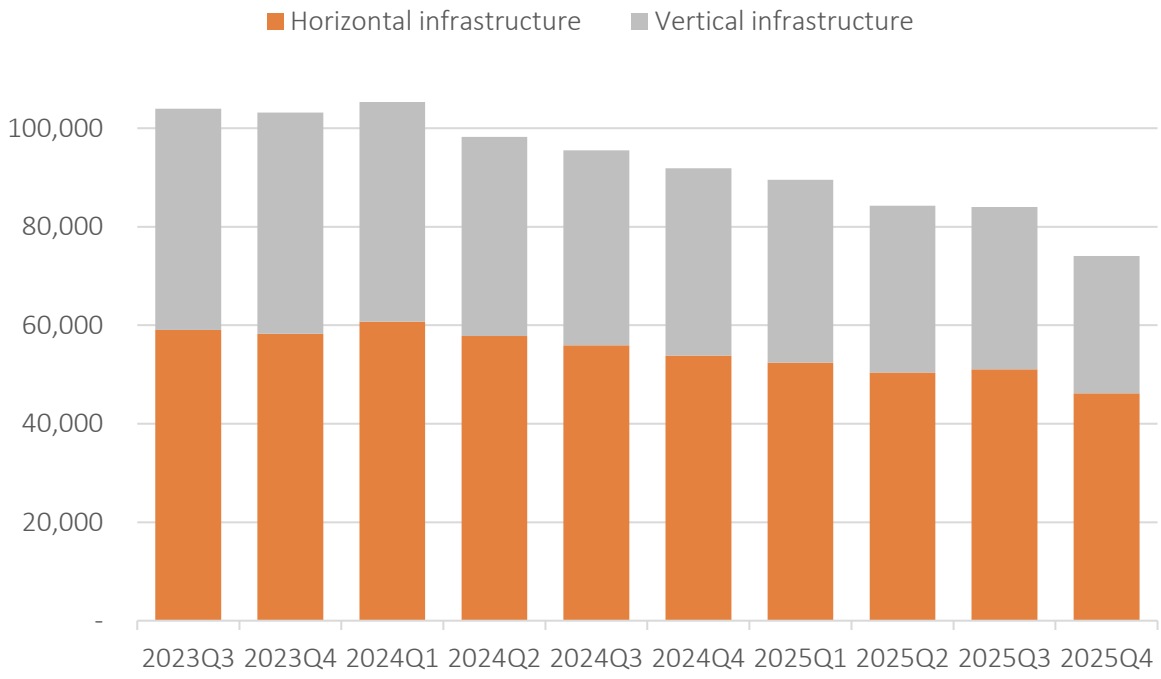


Figure 7: Infrastructure workforce demand (FTEs) by sector over time

Workforce demand by region

A breakdown of the 103,954 FTEs required in 2023 Q3 by region is shown in Figure 8 .

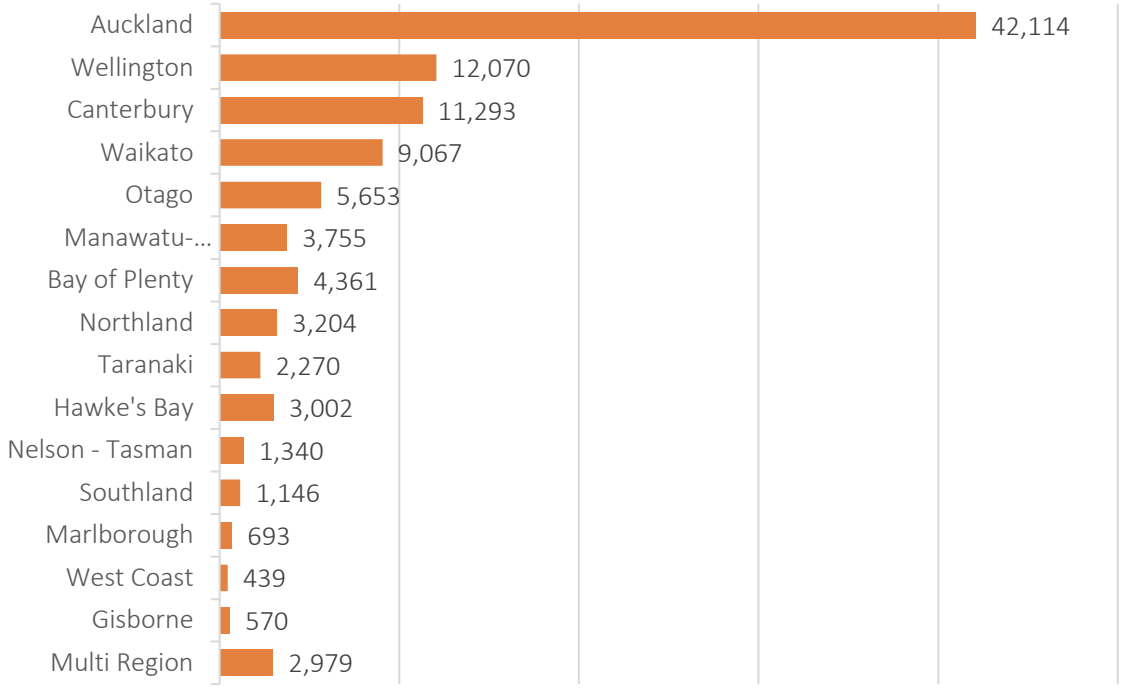


Figure 8: Current quarter (2023Q3) infrastructure workforce demand (FTEs) by region

Workforce demand by occupation

A breakdown of the top 10 occupations required in 2023 Q3 is illustrated in Figure 9.

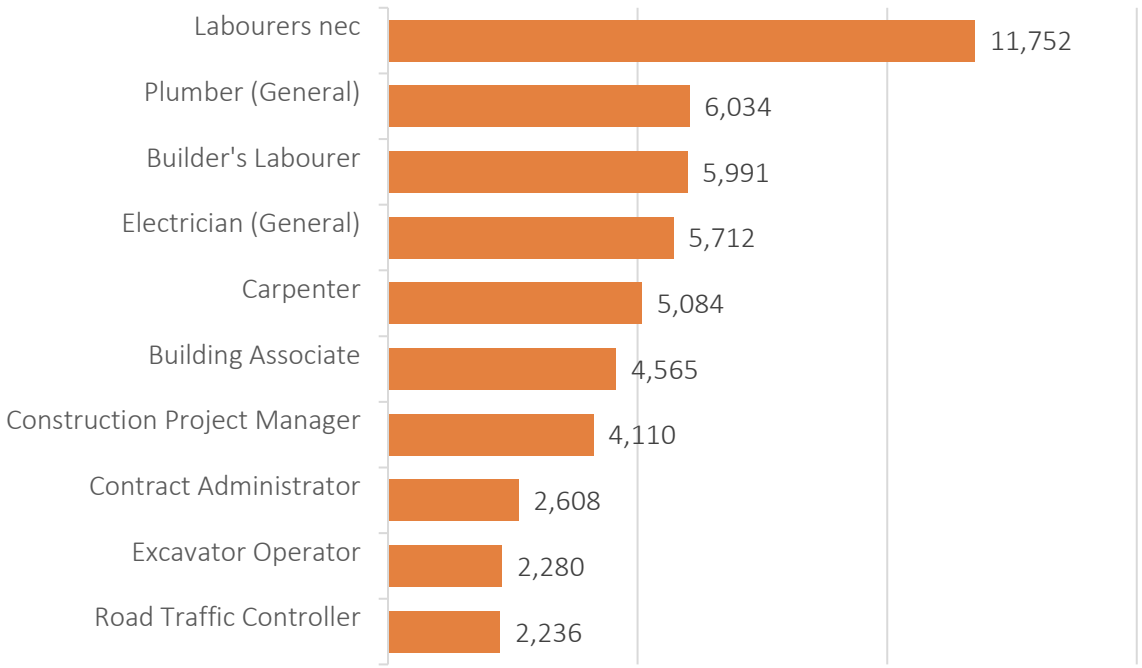


Figure 9: Current quarter (2023Q3) infrastructure workforce demand (FTEs) by occupation

Calibration of demand model estimates

Initial demand model input estimates

The initial set of inputs was adapted from an earlier project with Waihanga Ara Rau, which used a similar methodology to estimate construction workforce demand. The estimates in that model were developed by expert consultation, along with data from the QV CostBuilder tool (project-element breakdown and wage rates).

The occupations and project types defined for that model do not align with the workforce definition created for this project. Also, the scope of the model included only the construction phase and did not include any contribution from professional services workers. Therefore, the first step was to translate the work requirements from the original occupation list to align with the ANZSCO definition.

We then mapped the Te Waihanga project classifications (sector, sub-sector, and type) to the existing project types (see Table 6). We revisited the project element breakdown for those project types whose mapping was not perfect (e.g., communications -> energy transmission) and updated them to better reflect the new classification. We consider the project types used in the Waihanga Ara Rau model to now be superseded by the Te Waihanga project classification.

The ultimately desirable outcome would be for every level of the Te Waihanga project classification to be represented individually in the demand model. However, in some cases, it was felt that the distinction between project types within a sub-sector was either immaterial, or too far outside the expertise of Alta Consulting to be meaningful, and so we have defined these only at the sub-sector level.

Sector	Sub-sector	Type	Deprecated project type from Waihanga Ara Rau model
Vertical	Commercial sector	-	Building

Horizontal	Communications sector	-	Energy Transmission
Vertical	Community sector	-	Building
Vertical	Education and Research sector	-	Building
Horizontal	Energy sector	Electricity	Energy Transmission
Horizontal	Energy sector	Gas	Energy Generation
Horizontal	Energy sector	Liquid fuels	Energy generation
Vertical	Other	-	Building
Vertical	Social sector	-	Building
Horizontal	Transport sector	Road	Roading New
Horizontal	Transport sector	Port	Ports
Horizontal	Transport sector	Airport	Airports
Horizontal	Transport sector	Rail	Rail
Horizontal	Waste sector	-	Land Development and general civils
Horizontal	Water sector	Water supply	Water Transmission Pipelines
Horizontal	Water sector	Stormwater	Stormwater
Horizontal	Water sector	Wastewater	Water Treatment
Horizontal	Water sector	Irrigation	Water Transmission Pipelines
Horizontal	Water sector	Protection	Water Transmission Pipelines

Table 6: Mapping of Te Waihanga project classification to deprecated project types.

Calibrating demand model inputs with case studies

Appendix 8: Calibration of demand model contains examples of real-world project balance sheets (abridged and anonymised due to commercial sensitivity) used by Alta Consulting in the process of calibration, and brief commentary from Cameron Stewart of Alta Consulting on the calibration process, explaining the nature of the contracts used for reference. This followed an iterative process, wherein demand model outputs were compared to the case study projects, input matrices updated, and then outputs compared again.

Maintenance

As mentioned above, we assume that all major refits, renewals, and level of service improvements (LoSI) are captured in the Te Waihanga pipeline, and therefore do not model these under the 'Maintenance' stage. Unfortunately, this makes it difficult to capture exactly how much of the workforce is involved in maintenance work, since these projects are modelled as capital works in their own right: the output labour hours are tagged with the *planning and design* and *construction* stages, rather than the *maintenance* stage. The fact that approximately 8% of pipeline project value is associated with maintenance work suggests that it is reasonable to assume that about 8% of the labour hours output from the pipeline-driven demand model are involved in the maintenance of existing assets, rather than new builds.

To calibrate the maintenance work driven by the existing asset register, we undertook the following processes:

- Estimate the ratio of annual maintenance expenditure to asset value.
    - For subsectors where data on actual maintenance spend was available (e.g., transport, electricity, and gas), this was done by direct comparison of reported values.

- For subsectors with no accessible maintenance spend data, we used IRD and National Accounts depreciation rates as a proxy for this ratio.

