OECD Regional Development Papers

Migration and regional productivity: Evidence from individual wages in Australia

This paper examines the contribution of international migrants to regional differences in labour productivity in Australia. The study relies on individual-level administrative wage data from 2011 to 2018. It finds that a region with a 10% larger migrant share has, on average, a 1.3% larger regional wage difference, which indicates a positive link between migration and labour productivity. The presence of migrants benefits native workers with different skill levels residing in all types of regions. The positive effects of migrants are even more pronounced for higher-skilled migrants. Concretely, a region with a 10% larger share of higher-skilled migrants has, on average, a 1% higher regional productivity difference. However, these additional benefits mainly accrue to more productive regions and those with higher migrant shares than the median region.

JEL codes: R10, J61, R23
Keywords: Australia, migration, productivity, regional, wages

ABOUT THE OECD

The OECD is a multi-disciplinary inter-governmental organisation with member countries which engages in its work an increasing number of non-members from all regions of the world. The Organisation's core mission today is to help governments work together towards a stronger, cleaner, fairer global economy. Through its network of specialised committees and working groups, the OECD provides a setting where governments compare policy experiences, seek answers to common problems, identify good practice, and co-ordinate domestic and international policies. More information available: www.oecd.org.

ABOUT OECD REGIONAL DEVELOPMENT PAPERS

Papers from the Centre for Entrepreneurship, SMEs, Regions and Cities of the OECD cover a full range of topics including regional statistics and analysis, urban governance and economics, rural governance and economics, and multi-level governance. Depending on the programme of work, the papers can cover specific topics such as regional innovation and networks, sustainable development, the determinants of regional growth or fiscal consolidation at the subnational level. OECD Regional Development Papers are published on http://www.oecd.org/cfe/regional-policy.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcome, and may be sent to the Centre for Entrepreneurship, SMEs, Regions and Cities, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This paper is authorised for publication by Lamia Kamal-Chaoui, Director, Centre for Entrepreneurship, SMEs, Regions and Cities, OECD.

This document, as well as any statistical data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Note by the Republic of Türkiye: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Türkiye recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Türkiye shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Türkiye. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Executive summary

Migrants are an essential driver of population and economic growth in Australia. In 2021, 29% of the overall population had been born abroad, making Australia the country with the third-highest migrant share among all OECD countries. The migrant population in Australia is highly educated and well-integrated into the labour market. The level of education of migrants exceeds the average educational attainment of the native population.

The challenges and opportunities of migration differ across Australian regions because the places where migrants settle within the country depend in part on their socio-economic characteristics. For example, migrants in cities are, on average, more educated than those in rural areas. Migrants are particularly attracted to metropolitan areas: eight out of ten people born abroad live in metropolitan regions compared to seven out of ten native-born.

Measuring productivity in small regions poses significant challenges, but employing individual wages provides a practical solution. In the absence of a direct measure for regional labour productivity, this study uses a measure based on individual wages as an approximation. The idea is that higher nominal wages in a region suggest larger regional productivity. Despite its limitations, this approach offers a practical avenue for assessing the influence of regional drivers on productivity differences across regions.

This paper finds that Australian regions with larger shares of migrants tend to have higher regional wages, which provides evidence of a positive link between migration and labour productivity. The analysis shows that a region with a 10% larger migrant share (e.g., 33% instead of 30%) has, on average, a 1.3% larger regional wage difference. Moreover, lower-skilled natives benefit slightly more from migration than higher-skilled natives.

The positive effects of migrants are even more pronounced when migrants are higher skilled. Concretely, a region with a 10% larger share of higher-skilled migrants among the migrant population has, on average, a 1% larger regional labour productivity difference. This additional effect, however, only materialises in regions with above-median initial shares of migrants or productivity.

Acknowledgements

This paper was prepared by the OECD Centre for Entrepreneurship, SMEs, Regions and Cities (CFE) led by Lamia Kamal-Chaoui, Director. The work was conducted as part of the OECD Regional Development Policy Committee's program of work with financial support from the Centre for Population at the Australian Treasury.

Cem Özgüzel, Economist at CFE coordinated the preparation of the paper under the guidance of Ana Moreno-Monroy, Head of the Statistics and Territorial Analysis Unit (CFE) and the supervision of Rüdiger Ahrend, Head of the Economic Analysis, Data and Statistics Division at CFE. Jasper Hesse and Cem Özgüzel (all CFE) drafted the paper. Gabriel Chaves Bosch (CFE) provided drafting contributions throughout the entire working paper. The authors are thankful to Nadim Ahmad, Lukas Kleine-Rueschkamp, Alison Weingarden (all CFE) and Ana Damas de Matos (ELS) for their valuable comments. Alexandre Banquet, Diego Faurès, Eric Gonnard, and Claire Hoffmann (all CFE) provided statistical support.

The paper also benefited from valuable comments by Patrick Fazzone IV, Rodrigo Rodrigues, Ian South and Kasey Stanfield (all Centre for Population), Giovanni Peri (University of California, Davis), Rebecca Riley (London School of Economics), Tyler Reysenbach (Grattan Institute), Peter Varela (Australian National University), Andrew Birch, Taneem Muzaffar, Phillip Tzioumis (all Department of Home Affairs), Joe Castellino, Kathleen Cross, Linda Velzeboer (Department of Infrastructure, Transport, Regional Development, Communications and the Arts), conference participants at ERSA (Pécs, Hungary), KLEM (Manchester, UK), and APA (Canberra, Australia) and members of the Academic Expert Panel of the Centre for Population at the Australian Treasury. Abdurrahman Aydemir (Sabanci University), Clément Bosquet (University of Paris 1 Panthéon-Sorbonne), and Jan Stuhler (University of Carlos III Madrid) provided guidance throughout the project as scientific advisors.

Michel Beine (University of Luxembourg) and Christopher Parsons (University of Western Australia) kindly shared valuable data on the historical settlement of migrants across Australia. The Australian Bureau of Statistics (ABS) provided access to the Multi-Agency Data Integration Project (MADIP), which made the analysis possible.

This paper is the second output of a multi-annual collaboration between the Australian Centre for Population and the OECD. The first working paper of the project provides a detailed descriptive overview of migrants in Australia and the context of productivity and the labour market (OECD, 2023[1]). The third paper evaluates the impact of migration on regional labour markets (OECD, 2023[2]). The fourth paper assesses the impact of migration on regional innovation (OECD, 2024[3]).

Table of contents

FIGURES

TABLES

BOXES

1 Introduction

Migration is an integral part of both Australia's society and its economy. In 2021, 29% of the total population had been born abroad, making Australia the country with the third-highest migrant share among all OECD countries.[1] In the last hundred years, migration has been the main driver of Australia's population growth. Between 2000 and 2021, the share of migrants in the population increased by almost six percentage points (OECD, 2023[4]). Compared to other OECD countries, the migrant population in Australia is characterised by a higher level of education, which also exceeds the average educational attainment of the native population. Despite the significance of migration in the Australian economy and society, evidence on the economic impact of migration in Australia, especially at the subnational level, is scarce. This paper measures the contribution of migrants to regional labour productivity differences in Australian Statistical Area (SA4) regions, using individual wages to build a proxy of regional productivity.[2]

Whether migrants boost or lower regional labour productivity is ultimately a question that can only be answered empirically. Conceptually, there are channels through which migrants could increase or decrease labour productivity. Migrants may *enhance* labour productivity through the following three channels. Firstly, migrants may bring new ideas and skills from their countries of origin, which may complement or increase those in their host country. Secondly, migrants and natives usually choose and specialise in different jobs, especially when migrants do not speak the local language. Natives may choose communication-intensive jobs to exploit their superior language proficiency, while migrants prefer manual labour jobs, which are less communication-intensive. This division of work based on natives' and migrants' comparative advantages may improve labour efficiency and, finally, labour productivity. Lastly, migration may increase population density and lead to positive agglomeration effects stemming from improved labour market pooling and knowledge spillovers.

In contrast, migration may *lower* productivity through two channels. Firstly, migrants coming from different language backgrounds raise communication costs, which could reduce efficiency and productivity. Secondly, the arrival of migrants increases the labour supply, which might result in lower wages. As the increase in the labour supply may lower labour costs, firms might invest less in labour-saving technology or shift towards less productive, labour-intensive sectors. Given the complexity and interdependence of these channels, the net impact of migration on labour productivity requires a case-by-case empirical examination.

---

[1] The terms "migrants" and "foreign-born" are used interchangeably throughout this paper. Individuals born outside of their country of residence are considered migrants. Unlike citizenship, this criterion does not change over time, it is not subject to country differences in legislation, and it is thus adequate for international comparisons. In Australia, migration is primarily measured as Net Oversea Migration (NOM), which refers to the net increase or decrease in the Australian population resulting from immigration to and emigration from Australia, irrespective of the individual's country of birth or nationality.

[2] The Australian Bureau of Statistics (ABS) has designed Statistical Areas 4 (SA4) regions by considering a range of criteria that balance various factors. The two main criteria are population size and commuting patterns. As a result, the 88 SA4 regions considered in this analysis cover the whole of Australia and represent all regional labour markets. SA4 regions exhibit a functional characteristic in terms of capturing labour supply and demand. Throughout this paper, the term "region" refers to SA4 regions unless indicated otherwise. Please refer to Section 3 for more information.

The paper analyses the contribution of migrants to labour productivity differences across Australian regions. The analysis relies on administrative wage data with more than 27 million individual records covering the universe of workers from 2011 to 2018. Analysing the impact of migration on productivity at the regional level is essential as the presence of migrants, as well as the concentration of firms and productivity, is unevenly distributed across Australia. Consequently, national average effects would mask substantial regional heterogeneity. The analysis estimates regional productivity differences following the two-step approach of Combes et al. (2008[5]) using individual wages.[3] Specifically, the first step is an individual-level regression using wages, which accounts for differences in the regional industrial structure and workers' skills to calculate regional wage premiums that reflect regional productivity advantages.[4] The second step measures the contribution of migrants to regional productivity advantages (i.e., the regional wage premiums) while accounting for other regional factors that might be influential (e.g., employment density and market potential) and addressing various estimation challenges.

The paper makes three findings. First, it shows that the presence of migrants is associated with larger regional labour productivity differences in Australia. The impact of migration on regional wage differences is around 0.14, indicating that a region with a 10% larger migrant share (e.g., from 30% to 33%) has a 1.3% higher regional labour productivity. This effect is robust to alternative measurements of the regional migrant population and various identification strategies. Second, lower-skilled natives seem to benefit slightly more from migration than higher-skilled natives. This finding differs from previous research in other countries, where higher-skilled natives benefit more than lower-skilled natives. However, this differential effect is not altogether surprising given the high skill level of migrants stemming from the Australian migration system, which selects migrants based on the needs of the labour market. Third, the skill composition of migrants also contributes to regional productivity differences. Specifically, a region with a 10% larger share of higher-skilled migrants among the migrant population has, on average, a 1% larger labour productivity difference. However, this additional effect only materialises in regions with initial labour productivity or migrant shares above the national mean.

The paper is organised as follows. The next section presents a brief review of the relevant migration literature. Section 3 presents the data used in the analysis and descriptive findings. Section 4 elaborates on the empirical strategy. Section 5 presents the results of the analysis, and Section 6 concludes.

[3] Using individual wages to approximate individual productivity is standard in the economic geography and agglomeration economics literature that aims to explain the role and relevance of local factors (Combes, Duranton and Gobillon, 2008[5]; Combes, Démurger and Li, 2015[20]; De La Roca and Puga, 2016[53]; Quintero and Roberts, 2023[54]). This approach extends to measuring the positive productivity spillover effects of migration (Ottaviano and Peri, 2005[55]; Bakens, Mulder and Nijkamp, 2013[58]; Combes, Démurger and Li, 2015[20]; Combes et al., 2019[44]; Kemeny and Cooke, 2018[47]). Similar methods have also been used in previous OECD publications for comparable purposes, such as Ahrend et al. (2014[56]) or OECD (2020[57]).

