



Australian Government

2024 Occupation Shortage List Methodology

14 October 2024



Contents

| Overview | 3 |
|---|----|
| Introduction | 3 |
| How the OSL is used | 3 |
| Scope and granularity | 4 |
| Changes in the methodology for the 2024 edition | 4 |
| Caveats | 5 |
| Occupation shortage definition | 5 |
| OSL categorisations | 6 |
| OSL assessment framework | 6 |
| Stage 1: Generating preliminary occupation ratings | 7 |
| The GBM | 7 |
| Variable importance | 8 |
| Blending the GBM-based fill rates with SERA | 9 |
| Constructing confidence intervals | 10 |
| Considerations and limitations of the modelling framework | 11 |
| Converting fill rates into ratings | 11 |
| Stage 2: Verification – Human check and benchmarking | 13 |
| The OSL Stakeholder Survey | 14 |
| Labour Demand to Labour Supply ratios | 18 |
| Recruitment difficulty rates | 19 |
| SERA weighted by occupation similarity scores | 20 |
| Desktop research | 21 |
| Stage 3: Ratings feedback: Incorporate stakeholder feedback | 21 |
| Consultation with select external stakeholders | 21 |
| Appendices | 23 |
| Appendix A – Occupations out of OSL scope | 23 |
| Skill Level 5 occupations | 23 |
| New Zealand occupations | 24 |
| Occupations where the labour market is not open and contestable | 24 |
| Appendix B – Mapping to 2022 ANZSCO | 26 |
| Case 1: One-to-one (Full) occupation match | 26 |
| Case 2: One-to-multiple (Partial) occupation match | 26 |
| Case 3: Multiple-to-one (Mixed) occupation match | 26 |
| Case 4: New occupations | 27 |
| Appendix C – The GBM model performance | 27 |
| Appendix D – GBM input variables | 30 |

| Appendix E – SERA Methodology | 31 |
|--|----|
| E.1 Sampling Methodology | 31 |
| E.2 SERA Questionnaire | 32 |
| Appendix F – Blending the GBM-based fill rates with SERA | 33 |
| Appendix G – Constructing confidence intervals | 34 |
| Appendix H – Stakeholder Survey | 35 |
| Survey program | 35 |
| Appendix I – SERA weighted by occupation similarity scores | 37 |

Overview

Introduction

The Occupation Shortage List (OSL), produced by Jobs and Skills Australia, provides a list of occupations in shortage in Australia and in each state and territory. The OSL is released annually and as a point-in-time assessment of the shortage status of occupations. The findings and insights derived from the OSL products can be useful for informing labour market policies and programs.

The OSL was formerly known as the *Skills Priority List* (SPL). The product has been rebranded to better align the list's title to the definition of shortage used and the insights generated.

In this paper, Jobs and Skills Australia describes the methodology used to assess occupations and produce the 2024 OSL ratings. The scope of the OSL, the definition of occupation shortage, the evidence underpinning occupation assessments and the analytical framework used to determine occupation ratings are all covered in this paper.

The sources used to assess occupation shortages are diverse. These include data modelling, statistical analysis of the labour market, employer and stakeholder surveys, and additional engagements with various stakeholder groups. Stakeholders consist of peak bodies, industry groups, professional associations, unions, education providers and regional representative bodies in the Australian labour market. Feedback is also sought from the Commonwealth Government, state and territory governments, as well as <u>Jobs and Skills</u> <u>Councils</u>.

The paper has been written for a general audience, so that the content and key steps in the methodology are easily grasped, but there is content that may require some prior knowledge of statistics.

This report is intended to be read in its entirety. Whilst different sections of this report set out various aspects of the methodology, these individual sections could be misinterpreted if they are considered in isolation.

How the OSL is used

The OSL is a key source of labour market intelligence. Together with other analytical outputs produced by Jobs and Skills Australia, it informs advice on a range of Australian Government policy and program initiatives.

As the OSL database grows over the years, it will enable researchers and policy makers to distinguish between occupations that experience persistent shortages and those affected by economic fluctuations. This insight may inform the development of labour market and training policies and programs.

The OSL is available on the Jobs and Skills Australia website, along with additional reports that provide comprehensive analyses of various occupational groups.

Scope and granularity

The OSL covers over 900 occupations, defined in the 2022 *Australian and New Zealand Standard Classification of Occupations* (ANZSCO) at the occupation level.¹ Only ANZSCO Skill Levels 1 to 4 are included in the OSL due to their link to education and training. Occupations in shortage for several years, particularly those where qualifications are mandatory, have implications for the education and training sector in Australia.

Several occupation groups are excluded from the OSL:

- Skill Level 5 occupations.²
- New Zealand-based occupations.³
- Occupations where the labour market is not open and contestable.⁴

The full list of excluded occupations is in Appendix A.

Therefore, the OSL covers those occupations that require post-school qualifications and where there is a competitive process for selecting suitably skilled applicants for vacant positions.

Occupations are first assessed using various sources of input at the national level for occupations as defined in the 2013 ANZSCO. This is because most data from the Australian Bureau of Statistics (ABS) and Jobs and Skills Australia are based on this classification. This reliance enables greater use of data sources to support a thorough assessment of occupations. Once the assessment is complete on 2013 ANZSCO basis, the occupational ratings are then converted to the 2022 ANZSCO framework. The approach used to map occupation ratings from 2013 ANZSCO to 2022 ANZSCO is outlined in Appendix B.

Changes in the methodology for the 2024 edition

The 2024 OSL methodology has introduced significant improvements over the previous year's approach.

- A Gradient Boosting Machine (GBM) learning model has been adopted, replacing Generalised Linear Model (GLM) regression model and the human-written decision tree algorithms used to generate the preliminary ratings.
- Some data sources that were previously used to verify the indicative ratings generated from algorithms are now being used as part of the inputs into the GBM model. This helps to produce preliminary ratings more effectively.
- The OSL Stakeholder Survey data was used to verify the preliminary ratings and incorporate stakeholder qualitative feedback.⁵

¹ ANZSCO is a skill-based classification used to classify all occupations and jobs in the Australian and New Zealand labour markets. It is organised into a 5-level hierarchy: major groups, sub-major groups, minor groups, unit groups and occupations. Occupation refers to the most granular 6 digit ANZSCO and is the level that the report refers to. For more details on ANZSCO refer to the <u>ABS website</u>.

² The Australian Bureau of Statistics defines Skill Level 5 occupations as having a level of skill commensurate with Australian Qualifications Framework Certificate I or compulsory secondary education. For some occupations a short period of on-the-job training may be required in addition to or instead of the formal qualification. The Skill Level 5 occupations have fewer barriers to entry and unlike other in-scope occupations, they generally do not require significant post-school education and training.
³ Māori occupations are examples of New Zealand-based occupations only.

⁴ Judgement is used to determine occupations to exclude. The judgement is based on the definition of open and contestable labour market; job titles and descriptions; and tasks performed and specialisations within the occupations. Additional details are included in Appendix A.

⁵ The main reason for this change was that the stakeholder survey data was more high level, qualitative in nature and has small sample sizes. These limitations impacted the representativeness of the survey findings and thereby limited the ability to be used in the machine learning model.

• Definitions and rules for converting GBM model outputs to occupation ratings were also revised and streamlined.

Caveats

The production of occupation shortage ratings is not a straightforward process. The process includes modelling and implementation of definitional rules, which are largely automated; verifying occupation ratings using additional information, which is mostly an application of critical thinking and judgement; and incorporating feedback from stakeholders.

The abovementioned stages of the production process are all dependent on the quality of evidence available (quantitative and qualitative) and that shared by stakeholders.

Moreover, there are additional caveats:

- The analysis is based on ANZSCO, which does not identify all jobs in the labour market separately that employers seek skilled workers for. However, ANZSCO offers a standard way to collate, analyse and report data in consistent manner across time.
- There are time-lags in the data inputs. For instance, the *Survey of Employers who have Recently Advertised* (SERA) has a 3-month lag from the data collection to the implementation in the GBM. There are other data sources which are lagged 6 months to a year.
- The OSL is a point-in-time assessment of occupations in the labour market and is not forward looking.
- The OSL is an employer-focused view of shortages and is driven by data on employers' recruitment activity and other information on employing entities and not the self-employed.
- The OSL does not provide an instrument to inform occupations that are critical to the economy. But the insights from the OSL can be used, among other information, to develop an understanding of the importance of those occupations.
- At present, there is not a distinction between genuine vacancies and "Ghost vacancies". That is, vacancies which are falsely advertised by employers to test the labour market in their respective industry.