- Adjust this ratio to account for perceived underinvestment in renewal.
- Convert this to a labour cost ratio using the relevant assumption for that subsector or project type.
- Adjust this value to account for the labour cost of maintenance and renewal projects already represented in the pipeline.

Table 19: Maintenance labour cost assumptions by project type/subsector contains a full summary of these datapoints, and the resulting maintenance spend assumptions. These were reviewed and approved by Alta consulting, noting that uncertainty is still high due to lack of concrete data in many sectors. In a few cases, the reported asset value was small enough that even a single large renewal project in the pipeline would swamp the maintenance spend ratio. In these cases, the pipeline renewal projects were discounted.

In the output of the demand model, the 'maintenance' stage of work accounts for 38% of labour hours in the current quarter (Figure 10: Proportion of labour hours by stage of work over time). This value increases to 57% by 2025Q4, however this is primarily due to the drop-off of capital works in the pipeline, as opposed to an increase in maintenance work. As a point of reference, Waka Kotahi reports 31% of forecast funding on maintenance and operations (69% on capital works), and Watercare 45% on renewals (55% on growth and level of service improvements). We therefore feel that our estimate of maintenance work as a proportion of total workforce requirement is of the right order of magnitude. A more complete understanding of the national infrastructure pipeline's coverage of large renewal projects will help to improve confidence in this estimate.

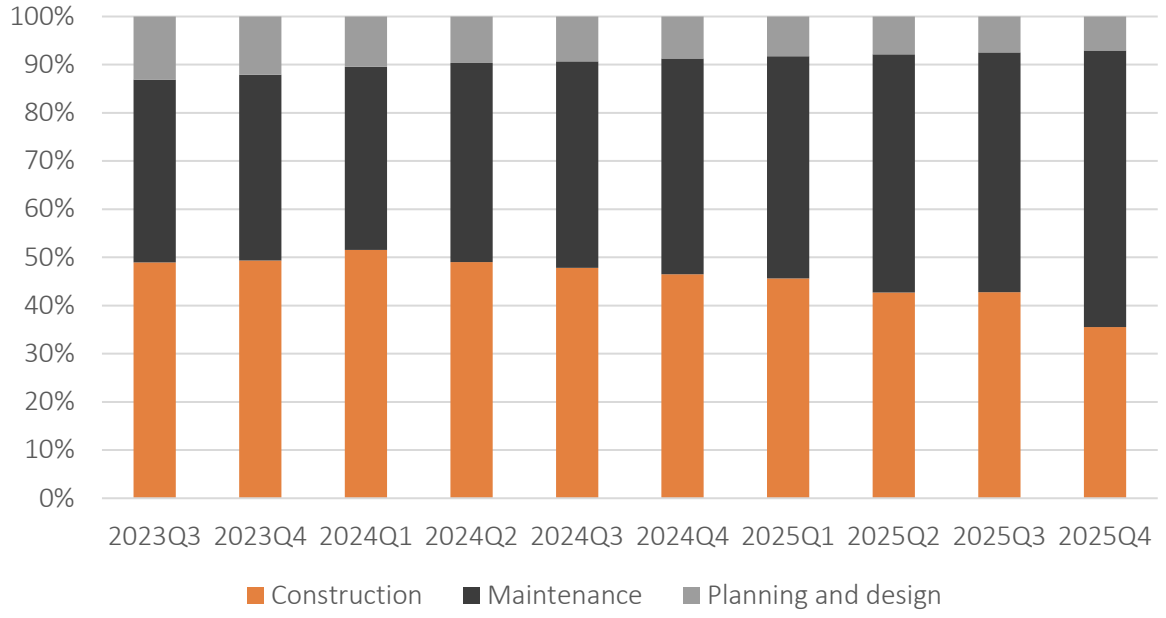


Figure 10: Proportion of labour hours by stage of work over time

Reconciliation of supply and demand models

Overall workforce size

This project developed separate supply and demand models, with the intention of undertaking a holistic calibration exercise to reconcile the two models and present a 'single source of truth' estimate for the current size of the infrastructure workforce. The supply and demand models semi-independently arrive at broadly similar numbers; 107,551 FTEs in the supply model and 103,954 FTEs in the demand model. The proportions of the workforce involved in vertical and horizontal infrastructure are similar, with the demand model (59,022 FTEs in horizontal, 44,932 vertical) slightly underpredicting vertical infrastructure compared to the supply model (54,448 horizontal, 53,063 vertical).

We note that both models are more likely to be under-estimates than over-estimates, since:

- The supply model relies on 2018 census data, and therefore does not take into account the growth of the infrastructure sector over the last 5 years. This effect is expected to be on the order of 5-10%, based on other analyses of workforce size over time[13].
- The demand model is driven by the Te Waihanga pipeline of work, which although thorough, cannot be expected to have 100% coverage. This is expected to be more significant in the vertical sector, where it may less clear if certain projects (e.g., shopping malls, apartment buildings, sport stadiums) belong to the pipeline or not.

Based on the level of reconciliation between the models, the estimated magnitude of underprediction described above, and the imprecision inherent in the definitions used, we judge the uncertainty[14] in our workforce estimate to be of the order of 15%. We note also that this estimate is heavily dependent on the exact definition of the workforce and should not be expected to reconcile to external figures, which are generally based on the size of a defined set of ANZSIC industries only. Since the supply model is anchored to a much more robust data source (i.e., the census) than the demand model, we treat this as the central estimate of workforce size.

We therefore present an indicative infrastructure workforce size of 107,000 FTEs, with a credible interval of 92,000 – 122,000.

This figure is a point-in-time estimate of the workforce as of the 2018 census. An estimate for the present-day workforce can be calculated by applying an exogenous growth rate to this figure, noting that uncertainty in the growth rate should also be considered. For example, based on a 2% p.a. growth rate, this estimate would become 118,000 (103,000 – 133,000) in 2023.

Workforce size by occupation

At the level of individual occupations, the reconciliation between the two models is typically not as good as it is for the overall workforce, although this varies by occupation (Figure 11 and Figure 12).

For the sake of discussion, we can divide the 112 modelled occupations into four categories:

- Those with good agreement between the two models (37 occupations)

[13] https://wip.org.nz/total-workforce-size?year=2021&workforce_definition=Core&industry_group=Infrastructure

[14] This is a holistic assessment only, in the absence of suitable analytical techniques for calculating uncertainty given the scale and complexity of these models. The credible interval can be expected to be refined by subsequent improvements to the model.

- Those with poor agreement between the two models (34 occupations)
- Those that are too small to be sensibly calibrated (30 occupations)
- Those that do not appear in both models (11 occupations).

See Appendix 9: Supply and Demand model comparison by occupation for a full table of these occupations by group, with relative and absolute differences.

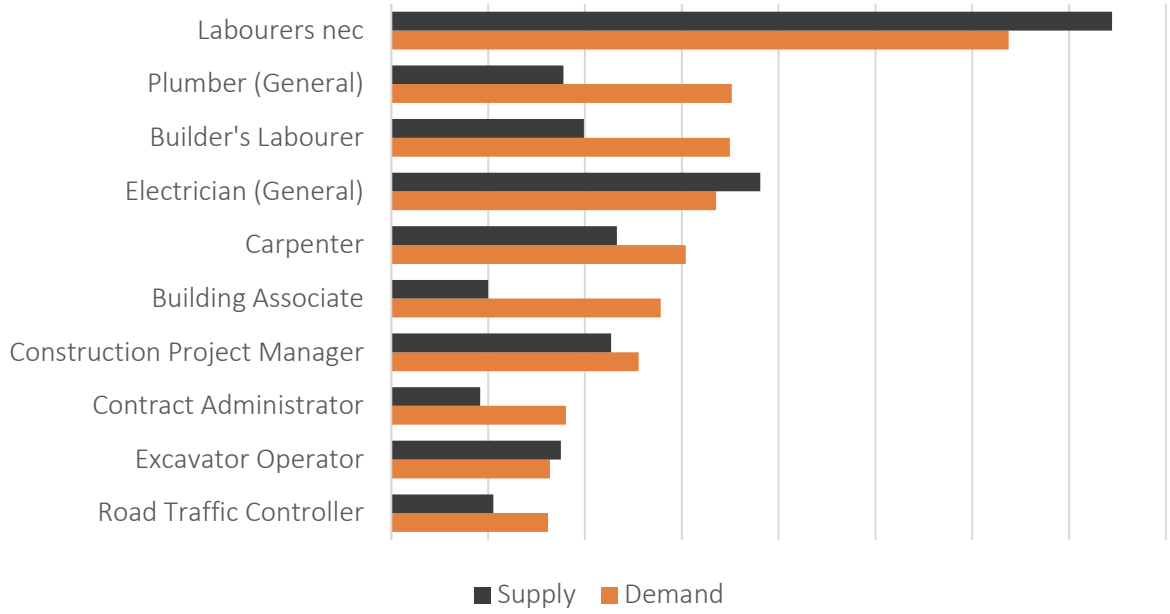


Figure 11: Supply vs demand FTEs by sector, top ten occupations

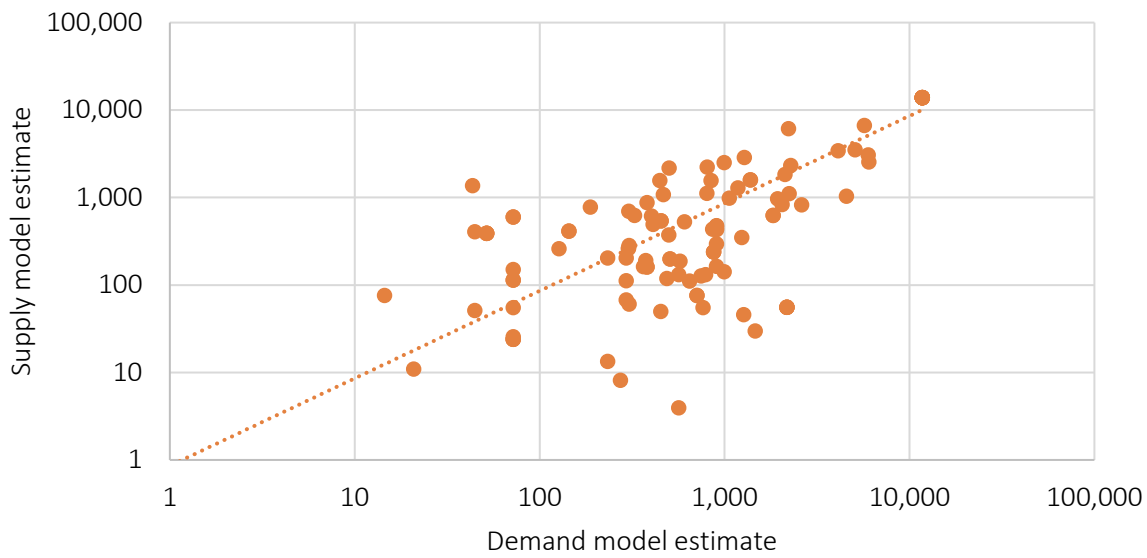


Figure 12: Supply vs demand model estimates for all occupations (excluding those not present in both models)

Occupations with good agreement between models

These occupations have at least 300 FTEs in at least one of the two models, and a maximum disparity of either 50% (calculated relative to the smaller estimate) or 500 FTEs. Many of the largest occupations fall into this category:

- Labourers n.e.c[15] (relative difference 18%)
- Construction project manager (relative difference 19%)
- Carpenter (relative difference 44%)
- Excavator operator (relative difference 2%)
- Electrician (relative difference 17%).

These occupations are more immediately familiar to stakeholders, and therefore tend to have received more attention throughout the iterative feedback process.

Occupations with poor agreement between models

These occupations have at least 300 FTEs in at least one of the two models, and a disparity greater than 50% or 500 FTEs.