[4] The analysis uses wages to measure the regional productive advantages (i.e., regional wage premia). This approach aligns with the notion that higher nominal wages in a region imply greater regional productivity (Combes, Duranton and Overman, 2005[50]). If workers were not inherently more productive, firms would leave high-wage regions and relocate to low-wage regions (Moretti, 2004[51]; Moretti, 2011[52]; De La Roca and Puga, 2016[53]). In essence, wages reflect the premium firms are willing to pay to workers in specific regions, indicating the comparative productivity of workers, even those of similar or identical qualifications. Yet, it's not entirely dismissible that variations in the cost of living across regions might play a role in the observed differences in nominal wages.

2 Related literature

Migration can affect regional labour productivity through various channels with positive or negative effects. Overall, the literature shows that the effect of migration depends on the skill and education levels of native and migrant workers (Bosetti, Cattaneo and Verdolini, 2015[6]; Fassio, Kalantaryan and Venturini, 2015[7]), as well as on the economic conditions in the receiving country and region (Alesina, Harnoss and Rapoport, 2016[8]; Bove and Elia, 2017[9]). As the size of the migrant population and their characteristics vary significantly among OECD countries, their impact on productivity is also context-specific (OECD, 2022[10]).

The literature has identified three main channels through which an increase in the number of migrants might foster regional productivity. Firstly, to avoid competing with migrants, natives may choose to specialise in communication-intensive occupations requiring high proficiency in the local language. In contrast, migrants may specialise in more manual-intensive tasks that require lower language proficiency (Peri and Sparber, 2009[11]). This division of labour can be mutually beneficial as it allows both groups to complement each others' skills, boosting overall productivity and resulting in higher wages and employment (Bosetti, Cattaneo and Verdolini, 2015[6]; Peri, 2012[12]; Mitaritonna, Orefice and Peri, 2017[13]; Docquier et al., 2020[14]). Secondly, migration and the cultural diversity among migrants can increase productivity by diversifying knowledge through the introduction of new ideas and know-how from their home countries (Alesina, Harnoss and Rapoport, 2016[8]; Bahar et al., 2022[15]). Furthermore, migrants may stimulate entrepreneurship (Kerr and Kerr, 2017[16]) and increase product varieties (Bahar and Rapoport, 2018[17]), innovation (Fassio, Montobbio and Venturini, 2019[18]), and trade (Felbermayr and Toubal, 2012[19]). Finally, more migrants might increase population density, resulting in positive agglomeration effects on regional productivity (Combes, Démurger and Li, 2015[20]).[5]

In contrast, migration can also have an adverse effect on productivity. Two channels stand out. Firstly, the arrival of predominantly lower-skilled migrants may increase the available labour supply, resulting in downward pressure on the wages of lower-skilled labour. Thus, lower wage costs might reduce firms' incentives to invest in labour-saving technical change or prompt them to prioritise industries that rely on low- or unskilled workers that have lower productivity levels (Ortega and Peri, 2011[21]). Secondly, migration can increase ethnic and cultural diversity in society and firms, yielding an increase in communication costs (Parrotta, Pozzoli and Pytlikova, 2014[22]) and cultural tensions (Hjort, 2014[23]).[6] These inefficiencies due to cultural or language barriers might hamper labour productivity.

Despite growing literature examining the empirical relationship between migration and productivity at the national level, evidence at the subnational level remains limited. Research on US states shows that the specialisation of migrants and natives in manual-intensive and communication-intensive tasks enhances

[5] Agglomeration effects are related to labour market pooling, knowledge spillovers, and matching of suppliers and buyers. While the research on agglomeration economies is well-established in the literature, only few studies examine the link between agglomeration economies and migration. Most prominent, previous research (Combes, Démurger and Li, 2015[20]) shows that internal migration in China generates positive effects on residents' wages, a proxy for productivity.

[6] Another channel is the migrant-induced change in the overall skill composition. If migrants are less skilled than native workers, their arrival can lower the total workforce's average skill level, resulting in a decline in economy-wide productivity. However, since this paper analyses the impact on native productivity, this channel is not further discussed.

productivity in the long run, as the task allocation in the workforce becomes more efficient (Peri, 2012[12]). In France, firm-level analysis shows that the arrival of predominately skilled migrants boosted productivity, especially in initially small and low-productive firms, by optimising the usage of technology and capital (Mitaritonna, Orefice and Peri, 2017[13]). These findings are also confirmed by cross-country research at the sector level covering France, Germany and the UK (Fassio, Kalantaryan and Venturini, 2015[7]).

Despite the importance of migrants in the Australian economy, empirical evidence, especially with a subnational focus, remains rather limited. Previous OECD work showed a positive association between regional migration and productivity, measured by gross value added per worker (OECD, 2023[1]). Similarly, a previous study in Australia finds a positive relationship between ethnic diversity in the population and regional wages. However, this positive effect disappears after accounting for individual characteristics that do not change over time, such as motivation, intelligence, or education (Elias and Paradies, 2016[24]).

3 Data and descriptive correlations

This section details the construction of the individual-level sample used in the first step of the analysis, followed by the construction of regional variables used in the second step of the analysis. Lastly, it gives a brief overview of migration patterns across Australia before discussing the correlation between regional migration and wages.

Constructing the individual-level sample

This study relies on rich Australian administrative data to analyse the effect of migration on regional productivity differences. The main data source is the Multi-Agency Data Integration Project (MADIP) dataset provided by the Australian Bureau of Statistics (ABS). MADIP combines information collected from different ministries related to health, education, government payments, income and taxation, employment, population demographics, and migration, as well as Census data. The data cover every Australian resident who was recorded by the Department of Social Services, paid income tax or interacted with the health system between 2006 and 2020, resulting in 27.1 million individual records. Its panel dimension allows tracking individuals over time and across Australia. Yet, the analysis is limited to 2011-2018 due to data limitations. Annex A provides a detailed description of the data sources.

The analysis combines several pieces of individual-level information on the native- and foreign-born population. It retrieves individual-level information on gender, age, annual wages, industry and occupation of employment, place of usual residence, and country of birth. The final sample is reduced to the employed native-born population aged 15-64. Following the literature, the analysis excludes public sector workers and health workers - as their wages do not necessarily follow market mechanisms - and agriculture and mining workers - as their productivity highly depends on natural resources.[7] After dropping individuals with missing information, the final dataset contains almost 26 million individual-year observations for 2011-2018.

The main analysis is conducted at the "Statistical Areas Level 4" (SA4). Australia is disaggregated into 89 SA4 regions with a population between 100 000 and 500 000. The ABS constructed the SA4 classification based on regional labour demand and supply data, aiming to mirror local labour market areas. Following the OECD territorial grid, the region "Other territories" is excluded, resulting in a total of 88 SA4 regions in the analysis.

Constructing regional variables and historical settlement patterns

The regional migrant share is the percentage of foreign-born residents in the regional population. The variable is calculated considering the total population of residents in Australia. Since MADIP does not

[7] The excluded sectors include the following Australian and New Zealand Standard Industrial Classification (ANZSIC) 2006 Divisions: Agriculture, Forestry and Fishing (A), Mining (B), Public Administration and Safety (O), Education and Training (P), and Health Care and Social Assistance.

provide the individual country of birth as a variable, the information is retrieved by combining data from the Department of Home Affairs and the 2016 Census.

The regional human capital of migrants is defined as the number of higher-skilled employed migrants over the total employed migrant population in the area. In the absence of educational or skill information for the total Australian population, the variables are calculated based on occupations using the employed tax-payer population. Occupation-level information following the Australian and New Zealand Standard Classification of Occupations (ANZSCO) serves as an approximation of the skill level.[8] Following the ABS, workers employed in the Major Groups "Managers" and "Professionals" are considered higher-skilled.

The density of employed workers relies on the headcounts of the employed working-age residents and the land area size provided by the ABS. The analysis further uses Domestic Market Potential (DMP) and Foreign Market Potential (FMP) measures. The construction of DMP combines the density of employed workers and the road distance in kilometres between population-weighted centroids of the individual regions. Lastly, FMP is the estimated travel duration between the population-weighted centroids and Australia's closest major mixed cargo naval ports, which are not restricted to special bulks, such as grains, coal, or iron ore.[9] The analysis uses the Global Human Settlement Layer (GHSL) to estimate population-weighted centroids and the Mapbox Directions API to determine the road distances and estimated travel time. Further, the Australian Department of Infrastructure, Transport, Regional Development, Communications and the Arts (DITRDCA) identified eleven nationally significant mixed ports. Section 4 explains the relevance of the individual variables for the analysis, and Annex B details the calculation of the individual variables.

Lastly, as explained in Section 4, establishing a causal relationship between migration and regional productivity differences requires using instrumental variables based on the historical settlement patterns of migrants. The information on the past settlement patterns is obtained from the Australian Census waves from 1981, 1986, 1991, and 2001. The historical data have been adjusted to 2016 borders by the ABS and provide data on the total employed population of the respective year disaggregated by country of birth, the industry of employment, and highest post-school qualification.

Descriptive evidence: The relevance of migrants in Australia

The Australian population is unevenly distributed across the country. In 2021, Australia had a population of 26 million residents living across 7.7 million square kilometres. While nine out of 88 regions reported a population density of less than one person per square kilometre, metropolitan areas, such as Melbourne (around 405 people per square km), Sydney (372 people per square km), and Brisbane (130 people per square km) are substantially more populated.

[8] The term higher-skilled throughout this paper refers to migrants or natives in employed in jobs within the Major Groups "Managers" and "Professionals". However, many national and regional statistics or other papers employ other classifications based on educational attainment. Consequently, the paper refers to migrants with tertiary education as "higher-educated". More information on the Occupation and Skill Classification are provided by the ABS (1220.0 - ANZSCO -- Australian and New Zealand Standard Classification of Occupations, 2013, Version 1.3 (abs.gov.au)).

[9] The Australian Department of Infrastructure, Transport, Regional Development, Communications and the Arts distinguishes between two types of ports: Specialised Bulk Ports and Mixed Ports.

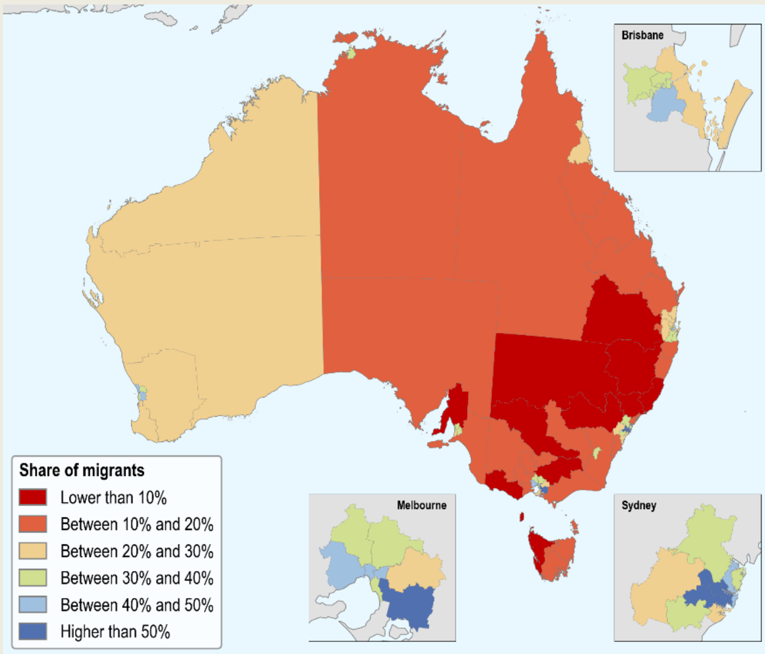
Box 1. The geographical distribution of migrants

Australia is one of the largest migrant-receiving countries in the OECD. In 2021, Australia had the third-highest share of migrants (29%) among OECD countries, after Luxembourg (49%) and Switzerland (30%). This share is substantially higher than the migrant share in Canada (21%), Germany (16%), the UK (14%), and the US (14%) (OECD, 2023[4]). Moreover, the share of migrants in Australia increased by six percentage points from 23% in 2000. Over the same period, the migrant share across the total OECD increased by just four percentage points from 10% to 14%.