Occupation shortage definition

Jobs and Skills Australia uses the following definition of shortage: when employers are unable to fill, or have considerable difficulty filling, vacancies for an occupation or cannot meet significant specialised skill needs within that occupation, at:

- Current levels of remuneration
- · Current conditions of employment; and
- In reasonably accessible locations.

Based on this definition, the primary measure of an occupational shortage is the ability of employers to fill vacancies or the *proportion of vacancies that are filled* (the fill rate). Therefore, the *fill rate* is the key proxy measure of shortage:

- Lower fill rates imply greater employer difficulty with filling vacant positions, indicating a shortfall in suitably skilled workers and a higher likelihood of occupations in shortage.
- Higher fill rates imply fewer challenges with filling vacancies and a lower likelihood of shortages.

This is illustrated below in Figure 1.

Figure 1. Occupation shortage broad definition



Source: Jobs and Skills Australia.

Another advantage of using fill rates is that the metric accounts for both employer demand for skilled workers and the supply of those workers. Analysis also demonstrates that the fill rate is a sound proxy measure of occupation shortage in the labour market.

The fill rate is also underpinned by a simple intuition. There is a high likelihood of an occupation being in shortage when:

- estimated fill rates are below 67%
- a sizeable proportion of potential fill rates (within an 80% confidence interval) falls below 67%.

However, to ensure assessments of occupations are comprehensive and accurate, the OSL methodology considers a broader set of evidence beyond fill rates and confidence intervals.

OSL categorisations

The OSL has four ratings to classify the shortage status on a specific occupation:

- Shortage (S): An occupation is in national shortage or overall shortage
- Metropolitan Shortage (M): An occupation is in shortage in a metropolitan area
- Regional Shortage (R): An occupation is in shortage in a regional area
- No Shortage (NS): An occupation is not in shortage.

OSL assessment framework

The 2024 OSL methodology used to determine occupation ratings is a mix-methods approach. The approach consists of 3 stages:

- 1. The preliminary occupation ratings produced via a machine learning model and rules to convert the model outputs and occupation employment size into occupation ratings.
- 2. The manual verification of preliminary occupation ratings using additional sources of information.
- 3. The incorporation of external feedback from the Commonwealth Government, state and territory governments and Jobs and Skills Councils.

These stages are shown in Figure 2 and explained in greater detail in the subsequent sections of the paper.

Figure 2: A 3-stage occupation assessment process



Source: Jobs and Skills Australia.

Stage 1: Generating preliminary occupation ratings

Preliminary occupation ratings are based on a modelling framework that includes a Gradient Boosted Machine learning model (GBM). The key elements of the modelling framework are summarised in Table 1.

Table 1: Key steps of the preliminary occupation rating modelling framework

Key elements of the modelling framework

- 1. Estimating occupation fill rates using a GBM model
- 2. Blending the GBM-based fill rates with weighted moving average SERA fill rates
- 3. Producing 80% confidence intervals for the blended occupation fill rates
- 4. Converting the blended fill rates, confidence intervals and occupation employment sizes into occupation ratings using a set of decision rules

The **GBM**

The GBM replaced the previously used Generalised-Linear Model (GLM). The GBM model uses decision trees to predict fill rates for each occupation. The GBM operates by consecutively fitting new decision trees to enhance the estimate of the response variable (the fill rate). The above process of fitting new models continues until a stopping rule is reached. For example, when the performance of the model on the data set on which it is trained ceases to improve. The final predicted fill rates are therefore the weighted sum of the predicted values made by the previous tree's models.⁶ Thus, the GBM produces the estimated fill rates in a gradual, additive and sequential manner.

⁶ More information on GBMs is available online, including high level introductions. An example is: Introduction to Boosted Trees — xgboost 2.0.3 documentation.

The GBM performs better than the previous GLM. This is discussed in more detail in Appendix C.

The final GBM model uses 24 indicators from 13 sources. Most indicators (16) are at the national level, 7 are at state-level and there is a metro/non-metro indicator for each statistical area level 4 (SA4) – a spatial unit defined under the <u>Australian Statistical Geography</u> <u>Standard (ASGS)</u>. Figure 3 shows the inputs used in GBM.

Figure 3: Inputs to the GBM



Source: Jobs and Skills Australia.

Note: IVI refers to the Jobs and Skills Australia's *Internet Vacancy Index*. Mean similarity refers to occupation similarity scores from the *Australian Skills Classification* (ACS), which was developed to provide a common language of skills to increase understanding and recognition of skills across occupations, sectors and contexts. More information on the ACS is available on the Jobs and Skills Australia <u>webpage</u>.

A more detailed list of data sources that are used in the GBM are in Table D1 in Appendix D.

Variable importance

Figure 4 shows the importance of the different variables that are used as inputs for the GBM. As can be seen, the unemployment rate, which is a measure of labour supply, provides the strongest signal to the model. Gender, measured by the share of hours worked by women, provides a strong signal.

Additional Jobs and Skills Australia data include occupation flows, from *Data on Occupation Mobility*; occupation similarity scores from *Australian Skills Classification*; and employment data from *Nowcast of Employment by Region and Occupation* (NERO) also provide significant predictive power. Information from Jobs and Skills Australia's *Recruiting Employer Outlook Survey* (REOS) also provides significant predictive power. However, as NERO

covers only a few financial years, its importance is dampened in the variable importance chart compared to variables which are available for all years.

Figure 4: Variable importance of new indicators in final model



Source: Jobs and Skills Australia.

Blending the GBM-based fill rates with SERA

A key variable used in the modelling framework, including featuring in the GBM as the target variable is the vacancy fill rate from SERA. Box 1 summarises SERA.

Box 1: SERA overview

SERA is the *Survey of Employers who have Recently Advertised*. These employers have had recent interactions with the labour market and are, therefore, able to provide information on current recruitment conditions and identify issues for selected occupations. The survey asks employers a range of questions regarding their recruitment experience for an advertised vacancy, collecting both quantitative and qualitative data.

SERA collects data on 9,000 employers annually on average, spread over approximately 350 occupations. Data is collected for ANZSCO defined occupations. It focuses on relatively large occupations (national employment of at least 1,500 people) and where enough job advertisements are available each year to conduct employer surveys.

SERA is designed for assessing occupational shortages and provides direct measures of employer experiences when recruiting.

Examples of the quantitative data collected include the proportion of vacancies filled, the number of applicants, qualified and suitable applicants. Examples of qualitative questions include reasons why vacancies are not filled, why applicants are considered unsuitable and the impact of recruitment challenges on employers.

More information on SERA is provided in Appendix E.

The estimated GBM and SERA fill rates, for each occupation are first converted to a linear predictor scale using a logit transformation and blended. The process is discussed in more detail in Appendix F.

Constructing confidence intervals

The modelling framework also produces a level of certainty around each estimated fill rate, which is used to derive an 80% confidence interval for the estimated fill rate. A confidence interval is the range of values that an estimated value is expected to fall between with a certain level of probability (or confidence). More details on the construction of confidence intervals can be found in Appendix G.

Figure 5 illustrates the modelling framework used to estimate occupation fill rates and confidence intervals prior to converting them to occupation ratings.

Figure 5: Overview of estimating occupation fill rates and confidence intervals



Source: Jobs and Skills Australia.

Considerations and limitations of the modelling framework

GBM is a useful tool for estimating occupation fill rates, which are then used to define occupation shortages. However, like all models, there are some limitations.

- Results are extrapolated to occupations different to those typically surveyed in SERA. The
 model is applied to all in-scope occupations. However, SERA has traditionally focused on
 larger, higher-skilled occupations, as these are most relevant for training. It is reasonable
 to assume that model performance will be weaker on categories of occupations that have
 never been surveyed or only included recently, particularly lower-skilled occupations.
 GBM includes a skill level indicator in the model which provides some protection against
 this risk.
- When there are no SERA results for an occupation, the indicator model has relatively
 wide confidence intervals. This can limit the ability for authoritative declarations of
 Shortage or No Shortage. In practice, those occupations with SERA results are much
 more likely to have an authoritative finding. However, despite this issue, the GBM
 performs well in ranking occupations by their fill rates from low to high.
- The model is reliant on the quality of the underlying data.
- An implicit requirement of the model setup is that data sources remain similar over time. Any structural change in the data will require the GBM to be retrained.