Notable occupations in this group include:

- Plumber (relative difference 136%)
- Contract administrator (relative difference 217%)
- Civil engineer (relative difference 176%)
- Building associate (relative difference 341%).

Early in this project we presented preliminary estimates from both models and identified some large discrepancies. This led to the creation of the occupation sector map and subsequent refinements to the supply model. We considered two case-study occupations in this process.

The initial estimates for road traffic controllers were 332 on the supply side, and 10,791 on the demand side – a discrepancy of 3250%. We immediately identified that the use of industry restrictions on the supply side IDI queries was a source of error here, as the figure of 332 road traffic controllers excluded more than 80% of the total size of the occupation, as the majority of these are technically employed in industries associated with transport rather than infrastructure (i.e., ANZSIC division I: transport, postal, and warehousing). However, even discarding this restriction, the total of 1701 road traffic controllers would still be only 20% of the demand side estimate.

While traffic management is often the single largest line-item in a roading project's labour budget, it is incorrect to assume that all traffic management labour is carried out by individuals who would self-identify as a "road traffic controller". In reality, a roading project traffic management team generally consists of one trained traffic controller, and a team of labourers. We updated the demand model inputs to reflect this, and reduced the discrepancy between models, both for road traffic controllers and for labourers. Currently, the discrepancy for road traffic controllers sits at 83%, which is not

[15] Not elsewhere classified

sufficient to exclude the occupation from the 'poor agreement' category but represents a significant improvement over the initial discrepancy of 3250%. Similar definitional challenges exist across many occupations.

Occupations that are too small to be sensibly calibrated

These occupations have fewer than 300 FTEs in both the demand and supply models. They tend to be specialisations or offshoots of more significant occupations (e.g., plumbing inspector, civil engineering draftsperson, grader operator), or occupations whose primary function is outside the infrastructure sector (e.g., hydrographer, surveying or spatial science technician, land economist). Little effort has been dedicated to calibrating these, as the impact of such calibration is unlikely to be impactful on the overall workforce estimates, which are the primary focus of this work.

They are included in the model as they can have critical impacts on the progress of infrastructure projects. For example, a survey by a hydrographer is necessary before beginning construction of a new port. However, we feel that focussing on the calibration of these occupations would not be an efficient use of project resources. Due to the small size of these occupations, it would be more appropriate to conduct an in-depth review for specific infrastructure projects as necessary.

Occupations that do not appear in both models

There are 26 occupations that appear only in one of the two models, but 15 of these have fewer than 300 total FTEs and are therefore fall into the 'too small' category. Of the remaining 11, four appear only in the demand model and seven appear only in the supply model.

On the supply side, occupations may be missing either due to differences between the ANZSCO specification used in IDI records and the definitions we have used in the development of the model, or because no-one uses that code when recording their occupation. These occupations are:

- Quality assurance manager
- Surveyor
- Airconditioning and mechanical services plumber
- Crane chaser.

On the demand side, there are a small number of occupations that have not been assigned any labour components, as their contribution to the elements could not easily be determined:

- Metal fabricator
- Telecommunications technician
- Sheetmetal trades worker
- Building and engineering technicians nec
- Forklift driver
- Fencer
- Electrical linesworker.

We note as an area for improvement that the workforce definition could be updated to either exclude these occupations (as they are primarily involved in manufacturing and logistics, rather than infrastructure delivery) or use the sub-occupation level to make more explicit their function in order to help assign them value in the demand model. It may also be worth investigating why the occupations missing from the supply side do not appear in the IDI queries.

Concluding thoughts and recommendations for use

There is a balance to be struck between reconciling the two models and overfitting to the median. While it would be possible to continually update the demand model inputs to match the supply model outputs, we feel that this approach would not necessarily improve the accuracy of the models and would lose some of the value offered by the alternative approaches. Where there is a large discrepancy between occupation figures, we suggest that the supply-side figure be treated as the source of truth, since the underlying census data is a more robust source than the expert estimation used in the demand model.

We now present some guidance as to how to use the two models for a variety of use-cases.

Providing descriptive statistics on the composition of the infrastructure workforce.

The supply model should be used as the source of truth for this task. Keep in mind the limitations on accuracy when the data is filtered at 3 or more dimensions and when the mean cell count drops to below ~ 50. Note also that the sector definitions are specific to the model and may not necessarily match those found in external sources. At this stage we do not recommend using the supply and demand models to identify regional workforce gaps due to the incomplete coverage of the pipeline.

Forecasting short- and long-term labour requirements

While the demand model produces a forecast as far out as projects are entered in the pipeline, the quality of this estimate degrades rapidly as the time horizon exceeds ~ 2 years due to lack of coverage. We therefore do not recommend that this tool be used to forecast long-term labour requirements, and that an econometric approach be used instead. For short-term forecasting, we believe the model estimates are reasonably robust for the next 4-8 quarters, but once again care should be taken in interpreting these results due to the incomplete coverage. As quarterly updates are made to the pipeline, a potentially useful analysis could involve indexing new demand model outputs against old ones, to understand the magnitude of changes. Since the demand model is self-consistent, this could provide better insight into the pressures on the workforce introduced by planned infrastructure work than considering only labour hour requirements in absolute terms. One drawback of this is that it may be difficult to separate increases in demand from improving pipeline coverage from genuine increases in planned work.

Undertaking 'what-if' analysis

The demand model can be used to examine the impact of hypothetical large capital works projects or policy directions, by manually appending these hypothetical projects to the pipeline. Since the model assumes linearity of project cost, it is sufficient to provide only a single project encompassing all potential work in a given timeframe (e.g., $1B of spending on stormwater upgrades does not need to be entered as 50 different $20M projects). The only caveat to this is in terms of timing – if entering a multi-year program as a single project, the model will assign the first few years solely to planning and design, which may be undesirable. Contextualising the results of these what-if analyses can be done by determining the relative increase in demand for each occupation and comparing this to the capacity of

the workforce as determined by the supply model. It may be valuable to compare the modelled increase in demand to the additional capacity for that occupation across the other sectors (i.e., *residential* and *other*).

Identifying or forecasting specific skills shortages

As noted in the previous section, many occupations have poor agreement between models or are too small (within the infrastructure workforce) to be meaningfully calibrated. We therefore do not recommend direct comparison of occupation counts between models as a means of identifying critical skills shortages. By the nature of the models and the imprecision of occupational definitions (even under a standardised framework like ANZSCO), it is unlikely that the models can ever be made accurate and consistent enough to serve this purpose. Furthermore, the very concept of a 'shortage' is highly dependent on one's interpretation of the pipeline's relationship to actual building work. Not all projects are delivered on time or even at all, and the reasons for this can include workforce capacity issues, funding problems, zoning restrictions and regulations, or lack of public support. Interpreting 'shortage' as the difference between 'workforce required to deliver all possible planned infrastructure projects' and 'current state of workforce' may be misleading.

However, where critical skills shortages are identified in other analyses or by consultation with industry groups, the models may be used to lend insight into how to address these, or to undertake situational analyses to examine how these shortages are expected to change over time. On the supply side, understanding the breakdown of age and tenure within an occupation can help guide industry response in terms of training new employees and retaining existing ones. Ethnicity and visa status can be used as a lens to examine the impact of immigration policy on these shortages. On the demand side, 'what-if' analyses can be used to explore how these shortages may be affected by variations in the composition of the infrastructure pipeline, or conversely, how these shortages may limit the ability of the workforce to deliver certain types of projects.

Recommendations for refining methodology and inputs

The methodologies presented in this report have been used to establish a 'point in time' estimate of the supply of and demand for the infrastructure workforce. Several of the inputs used in the methodology still have some uncertainty and could benefit from refinement to improve the analysis outputs over time. While these refinements fall out of scope for this initial iteration of the infrastructure supply and demand estimate, we have provided recommendations for improving the models' outputs below by considering:

1. How sensitive the outputs from the supply and demand models are to the specified inputs (e.g. sensitivity of outputs to inputs),
2. To what level would improving the specified inputs be expected to improve the supply and demand models' outputs (e.g. level of output improvement)

Input	Area of improvement	Sensitivity of outputs to inputs	Level of output improvement
Supply side inputs			
IDI data	Limitation of input The current IRD and census data extracted from the IDI to estimate the supply of the workforce is from 2018. Recommendation for improving input The outputs from the supply model could be improved by extracting new releases of IRD and census data (e.g. from the 2023 census) as they are added to the IDI.	High	Low
Occupation sector mapping	Limitation of input The total counts of relevant occupations have been manually mapped between the sectors they are primarily employed with support from the reference group and industry experts. Responses were collated and averaged as inputs to the supply model. The reference group members engaged during this process provided their best guess estimate of the sector splits for all 100+ relevant occupations. Where necessary, industry experts with targeted expertise were engaged to fill any knowledge gaps. Despite efforts, not all gaps were addressed and there are occupations which could be further reviewed. However, we feel that averaging responses mitigates the risk of modelling error. Recommendation for improving input Further review of these inputs with a broader stakeholder group with relevant industry experience is not expected to have a significant improvement on the supply model's outputs.	High	Moderate

	We would recommend starting with the occupations where the variance of sector specificity is considered high.		
Demand side inputs			
Pipeline data	**Limitation of input** Te Waihanga's infrastructure project pipeline is in a state of continuous improvement where new projects are being added and the details of existing projects are being refined and updated. There are currently a few limitations of the pipeline data used for the demand model: 1. The pipeline has an absence of project data from 2+ years from now causing workforce demand to decrease unrealistically over time. 2. The pipeline is still in a developmental phase and does not yet have 100% coverage of current and planned infrastructure projects. **Recommendation for improving input** As Te Waihanga updates and adds details of current and future projects in its pipeline the updated pipeline should be referenced by the demand model to better reflect realistic demand.	High	High
Hourly charge-out rates	**Limitation of input** Hourly charge out rates for each occupation are used to translate project labour costs into FTE requirements. There is a high chance that the competitive labour market causes the current rates used in the modelling to become out of date. Using charge-out rates lower than the market rate will result in an overestimation of the infrastructure FTE requirements (demand). **Recommendation for improving input** As hourly charge-out rates are understood to be changing, we would recommend the rates use in the demand modelling for the infrastructure workforce be reviewed and updated. QV's CostBuilder has been used as a source for this information previously.	High	High
Demand side estimates	**Limitation of input** The approach used to model the infrastructure workforce demand starts with a top-down where a project's total cost is broken down by: 1. The proportion of the total project cost directly attributed to labour,	High	Low

	2. The proportion of a project's labour costs required to complete its key elements 3. The phasing of when a project's key elements are expected to be completed over the project's life 4. The proportion of the project's labour by occupation(s) required to complete the project's key elements 5. The hourly charge out rates (as above) required to translate the occupation cost by element, into FTE by element. The assumptions to breakdown the total project's costs through these five steps into an FTE requirement have been provided by Alta Consulting. As the dynamics of construction and infrastructure change, so would the elements and their costs. Recommendation for improvement As the industry grows, develops and innovates, it is likely the inputs and assumptions required to break down a project's cost to FTE requirements (as above) will need to be reviewed with appropriate industry stakeholders for their future ongoing relevance.		
Existing asset register database	Limitation of input The current valuation of the existing asset database is $440 billion. While this number has been triangulated with available sources, differences in valuation and accounting practices across organisations reporting asset values vary and mean calibration to date is only approximate. Recommendation for improvement We believe our current estimate is of the right order of magnitude and that any changes to the total asset value would only marginally improve the quality of the demand model's outputs. That being said, this is a new analysis and could benefit from further calibration and review of individualised asset values where possible. Where changes are made, the demand model should reference the latest version.	Moderate	Low
Maintenance model	Limitation of input The maintenance model assumes that all major refits and renewals are entered in the project pipeline and therefore errs on the side of underestimation to avoid double-counting of work. The model is simplistic and may not fully	High	Moderate

	account for the work requirements of some maintenance projects. Recommendation for improvement This assumption should be reviewed with reference to the actual coverage of the pipeline regarding maintenance work. More detailed calibration of maintenance spend with respect to depreciation rates of assets by sector may also be beneficial. It may also be beneficial to update this model such that it is entirely driven by forecast actual maintenance spending, as opposed to top-down estimates of existing asset value. In principle this is a relatively simple change, but the calibration of maintenance spend may be a significant task.		