The presence of migrants has a pronounced regional dimension in Australia. About 82% of all Australian migrants concentrate in large and midsize metropolitan areas, such as Brisbane, Melbourne, Perth, and Sydney, compared to 66% of natives. Consequently, only 18% of the migrant population lives in non-metropolitan areas, compared to almost one-third (33%) of natives. As a result, migrants constitute a high share of the population in large metropolitan areas (40%) such as Brisbane, Melbourne, Perth, and Sydney. Similarly, in midsize metropolitan regions, the migrant share is around 29%. In non-metropolitan areas, however, less than one-fourth of the population is born abroad, with some regions in the southeast exhibiting values of less than 10% (OECD, 2023[1]).

Figure 1. Share of migrants across Australian regions

Share of the foreign-born population across Australian regions, 2016



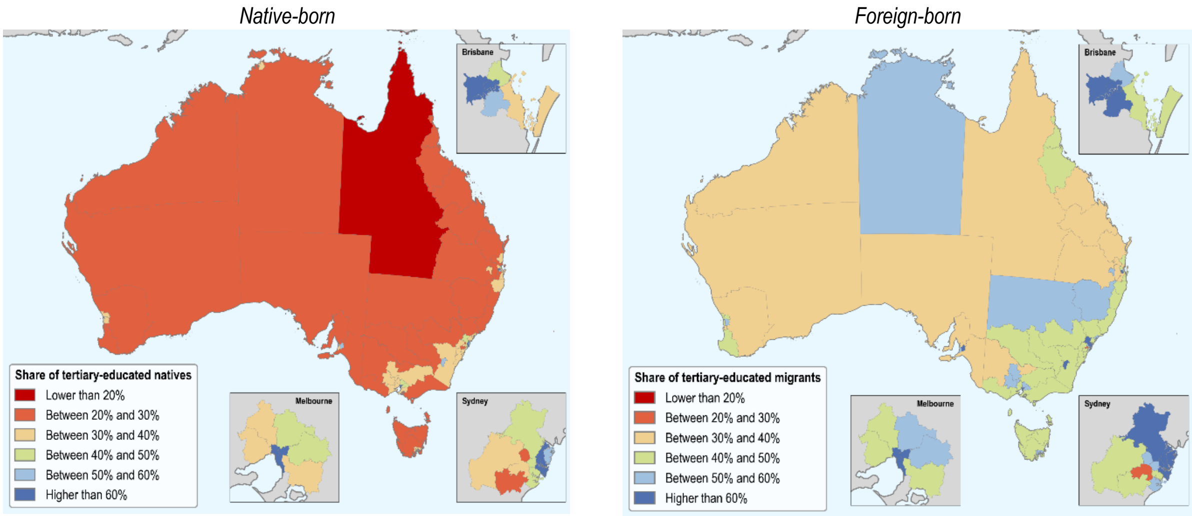
Note: The figure presents the share of foreign-born among the working-age population (15-64 years) in Australia disaggregated by regions. Data are for 2016.
Source: OECD calculations based on the Australian Census of Population and Housing 2016 accessed via ABS Census TableBuilder (accessed May 2022).

In Australia, higher-educated individuals tend to concentrate in densely populated regions. Figure 2 presents the share of the tertiary-educated population across regions, separately for natives (left panel) and migrants (right panel). The left panel shows that tertiary-educated natives are concentrated in metropolitan regions, which are generally located on the coast. The concentration is highest in cities such as Brisbane, Melbourne, and Sydney. In contrast, the share is lowest in sparsely populated regions located in Western Australia and the northeast. The pattern is somewhat similar for the migrant population (right panel). However, regional differences are less pronounced. While migrants in cities are, on average, more educated than in rural regions, the regional share of higher-educated migrants is rarely below 30%.

The migrant population is, on average, more educated than their native counterparts. Across all regions, the share of higher-educated among the migrant population is 13 percentage points higher compared to natives, on average. However, the gap varies between 33 percentage points in the Outback of the Northern Territory and -1 percentage points in South West Sydney, which is the only region with a slightly higher average education of natives compared to the foreign-born population.

Figure 2. Share of tertiary-educated natives and migrants

Share of tertiary-educated natives (left panel) and migrants (right panel) across Australian regions, 2016



Note: The figure presents the share of natives (left panel) and migrants (right panel) with tertiary education among the respective population aged 25-64 years. Data are for 2016.
Source: OECD calculations based on the Australian Census of Population and Housing 2016 accessed via ABS Census TableBuilder (accessed May 2022).

Box 2. Measuring productivity differences across regions through wages

Measuring productivity presents significant challenges, as it involves assessing the efficiency of firms and their workers in converting inputs into outputs. Although productivity is a crucial aspect of economic models, its measurement remains a complex task. The two measures most commonly used to approximate productivity are Total Factor Productivity (TFP) and labour productivity. The preferred approach highly depends on the scope of the research and the available data (OECD, 2001[25]).

TFP can be measured at the firm level using either index or estimation-based approaches. Index measures directly calculate TFP based on the difference between output and input costs. Alternatively, estimation-based measures employ econometric techniques to estimate TFP (Gal, 2013[26]). While estimation-based measures offer more flexibility by considering non-linear relationships between inputs and production, index measures are simpler to calculate.

Labour productivity can be approximated through either output per worker or wages. Output per worker is determined by dividing the total output of a firm by the number of workers involved in the production process using firm-level data. Conversely, individual wages might also provide a measure of productivity, as wages are often proportional to workers' productivity (Borjas, 2013[27]).

This study requires individual-level measures of productivity, which allows accounting for individual and regional characteristics that may affect labour productivity differences across regions. In the absence of a perfect measure, this study uses individual wages as a proxy for individual labour productivity. There are four reasons why individual wages are useful for such analysis. Firstly, using individual-level wages provides a large number of observations, which is crucial for approximating regional productivity in small geographical areas where the availability of alternative measures, such as aggregate gross value-added or firm-level data, is limited. Secondly, individual wage data permits measuring uneven effects of migration on specific subgroups of workers (e.g., highly skilled workers, women, younger workers), which is hardly feasible with other measurement methods. Thirdly, unlike alternative measures, individual-level panel data allows accounting for individual characteristics that are observable (e.g., age, education, and occupation) and unobservable (e.g., grit, motivation). Accounting for individual characteristics that influence regional productivity is vital when measuring the role of regional factors, especially in developed economies. Lastly, since reliable wage information is more commonly accessible than firm-level data, it is widely used in research which makes international comparisons with existing evidence from other OECD countries possible.

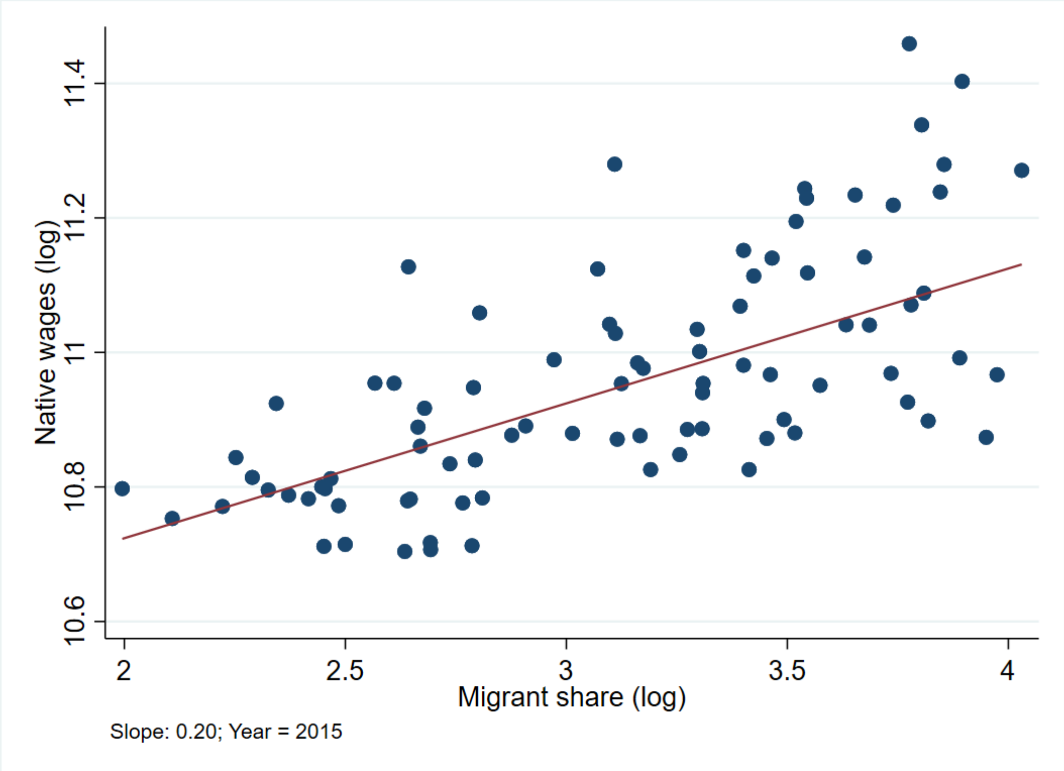
Approximating labour productivity based on individual-level wages also has its limitations. This approach enables the approximation of relative productivity differences across regions. As such, any national-level factor affecting the relationship between productivity and wages (e.g., declining labour shares, legal framework) will affect all regions. However, this approach does not allow accounting for the differences in the capital endowment or intangible assets that may exist across areas and which may affect labour productivity.

Correlation between regional average wages and the share of migrants

Regions with higher migrant shares also report higher average wages. The scatterplot in Figure 3 presents the estimated linear relationship between the logarithm of the average wages of workers in a region and the logarithm of the migrant share in 2015. The positive correlation suggests that workers receive substantially higher wages in regions with relatively more migrants. The slope of the fitted line would suggest that a 1% higher share of migrants is associated with, on average, 0.2% higher (annual) wages. However, the positive correlation between migration and native wages is most likely inflated by reverse causality and other factors correlated with the migration share (i.e., “confounders”). Migrants tend to settle in economically more prosperous and dense areas, creating a reverse causality issue. Additionally, these areas often have a more educated workforce or more productive industries. Consequently, if these regional characteristics are not taken into account, their positive effects on wages would be attributed to the presence of the migrant population, leading to the so-called “omitted variable problem”. Advanced econometric methods are required to account for these empirical challenges and establish a causal relationship between migration and productivity. The following section details these methods and how the analysis deals with these empirical challenges.

Figure 3. Regional average wages are highly correlated with the share of migrants

Relationship between the share of migrants and the average nominal wages across Australian regions, 2015



Note: The figure plots the relationship between the log share of migrants in the region (horizontal axis) and the log average individual wages among natives (vertical axis). Dots correspond to regions. The red line represents the line of best fit. The note on the bottom left corner of the figure provides the slope of the line of best fit. The estimated line of best fit is weighted by the total native population size.
Source: OECD calculations based on MADIP (accessed June 2023).