Converting fill rates into ratings

To obtain preliminary ratings, Jobs and Skills Australia applies a set of decision rules based on vacancy fill rates and other criteria.

National or overall shortages (S)

An occupation is rated as in national Shortage (S) if:

- 1. The occupation fill rate is below 67% in both metro and regional areas, and
- 2. The occupation fill rate is below 67% in an area that has a majority employment share, which is defined as 80% or more.

Example: An occupation would be rated as **S** if the employment share in a metropolitan area was 85% and the occupation fill rate in that area was 62%.

Regional/Metropolitan shortages (R/M)

An occupation is rated as in regional shortage (R) if:

- 1. The occupation fill rate is below 67% in a regional area, and
- 2. The employment share of the occupation in the regional area was between 20% and 80%.

A similar rule holds for Metropolitan shortage (\mathbf{M}) if the above criteria are met for a metropolitan area.

Example: An occupation would be rated as R if the employment share in the regional area is 40% and the occupation fill rate in the regional area is 57%.

Additional criteria for occupations to be rated as S, M, or R

- 1. 90% or more of an occupation's fill rates fall below 67% within an 80% confidence interval
- 2. At least a 20 percentage point change as compared to last year in the percentage of fill rates that fall below 67% within an 80% confidence interval.
- 3. For an S, M or R to apply, the employment size for an occupation must be above the bottom 25th percentile (approximately 500 workers).
- 4. If the employment size is between the 10th and 25th percentiles, and the occupation satisfies the above definitions and rules, an uncertainty flag is applied.

Criteria for occupations to be rated as NS

- 1. If the above rules are not satisfied an NS rating is applied.
- 2. If the employment size is below the 10th percentile, an NS rating is also applied.

The above parameters and rules were based on sensitivity tests to produce the most stable and consistent results overtime.

These rules and thresholds are applied to ensure that:

- the preliminary ratings are based on strong data-evidence; and
- any changes in ratings reflect, as best as possible, changes in labour market conditions rather than being influenced by sudden shifts or data volatility.

These initial occupation ratings are further verified in the second stage of the OSL assessment process.

Figure 6 below illustrates the occupation shortages scenarios





Source: Jobs and Skills Australia.

Stage 2: Verification – Human check and benchmarking

In this stage of the assessment process, the preliminary occupation ratings are subjected to a human review and verification. This second stage is designed to ensure that the ratings generated by the modelling framework, described in the previous section, are sensible and reliable.⁷

A human check is needed as the GBM and the rules used to generate preliminary occupation ratings:

- do not consider all important sources of information, particularly qualitative information
- rely on primary data sources that can be limited for some occupations, leading to increased uncertainty in the preliminary ratings.

Preliminary occupation ratings that present uncertainty, such as those with wide confidence intervals or those that have experienced a change from the rating from last year, are prioritised for review.

In this stage, additional analysis is undertaken on a checklist of quantitative and qualitative data that were not captured by the GBM model to verify the preliminary occupation ratings. Unlike the automated initial assessment, this stage relies on human judgment to ensure a nuanced and accurate evaluation.

The data sources for the verification stage are:

- Quantitative and qualitative results from the OSL Stakeholder Survey and additional evidence from non-survey submissions by stakeholders.
- Analysis of *Internet Vacancy Index* (IVI) to unemployment ratios (labour demand to labour supply ratio).
- Analysis of Recruitment difficulty rates from REOS.
- Additional metrics from SERA data such as qualified applicants per vacancy, suitable applicants per vacancy and vacancy duration.
- Desktop research of occupations and industries.

The various outputs and metrics used from the above listed sources are described further below.

For each occupation, these additional metrics are compared with the preliminary rating and the elements that generated the rating – the fill rate, the confidence interval and employment size. Collaborative discussions and the application of judgment are facilitated by using a verification dashboard that provides a clear summary of the modelling framework outputs and the additional metrics from the above listed sources at the occupation level.⁸ The collaborative approach ensures that a consensus is reached, even in cases of ambiguity.

Simple examples of applications of judgment are below.

The preliminary rating for an occupation was:

• **NS** but it only just failed to meet the rules and conditions listed in the previous section. Further, the additional sources listed provided evidence of the occupation being in shortage. The preliminary rating would be changed to a Shortage.

⁷ Testing on past year's OSL outcomes shows that, on average, there is an 85% to 90% match rate between the GBM and rules-based ratings and human-based occupation assessments.

⁸ The verification dashboard is an analytical tool developed by Jobs and Skills Australia to check the preliminary ratings against a set of indicators which were not used when generating the preliminary occupation ratings or were used in a different form (for example, ABS, *2021 Census of Population and Housing* data) but are considered useful to make a final decision on ambiguous cases.

• **S** but it only just met the rules and conditions listed in the previous section. Further, the employment level was below the 25th percentile. That is, employment size was small. Further, the statistical tests performed in the previous section (see Table 4 above) failed to reject the null hypothesis. Moreover, the additional sources listed provided no or very little evidence of the occupation being in shortage. The preliminary rating would be changed to No Shortage.

The verification process on average results in 10% to 15% of occupations from the preliminary ratings being changed. Such changes occur most frequently for the occupations where ratings were considered low-confidence or inconclusive in the preliminary stage.

When all available sources of information are exhausted and no convincing evidence of shortages is found in either the preliminary ratings or the verification assessments, the occupation is rated as No Shortage.

The OSL Stakeholder Survey

Jobs and Skills Australia's *OSL Stakeholder Survey* (stakeholder survey) is an annual survey open to the public between November – February each year. The survey primarily targets peak bodies, industry groups, professional associations, unions, regional representative bodies, and education and training advisory bodies. However, government agencies, recruitment agencies/labour hire firms, employers and Jobs and Skills Councils are welcome to complete the survey.

Stakeholders provide information and evidence of their recruitment experience and/or that of the members they represent. The stakeholder survey is complementary to SERA. Participants are asked to provide information for occupations defined in the 2022 ANZSCO.

Box 2: Rationale for use

The stakeholder survey provides qualitative and additional information that is not captured in data that features in the preliminary ratings and other data considered in the verification stage of assessments. For example, peak-body and/or survey participants that represent employers or workers have extensive engagements with said cohorts.

The stakeholder survey focuses on collecting quantitative and qualitative information on:

- Whether the proportion of vacancies filled is low, moderate or high, by national level, state or territory, and regionality (metropolitan or regional area or both).
- Where possible, the number of vacancies advertised, vacancies filled, applicants and reasons vacancies were not filled.
- The evidence supporting the above.

The stakeholder survey questions are further detailed in Table 2.

Table 2: Main variables used in the stakeholder survey scoring system

| Variable | Response type | |
|---|----------------------------|--|
| Proportion of vacancies filled | Multiple choice: | |
| | High proportion filled | |
| | Moderate proportion filled | |
| | Low proportion filled | |
| | Unsure | |
| What evidence did you consider to make your selection (Proportion of vacancies filled)? | Free text | |
| Which level best describes the supply of workers? | Multiple choice: | |
| | Oversupply | |
| | Meeting demands | |
| | Undersupply | |
| | Unsure | |
| What evidence did you consider to make your selection (supply of workers)? | Free text | |
| Which level best describes the use of Employer | Multiple choice: | |
| Sponsored Visas? | High use | |
| | Moderate use | |
| | Low use | |
| | No use | |
| | Unsure | |
| What evidence did you consider making your selection (the use of Employer Sponsored Visas)? | Free text | |
| How many vacancies were advertised? | Numeric free text | |
| How many advertised vacancies were filled? | Numeric free text | |
| Provide any comment for your selections (number of vacancies filled and advertised, number of applicants and reason for lack of applicants) | Free text | |

Source: Jobs and Skills Australia, 2024 Occupation Shortage List Stakeholder Survey.

The full stakeholder survey questionnaire is provided in Appendix H.

Jobs and Skills Australia publishes a report, *Snapshot of the 2024 Occupation Shortage List Stakeholder Survey Results*, which summarise the results from the *2024 Occupation Shortage List Stakeholder Survey*.

