Determinants of the economic outcomes of Australian permanent migrants

Peter Varela[*] Robert Breunig

Tax and Transfer Policy Institute
Crawford School of Public Policy
Australian National University

[*] We would like to thank the Department of Home Affairs and the Commonwealth Treasury for financial support to undertake this research. The authors have benefited substantially from comments from the Department of Home Affairs and Treasury, as well as from comments from Cem Ozguzel, Jasper Hesse, Gabriel Chavesbosch, Lachlan Vass, Tyler Reysenbach, Trent Wiltshire, Nick Husek and Thom Williams. All errors and omissions remain our own.

1. Executive summary

This report uses data from the Multi-Agency Data Integration Project (MADIP) to better understand the factors that are associated with economic outcomes of Australian permanent migrants. MADIP is a rich administrative dataset that includes a wider range of important covariates than were available in previous studies of this type. For instance, this study includes information on age, gender, education and English language skills, geographic location, permanent visa category, temporary visa history and country of birth. MADIP includes data over a longer period than other data sources, allowing the examination of long-term trends.

The analysis in this paper uses three main techniques:

- **Descriptive analysis** that visually presents the aggregate economic outcomes of migrants in the years following the grant of a permanent visa. We examine taxable income from all sources and a proxy for labour force participation based upon taxable income.
- **'Mincer' regressions** that estimate the marginal impact of covariates on migrant outcomes. These regressions can answer questions such as "If other factors were held equal, how much more does a migrant with a university degree earn compared to one that has not completed high school?"
- **Oaxaca-Blinder analysis** that decomposes the difference in economic outcomes between groups of migrants into components that can be explained by different observable characteristics and unobservable and unexplained components. These decompositions can answer questions such as "how much of the higher income level of Employer Sponsored Migrants can be explained by higher levels of education or English language skills?" or "How much more do Employer Sponsored migrants earn compared to other migrants that can't be explained by observable characteristics?"

The paper then contains two thematic chapters. The first compares the determinants of migrants' outcomes with similar analysis for the non-migrant population. The second examines how well nominated income predicts the incomes of Employer Sponsored migrants beyond what can be explained through observable characteristics. This addresses the question of whether nominated income provides additional information beyond that which is normally used to select migrants.

The paper concludes with a discussion of how our results can be used to inform the design and evaluation of the Australian migration program.

Key empirical findings

On average, permanent migrants have slightly weaker economic outcomes than the non-migrant population in the years following permanent migration. However, they converge to the outcomes of the non-migrant population over a period of around 10 years.

- There is considerable variation in economic outcomes across visa programs. Skilled visa holders have stronger economic outcomes than non-migrants. Family and Humanitarian visa holders have weaker economic outcomes than non-migrants.
- These migrant programs also have very different integration profiles. Some groups start with strong outcomes and then drop off in the years following migration, while other groups (particularly secondary migrants) have low initial outcomes but improve over a period of decades.

Mincer regression analysis shows that the determinants of economic migrant outcomes defy simple characterisation. Education, English language skills, migrant occupation, demographics, visa characteristics, temporary visa history and country of birth all play an important role in determining migrant outcomes.

- Migrants with higher levels of education enjoy an earnings premium relative to migrants with a high school degree of 11% for an undergraduate degree, 17% for a master's degree and 41% for a PhD.
- Compared to migrants that "only speak English", migrants that report speaking English "Not Well" face a 28% income penalty and are less than half as likely to report an income over $20,000.
- Migrants that have previously held a temporary skilled visa have better earnings outcomes than migrants that have held other types of temporary visas or no temporary visa.
- Migrants that transition through student visas have similar economic outcomes to the average permanent migrant, although this effect varies significantly by the type of education undertaken while on the student visa.
- Migrants enjoy an income premium from living in cities but are also less likely to be employed when living in a city.
- Migrants born in English speaking countries have higher income (even when controlling for English language capability).

Different factors explain short- and long-term outcomes with different importance:

- Education has a larger impact on long-term outcomes than short-term outcomes.
- Visa program and visa history are more important in determining short-term outcomes than long-term outcomes.
- English language skills have similar impacts in the short and long run.

This paper also estimates Mincer regressions separately for each visa group to identify whether the determinants of migrant earnings vary by visa group.

- Employer Sponsored visa holders have the highest returns to education.
- Transitioning through a temporary Skilled visa generates an income premium for all primary visa holders (including Family and Humanitarian visa holders) but this premium is much weaker for secondary visa holders.

Oaxaca-Blinder analysis is then used to compare the economic outcomes of different migrant groups This approach can decompose the average incomes of a group of migrants into two components:

- Returns to observable characteristics (e.g., age, gender, education or English language skills)
- An unexplained component, which captures unobserved characteristics including motivation and attitudes, ease of labour market integration and discrimination.

Around half of the variation in migrant outcomes by visa stream can be explained by observable characteristics.

This approach can determine the extent to which strong economic outcomes are driven by factors that the migration system is designed to select on (such as education levels or English Language skills) as well as the extent these outcomes are driven by characteristics that are not part of selection criteria, but that may have an impact on economic outcomes, such as age and gender.

Taken as a whole, the factors that determine migrant income and employment are similar to the factors that determine outcomes for the non-migrant population. However, some differences exist:

- Permanent migrants have a lower return to education than the non-migrant population.
- Relative to non-migrants, living outside of cities has a lower impact on incomes and employment.

This paper then compares the outcomes of Employer Sponsored migrants that report a nominated income.

- Nominated income, on its own, is a better predictor of migrant income than a regression that includes all other observable characteristics but excludes nominated income.
- Nominated income is based on the market wages that a migrant has agreed with a sponsor and reflects individual characteristics such as age and education levels. It also provides information about the 'unobserved characteristics' of migrants, which can improve the targeting of the visa program (particularly in relation to high income earners).
- Primary migrants with higher levels of nominated income are also older, more likely to be male and have more secondary migrant applicants.

Key policy conclusions

Given the 'non-causal' nature of the analysis in this paper, care must be taken in applying the results to policy. In particular, it is important to ask whether any policy change would affect the underlying dynamics of the visa program. With this key caveat in mind, the main implications for policy from this paper are that:

- The outcomes of migrants from different visa categories (partially) converge in the years after permanent migration. The existing measures of the economic impact of migration (which are calibrated on short-term economic outcomes) are likely to overstate the difference in outcomes between visa groups.
- Estimates from the Mincer regressions in Section 4 can be compared to the Skilled Independent visa points test to understand whether the existing points allocation reflect factors that contribute to economic outcomes of migrants. However, significant further work would be needed to calibrate an 'optimal' points test.
- Occupation is quite important in explaining the different outcomes of visa streams. Any change to the visa system and Skilled Occupation Lists should try to maintain (and improve) the targeting of migrants towards high-earning occupations.
- The economic outcomes of migrants that have previously held a temporary visa are stronger than those that arrive from offshore. This effect is strongest for those that have held a temporary Skilled Visa, but is seen across all temporary visa categories. This suggests that the design of the temporary migration program has large long-term economic outcomes that should be considered in the design of the overall migration program.
- The nominated income reported by Primary Employer Sponsored Migrants is a strong predictor of observed outcomes (and is particularly good at predicting the outcomes of migrants in the upper part of the income distribution). This supports the idea of a wage floor. However, migrants with higher levels of nominated income are also older and have a higher number of secondary migrants on average. These factors reduce the economic impact of the program as older migrants have a lower number of years before retirement and secondary migrants typically earn less than primary migrants. Finally, the finding that nominated income predicts actual income is based on a system without a wage floor and this relationship may change significantly if a wage floor were implemented.

2. Introduction

Successful economic integration is a key goal of the Australian migration program. For migrants, higher degrees of economic success represent significant direct welfare gains. For non-migrants, the economic success of migrants is also a desirable outcome. Economically successful migrants pay taxes and use fewer government services, and therefore do not represent a 'fiscal burden' to the existing population.

Improving the economic outcomes associated with Australian migrants is an ongoing policy goal. For instance, the recent review of the Australian migration system (Parkinson et al. 2023) calls for a migration system that can "Enrich the economy, with a focus on productivity growth" and that "Unlocks the potential of migrants". In its 5-yearly Productivity Review, the Australian Productivity Commission identifies a better targeted Skilled migration system as a reform that has potential to significantly improve Australian productivity.

The goal of this paper is to better understand the factors that contribute to strong economic outcomes among migrants to inform the design of the Australian migration program. It approaches this goal in three steps.

The first step (Section 3) is to compare the aggregate economic outcomes of permanent migrants in the years after receiving a permanent visa to the economic outcomes of the non-migrant population. This analysis is largely descriptive – it controls for migrant age and time/inflation effects but does not otherwise attempt to control for observed individual characteristics. This first set of analysis shows that different groups of migrants have different economic outcomes. It also shows that different groups of migrants have varying periods of economic integration, with some groups having outcomes that continue to improve over a period of decades.

The second step (Section 4) is to estimate a series of 'Mincer' regressions that examine the impact of observable migrant characteristics on economic outcomes. Specifically, this paper considers visa category, previous history on a temporary visa, gender, education level, English language skills, occupation, geographic location within Australia, country of birth, age at migration and the number of secondary migrants as potential factors that could predict the outcomes of migrants. Permanent visas allow migrants to sponsor eligible relatives for permanent residency. These relatives enter Australia under the same visa category as the primary applicant and are known as secondary migrants.

The outcomes from these regressions can be interpreted as the earnings premium related to different migrant characteristics (holding other characteristics constant). For instance, these regressions suggest that amongst permanent migrants, the average earnings premium for an undergraduate degree (compared to high school education) is around 8%. The regression results do not, however, represent the causal impact of characteristics, as the estimates depend both upon individual choices and the migration program's selection of immigrants.

The third step (Section 5) is to use Oaxaca-Blinder decompositions to compare the average outcomes of different groups of migrants. Oaxaca-Blinder analysis decomposes the difference in outcomes across migrant groups into a share than can be explained by observable characteristics and a share that is unexplained. The motivation behind this approach is that while visa characteristics produce large effects in a Mincer Regression, visa programs are also designed to select migrants on observable characteristics (e.g., Skilled migrants are selected in part, on education levels). When comparing the outcomes of visa programs, it is desirable to look at both the observed differences (such as different education levels selected by different visa programs) and the returns to migrants that can't be explained by observable characteristics.

For instance, Primary Employer Sponsored visa holders earn 48.1 per cent more than the average migrant. Of this difference:

- 1.0 percentage point is explained by age (primary Employer Sponsored migrants are older, on average, than other permanent migrants),
- 7.4 percentage points are explained by gender (primary Employer Sponsored migrants are more likely to be male than the average migrant)
- 11.3 percentage points is explained by different occupational patterns.
- 7.4 percentage points is explained by the returns to temporary visa history (Employer Sponsored visa holders have a large share of migrants transitioning from Temporary Skilled visas)
- 3.0 percentage points are explained by English Language Skills
- 1.3 percentage points are explained by Education levels.
- 18.0 percentage points remain unexplained.

A Oaxaca-Blinder decomposition is calculated for each visa category, for migrants with different visa histories and for migrants born in different countries.

A key advantage of the Oaxaca-Blinder approach is that while the migration system is explicitly designed to select on some observable characteristics (like education), it may also select on other characteristics (such as age or gender) that predict higher earnings but may also be undesirable for other policy reasons.[1] Selection on observable characteristics may also work to select on unobservable characteristics such as motivation and attitude. The Oaxaca-Blinder approach can determine which types of selection are occurring within different visa groups and the extent to which these drive aggregate economic outcomes.

Section 6 compares Mincer regressions for the migrant and non-migrant population to determine whether the factors that predict migrant economic outcomes are similar to those that predict non-migrant outcomes. At an aggregate level, determinants of migrant outcomes are reasonably 'similar' to those of the non-migrant population. However, notable differences exist, including a lower return to education for migrants and a lower labour market penalty to living in regional areas for migrants.

Section 7 uses the Oaxaca-Blinder framework to compare the outcomes of Employer Sponsored migrants with different levels of Nominated income. This analysis is directly related to the policy debate about the merits of using an income threshold as the basis of admission for Employer Sponsored visas. This analysis finds that nominated income is a good predictor of observed income (that migrants with higher levels of nominated income earn more). It also shows that this higher level of earnings includes returns to both observable and unobservable characteristics. At high income levels (nominated income over $100,000), there is a large share of income that is unexplained by migrant characteristics, suggesting that nominated income could be an effective mechanism to target very high-income workers that would otherwise be difficult to target using 'observable' migrant characteristics.

[1] For instance, the Australian Treasury's FIONA model (Varela et al. 2021) shows that the Australian migration program is heavily concentrated between the ages of 25-35 and that age is the most important determinant of the lifetime fiscal impact of migrants. A migration program that targeted 40-year-old migrants would likely achieve higher short-term outcomes, but could be undesirable from a longer-term economic perspective.

However, this analysis also shows that migrants with higher levels of nominated income are older, more likely to be male and have more secondary migrants than migrants with lower levels of nominated income. These factors would need to be considered if nominated income were to be used to admit Employer Sponsored visa holders.

Section 8 summarises how the key findings from this paper can be used to improve the design of the Australian migration system.

Interpreting regression results

The analysis in this paper identifies factors that are related to strong economic outcomes of permanent migrants. This analysis can be used to inform two types of question:

- How can the migration system select migrants that will have better economic outcomes when they arrive in Australia? This question is most relevant for the Skilled migration program.
- What factors are most important in assisting migrants fulfil their potential? This question is relevant for all migrants.

However, these regressions are not 'causal' in the sense that they reflect the outcomes of the existing migration system and may not represent a structural relationship. Observed characteristics also reflect the decisions of individual migrants to pursue education or language acquisition which may be correlated with economic outcomes and with unobservable factors such as ambition or motivation. When interpreting these regressions, it is important to question whether the underlying relationships between variables is likely to change in response to a change in migration policy or whether the observed relationship is a result of 'reverse causality'.

Further research that uses an experimental research design based on policy variation would complement this research.[2]

Interpreting analysis from this paper within a broader economic framework

The analysis in this paper identifies the outcomes that are related to strong outcomes of Australian permanent migrants. While the analysis in this paper informs a broad range of migration-related policy questions, this analysis is best interpreted within the context of a broader economic framework. For instance, analysis conducted with the Australian Treasury's FIONA[3] and OLGA[4] models highlights the importance of demographics in determining long-run economic outcomes and, while the analysis in this paper 'controls' for demographic characteristics,[5] it does not capture the effect of younger migrants having more years in the Australian labour market. In addition, as the focus in this paper is on the outcomes of individual migrants, it will not be able to identify any

[2] See McKenzie and Yang (2010) for a review of experimental research designs used to study the economics of migration.

[3] For instance, in the FIONA model, variation in demography is twice as important as variation in income in explaining the differences in lifetime fiscal impact between migrants and non-migrants (Table 7 of Varela et al. 2021).

[4] Described in the 2021 Intergenerational Report. Commonwealth of Australia (2021).

[5] For instance, due to the accumulation of skills and experience, individuals will typically earn more as they age. This paper identifies how large this effect is so that the incomes of different migrants of different ages can be compared.

potential spill-over effects on the wages of Australian workers or the profitability of Australian businesses.[6]

While the analysis in this paper can provide guidance towards improving the targeting of the permanent migration system (i.e. a visa system that selects migrants with strong economic outcomes), the circumstances under which the migration system should try to target high-skilled migrants is a nuanced policy question that is highly situationally dependant. For instance, a policy designed to increase the economic outcomes within the Skilled migration stream is very different to one that increases the relative share of Skilled migration (at the expense of places within the Family or Humanitarian programs). Similarly, a policy that improved migrant incomes as a result of better accreditation and utilisation of foreign qualifications has a different policy implication to one that restricts visas to migrants working in high-paying occupations.

Data used in this project

The project has been conducted using datasets made available through the Multi-Agency Data Integration Project (MADIP). These datasets include:

- All Australia tax returns and payment summaries from 2010-11 to 2020-21
- Visa records (including from visa applications and settlement data)
    - Visa applications from 1990 onwards
    - Settlement data for permanent migrants from 2000 onwards
    - ‘Travellers data’ containing information on movements into or out of Australia.
- MADIP location data (derived from multiple sources)
- The 2016 Census

Further details about the MADIP dataset can be found in the appendix to this paper.

MADIP has several key advantages over alternative data sources (such as census data and migrant surveys) that have been used in previous research on this topic in Australia.

- MADIP has 100% coverage of temporary and permanent visa granted from 1990 onwards.
    - This allows results at a highly disaggregated level.
    - This allows long-run outcomes of migrants to be studied.
    - Administrative data in MADIP does not suffer from attenuation bias that occurs in survey data where some groups of migrants stop responding to follow-up surveys.
- The visa data in MADIP includes a mapping of visas that groups together visas with similar policy intent across years.
- MADIP allows for covariates to be merged from other datasets.
- MADIP allows for comparison with the non-migrant population.

The visa mapping is necessary to understand how the long-run outcomes of migrants vary by visa category. This variation (shown in Section 3) is substantial and would not be possible to analyse with most migration datasets.

MADIP has two notable disadvantages in the context of this study. The first is that earning zero income is not explicitly measured in administrative data. This study allocates an income of zero to individuals with no tax return or payment summary who have not been observed to leave the country (Appendix A contains further details of this calculation). This is not as problematic as it may

[6] The impact of migration on Australian productivity is the focus of Andrews et al. (2023) and OECD (2023).

first appear. The tax-free threshold creates a situation where individuals with income less than approximately $20,000 do not need to file a tax return. However, individuals with low income who do not file a tax return still appear in payment summary data, and therefore in our data, if they are paid a wage which is reported to the Australian Taxation Office at some point during the year.

The second drawback is that some covariates in this study vary less over time compared to the reality of an individual's experience. Census data are used for English language and education and the census is only available for one year in MADIP--2016. Thus, evolution in language ability and changes in educational levels are not captured in our approach. This is not necessarily problematic but English language ability and education need to be interpreted as being the level of language ability/education in 2016, not the contemporaneous level of language ability/education. In addition, there is evidence that individuals do not frequently update their personal details in administrative databases. For instance, Hathorne and Breunig (2022) find that reported occupation switching rates in administrative ATO data are significantly lower than the rate of reported occupation switching in HILDA. For this reason, this paper uses research designs that exploit variation in outcomes across migrants, rather than panel research designs such as fixed or random effects that are based on variation of income across years for a given individual.

Permanent migrants, temporary migrants and the non-migrant population

This paper separates tax records into three groups:

- Those which are linked to a permanent visa (the main focus of this study). Many in this group have previously held a temporary visa.
- Those which can be linked to a temporary visa but not to a permanent visa. This group are not included in this study. Future research on this group would be worth conducting.
- Those that are not linked to a temporary or permanent Australian visa[7] – referred to in this paper as the `non-migrant' population.

Defining non-migrants in this way is equivalent to the approach used in Mackey et al. (2020). However, some studies[8] base the comparison on the average person in Australia (which includes all three groups above). This choice of comparison group is important as temporary migrants earn less than either permanent migrants or non-migrants.[9] This choice will affect the assessment of the gap between `migrant' outcomes and comparison group outcomes.

Categorisation of permanent visas

The Australian migration system is complex. There are a large number of different visa categories that change over time. In order to make the analysis in this paper tractable, it is necessary to aggregate visas into categories.

This paper aggregates visas into five `Skilled' visa categories: Employer Sponsored, Skilled Independent, Business, Distinguished Talent and Regional. We consider two other visa categories: Family and Humanitarian visas. In addition to these seven categories, this paper distinguishes

[7] This will include individuals born in Australia as well as migrants that received a visa more than 20 years ago.

[8] Including Australian Treasury's FIONA model (Varela et al. 2021).

[9] For instance, the finding that permanent migrants earn less than the non-migrant population is in contrast to the headline finding in Varela et al. (2021). While there are several differences in the calculation, the difference is most likely explained by the difference in comparison group.

between the outcomes of primary applicants and secondary applicants (typically family members of the primary applicant). This categorisation is an aggregation of the categories used by the Department of Home Affairs Migration Program Statistics (Department of Home Affairs 2021). It is also designed to replicate the approach of Varela et al. (2021), which was the outcome of extensive consultation between the Australian Treasury and the Commonwealth Department of Home Affairs.

3. Aggregate economic outcomes in the years following migration

This section presents aggregated migrant outcomes in the years following the grant of a permanent visa. This analysis is based on two measures of migrant outcomes:

- Average standardised income, which is defined as income divided by the average income of non-migrants of the same age (this measure includes those that earn below $20,000).[10]
- The share of the population with an income below $20,000 (in 2019-20 dollars). This is a measure of non-participation designed to capture those with zero or minimal income (from labour or other sources).[11]

In all Figures, data are pooled across all years and analysis is restricted to individuals between the ages of 25 and 60, inclusive. Results are shown for the 20 years following a permanent visa grant.

These results show significant variation in the profile of earnings in the years following migration. Some groups of visa holders start at a high level of earnings (relative to `non-migrants') which reduces over time, while others start with low levels of relative earnings which increase over time.

One important insight of this initial analysis is that short-term outcomes of visa programs are likely to overstate the difference in earnings outcomes across different visa programs and differences relative to native-born Australians.

Results in this section should be interpreted as gross effects that do not distinguish between changes in migrant outcomes and changes in the underlying cohort. These cohort effects include migrants that arrive before the age of 25 and then enter into the age range covered by the Figure, outmigration, and differences in cohort quality across years. Cohort effects are investigated further in Appendix C using a fixed effects research design. That analysis suggests that cohort effects are unlikely to be a significant factor in explaining the results in this section.