Box 3. Migration and regional innovation in Australia

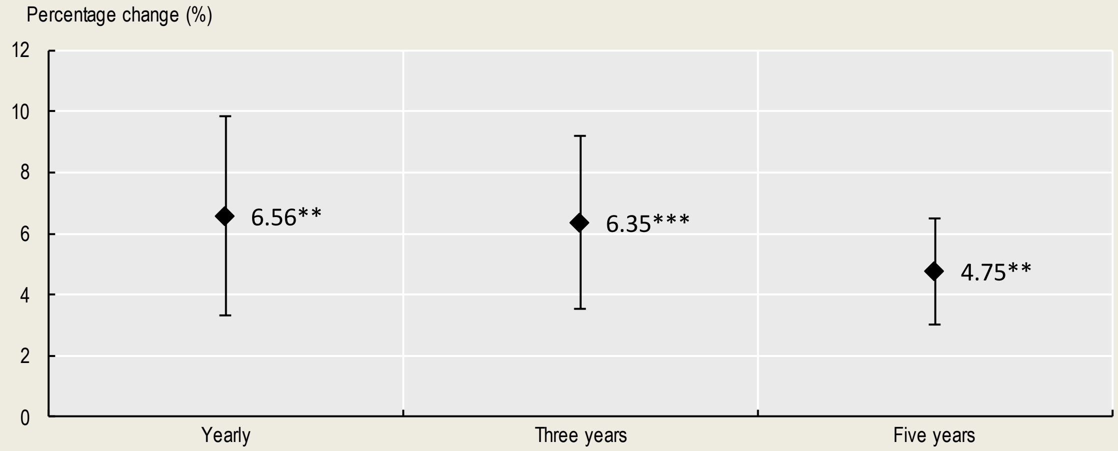
The regional impact of migration is not limited to the labour market. Migrants also play a prominent role in driving economic growth through various channels, for instance, by bringing in new skills, ideas, and fostering innovation. A companion paper evaluates the impact of migration on regional innovation in Australia (OECD, 2024[3]).

Migrants have a positive effect on patent applications across Australian regions. Figure 4 shows that, on average, a one percentage point increase in the regional employment share of higher-educated migrants relative to total employment leads to a 6.6% rise in regional patent applications in the short run (one year). These effects persist in the medium run (five years).

Patent applications typically encapsulate innovation in STEM industries. However, there is no effect on other types of innovation, such as trademarks or design rights, used more intensively by other industries. Additionally, the paper shows that the effect of migration is positive across migrants of all backgrounds, although those in scientific occupations have the largest effect.

Figure 4. The regional effects of migration on patent applications in Australia

Estimated effect of a one percentage point increase in employment due to highly educated migration on regional patent application across Australian regions, 2011-2018



Note: The figure presents IV estimates for the impact of a one percentage point increase in the workforce due to highly educated migration on regional patent applications per worker, annually and using 3 and 5 years. IV estimations use the predicted increase in the workforce due to highly educated migrants (i.e., the shift-share) as the instrument. All specifications are weighted by the number of employed natives in the considered region. Time fixed-effects are applied to account for time-varying events that might affect the entire country or economy. Standard errors are clustered at the regional level in all specifications. Standard errors are clustered at the regional level in all Columns. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively.
Source: OECD calculations based on MADIP (accessed July 2023).

4 Empirical strategy

This section presents the empirical strategy used to estimate the role of migration on labour productivity differences across Australian regions. First, it explains the empirical model used in estimation. Second, it discusses the empirical challenges in identifying the true impact of migration on regional productivity and explains methods used to address them.

Empirical model

To evaluate the impact of migration on regional labour productivity differences, the analysis uses a two-step estimation strategy commonly used in the literature examining the role of regional factors on productivity (Combes, Duranton and Gobillon, 2008[5]). The first step uses individual wages to calculate regional wage premiums for native workers, which reflects regional productivity advantages. The second step employs these estimated regional productivity advantages (i.e., regional wage premiums) to understand the role of regional characteristics. Specifically, the analysis aims to identify the impact of migrants on regional productivity advantages while accounting for workforce characteristics and other regional factors that might affect regional productivity, such as native human capital, population density, and market potential.

The two-step approach offers two main advantages. First, it allows for distinguishing the effect of individual characteristics (estimated in the first step) and regional characteristics (estimated in the second step). This allows to identify and differentiate the relevance of individual characteristics of workers in the workforce and regional characteristics. Second, it offers advantages from an econometric perspective. Splitting the estimation into two steps allows for obtaining two separate error terms. The first-step regression provides an error term at the region-sector-year level, while the second-step regression provides the error term at the region-year level. As the latter does not contain a sectoral dimension, the endogeneity concerns can be addressed without considering sector-specific endogeneity.

First step: Accounting for individual and industry characteristics

The first step estimates productivity differences – net of the skill and industry composition – across regions and over time using individual-level microdata for native workers. It relies on a standard Ordinary Least Square (OLS) regression with individual-level annual nominal wage as the dependent variable (left-hand side of the equation) and industry specialisation, individual-level characteristics, and a set of fixed-effects as independent variables (right-hand side).[10] It estimates the following regression:

$$y_{irst} = \alpha + \beta Spe_{rst} + \Phi X_{it} + \gamma_s + \gamma_i + \gamma_{rt} + \varepsilon_{irst} \quad (1)$$

where y_{irst} is the natural logarithm of the annual wage of native workers i in region r working in industry sector s in year t. Spe_{rst} is the share of workers employed in sector s over the total employed population

[10] Due to limited data availability, the analysis considers annual wages rather than hourly or daily wages. Consequently, the changes in the annual wage might also be due to changes in the labour supply.

in the region *r*. X_{it} is a set of variables capturing the effect of observable individual-level characteristics such as age and occupation on wages. γ_s, γ_i, and γ_{rt} are sector, individual and region-year fixed-effects, respectively. ε_{irst} is the error term at the region-sector-year level.

High regional industrial specialisation might determine individual wages as it creates positive spillover effects for those working in that sector (Combes and Gobillon, 2015[28]). Individuals working in these specialised industries can benefit from better skill matches in the labour market and a higher spread of information, which may boost productivity in that sector and, therefore, their wages (Özgüzel, 2022[29]).

A set of individual-level variables ($\mathrm{X_{it}}$) is included in Equation (1) to account for differences in the individual-level characteristics that might affect individual wages and productivity. These variables include *age* as a proxy for experience, as workers with more experience are expected to be more productive and receive higher wages. However, an additional year of experience might be more relevant in the early stages of the career, while the value of one additional year of experience might decrease over the years. To capture the diminishing effect of experience (i.e., the non-linear effects), Equation (1) also introduces *age²* (*age*age*).

Education is a key determinant of individual wages and productivity. Higher-skilled workers may generate more value-added and consequently might be more productive. In the absence of reliable information on educational attainment, *occupation* can provide information on the assumed skill or education level of workers. Hence, it serves as a proxy for the skill level (Combes, Duranton and Gobillon, 2008[5]).

The first step of the regression also includes individual fixed-effects (γ_i), which account for individual characteristics that are not observable and fixed over time. Unobservable characteristics such as personal grit and motivation may substantially impact individual productivity. Introducing individual fixed-effects is feasible due to the long (7-year) time dimension of observations and addresses the "sorting bias", which is further discussed in the next subsection.

The regression includes sector fixed-effects (γ_s) to absorb structural productivity differences between industries that are constant over time. Some industries, such as manufacturing, are more productive and pay higher wages than others, such as IT. Introducing these fixed-effects helps net out wage differences between individuals sourcing from their employment sector. It allows to compare the impact of regional characteristics on workers living in the same area but employed in different sectors.

Finally, the equation includes the region-year fixed-effects (γ_{rt}). Without additional controls, the region-year fixed-effects would be differences in the average wage by region and year. However, following the introduction of the abovementioned control variables and fixed-effects employed in the regression, the remaining variation captured by $\widehat{\gamma_{rt}}$ encompasses differences in regional labour productivity net of compositional differences.

Second Step: Regional drivers of productivity differences

The second step of the analysis aims to explain the impact of the migrant population and regional characteristics on regional labour productivity advantages. To do so, it uses the region-year fixed-effects ($\widehat{\gamma_{rt}}$) estimated in the first step as the dependent variable and regresses it on the regional migrant share, a measure to capture their skill composition and other regional variables that may affect productivity. Specifically, it estimates the following equation:

$$\widehat{\gamma_{rt}} = \alpha + \beta\ migrant\ share_{rt} + \gamma\ skilled\ migrants_{rt} + \kappa\ skilled\ natives_{rt} + \theta X_{rt} + \gamma_t + \varepsilon_{rt} \qquad (2)$$

where $migrant\ share_{rt}$ is the main variable of interest and corresponds to the logarithm of the share of migrants over to the total population in region r at time t. Migrants might boost or restrain labour productivity through various channels. As discussed in Section 2, the net effect of migrants on productivity can be positive or negative depending on the context. The variable of interest is defined as the number of migrants (M_{rt}) relative to the sum of native (N_{rt}) and migrant population, in region r at time t, in logarithm ($log\ \frac{M_{rt}}{N_{rt}+M_{rt}}$).

The presence of more higher-skilled workers in a region can foster regional labour productivity through positive externalities. The productivity of workers might increase when they are surrounded by higher-skilled workers in their area. To capture human capital externalities arising from the average skill level of migrants, the analysis considers the logarithm of the share of higher-skilled migrant workers, who are employed in occupations such as “managers” and “professionals”, relative to all employed migrant workers in the region *r* at time *t*, ($log\ \frac{M_{rt}^{high}}{M_{rt}^{high}+M_{rt}^{low}}$). As higher-skilled migrants tend to locate in areas with more higher-skilled natives, the same measure is employed for natives ($log\ \frac{N_{rt}^{high}}{N_{rt}^{high}+N_{rt}^{low}}$) (OECD, 2022[10]). Available education information is only time-invariant and restricted to participants of the 2016 Census. The regional share of higher-educated migrants, referring to migrants with tertiary education in 2016, and the share of higher-skilled is highly correlated, supporting the validity of the proxy.

The second step also accounts for time-varying regional characteristics (X_{rt}) that may affect productivity. First, the estimation includes *population density* to control for positive agglomeration externalities. The underlying idea is that larger markets benefit from more intensive input-output linkages and that thicker local labour markets, as well as knowledge and technological spillovers between firms, increase average regional productivity. The large theoretical and empirical literature on agglomeration economies shows that larger cities have higher productivity as their market size facilitates sharing, learning, or matching across and within sectors (Duranton and Puga, 2001[30]).[11]

Domestic and international trade are vital drivers of regional productivity. While firms that trade exhibit higher average productivity, the opportunity of trade might also foster firms’ productivity (Hering and Poncet, 2010[31]). Therefore, firms located in places with a better connection to other markets benefit from lower transportation costs and trade more easily (Krugman and Venables, 2006[32]). To acknowledge the relevance of geographic accessibility, the second-step regression includes the *Foreign Market Potential (FMP)* and the *Domestic Market Potential (DMP)*. The FMP is the travel duration (by car) to the closest major mixed cargo naval port in Australia.[12] The DMP is defined by the population density in surrounding areas and the road distance.

γ_t refers to time fixed-effects, which capture variation from time-varying shocks affecting the whole country. Further, the second-step estimation uses population weights, which reflect the number of employed natives in the region (Moulton, 1990[33]). The population weights correspond to the number of observations by region and year in the sample of the first step. Finally, ε_{rt} is the error term at the region-year level. Clustering the standard errors at the regional level addresses econometric concerns, including heteroscedasticity and arbitrary autocorrelation of the error term over time.

Estimation issues

Two potential estimation issues can lead to misleading results in the second step of the analysis. First, workers with higher abilities may choose to live in larger cities, which can create an overestimation of the

[11] Alternatively, controlling for the number of employed people in an area would partially capture this variation, yet it does not account for the stark differences in land area size across Australian regions. While the land area remains constant over time, the total number of employed in the region varies.