The main outputs used in the verification are a Shortage or No Shortage indicator, the strength of the indicators and qualitative information such as free text provided by survey participants. The method used to generate the Shortage or No Shortage indicator and the strength of the indicators are described below.

Stakeholder survey scoring method

Responses from the stakeholder survey are used, along with other indicators, to verify the preliminary occupation ratings. To simplify the verification process, responses are collated, scored and aggregated to produce a signal (Shortage/No Shortage) and a signal strength (weak, moderate, strong) for each occupation reported on.

Both the signal and the signal strength for each occupation, in principle, are determined by the number of survey responses, the completeness of each response, the evidence supporting the response and the representativeness of the response.

Scoring each response for an occupation

For each survey response received for an occupation, Jobs and Skills Australia allocates a score based on the completeness of the survey and the evidence provided.

- Scores given to survey responses, for an occupation, have values ranging between -13.5 to +13.5.
- The sign of the survey response score, for an occupation, is a potential indicator of a Shortage (positive sign) or No Shortage (negative sign).
- The absolute value of the magnitude of score signals the completeness of the survey response and the consistency with which the questions are answered. It reflects the strength of the Shortage or No Shortage indication and the reliability placed on the survey participants' response for the specific occupation.

For example:

A score of +13.5 for a survey response is a fully completed survey and indicates a potential shortage for an occupation. Similarly, a -13.5 score for a survey response indicates a fully completed survey and indicates, potentially, no shortage for the occupation. These scores also mean that the strength of the shortage or no shortage indicators are strong. That is, a high level of reliability may be placed on the survey responses.

A score of -0.3 or +0.3 indicates an incomplete survey and is a relatively weaker No shortage/Shortage signal for an occupation.

Geography and size

Each overall score for a survey response for an occupation is weighted by 2 factors: geography for which the survey responses reflect and size of the organisation completing the survey.

- The geography weighting is based on the employment size of the occupation covered in the survey response in the geographical areas (state or territory, and/or metropolitan or regional) the survey participant has selected.⁹
- The organisation size weighting is based on how representative the organisation is in terms of other organisations, members and employees.¹⁰

These weightings are applied to better reflect the representativeness of one survey response relative to the labour market and to other submissions for a single occupation.

⁹ Metropolitan refers to Capital City and Regional refers to Rest of State as defined in the Australian Bureau of Statistics *Australian Statistical Geography Standard*.

¹⁰ Survey respondents are asked on their number of employees, the size of their membership base (for unions and employeerepresenting bodies), and the number of employers they represent (for employer-representing bodies e.g. industry groups).

For example, a small carpentry business in Tasmania will not be as representative of carpentry in Australia compared to a large national body which surveyed its members.

Creating an overall score for an occupation

An occupation may have multiple survey responses. Each of the survey response scores for an occupation are then aggregated. Similar to the scores for an individual survey response:

- A positive aggregate score indicates a Shortage signal for an occupation, while a negative aggregate score indicates a No Shortage signal.
- The magnitude of the aggregate score reflects the strength of the Shortage or No Shortage indication for an occupation.

For example:

For an occupation, suppose we received 4 survey responses with weighted scores of - 3, -2, -1 and 7 then we would calculate an aggregate score by adding these together:

(-3) + (-2) + (-1) + 7 = 1

Even though 3 responses indicated a *No Shortage* rating (indicated by the negative number), the overall score indicated a *Shortage* rating due to the fourth response strength.

Similarly, if we received 6 survey responses for an occupation with the weighted scores of -4, -2, -2, 1, 1 and 3 then the addition of these scores would give the aggregate score below:

(-4) + (-2) + (-2) + 1 + 1 + 3 = -3

Even though 3 responses indicated a No Shortage rating, and 3 responses indicated a Shortage rating, the overall score indicated a No Shortage rating (as it was a negative overall score).

After generating the final aggregate score for an occupation, the absolute value of the score is used as the signal strength of the Shortage or No Shortage indication. The absolute value of the score is then converted into weak, moderate or strong category.

The conversion criteria to turn the absolute value of the aggregate scores into a weak, moderate and high signal strength category is shown in Table 3.

Table 3. Numerical signal strength conversion

| Score | Category |
|-------|--------------------------|
| 0–3 | Weak signal strength |
| 3–4 | Moderate signal strength |
| 4+ | High signal strength |

For example:

Following on the aggregate scores from the previous example, an overall score of 1 for an occupation indicates stakeholders are providing a Weak Shortage signal.

Whereas an overall score of -3, for an occupation, indicates that stakeholders are providing a Moderate No Shortage signal.

Additional evidence from non-survey submissions by stakeholders

Some stakeholders submit evidence and documentation on occupations and industries either via email or as part of their response to the stakeholder survey. This evidence is coded into the appropriate ANZSCO occupation and treated in-line with the desktop research (discussed below).

Labour Demand to Labour Supply ratios

The ratio of Jobs and Skills Australia's IVI to unemployment from the ABS *Labour Force Statistics* were used as a proxy measure of labour demand to labour supply ratios. Both measures are financial year based (July to June).

Rationale for use

Labour demand to supply ratios provide a proxy for shortage pressures in the labour market. If the ratio is greater than 1, then the demand for the occupation is higher than the potential workers available. Conversely, a ratio below 1 indicates an oversupply of potential workers for the occupation

Table 4 below lists the metrics use in the verification analysis.

Table 4: Summary of the IVI to unemployment ratios

| Metric | Description | | |
|--|---|--|--|
| Raw average ratio | The average across 3 financial years, 2021–22 to 2023–24. | | |
| The annual change in the raw ratio | The annual changes from: | | |
| | • 2021–22 to 2022–23; and | | |
| | • 2022–23 to 2023–2024. | | |
| The Z-scores of the ratio and their groupings according to their deviation from the mean ¹¹ | The 3 year average ratio and 2023–24 ratio were converted into Z-scores and then grouped into whether they were 0.5 standard deviations below, within or above their respective mean. | | |

How the metrics were used in the analysis can be demonstrated through a simple example:

¹¹ The Z-scores are based on the log-transformed rates. Z-scores are a way to standardise raw figures. It is a statistical measure that quantifies the distance between a data point and the mean of a dataset. It's expressed in terms of standard deviations. It indicates how many standard deviations a data point is from the mean of the distribution. Z-scores above the mean are positive, while those below the mean are negative.

- a positive annual change in the ratio from 2022–23 to 2023–24, including Z-scores of the 3 year average and Z-scores of the 2023–24 ratios lying above 0.5 standard deviations of the mean would provide a signal of shortage pressures for an occupation.
- a negative annual change in the ratio from 2022-23 to 2023-24, including Z-scores of the 3-year average and Z-scores of 2023-24 ratios lying below 0.5 standard deviations of the mean would provide a weak signal of shortage pressures for an occupation.

Recruitment difficulty rates

Data from Jobs and Skills Australia's REOS captures a wealth of information, including on recruiting difficulty rates.¹²

Rationale for use

Like fill rates, recruitment difficulty rates provide an indication of the recruitment challenges of employers. Recruitment difficulty rates of occupations are inversely related to occupation fill rates. Generally, the higher the difficulty rate, the lower the fill rate for an occupation. This metric provides valuable insights on employer's recruitment challenges.

REOS has a question on whether the respondent has a recruitment difficulty issue or not. The response recorded is either Yes or No. Converting "Yes" to 1 and "No" to 0, the values are then grouped by occupation ANZSCO codes to produce recruitment difficulty scores ranging from 1: "Very difficult to recruit" to 0:" Very easy to recruit".

Table 5 below lists the metrics use in the verification analysis.

Table 5: Summary of recruitment difficulty rate metrics

| Metric | Description | | |
|---|---|--|--|
| Raw average rate | The average across 3 financial years, 2021–22 to 2023–24. | | |
| The annual change in the raw rate | The annual changes from: | | |
| | • 2021–22 to 2022–23; and | | |
| | • 2022–23 to 2023–2024. | | |
| The Z-scores of the rate and their groupings according to their deviation from the mean ¹³ | The 3 year average rate and 2023–24 rate were converted into Z-scores and then grouped into whether they were 0.5 standard deviations below, within or above their respective mean. | | |

The metrics are analysed in an analogous way to that described for IVI to unemployment ratios.