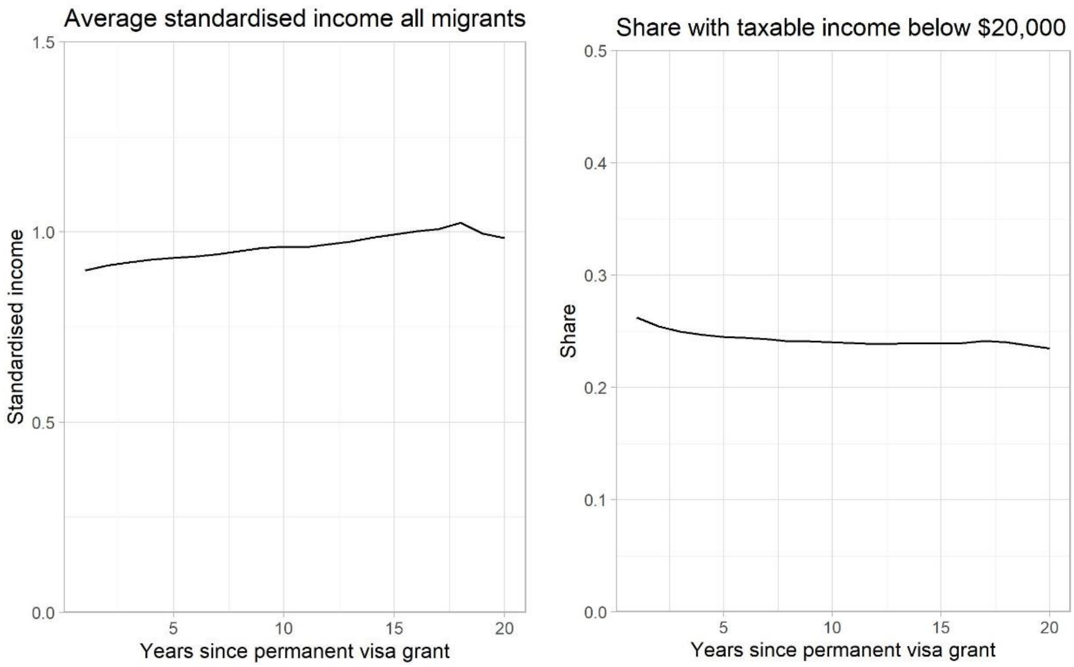
Aggregated results for all permanent migrants

Figure 3.1 presents economic outcomes aggregated across all permanent migrants. This shows that migrants earn slightly less, on average, than non-migrants in the years following permanent migration, before converging to the non-migrant level around 10 years after migration. There is also a slight reduction in the share of migrants with income below $20,000 in the years following permanent migration.

[10] For instance, if an individual has an income of $75,000 and the average income of all Australians of the same age is $50,000, then that individual has a standardised income of 1.5.

[11] In addition, the figure of $20,000 was chosen as it is near the tax-free threshold and therefore represents a group that pay minimal taxes. An individual with labour income of $20,000 would pay zero taxes given the combination of the tax-free threshold and the low-income tax offset.

Figure 3.1: Aggregate economic outcomes of Australian permanent migrants



Aggregated results by stream

Figure 3.2 shows standardised income by aggregate visa category split by whether an individual is a primary or secondary migrant. Consistent with other analysis conducted on the Australian migration system, the economic outcomes of the Skilled migration program are stronger than the Family program, which are in turn stronger than the Humanitarian program. There are also significant differences between primary and secondary migrants across all three programs.

Figure 3.2 Standardised income in the years following permanent migration, by Visa Stream

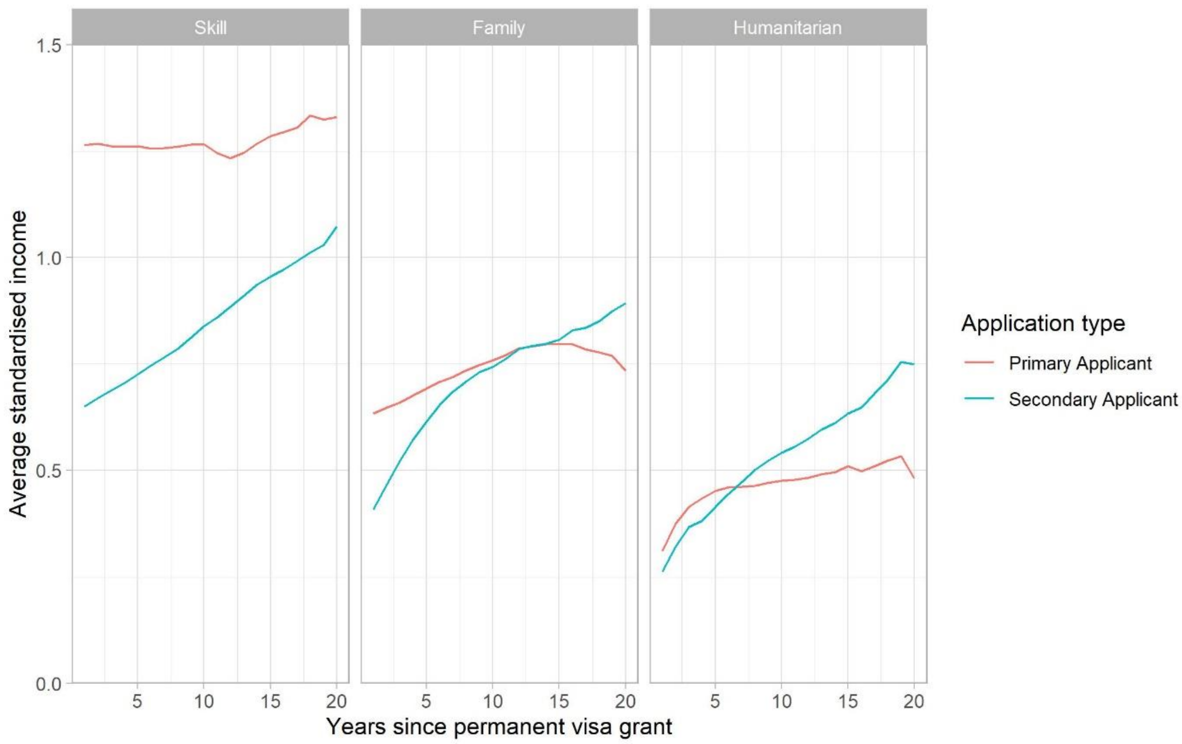
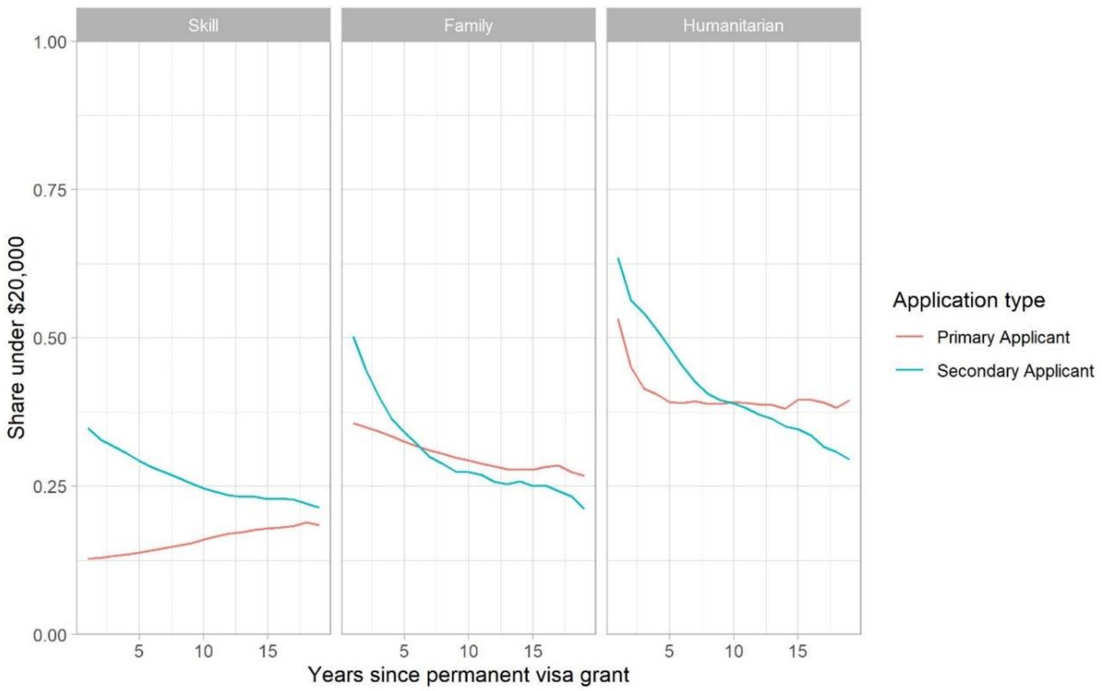


Figure 3.2 shows that the aggregate economic outcomes of migrants partially converge over time. For instance, secondary migrants initially have lower outcomes than primary migrants in all three streams but improve faster than primary migrants (surpassing outcomes for Family and Humanitarian Streams). While noting that this is partially driven by improved outcomes of migrants and an increasing share of the children of migrants (who enter as secondary migrants) in later years (explored further in Figures 3.22 and 3.23), it highlights the potential dangers of evaluating the economic outcomes of the migration system exclusively using short-term economic outcomes.

Figure 3.3 shows the share of migrants with an income below $20,000. This shows similar features to the standardised income Figure. For instance, migrants from the Skilled stream are more likely to have income over $20,000 than Family or Humanitarian visa holders. There is also a 'cross-over' point on both the Family and Humanitarian figures beyond which secondary migrants begin to have a higher percentage of people making more than $20,000 relative to primary migrants.

Figure 3.3: Share of migrants with taxable income less than $20,000 in the years following permanent migration, by Visa Stream



Results by Skilled migration program

Figure 3.4 and Figure 3.5 present standardised income and the share of income below $20,000 by permanent visa type. Consistent with other analysis of the Australian migration program, this shows that Employer Sponsored and Skilled Independent visa holder have the strongest economic outcomes among Australian permanent migrants.

These Figures also show some degree of convergence over time. For instance, Regional visa holders, Business visa holders and secondary Skilled visa holders have better outcomes after 10 years than in the first 5 years.

Figure 3.4 Standardised income in the years following permanent migration, by Skilled visa type

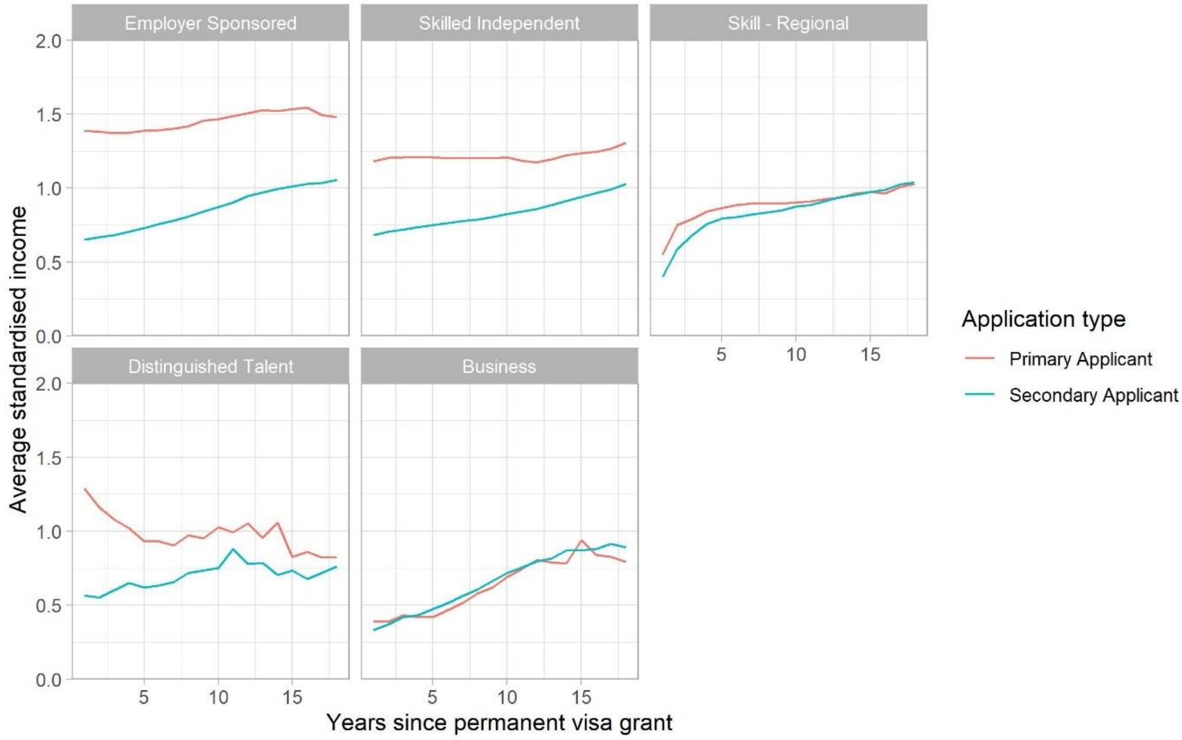
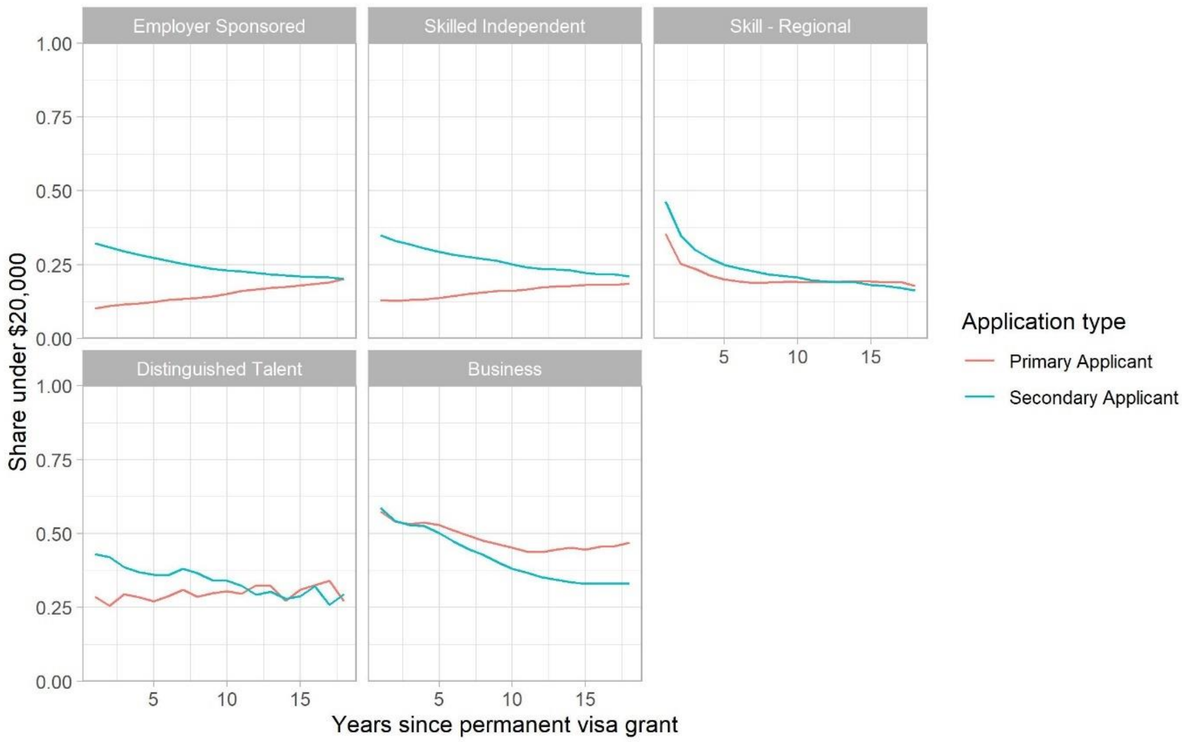


Figure 3.5 Share of migrants with taxable income less than $20,000 in the years following permanent migration, by Visa Stream



How do aggregated results vary across key migrant characteristics?

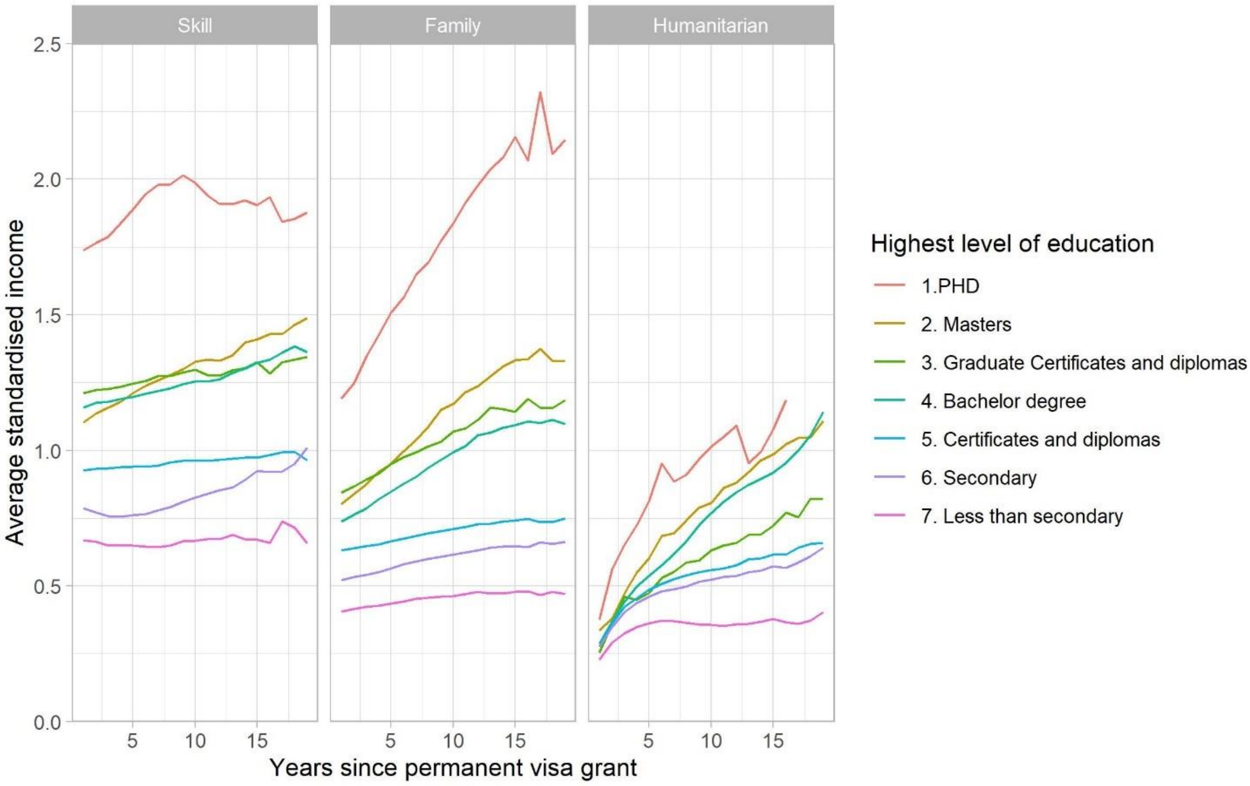
The following section shows how standardised income and the share of migrants with an income below $20,000 varies by education level, English language skill, geography and visa history. In these Figures, we do not control for any other factors. As a result, they are best interpreted alongside the regression results in the following section.[12]

Education

Figure 3.6 and Figure 3.7 present standardised income and the share of income below $20,000 by migrant stream and level of education. These Figures show that:

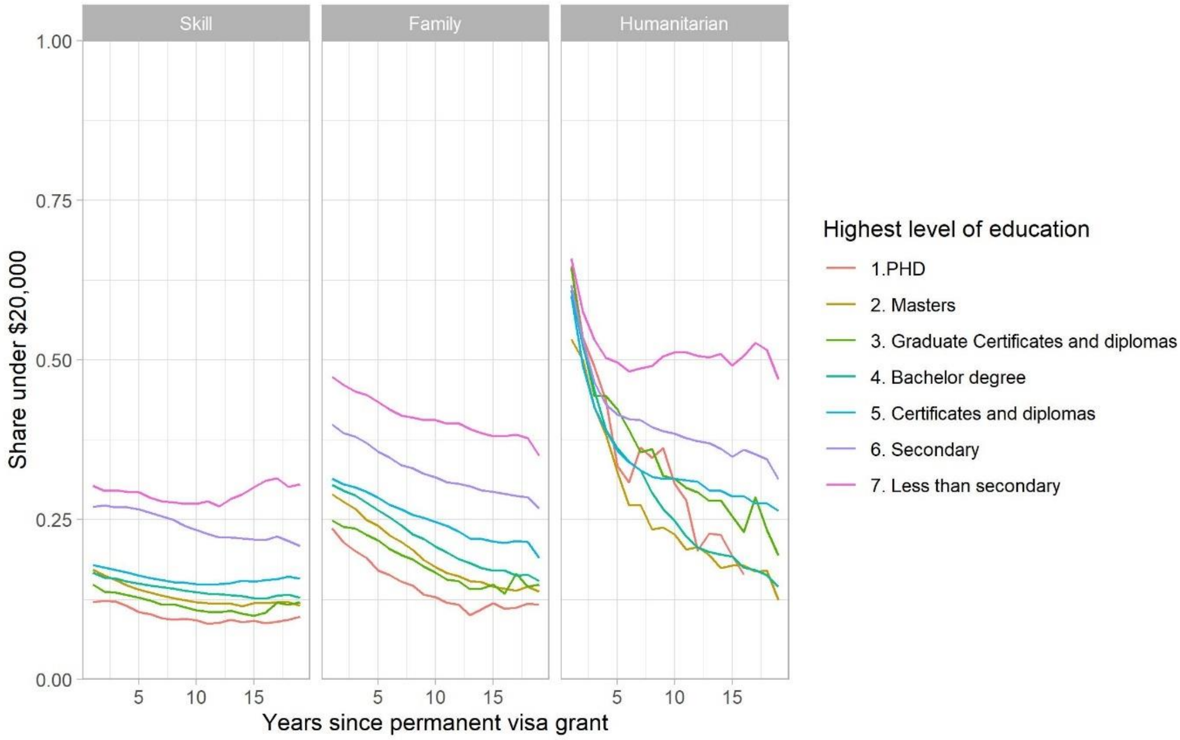
- Within each visa stream, migrants with higher levels of education have higher levels of income.
- The earnings gap between education levels widens over time.[13]
- The positive impact of higher levels of education is seen immediately for Skilled and Family stream migrants, but not for Humanitarian migrants.[14]

Figure 3.6 Standardised income in the years following permanent migration, by Education level



[12] In particular, these figures show how the 'settlement pattern' of migrants varies by migrant characteristics, which can inform specification and interpretation of regression equations.
[13] This is consistent with the Mincer regressions presented below comparing short- and long-run outcomes.
[14] This may reflect a period for these skills to be recognised or utilised. However, it may reflect migrants that are still studying for this qualification as education in this figure is based on the highest level of education from the 2016 census.