[12] Market potential might be measured by the presence or number of ports and airports contributing to a region's infrastructure and connectivity. However, authorities do not randomly provide such infrastructure. More productive places are more likely to receive financial support for infrastructure projects, making them even more attractive to existing and newly settling firms. This might even lead to more need and investment in infrastructure. The rational consideration of authorities and the consequential settlement behaviour of firms introduce reverse causality to the impact of more infrastructure on productivity (Combes and Gobillon, 2015[28]).

effect of regional factors on productivity. Secondly, natives and migrants may prefer to settle in dense places with, on average, higher wages or offer better employment opportunities, creating a problem of reverse causality between density and wages. The remainder of this section elaborates on these issues and how they are addressed empirically.

Estimation issue 1: Sorting of workers with higher ability into larger areas

Workers with higher abilities, such as better formal education, more experience or higher motivation, may prefer to live in larger and denser cities. The disproportionate presence of these skilled workers in more productive areas, the so-called *sorting effect*, potentially inflates the correlation between regional characteristics such as density and productivity, resulting in an overestimation of their contribution to regional productivity.

Some factors influencing individuals' abilities, such as education or experience, are observable in the data and can be captured in the estimation by introducing individual controls (e.g., age, education or occupation). However, some unobservable factors, such as motivation or grit, might also drive individual productivity. While the presence and relevance of sorting effects are country-specific, their role can be substantial and result in an overestimation of the impact of regional characteristics (Combes and Gobillon, 2015[28]). Following the seminal paper of Combes et al. (2008[5]), the analysis addresses this issue by introducing individual fixed-effects in the first step of the analysis to account for time-invariant unobservable individual characteristics.

Estimation issue 2: Reverse causality between regional characteristics and productivity

Characteristics of the regional economy may also be correlated with regional productivity, leading to an endogeneity or reverse causality problem. For example, cities are denser and also pay higher wages. As higher wages also attract workers, the arrival of these additional workers further increases the regional density, leading to higher productivity and, consequently, higher wages. Similarly, places that pay higher average wages also attract individuals with higher education who further increase the average salaries.

The well-established approach to address potential reverse causality uses instrumental variables based on historical data (Ciccone and Hall, 1996[34]). The underlying idea is that the regional population density or human capital levels in the past are persistent over time and, therefore, affect the levels even years later. However, factors that have affected these patterns in the past (such as the industrial structure of the regional economy) are not likely to be related to regional productivity levels today, especially if the time lag between the two periods is large enough. Evidence from other countries indicates that addressing this bias due to reverse causality using instruments slightly reduces the magnitude of the estimated effects, although the results remain unchanged. Following the literature, the analysis relies on the historical data from the 1981 Census (the oldest year that can be matched to current administrative borders) to instrument the levels of regional characteristics during the period of analysis and eliminate potential endogeneity in the analysis. The analysis tests the robustness of the results by using alternative instruments built from more recent census years, as well as not instrumenting or excluding the regional variables from the analysis.

Similarly, migrants do not settle randomly across regions in their host country. The higher likelihood of migrants settling in economically more dynamic places – so-called "endogenous sorting" – makes it challenging to estimate the causal effect of migration on the regional economy. For example, as migrants tend to settle in productive places, it is unclear whether migrants boost regional productivity or whether the high productivity results from other regional characteristics. If other characteristics drive productivity, this could create spurious correlations or reverse causality between migration and productivity. Thus, the "endogenous sorting" of migrants across regions creates a positive correlation between migration and regional labour productivity, contaminating the "average causal" effects of migration on productivity (Peri, 2016[35]; Card, 2001[36]). Accounting for the positive correlation between regional characteristics and migrants' presence requires more elaborate empirical strategies, as explained below. Not addressing this endogeneity concern inflates the relationship (i.e., upward bias).

Shift-share instrument based on the 1981 Census

A standard approach to address this endogeneity issue relies on an instrumental variable approach using the network instrument or the "shift-share" instrument (Altonji and Card, 1991[37]). The instrument builds on the idea that immigrants tend to settle in regions with pre-existing co-national or co-ethnic migrant networks (Gross and Schmitt, 2003[38]; Epstein and Gang, 2010[39]). Therefore, the shift-share approach leverages the historical settlement patterns of migrant groups to isolate the changes in migrant allocation resulting from past settlements. By doing so, it effectively eliminates the estimation bias that arises from migrants' higher likelihood to settle in regions with better employment opportunities.

The instrument is constructed in several steps. First, the migrant population is split into 60 origin groups (See Annex A for the country list). Second, the distribution of each of these groups across regions (i.e., the share) is calculated based on their distribution in 1981 using the 1981 Census. The *share* component of the instrument is calculated as follows:

$$Share_{n,j}^{1981} = \frac{Migrants_{n,j}^{1981}}{Migrants_{n}^{1981}} \quad (3)$$

The numerator, $Migrants_{n,j}^{1981}$, is the number of employed migrants in 1981 by 60 national grouping *n* in region *j*. The denominator, $\sum Migrants_{n}^{1981}$, refers to the total employed migrant population by national grouping *n* in 1981 across Australia.

Next, migrants from each national grouping *n* living in Australia during the period of analysis *t* ($Migrants_{n}^{t}$), are distributed across regions using their share in the past ($Share_{n,j}^{1981}$):

$$\widehat{Migrants}_{n,j}^{t} = Share_{n,j}^{1981} * Migrants_{n}^{t} \quad (4)$$

Distributing migrants from each country of origin across regions yields their distribution "predicted" by their settlement patterns in 1981. Next, the predicted number of migrants ($\widehat{Migrants}_{n,j}^{t}$) are aggregated by region to obtain the predicted number of total migrants:

$$\widehat{Migrants}_{j}^{t} = \sum_{n}^{N} \widehat{Migrants}_{n,j}^{t} \quad (5)$$

Similar to the migrant population, the settlement decision of natives may not be random as they can also be attracted to more productive places offering better wages. Furthermore, natives may also react to the arrival of migrants by moving out of more affected regions. Therefore native population numbers, used in the denominator of the migrant share $\left(\text{i.e.}, log\frac{M_{rt}}{N_{rt}+M_{rt}}\right)$ may also suffer from endogeneity problems. To address this concern, the current native population ($Natives_{j}^{t}$) is also distributed based on the predicted settlement patterns in 1981 ($Natives_{j}^{1981}$) to obtain predicted numbers of natives by region ($\widehat{Natives}_{j,t}$):

$$\widehat{Natives}_{j,t} = \frac{Natives_{j}^{1981}}{\sum Natives_{j}^{1981}} * Natives_{j}^{t} \quad (6)$$

Finally, the predicted numbers of migrants and natives are used to calculate the predicted share of migrants in the predicted total (native and migrant) population, which is used to instrument migrant share.

$$\widehat{foreign}_{j,t} = \frac{\widehat{Migrants}_{j,t}}{\widehat{Migrants}_{j,t}+\widehat{Natives}_{j,t}} \quad (7)$$

Validity of the instrumental variable approach

To obtain causal evidence, the instrument exogeneity and relevance assumptions must be valid. The former entails that historical settlement patterns must not be associated with current native labour productivity. For the latter, the instrument must be associated with changes in migration. Taken together, these assumptions imply that the instrumental variable must affect current productivity differences only through its effect on current migration inflows. In the context of shift-share instruments, recent literature has shown that instrument exogeneity can be satisfied from either exogeneity of the aggregate "shifts" (Borusyak, Hull and Jaravel, 2022[40]) or exogeneity of the baseline "shares" (Goldsmith-Pinkham, Sorkin and Swift, 2020[41]). This study relies on identification based on exogeneity of the baseline shares, which means that the initial settlement of migrants across regions in 1981 is not correlated with persistent omitted factors that could also determine regional productivity differences.[13]

A longer period between the baseline year and the beginning of the analysis period alleviates concerns that the historical presence of migrants impacts current regional productivity differences. Recent papers often select baseline years containing a sufficient time lag to counteract the potential correlation between previous settlement patterns used to construct the shift-share and current labour market outcomes (Dustmann, Fabbri and Preston, 2005[42]). However, recent literature shows that choosing a reasonably distant time lag does not provide sufficient support for instrument exogeneity, and additional evidence must be provided (Jaeger, Ruist and Stuhler, 2018[43]; Goldsmith-Pinkham, Sorkin and Swift, 2020[41]).

Examining the correlation between past settlement patterns and past regional characteristics provides supporting evidence for the instrument exogeneity assumption. Previous settlement patterns should ideally not be associated with regional characteristics (Goldsmith-Pinkham, Sorkin and Swift, 2020[41]). Table C.1. in Annex C shows that the shares of the nationalities driving most of the migration increase during the 2011-2018 period, i.e., India, China, Philippines, and Korea, as well as the instrument built upon these shares, are not associated with a set of regional characteristics and the industry composition. Annex C provides an additional discussion of the results of these tests.

In conjunction, the test, as well as the long period between the baseline year and the analysis, support the exogeneity of the baseline shares in 1981. It is likely that the instrument is uncorrelated with other characteristics determining the impact of migration on regional productivity differences. Therefore, it is plausible to assume that the changes in migration, as predicted by the instrument, affect productivity differences only through their effect on actual changes in migration.

Other instrumental variables

The instrumental variables for the human capital of migrants or natives follow a similar logic to the instrument used to calculate the migrant share. For the instrumental variable, the number of higher- and lower-skilled migrants is separately predicted following Equations 4 and 5. However, instead of the number of all employed migrants, the migrant headcounts in 1981 and the year of analysis are restricted to higher- and lower-skilled migrants, respectively. The instrument for human capital among natives is constructed similarly.

$$skilled\ \widehat{migrants}_{rt} = \frac{\widehat{Migrants}_{J,t}^{high}}{\widehat{Migrants}_{J,t}^{high} + \widehat{Migrants}_{J,t}^{low}} \quad (8)$$

[13] The number of migrants increased as of 2005 due to the reforms in the migration policies (Nguyen and Parsons, 2018[48]). While the dramatic increase also creates an exogenous "shift", this study relies on the exogeneity of the "shares".

5 Results

This section presents the results of the analysis. It examines the relevance and impact of regional and individual factors on individual wages. Moreover, it estimates the effect of migration and regional characteristics on regional productivity differences and discusses the robustness of the results. This analysis is further nuanced by looking at uneven effects for different subsamples of regions and native workers. Lastly, the section elaborates on the relevance of controlling for the sorting of skilled workers into cities.

Regional and individual drivers of labour productivity

Individual labour productivity is the result of a complex set of factors. It is driven by individual characteristics, as well as the industry and the region of employment. These factors might also be correlated with one another. Understanding the extent to which each factor matters for individual productivity is relevant for informing and formulating policies. However, the relative importance of these factors is context-specific and varies across countries, requiring case-to-case examination.

Examining the sources of variation in individual productivity is one way to understand the role and relevance of these factors. Table 1 presents the R^2 values of the regressions with individual native wages as the dependent variable. For each specification (row), the R^2 shows how much of the variation in individual wages is explained by the specific variables and fixed-effects in the regression. *Individual Characteristics* (Row 1) include the standard Mincerian controls such as age, age squared, occupation, and gender at the individual level.[14] These variables capture the importance of observable individual characteristics on wages. *Age* and *age^2* serve as a proxy for experience. As discussed in Section 3, *occupation* is used as a proxy for skills in the absence of available education data. *Industry Characteristics* are proxied by using indicator functions for the industrial sector (ANZSIC rev. 1 classification). Regional fixed-effects at the regional level and an industrial *specialisation index* are used to measure the *regional effects*. Annex B contains the exact definition of the individual variables.