¹² Please refer to <u>REOS methodology and questionnaire</u> for additional information.

¹³ As per footnote 12.

SERA weighted by occupation similarity scores

As the scope of SERA is limited to approximately 350 occupations, Jobs and Skills Australia uses similarity scores to identify occupations that have skills similar to ones we have coverage for. The SERA data is weighted by occupation similarity scores from the *Australian Skills Classification* (ASC).¹⁴

The weighted measure includes:

- Qualified applicants per vacancy
- Suitable applicants per vacancy
- Vacancy duration (weeks unit, yearly from May 2023 to May 2024)
- The gap between weighted qualified applicants per vacancy and weighted suitable applicants per vacancy.

In cases where the initial occupation has SERA survey observations more than or equal to 30, the original SERA data is used. Where this is not the case, only occupations with a similarity score above 80% were used to weight the SERA data. More details are found in Appendix I.

Table 6 below lists the metrics used in the verification analysis.

Table 6: Summary of additional SERA data metrics

| Metric | Description | | |
|--|--|--|--|
| Z-scores of the average | The average across 3 financial years, 2021-22 to 2023-24. | | |
| The annual change in the metrics | The annual changes from: | | |
| | • 2021–22 to 2022–23; and | | |
| | • 2022–23 to 2023–2024. | | |
| The Z-scores of the metrics and their groupings according to their deviation from the mean ¹⁵ | The 3-year average and 2023-24 metrics were standardised and then grouped into whether they were 0.5 standard deviations below, within or above their respective mean. | | |

The additional SERA metrics reveal insights on the potential shortage pressure or lack thereof for an occupation. For example, an occupation may have:

- qualified applicants per vacancy that are 0.5 standard deviations above the average but suitable applicants per vacancy that are 0.5 standard deviations below or within the average. This may suggest employer difficulty finding skilled workers to fill vacant positions.
- vacancy duration that is falling over time, suggests that vacancies are becoming easier to fill. This may suggest that employers may be finding it easier to fill vacancies.

¹⁴ The ASC was developed to provide a common language of skills to increase understanding and recognition of skills across occupations, sectors, and contexts. While, ASC has been decommissioned and is to be replaced with a new National Skills Taxonomy, it remains available for use in analysis, particularly a matrix of scores which shows the commonality of occupation in terms of skills and tasks performed. This data is valuable for imputing missing SERA data and buttressing it to strengthen the analysis. Please refer to the <u>Australian Skills Classifications</u> for more details.

¹⁵ As per footnote 12.

• strong growth in both qualified and suitable applicants per vacancy in line with softening labour market conditions.

Desktop research

For each occupation, where possible, a summary of claims is made through relevant media articles, industry reports and other publications. This information is principally used as part of judgement confirmation.

Stage 3: Ratings feedback: Incorporate stakeholder feedback

Consultation with select external stakeholders

The aim is to seek additional context or evidence that may not have been considered under stages 1 and 2 of the occupation assessments.

Feedback is sought from:

- Federal Government agencies on the national occupation ratings
- Jobs and Skills Councils (JSC) on national and on state and territory occupation ratings for occupations that relate to each JSC's areas of industry expertise
- State and territory government agencies on where national occupation ratings do not apply to their respective state or territory, reflecting the unique labour market characteristics that are observed in each state and territory, the difference in the methodologies used by them when conducting their own occupation shortage analysis and the difference in stakeholder feedback that they each receive.

Several criteria are applied to the feedback received to assist with incorporating and accepting feedback on occupation ratings, including the following:

- Stakeholder feedback on occupations aligns with the scope of the OSL (that is, Skill Level 1 to 4 occupations and those not excluded for additional reasons as explained in Appendix A).
- Stakeholder feedback that is focused on the current status of an occupation rather than its possible future status.
- Where stakeholder-suggested ratings differ from preliminary occupation ratings, several factors are considered:
 - Preliminary occupation ratings may be based on data that is not robust or has wide confidence intervals.
 - Preliminary occupation ratings that are on the borderline of being classified as S, M, R or NS.
 - Stakeholder-suggested ratings that are underpinned by compelling evidence that were not previously considered by Jobs and Skills Australia.

Similar to the stage 2 verification stage, described in the previous section, further collaborative discussions and application of judgment is used to reach a consensus when accepting the feedback from the stakeholders.

When incorporating the ratings, including the state and territory ratings, the feedback is assessed against the strength of the Jobs and Skills Australia preliminary ratings and the stage 2 verified ratings and the evidence underpinning them.

Regarding the state and territory ratings, the feedback is compared with the ratings generated by the GBM model at state level where the GBM results are robust and reliable.

Additional rules are applied to deduce the occupation rating at state and territory level and verify the occupation rating at national level.

- **Rule 1:** In cases where feedback from a state or territory is absent, the state or territory will adopt the stage 2 verified national level occupation rating.
- **Rule 2:** A state or territory will be assigned a No Shortage rating if an occupation has zero employment size in that state or territory.
- **Rule 3:** This rule is to ensure the national ratings reflect the cumulative impact of incorporated state and territory ratings.

If states and territories with a Shortage rating for an occupation make up at least two-thirds of the national employment level for that occupation, the national rating will also be a Shortage.

In some cases, if this two-thirds threshold is not satisfied but the employment share of the states and territories for the occupation is relatively close to two-thirds, judgement is applied when deciding whether or not the national level rating is changed.

The above-described approach to accepting state and territory feedback or allocating ratings to each state and territory ensures that the national ratings reflect the cumulative impact of state and territory ratings.

For example:

The national total employment for occupation A is 1,000. Occupation A is rated as Shortage in NSW, VIC, and QLD with employment figures of 400, 200 and 100, respectively. It is rated as No Shortage in the remaining state and territories with combined employment size of 300. The combined employment share of Shortage ratings (NSW, VIC and QLD) is 70%, while the No Shortage rating (all other state and territories) make up 30%. Thus, the national rating for occupation A will be Shortage rating following by the rule of majority employment share.

Appendices

Appendix A – Occupations out of OSL scope

There are several occupations that are excluded from the OSL, including Skill Level 5 occupations, occupations only present in New Zealand and occupations where the labour market has been judged to be not open and contestable.

Skill Level 5 occupations

The table below shows the ABS ANZSCO skill level classification. There are 156 Australian Skill Level 5 occupations that are excluded from the OSL scope.

| Skill Level | Definition |
|----------------|--|
| Level 1 | Occupations at Skill Level 1 have a level of skill commensurate with a bachelor degree or higher qualification. At least five years of relevant experience may substitute for the formal qualification. In some instances, relevant experience and/or on-the-job training may be required in addition to the formal qualification. |
| Level 2 | Occupations at Skill Level 2 have a level of skill commensurate with one of the following: |
| | NZ Register Diploma or |
| | AQF Associate Degree, Advanced Diploma or Diploma. |
| | At least three years of relevant experience may substitute for the formal qualifications listed above. In some instances, relevant experience and/or on-the-job training may be required in addition to the formal qualification. |
| Level 3 | Occupations at Skill Level 3 have a level of skill commensurate with one of the following: |
| | NZ Register Level 4 qualification |
| | AQF Certificate IV or |
| | AQF Certificate III including at least two years of on-the-job training. |
| | At least three years of relevant experience may substitute for the formal qualifications listed above. In some instances, relevant experience and/or on-the-job training may be required in addition to the formal qualification. |
| Level 4 | Occupations at Skill Level 4 have a level of skill commensurate with one of the following: |
| | NZ Register Level 2 or 3 qualification or |
| | AQF Certificate II or III. |
| | At least one year of relevant experience may substitute for the formal qualifications listed above. In some instances, relevant experience may be required in addition to the formal qualification. |

Table B1: ABS skill level classification

| Skill Level | Definition |
|----------------|--|
| Level 5 | Occupations at Skill Level 5 have a level of skill commensurate with one of the following: |
| | NZ Register Level 1 qualification |
| | AQF Certificate I or |
| | compulsory secondary education. |
| | For some occupations, a short period of on-the-job training may be required in addition to or instead of the formal qualification. |
| | In some instances, no formal qualification or on-the-job training may be required. |

Source: ABS, Australian and New Zealand Standard Classification of Occupations.