Table 7: Recommendations for refining methodology and inputs

Appendices

Appendix 1: Occupation and workforce size

ANZSCO code	Occupation	Degree of sector specificity	Full occupation count	Breakdown of an occupation by sector:			
				Infrastructure - Vertical	Infrastructure - Horizontal	Residential construction	Other
899999	Labourers nec	Low	43,977	20%	18%	22%	40%
133112	Project Builder	Moderate	34,362	5%	5%	90%	0%
341111	Electrician (General)	Low	19,989	25%	15%	50%	10%
511112	Program or Project Administrator	Low	16,074	7%	5%	20%	68%
331212	Carpenter	Moderate	14,529	22%	8%	67%	3%
332211	Painting Trades Worker	Low	12,306	17%	1%	80%	3%
233512	Mechanical Engineer	Moderate	12,174	9%	7%	1%	82%
821111	Builder's Labourer	Low	11,379	27%	5%	63%	5%
133111	Construction Project Manager	Moderate	9,759	23%	20%	53%	3%
334111	Plumber (General)	Low	9,036	25%	8%	63%	3%
233211	Civil Engineer	Moderate	8,382	13%	73%	12%	2%
232111	Architect	Moderate	6,471	33%	8%	57%	2%
721311	Forklift Driver	Low	6,360	5%	8%	10%	77%
322211	Sheetmetal Worker	Moderate	4,314	12%	5%	2%	82%
322311	Metal Fabricator	Low	4,224	25%	12%	5%	58%
721214	Excavator Operator	Low	4,143	10%	60%	27%	3%
721211	Earthmoving Plant Operator (General)	Low	4,002	7%	70%	20%	3%
333311	Roof Tiler	Low	3,705	10%	0%	90%	0%
233214	Structural Engineer	Moderate	3,672	60%	13%	23%	3%
312999	Building and Engineering Technicians nec	Moderate	3,639	43%	13%	43%	0%
233311	Electrical Engineer	Moderate	3,570	25%	12%	20%	43%
233213	Quantity Surveyor	Moderate	3,345	25%	23%	50%	2%
331213	Joiner	Low	3,273	10%	0%	60%	30%
312112	Building Associate	Moderate	3,123	27%	13%	55%	5%
342414	Telecommunications Technician	High	3,099	8%	32%	15%	45%
322313	Welder (First Class) (Aus) / Welder (NZ)	Low	3,087	20%	10%	20%	50%
511111	Contract Administrator	Moderate	3,048	17%	17%	17%	50%
821712	Scaffolder	Low	3,036	27%	15%	50%	8%

232611	Urban and Regional Planner	High	3,006	15%	50%	35%	0%
332111	Floor Finisher	Low	2,982	25%	0%	75%	0%
334113	Drainer	Low	2,937	7%	50%	37%	7%
342211	Electrical Linesworker / Electrical Line Mechanic	High	2,514	5%	52%	5%	38%
312312	Electrical Engineering Technician	High	2,496	10%	45%	0%	45%
251312	Occupational Health and Safety Adviser	Low	2,400	17%	20%	22%	42%
333212	Renderer (Solid Plaster)	Low	2,400	20%	3%	75%	2%
821311	Fencer	Low	2,379	7%	20%	13%	60%
333211	Plasterer (Wall and Ceiling)	Low	2,229	25%	0%	75%	0%
133211	Engineering Manager	Moderate	2,124	17%	37%	2%	45%
312111	Architectural Draftsperson	High	2,067	30%	5%	63%	2%
333111	Glazier	Low	2,046	27%	0%	57%	17%
312512	Mechanical Engineering Technician	Moderate	1,833	15%	10%	2%	73%
331111	Bricklayer	Low	1,806	30%	0%	68%	2%
333411	Wall and Floor Tiler	Low	1,803	10%	0%	90%	0%
232511	Interior designer	Low	1,785	13%	0%	73%	13%
233999	Engineering Professionals nec	Moderate	1,749	15%	25%	5%	55%
899923	Road Traffic Controller	Low	1,701	7%	70%	7%	17%
821211	Concreter	Low	1,683	20%	23%	57%	0%
712111	Crane, Hoist or Lift Operator	Low	1,569	40%	25%	15%	20%
342111	Airconditioning and Refrigeration Mechanic	Low	1,479	20%	0%	20%	60%
133612	Procurement Manager	Moderate	1,425	7%	7%	20%	67%
721216	Loader Operator	Low	1,293	10%	40%	10%	40%
263312	Telecommunications Network Engineer	Moderate	1,287	3%	37%	3%	57%
721913	Paving Plant Operator	Low	1,272	10%	77%	13%	0%
312611	Safety Inspector	Moderate	1,248	17%	20%	23%	40%
234312	Environmental Consultant	Low	1,242	7%	23%	0%	70%
312114	Construction Estimator	Moderate	1,194	23%	23%	53%	0%
224512	Valuer	Moderate	1,158	5%	3%	40%	52%
312113	Building Inspector	High	1,017	20%	2%	60%	18%
232112	Landscape Architect	Low	966	17%	17%	57%	10%
233215	Transport Engineer	Moderate	963	2%	63%	3%	32%
312212	Civil Engineering Technician	Moderate	963	10%	67%	23%	0%
821711	Construction Rigger	Low	921	38%	35%	27%	0%
312911	Maintenance Planner	Moderate	882	20%	13%	17%	50%

263311	Telecommunications Engineer	High	795	7%	23%	10%	60%
331112	Stonemason	Low	753	10%	5%	65%	20%
233212	Geotechnical Engineer	Low	708	10%	53%	20%	17%
821411	Building Insulation Installer	Low	687	10%	0%	90%	0%
821713	Steel Fixer	Low	624	33%	33%	33%	0%
149411	Fleet Manager	Low	540	7%	17%	15%	62%
334114	Gasfitter	Low	456	27%	7%	65%	2%
312116	Surveying or Spatial Science Technician	Low	417	7%	27%	7%	60%
312199	Architectural, Building and Surveying Technicians nec	Low	417	32%	3%	61%	3%
331211	Carpenter and Joiner	Low	405	10%	0%	90%	0%
139912	Environmental Manager	Low	366	3%	30%	3%	63%
233915	Environmental Engineer	Low	363	3%	17%	3%	77%
821611	Railway Track Worker	High	297	0%	80%	0%	20%
712915	Concrete Pump Operator	Low	282	40%	33%	23%	3%
821511	Paving and Surfacing Labourer	Low	282	10%	57%	33%	0%
342212	Technical Cable Jointer	Low	273	2%	35%	0%	63%
721215	Grader Operator	Low	270	10%	70%	12%	8%
711913	Sand Blaster	Low	267	10%	10%	10%	70%
312211	Civil Engineering Draftsperson	Moderate	261	12%	62%	23%	3%
721999	Mobile Plant Operators nec	Low	261	23%	37%	20%	20%
312511	Mechanical Engineering Draftsperson	Moderate	258	13%	13%	2%	72%
721915	Road Roller Operator	Low	258	10%	80%	10%	0%
821113	Earthmoving Labourer	Low	234	10%	80%	10%	0%
341113	Lift Mechanic	Low	225	70%	2%	20%	8%
342411	Cabler (Data and Telecommunications)	High	210	12%	27%	15%	47%
821112	Drainage, Sewerage and Stormwater Labourer	Moderate	183	7%	63%	30%	0%
721916	Streetsweeper Operator	Low	165	0%	43%	5%	52%
233914	Engineering Technologist	Moderate	159	5%	30%	8%	57%
342412	Telecommunications Cable Jointer	High	153	0%	45%	0%	55%
821714	Structural Steel Erector	Low	135	57%	13%	27%	3%
313212	Telecommunications Field Engineer	Moderate	132	3%	37%	3%	57%
821912	Driller's Assistant	Low	126	8%	53%	8%	30%
821915	Surveyor's Assistant	Low	111	13%	43%	23%	20%
821114	Plumber's Assistant	Low	108	23%	10%	67%	0%

721213	Bulldozer Operator	Low	96	8%	73%	8%	10%
313214	Telecommunications Technical Officer or Technologist	High	93	2%	30%	2%	67%
312311	Electrical Engineering Draftsperson	Moderate	81	27%	32%	13%	28%
313213	Telecommunications Network Planner	High	63	0%	33%	0%	67%
821913	Lagger	Low	51	37%	27%	10%	27%
224511	Land Economist	Low	42	10%	22%	20%	48%
342413	Telecommunications Linesworker / Telecommunications Line Mechanic	High	39	3%	38%	3%	55%
312115	Plumbing Inspector	Low	36	20%	0%	57%	23%
334115	Roof Plumber	Low	33	20%	0%	70%	10%
311415	Hydrographer	High	30	0%	22%	0%	78%
721914	Railway Track Plant Operator	High	24	0%	72%	0%	28%
334112	Airconditioning and Mechanical Services Plumber	Low	21	20%	0%	68%	12%
232214	Other Spatial Scientist	Moderate	15	0%	7%	0%	93%
821911	Crane Chaser	Moderate	15	50%	17%	30%	3%

Table 8: Occupation sector mapping

Appendix 2: Project sectors

Sector - Level 1	Sub-sector - Level 2	Type - level 3
Water sector	Irrigation	Distribution network
		Storage
	Protection	Erosion control
		Flood protection
		Seawall
		Stopbank
		Wetland
	Stormwater	Collection network
		Overflow management facility
		Storage facility
		Treatment and outflow facility
	Wastewater	Collection network
		Disposal facility
		Storage facility
		Treatment plant
	Water supply	Distribution
		Firefighting reticulated supply
		Storage facility
		Treatment plant
		Water collection or abstraction facility
Waste sector	Carbon management	Carbon capture
		Carbon storage
	Collection	Storage and yard facility
		Transfer station
	Landfill	Cleanfill site
		Construction and demolition landfill site
		Managed landfill site
		Municipal solid waste landfill site
	Recycling	Recovery facility
		Sorting facility
	Waste processing	Composting facility
		Hazard waste treatment plant
Transport sector	Airport	Apron
		Auxiliary building
		Freight facility
		Hangar
		Passenger terminal
		Runway
	Port	Auxiliary building
		Cruiseship terminal
		Ferry terminal
		Freight facility
		Harbour

		Marina
		Shipping berth
		Wharf
	Rail	Auxiliary building
		Freight management
		Passenger terminal
		Rail yard
		Tracks - heavy rail
		Tracks - light rail
	Road	Bus terminal
		Car parking
		Footpaths
		Local road
		Separated busway
		Separated cycleway
		State highway
Social sector	Emergency services	Administration building
		Emergency services facility
		Training facility
	Health	Administration building
		Aged care facility
		Auxiliary building
		Health centre facility
		Hospital facility
		Maternity facility
		Plant facility
	Housing	Private housing
		Residential accommodation facility
		Social housing
		Sub-division site
	Justice	Administration building
		Community probation facility
		Court building
		Prison facility
		Rehabilitaion or work facility
		Staff training facility
	Other	Conservation estate
		Other facility
Energy sector	Electricity	Distribution network
		Geothermal generation
		Hydro generation
		Other
		Solar generation
		Thermal fuel plant
		Transmission network