Individual characteristics can explain almost a quarter of the individual wages

Individual characteristics are the most important drivers of differences in individual wages in Australia. Row 1 shows that observable individual characteristics explain 22.8% of the variation in individual wages. Industrial characteristics that do not change over time (Row 2) explain 6.6% of the variation, whereas time-invariant regional characteristics and industrial specialisation in the area (Row 3) explain 1.4%.[15]

The relationship between the different specifications and variables can be examined by comparing the explanatory power of joint analyses with the sum of their individual explanatory power (Combes et al.,

[14] Gender is not considered in the first step of the main analysis, since its variation is entirely captured by individual fixed-effects.

[15] There are many factors that may lead to a low explanatory power of the local characteristics. However, it should be noted that the use of smaller geographical units, like SA4 areas, may reduce the importance of local features (McCann, 2023[46]).

2019[44]). Rows 4-7 present the explanatory power of the combined specifications, i.e., when including and combining several specifications in one regression. If the variables of the different specifications are uncorrelated, the R^2 of the joint regression (Rows 4-7) should be similar to the sum of R^2 in the separate stand-alone specifications (Rows 1-3).

Individual characteristics are partially correlated to regional and industry attributes

The individual, industry, and regional characteristics are highly correlated and overlap to a large extent. Jointly analysing regional effects and individual characteristics (Row 4) has an explanatory power of 23.5%, which is slightly higher than the sum of both individual R^2 (1.4% + 22.8% = 24.2%), indicating that regional and individual characteristics are correlated to a certain degree. In contrast, combining regional effects and industry characteristics in one regression results in a power of 7.8% (Row 5), which is almost identical to the sum of both individual R^2 (1.4% + 6.6% = 8.0%). The similarity between both values suggests that the regional effects are rather uncorrelated (orthogonal) to individual characteristics.

Further, the R^2 for the joint regression of industry and individual characteristics (24.5%) is 4.9 percentage points lower than the sum of individual R^2 (22.8% + 6.6% = 29.4%), indicating that the sets of variables are somewhat correlated. Lastly, joining all specifications in one regression (by adding regional characteristics) raises the explanatory power of the estimation by 0.8 percentage points from 24.5% to 25.3% (row 7). The sum of individual specifications yields 30.8% (22.8% + 6.6% + 1.4%). Comparing both values indicates that the separate sets are partially correlated with each other. Notwithstanding, the comparison might be less meaningful than in other studies due to the minor explanatory power of regional effects.

Unobservable individual characteristics have high explanatory power. In addition to the observable characteristics, *occupation, gender, age, and age^2*, the analysis also considers unobservable characteristics by leveraging the panel structure of the data and using individual fixed-effects. When combining the unobservable and observable characteristics, the explanatory power lies at 74%. Further, R^2 does not increase when adding additional variables, such as industry or regional characteristics. The equivalent table to Table 1, including individual fixed-effects along with observable characteristics, is displayed in Annex D.

Table 1. The explanatory power of different specifications

R^2 for individual wages regressed on different specifications, 2011-2018

Row	Specification	R^2
(1)	Individual Characteristics	22.8%
(2)	Industry Characteristics	6.6%
(3)	Regional Effects	1.4%
(4)	Regional Effects + Individual Characteristics	23.5%
(5)	Regional Effects + Industry Characteristics	7.8%
(6)	Individual Characteristics + Industry Characteristics	24.5%
(7)	All three sets of variables	25.3%
	Number of observations	25 845 298

Note: The table presents the adjusted R^2 from an OLS estimation where the dependent variable is the natural logarithm of individual yearly wage. Different rows include different sets of independent variables. In Row 1, individual characteristics (sex, age, age^2, occupation) are included. In Row 2, industrial sector dummies (ANZSIC rev. 1 classification) are included. In row 3, industrial specialisation and regional dummies are considered. Rows 4-7 combine the independent variables from Rows 1-3. The analysis is based on 25 845 298 observations. Standard errors are clustered at the regional level in all rows.
Source: OECD calculations based on MADIP (accessed June 2023).

First Step: Calculating regional productivity differences net of skill and industrial composition

In the first step of estimation, individual wages of native workers are regressed on a set of individual characteristics, industrial specialisation, as well as sectoral and individual fixed-effects. While each Column in Table 2 presents results from different specifications, all regressions include industry and region-year fixed-effects. In Column 1, *industrial specialisation* is the only explanatory variable. Column 2 also includes observable individual characteristics. Lastly, Column 3 adds individual fixed-effects, capturing the variation in individual wages due to unobservable characteristics of workers.

Higher regional industrial specialisation benefits the productivity of individuals employed in the sector. The first-step regression shows a significant positive effect of regional industrial specialisation on individual wages (Column 1). The positive estimate indicates that a 1% increase in the regional share of workers in an industry increases individual wages of workers employed in the same sector by 0.14%. The effect is significant at the 1% level and holds when controlling for observable (Column 2) and unobservable individual characteristics (Column 3).

Men and more experienced workers receive substantially higher wages

Observable individual characteristics have a significant impact on wages. Column 2 repeats the regression of Column 1 and adds observable individual characteristics, such as gender, age, age^2, and occupation. The positive and statistically significant estimate of the male dummy suggests that male workers earn almost 38% higher wages than female workers with similar observable characteristics employed in the same sector, region and year. In addition, experience, proxied by age, indicates that older and more experienced workers receive higher wages, although this effect decreases with increasing age, as the negative estimate of age^2 indicates. Lastly, the most skill-demanding jobs are compensated with higher earnings: for instance, workers with the highest skills, i.e., working in Occupational Level 1, earn wages more than 44% higher than workers in omitted lower-skilled occupations (Occupation Levels 7-9). These findings align with the literature on the determinants of individual wages.

Adding individual fixed-effects reduces the positive impact of working in more skill-demanding occupations on wages (Column 3). The drop in the estimates indicates that positive unobservable characteristics, such as motivation, ambition, education and grit, are correlated with individual occupation groups. Gender is not included in Column 3 as its variation is entirely captured by individual fixed-effects, which account for characteristics that do not change over time.

Gradually adding variables increases the explanatory power of the estimations. Industrial specialisation, combined with industry and region-year fixed-effects (Column 1), can explain about 9% of the variation in wages. Including observable characteristics raises the explanatory power to 26% (Column 2). After further controlling for unobservable characteristics through individual fixed-effects (Column 3), the explanatory power rises to 73%, indicating that the model can explain an important part of the differences in wages. The remaining variation that the analysis is unable to account for might be due to time-varying regional or individual factors or differences in firm characteristics.

The estimated region-year fixed-effects in the first step are used as a measure of regional native labour productivity in the second step. In all specifications (Columns 1-3), region-year fixed-effects capture differences in individual wages, which can be attributed to a specific region in a given year net of the factors that are included in the regression. For example, region-year fixed-effects recovered from Column 1 would measure productivity differences across regions after accounting for differences in the industrial structure. This analysis recovers region-year fixed-effects estimated in Column 3, which allows measuring differences in the regional labour productivity of natives net of industrial and workforce composition. These estimates are used in the second step of the analysis.

Table 2. Drivers of individual wages in Australian labour markets

First-step regression for nominal annual wages at the individual level, in logarithm, 2011-2018

	No Mincerian	Mincerian	Mincerian + Ind. FE
	(1)	(2)	(3)
Male		0.377***	
		(0.000)	
Age		0.133***	
		(0.000)	
Age^2		-0.001***	
		(0.000)	
Occ. Level 1		0.442***	0.219***
		(0.001)	(0.001)
Occ. Level 2		0.383***	0.299***
		(0.001)	(0.001)
Occ. Level 3		0.238***	0.181***
		(0.001)	(0.001)
Occ. Level 4		-0.089***	-0.037***
		(0.001)	(0.001)
Occ. Level 5		0.121***	0.120***
		(0.001)	(0.001)
Occ. Level 6		-0.062***	-0.088***
		(0.001)	(0.001)
Log (specialisation)	0.141***	0.111***	0.150***
	(0.001)	(0.001)	(0.001)
Area-Year FE	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes
Individual FE	No	No	Yes
N	25 845 298	25 845 298	25 845 298
R^2	0.085	0.259	0.725

Note: The table presents the point estimates from an OLS estimation where the dependent variable is the natural logarithm of individual yearly wage, and the independent variables are industrial specialisation, individual characteristics (sex, age, age²), sectoral and individual fixed-effects, grouped in three different specifications (Columns). Every specification includes region-year fixed-effects. The analysis is based on 25 845 298 observations. Standard errors are clustered at the regional level in all Columns. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively.
Source: OECD calculations based on MADIP (accessed June 2023).

Second Step: Migration and other drivers of regional productivity differences

The second step of the analysis assesses the importance of migration on regional productivity advantages (i.e., regional wage premiums) following Equation 2. The region-year fixed-effects of the first step (Table 2, Column 3) are used as the dependent variable in the second step, estimating the regional productivity differences of native workers net of skill and industrial composition. Table 3 shows the role of migration and regional characteristics in explaining the regional productivity differences. Panel A displays the OLS estimations, while Panel B shows the Two-Stage-Least-Square (2SLS) estimates using the instruments discussed in Section 4. Columns 1-5 gradually include additional variables, with Column 5 being the preferred specification. All estimations are weighted by the number of employed native workers used in the first step. Year fixed-effects are added to account for any time-varying factor that might affect the productivity of all regions equally (e.g., currency devaluation, trade shock).

The presence of migrants is positively correlated with the regional productivity of natives

The presence of migrants is positively associated with regional productivity differences. When controlling for time-variant effects only (Column 1), the estimated effect of migration on labour productivity is 0.11, which would suggest that a region with a 10% larger migrant share (e.g., 33% instead of 30%) has a 1.1% ($1.1^{0.11} - 1 = 1.1\%$) larger labour productivity difference, on average. As migrants tend to settle in cities or denser places, which may also be more productive, Column 2 controls for regional population density, which slightly reduces the effect of migration to 0.9% ($1.1^{0.09} - 1 = 0.9\%$).

Column 3 also accounts for the market potential of the region, while Columns 4 and 5 additionally include measures of migrants' and natives' human capital to account for the skill composition of migrant workers. The results in Column 4 show that while a larger share of the migrant population is positively associated with regional productivity, these positive effects are even larger when the migrants are more skilled. However, as discussed earlier, higher-skilled migrants tend to live in places where natives are also more skilled. Therefore, Column 5 also accounts for the skill composition of the native population.

Going beyond associations and understanding the causal relationship between migration and regional productivity advantages requires addressing the estimation bias due to reverse causality. As discussed in Section 4, migrants tend to settle in economically more attractive and productive places. Their arrival might further boost regional productivity. Hence, identifying the causal impact requires an instrumental variable strategy that teases out the biases due to reverse causality and omitted variables.

More and higher skilled migrants boost native regional productivity differences

The estimated impact of migrants on regional productivity advantages slightly increases and remains significantly positive when addressing the endogeneity bias. Panel B follows the same structure as Panel A but uses the instrumental variable strategy presented in Section 5 to account for the bias. The estimate in the preferred specification (Column 5) is 0.14, suggesting that a region with a 10% larger share of migrants, on average, has a 1.3% ($1.1^{0.14} - 1 = 1.3\%$) larger labour productivity difference. Annex C elaborates on the validity of the instrument. Annex F shows the estimates of the first stage.

The skill level of the migrant workforce is another important driver of regional productivity differences. When using the instrumental variable strategy, the effect of migrants' human capital is substantially higher and statistically significant in the preferred specification, suggesting that the skill level of migrants generates positive spillover effects for natives. Specifically, the estimate (0.10) suggests that a region with a 10% larger share of higher-skilled migrants has, on average, a 1% larger native productivity difference.