Figure 3.7 Share of migrants with taxable income less than $20,000 in the years following permanent migration, by Education level



English Language skills

Figure 3.8 and Figure 3.9 present standardised income and the share of income below $20,000 by migrant stream and level of education. These Figures show that:

- There is a large premium for language skills across all visa streams.
- There is a significant distinction between those that "speak English very well" and those that "speak English Only".
- Humanitarian migrants with higher levels of English Language skills have outcomes that improve over time, but we do not observe the same improvement for those with lower levels of English language skills.

Figure 3.8 Standardised income in the years following permanent migration, by English Language Skills

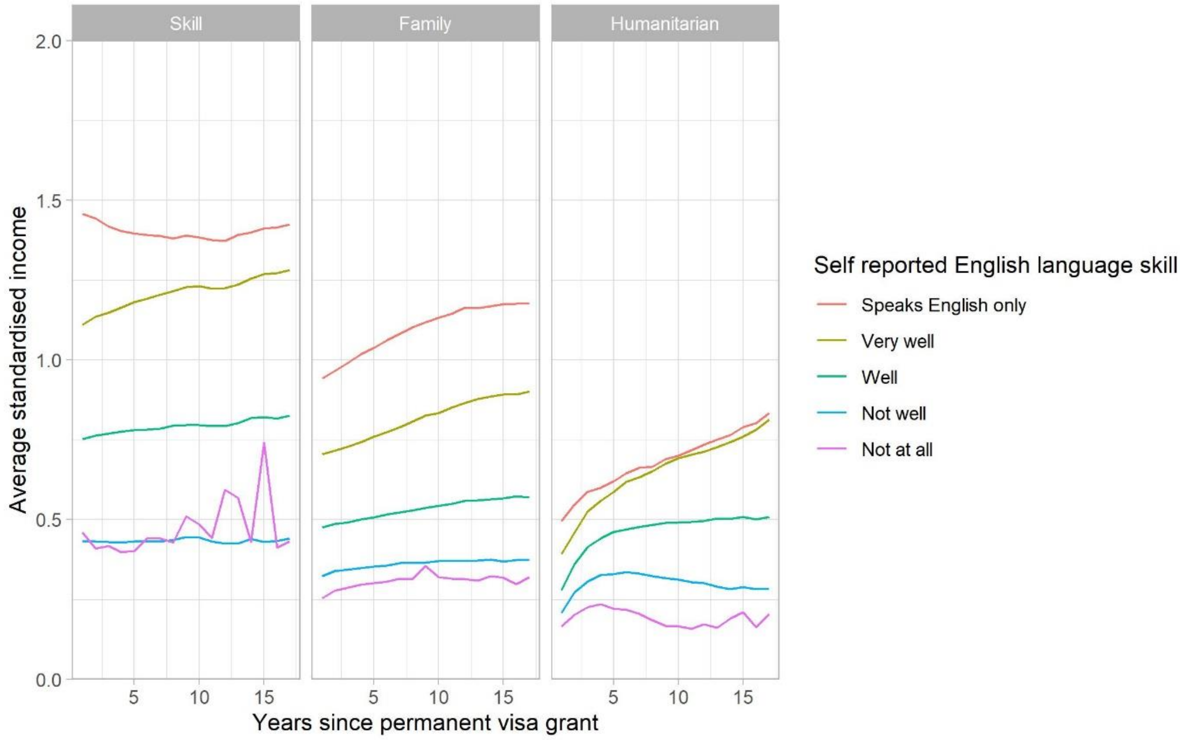
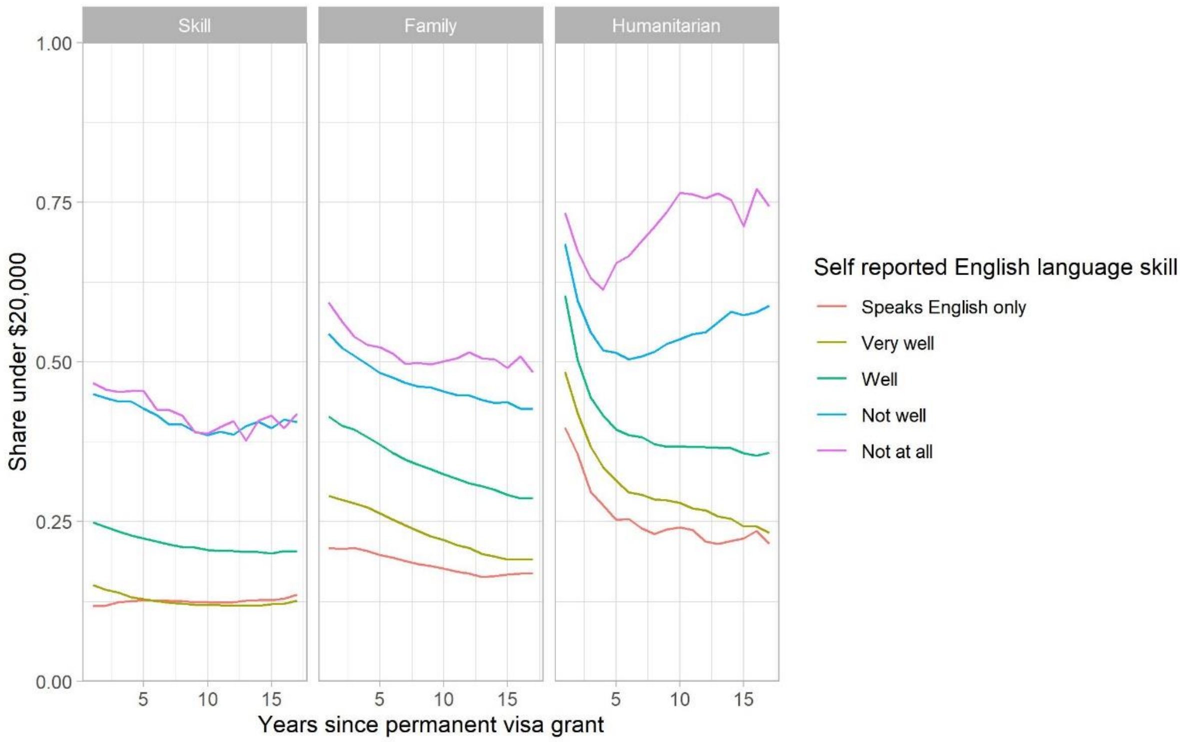


Figure 3.9 Share of migrants with taxable income less than $20,000 in the years following permanent migration, by English Language Skills



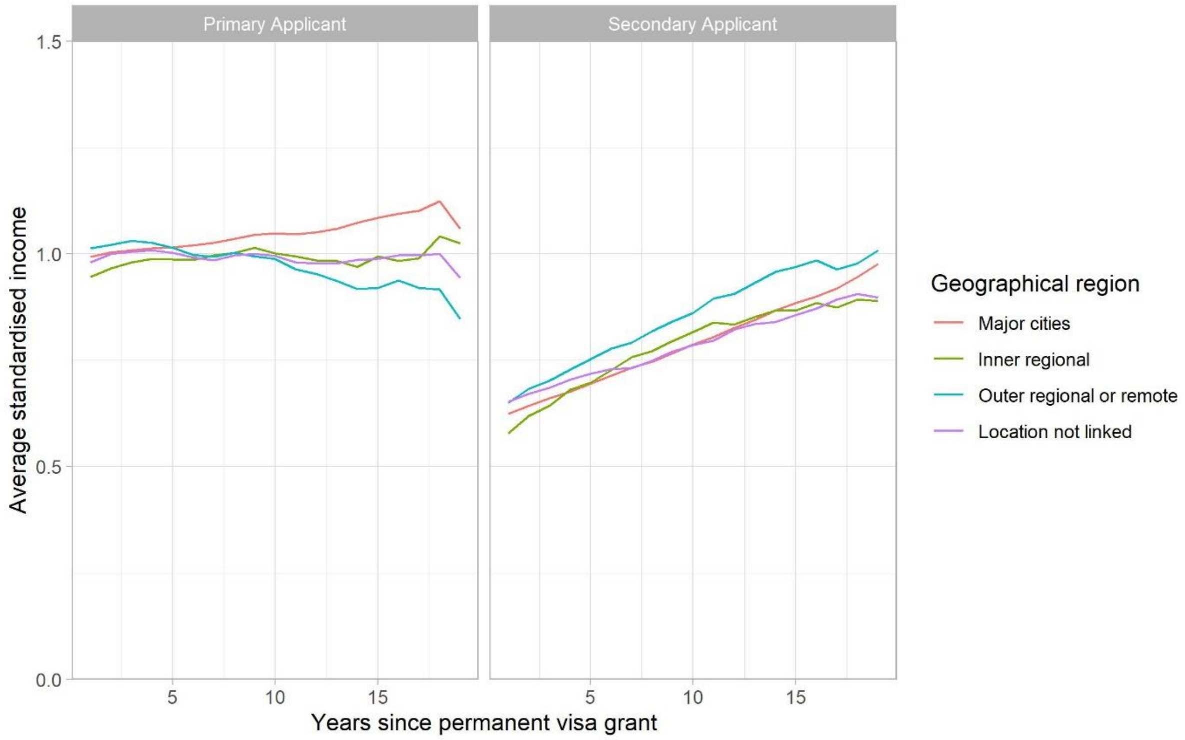
Geography

Figure 3.10 and Figure 3.11 present standardised income and the share of income below $20,000 by primary applicant status and level of education. These Figures show that:

- Migrants have similar economic outcomes and integration patterns in cities and regional areas.
- Secondary migrants are more likely to have income above $20,000 in regional areas than in cities.

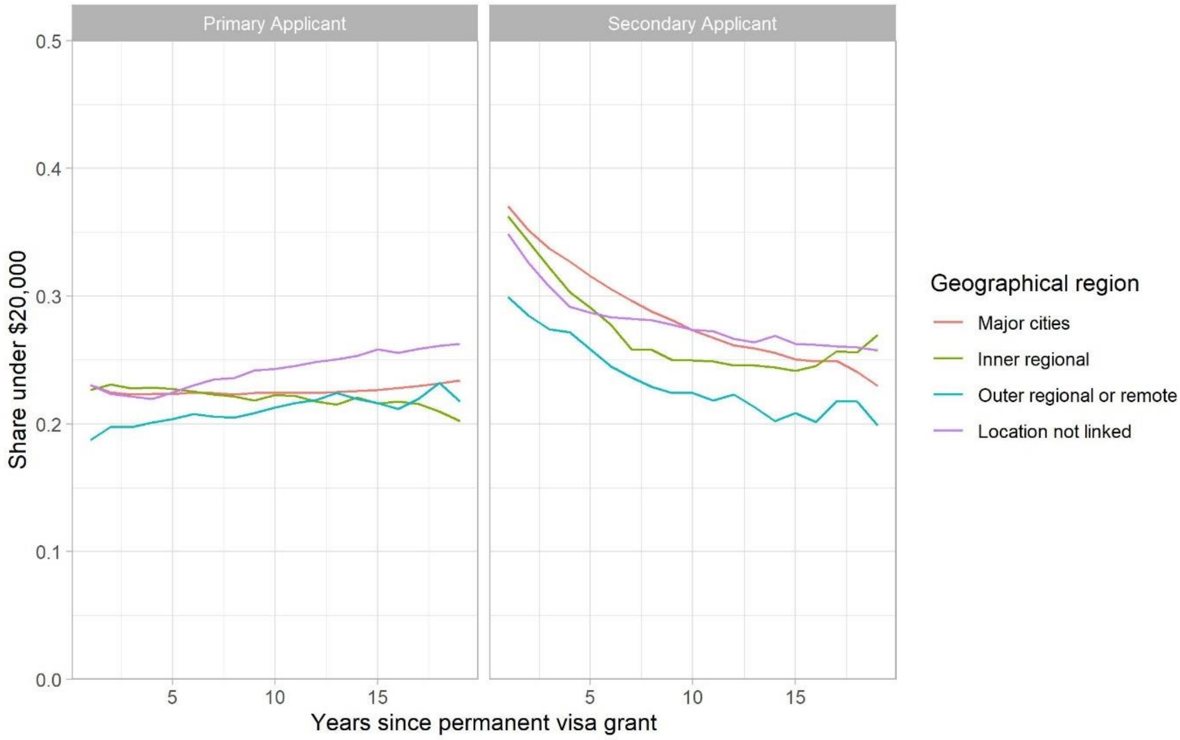
The patterns in these Figures show a relatively small relationship between migrant outcomes and geography. This is in contrast to the non-migrant population (outcomes for migrants and non-migrants are compared in Section 6) where economic outcomes are very different when comparing urban areas to regional and remote areas.

Figure 3.10: Standardised income in the years following permanent migration, by Region



The difference in outcomes between migrants and non-migrants in regional areas is likely to reflect a different set of factors underpinning the decision-making process for migrants and non-migrants moving between regional or metropolitan areas (rather than a labour market premium for migrants in regional areas). For instance, non-migrants will have a different profile of personal connections such as friends and families in different regions that will strongly influence movement decisions. Migrants may also be less tied to a specific regional location and therefore more willing to move to regional areas with employment opportunities. Further work to understand the location decision of migrants is beyond the scope of this project, but an exciting possibility for future work with the MADIP dataset.

Figure 3.11 Share of migrants with taxable income less than $20,000 in the years following permanent migration, by region



Visa History

Figure 3.12 and Figure 3.13 present standardised income and the share of income below $20,000 by temporary visa history. These Figures show that:

- Migrants that have previously held a Temporary Skilled visa have better economic outcomes than migrants with other visa histories.
- There is some degree of convergence amongst student visa holders.

Figure 3.12 Standardised income in the years following permanent migration, by visa history

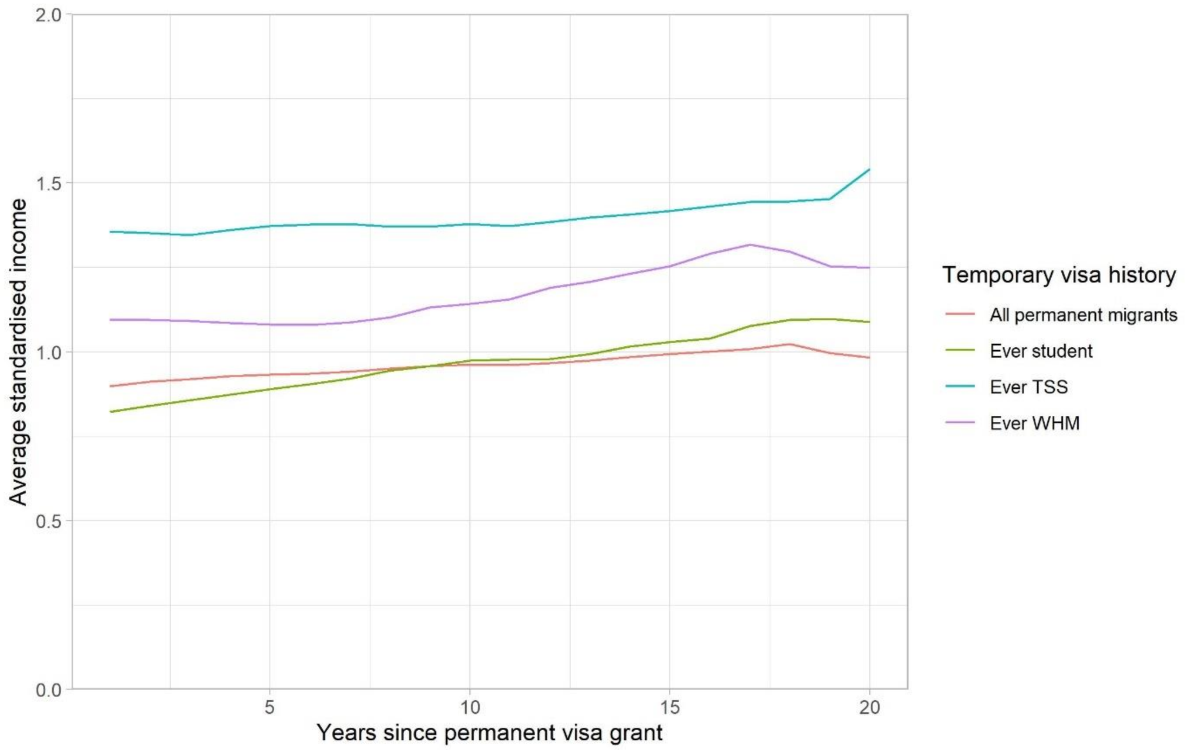
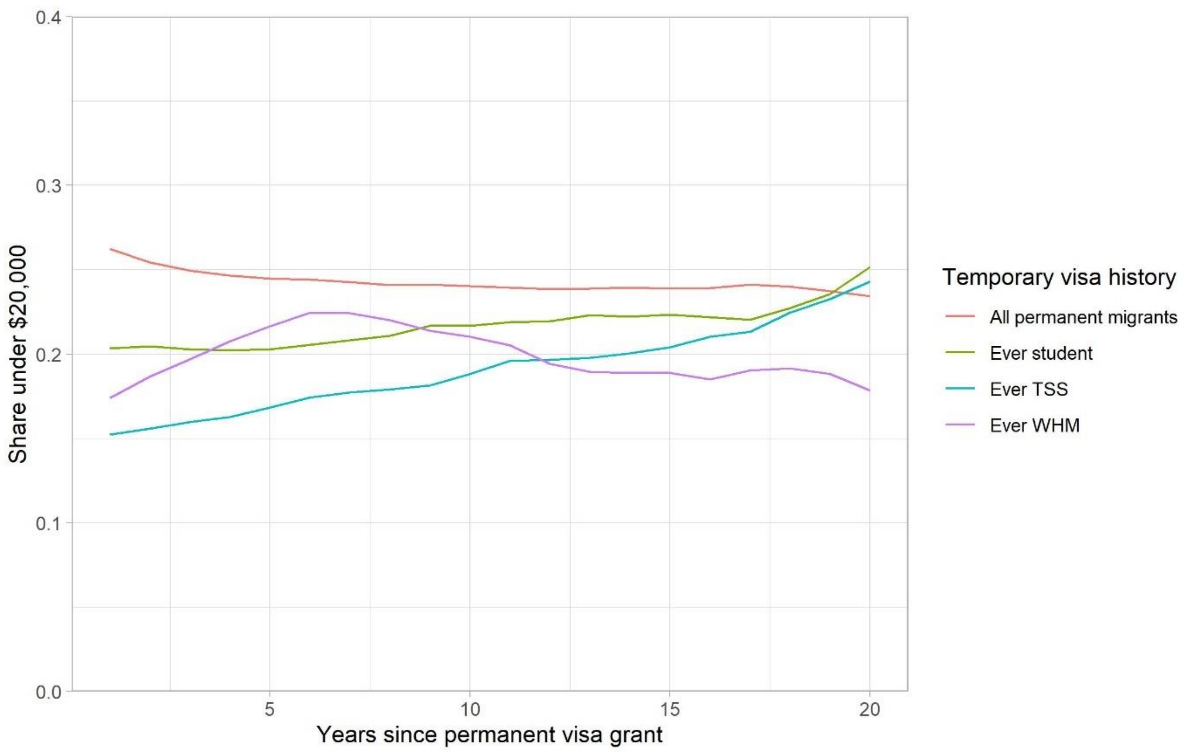


Figure 3.13 Share of migrants with taxable income less than $20,000 in the years following permanent migration, by visa history



Disaggregated student visas

Figure 3.14 and Figure 3.15 present standardised income and the share of income below $20,000 by temporary visa history, disaggregated by the highest level of education studied while on a student visa. These Figures show that there is a significant difference in outcomes between permanent migrants that have previously held different types of student visas. This difference is intuitive (migrants with higher levels of education earn more in the Australian labour market). However, this highlights the importance of disaggregating this group of migrants when thinking about the long-term economic impacts of student visas.