		Vehicle charging station
		Wind generation
	Gas	Distribution network
		Exploration facility
		Extraction facility
		Production facility
		Storage facility
		Transmission network
	Liquid fuels	Distribution pipeline
		Exploration facility
		Extraction facility
		Production
		Storage facility
Education and Research sector	School	Administration facility
		Indoor shared facility
		Outdoor facility
		Plant facility
		Preschool facility
		Teaching facility - regular
		Teaching facility - specialist
	Science and research	Administration facility
		Geotech equipment facility
		Land or agriculture facility
		Observatory or astronomy facility
		Research centre facility
		Satellite facility
		Warehouse
	Tertiary	Administration facility
		Indoor shared facility
		Outdoor facility
		Plant facility
		Teaching facility - regular
		Teaching facility - specialist
Community sector	Community	Clubroom or hall
		Library
		Marae
		Museum or art gallery
		Streetscape
		Toilet or changing room facility
	Recreation	Conservation facility
		Park or reserve
		Playground
		Retired land
		Sports ground
		Swimming facility
	Venue	Conference centre

		Indoor arena
		Stadium
Communications sector	Data centre	Hosting
	Switching	Local exchange or cabinet
	Transmission	Backbone
		Copper
		Fibre network
		Microwave
		Navigation aid
		Satellite
		Submarine cable
		Tower network
Commercial sector	Commercial	Industrial park
		Mixed use building
		Office building
		Shopping centre
	Logistics	Transport yard
		Warehouse facility
	Manufacturing	Building supply plant
		Consumer goods plant
		Food processing plant
		Heavy industry plant
		Minerals chemicals or metals plant
		Other plant
		Transport and equipment plant
	Space	Rocket facility
	Tourism	Accommodation
		Attraction

Table 9: Full Te Waihanga project sector classification

Appendix 3: IDI queries

Dimension	Data source	Data table	Data columns	Key definitions	Comments
Occupation codes	Census 2018	cen_clean.census_individual_2018	• cen_ind_occupation_code	ANZSCO (6 digit)	• Only included occupations relevant to the Infrastructure Commission. • Excluded: ANZSCO code '111111' Chief Executive or Managing Director'
Occupation names	IDI Metadata	[IDI_Metadata].[clean_read_CLASSIFICATIONS].[CEN_ANZSCO]	• cat_code, • descriptor_text		
Industry codes	IRD	ir_clean.ird_ems	• ir_ems_ent_anzsic06_code	ANZSIC (4 digit)	• Only included industries relevant to the Infrastructure Commission.
Industry names	IDI Metadata	[IDI_Metadata].[clean_read_CLASSIFICATIONS].[CEN_ANZSIC06]	• cat_code, • descriptor_text		
Region	Address notification data			16 regional councils	• Most recent region a person resided in. • Regions include: Auckland, Bay of Plenty, Canterbury, Gisborne, Hawke's Bay, Manawatu-Wanganui, Marlborough, Nelson, Northland, Otago, Southland, Taranaki, Tasman, Waikato, Wellington, West Coast, Area Outside Region
Tenure	IRD	ir_clean.ird_ems	• snz_uid • ir_ems_return_period_date		• Count of the distinct number of month an individual has worked in the same industry. • 0-1 years, 1-2 years, 2-5 years, 5-10 years, 10+ years
Highest qualification	Census 2018	cen_clean.census_individual_2018	• cen_ind_standard_hst_qual_code	NZQA levels	• No Qualification, • Level 1 Certificate, • Level 2 Certificate, • Level 3 Certificate,

					• Level 4 Certificate, • Level 5 Diploma, • Level 6 Diploma, • Bachelor Degree or Level 7 Qualification • Post-graduate and Honours Degrees, • Masters Degree, • Doctorate Degree • Overseas Secondary School Qualification
Age	Personal details	data.personal_detail ir_clean.ird_ems	• snz_birth_year_nbr • ir_ems_return_period_date	10-year bands	• Calculated age based on the difference between tax month/year and birth year. • 15-24, • 25-34, • 35-44, • 45,54, • 55-64, • 65+
Gender	Personal details	data.personal_detail	• snz_sex_gender_code		• Male, female
Ethnicity	Personal details	data.personal_detail	• snz_ethnicity_grp1_nbr (European) • *snz_ethnicity_grp2_nbr (Maori)* • *snz_ethnicity_grp3_nbr (Pacific)* • *snz_ethnicity_grp4_nbr (Asian)* • *snz_ethnicity_grp5_nbr (MELAA)* • *snz_ethnicity_grp6_nbr (Other)*		• Participants in Stats NZ surveys (e.g., census, HLFS) are able to self-report multiple ethnicities. These responses are available as standalone columns under the personal details table. Using these data, ethnicity is derived by allocating a single ethnicity to each unique individual using the following prioritisation sequence: Māori, Pacific, Asian, MELAA, European, Other. For example, if a survey participant reports as identifying as Māori and European, their allocated ethnicity will be Māori.

Size of employing firm	IRD	ir_clean.ird_ems	• snz_employer_ird_uid • snz_uid		• Count of the number of distinct individuals associated with each employer. • Firm size grouped as: • 1-5, 6-15, 16-50, 51-100, 101+
Visa status	Visa applications	dol_clean.decisions ir_clean.ird_ems	• dol_dec_application_stream_text, • *dol_dec_application_criteria_text,* • *dol_dec_decision_date, dol_dec_nationality_code, dol_dec_sex_snz_code, dol_dec_application_type_code, dol_dec_decision_type_code, ir_ems_return_period_date*		• dol_dec_decision_date is used to exclude transit visa holders (codes 20, 21). • Dol_dec_application_stream_text is used to exclude transfer visa holders. • dol_dec_decision_type_code exludes records where visa applications were denied. • Ir_ems_return_period_date ensures visa approval occurred prior to the starting month of working in an industry. Visa's grouped as: • Resident, • Working holiday, • Work, • Student • NZ citizen (where there was no visa application)
Hours worked	HLFS	hlfs_clean.data	• hlfs_urd_quarter_nbr • hlfs_urd_quarter_date • hlfs_urd_usual_hr_main_nbr • hlfs_urd_occ_main_code • hlfs_urd_ind_main_code • hlfs_urd_emp_status_main_code		• Self-reported usual hours worked per week for each quarter of 2018. • Grouped as: • 0-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45+

Table 10: Detailed description of codes and definition used in IDI queries.

As well as the total for each dimension on its own, we supply marginal totals across the following groups of occupations to the IPFP model:

- Occupation x region x gender x ethnicity
- Occupation x ethnicity x visa
- Industry x ethnicity x visa
- Occupation x employment type
- Industry x employment type
- Region x gender x ethnicity
- Ethnicity x visa
- Industry x Occupation x region
- Industry x Occupation x visa
- Industry x Occupation x tenure
- Tenure x age x gender
- Tenure x age x industry

For each supplied marginal, the supply model records the root-mean-square error (RMSE) and mean-absolute error (MAE) with respect to the IDI data.

Appendix 4: Detailed analysis of IPFP

The python IPFN package implements a convergence parameter, which measures the 'similarity' of the modified matrix to the target. Each iteration of the algorithm updates this convergence, and the algorithm terminates when the rate of change of convergence reaches a specified threshold. It also allows the user to set a maximum number of iterations. From testing, a maximum of 30 iterations seems to produce good results, but a thorough investigation of how goodness-of-fit improves with subsequent iterations could be an area for improvement. Frick and Axhausen[16] suggest that 20 iterations are sufficient for convergence for a similar problem, albeit with fewer dimensions.

We validate the IPFP by comparing output marginals across up to 4 dimensions. There is an important distinction to draw here, which is the difference between goodness-of-fit to *supplied* vs *unsupplied* marginal totals (this can be understood as analogous to 'in-sample' and 'out-of-sample' fit). The first of these is somewhat trivial, as it is already in some sense measured by the IPFP as the 'similarity' maximised by the algorithm. The charts below show the correlation of cell counts for 1, 2, and 3 dimensions between *explicitly supplied* marginals and IPFP outputs.

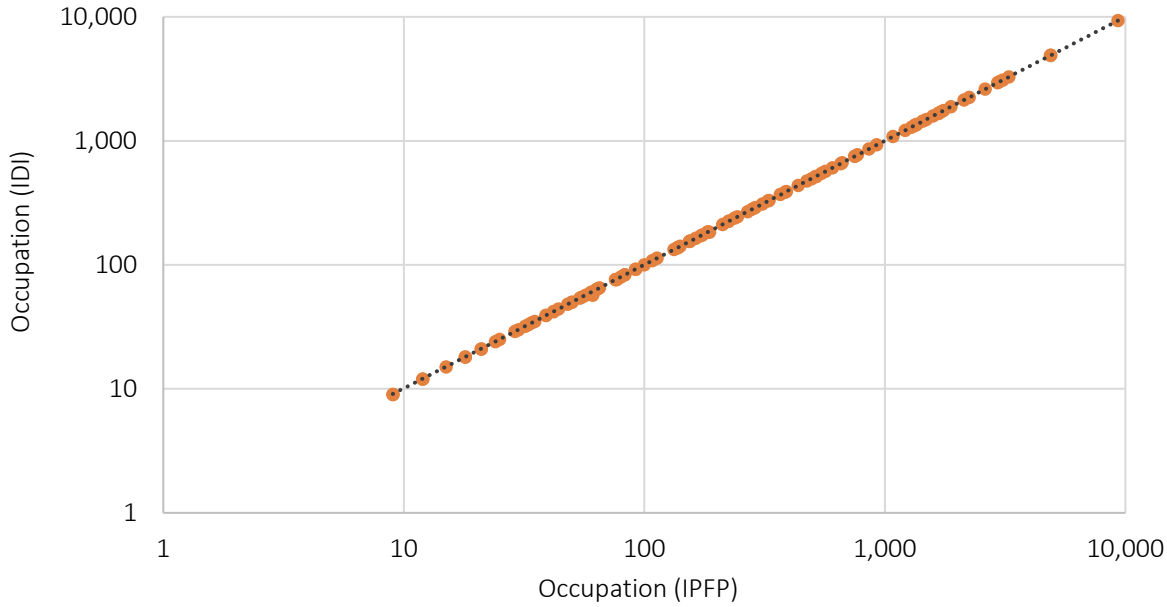


Figure 13: Model vs IDI correlation, 1 dimension

[16] https://ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/reports/101-200/ab150.pdf

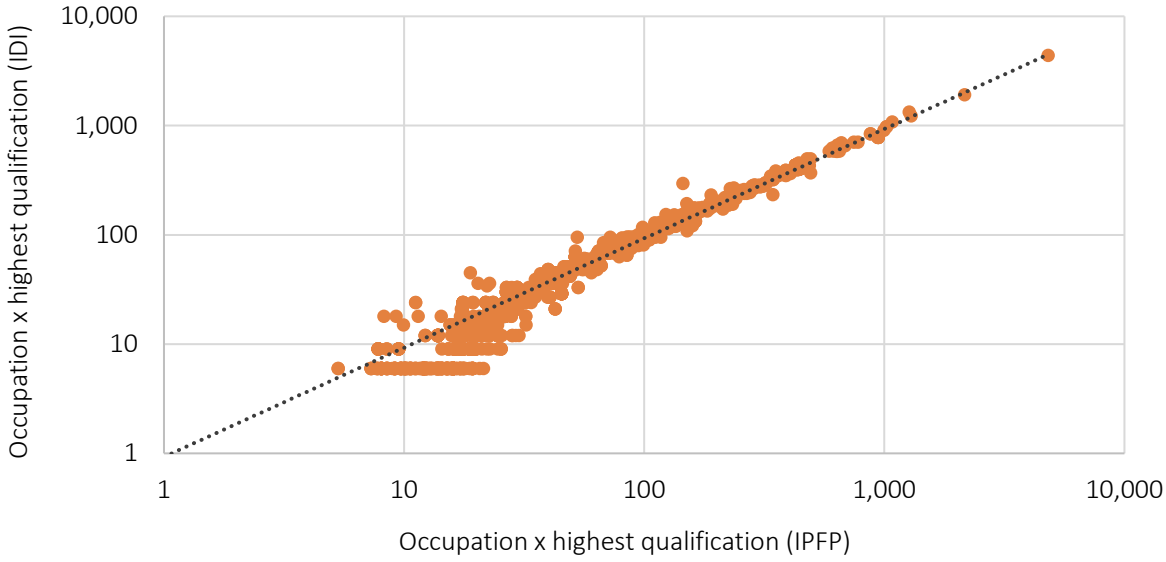


Figure 14: Model vs IDI correlation, 2 dimensions

Figure 15: Model vs IDI correlation, 3 dimensions

To assess the goodness-of-fit, we calculate the root-mean-square error and mean absolute error for each of these datasets.