The ABS defines Skill Level 5 occupations as having a level of skill commensurate with *Australian Qualifications Framework* Certificate I or compulsory secondary education. For some occupations, a short period of on-the-job training may be required in addition to or instead of the formal qualification. The skill level occupations have fewer barriers to entry and unlike other in-scope occupations, they generally do not require significant post-school education and training.

New Zealand occupations

There are seven Māori specific occupations that are excluded from the OSL. Māori specific occupations are examples of New Zealand occupations only.

Table B2: List of New Zealand occupations

| ANZSCO | Description |
|--------|--|
| 241112 | Kaiako Kohanga Reo (Māori Language Nest Teacher) |
| 241211 | Kaiako Kura Kaupapa Māori (Māori-medium Primary School Teacher) |
| 241212 | Pouako Kura Kaupapa Māori (Māori-medium Primary School Senior Teacher) |
| 252215 | Traditional Māori Health Practitioner |
| 411512 | Kaiāwhina (Hauora) (Māori Health Assistant) |
| 422113 | Kaiāwhina Kohanga Reo (Māori Language Nest Assistant) |
| 422114 | Kaiāwhina Kura Kaupapa Māori (Māori-medium School Assistant) |
| | |

Source: ABS, Australian and New Zealand Standard Classification of Occupations.

Occupations where the labour market is not open and contestable

Due to some regulations or complex market mechanisms, there are 16 occupations where the labour market is not open and contestable. For example, defence force roles, judges and detectives will not be covered as recruitment/appointment is mainly conducted internally.

Defining an open and contestable labour market

An open labour market, as defined by the <u>Australian Human Rights Commission</u> is a market 'open to everyone' such that there is an absence of discrimination in the access to and maintenance of employment on grounds of race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, which has the intention

or effect of impairing or nullifying exercise of the right to work as enumerated in article 2 of *International Covenant on Economic, Social and Cultural Rights* (ICESCR). The Commission also suggests that age should be a factor that is protected from discrimination, but restricting the work rights of non-citizens would not violate article 2 of ICESCR.

A contestable labour market refers to a labour market where there are no significant barriers to entry or exit for firms and workers. This type of market is characterized by a competitive environment where workers are free to move between jobs and employers compete to attract and retain employees.

It is typically associated with a high level of transparency in job vacancies and recruitment processes, ensuring that the best candidate is hired based on merit.

In the context of the OSL, occupations that are considered to have an open and contestable labour market take into consideration the above definitions and those that are publicly advertised and require a competitive process for selecting suitably skilled applicants for vacant positions. This ensures that the OSL covers occupations that tend to require post-school qualifications and where there is a fair and competitive selection process.

Further judgement is applied on the job titles and descriptions; and tasks performed and specialisations within the occupations when deciding on occupations to exclude.

| ANZSCO | Description |
|--------|--|
| 111212 | Defence Force Senior Officer |
| 111311 | Local Government Legislator |
| 111312 | Member of Parliament |
| 111399 | Legislators nec |
| 139111 | Commissioned Defence Force Officer |
| 139112 | Commissioned Fire Officer |
| 139113 | Commissioned Police Officer |
| 139211 | Senior Non-commissioned Defence Force Member |
| 271211 | Judge |
| 271212 | Magistrate |
| 271213 | Tribunal Member |
| 441111 | Defence Force Member – Other Ranks |
| 441311 | Detective |
| 452411 | Footballer |
| 452412 | Golfer |
| 452499 | Sportspersons nec |

Table B3. List of occupations not considered to be open and contestable

Source: Jobs and Skills Australia.

Appendix B – Mapping to 2022 ANZSCO

Assessments and occupation ratings are initially based on the 2013 version of ANZSCO. This is because many data sources currently published at Jobs and Skills Australia (such as the IVI) are published on a 2013 ANZSCO basis.

The occupation ratings from 2013 ANZSCO are then mapped onto the 2022 ANZSCO occupations to produce the OSL, reflecting the newest occupational framework. This mapping is guided by the ANZSCO correspondence tables, from 2013 to 2021 and from 2021 to 2022, as published by ABS. The interactions between the different frameworks can be complex, but they can be reduced to the four following scenarios.

Case 1: One-to-one (Full) occupation match

This is when there is no difference between an occupation on the 2013 basis and the 2022 basis, which is the case for most occupations. In this case the rating from the Stage 1 assessment based on 2013 ANZSCO is carried across.



Case 2: One-to-multiple (Partial) occupation match

This is when an occupation on the 2013 basis branches out into many occupations on the 2022 basis. In this case the rating from the Stage 1 assessment based on 2013 ANZSCO is carried across for all matches found.



Case 3: Multiple-to-one (Mixed) occupation match

This is when multiple occupations on the 2013 basis map to the same occupation on the 2022 basis. This is the most difficult case, as the constituent ratings may disagree with each other. In this case the employment size for each 2013 ANZSCO occupation is used as a weight. The rating with the highest weight becomes the 2022 ANZSCO occupation rating.



Case 4: New occupations

Fifteen occupations were introduced in 2022 ANZSCO and do not map from any occupation on the 2013 basis. In this case, there is no historical data from SERA and REOS. Therefore, for these occupations the previous year's ratings are maintained in Stage 1 OSL assessment. The ratings are then verified against other sources of evidence and adjusted accordingly in Stages 2 and 3 as described in the body of the report.

Appendix C – The GBM model performance

The GBM performance results are summarised in Table C1 below.

| | Accuracy (higher is better) | | Gains ratio (higher is better) | | RMSE (lower is better) | |
|-------------------------------|--------------------------------|---------|-----------------------------------|---------|---------------------------|---------|
| Model | All years | 2022–23 | All years | 2022–23 | All years | 2022–23 |
| Constant model | 52.5% | 54.7% | 0% | 0% | 14.3% | 20.9% |
| 2021 GLM | 59.7% | 62.1% | 53.3% | 54.3% | 11.28% | 16.5% |
| GBM model | 66.2% | 70.5% | 66.7% | 66.3% | 10.30% | 15.5% |
| Improvement of GBM vs. GLM | 10.9% | 13.5% | 25.1% | 22.1% | -8.7% | -6.1% |
| (GBM ÷ GLM - 1) | | | | | | |

Table C1: Comparison of GBM with the GLM at the National level¹⁶

Source: Jobs and Skills Australia.

Compared to the 2021 GLM, the new GBM model has better performance results across all 3 metrics (accuracy, gains ratio and RMSE) at the national level. This improvement is observed for:

- occupations with SERA data and occupations where there is no SERA data.
- unseen occupations across all years and unseen occupations in 2022–23 only. That is, for data that was not included in the set used to train the GBM.

The GBM model is better for capturing historical variations in the data while still demonstrating superior performance. The GBM gives the best combination of predictive accuracy compared to other tested models (GLM, neural network and penalised regression structures). It also has a better ability to capture complex relationships between labour market indicators and fill rates. This includes indicators that were previously used in the GLM as well as the new labour market indicators included in the GBM.

Figure C1 below provides an illustration of a decision tree. This is a hierarchical structure which splits the full model data (all training data) based on a series of rules (or decisions).

¹⁶ Accuracy is defined by how well the model predicts fill rates compared to SERA fill rates. The gains ratio is a measure which summarises how well the predicted ordering of fill rates (from highest to lowest) matches the observed ordering of fill rates. Root mean square error (RMSE) is a measure of the average absolute deviation between predicted and observed fill rates.





Source: Jobs and Skills Australia.

In Figure C1:

- The training dataset is first split on the national unemployment rate being above or below 5%.
- Where the national unemployment rate is above 5%, there is a second split based on the number of vacancies per employed person being above or below 0.01.¹⁷

After these 2 splits, the training dataset has now been divided into 3 groups based on the unemployment rate and the number of IVI vacancies per employed person. The depth of this decision tree is 2 as there are (at most) 2 decision rules allocating the data into the groups. The algorithm selects the best decision rule for the data in each group. This is the split, which maximises the homogeneity of the modelling target within each split.

As mentioned above, GBM fits many decision trees, with each subsequent tree fit to the residual (target – prediction) to improve the current prediction.

The GBM estimates fill rates over a 12-month period from April of one year to April of the following year for all occupations defined by 2013 ANZSCO.

Table C2 below provides a summary of the relative strengths and weaknesses of GBMs.