Figure 3.14 Standardised income in the years following permanent migration, by disaggregated student visa history

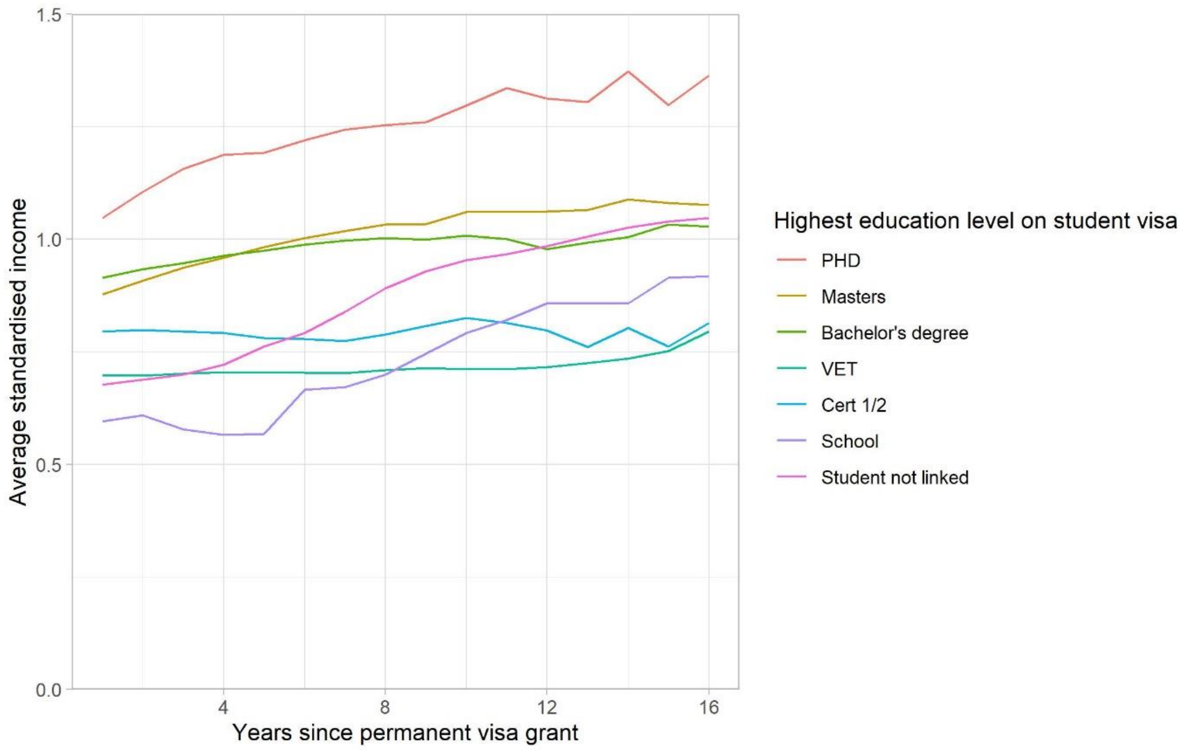


Figure restricted to 16 years due to small sample sizes.

Figure 3.15 Share of migrants with taxable income less than $20,000 in the years following permanent migration, by disaggregated student visa history

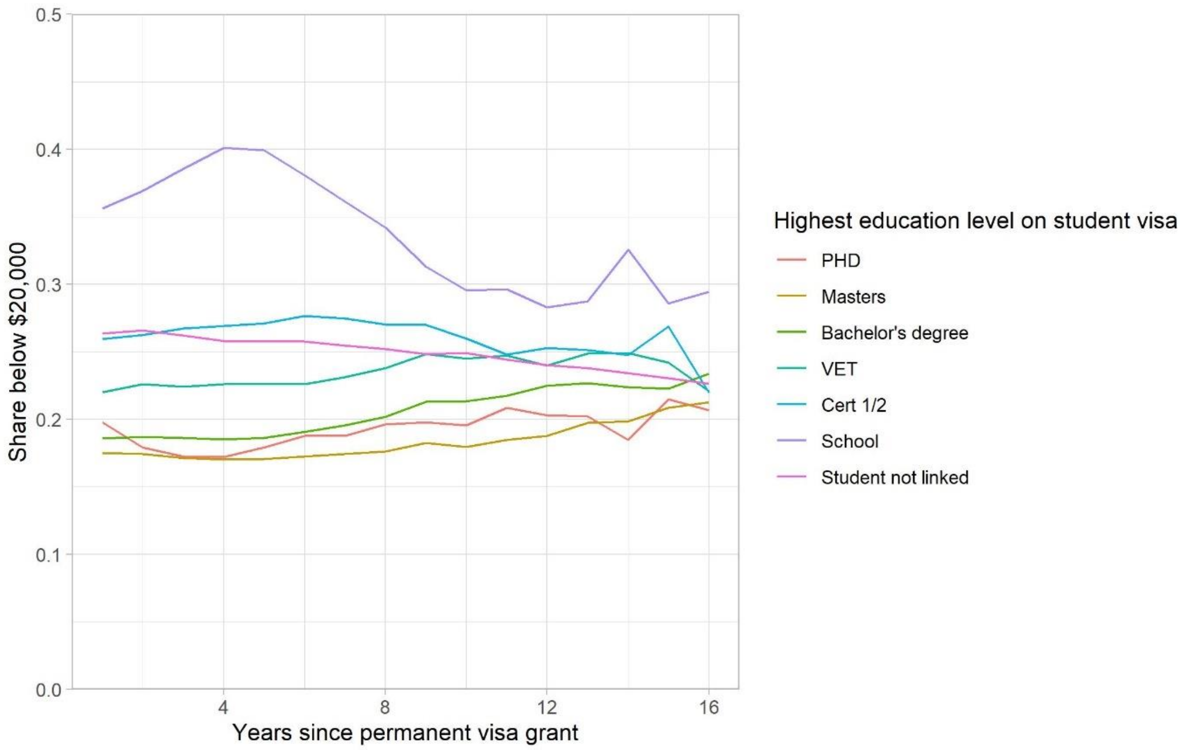


Figure restricted to 16 years due to small sample sizes.

Have migrant outcomes changed between cohorts?

One potential shortcoming of the descriptive analysis presented in this section is pooling together migrants that arrived in different years which potentially conflates integration patterns (how well a migrant *performs X* years after arrival) with cohort effects (have migrants granted visas in more recent years had better or worse outcomes than migrants from previous cohorts?).

To investigate whether there are strong cohort effects, Figures 3.16 and 3.17 present standardised income and the share of individuals with income below $20,000 grouped by migrant cohort. These Figures show that:

- Average earnings of skilled migrants may have fallen slightly relative to earlier cohorts.
    - This is not apparent in the patterns of labour market participation (as captured by the share of migrants with income below $20,000).
- Family and Humanitarian outcomes have improved in more recent cohorts.
    - This is seen in both standardised income and share of migrants with income below $20,000.

Figure 3.16 Standardised income in the years following permanent migration, by migrant cohort

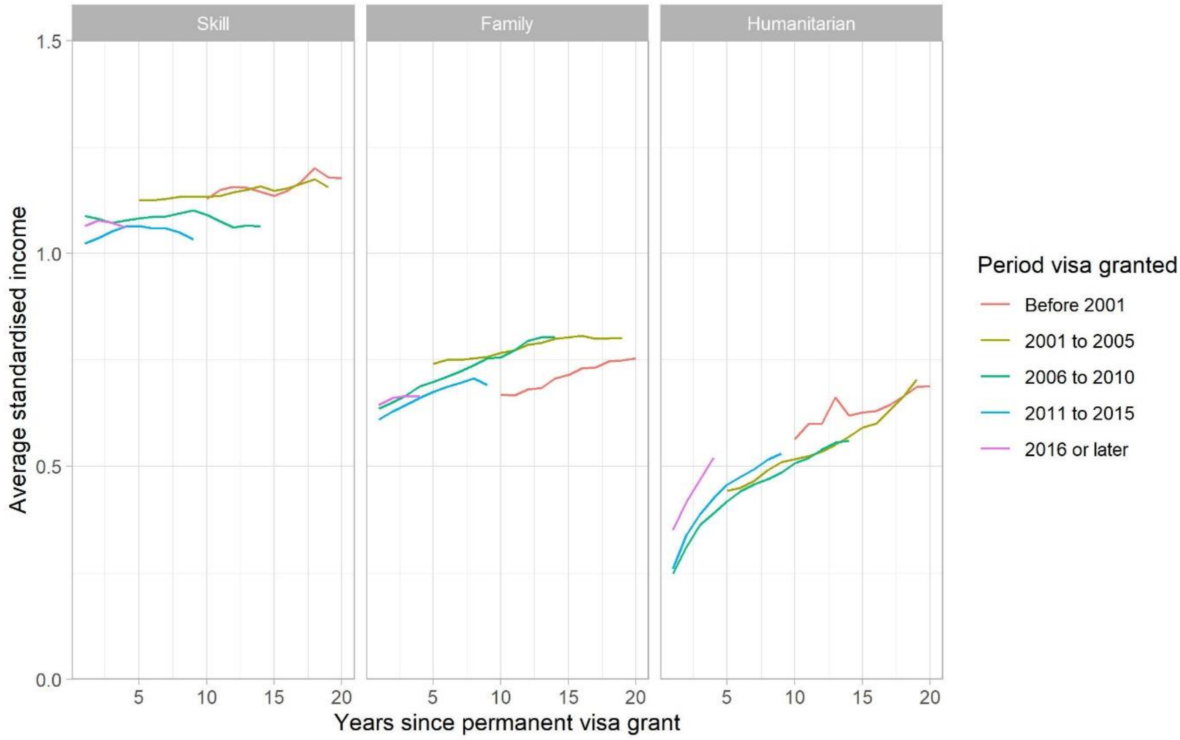
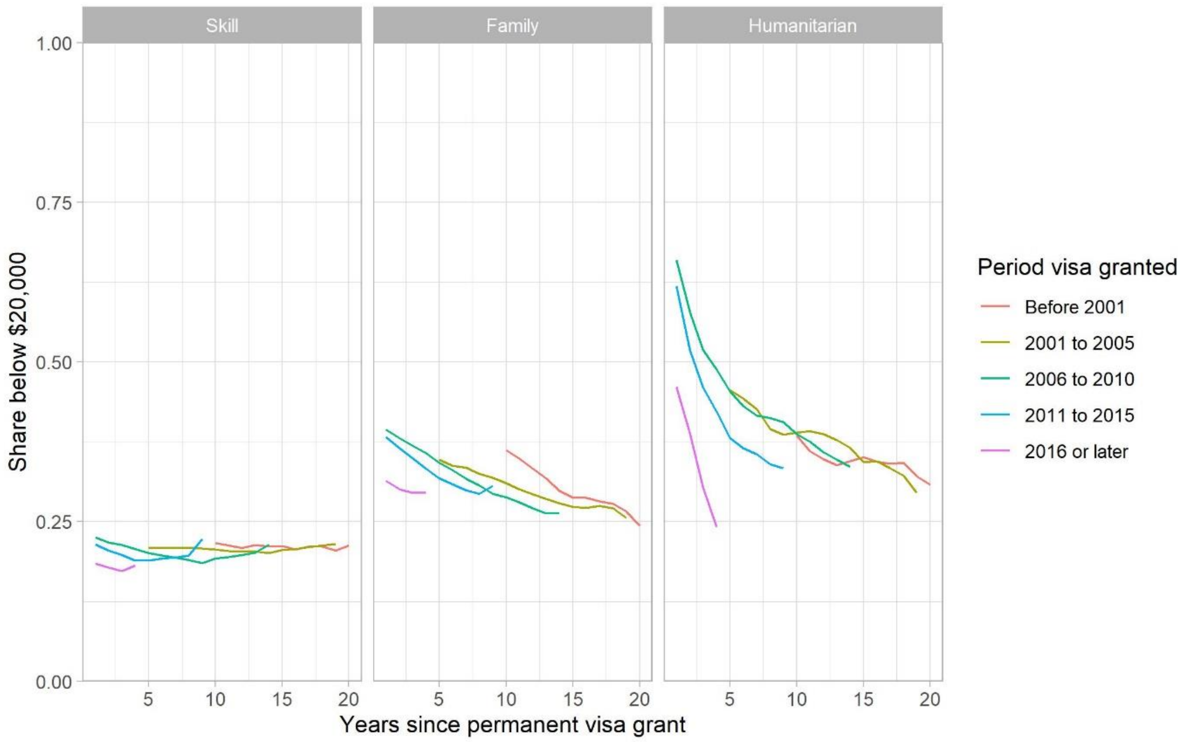


Figure 3.17 Share of migrants with taxable income less than $20,000 in the years following permanent migration, by migrant cohort



We conduct a similar exercise to check for cohort effects within categories of skilled migrants. These results are presented in Figures 3.18 and 3.19, and show:

- Employer Sponsored visa holders have lower incomes now than in previous cohorts.
- Recent regional visa holders have had worse outcomes than previous cohorts.

Figure 3.18 Standardised income in the years following permanent migration, by Skilled visa cohort

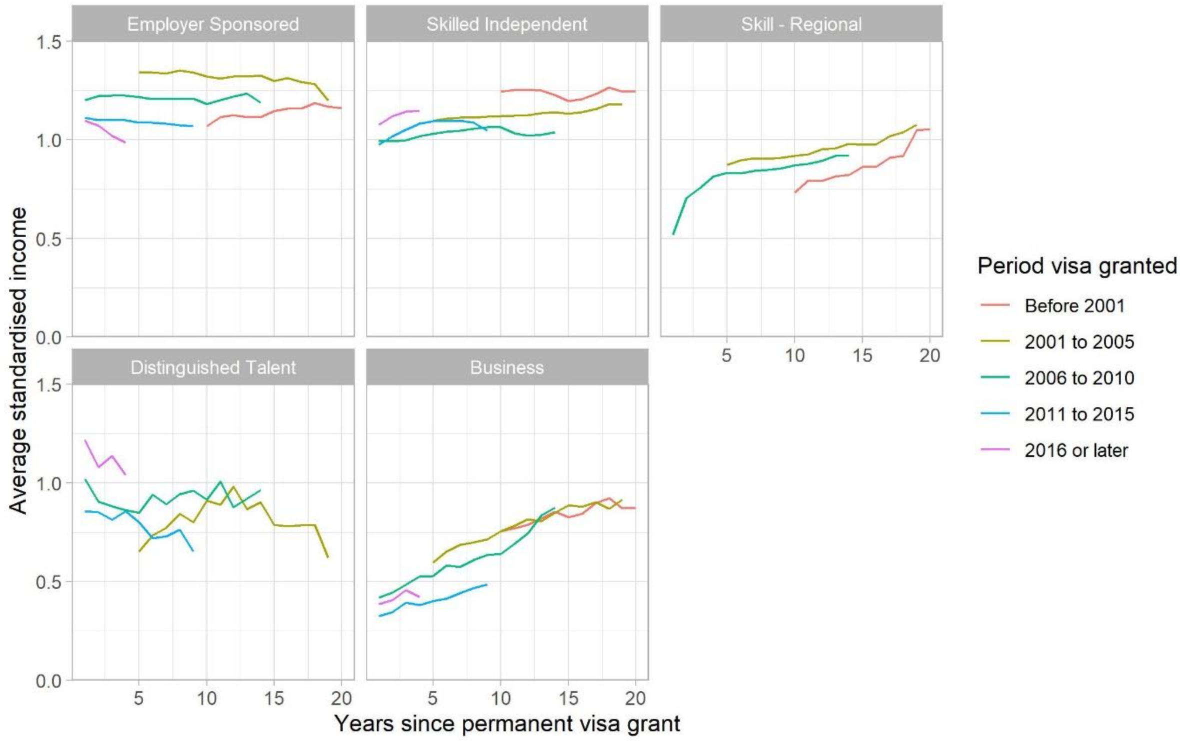


Figure excludes "Skill – Regional" after 2011 due to the small cohort size which may reflect an issue with the visa concordance used in this study.

Figure 3.19 Share of migrants with taxable income less than $20,000 in the years following permanent migration, by Skilled visa cohort

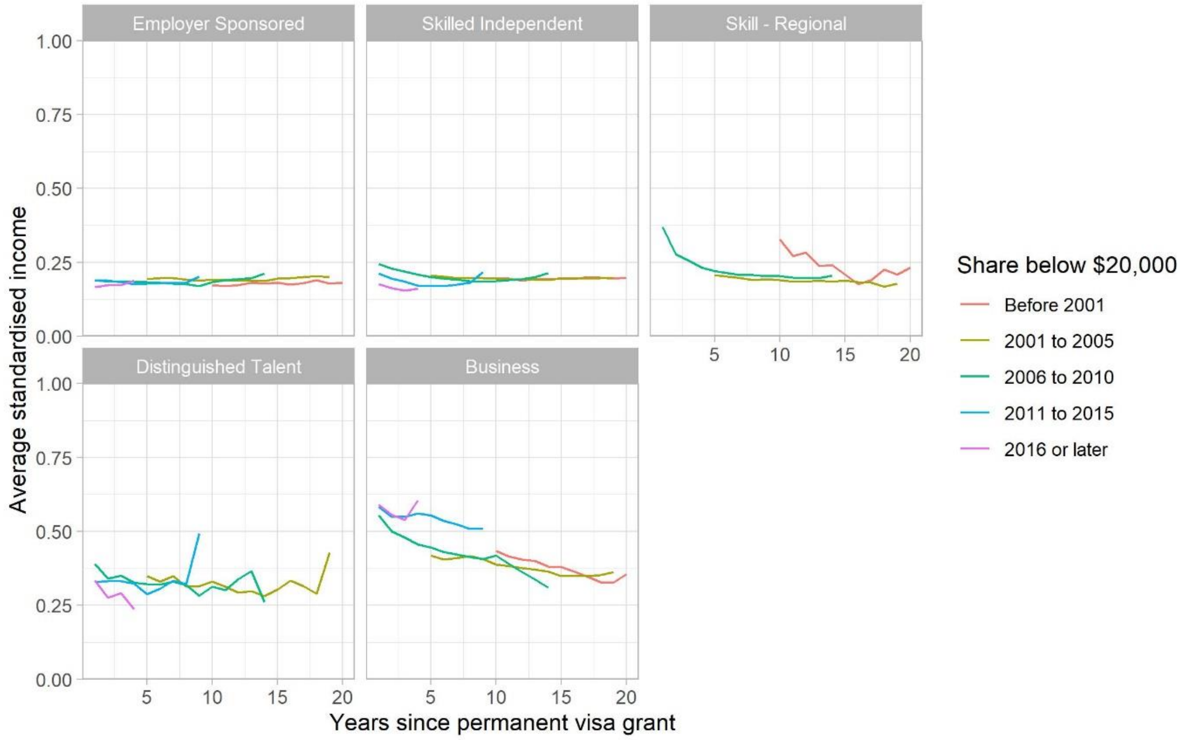


Figure excludes "Skill – Regional" after 2011 due to the small cohort size which may reflect an issue with the visa concordance used in this study.

How different are the outcomes of migrants that arrive as children?

Another potential question of interest is to split the outcomes of secondary migrants between partners and children.[15] Figures 3.20 and 3.21 show the standardised income and share of migrants below $20,000 for secondary migrants above and below the age of 18 (these Figures are analogous to Figures 3.2 and 3.3 above). Better outcomes of secondary migrants over time are driven by both improved individual outcomes and an increased share of migrants that arrive before the age of 18.

[15] The economic outcomes of migrants that arrive early in life is an important policy question in its own right. For instance, the calculations of lifetime fiscal impact in Varela et al. (2021) assume that migrants that arrive below the age of 18 have economic outcomes equivalent to the Australian population. It is important to note that estimates of fiscal impact produced in Varela et al. (2021) were sensitive to this assumption.