Dimensions	Mean	RMSE	RMSE (%)	MAE	MAE (%)
1 (occupation)	695.9	1.6	0.2%	0.9	0.1%
2 (occupation x highest qualification)	62.4	20.0	32.1%	9.0	14.4%
3 (occupation x highest qualification x tenure)	16.5	8.8	53.3%	3.8	23.0%

Table 11: Analysis of error at each dimension

Note that in each case, the slope of the line is slightly lower than 1, that is, the model appears to be overpredicting counts in each cell. This is because the IDI data contains suppressed rows, and that the extent of this suppression varies between queries. For example, the sum of cells from the occupation x tenure x highest qualification query is only 89% of the sum from the industry x occupation query, despite both measuring the same exact underlying population.

To correct for this, we define the true population size to be the size of the population from the least-suppressed query. For each other marginal, we calculate the difference between its size and the true population size and the number of suppressed rows and replace the suppressed rows with the value $\frac{size\ difference}{number\ of\ suppressed\ rows}$ such that the total population size is the same for all marginals.

We do this on the understanding that a suppression occurs only if there is at least one person who fits the criteria (i.e., the row simply does not exist if there are no people in the relevant attribute bucket).

Next, we assess the ability of the IPFP model to preserve correlations between marginals which are not explicitly supplied. To do this, we withhold the industry x occupation x region marginals, and supply only industry x occupation and occupation x region. We then compare each of the 3 resulting 2-dimensional distributions to the 3-dimensional industry x occupation x region query and calculate the RMSE and MAE as above.

Dimensions	Mean	RMSE	RMSE (%)	MAE	MAE (%)
Industry x occupation (supplied)	56.0	23.7	42.3%	8.6	15.4%
Occupation x region (supplied)	73.1	25.6	35.0%	12.1	16.6%
Industry x region (withheld)	321.4	123.8	38.5%	54.6	17.0%
Industry x region x occupation (withheld)	7.2	6.5	90.3%	2.2	30.6%

Table 12: Analysis of error for supplied and withheld marginals (2 dimensions)

In relative terms, the RMSE and MAE for the withheld 2D marginal is very similar to those of the supplied 2D marginals (noting that the absolute values are larger due to the larger mean, resulting from the fact that there are ~ 100 occupations but only ~20-30 industries and regions).

The story is similar even in 3 dimensions, although the errors overall are significantly larger (noting the higher impact of suppression), they are of the same order for both supplied and withheld marginals. We therefore feel confident in stating that the IPFP is good enough at preserving unsupplied correlations as to produce useful outputs.

Dimensions	Mean	RMSE	RMSE (%)	MAE	MAE (%)
Occupation x age x gender (supplied)	74.7	33.7	45.1%	16.6	22.2%
Region x age x gender (supplied)	379.6	84.2	22.2%	61.1	16.1%
Occupation x region x age (withheld)	9.1	7.1	78.0%	3	33.0%
Occupation x region x gender (withheld)	26.2	14.5	55.3%	5.8	22.1%
Occupation x region x gender x age (withheld)	5.8	5.5	94.8%	2.1	36.2%

Table 13: Analysis of error for supplied and withheld marginals (3 dimensions)

Note also that in this case the means of the withheld marginals are smaller than the means of the supplied marginals, which we believe contributes to the difference in the size of relative error. As demonstrated in the chart below, it appears that relative RMSE is strongly correlated with mean (and therefore number of attribute buckets). For instance, the relative RMSE for the 3-dimensional region x

age x gender is lower than any of the relative RMSEs for the 2-dimensional totals above, all of which have smaller means. Therefore, it may also be worth considering mean bucket-size when evaluating whether to use a set of outputs.

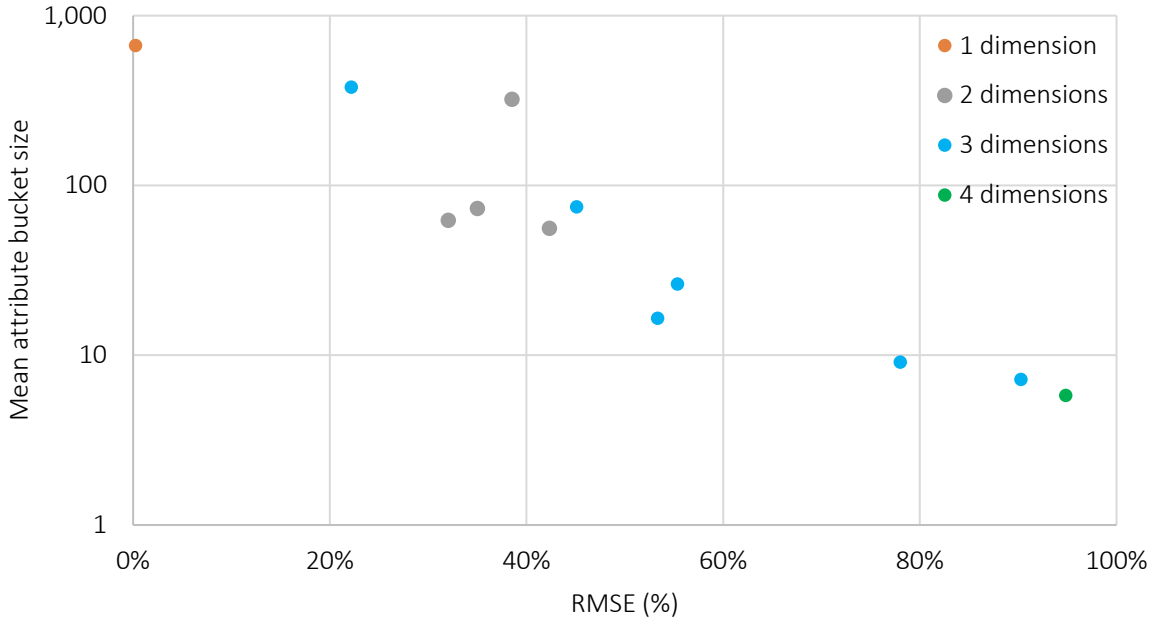


Figure 16: Mean attribute bucket size vs relative RMSE for all dimensions analysed.

Appendix 5: Demand model input tables

Project type	Percentage of total project cost associated with labour in stage	
	Planning and design	Construction
Water supply	12%	33%
Stormwater	12%	33%
Wastewater	12%	33%
Road	12%	27%
Port	10%	29%
Airport	14%	36%
Commercial	14%	36%
Waste	14%	31%
Electricity	12%	32%
Gas	12%	24%
Rail	14%	32%
Social	13%	36%
Community	13%	36%
Education and Research	13%	36%
Other	13%	36%
Communications	12%	32%
Liquid fuels	14%	32%
Irrigation	12%	32%
Protection	12%	32%

Table 14: Project labour cost breakdown

Stage	Element	Water supply	Road	...
Planning and design	Property and procurement	5%	10%	...
	Consenting	15%	20%	...
	Design	80%	70%	...
Construction	Undertake earthworks - bulk earthworks, cutting and filling, trimming, undercuts, landscaping, finishing	5%	25%	...
	Build structures - includes bridges, tunnels, retaining structures, piling and foundations	9%	20%	...
	Place pavement - pavement, seal, marking, barrier, street furniture	5%	27%	...
	Lay pipe - includes trench excavation and backfill	62%	8%	...
	Construct buildings - amenity and facility buildings associated with the infrastructure	0%	2%	...
	Install mechanical plant - pumps, internal pipes, valves and steelwork	2%	0%	...
	Install electrical equipment - includes cables, equipment and comms	2%	3%	...
	Traffic Management	15%	15%	...
	Building finishes	0%	0%	...
Maintenance	Annual minor maintenance	0.16%	0.16%	
	5-yearly maintenance	0.8%	0.8%	

Table 15: Project element breakdown (abridged – 17 columns omitted)

Note that the percentages sum to 100% within each project stage (e.g. planning and design, construction), but for maintenance they refer to a percentage of the initial project value.

Stage	Element	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8	Phase 9	Phase 10
Planning and design	Property and procurement	30%	6%	6%	20%	21%	3%	3%	3%	3%	5%
	Consenting	7%	40%	30%	5%	5%	5%	1%	1%	1%	5%
	Design	3%	5%	25%	25%	20%	10%	5%	3%	3%	3%
Construction	Undertake earthworks						60%	25%	15%		
	Build structures						20%	35%	35%	10%	
	Place pavement								20%	30%	50%
	Lay pipe -						25%	45%	30%		
	Construct buildings						10%	20%	30%	25%	15%
	Install mechanical plant								20%	50%	30%
	Install electrical equipment								20%	50%	30%
	Traffic Management						12%	22%	22%	22%	22%
	Building finishes							10%	20%	30%	40%

Table 16: Project phasing matrix

Stage	Element	Construction Project Manager	Project Builder	Engineering Manager	...
Planning and design	Property and procurement	18%	12%	23%	...
	Consenting	9%	4%	9%	...
	Design	5%	7%	5%	...
Construction	Undertake earthworks	5%	5%	1%	...
	Build structures	8%	8%	2%	...
	Place pavement	7%	7%	1%	...
	Lay pipe -	5%	5%	1%	...
	Traffic Management	1%	1%	0%	...
Construction & Maintenance	Construct buildings	3%	0%	1%	...
	Install mechanical plant	9%	1%	2%	...
	Install electrical equipment	9%	1%	2%	...
	Building finishes	2%	0%	0%	...

Table 17: Element occupation breakdown (abridged – 109 columns omitted)

Appendix 6: Example project calculation

The Sankey diagram below[17] visualises the process of the demand model for a $10M roading project, for a single occupation (construction project manager). For visual simplicity, this is a whole-of-project lifetime breakdown, that is, we do not show the application of the phasing matrix. We also do not show the maintenance phase. From this, we can see that a total of $235,935 of the project value goes towards paying the labour costs of construction project managers. Assuming a charge-out rate of $100/hr, this becomes 2,359 labour hours over the lifetime of the project. Table 18 presents a mock-up of the resulting output database rows.