¹⁷ As per Jobs and Skills Australia's Internet Vacancy Index (IVI).

Table C2: Summary of relative strengths and weaknesses of GBMs

| Strengths | | Weaknesses | | |
|-----------|---|------------|---|--|
| • | High predictive accuracy – widely regarded as one of the best predictive models for tabular data. | • | Black-box nature with: | |
| | | | Limited control over the model structure | |
| | | | Greater effort required in understanding the effects of individual predictors | |
| | | | More difficulty in determining whether a predictor is statistically significant | |
| • | Quick to build – the model can implicitly capture: | • | Risk of overfitting (where the model works well for the training dataset and | |
| | Non-linear effects of predictors | | poorly for other datasets) | |
| | Interactions between predictors | | | |
| • | Robust to outliers without needing to explicitly cap or cup variables | • | More computationally intensive, leading to longer run time for fitting the model and using it to make predictions compared to regression models. | |

Source: Jobs and Skills Australia.

There are several parameters in a GBM which specify aspects of the fit or place constraints on the model. Table C3 below provides a summary of some these parameters, and the value used in the GBM. Some parameters are fixed for a given modelling context. For others, the parameters can be tuned to optimise performance in the context and there are generally a range of acceptable values. In tuning parameters there is a trade-off between:

- Providing a better fit to the training data
- Overfitting to the training data leading to poorer generalisability.

Table C3: Some GBM parameters

| Parameter | Description | Value |
|--|--|-------|
| Number of trees | This is the maximum number of decision trees to be used in the ensemble. A higher total number of trees gives a more complex ensemble that can better capture the nuances of the data. However, this also takes a longer time to fit and run. | 1,000 |
| Depth | This is the maximum number of decision trees to be used in the ensemble. A higher total number of trees gives a more complex ensemble that can better capture the nuances of the data. However, this also takes a longer time to fit and run. | 3 |
| Minimum number of observations in a node | This is the minimum number of observations within each final grouping of the individual decision trees. More granular tree structures capture patterns in smaller subgroups of the data, at the risk of overfitting. | 30 |

| Parameter | Description | Value |
|----------------------------|---|-------|
| Shrinkage or learn rate | This parameter controls how fast the model learns, decreasing the shrinkage rate will reduce the contribution of individual trees to the prediction and reduce the chance of biased estimates. However, this then requires a higher number of trees to learn the data, leading to longer run times. | 0.02 |

Source: Jobs and Skills Australia.

Appendix D – GBM input variables

The following table outlines the indicators used in the model.

The final GBM model uses 24 indicators from 13 sources. Most indicators (16) are at the national level, however there are 7 state-level indicators, and a metro/non-metro indicator for each SA4. Table D1 below lists the labour market indicators and the underlying data sources. Additionally, the SERA itself is a critical data input into the SPLIM as it provides the 'true' fill rate of vacancies against which the model is calibrated.

Table D1: GBM input variables

| Category | Variable | Lowest level used | Sources and notes |
|-----------------------|---|-------------------------|---|
| IVI vacancies | IVI vacancies per employed | State | |
| | IVI vacancies per employed 3 year change | State | Jobs and Skills Australia |
| NERO | Number employed | State | Jobs and Skills Australia; only available at ANZSCO unit group, using ABS data to fill in missing and |
| | Employed 3 year change | National | older values |
| Employment | Unemployment rate | State | |
| market | Unemployment rate 1 year change | National | ABS Labour Force Survey |
| | Seek weeks 3 year change | National | |
| Visas | Visas granted per employed | National | Home Affairs |
| Demographics | Proportion of hours worked by people aged under 25 | State | |
| | Proportion of hours worked by people aged 55 and over | State | ABS, 2021 Census of Population and Housing |
| | Proportion of hours worked by females | National | |
| Geography | Proportion employed in metro areas | State | ABS, 2021 Census of Population and Housing |
| | Whether the SA4 is in a metro or non-metro area | SA4 | ABS, ASGS |
| ACS Similarity matrix | Mean similarity across top 3 most similar occupations | National | Jobs and Skills Australia |

| Category | Variable | Lowest level used | Sources and notes |
|--------------------------------------|--|-------------------------|---|
| Skill level | Skill level of occupation (1–5); Skill Level 1 is the highest skill level | | ABS, ANZSCO |
| REOS | Scaled fill rate at ANZSCO unit group – the difference in the raw fill rate from the ANZSCO major group level mean is scaled down where there are few surveys contributing to the data point | National | Jobs and Skills Australia; annual data, limited historical series as only started in 2022 |
| Occupation flows | Total size of inflow to the occupation | National | Jobs and Skills Australia |
| | Inflow yearly change | | |
| | Percentage net flow; net flow relative to occupation size | | |
| | • Outflow diversity. Entropy of the previous occupations of those in the occupation. A higher entropy means employees come from a more diverse range of previous occupations. | | |
| Lightcast ¹⁸ vacancies | Proportion requiring bachelor and above | National | Lightcast vacancies |
| Occupation training | Study diversity | National | ABS, 2021 Census of Population and Housing |
| Unemployment flows | Proportion of unemployment flows into occupation | National | Jobs and Skills Australia |
| Unemployment transitions | Proportion of all unemployment transitions which are into a given occupation | | Jobs and Skills Australia |

Appendix E – SERA Methodology

E.1 Sampling Methodology

A sample of employers for selected occupations are surveyed through a structured, telephone-based survey. Employers are asked about their experiences recruiting for specific advertised vacancies in particular occupations (see Appendix E.2 for the SERA questionnaire).

¹⁸ Formerly Burning Glass.

Main points

- Vacancies are surveyed across all states and territories.
- The samples are based on employment levels in the regions such that they proportionally
 match the population as per ABS, 2021 Census of Population and Housing. However, to
 have sufficient data to draw conclusions from smaller jurisdictions and occupations, more
 data than what would be proportionally accurate is collected.¹⁹
- Surveyed vacancies are for specific positions offered for paid work of 15 hours or more per week and at least 3 months' duration.²⁰
- Vacancies advertised by recruitment agencies are included in SERA if they are for an actual vacancy with an employer rather than a general 'canvassing' advertisement.
- Advertisements for self-employment or partnerships are generally excluded. However, in industries where there is significant subcontracting (for example, construction), such positions may be included.

Attempts are made to survey an appropriate number of employers from both metropolitan and regional areas. The proportion of vacancies outside metropolitan areas depends on the state or territory's employment profile (as per ABS, *2021 Census of Population and Housing*), as well as the profile for the particular occupations. The term metropolitan area refers to state and territory Capital City and regional refers to the Rest of the State as defined under the <u>Australian Statistical Geography Standard (ASGS)</u>.

Coverage

SERA data is now collected for over 350 occupations. Over half of the occupations covered by SERA are Professionals major group 2 and Technicians and Trades Workers major group 3. The survey also includes Manager major group 1 and Community and Personal Service Workers major group 4, as well as a number of targeted occupations within the remaining ANZSCO major groups 5 to 8.

E.2 SERA Questionnaire

SERA Questionnaire

What is the postcode where the vacancy is located?

How many positions for [target occupation] were you attempting to fill?

How many of these positions were filled?

How many people applied for the position(s)?

Are formal qualifications required for the position?

- (If yes) what qualification are required?
- (If yes) how many applicants had the required qualification?
- (If no) how many applicants held a relevant, formal qualification?

Is relevant experience required for the position?

• (If yes) what is the minimum length of experience required for the position?

¹⁹ For example, with the limited number of employers that can realistically be surveyed, a given occupation may only represent a small proportion of the labour market such that 10 vacancies would be collected. However, as 10 survey responses are not sufficient for robust analysis, a minimum of 40 vacancies for many occupations (and 20 vacancies in limited circumstances) is set instead.

²⁰ The thresholds used are to ensure consistency with past approaches.

SERA Questionnaire

Are any specific skills or specialised experience required for the position?

• (If yes) what specific skills or specialised experience are required?

How many applicants were suitable?

That is, they had the qualifications, skills, and experience to do the job?

What were the main reasons applicants were considered unsuitable for the position?

How long have you been trying to fill the vacancy/ies? or

How long did it take you to fill the vacancy/ies?

If insufficient suitable applicants were attracted, why?

If suitable applicants were attracted, but the vacancy was not filled, why?