Figure 3.20 Standardised income of secondary migrants in the years following permanent migration, by age at arrival

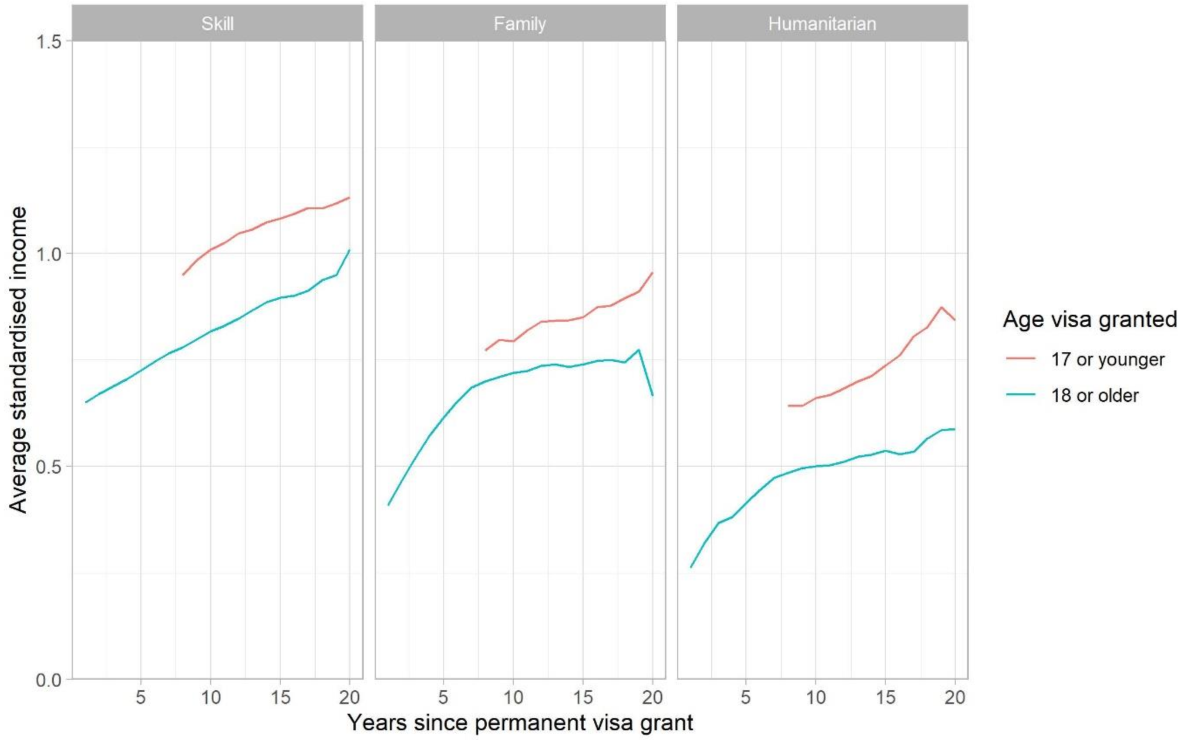
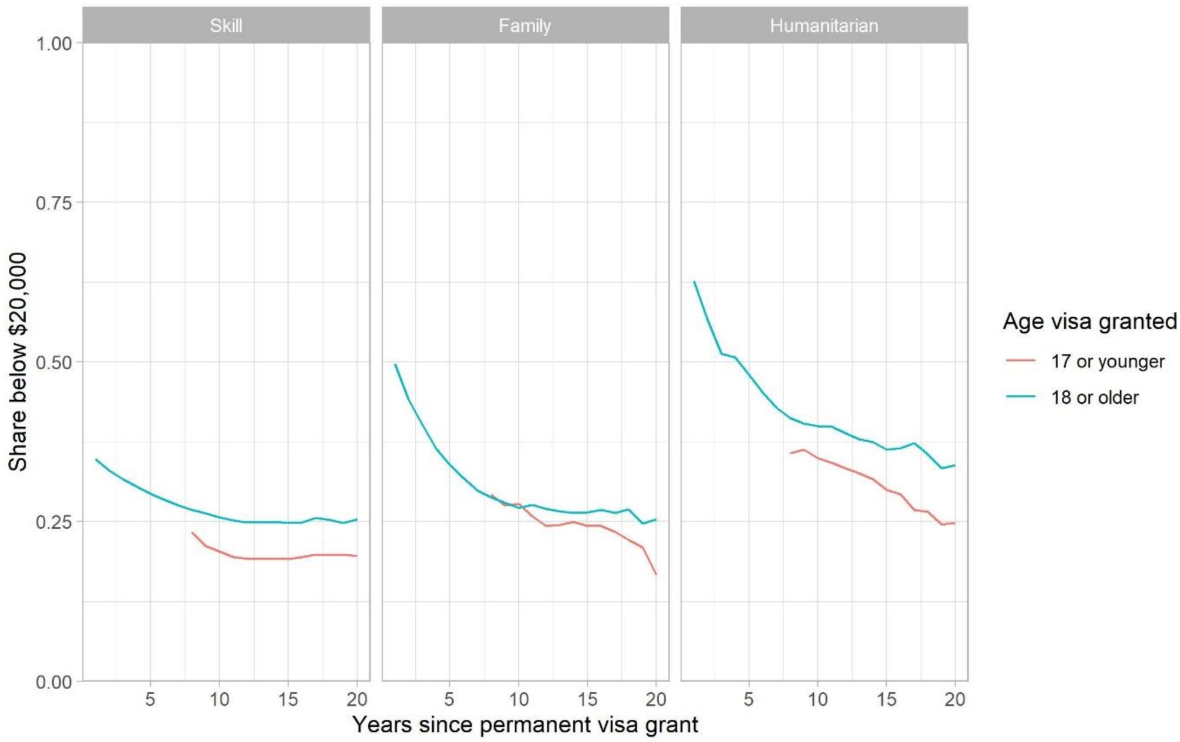


Figure 3.21 Share of secondary migrants with taxable income less than $20,000 in the years following permanent migration, by age at arrival



4. Mincer regressions of migrant outcomes

This section reports the results of a series of Mincer regressions (Mincer 1974). The estimated coefficients from this analysis can be interpreted as the returns to different migrant characteristics. Mincer regressions are among the most common models in empirical economics, with regular application in the economics of migration, labour and education economics.

This section of the paper contains three sets of Mincer regressions:

- A pooled regression that includes all permanent migrants pooled across all years.
  - This estimates the impact of different characteristics across all migrants averaged over all years.
- A regression that includes all migrants split between recently arrived migrants (those who received a permanent visa no more than five years earlier) and established migrants (those who received a permanent visa more than 5 years earlier).
  - This sheds light on which factors are more important in determining migrant outcomes in the short term versus the long term.
- A series of regressions that are estimated, in turn, on each group of permanent visa holders.
  - We use this to examine whether different factors are important for different groups of migrants by visa type.

In each case, separate regressions are estimated on the intensive and extensive margins. Intensive regressions are based on linear regressions on log earnings, while extensive margin regressions are logit models based on whether a migrant has taxable income above $20,000 (in 2020-21 dollars).[16]

Pooled regression of all permanent migrants

The first regression is a pooled Mincer regression that includes all permanent migrants aged 25-60 of the following form:

$$Y_{it} = \beta_0 + \beta_1 X_i^{vclass} + \beta_2 X_i^{vhistory} + \beta_3 X_{it}^{location} + \beta_4 X_{it}^{occupation} + \beta_5 X_i^{English} + \beta_6 X_i^{gender} + \beta_7 X_i^{partysize} + \beta_8 X_i^{education} + \beta_9 X_i^{countrybirth} + \beta_{10} X_i^{arrivalage} + \beta_{12} X_t^{year} + \beta_{12} X_{it}^{age} + \beta_{13} X_t^{ysa} + \varepsilon_{it} \qquad (1)$$

Where:

- Y_{it} is the economic outcome of interest (log income or an indicator for income over $20,000).
- X_i^{vclass} is a set of 13 dummy variables for visa class. These include the 5 categories of Skilled visa along with Family and Humanitarian visas, all divided by primary and secondary applicants. Family primary visa holder is the reference category in the results presented below.
- $X_i^{vhistory}$ is a set of 10 dummy variables for temporary visa history. These groups are defined as ever having a Skilled temporary visa, a working holiday visa or a student visa. Student visas are split into 8 categories reflecting the highest level of education on a visa. Categories are not mutually exclusive, meaning that someone can be in multiple visa history categories. No temporary visa is the reference category.

[16] Taxable income is the main outcome of interest in this paper. The decision to use a cut-off of $20,000 was made to exclude individuals with a small amount of taxable income from non-labour sources (such as interest), with $20,000 being near the effective tax-free threshold in the Australian tax system. See discussion in Section 2 above. Sensitivity tests using wages and salaries and a threshold of $1 were conducted and are included with other results in appendix C. The results are substantially the same.

- $X_{it}^{location}$ is a set of 12 dummy variables for geographic location (by state and whether or not an individual is in a metropolitan area). Tasmania, ACT and NT are each treated as a single region. Sydney is the reference category.
- $X_{it}^{occupation}$ is a set of 43 dummy variables for 2-digit ANZSCO occupation. Education is the reference category.
- $X_{i}^{English}$ is a set of 5 dummy variables for self-reported English language skills from the 2016 census.[17] Speaks English only is the reference category.
- X_{i}^{gender} is a dummy variable equal to one if an individual is male.
- $X_{i}^{education}$ is a set of 7 dummy variables for highest level of education. High school education is the reference category.
- $X_{i}^{countrybirth}$ is a set of 20 dummy variables for country of birth, based on the 2016 census. Born in India is the reference category.
- $X_{i}^{partysize}$ is a set of 4 dummy variables representing the number of secondary migrants who accompany the primary migrant. Having no secondary migrants is the reference category.
- $X_{i}^{arrivalage}$ is a set of 3 dummy variables reflecting the age at arrival of migrants. Receiving a permanent visa aged 25-32 is the reference category.
- X_{t}^{year} is a set of 10 dummy variables for financial year. 2010-11 is the reference category.
- X_{it}^{age} is a set of 35 dummy variables for single year of age. 25 years old is the reference category.
- X_{t}^{ysa} is a set of 19 dummy variables for years since arrival. 1 year since arrival is the reference category. We exclude the year in which a migrant is granted a visa.

Note that when regressions include dummy variables for a set of categories, one category needs to be used as the "reference" category and the dummy variable for that category is omitted in the regression specification. When interpreted correctly, results are insensitive to the choice of reference/omitted category. Included indicators can be interpreted as the difference between that category and the reference category.

Data on English Language Skills, Education and Country of birth are available in MADIP at a highly disaggregated level and were aggregated for this project. Details of this aggregation process are included in Appendix A.

The average earnings model (intensive margin) is estimated using a log-linear form with log (taxable income) as the dependant variable. For extensive margin models, the functional form of the model is a logit model (conditional on income greater than $20,000). Throughout the paper, logit model output is reported as odds ratios. Results from this model are presented in Table 4.1 (with full results for coefficient estimates and standard errors provided in Appendix C).

Table 4.1: Pooled Mincer regressions

	Intensive margin	Extensive margin
male	0.370***	1.970***

[17] This paper uses census information for English Language and Education rather than information from the visa application process. While the administrative visa data are richer, the type of information collected varies across visa categories and years. If missing information can't be fully identified, it could introduce bias into our analysis. An additional advantage to using census data is that these variables are comparable across migrants and non-migrants.

PHD	0.411***	2.031***
Masters	0.166***	1.565***
Graduate certificate of diploma	0.123***	1.862***
Undergraduate	0.110***	1.481***
Cert 3or4	-0.00755***	1.313***
Less than year 12	-0.00625**	0.824***
English very well	-0.0606***	0.979***
English well	-0.187***	0.647***
English not well	-0.282***	0.437***
English not at all	-0.307***	0.391***
NSW regional or remote	-0.0925***	1.070***
Victoria city	-0.0533***	1.004
Victoria regional or remote	-0.0912***	1.089***
Queensland city	-0.0657***	1.071***
Queensland regional or remote	-0.0611***	1.109***
WA city	0.00638***	1.078***
WA regional or remote	0.00883*	1.286***
SA city	-0.106***	1.004
SA regional or remote	-0.0691***	1.223***
Tasmania	-0.154***	1.101***
ACT	0.0292***	1.348***
NT	0.104***	1.775***
Party size of 2	0.00385**	1.008
Party size of 3	-0.0235***	0.917***
Party size above 3	0.00683***	0.935***
Financial year dummies	✓	✓
Year of age dummies	✓	✓
Years since migration dummies	✓	✓
Occupation dummies	✓	
Visa stream dummies	✓	✓
Visa history dummies	✓	✓
Party size	✓	✓
Grant under 18	✓	✓
Country of birth dummies	✓	✓
Observations	12,158,658	14,084,627

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%
✓ indicates that controls for these characteristics are included in the model.
Regressions are run on all permanent migrants aged 25-60 in all years.

Results from the intensive model can be interpreted as the percentage income premium holding other factors equal. For instance, all else equal, migrants with a PhD earn 41.1 per cent more than a migrant with high school level education. Results from the extensive model are reported as odds ratios (all else equal, a migrant with a PhD has double the "odds"[18] of a positive income than an individual with a high school education).

[18] Odds are defined as the probability of success (having income above $20,000) divided by the probability of failure, and the odds ratio is the estimated impact on these odds. For instance, if the underlying probability of having income over $20,000 is 0.8, then the odds are 0.8/0.2 = 4. An odds ratio coefficient of 2 would mean that the odds for an individual with that characteristic become 8:1, or 87.5%. An odds ratio equal to one has no estimated impact on outcomes while an odds ratio below one is associated with a lower likelihood of the outcome.

Table 4.1 shows that migrants with higher levels of education and better English language skills have higher incomes and are more likely to be employed. There is also a premium associated with being male.

The returns to geography are mixed, with most regions having a negative coefficient in the intensive margin equation (all else equal, migrants earn more in Sydney than in other areas) but also have a coefficient greater than one in the extensive margin (all else equal, migrants are more likely to have a positive income in regional areas than in cities). The returns to geographical location are different to those of the non-migrant population (discussed below in section 6).

Party size has a relatively small impact on income and a negative impact on the likelihood of a positive income (noting that this regression also has dummy variables for primary and secondary migrants by stream). This means that assuming that the outcomes of primary and secondary migrants are independent when conducting aggregate analysis of the migration system may not be innocuous.[19]

The returns to visa history, visa stream and country of birth are all included in the model and play an important role in explaining migrant economic outcomes. The coefficients from these variables are included in Appendix C. However, given that these variables are designed to select on migrant characteristics, the returns to visa history, visa stream and country of birth are best interpreted using the Oaxaca-Blinder framework in Section 5 below.

Short-term versus long-term effects

Another potential question of interest that can be addressed using this research design is to separate the factors that predict short-term migrant outcomes from those that predict long-term migrant outcomes. This distinction is important because it can indicate factors that may assist with migrant settlement.

Another reason that it is important to compare short-term and long-term determinants of migrant outcomes is that the economic impacts of migration play out over a long timeframe. Unfortunately, policy analysis is often based on shorter-term outcomes. Therefore, in addition to providing a richer understanding of the migrant settlement process, this analysis can distinguish between factors that may give misleading outcomes when analysis is conducted on shorter-term datasets.

The model used in this analysis is described in Equation (1). For the short-term, we restrict the data on migrant outcomes to the first five years after receiving a permanent visa while the long-term regression is estimated on all years beyond the fifth year.

Selected outcomes of the regression are presented in Table 4.2, with all coefficients and full results included in Appendix C.[20]

Table 4.2: Mincer regressions split by short-term and long-term outcomes

[19] For instance, in Varela et al. (2021) the lifetime fiscal impacts of primary and secondary migrants are calculated independently, and the combined impact taken as the sum of primary and secondary impacts. However, the results in this paper show that secondary migrants are less likely to have a positive income when arriving in larger groups.

[20] This spreadsheet also contains a version of this model run on migrants in the short and long term that excludes migrant-specific variables (such as visa category and temporary visa history) that can be compared with regression results for the non-migrant population in Section 6.

	Short-term intensive	long-term intensive	Short-term extensive	Long-term extensive
Male	0.382***	0.358***	2.285***	1.769***
PHD	0.328***	0.471***	1.710***	2.377***
Masters	0.106***	0.218***	1.316***	1.876***
Graduate certificate or diploma	0.106***	0.141***	1.580***	2.128***
Undergraduate	0.0675***	0.143***	1.319***	1.668***
Cert 3or4	-0.0174***	0.00146	1.239***	1.365***
Less than year 12	0.0107***	-0.0159***	0.896***	0.770***
English very well	-0.0607***	-0.0593***	0.994	0.998
English well	-0.179***	-0.194***	0.656***	0.682***
English not well	-0.272***	-0.292***	0.482***	0.443***
English not at all	-0.307***	-0.316***	0.469***	0.371***
NSW regional or remote	-0.0878***	-0.0971***	1.069***	1.097***
Victoria city	-0.0482***	-0.0569***	1.039***	0.991
Victoria regional or remote	-0.0838***	-0.0979***	1.138***	1.080***
Queensland city	-0.0619***	-0.0691***	1.071***	1.090***
Queensland regional or remote	-0.0457***	-0.0759***	1.129***	1.136***
WA city	0.0214***	-0.00724***	1.112***	1.082***
WA regional or remote	0.0247***	-0.00784	1.359***	1.284***
SA city	-0.108***	-0.100***	1.008	1.035***
SA regional or remote	-0.0405***	-0.0962***	1.299***	1.217***
Tasmania	-0.141***	-0.164***	1.126***	1.111***
ACT	0.0065	0.0475***	1.327***	1.396***
NT	0.110***	0.0927***	1.848***	1.751***
Financial year dummies	✓	✓	✓	✓
Year of age dummies	✓	✓	✓	✓
Years since migration dummies	✓	✓	✓	✓
Occupation dummies	✓	✓		
Visa stream dummies	✓	✓	✓	✓
Visa history dummies	✓	✓	✓	✓
Party size	✓	✓	✓	✓
Grant under 18	✓	✓	✓	✓
Country of birth dummies	✓	✓	✓	✓
Observations	4,539,545	5,327,662	6,317,043	7,767,584

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%

✓ indicates that controls for these characteristics are included in the model.

Regressions are run on all permanent migrants aged 25-60. Short-term regression includes the first 5 years after the grant of a permanent visa, while the long-term regression captures all years from 6 onwards.

These regressions show that:

- The return to education is larger in the long term.
- The return to English language skills is persistent (there is a similar effect in the short and long term)[21]

[21] Data on English language skills are taken from the census and do not vary over time for each individual. Therefore, the interpretation to this regression is that having a lower level of English language skills at a point in time has a long-term negative impact on earnings and participation (even noting that migrants' English skills typically improve as time passes after migration).

- Visa stream and temporary visa history become less important over time.[22] This is consistent with the descriptive Figures in section 3 which shows that the difference in standardised income between migrant groups partially converges over time.
- Country of birth effects become less important, on average, but with significant variation across countries.
- Gender effects on participation are stronger in the short term than in the long term. This may be driven by a fertility effect around the period of migration.

Mincer regressions by visa type

Another potential question of policy interest is to understand whether the returns to migrant characteristics vary across visa categories. For instance, in designing the points test for the Skilled Independent Visa, it is informative to know whether the determinants of outcomes for the Employer Sponsored migrants vary from other migrant groups. Similarly, Mincer regressions restricted to secondary migrants may uncover factors that are important in improving the outcomes of this group of migrants.

The model used in this analysis is the same as the pooled regression model in Equation (1). These regressions are estimated for 14 groups of migrants on the intensive and extensive margin (for a total of 28 regressions). Table 4.3 presents selected results from 6 of these regressions; complete results from all regressions are included in Appendix C.

Table 4.3: Mincer regressions by visa program

	Emp. Sp. primary	Emp Sp. Secondary	Skill Ind. Primary	Skill Ind. Secondary	Family Primary	Humanitarian Primary
Male	0.338***	0.380***	0.298***	0.441***	0.432***	0.200***
NSW regional or remote	-0.113***	-0.0789***	-0.0468***	-0.0393**	-0.119***	-0.0278
Victoria city	-0.0891***	-0.0678***	-0.0657***	-0.0383***	-0.0223***	0.0044
Victoria regional or remote	-0.112***	-0.0945***	-0.0576***	-0.0492***	-0.104***	-0.0336*
Queensland city	-0.113***	-0.0895***	-0.0647***	-0.0606***	-0.0390***	0.0486***
Queensland regional or remote	-0.0789***	-0.0549***	-0.00506	-0.0584***	-0.0850***	0.00779
WA city	-0.0189***	-0.0361***	0.0320***	-0.00455	0.0183***	0.0932***
WA regional or remote	-0.0045	-0.0278***	0.0861***	0.0178	-0.0144	0.142***
SA city	-0.160***	-0.121***	-0.125***	-0.101***	-0.0660***	-0.0115
SA regional or remote	-0.0759***	-0.0551***	-0.0384*	-0.0745**	-0.0906***	-0.0217
Tasmania	-0.152***	-0.164***	-0.132***	-0.143***	-0.169***	-0.136***
ACT	-0.00843	0.0297***	0.0576***	0.0307***	0.0274***	0.0221
NT	0.0127	0.133***	0.122***	0.130***	0.145***	0.276***
PHD	0.463***	0.465***	0.289***	0.407***	0.419***	0.217**
Masters	0.228***	0.142***	0.136***	0.127***	0.170***	0.109***
Graduate certificate or diploma	0.105***	0.138***	0.0798***	0.104***	0.134***	0.0692*
Undergraduate	0.146***	0.0911***	0.0918***	0.0898***	0.0920***	0.0390***
Cert 3or4	-0.00448	-0.00487	-0.0440***	-0.000629	-0.00765**	0.00906
Less than year 12	-0.0763***	-0.0000629	-0.0601***	-0.0383***	-0.0212***	-0.00385
English very well	-0.0397***	-0.0491***	-0.0642***	-0.0434***	-0.0578***	-0.0275
English well	-0.172***	-0.174***	-0.195***	-0.202***	-0.165***	-0.0714***
English not well	-0.361***	-0.242***	-0.365***	-0.370***	-0.261***	-0.109***

[22] Visa effects and country of birth effects are not reported in Table 4.2 but included in Appendix C.

English not at all	-0.423***	-0.182***	-0.189***	-0.332***	-0.296***	-0.194***
Financial year dummies	✓	✓	✓	✓	✓	✓
Year of age dummies	✓	✓	✓	✓	✓	✓
Years since migration dummies	✓	✓	✓	✓	✓	✓
Occupation dummies	✓	✓	✓	✓	✓	✓
Visa stream dummies	✓	✓	✓	✓	✓	✓
Visa history dummies	✓	✓	✓	✓	✓	✓
Party size	✓	✓	✓	✓	✓	✓
Grant under 18	✓	✓	✓	✓	✓	✓
Country of birth dummies	✓	✓	✓	✓	✓	✓
Observations	2,115,809	1,251,384	2,907,083	1,236,036	3,173,799	246,614

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%
✓ indicates that controls for these characteristics are included in the model.
Regressions are run on all permanent migrants aged 25-60. Visa programs are defined using the concordance described in Appendix A.