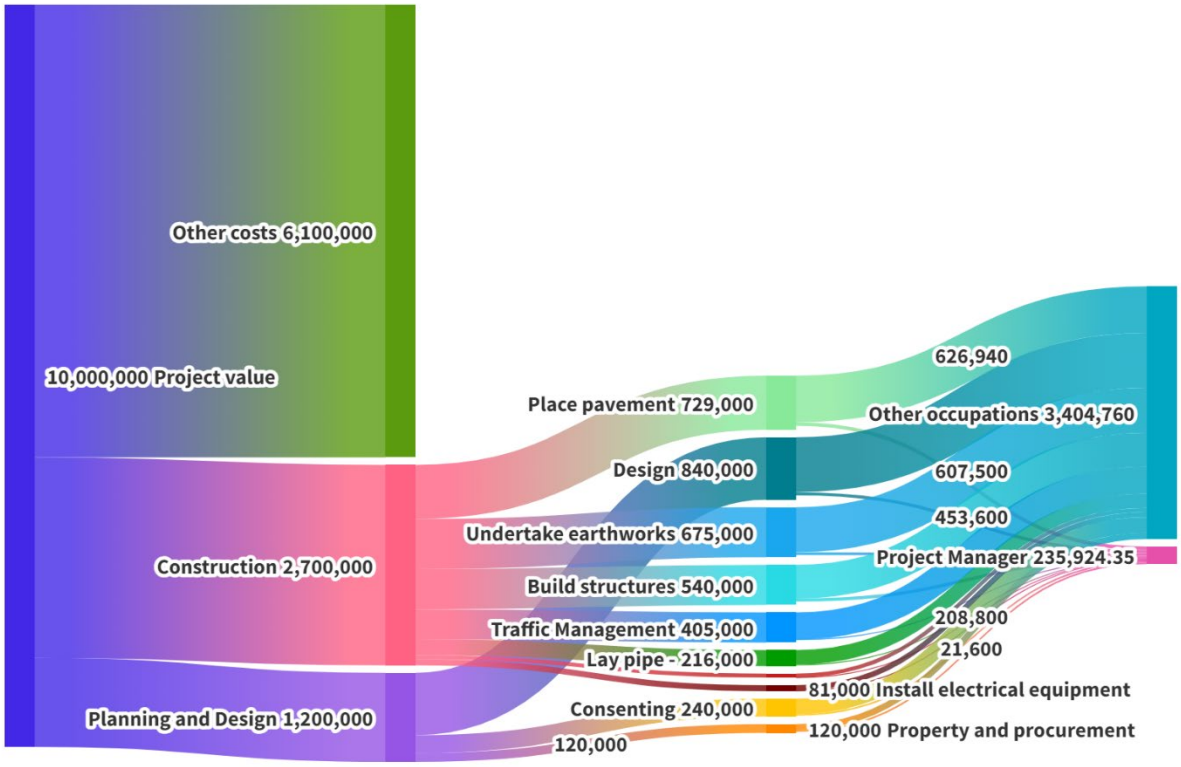


Figure 17: Sankey diagram visualisation of demand model

[17] Interactive version available at: https://public.flourish.studio/visualisation/14602947/

Occupation	Quarter	Stage	Project type	Sector	Labour hours	FTE
Construction Project Manager	2023Q1	Planning and design	Road	Horizontal	118.47	0.30
Construction Project Manager	2023Q2	Planning and design	Road	Horizontal	172.85	0.43
Construction Project Manager	2023Q3	Planning and design	Road	Horizontal	200.56	0.50
Construction Project Manager	2023Q4	Planning and design	Road	Horizontal	172.64	0.43
Construction Project Manager	2024Q1	Construction	Road	Horizontal	410.16	1.03
Construction Project Manager	2024Q1	Planning and design	Road	Horizontal	63.16	0.16
Construction Project Manager	2024Q2	Construction	Road	Horizontal	405.61	1.01
Construction Project Manager	2024Q2	Planning and design	Road	Horizontal	30.38	0.08
Construction Project Manager	2024Q3	Construction	Road	Horizontal	358.00	0.90
Construction Project Manager	2024Q3	Planning and design	Road	Horizontal	23.02	0.06
Construction Project Manager	2024Q4	Construction	Road	Horizontal	333.48	0.83
Construction Project Manager	2024Q4	Planning and design	Road	Horizontal	35.63	0.09

Table 18: Output of demand model for $10M roading project, construction project manager only.

Appendix 7: Breakdown of existing asset register by asset type

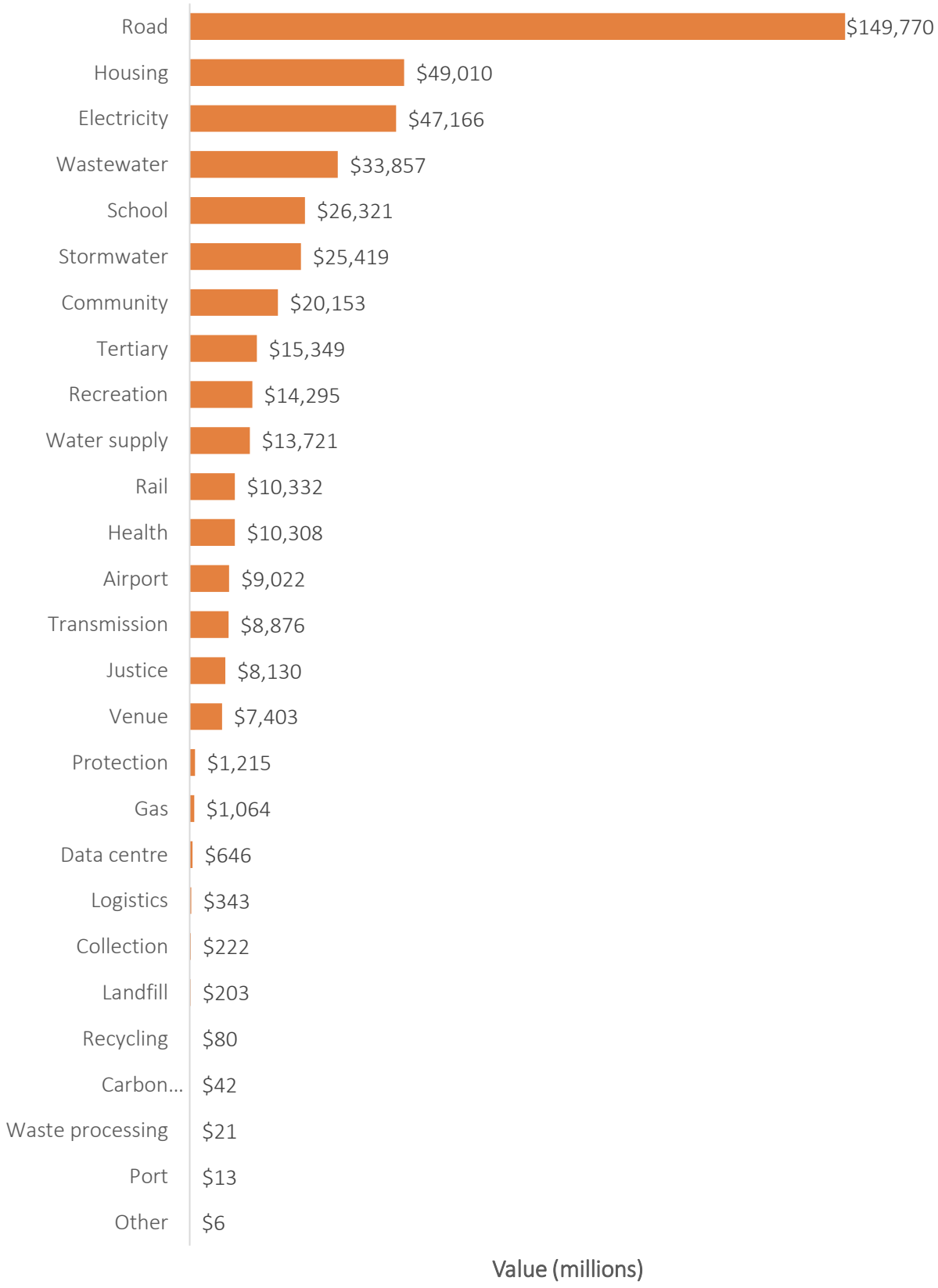


Figure 18: Breakdown of existing asset register by asset type ($M)

Asset type/subsector	Total asset value ($)	Actual ratio of maintenance / renewal spend to asset value	Estimated ratio of maintenance / renewal spend to asset value, based on depreciation trends	Ratio of labour cost to total project cost	Adjustment for maintenance / renewal projects in NIP	Average annual maintenance / renewal labour cost as % of asset value	Notes/sources
Water supply	13,721,287,059	NA	1.40%	45%	25,905,965	0.44%	National Accounts average CFK/NKS[18] ratio for local government water, sewerage, drainage, and waste services for 2013-2022 period, marked down by 40% to account for under-renewal trend observed in Council LTPS
Stormwater	25,418,960,599	NA	1.40%	45%	12,057,357	0.58%	National Accounts average CFK/NKS ratio for local government water, sewerage, drainage, and waste services for 2013-2022 period, marked down 40% to account for under-renewal trend observed in Council LTPS
Wastewater	33,857,306,326	NA	1.40%	45%	29,742,644	0.54%	National Accounts average CFK/NKS ratio for local government water, sewerage, drainage, and waste services for 2013-2022

[18] CFK = consumption of fixed capital; NKS = net capital stock

							period, marked down 40% to account for under-renewal trend observed in Council LTPS
Road	170,855,574,604	1.40%		39%	84,992,812	0.50%	Waka Kotahi actual local road and state highway maintenance spending
Port	13,000,000	NA	2.00%	39%		0.78%	No relevant data on either depreciation or maintenance/renewal expenditure. Midpoint of roads and electricity used.
Airport	9,021,900,000	NA	2.80%	50%	9,965,538	1.29%	No relevant data on either depreciation or maintenance/renewal expenditure. Midpoint of roads and education and research used.
Commercial	343,000,000	NA	2.50%	50%	80,669	1.23%	No relevant data, used mean of other sectors
Waste	218,977,161	NA	4.50%	45%	14,078	2.02%	National Accounts average CFK/NKS ratio for commercial/private water, sewerage, drainage, and waste services for 2013-2022 period, marked down by ~15% to account for under-renewal trend observed in Council LTPS
Electricity	47,166,000,000	2.60%		44%	16,229,499	1.12%	Actual maintenance spend from infrastructure network capital stock and investment baseline data

Gas	1,064,000,000	1.00%		36%	657,390	0.30%	Actual maintenance spend from infrastructure network capital stock and investment baseline data
Rail	10,332,000,000	NA	1.40%	46%	31,117,435	0.34%	National Accounts data includes vehicles and above-track infrastructure, resulting in a higher depreciation ratio. Value for roads used as a proxy.
Social	67,454,368,202	NA	3.40%	49%	156,201,846	1.43%	National Accounts average CFK/NKS ratio for central government hospitals for 2013-2022 period, marked down by ~15% to account for under-renewal
Community	20,153,012,989	NA	3.80%	49%	73,662,340	1.50%	Midpoint of social and education/research
Education and Research	15,350,000,000	NA	4.20%	49%	15,829,302	1.95%	National Accounts average CFK/NKS ratio for central government preschool, school education and tertiary education for 2013-2022 period, marked down by ~15% to account for under-renewal
Other	6,000,000	NA	2.50%	49%		1.23%	Unsure what this category includes – use whatever seems most relevant.
Communications	9,521,399,756	NA	7.00%	44%	388,202	3.11%	National Accounts data for telecommunications services suggests a depreciation ratio of 9.5%, reflecting greater role of software and shorter cycles of

							technological change. A lower figure has been used to better align with the IRD's range of asset life estimates for fixed telecommunications assets.
Liquid fuels	607,260,188	NA	2.60%	46%	0	1.20%	Use figure for electricity sector
Irrigation	607,260,188	NA	2.00%	44%	0	0.88%	Use figure for water supply
Protection	1,214,520,377	NA	2.00%	44%	3,803,540	0.57%	Use figure for water supply

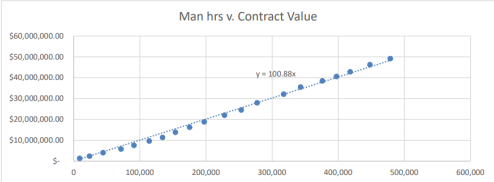
Table 19: Maintenance labour cost assumptions by project type/subsector