The slight rise in the estimated migration effect when addressing endogeneity yields few findings. First, once the endogeneity bias is addressed, the estimated effects of migrants get slightly larger, indicating that OLS estimates suffer from a negative bias. This contradicts the general expectation of an upward bias due to migrants moving to places with higher productivity levels. However, the unique Australian context might explain this phenomenon[16]. For instance, the high costs of housing and living in metropolitan regions might lead migrants to settle in rural regions with lower costs of living but also lower regional productivity levels. Moreover, rural regions in Australia might attract migrants by offering them beneficial employment opportunities in the agriculture or mining sector (OECD, 2023[45]). Relatively larger shares of migrants in such rural regions can also create a negative bias between migrant shares and productivity levels. In addition, the Australian government issues specific visa classes that require migrants to settle in remote areas or limit their mobility, which might explain the negative bias. Nonetheless, the bias might also be driven by a complex set of factors that are potentially unique to the Australian context and require further analysis beyond this paper's scope.

Table 3. The effect of regional determinants and migration on regional productivity differences

Second-step regression for regional productivity differences, 2011-2018

	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Log Migrant Share	0.11421***	0.08930***	0.09567***	0.09614***	0.09331***
	(0.010)	(0.01)	(0.015)	(0.014)	(0.012)
Log Share Higher-Skilled Migrants				0.05599***	0.03551
				(0.022)	(0.048)
Log Share Higher-Skilled Natives					0.02467
					(0.056)
Time FE	Yes	Yes	Yes	Yes	Yes
Pop. weights	Yes	Yes	Yes	Yes	Yes
Control	None	Dens	Dens, MP	Dens, MP	Dens, MP
N	704	704	704	704	704
R^2	0.889	0.902	0.903	0.907	0.907
Panel B: IV					
Log Migrant Share	0.12867***	0.11045***	0.13119***	0.12947***	0.13545***
	(0.011)	(0.013)	(0.018)	(0.016)	(0.017)
Log Share Higher-Skilled Migrants				0.05689***	0.10413**
				(0.020)	(0.052)
Log Share Higher-Skilled Natives					-0.05691
					(0.060)
Time FE	Yes	Yes	Yes	Yes	Yes
Pop. weights	Yes	Yes	Yes	Yes	Yes
Control	None	Dens	Dens, MP	Dens, MP	Dens, MP
N	704	704	704	704	704
R^2	0.887	0.899	0.897	0.902	0.902
F-stat	830.658	537.891	311.647	330.441	254.251

[16] SGS Economics & Planning presents several reports and interactive maps, showcasing rental affordability in Australian regions: Rental Affordability Index | SGS Economics & Planning (sgsep.com.au).

Note: The table presents both OLS estimates (Panel A) and IV estimates (Panel B) of the second-step regression for regional productivity differences. Region-time fixed-effects estimated in Equation 1 are used as the dependent variable. The logarithm of the migrant share, the share of higher-skilled migrants and the share of higher-skilled natives are the independent variables. Additional controls are Population Density (Columns 2-5), Domestic Market Potential (Columns 3-5), and Foreign Market Potential (Columns 3-5). In Panel B, the endogenous variable "Log Migrant Share" is instrumented as described in the empirical strategy. All the specifications are weighted by the number of employed natives in the region. Time fixed-effects are applied to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications. The analysis considers 88 regions over eight years, yielding 704 observations.
Source: OECD calculations based on MADIP (accessed June 2023).

The positive effect of migrants' presence on native labour productivity differences exhibits robustness across various dimensions. Firstly, the findings are consistent and unaffected by alternative measures of the migrant share, as explored in different specifications. Independent of the definition of the migrant share measurement, the positive effect on productivity remains stable. Secondly, the analysis accounts for different identification strategies, including a past-settlement instrument to address spatial sorting and reverse causality concerns for various regional characteristics. Despite these methodological adjustments, the estimated impact of migrants on native productivity remains statistically significant. Moreover, the study examines the influence of alternative definitions of population density on the results. By considering various measures, such as population density based on the surface area, populated area, or built-up area, the robustness of the findings is confirmed, highlighting the robust and positive impact of migrants on native labour productivity. Annex G presents comprehensive details and results of these analyses for further reference.

Box 4. Sorting effects in Australia

Identifying sorting effects in Australia is valuable for understanding the role of urban areas in the national economy and the gains associated with them. This subsection aims to detect if sorting effects exist in Australia. Moreover, it seeks to understand its importance when estimating the contribution of migrants to regional productivity.

Native workers with higher abilities tend to live in denser and more productive places. An important empirical challenge in understanding the drivers of regional productivity differences is the sorting of native workers into regions based on their characteristics. Drivers might be observable characteristics, such as age, experience, and skill or unobservable ones, like grit, motivation, and intelligence. As discussed in Section 4, if native workers with higher abilities settle in Australian regions that are denser or more productive, it can lead to an overestimation of the importance of regional factors, although place-based and personal-based attributes are often closely linked and difficult to distinguish (McCann, 2023[46]). To address this potential bias due to so-called *sorting effects*, the analysis in the paper accounts for individual characteristics using individual control variables and fixed-effects.

A common way to detect sorting effects is to compare second-step results, which use different sets of controls in the first step. Table 4 displays the second step of the analysis for different first-step specifications. The regional productivity measure used in Columns 1 and 2 is estimated, controlling for industry fixed-effects and industrial specialisation. For Columns 3 and 4, the first step further nets out observable individual characteristics. Lastly, Columns 5 and 6 present the preferred specification in the first step, including the full set of variables and individual fixed-effects. The odd-numbered Columns regress regional productivity on migration with no controls except for time-varying factors, while the even-numbered Columns present the preferred specification. Further, Annex H presents the second stage estimates without controlling for the occupations in the first stage.

Accounting for sorting based on individual characteristics matters when estimating the contribution of migration to regional productivity. The estimated effect of migration of 0.15 in Column 1, which does not account for any individual characteristics, reduces to 0.13 in Column 5 once observable and unobservable individual characteristics are considered. However, accounting for regional characteristics such as skill composition of workers, regional density, or market access in the second step (Columns 2, 4 and 6) addresses the measurement issue and leads to results ranging between 0.14 and 0.15.

Controlling for observable and unobservable characteristics in the native population substantially reduces the relevance of migrants' human capital. Without controlling for individual characteristics in the first step, the impact of human capital among migrants is at 0.22 (Column 2), suggesting that a region with a 10% larger share of higher-skilled migrants among the migrant population has a 2.1% higher labour productivity difference. This effect decreases to 0.18, corresponding to 1.7% (Column 4) when controlling for observable characteristics and almost halves to 0.10 (1%) when also controlling for unobservable characteristics (Column 6).

In Australia, the sorting of higher-skilled workers into better-paying regions is less prevalent than in other countries. The literature suggests that higher-skilled workers tend to sort into cities and denser areas (OECD, 2022[10]). As migrants also tend to allocate in denser areas, analysing the impact of migration on productivity without controlling for skill-related characteristics causes an endogeneity problem and potential bias. However, the findings in Table 4 do not support this concern. While the estimated impact of regional migration is higher without addressing regional differences in skills among workers (Column 1) compared to the preferred specification in the first step (Column 5), this gap diminishes when controlling for regional characteristics in the second step (Columns 2 and 6).

Table 4. The effect of sorting on the second step

Second-step regression for regional productivity differences based on the different specifications in the first step, 2011-2018

	No Minc.	No Minc.	Mincerian	Mincerian	Minc. + Ind FE	Minc. + Ind. FE
	(1)	(2)	(3)	(4)	(5)	(6)
Log Migrant share	0.14935***	0.13813***	0.10844***	0.14607***	0.12867***	0.13545***
	(0.020)	(0.026)	(0.016)	(0.026)	(0.011)	(0.017)
Log Share Higher-Skilled Migrants		0.21642***		0.18142**		0.10413**
		(0.082)		(0.080)		(0.052)
Log Share Higher-Skilled Natives		-0.05516		-0.14116		-0.05691
		(0.098)		(0.095)		(0.060)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Population weights	Yes	Yes	Yes	Yes	Yes	Yes
Regional controls	None	All	None	All	None	All
N	704	704	704	704	704	704
R2	0.594	0.704	0.605	0.627	0.887	0.901
F-stat	830.658	254.251	830.658	254.251	830.658	254.251

Note: The table presents IV estimates of the second-step regression for different regional productivity differences measures. Region-time fixed-effects estimated in the first step include different controls. In Columns 1 and 2, the first-step regression only controls industry fixed-effects and industry specialisation, and does not account for individual Mincerian controls ("No Minc."). In Columns 3 and 4, the first-step regression further controls for age, age^2, male and occupation. In Columns 5 and 6, the first-step regression further controls for individual fixed-effects. In the second step, the logarithm of the migrant share, the share of higher-skilled migrants and the share of higher-skilled natives are the independent variables. The migrant share variable is instrumented. Additional controls are Population Density, Domestic Market Potential and Foreign Market Potential. All the specifications are weighted by the number of employed natives in the region. Time fixed-effects are present to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively. The analysis considers 88 regions over eight years, yielding 704 observations.
Source: OECD calculations based on MADIP (accessed June 2023).

Uneven effects across people and places

The estimated average effects potentially mask uneven impacts of migration on natives and regions. This subsection goes beyond the average effect and disaggregates the effect by characteristics of natives and regions.

Uneven effects across natives with different skill levels

Migrants may have uneven effects on the labour productivity of natives. In particular, while all natives may benefit from synergies and skill complementarities that may boost their productivity, these positive effects are expected to be higher for natives with different skills than migrants (Kemeny and Cooke, 2018[47]). Evidence from most OECD countries suggests that higher-skilled natives are more likely to benefit from such synergies as migrants tend to be relatively less educated than natives (OECD, 2022[10]). However, given the substantial share of higher-educated migrants, such findings might not apply to Australia.

The two-step approach allows to estimate the productivity effects of migrants on higher- and lower-skilled natives separately. In the baseline, the analysis calculates region-year fixed-effects using the total sample of workers. Hence, to investigate uneven effects across different native subgroups, higher- and lower-skilled, the first step separately estimates region-year productivity differences for both groups. Table 5 presents the second-step results for higher-skilled natives (Columns 1 and 2) and lower-skilled natives (Columns 3 and 4). Further, Annex G excludes native workers below 25 as well as workers below 25 and above 54 to support the robustness of the estimates.

The presence of migrants boosts the productivity of higher- and lower-skilled natives, with slightly higher impacts for lower-skilled natives

The productivity-enhancing impact of migration applies to high and lower-skilled natives. When only controlling for time-variant characteristics (Columns 1 and 3), the estimates are very similar for higher- and lower-skilled natives. However, in the preferred specification, controlling for the full set of regional characteristics, the estimated impact of migrants on the productivity of higher-skilled natives is slightly lower (0.12) compared to the estimate for lower-skilled natives (0.14). These estimates indicate that a region with a 10% larger migrant share has a 1.1% higher labour productivity difference for higher-skilled natives and a 1.3% higher labour productivity difference for lower-skilled natives. In contrast, the presence of higher-skilled migrants has a similar effect on both native subgroups.

Natives of all skill groups seem to benefit roughly equally from migration. This differs from evidence for other OECD countries where higher-skilled natives benefit more from the presence of migrants (Kemeny and Cooke, 2018[47]). Several factors related to the unique features of Australia can possibly explain these results. First, as noted earlier, Australia has one of the largest high-educated migrant populations among OECD countries, with average education shares surpassing those of the native population. This stands in stark contrast with most OECD countries where migrants have lower levels of education than natives (OECD, 2023[1]). Second, migration policies in Australia aim to address skill shortages. Migration may alleviate skill shortages in regions, which may improve overall productivity and benefit all groups.