Table 4.3 shows that in aggregate, the determinants of income are similar across visa streams.

- The relationship between outcomes and geography is different for Humanitarian migrants compared to other groups, with comparably better earnings outside of cities (this is also reflected in the extensive margin results in Appendix C).
- Primary Skilled Independent visa holders have a lower return to education than other migrant groups.
- The returns to different visa histories vary significantly by type of visa.
    - Having previously held a temporary Skilled visa has large positive returns when transitioning to a permanent visa, but a much smaller effect for secondary permanent migrants.
    - Having previously held a working holiday maker visa has a negative effect (compared to no temporary visa) for primary Employer Sponsored visa holder and primary Skilled Independent visa holders but a positive effect for most other visa groups.

5. Oaxaca-Blinder decompositions of migrant earnings

This section of the paper uses Oaxaca-Blinder analysis to compare the economic outcomes of migrants (Oaxaca (1973), Blinder (1973). Three separate comparisons are included in this section:

- Oaxaca-Blinder analysis comparing the outcomes of different migrant streams.
- Oaxaca-Blinder comparing the outcomes of migrants with different temporary visa histories.
- Oaxaca-Blinder comparing the outcomes of migrants with different countries of birth.

Oaxaca-Blinder decompositions compare the average outcomes from two regressions and decompose the mean outcome into components that are explained by observable characteristics and those that are not. Specifically, this paper uses the two-part Oaxaca-Blinder decomposition with the difference explained by covariates based on a pooled regression as suggested by Neumark (1988) to compare different groups of migrants.

Following the derivation provided in Jann (2008), when comparing the average outcomes of two groups (A and B) using linear models of the form:[23]

$$Y_i^A = \beta X_i^A + \varepsilon_i^A \qquad (2)$$

[23] Note that the algebra here uses vector notation.

$$Y_i^B = \beta X_i^B + \varepsilon_i^A \quad (3)$$

The difference in the average outcomes ($E(Y_A)$,$E(Y_B)$) between two groups can be decomposed into an explained component and an unexplained component. Specifically,

$$E(Y_A) - E(Y_B) = \underbrace{\{E(X_A)-E(X_B)\}.\beta^*}_{\text{Explained}} + \underset{\text{Unexplained}}{E(X_A)(\beta_A - \beta^*) + E(X_B)(\beta^* - \beta_B)} \quad (4)$$

The share of the outcome variable that is 'explained' by the covariates of the two groups (X_A,X_B) is intuitive. For instance, the amount of earnings (Y) explained by gender is equal to the difference in the share of men, multiplied by the return to gender from a pooled model. The "unexplained" portion arises from two sources. One is differences in unobservable characteristics. The other is in differences to return to characteristics. Differences in returns to characteristics could arise from discrimination but it could also arise from other sources such as unobserved differences in the quality of characteristics (e.g., education obtained overseas as opposed to education obtained in Australia.)

The Oaxaca-Blinder analysis in this report is based on the Mincer regressions from the previous section (the pooled regression is the same as the main pooled model in Equation 1). Oaxaca-Blinder analysis is conducted for both the intensive and extensive margin.[24]

When interpreting the results from Oaxaca-Blinder analysis in this migration context, it is useful to interpret the results using the following categories:

- Returns to covariates that are explicitly targeted by the visa system, such as education, English language or occupation.
  - The amount of income explained by these characteristics can be interpreted as the returns to this selection process.
- Returns to covariates that the migration system is not designed to select for, such as age and gender.
  - This type of selection is often not desirable. For instance, the migration system targets younger migrants that can spend a long period in Australia before retirement. Therefore, a visa program that achieves higher short-term economic outcomes by targeting older workers goes against this policy goal.[25]
- Returns that are not explained by observable characteristics. This captures everything that is not explained by observed characteristics, including:
  - The extent to which visa streams are able to select migrants with 'unobservable' attributes, such as ambition.
  - It could also occur if the design of the visa program better utilises the skills of the visa holder.
  - Differences in the `quality' of characteristics, such as education.
  - Labour market discrimination or skill underutilisation/downgrading.

[24] Oaxaca-Blinder analysis is conducted using the Oaxaca command in Stata described in Jann (2008). Extensive margin decompositions use the logit option.
[25] Similarly, a visa program that explicitly targets high economic outcomes may indirectly target men over women in the migration system. See, for instance Boucher (2008).

Oaxaca-Blinder decomposition by visa type

The following Oaxaca-Blinder decompositions are based upon regressions that include all migrant groups but exclude 'non-migrants'. The first Oaxaca-Blinder decomposition compares the outcomes of permanent migrants in different visa programs. In line with the analysis in the preceding section, migrants are categorised into 14 groups (5 Skilled visa categories, Family and Humanitarian with each group split by primary/secondary applicant status). Regressions are based on all migrants aged 25-60 and are pooled across all years in the sample. Table 5.1 shows the Oaxaca-Blinder results for 6 migrant groups, with the results for all migrant groups in Appendix C.

Our Oaxaca-Blinder analysis compares the incomes of each group with all other permanent migrants. For instance, the first column in Table 5.1 below shows the outcomes of primary Employer Sponsored visa holder compared to all other permanent visas. It shows that they earn 48.1% more (on average) than other visa holders, with 30.1 percentage points explained by observable characteristics and 18.0% 'unexplained'.

Table 5.1: Oaxaca-Blinder decomposition of migrant income by visa type

	Emp. Sp. primary	Emp Sp. Secondary	Skill Ind. Primary	Skill Ind. Secondary	Family Primary	Humanitarian Primary
Ln (income) of visa group	11.16***	10.59***	11.04***	10.60***	10.53***	10.33***
Ln (income) of all other perm. migrants	10.68***	10.78***	10.68***	10.78***	10.85***	10.77***
Difference	0.481***	-0.192***	0.367***	-0.182***	-0.316***	-0.444***
Explained	0.301***	-0.0132***	0.233***	-0.0778***	-0.230***	-0.326***
Unexplained	0.180***	-0.179***	0.134***	-0.104***	-0.0866***	-0.118***
Explained by						
Year	0.00250***	0.0122***	-0.00717***	0.00623***	-0.00145***	-0.00974***
Age	0.0102***	-0.00109***	-0.0000157	0.00227***	0.00360***	0.00645***
Years since grant	-0.0177***	-0.00608***	0.00216***	0.00640***	-0.00266***	0.0126***
Male	0.0738***	-0.0769***	0.0729***	-0.0713***	-0.0704***	0.0835***
Occupation	0.113***	-0.0507***	0.175***	-0.0113***	-0.132***	-0.202***
Geography	0.00228***	0.00356***	0.000175***	0.000643***	-0.00111***	-0.00698***
Education	0.0132***	-0.0110***	0.0397***	0.00571***	-0.0252***	-0.0547***
English	0.0295***	0.0165***	0.0144***	0.0126***	-0.0122***	-0.120***
Visa History	0.0743***	0.103***	-0.0662***	-0.0204***	-0.0351***	-0.0357***
Party size	-0.00331***	-0.00130***	0.00312***	-0.00618***	0.0442***	-0.00103***
Grant age	0.00356***	-0.00133***	-0.000531***	-0.00244***	0.00234***	0.00215***

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%

The share explained by the OB decomposition is based on a pooled regression on all permanent migrants aged 25-60 based upon equation (1). Visa programs are defined using the concordance described in Appendix A.

Table 5.1 shows that around half of the variation in average earnings between visa groups is explained by observable characteristics, while the remaining half is unexplained. Notable features of Table 5.1 include:

- Occupation is the largest determinant of earnings differences across visa groups.
- There is a significant gender divide across visa groups and this explains a sizeable share of the earnings difference across visa groups.
- There is a relatively small impact based on differences in age.
- There is a large difference between Employer Sponsored visas and Skilled independent visas driven by visa history. This reflects the larger share of Employer Sponsored migrants having previously held a temporary Skilled Visa and a larger share of Skilled Independent migrants having previously held a student visa.

- Education plays a relatively small role compared to other factors.

The Oaxaca-Blinder Framework can also be applied on the extensive margin. In this case, the variable being compared is the probability of having income above $20,000. Table 5.2 shows selected results, with all coefficients and standard errors provided in Appendix C.

Table 5.2: Oaxaca-Blinder decomposition of migrant share above $20,000 by visa type

	Emp. Sp. primary	Emp Sp. Secondary	Skill Ind. Primary	Skill Ind. Secondary	Family Primary	Humanitarian Primary
Positive income share of visa group	0.933***	0.863***	0.923***	0.854***	0.819***	0.731***
Positive income share of all other perm. migrants	0.850***	0.863***	0.846***	0.864***	0.880***	0.867***
Difference	0.0833***	-3.38e-05	0.0769***	-0.0101***	-0.0609***	-0.136***
Explained	0.0492***	0.00611***	0.0386***	-0.00517***	-0.0287***	-0.0734***
Unexplained	0.0341***	-0.00615***	0.0383***	-0.00489***	-0.0321***	-0.0622***
Explained by						
Year	0.000317***	0.000709***	-0.000359***	0.000279***	-0.000241***	-0.000670***
Age	-0.000949***	-0.000162***	0.00214***	0.00124***	-0.000745***	-0.00192***
Years since grant	0.00106***	0.000548***	-0.000328***	-0.000412***	0.000452***	-0.000638***
Male	0.0157***	-0.0145***	0.0148***	-0.0136***	-0.0137***	0.0125***
Geography	0.00167***	0.00242***	-0.00178***	-0.000376***	-0.000122***	0.000314***
Education	0.00747***	-0.000128***	0.0123***	0.00472***	-0.00852***	-0.0267***
English	0.0102***	0.00863***	0.00937***	0.00666***	-0.00830***	-0.0395***
Visa History	0.0122***	0.00870***	0.00349***	-0.00525***	-0.00919***	-0.0154***
Party size	-0.000568***	-0.00113***	0.000803***	0.000738***	0.0113***	-0.00135***
Grant age	0.00208***	0.00104***	-0.00187***	0.000848***	0.000404***	-0.000128***

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%

The share explained by the OB decomposition is based on a pooled regression on all permanent migrants aged 25-60 based upon equation (1). Visa programs are defined using the concordance described in Appendix A.

Table 5.2 shows similar aggregate characteristics to the previous table. Roughly half of the variation between visa streams is due to observed characteristics, while the other half is unexplained. Gender, English Language skills and visa history all play an important role.

Oaxaca-Blinder Decompositions by visa history

We next compare the outcomes of migrants with different temporary visa histories. Again, the sample for the pooled regression is all migrants with 'non-migrants' excluded. Each column of the Table compares the specified group to all other migrants. For instance, the first column compares the outcomes of any permanent visa holder that has previously held a Working Holiday Maker visa with the outcomes of all other permanent visa holders. The final five columns relate to permanent visa holders that have previously held a student visa. For instance, the column labelled PhD compares the outcomes of all permanent migrants that have previously held a Student visa to study a PhD with all other permanent migrants. [26]

The majority of Australian permanent migrants have previously held a temporary visa and the outcomes of permanent migrants in Australia vary significantly based on their temporary visa history. This means that the design of the permanent visa system must ideally include the

[26] Appendix C includes OB decompositions for pre-tertiary student visa holders and student visa holders where the highest level of education is not reported in the data.

development of a pipeline of temporary migrants that are likely to succeed long term. The Oaxaca-Blinder decomposition below informs this policy objective.

The results from the OB decomposition on earnings are shown in Table 5.3.

Table 5.3: Oaxaca-Blinder decomposition of migrant outcomes by visa history

	WHM	Temp skilled	Student visas				
			PhD	Masters	Grad Dip.	Undergrad	VET
Ln (income) of visa history	10.93***	11.10***	11.04***	10.84***	10.83***	10.79***	10.55***
Ln (income) of all other perm. migrants	10.75***	10.67***	10.76***	10.76***	10.76***	10.76***	10.78***
difference	0.185***	0.431***	0.277***	0.0802***	0.0689***	0.0266***	-0.227***
explained	0.116***	0.199***	0.429***	0.212***	0.123***	0.101***	-0.117***
unexplained	0.0687***	0.231***	-0.151***	-0.131***	-0.0542***	-0.0742***	-0.111***
Explained by:							
Year	0.00666***	0.00850***	0.0199***	0.00217***	-0.00855***	0.00540***	0.0211***
Age	-0.00286***	0.0121***	0.00838***	-0.0187***	-0.00165***	-0.0421***	-0.0158***
Years since grant	-0.0121***	-0.0149***	-0.0178***	-0.00748***	0.00462***	-0.0100***	-0.0174***
Male	-0.00176***	0.0140***	0.0323***	0.0271***	-0.00138*	-0.00441***	0.0137***
Occupation	0.0228***	0.0986***	0.141***	0.104***	0.0382***	0.0936***	-0.111***
Geography	0.00503***	0.00537***	-0.00527***	-0.00136***	0.00210***	-0.00191***	0.00122***
Education	-0.00985***	0.00603***	0.204***	0.0629***	0.0351***	0.0187***	-0.0213***
English	0.0957***	0.0417***	0.000493	-0.0208***	0.0114***	-0.0212***	-0.0328***
Visa History	0.0276***	0.0222***	0.0500***	0.0740***	0.0553***	0.0689***	0.0533***
Party size	-0.0106***	0.00516***	-0.00306***	-0.00593***	-0.00925***	-0.00728***	-0.00620***
Grant age	-0.00454***	0.000395***	-0.00173***	-0.00404***	-0.00282***	0.00106*	-0.00168***

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%
The share explained by the OB decomposition is based on a pooled regression on all permanent migrants aged 25-60 based upon equation (1). Visa programs are defined using the concordance described in Appendix A.

These show that temporary Skilled visa holders and migrants that study a PhD on a student visa have the strongest earnings outcomes after transition to a permanent visa. These migrants are both older and more likely to be male than the permanent migrant cohort, although this only explains a small share of the higher earnings in these groups.

Unsurprisingly, student visas holders have a higher level of education than other permanent migrants, which results in a large share of income 'explained' by education levels. This results in a negative 'unexplained' component for migrants that have previously held a student visa.[27]

Table 5.3 also shows significant variation between types of student visa. Most notably, the outcomes of students undertaking higher education are qualitatively different to those studying through the VET system. This nuance is lost in analysis that pools all student visa outcomes together.

[27] This demonstrates the value of the OB framework. The unexplained component is closely related to the estimated coefficients from the Mincer regressions in Section 3. On its own, this value would suggest that student visa holders perform poorly compared to other visa classes. However, the OB results show that this is not the case.

A similar pattern can be seen in Table 5.4, which presents the results of a Oaxaca-Blinder decomposition on the binary outcome of having income over $20,000. In this case, master's degree students have the highest rate of positive income after transitioning to a permanent visa.

Table 5.4: Oaxaca-Blinder decomposition of migrant share above $20,000 by visa history

	WHM	Temp skilled	PhD	Masters	Grad Dip	Undergrad	VET
Positive income share of visa history	0.889***	0.907***	0.904***	0.914***	0.906***	0.902***	0.895***
Positive income share of all other perm. migrants	0.861***	0.852***	0.863***	0.859***	0.863***	0.860***	0.862***
Difference	0.0284***	0.0547***	0.0415***	0.0555***	0.0434***	0.0420***	0.0335***
Explained	0.0232***	0.0392***	0.0523***	0.0538***	0.0427***	0.0384***	0.0252***
Unexplained	0.00519***	0.0155***	-0.0108***	0.00175***	0.000771	0.00361***	0.00825***
Explained by							
Year	0.000484***	0.000695***	0.00129***	0.000192***	-0.000475***	0.000629***	0.00225***
Age	0.00216***	-0.00105***	0.000598***	0.00280***	0.00168***	0.00466***	0.00298***
Years since grant	0.00162***	0.00149***	0.00201***	0.00104***	-0.000731***	0.00164***	0.00297***
Male	0.0000103	0.00336***	0.00621***	0.00597***	-0.000687***	0.000458***	0.00355***
Geography	0.00156***	0.00151***	0.000577***	-0.00197***	-0.000726***	-0.00154***	-0.00127***
Education	-0.0000486	0.00336***	0.0203***	0.0132***	0.0140***	0.00786***	0.00170***
English	0.0126***	0.0103***	0.00253***	0.00329***	0.00617***	-0.000108	-0.00296***
Visa History	0.00307***	0.0203***	0.0179***	0.0291***	0.0216***	0.0275***	0.0158***
Party size	0.00243***	-0.00231***	0.000184***	0.00215***	0.00201***	0.00228***	0.00168***
Grant age	-0.000686***	0.00165***	0.000647***	-0.00197***	-0.00156***	-0.00492***	-0.00150***

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%
The share explained by the OB decomposition is based on a pooled regression on all permanent migrants aged 25-60 based upon equation (1). Visa programs are defined using the concordance described in Appendix A.

Oaxaca-Blinder Decompositions by country of birth

The Oaxaca-Blinder approach can be used to compare the outcomes of migrants with different countries of birth. Table 5.5 shows these estimates for 6 countries. Each column of Table 5.5 compares the outcomes of permanent migrants born in one country with the outcomes of permanent migrants from all other countries. Results for other countries and country groupings are included in Appendix C. Appendix A explains how we group countries together.

Table 5.5: Oaxaca-Blinder decomposition of migrant outcomes, by country of origin

	China	India	Indonesia	North America	United Kingdom	Vietnam
Ln (income) of country	10.70***	10.99***	10.85***	11.12***	11.21***	10.57***
Ln (income) of all other permanent migrants	10.96***	10.93***	10.94***	10.94***	10.90***	10.95***
Difference	-0.263***	0.0569***	-0.0905***	0.182***	0.307***	-0.379***
Explained	-0.151***	0.0573***	-0.0874***	0.182***	0.266***	-0.422***
Unexplained	-0.112***	-0.000374***	-0.00312***	-4.21e-05	0.0408***	0.0433***
Explained by						
Year	-0.000438***	0.00375***	-0.0115***	-0.00409***	-0.0116***	0.00688***
Age	-0.00965***	-0.00568***	-0.00402***	0.00613***	0.0173***	-0.0103***
Year since grant	-0.000533***	-0.00693***	0.0187***	-0.00167***	0.00745***	0.000943***
Male	-0.0298***	0.0205***	-0.0266***	-0.00933***	0.0124***	-0.0432***
Occupation	-0.000875***	0.0319***	-0.0250***	0.0328***	0.0440***	-0.149***
Geography	-0.000322***	-0.00389***	0.00731***	-0.00543***	0.00165***	-0.00339***

Education	0.0221***	0.0222***	0.0135***	0.0213***	-0.00999***	-0.0189***
English	-0.0935***	-0.00710***	-0.0274***	0.135***	0.147***	-0.151***
Visa category	0.0121***	0.0264***	0.0136***	-0.0166***	-0.000745***	-0.0261***
Visa history	-0.0475***	-0.0194***	-0.0463***	0.0244***	0.0545***	-0.0272***
Party size	-0.00149***	-0.00258***	-0.00109***	-0.00145***	0.00272***	-0.00204***
Grant age	-0.000687***	-0.00193***	0.00132***	4.91e-06	0.000987***	0.00120***

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%
The share explained by the OB decomposition is based on a pooled regression on all permanent migrants aged 25-60 based upon equation (1). Visa programs are defined using the concordance described in Appendix A.

Table 5.5 shows that there is significant variation in incomes across countries. However, for all countries except China, the explained portion of the variation is much larger than the unexplained portion. This was not necessarily the case with visa history and visa type. It also shows that migrants from different countries have greater variation in covariates, with gender, English language and visa history and visa category all explaining large differences in earnings.

Table 5.6 presents the Oaxaca-Blinder results on the extensive margin. Similar to the results for the intensive margin, the differences in observable characteristics account for most of the variation in outcomes by country of birth, although significant unexplained variation remains for some countries.

Table 5.6: Oaxaca-Blinder decomposition of migrant share above $20,000, by country of birth

	China	India	Indonesia	North America	United Kingdom	Vietnam
Positive income share of birth country	0.866***	0.928***	0.886***	0.905***	0.941***	0.805***
Positive income share of all other migrants	0.863***	0.854***	0.863***	0.863***	0.854***	0.864***
Difference	0.00349***	0.0743***	0.0229***	0.0426***	0.0864***	-0.0599***
Explained	-0.00914***	0.0616***	0.0271***	0.0616***	0.0689***	-0.0808***
Unexplained	0.0126***	0.0127***	-0.00421***	-0.0190***	0.0175***	0.0210***
Explained by						
Year	0.000137***	0.000749***	-0.00135***	-0.000***	-0.000736***	0.000315***
Age	0.000421***	0.00153***	0.00144***	-0.00167***	-0.00210***	0.000487***
Years since grant	0.000274***	0.000990***	-0.00289***	0.000284***	-0.000703***	-0.0000108
Male	-0.00515***	0.00516***	-0.0103***	-0.00190***	0.00423***	-0.0131***
Geography	-0.00199***	-0.00125***	-0.00152***	0.00194***	0.00276***	-0.00164***
Education	0.00531***	0.0149***	0.0145***	0.0119***	0.00622***	-0.0147***
English	-0.0105***	0.0251***	0.0209***	0.0590***	0.0508***	-0.0316***
Visa Category	0.00315***	0.0156***	0.00667***	-0.0114***	0.00602***	-0.0170***
Visa History	-0.000628***	0.000713***	-0.000148	0.00107***	0.00127***	-0.00367***
Party size	0.000598***	-0.000769***	0.00275***	0.00148***	-0.000925***	0.00174***
Grant age	-0.000725***	-0.00119***	-0.00294***	0.00100***	0.00214***	-0.00163***

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%
The share explained by the OB decomposition is based on a pooled regression on all permanent migrants aged 25-60 based upon equation (1). Visa programs are defined using the concordance described in Appendix A.