If vacancy unfilled, what will be done now?

What are the main tasks and duties of the position(s)?

Is the position full-time, part-time, or casual?

(If part-time or casual) for how many hours per week, on average, is the position?

Is the position permanent (ongoing) or a contract (fixed term)?

(If contract) what is the length of the contract?

What goods or services does your organisation mainly produce or supply (what industry is your organisation in)?

How many staff are currently employed in this organisation in Australia?

Do you have any other comments in relation to this recruitment round or the labour market for this occupation?

Would you like to receive a copy of the report once it is published?

Appendix F – Blending the GBM-based fill rates with SERA

The base GBM and SERA fill rates, for each occupation, are first converted to a linear predictor scale using a logit transformation. For example, for occupation *j*, the GBM-based fill rates, $p_{i,GBM}$ are transformed to:

$$LP_{j,\text{GBM}} = \log\left(\frac{p_{j,\text{GBM}}}{1 - p_{j,\text{GBM}}}\right)$$

The *LP* refers to linear predictor. The SERA fill rates, $p_{j,SERA}$, are similarly transformed.

GBM fill rates are then blended with 5 year weighted moving average fill rates from SERA. The blending is based on a Bayesian credibility approach, where more weight is applied to the source (GBM or SERA) with the lower variance. The reasons for blending are as follows:

- It is assumed SERA data is the 'ground truth', so the two sources are aligned.
- The blending process reduces the overall standard error and narrows the width of the confidence intervals, providing more confidence for the likely occupation ratings.
- Partial weight can be given to SERA data in situations where only a small to moderate number of vacancies have been surveyed.

The credibility weight used to determine the blended fill rate for occupation *j* is the contribution to uncertainty of the GBM model and SERA fill rates and is shown below.

$$\phi_j = \frac{\sigma_{GBM}^2}{\sigma_{GBM}^2 + \sigma_{j,SERA}^2}$$

Therefore, if the number of SERA vacancies surveyed is very large, $\sigma_{j,SERA}$ will be small and the credibility factor ϕ_j will be close to 1. On the other hand, if there are very few SERA vacancies surveyed, $\sigma_{i,SERA}$ will likely be large and the credibility factor ϕ_j will be closer to 0.

The blended fill rate for occupation *j*, then takes the below form.

$$LP_j^* = \phi_j LP_{j,SERA} + (1 - \phi_j) LP_{j,GBM}$$

The blending is conducted on the linear predictor scale, which transforms the fill rate (usually a proportion bounded between 0 and 1) to the full number line.

The blended fill rate is then converted back into probability or a proportion fill rate figure using the inverse logit transformation:

$$1/\{1 + \exp(-LP_i^*)\}$$

Regarding SERA fill rates, instead of using the observed fill rates from a particular year, a 5year weighted moving average is used. This is due to the strong autocorrelation in SERA. That is, correlation between current and past years' data. The higher the autocorrelation, the more relevant SERA results from older years are in estimating the current fill rate for the occupation.

The weight for a survey result that is *t* years before the current year is ρ^t , where $\rho < 1$ and is the decay factor. This reflects that SERA results from older years will carry some information on fill rates, however results from more recent years are more relevant and so given more weight in the final estimate.

The observed SERA fill rate is calculated as the weighted number of filled vacancies (from the survey) divided by the weighted number of total vacancies (from the survey) in the last 5 years. A decay factor (ρ) of 0.75 is used.

Therefore, the SERA fill rates in 2024 for a particular occupation *j* is:

$$p_{j,SERA,2024} = \frac{m_{j,2024} + \rho m_{j,2023} + \rho^2 m_{j,2022} + \rho^3 m_{j,2021} + \rho^4 m_{j,2020}}{n_{j,2024} + \rho n_{j,2023} + \rho^2 n_{j,2022} + \rho^3 n_{j,2021} + \rho^4 n_{j,2020}}$$

Where $m_{j,t}$ is the number of vacancies filled in occupation *j* and year *t* and $n_{j,t}$ is the number of vacancies in SERA in the year *t* for occupation *j*.

Additional statistical tests are performed on the estimated fill rates to validate the initial rating generated by the GBM model. These tests are to confirm the confidence in the preliminary ratings or flag occupations with uncertain ratings. The tests assess whether there is a statistically significant difference in the fill rates:

- · between an occupation and the national average, and
- between two years for an occupation.

Appendix G – Constructing confidence intervals

The modelling framework produces a level of certainty around the estimated fill rate, which is used to derive an 80% confidence interval for the estimated fill rate. A confidence interval is the range of values that an estimated value is expected to fall between with a certain level of probability (or confidence).

The formula used to calculate the upper and lower bounds of the confidence interval (CI) for each occupation is:

Fill Rate Upper/Lower
$$CI = Fill Rate \pm t_{\alpha/2,n-1}\sigma_{fill rate}$$

where $t_{\alpha/2,n-1}$ is the two-sided critical value of the Student's t-distribution at $100(1 - \alpha)\%$ confidence and with *n*-1 degrees of freedom.

The standard errors used in producing the confidence intervals are as follows.²¹

$$\left(\frac{1}{\frac{1}{\sigma_{GBM}^2} + \frac{1}{\sigma_{j,SERA}^2}}\right)^{\frac{1}{2}}$$

For occupations with higher numbers of SERA survey results, this standard error shrinks the confidence interval substantially.

Appendix H – Stakeholder Survey

Survey program

Jobs and Skills Australia's stakeholder survey seeks to capture information from stakeholders on recruitment challenges for a wide range of occupations.

Survey respondents can provide additional documentation to support their claims, such as member surveys, industry reports or their own internal modelling or data. Information gathered through the surveys is then considered during the assessment process.

Table H1: Stakeholder Survey Questionnaire

Stakeholder Survey questionnaire

About the organisation (Part A)

What is the name of the organisation?

Which best describe your organisation?

How many people does your organisation employ?

How many members does your organisation represent?

How many other organisations does your organisation represent?

In a few words, please describe the products and services your organisation offers

About the organisation (Part B)

Which industry does your organisation represent?

About the organisation (Part C)

Which states and territories does your organisations operate in?

Which state/territory is considered the primary location of the organisation?

What is the suburb/town and postcode of the primary location of the organisation?

About you

²¹ This is directly computable by taking the variance of the updated linear predictor $LP_j^* = \phi_j LP_{j,SERA} + (1 - \phi_j)LP_{j,model}$ and the formula for weights ϕ_j .

Stakeholder Survey questionnaire

Name

Job title

Best contact

Work email

Your advice on proportion of vacancies filled - Australia-wide

For all the members and organisations you represent (including your own organisation), which proportion of vacancies were filled in the period January – December 202X, Australia-wide?

Was the selected proportion specific to the whole of Australia or only metro/regional areas of Australia?

Provide a summary of the evidence you considered to make your selections.

Your advice on proportion of vacancies filled - By states and territories

Same 3 questions as above but by states and territories (NSW, QLD, VIC, ...)

Skill supply of selected occupations

Select the level of suitably skilled workers for the period January – December 202X.

- Undersupply
- Meets Demand
- Oversupply

Provide a summary of the evidence you considered to make your selection.

Use of visas in selected occupations

Select the level of Employer Sponsored Visas best applies for the period January – December 202X.

- No use
- Low use
- Moderate use
- High use

Provide a summary of the evidence you considered to make your selections.

Vacancies filled for selected occupations

How many vacancies were advertised for the period January - December 202X?

How many of those advertised vacancies were filled?

How many people applied for the advertised vacancies?

Where were the vacancies advertised?

How many years of experience were required for the role?

Which reasons were the main ones the vacancies were not filled?

Provide any comments, including a summary of evidence you considered.

Appendix I – SERA weighted by occupation similarity scores

Using qualified applicants per vacancy for occupation j as an example, the following is the approach used to weight SERA data with the similarity scores.

 $Qualified \ applicants \ per \ vacancy_{j} = \frac{\sum_{i=0}^{n} similarity_{i} * qualified \ applicants_{i}}{\sum_{i=0}^{n} similarity_{i} * vacancies_{i}}$

- *i* represents the index of each similar occupation. When *i* is zero, it is referring to the original SERA occupation itself.
- *n* is the number of similar occupations.
- *vacancies* are the total number of vacancies employers were aiming to fill with their recruitment.