Appendix 8: Calibration of demand model

Example Roadworks Project - Manhour Summary

Prepared: C. Stewart 3/3/2023
Revised: 31/5/2023
Ammended: 28/7/23

DATA SET	Total Hours			Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10
Total Direct Labour	***292,007***			***5,432***	***9,145***	***12,480***	***16,566***	***11,935***	***14,103***	***12,211***	***11,813***	***13,178***	***13,536***
Subcontract adjustment assume cost t	186,693	0	0	3,473	5,847	7,979	10,591	7,631	9,017	7,807	7,552	8,425	8,654
TOTAL ESTIMATE MANHOURS	**478,701**	**100%**		**8,904**	**14,992**	**20,458**	**27,157**	**19,566**	**23,120**	**20,018**	**19,365**	**21,603**	**22,189**
				8,904	23,896	44,355	71,511	91,077	114,197	134,215	153,580	175,183	197,372
			PP1	PP2	PP3	PP4	PP5	PP6	PP7	PP8	PP9	PP10	
			30/09/2021	31/10/2021	30/11/2021	31/12/2021	31/01/2022	28/02/2022	31/03/2022	30/04/2022	31/05/2022	30/06/2022	
PROJECT MAIN CONTRACT COST		**Claim**	**$ 1,342,314.64**	**$ 2,368,155.16**	**$ 4,030,468.20**	**$ 5,768,948.49**	**$ 7,530,530.38**	**$ 9,671,127.39**	**$ 11,370,559.55**	**$ 13,848,177.14**	**$ 16,224,892.35**	**$ 18,831,017.80**	



Project Manhours - Stormwater Projects

Prepared: C. Stewart
Date 31-May-23
ammended 28-Jul-23

Description	Ave manhrs	FTE Equivalent	Contractor	Main contract value	duration (months)	approx. total manhrs	PV/manhr
Open cut trenching - Greenfield large pipe diameter - ~2.4m dia	1600	7.39	[redacted]	$ 3,400,000	6	9,600	$ 354.17
Micro-tunnel - 2.1m dia 700m drive Shaft	4250	19.63	[redacted]	$ 28,600,000	20	85,000	$ 336.47
Micro-tunnel - 1.2 dia pipejack 60m - space constrained TTM	1350	6.24	[redacted]	$ 3,650,000	12	16,200	$ 225.31

NB: Hours are based upon H&S reporting and represent construction phase and site workforce.

Figure 19: Confidentialised excerpts of real-world project data used in calibration.

Accompanying commentary provided by Cameron Stewart from Alta Consulting:

- *To calibrate and validate model outputs we reviewed projects in the NZIC pipeline to which Alta have direct involvement. We selected a live roadworks project as a key case study project. We chose this particular project as it uses a cost re-imbursement contract. In this form of contract, the main contractor records all manhours and costs incurred and adds an agreed profit margin which the client is liable. To substantiate the cost claims, the contractor must submit their accounting records. This means there is a much greater level of detail available for tracking manhours than is typical with other contract forms. As the project is still under construction, total manhours were extrapolated. Order of magnitude multipliers were applied to subcontractor costs to account for subcontract costs where direct manhours aren't recorded. The forecast total manhours were compared to the simulated outputs from the demand model. Adjustments were made to the weights of input matrices to compensate for discrepancies and to correlate outputs.*

- *Other projects considered included a range of recent Stormwater Infrastructure projects. For the reference projects, we reviewed the summary information of program, manhours and project values. These contracts were of a more typical configuration where there is generally less granular data available to the managing client teams. As a result, the same level of interrogation as conducted on our study roadworks project was not possible. On projects such*

as these, reported manhours are generally collected by the main contractor for Health and Safety reporting. These figures would not typically include hours for off-site labour, overheads or design and planning.

Appendix 9: Supply and Demand model comparison by occupation

Occupation code	Occupation name	Demand estimate	Supply estimate	Difference (absolute)	Difference (relative)	Group
133111	Construction Project Manager	4110	3446	664	19%	Good agreement
232112	Landscape Architect	302	262	39	15%	Good agreement
233212	Geotechnical Engineer	499	374	125	33%	Good agreement
233213	Quantity Surveyor	1180	1292	112	9%	Good agreement
233215	Transport Engineer	411	494	83	20%	Good agreement
233311	Electrical Engineer	802	1122	320	40%	Good agreement
233915	Environmental Engineer	305	61	244	403%	Good agreement
234312	Environmental Consultant	305	283	22	8%	Good agreement
251312	Occupational Health and Safety Adviser	303	699	396	131%	Good agreement
263312	Telecommunications Network Engineer	144	415	271	189%	Good agreement
312113	Building Inspector	364	163	201	123%	Good agreement
312114	Construction Estimator	863	436	427	98%	Good agreement
312212	Civil Engineering Technician	325	626	301	93%	Good agreement
312512	Mechanical Engineering Technician	45	406	361	811%	Good agreement
312611	Safety Inspector	52	391	339	656%	Good agreement
331111	Bricklayer	904	433	471	109%	Good agreement
331212	Carpenter	5084	3522	1562	44%	Good agreement
332111	Floor Finisher	402	616	214	53%	Good agreement
332211	Painting Trades Worker	2124	1838	286	16%	Good agreement
333111	Glazier	904	454	451	99%	Good agreement
333211	Plasterer (Wall and Ceiling)	904	478	427	89%	Good agreement
333212	Renderer (Solid Plaster)	904	474	430	91%	Good agreement
334114	Gasfitter	564	132	433	329%	Good agreement

341111	Electrician (General)	5712	6685	973	17%	Good agreement
511112	Program or Project Administrator	1381	1589	208	15%	Good agreement
712915	Concrete Pump Operator	573	187	386	206%	Good agreement
721214	Excavator Operator	2280	2329	49	2%	Good agreement
721216	Loader Operator	454	543	89	20%	Good agreement
721913	Paving Plant Operator	381	879	499	131%	Good agreement
721915	Road Roller Operator	506	199	308	155%	Good agreement
821112	Drainage, Sewerage and Stormwater Labourer	487	119	368	309%	Good agreement
821113	Earthmoving Labourer	374	190	184	97%	Good agreement
821411	Building Insulation Installer	452	50	402	804%	Good agreement
821511	Paving and Surfacing Labourer	381	160	220	137%	Good agreement
821711	Construction Rigger	608	527	81	15%	Good agreement
821712	Scaffolder	1060	989	71	7%	Good agreement
899999	Labourers nec	11752	13880	2128	18%	Good agreement
133112	Project Builder	1279	2873	1593	125%	Poor agreement
133211	Engineering Manager	1934	968	966	100%	Poor agreement
133612	Procurement Manager	993	142	851	601%	Poor agreement
139912	Environmental Manager	648	111	537	483%	Poor agreement
232111	Architect	807	2240	1433	178%	Poor agreement
232611	Urban and Regional Planner	845	1566	721	85%	Poor agreement
233211	Civil Engineer	2219	6132	3912	176%	Poor agreement
233214	Structural Engineer	500	2171	1670	334%	Poor agreement
233512	Mechanical Engineer	446	1565	1119	251%	Poor agreement
233999	Engineering Professionals nec	72	600	528	735%	Poor agreement
312111	Architectural Draftsperson	72	601	529	737%	Poor agreement

312112	Building Associate	4565	1034	3530	341%	Poor agreement
312312	Electrical Engineering Technician	466	1083	617	133%	Poor agreement
322313	Welder (First Class) (Aus) / Welder (NZ)	188	778	590	314%	Poor agreement
331211	Carpenter and Joiner	1271	46	1225	2669%	Poor agreement
333311	Roof Tiler	904	295	609	206%	Poor agreement
333411	Wall and Floor Tiler	904	164	741	452%	Poor agreement
334111	Plumber (General)	6034	2558	3476	136%	Poor agreement
334113	Drainer	43	1367	1323	3049%	Poor agreement
334115	Roof Plumber	564	4	560	14180%	Poor agreement
341113	Lift Mechanic	790	132	658	499%	Poor agreement
342111	Airconditioning and Refrigeration Mechanic	871	239	631	264%	Poor agreement
342411	Cabler (Data and Telecommunications)	2174	56	2118	3808%	Poor agreement
511111	Contract Administrator	2608	823	1785	217%	Poor agreement
712111	Crane, Hoist or Lift Operator	2045	833	1212	146%	Poor agreement
721211	Earthmoving Plant Operator (General)	996	2508	1512	152%	Poor agreement
721999	Mobile Plant Operators nec	746	127	620	489%	Poor agreement
821111	Builder's Labourer	5991	3071	2920	95%	Poor agreement
821114	Plumber's Assistant	1467	30	1437	4805%	Poor agreement
821211	Concreter	1832	624	1208	194%	Poor agreement
821713	Steel Fixer	1237	349	888	254%	Poor agreement
821714	Structural Steel Erector	709	76	633	833%	Poor agreement
821915	Surveyor's Assistant	764	55	708	1282%	Poor agreement
899923	Road Traffic Controller	2236	1108	1129	102%	Poor agreement
149411	Fleet Manager	294	112	182	162%	Too small
224511	Land Economist	21	11	10	91%	Too small
224512	Valuer	14	76	62	427%	Too small

232511	Interior Designer	Not in model	211	211	-	Too small
233914	Engineering Technologist	72	55	17	30%	Too small
263311	Telecommunications Engineer	233	205	28	14%	Too small
311415	Hydrographer	Not in model	3	3	-	Too small
312115	Plumbing Inspector	274	8	266	3262%	Too small
312116	Surveying or Spatial Science Technician	72	114	42	58%	Too small
312199	Architectural, Building and Surveying Technicians nec	Not in model	114	114	-	Too small
312211	Civil Engineering Draftsperson	72	151	79	110%	Too small
312311	Electrical Engineering Draftsperson	72	26	46	181%	Too small
312511	Mechanical Engineering Draftsperson	45	51	7	15%	Too small
312911	Maintenance Planner	Not in model	221	221	-	Too small
313212	Telecommunications Field Engineer	Not in model	46	46	-	Too small
313213	Telecommunications Network Planner	233	13	219	1631%	Too small
313214	Telecommunications Technical Officer or Technologist	72	24	48	199%	Too small
331112	Stonemason	Not in model	94	94	-	Too small
331213	Joiner	127	261	134	106%	Too small
342212	Technical Cable Jointer	Not in model	85	85	-	Too small
342412	Telecommunications Cable Jointer	Not in model	57	57	-	Too small
342413	Telecommunications Linesworker (Aus) / Telecommunications Line Mechanic (NZ)	Not in model	14	14	-	Too small
711913	Sand Blaster	Not in model	49	49	-	Too small
721213	Bulldozer Operator	294	68	226	334%	Too small
721215	Grader Operator	294	204	91	44%	Too small
721916	Streetsweeper Operator	Not in model	52	52	-	Too small
821611	Railway Track Worker	Not in model	206	206	-	Too small

821912	Driller's Assistant	Not in model	56	56	-	Too small
362512	Tree Worker	147	Not in model	147	-	Too small
821913	Lagger	181	Not in model	181	-	Too small
312999	Building and Engineering Technicians nec	Not in model	1697	1697	-	Not in both models
322211	Sheetmetal Trades Worker	Not in model	594	594	-	Not in both models
322311	Metal Fabricator	Not in model	1294	1294	-	Not in both models
342211	Electrical Linesworker (Aus) / Electrical Line Mechanic (NZ)	Not in model	1138	1138	-	Not in both models
342414	Telecommunications Technician	Not in model	1033	1033	-	Not in both models
721311	Forklift Driver	Not in model	725	725	-	Not in both models
821311	Fencer	Not in model	476	476	-	Not in both models
139916	Quality Assurance Manager	491	Not in model	491	-	Not in both models
232212	Surveyor	1424	Not in model	1424	-	Not in both models
334112	Airconditioning and Mechanical Services Plumber	1041	Not in model	1041	-	Not in both models
821911	Crane Chaser	1060	Not in model	1060	-	Not in both models

Table 20: Comparison of supply and demand model estimates for all occupations.