6. Comparing the determinants of economic outcomes of permanent migrants and the non-migrant population

This paper explores the determinants of economic outcomes amongst Australian permanent migrants. A natural benchmark for this exercise is to compare determinants of migrant outcomes with those of the non-migrant population. In other words, are the outcomes of the Mincer regressions in Section 3 specific to migrants, or are they describing returns that are common to all participants in the Australian labour market?

In order make this comparison, we estimate separate Mincer regressions for the migrant and non-migrant populations. For the sake of this comparison, variables which are undefined for the non-migrant population, such as years since migration, are excluded from the regressions. Specifically, the regression used in this section is of the form:

$$Y_{it} = \beta_0 + \beta_1 X_{it}^{location} + \beta_2 X_{it}^{occupation} + \beta_3 X_i^{English} + \beta_4 X_i^{education} + \beta_5 X_t^{year} + \beta_6 X_{it}^{age} + \varepsilon_{it} \quad (5)$$

where the variables are defined as described below Equation (1) in Section 4. As described in Appendix A, the non-migrant regressions use a 10 per cent sample of tax returns to reduce computational burden. Results are presented in Table 6.1.

Table 6.1: Mincer Regressions for the migrant and non-migrant population

	Migrant	Non-Migrant	Migrant	Non-Migrant
Outcome variable	Ln(wages)	Ln(wages)	Income > $20000	Income > $20000
male	0.417***	0.387***	2.140***	1.451***
PHD	0.428***	0.473***	2.457***	3.230***
Masters	0.160***	0.403***	2.016***	3.065***
Graduate certificate or diploma	0.135***	0.320***	2.147***	4.135***
Undergraduate	0.137***	0.264***	1.771***	2.451***
Certificate 3 or 4	0.0265***	0.0216***	1.414***	1.301***
Less than year 12	0.00933***	-0.143***	0.719***	0.584***
English very well	-0.187***	-0.0719***	0.809***	0.714***
English well	-0.386***	-0.211***	0.479***	0.544***
English not well	-0.548***	-0.285***	0.300***	0.422***
English not at all	-0.604***	-0.620***	0.296***	0.262***
NSW regional or remote	-0.0782***	-0.157***	1.104***	0.794***
Victoria city	-0.0648***	-0.0471***	1.007	1.060***
Victoria regional or remote	-0.0862***	-0.177***	1.107***	0.867***
Queensland city	-0.0627***	-0.0683***	1.110***	0.964***
Queensland regional or remote	-0.0298***	-0.106***	1.222***	0.793***
WA city	0.0294***	0.0314***	1.155***	1.002
WA regional or remote	0.0670***	-0.00768***	1.474***	0.872***
SA city	-0.122***	-0.102***	1.018**	0.976***
SA regional or remote	-0.0498***	-0.136***	1.313***	0.848***
Tasmania	-0.167***	-0.172***	1.088***	0.863***
ACT	0.0226***	0.0915***	1.386***	1.452***
NT	0.140***	-0.0348***	1.986***	0.842***
Financial year dummies	✓	✓	✓	✓
Year of age dummies	✓	✓		
Years since migration dummies	12,158,658	8,252,592	14,084,627	9,346,968

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%

✓ indicates that controls for these characteristics are included in the model.

Regressions are run on all permanent migrants aged 25-60 in all years.

These results show that:

- Migrants have a lower return to education than the non-migrant population.
- Occupation plays a similar role in determining migrant and non-migrant incomes.
- Migrants have a similar gender pay gap (intensive margin), but gender plays a larger role in determining whether a migrant has a positive income (extensive margin) than for non-migrants.
- Non-migrants are more likely to work and are paid more in cities. However, this pattern is much weaker among migrants.
    - The income premium for cities still exists, but is smaller than for non-migrants.
    - Other things equal, migrants are more likely to work in regions than in cities.

Comparing migrants and non-migrants using an Oaxaca-Blinder decomposition

Table 6.2 presents the results of a Oaxaca-Blinder Decomposition that compares the outcomes of the migrant and non-migrant population.[28] It shows that migrants and non-migrants have similar aggregate outcomes, and that relative to the non-migrant population:

- Migrants work in higher paid occupations.
- Migrants earn more due to their geographic location (they are more likely to live in cities)
- Migrants have higher levels of education.
- Migrants have worse English language skills.

Table 6.2: Oaxaca-Blinder comparison of migrant and non-migrant economic outcomes

	Log income	Share above $20,000
Migrants	10.94***	0.863***
Non-migrants	10.97***	0.883***
Difference	-0.0270***	-0.0197***
Explained	0.0224***	-0.0198***
Unexplained	-0.0494***	0.000108
Explained by		
Year[29]	0.0327***	-0.00134***
Age	-0.00764***	0.00253***
Male	0.00384***	-0.00140***
Occupation	0.0431***	
Geography	0.0166***	0.00188***
Education	0.0380***	0.0207***
English	-0.104***	-0.0421***
Observations	16,456,389	23,431,595

Levels of statistical significance are indicated by : * 10%; ** 5%; *** 1%. The share explained by the OB decomposition is based on a pooled regression on all permanent migrants and non-migrants aged 25-60.

[28] For consistency with other analysis in this paper, Table 6.2 present the results of a 2-part Oaxaca-Blinder decomposition. A three-part Oaxaca-Blinder decomposition is provided in Appendix C. The three-part decomposition includes the impact of different returns to characteristics, while the two-part approach assumes that these returns are fixed. In this case, the distinction is most relevant to the returns to education as migrants have a lower estimated return to education (Table 6.1).

[29] The high return for year reflects that there is a greater share of migrants in the dataset in later years. Were the Oaxaca-Blinder decomposition adjusted for average income growth, the unexplained gap would be 4.94 + 3.27 = 8.2 per cent.

It is interesting that the explained and unexplained components have different signs in column two of Table 6.2. This means that, based upon observable characteristics, migrants should have higher incomes than non-migrants. But the negative, and larger, unexplained term means that there is a gap between migrant and non-migrant wages and it is driven entirely by unexplained factors.

The unexplained term is around 5 per cent. This term captures lower returns to migrants that are not captured by the earnings equation and captures the aggregate impact of:

- Discrimination against migrant workers.
- Lower returns as migrants settle into the country and become culturally assimilated (the aggregate patterns shown in Section 3).
- Skill down-grading or underutilisation.
- Difference in measurement of variables or quality of characteristics (for instance, are education levels equivalent if studied overseas and do migrants report their English language skills in the same manner as non-migrants?).

The importance of controlling for unobservables when examining wage gaps between migrants and non-migrants is also stressed by Breunig, Hasan and Salehin (2013). They find that wage gaps are underestimated when unobservables are not controlled for. Their result is consistent with what is observed in Table 6.2.

7. How well can nominated income from visa applications predict migrant income?

When applying for a primary Employer Sponsored visa, a prospective migrant must state what he/she will be paid in the nominated role and provide documentation from their employer. This section examines how well this nominated income predicts observed migrant incomes after visa receipt.

This question is of direct policy relevance. As suggested by the Grattan Institute (2022) and the Productivity Commission (2023), using an income threshold to admit Employer Sponsored migrants, as opposed to a skills list, is appealing because it directly targets high income workers. It also has fewer information requirements than the current admission system and provides more certainty for migrants applying for visas.

The analysis that follows shows that:

- Nominated income from a visa application is a good predictor of actual income.
- Higher levels of nominated income reflect both observable and unobservable characteristics. In other words, some of the outcomes of migrants with high nominated incomes can be explained by higher levels of education, language skills, etc. but a significant amount remains ‘unexplained’.
    - This suggests that nominated income contains a significant amount of information about a potential migrant and that incorporating nominated income into the visa system has the potential to improve the targeting of the migration system.

- The share of unexplained income is particularly high among high income earners, suggesting that nominated income may be an effective way to target these migrants.
- Migrants with higher levels of nominated income are also older, have more secondary migrants and are more likely to be male. These factors would need to be considered in any decision to base visa decisions on nominated income.

In interpreting the analysis that follows, a key caveat is that the outcomes are estimated under a system in which visa applications were not based on an income threshold and the introduction of such a rule may significantly change the composition of migrants applying for the program. It would also create an incentive to misreport nominated income if this was not carefully audited.

How well does nominated income predict actual income?

As an initial step, it is useful to examine how well nominated income predicts observed income. To answer this question, we compare the predictive power of two regression models:

The first is a simple regression estimating log income using log nominated income:

$$Y_{it} = \beta_0 + \beta_1 X_t^{nominc} + \beta_2 X_t^{ysa} + \varepsilon_{it} \qquad (6)$$

The second model is the pooled Mincer model of Equation (1), but estimated using only primary Employer Sponsored migrants who report nominated income.

In both cases, the outcome variable is log income that has been 'detrended' based on average income growth.[30]

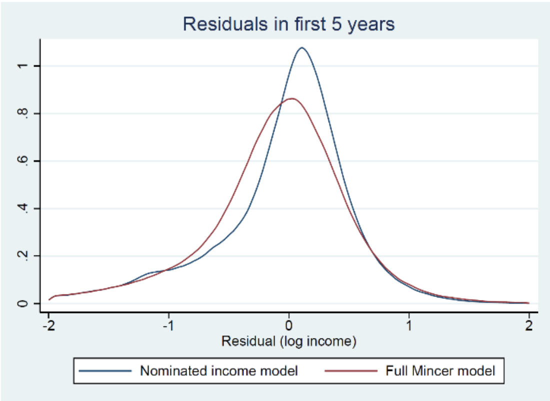
One way to compare the predictive capabilities of the two models is to compare the residuals from the two models. Figure 7.1 shows the residuals from these two model in the first 5 years and all years after 5 years. It shows that nominated income is better at predicting migrant income[31] than a Mincer regression (that includes all variables other than nominated income) and that this is true both in the short and long term.[32] Another option is to add nominated income to equation (1). If we do this, the explanatory power of the model increases substantially. Including nominated income alongside regressors which predict income produces coefficients that are difficult to interpret. We prefer the comparison that we present here.

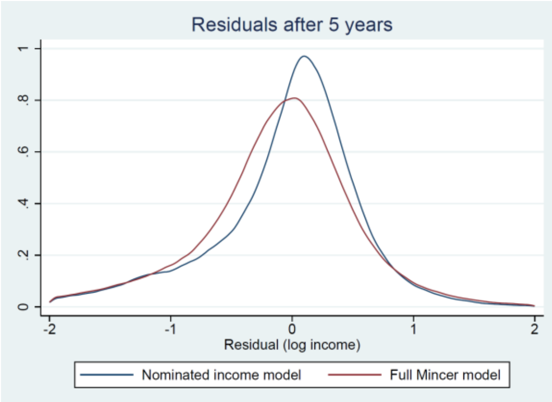
Figure 7.1: Residuals from regressions predicting incomes of primary Employer Sponsored migrants

[30] It is necessary to detrend the outcome variable as nominated income is reported in nominal terms across multiple years. The detrending calculation was based on full time average weekly earnings (ABS Catalogue Number 6302.0, Table 2.

[31] This regression is run on the intensive margin. In contrast, there is very little relationship between nominated income and extensive margin outcomes (the likelihood of having income above $20,000). This can be seen in Table 7.2.

[32] An alternative specification that includes the log of nominated income and the squared log provides a significantly better prediction, based on the stronger outcomes (relative to nominated income) among those with higher levels of nominated income. See Table 7.2.





Oaxaca-Blinder Decomposition by groups of nominated income

The analysis above shows that nominated income is a good predictor of actual income. In other words, people with higher levels of nominated income also earn higher levels of actual income. However, the desirability of using nominated income as a screening tool within the visa system also depends on why migrants are earning higher incomes. As with the Oaxaca-Blinder analysis in the previous section, it is useful to distinguish between three cases:

- Where higher incomes are explained by characteristics that the Australian migration currently selects for, such as education, English language or occupation:
  - In this case, using nominated income in the visa system could be a simple mechanism to select migrants without the need to verify migrant characteristics which can take time and money.
- Where higher incomes are explained by characteristics that the Australian migration system does not select for, such as gender or age:
  - This could be an unintended, potentially undesirable, outcome of an income floor.
- Where higher incomes are not explained by observable characteristics and nominated income contains information that is not otherwise available within the visa system.
  - In this case, using an income floor has the potential to increase the targeting of the visa system as these ‘unobserved’ characteristics can be targeted by an income floor.

Table 7.1 decomposes the earnings outcomes of migrants by nominated income into the categories above using a Oaxaca-Blinder framework. Specifically, migrants with primary Employer Sponsored visas are split into bins based upon nominated income, and then the outcomes of each bin are compared with **all other primary Skilled migrants** (including those outside of the Employer Sponsored program). For instance, the first column below shows the outcomes of migrants with a nominated income below $50,000. It shows that they earn 56.1% less than the typical primary Skilled migrant, of which 40.5% can be explained by observable characteristics (for instance, they have lower levels of education and worse English language skills than the average migrant), while 15.6% remains unexplained.

Table 7.1: Oaxaca-Blinder decomposition of log income by nominated income range (000s)

	Under $50	$50-$60	$60-$70	$70-$80	$80-$90	$90-$100	$100-$110	$110-$120	$120-$150	Over $150
Ln(income) of income group	10.65	10.71	10.89	11.12	11.36	11.46	11.54	11.61	11.77	12.16
Ln(income) of all other primary migrants	11.21	11.22	11.22	11.21	11.21	11.21	11.21	11.21	11.20	11.17

Difference	-0.561	-0.506	-0.330	-0.0895	0.152	0.252	0.334	0.401	0.570	0.993
Explained	-0.405	-0.367	-0.234	-0.0530	0.116	0.192	0.218	0.239	0.282	0.266
Unexplained	-0.156	-0.139	-0.0954	-0.0365	0.0363	0.0602	0.116	0.162	0.288	0.727
Explained by:										
Age	-0.0562	-0.0510	-0.00847	0.00709	0.0179	0.0350	0.0353	0.0349	0.0426	0.0457
Years since grant	0.0153	0.0180	0.0159	0.00912	0.00711	0.00620	0.00670	0.00613	0.00540	-0.00126
Male	-0.00177	0.00292	0.0263	0.00431	-0.0222	-0.0507	-0.0239	0.0106	0.0316	0.0555
Occupation	-0.243	-0.219	-0.142	-0.0105	0.0942	0.157	0.131	0.0936	0.0988	0.0558
Geography	-0.0102	-0.00127	0.00700	0.0124	0.00945	0.00672	0.0101	0.0108	0.00861	0.0115
Education	-0.0435	-0.0440	-0.0440	-0.0330	-0.0195	-0.00613	0.00531	0.0197	0.0240	0.0159
English	-0.0662	-0.0730	-0.0887	-0.0425	0.0288	0.0438	0.0536	0.0628	0.0710	0.0825

Note: Oaxaca-Blinder regression is based on the Mincer regression above. The decomposition is based on the two-part decomposition, with the share explained by observed characteristics based on a pooled regression of all primary Skilled migrants.
Levels of statistical significance are indicated by: * 10%; ** 5%; *** 1%

Table 7.1 shows that as we move from migrants with lower levels of nominated income to those with higher levels of nominated income, we see higher levels of observed income. We also see that part of this increase is explained by observed characteristics and part is unexplained. Interestingly, at higher income levels, the share of the observed income gap between migrants with primary Employer Sponsored visas and other primary skilled migrants which is explained by observable characteristics decreases. The unexplained share grows monotonically with nominated income.

It is notable that the incomes of those in the highest groups of nominated income are largely unexplained by a Mincer regression. This suggests that nominated income may be a mechanism to directly target very high skill workers that would otherwise be unidentifiable through observable characteristics.

How do other migrant characteristics vary with higher levels of nominated income?

While an income threshold would select migrants with higher levels of income relative to alternative systems, it is also likely to indirectly select on other migrant characteristics, such as age, gender composition and the share of primary migrant applicants. To inform any such potential policy, Table 7.2 presents key summary statistics from the MADIP dataset, split by group of nominated income.

Table 7.2 Averages of selected variables, by nominated income group

Nom. Inc. group	Observations	Nominated income	Income	Standardised income	Share below 20000	Party size	Secondary standardised income	Share male	Grant age
Under $50000	6318	47566	47977	0.70	17%	1.9	0.56	60%	30.7
$50000-$60000	37065	56176	50942	0.72	17%	2.0	0.57	66%	32.3
$60000-$70000	122191	65344	62610	0.85	14%	2.2	0.57	72%	34.4
$70000-$80000	90977	74045	79250	1.05	11%	2.3	0.64	66%	34.9
$80000-$90000	62737	84919	97062	1.28	9%	2.4	0.75	59%	35.4

$90000-$100000	58112	94940	105870	1.37	7%	2.5	0.77	50%	36.3
$100000-$110000	38953	104592	115900	1.51	8%	2.4	0.80	59%	36.4
$110000-$120000	29507	114875	126561	1.65	8%	2.3	0.85	68%	36.1
$120000-$150000	60414	133876	148197	1.91	9%	2.3	0.83	75%	36.6
Over $150000	117435	249441	245760	3.09	10%	2.8	0.90	82%	39.7

Table 7.2 shows that:

- Migrants with lower levels of nominated income have incomes which grow more slowly than Average Weekly Earnings. The opposite is true for migrants with higher nominated incomes.
- Migrants with higher nominated incomes are, on average, older.
- Migrants in the highest categories of nominated income are more likely to be men, although this pattern is not linear across categories.
- Migrants with the lowest levels of nominated income are more likely to report income below $20,000, although this pattern flattens out for migrants above $70,000.
- Migrants with higher levels of nominated income have more secondary migrants.
- There is some degree of assortative matching, with the income of secondary migrants correlated with the nominated incomes of primary applicants.

8. Implications for migration policy in Australia

The estimates in this paper can be used to inform the design of the Australian migration program. However, given the non-causal nature of the analysis, care must be taken when interpreting the results. Two important concepts to assist in this interpretation are:

- The different empirical approaches used in this paper are complimentary. For instance, the descriptive charts in section 3 inform the Mincer regressions in Section 4, which in turn are the basis for the Oaxaca-Blinder analysis in Section 5. Therefore, the results are best interpreted together.
- The analysis is based on outcomes from the existing migration program. Where a policy significantly changes the design of the migration program, the underlying relationships which have been highlighted in this paper may no longer hold.

To assist with the interpretation of the results, this section discusses the key implications for Australian migration policy.

Implications for the design of the Skilled Independent points test.

The Skilled independent visa allocates points to different migrant characteristics such as education and age that are associated with strong economic outcomes. In principle, the coefficients from the Mincer regressions from Table 4.1 could be used to adjust the points allocated to different migrant characteristics and improve the targeting of the points test.

Such a calculation would face various conceptual challenges:

- First, it would be necessary to make value judgements regarding the relative importance of:

    - intensive and extensive margin results.
    - absolute migrant outcomes compared to migrants achieving their potential.
- Second, it would be necessary to incorporate the value of migrant age in the points test. This would require a demographic model such as the FIONA model (Varela et al. 2021).
- Third, migrants applying for the Skilled Independent points test must also have a "suitable skills assessment" on a Skilled Occupation List, pay an application fee, and wait a number of months for the visa to be processed. An optimal points test calibration will also depend on these design features.
- Fourth, the regressions reflect the marginal return to characteristics within the current system and are likely to change if the points test changes. However, some insight into how stable these effects are can be gained by comparing the coefficients in the regressions split by migrant stream presented above.
- Fifth, the points test contains several categories that are not included in this study, including years of industry experience and study in a regional area.

Therefore, while the Mincer coefficients from Table 4.1 represent a useful starting point when thinking about the appropriateness of the current points test, there is significant further analysis that would need to be done to 'optimise' the points allocation.

Implications for designing a 'wage floor' for permanent employer sponsored visas

Using a wage floor as a basis for admission to the Employer Sponsored visa program has been proposed in Grattan Institute (2022) and the Productivity Commission (2023). The appeal of a wage floor is based on two ideas:

- That a migrant's current wage is a good measure of their future earnings potential.
- That it would be administratively simpler to design a visa system around a single value (nominated wage) rather than having to provide evidence of Education/English language/ Occupational qualifications.

The analysis in section 7 shows that the first claim is true (within the current system). In addition, for migrants with a high level of nominated income (above $120,000), the nominated value provides substantial information about migrant outcomes that cannot be observed through other characteristics. Therefore, the argument for a wage floor is particularly strong for high-income migrants.[33] However, Table 7.2 shows that on average, migrants with higher levels of nominated income are older, are more likely to be male and have more secondary migrants. These factors and their impact on the fiscal impact of immigrants would all need to be considered in the design of a wage floor.

Perhaps the most important caveat to this analysis is that it is conducted in a system in which the level of nominated income is not used as an entry mechanism. If a wage floor were implemented, it would change the characteristics and behaviour of migrants applying for the program.

[33] High-income migrants are also targeted by the Global Talent Visa. Therefore, a wage floor designed to target high-income migrants would need to consider the interaction with this program.

Implications for evaluating the economic impact of changes to the composition of the permanent visa program

Each year, the Commonwealth Government allocates places within the permanent visa program among different visa categories. To support this decision process, economic analysis is conducted based on the past performance of migrants in these streams such as the FIONA (Varela et a. 2021) and OLGA models (Commonwealth of Australia (2021)) maintained by the Australian Treasury. However, these models are based on short-term outcomes of migrants.

Figure 3.2 and Figure 3.4 in this paper show that these short-term outcomes overstate the long-term differences in economic outcomes between visa categories. As a result, this difference is likely to flow through to aggregate modelling results. This will not change the key qualitative findings of those models (that the economic benefits of Skilled migrants are greater than Family or Humanitarian migrants) but it will reduce the magnitude of the difference.

Occupation plays a strong role in explaining the economic outcomes of migrants

The role of Skilled Occupation Lists in the design of the Australian migration program has been criticised. For instance, the recent review of the Australian migration system (Parkinson et al. 2023) found that "the occupation lists underpinning the employer sponsored visas are unresponsive and outdated".

While there is strong evidence to support a reduced role for Skilled Occupation Lists, this paper shows that occupational sorting is one of the main factors driving different economic outcomes across permanent visa categories (Table 5.1). Therefore, any reform to the Skilled Occupation Lists should ensure that the migration system maintains a focus (directly or indirectly) on high-income occupations.[34] If lists are to be used, improving flexibility and adaptability of such lists would also be desirable.

Implications for the role of the temporary visa program as a pipeline for permanent migrants

One of the key functions of the temporary migration program is to provide a 'pipeline' for potential long-term migrants wishing to immigrate to Australia. However, despite this key role, relatively little is known about which types of temporary visa pathways deliver strong outcomes. This paper contributes several key insights:

- Migrants that have previously held a temporary visa have stronger outcomes than those that apply from offshore:
    - This effect is strongest for migrants that have previously held a Temporary skilled visa, but is seen across all temporary visa categories (Table 5.3 and 5.4).
    - All temporary visa histories have a higher share of migrants with an income above $20,000 than offshore applicants.
    - All categories (other than student visas studying VET) have a higher average income than offshore applicants.

[34] For instance, a points test that included points for high-skilled occupations would directly target occupation, while a wage floor set at high wages would indirectly target migrants from high earnings occupations.

- Migrants that enter Australia on a student visa have strong long-term economic outcomes. However, these outcomes are not as strong as would be expected given their observable characteristics (this could be the result of selection on unobservable characteristics, discrimination or challenges entering the Australian labour market).[35] This suggests that long-term outcomes of student migrants are good, there is the potential to improve outcomes amongst this group, either by selecting migrants that better integrate into the Australian labour market or .

There is significant further work that could be done in this area, including:

- Understanding in more depth the relationship between visa history and the age at which a migrant receives a permanent visa. For instance, student visa holders gain permanent visas at a younger age than temporary skilled visa holders and have more future years in the labour force. The analysis in this paper controls for the different ages at which migrants arrive, but in order to capture the economic value of additional years in the labour force it is necessary to use a structural model such as FIONA or OLGA.
- Further disaggregating the visa pathways to capture more of the complexity of the visa system.
- Capturing the economic outcomes in the years that migrants held a temporary visa (this analysis only considers the years after a migrant receives a permanent visa).
- Identifying the impact of changes to the temporary migration program (such as the uncapping of student visas or the freezing of the TISMIT) on the characteristics and outcomes of the Australian permanent migration program.

Appendix A: Data used in this report

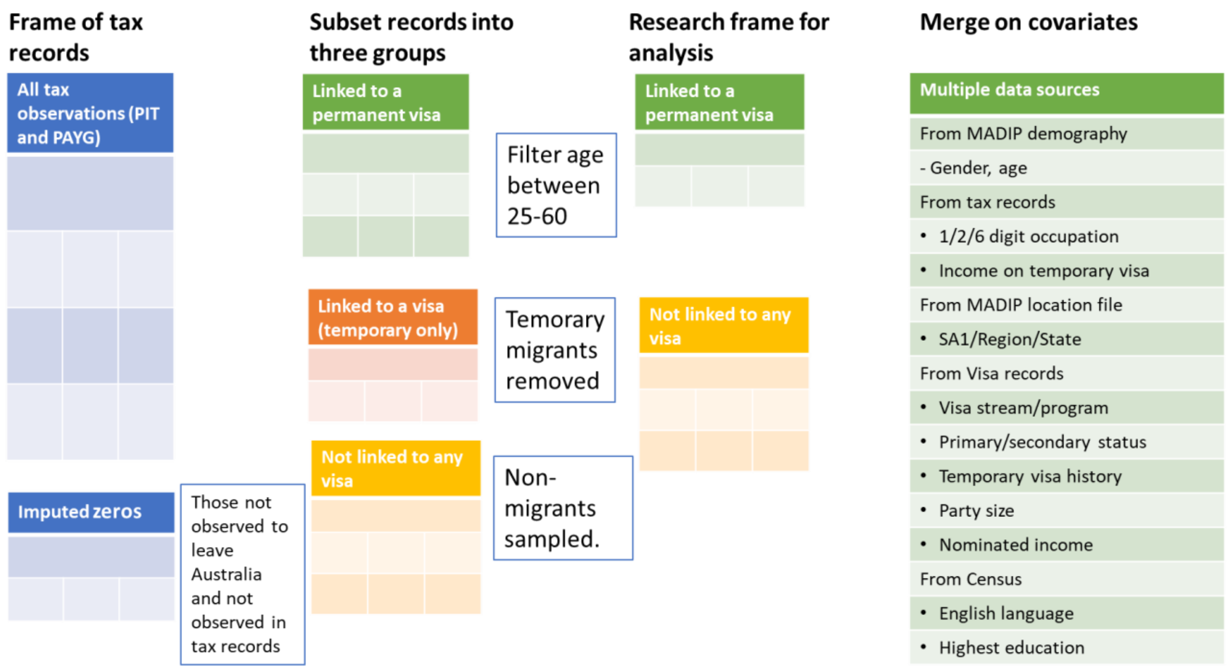
The project has been conducted using datasets made available through the Multi-Agency Data Integration Project (MADIP). These datasets include:

- All Australia Tax Returns and Payment Summaries from 2010-11 to 2020-21
- Visa records (including from visa applications and settlement data)
  - Visa applications data from 1990 onwards
  - Settlement data for permanent migrants from 2000 onwards
  - 'Travellers data' containing information on when an individual moves into or out of Australia.
- MADIP location data (derived from multiple sources)
- The 2016 Census

This project is conducted by converting these datasets into a panel data format, which are then used to conduct the analysis in this paper. An overview of the steps used to create the datasets is shown in Figure A.1 below, with a detailed description following below.

Figure A.1: Overview of data construction

[35] Understanding what is driving this effect is beyond the scope of this study but would be a strong candidate for future research.



Income

The income variable for this project is created for the years 2010-11 to 2020-21 using the following hierarchy.

- If an individual is observed to file a tax return a given financial year, then income is equal to net taxable income from the tax return.
- If no tax return is filed, but one or more PAYG payment summaries are observed, then income is equal to the sum of all payment summaries.
- If no tax return or summary statement is observed and that person is 'in Australia' then that person is allocated zero income.
- If no tax return or summary statement is observed and that person is 'outside of Australia' then that person is allocated zero income.

Defining inside/outside of Australia is described in the next section.

Process for inferring zero income

In the final step listed above, it is necessary to identify individuals in MADIP that are not in the country. This step is important as different visa classes have different shares of migrants in the country, either because they never turn up after receiving a visa or because they arrive and then migrate away from Australia (Varela et al. 2021).

This paper creates a dummy variable to indicate whether an individual in MADIP is in the country using information from the 'Travellers' dataset that shows when an individual leaves/enters the country. This variable is then used to remove any observation in which:

- No tax information is available and the next NOM movement is to move into Australia.
- No tax information is available and the previous NOM movement was to move outside of Australia.
- No tax information is available for an individual in any year and no NOM movement was observed in the dataset.

Figure A.2: Stylised example of process used to infer zero income

Individual ID	Year	Tax record	NOM variable	Tax variable
1	2011	Not observed	No observation	Excluded
1	2012	Not observed	No observation	Excluded
1	2013	Not observed	Into Australia	Excluded
1	2014	Not observed	No observation	Imputed zero
1	2015	Not observed	No observation	Imputed zero
2	2011	Not observed	No observation	Imputed zero
2	2012	Not observed	No observation	Imputed zero
2	2013	Not observed	Out of Australia	Excluded
2	2014	Not observed	No observation	Excluded
2	2015	Not observed	No observation	Excluded

There are some minor limitations to this approach. For instance, any individual who is in Australia over the period (doesn't have a NOM event) and has no tax record will be excluded from the analysis instead of captured as a zero. This is primarily a concern for the non-migrant population. In addition, this approach does not drop or adjust any tax observations for individuals that are in the country for part of the year.

Permanent visa holders used in this study

The sample of permanent visa holders in this study includes all permanent migrants from 2000 onwards. This sample is based on the 'Settlements' dataset in MADIP.

The Australian migration program also includes so called 'two stage' visas, in which a permanent visa is granted in two parts (such as 820/801 partner visas). This paper follows the convention used by the Department of Home Affairs in their annual reports that these visas are considered to be granted when an individual receives their first visa.[36]

Visa definitions used in this study

This paper groups individuals into three visa 'programs' (Skill, Family and Humanitarian) and present disaggregated results across 5 categories of Skilled visas (Employer Sponsored, Skilled Independent, Regional, Business and Distinguished Talent). This analysis requires grouping of visas together from different time periods with slightly different rules. For instance, the category of Employer Sponsored visas contains current Employer Sponsored visas (subclasses 186 and 187) as well as previous Employer Sponsored visa (subclass 186) that closed to new nominations in 2012.

These groupings are an aggregated version of the classifications used by the Department of Home Affairs and are similar to those used in Varela et al. (2021) which were based on ongoing consultation between the Australian Treasury and the Commonwealth Department of Home Affairs. The 'Family' category includes Child, Partner and Parent visas, noting that parent visas will be excluded by the age category.

Defining non-migrants

This study separates all individuals observed in the tax records into three groups:

- Any individual that can be linked to a permanent visa
- Those that can be linked to a temporary visa (but not a permanent visa).

[36] The calculation also excludes resident return visas, such as subclass 155 and 157 which allow permanent visa holders to enter/exit Australia.

- Those that cannot be linked to any visa using the MADIP linking key.

The first group are the main focus of this study and are referred to as permanent visa holders in this paper. The third group are used as the 'non-migrant' comparison group. The second group are not considered in this study.[37]

A random sample of the non-migrant population is used to reduce the computation burden required for the analysis in this paper. We pool all non-migrants across the 11 years of our dataset and take a random 10 per cent sample from this pooled sample. Individuals could thus be included in more than one year, but we do not track individuals overtime. The full sample of permanent migrants is included in the analysis.

Visa History

The visa history variables used in this study are dummy variables about whether an individual has, in their past, held a particular type of temporary visa. These visa history variables are not mutually exclusive, meaning that an individual can fall into more than one visa history category. These variables are created using the visa data in MADIP.[38] The categories used in the paper are:

- Ever held a Temporary Skilled visa
- Ever Held a Working Holiday visa
- Ever Held a Student visa for study towards:
  - PhD
  - Master's
  - Graduate Certificate or Diploma
  - Bachelor's Degree
  - Certificates and diplomas (excluding Cert 1 and 2)
  - High school

Student visa history is defined as the highest level of education that an individual studied while on a student visa.

Temporary visa pathways can be very complex and the categorisation of pathways is a simplification of this complex system. For instance, this categorisation does not distinguish between those that have multiple visa 'pathways', nor does it distinguish between the period spent on a permanent visa. Such extensions would be valuable topics for future work.

The choice of temporary visa categories was influenced by analysis of temporary visa pathways conducted by the Centre for Population (2023) at the Australian Treasury.

Geography

This paper uses the location information from the MADIP geography module. This information is derived by the Australian Bureau of Statistics using a combination of Census, DOMINO Centrelink Administrative Data, Medicare Consumer Directory, and Personal Income Tax data. From these data

[37] The linkage rates for temporary migrants are lower than for other groups. Therefore, some non-migrants are likely misallocated to the general population. However, given the relative size of these groups, it is unlikely that this would impact the results of this paper.

[38] The visa applications data used to create visa histories cover the period from 1990 onwards, while the permanent migrant data cover the period from 2000 onwards. Therefore, an individual with a temporary visa could potentially be missed if they had a temporary visa more than 10 years before being granted a permanent visa.

sources, the MADIP geography module prioritises the data source where the address has been most recently updated (rather than prioritising one data source over another).

MADIP includes geographic data at the SA1 level, which is more disaggregated than required for this study. Therefore, this information is aggregated to a combination of State and Greater Capital City Statistical Areas (ABS 2021).

Education

The Education variable in this study is based on the Highest level of educational attainment from the 2016 census.[39] This information does not vary over time for an individual.[40] Education is aggregated to 7 categories:

- PhD
- Masters
- Grad certificates and diplomas
- Bachelor's degree
- Certificates and diplomas (excluding Cert 1 and 2)
- Completed high school
- Less than high school (including Cert 1 and 2).

Occupation

Occupation is based on self-reported occupation from tax return data. These data are collected on tax returns at the six-digit level, but is aggregated to the two-digit level for analysis in this paper. The two-digit occupation categories generate 45 categories. The dataset contains missing observations which are included in regressions as their own category.

The strengths and limitations of using self-reported occupation from administrative tax data are considered in Hathorne, C. and Breunig, R. (2022). This paper finds that analysis that relies on cross-sectional variation (such as the analysis in this paper) will produce results that are nationally representative but caution against using this data to look at dynamic changes over time.

Country of birth

The country of birth variable is based on the 2016 census. This variable is reported at the country level in census data available in MADIP and is aggregated for the purpose of this study. The 10 countries with the largest number of observations are included as their own categories, while all other countries are aggregated on a regional/continental basis. The categories used in the analysis in this paper are China, India, Indonesia, Ireland, Japan, South Korea, Malaysia, Philippines, South Africa, Thailand, United Kingdom, Middle East, North America, Other Africa, Other Asia, Other Europe, Other Oceania, South/Central America.

Approach to MADIP Linkage

Different datasets are linked together using the MADIP linkage keys. This analysis was conducted using version 5 of the MADIP spine. In some cases, these linkage keys are not unique (more than one observation from a dataset is linked to an individual on the MADIP spine). In these cases, an observation is selected at random from the MADIP dataset. Where a variable is unable to be linked,

[39] Q29. "What is the level of the highest qualification the person has completed?"
[40] The decision to restrict the sample to those 25 years and older limits the impact of those who are currently studying on the results.

the observation is included in the regression analysis and a categorical variable for missingness of that variable is created and included for such observations.

Appendix B: Understanding cohort effects by comparing OLS and fixed effects estimates

The descriptive Figures in Section 3 of this report present pooled outcomes of migrants in the years following the grant of a permanent visa. This report interprets this effect as an integration pathway (some migrants take time to settle into the labour market while other don't). However, these high-level patterns could also occur as a result of a change in the underlying cohort covered by the Figures. For instance:

- The years to the right-hand side of the Figures (more years after arrival) have a greater share of migrants that arrived longer ago. Therefore, if the underlying quality of migrants has improved over time, then it will appear as if migrants have worse economic outcomes over time.
- The Figures only show migrants between the ages of 25 and 60. This means that in the early years, the Figures show only migrants that arrived as adults, while in later years, it captures the outcomes of migrants that arrived as children, primarily as secondary migrants.
- Migrants with strong early economic outcomes may be more likely to stay in the country than those that do not.

To test whether such cohort effects are a significant factor in explaining the patterns in Section 3, we estimate two regression models.

- The first model is an ordinary least squares (OLS) regression model which includes all individuals, migrant and non-migrant.
- The second model is similar but also incorporates individual fixed effects.

These models are of the form:

$$Y_{it} = \beta_0 + \beta_1 X_t^{ysa} + \beta_1 X_t^{year} + \beta_2 X_i^{age} + \alpha_i + \varepsilon_{it} \qquad (B1)$$

Where:

- Y_{it} is the economic outcome of interest (log income in the intensive model and a binary measure of income over \$20,000 in the extensive margin model).
- X_t^{ysa} is a set of dummy variables for years since arrival
- X_t^{year} is a set of dummy variables for financial year
- X_i^{age} is a set of dummy variables for single year of age
- α_i are individual fixed effects that vary by individual but do not vary by time
- ε_{it} capture all other unobservable effects that vary across both individuals and time

The intuition behind running these two models is that the OLS model is comparable to the Figures in Section 3, while the fixed effects model will only capture variation that occurs 'within' an individual's set of observation. The OLS model treats the combined term, $\alpha_i + \varepsilon_{it}$, as the unobserved component of the model. The fixed effect estimation controls for α_i.

Results from the two models are presented in Appendix C.[41] They show that the estimated coefficients from the OLS and fixed effects models are similar. Therefore, at an aggregate level, the patterns in Section 3 of this report can be interpreted as patterns of migrant integration. Further analysis of this type could identify whether this is also true for migrant subgroups.

Appendix C: Additional regression results

Additional regression results are included in a spreadsheet that accompanies this report. These include:

- Standard errors for all coefficients reported in this paper.
- Pooled regressions estimating Equation (1) using wages and salary (rather than taxable income) as the outcome variable.
- Pooled regressions estimating Equation (1) using standardised income (rather than taxable income) as the outcome variable.
- Logit regression estimates of Equation (1) that use different definitions of positive income, including:
    - Taxable income over $1
    - Salary and wage over $20,000
    - Salary and wage over $1
- Mincer regression results by 14 Permanent visa group (6 of these groups are reported in Table 4.3).
- Oaxaca-Blinder decomposition results by 14 Permanent visa group (6 of these groups are reported in Table 5.1).
- Oaxaca-Blinder decomposition results by 20 birth groups (6 of these groups are reported in Table 5.1).
- A 'three-part' Oaxaca-Blinder decomposition comparing the migrant and non-migrant populations.
- Results from the fixed effects model described in Appendix B.
- Charts that visually present the results from Section 5 (Oaxaca-Blinder decompositions) and Section 6 (comparing Mincer regression result between the migrant and non-migrant population).

[41] The spreadsheet also contains OLS models estimated separately on migrant populations aged above 18 and below 18 when a permanent visa is granted.

References

ABS (Australian Bureau of Statistics) (2021), “Australian Statistical Geography Standard Edition 3”. https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026

Andrews, D., Clarke, E., Vass, L. and Wong, A. (2023), “Misallocated migrants: Immigration and firm productivity in Australia”. e61 Research Note No. 5, https://e61.in/misallocated-migrants-immigration-and-firm-productivity-in-australia/

Blinder A. (1973), “Wage Discrimination: Reduced Form and Structural Estimates.” Journal of Human Resources, 8(4), 436–455.

Boucher, A. (2007), Skill, “Migration and Gender in Australia and Canada: The Case of Gender-based Analysis”. Australian Journal of Political Science, 42:3, 383-401.

Breunig, R., S. Hasan and M. Salehin (2013), “The immigrant wage gap and assimilation in Australia: does unobserved heterogeneity matter?”. Economic Record, Volume 89, Number 287, pp. 490-507.

Centre for Population (2023), “Pathways from temporary visas to permanent residency: a case study of migrants that arrived in 2006-07”. Australian Treasury.

Commonwealth of Australia (2021), “2021 Intergenerational Report: Australia over the next 40 years”.

Department of Home Affairs (2021), “Permanent Migration Program (Skilled & Family) Outcomes Snapshot Annual Statistics Program years 2011-12 to 2020-21”. https://www.homeaffairs.gov.au/research-and-statistics/statistics/visa-statistics/live/migration-program

Hathorne, C. and Breunig, R. (2022), “Occupational Mobility in the ALife Data: How Reliable are Occupational Patterns from Administrative Australian Tax Records?”. Economic Papers, 41(4), pp. 297-324.

Jann, B. (2008), “The Blinder–Oaxaca Decomposition for Linear Regression Models”. The Stata Journal, 8(4), 453–479.

Mackey, W., Coates, B., and Sherrell, H. (2022), “Migrants in the Australian workforce”. Grattan Institute Guidebook No. 2022-06, May 2022.

McKenzie, D and Yang, D. (2010), “Experimental Approaches in Migration Studies”, World Bank Policy Research Working Paper No. 5395.

Mincer, J. (1974), “Schooling, Experience, and Earnings”. Columbia University Press for the National Bureau of Economic Research.

Neumark, D. (1988), “Employers’ Discriminatory Behavior and the Estimation of Wage Discrimination”. The Journal of Human Resources 23: 279–295.

Oaxaca R. (1973), “Male-Female Wage Differentials in Urban Labor Markets.” International Economic Review, 14(3), 693–709.

OECD (Organisation for Economic Cooperation and Development) (2023), "The impact of migration on local productivity in Australia", Paper 2 of OECD's Migration and Regional Development Project conducted for the Australian Treasury (due for publication late 2023).

Parkinson, M. Howe, J. and Azarias, J. (2023), "A migration system for Australia's future", Discussion Paper.

Productivity Commission (2023), "5-year Productivity Inquiry: Advancing Prosperity", Vol. 1, Inquiry Report no. 100, Canberra.

Varela, P., Husek, N. Willliams, T. Maher, R. & Kennedy, D. (2021), "The lifetime fiscal impact of the Australian permanent migration program", Treasury Working Paper.