Table 5. Uneven effects across natives with different skill levels

Second-step regression for regional productivity differences for higher- and lower-skilled natives, separately, 2011-2018

	Panel A: Higher Skilled		Panel B: Lower Skilled	
	(1)	(2)	(3)	(4)
Log Migrant Share	0.12339***	0.11752***	0.12971***	0.13538***
	(0.011)	(0.017)	(0.011)	(0.017)
Log Share Higher-Skilled Migrants		0.12053**		0.12173**
		(0.053)		(0.052)
Log Share Higher-Skilled Natives		-0.07097		-0.05120
		(0.058)		(0.061)
Time FE	Yes	Yes	Yes	Yes
Pop. weights	Yes	Yes	Yes	Yes
Regional controls	None	All	None	All
N	704	704	704	704
R^2	0.774	0.826	0.896	0.912
F-stat	973.078	250.568	749.188	251.243

Note: The table presents IV estimates of the second-step regression for regional productivity differences for higher-skilled (Panel A) and lower-skilled (Panel B) natives. Region-time fixed-effects estimated in Equation 1 are used as the dependent variable. The region-time fixed-effects are estimated using only the higher-skilled native population (Columns 1 and 2) or the lower-skilled native population (Columns 3 and 4). The logarithm of the migrant share, the share of higher-skilled migrants and the share of higher-skilled natives are the independent variables. The migrant share variable is instrumented. Additional and undisplayed controls are Population Density, Domestic Market Potential and Foreign Market Potential. All the specifications are weighted by the number of employed natives in the region. Time fixed-effects are present to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively. The analysis considers 88 regions over eight years, yielding 704 observations.
Source: OECD calculations based on MADIP (accessed June 2023).

Uneven effects across places

Migration may have uneven impacts on native workers depending on regional conditions. For example, if migrants disproportionally benefit places with lower productivity, it may contribute to convergence in productivity levels across places. Similarly, places with few migrants may benefit more from an increase in the migrant population than places with a large migrant population. In this sense, the analysis needs to ensure that certain types of places do not drive the estimated average effects.

The impact of migration on productivity in different regional subsamples might unveil regional differences and nuances the analysis. Table 6 presents the estimates of the preferred specification for subgroups of regions disaggregated based on their migrant share (Panel A) and productivity (Panel B). Both panels follow the same structure. Colum 1 shows the baseline result, i.e., the estimates for the whole universe of considered regions. Columns 2 and 3 divide the sample along the median of the migrant share or estimated productivity, with Column 2 displaying the effect on all regions above the respective median. Column 4 excludes all regions in the bottom or top 5th percentile. The estimations in Columns 6 and 7 exclude areas above the 90th percentile and outside the 25th and 75th percentile. Further, Panel C presents additional regressions, excluding Western Australia or regions located in the five and three largest metropolitan areas.

While the presence of migrants boosts productivity everywhere, human capital effects require a minimum of productivity and migration

Migration enhances regional productivity regardless of the initial regional migrant share (Panel A). The positive effect of migration on productivity is highly significant and of similar size for all types of regions and is robust to excluding outliers. In contrast, the positive effect of human capital among migrants is less robust. The effect of more higher-skilled migrants among migrants is positive for all specifications except for regions with initially below-median migrant populations. In addition, not all positive estimates are statistically significant.

The positive impact of migration on productivity does not depend on the initial productivity advantages of the region, yet the positive effect of migrants' human capital requires a minimum of initial regional productivity advantages and share of migrants in the area to materialise. Although regional migrant share has a positive and significant impact on all subgroups, the effect is slightly higher for more productive and migrant-hosting regions (Columns 2, Panels A and B) than those in less productive regions and with below-median shares of migrants (Columns 3, Panels A and B). The estimated effect of migrants' human capital is positive and significant for all specifications except for regions with below-median migrant populations and below-median productivity. This is also reflected in Panel C, where the estimates for human capital among migrants are non-significant when excluding the largest three and five metropolitan areas, which are characterized by high migrant shares and high productivity.

Table 6. Uneven effects across places

Second-step regression for regional productivity differences disaggregated by initial regional productivity differences and regional migrant shares, 2011-2018

	Baseline	> Median	< Median	5th -95th	< 90th	25th -75th
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Migrant Share						
Log Migrant Share	0.13545***	0.20631***	0.17661***	0.14534***	0.15964***	0.22780***
	(0.017)	(0.044)	(0.036)	(0.019)	(0.021)	(0.066)
Log Share Higher-Skilled Migrants	0.10413**	0.17985***	-0.07031	0.13343**	0.13178**	0.28930*
	(0.052)	(0.064)	(0.091)	(0.057)	(0.057)	(0.158)
Log Share Higher-Skilled Natives	-0.05691	-0.11800*	0.12457	-0.07612	-0.07797	-0.25016
	(0.060)	(0.065)	(0.106)	(0.066)	(0.067)	(0.194)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Pop. weights	Yes	Yes	Yes	Yes	Yes	Yes
Regional controls	All	All	All	All	All	All
N	704	352	352	632	634	352
R^2	0.901	0.926	0.849	0.894	0.897	0.842
F-stat	254.251	79.555	130.105	204.332	185.110	63.614
Panel B: Productivity Advantages						
	Baseline	> Median	< Median	5th -95th	< 90th	25th -75th
	(1)	(2)	(3)	(4)	(5)	(6)
Log Migrant Share	0.13483***	0.13553***	0.09375***	0.12272***	0.13258***	0.09840***
	-0.017	(0.016)	(0.016)	(0.016)	(0.017)	(0.013)
Log Share Higher-Skilled Migrants	0.34375**	0.09060*	0.05657	0.09525*	0.11951**	0.08616**
	-0.145	(0.051)	(0.053)	(0.051)	(0.048)	(0.040)
Log Share Higher-Skilled Natives	-0.23307	-0.03798	-0.03139	-0.04774	-0.07300	-0.06506

	-0.179	(0.056)	(0.060)	(0.060)	(0.057)	(0.049)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Pop. weights	Yes	Yes	Yes	Yes	Yes	Yes
Regional controls	All	All	All	All	All	All
N	704	352	352	632	633	352
R²	0.902	0.800	0.676	0.871	0.889	0.654
F-stat	276.366	257.047	266.775	252.730	248.601	302.605
Panel C: Outliers						
	Baseline	w/o WA	NoTop3	NoTop5	Top5	
	(1)	(2)	(3)	(4)	(5)	
Log Migrant Share	0.13483***	0.12546***	0.13149***	0.12704***	0.11440***	
	-0.017	(0.021)	(0.022)	(0.040)	(0.021)	
Log Share Higher-Skilled Migrants	0.34375**	0.10777**	0.06608	0.08709	0.10350***	
	-0.145	(0.052)	(0.100)	(0.108)	(0.035)	
Log Share Higher-Skilled Natives	-0.23307	-0.06460	-0.01424	0.00008	-0.07313	
	-0.179	(0.060)	(0.098)	(0.114)	(0.046)	
Time FE	Yes	Yes	Yes	Yes	Yes	
Pop. weights	Yes	Yes	Yes	Yes	Yes	
Regional controls	All	All	All	All	All	
N	704	624	440	360	344	
R²	0.902	0.910	0.846	0.820	0.952	
F-stat	276.366	215.490	204.403	85.341	115.895	

Note: The table presents IV estimates of the second-step regression for regional productivity differences for different subgroups of the migrant share (Panel A) and productivity advantages (Panel B). Region-time fixed-effects estimated in Equation 1 are used as the dependent variable. The logarithm of the migrant share, the share of higher-skilled migrants and the share of higher-skilled natives are the independent variables. The migrant share variable is instrumented. Additional and undisplayed controls are Population Density, Domestic Market Potential and Foreign Market Potential. All the specifications are weighted by the number of employed natives in the considered regions. Time fixed-effects are present to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively. The analysis considers 88 regions over eight years, yielding 704 observations.
Source: OECD calculations based on MADIP (accessed June 2023).

6 Concluding remarks

The analysis uses individual-level administrative data covering almost the entire Australian population. The paper presents the first evidence of the impact of migration on regional labour productivity in Australia. It examines the effects of migrants and their skill composition on productivity while accounting for other regional characteristics and estimation concerns.

Using a standard two-step regression approach and instrumental variable strategy to address the endogeneity concerns, it shows that the presence of migrants is associated with higher labour productivity differences across Australian regions. Natives of all skill groups living in all types of regions benefit from these positive gains due to the presence of migrants. Moreover, these positive gains are further amplified when migrants are more skilled. These additional gains, however, materialise only in already productive regions and regions with above-median migrant shares.

While this analysis contributes to the evidence base on migration's contribution to regional productivity in Australia, it also leaves scope for future research. Such work could examine how the regional industrial structure affects regions' capacity to benefit from migrants and their skills in boosting regional productivity. Furthermore, understanding how migration affects firms could complement this evidence and provide insights into the underlying mechanisms driving the positive effects found in this analysis. Finally, future analysis could examine how migration's contribution to regional productivity may depend on the specific visa status or the length of stay of migrants. Such analysis may help to enrich further the empirical evidence on the regional economic effects of migration in Australia and support more effective policy design.

References

Ahrend, R. et al. (2014), “What Makes Cities More Productive? Evidence on the Role of Urban Governance from Five OECD Countries”, *OECD Regional Development Working Papers*, No. 2014/5, OECD Publishing, Paris, https://doi.org/10.1787/5jz432cf2d8p-en. [56]

Alesina, A., J. Harnoss and H. Rapoport (2016), “Birthplace diversity and economic prosperity”, *Journal of Economic Growth*, Vol. 21/2, pp. 101-138, https://doi.org/10.1007/s10887-016-9127-6. [8]

Altonji, J. and D. Card (1991), “The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives”, *NBER Working Paper*, Vol. Immigratio/January, pp. 201-234, https://doi.org/10.1017/CBO9781107415324.004. [37]

Bahar, D. et al. (2022), “Migration and post-conflict reconstruction: The effect of returning refugees on export performance in the former Yugoslavia”, *The Review of Economics and Statistics*, https://doi.org/10.1162/rest_a_01165. [15]

Bahar, D. and H. Rapoport (2018), “Migration, knowledge diffusion and the comparative advantage of nations”, *The Economic Journal*, Vol. 128/612, pp. 273-305. [17]

Bakens, J., P. Mulder and P. Nijkamp (2013), “Economic impacts of cultural diversity in the Netherlands: Productivity, utiliy and sorting”, *Journal of Regional Science*, Vol. 53/1, pp. 8-36, https://doi.org/10.1111/jors.12012. [58]

Borjas, G. (2013), *Labor Economics*, http://www.mhhe.com (accessed on 20 April 2023). [27]

Borusyak, K., P. Hull and X. Jaravel (2022), “Quasi-Experimental Shift-Share Research Designs”, *The Review of Economic Studies*, Vol. 89, pp. 181-213, https://doi.org/10.1093/restud/rdab030. [40]

Bosetti, V., C. Cattaneo and E. Verdolini (2015), “Migration of skilled workers and innovation: A European Perspective”, *Journal of International Economics*, Vol. 96/2, pp. 311-322, https://doi.org/10.1016/J.JINTECO.2015.04.002. [6]

Bove, V. and L. Elia (2017), “Migration, Diversity, and Economic Growth”, *World Development*, Vol. 89, pp. 227-239, https://doi.org/10.1016/J.WORLDDEV.2016.08.012. [9]

Card, D. (2001), “Immigrant inflows, native outflows, and the local labor market impacts of higher immigration”, *Journal of Labor Economics*, Vol. 19/1, https://doi.org/10.1086/209979. [36]

Ciccone, B. and R. Hall (1996), “Productivity and the Density of Economic Activity”, *The American Economic Review*, Vol. 86/1, pp. 54-70. [34]

Combes, P., S. Démurger and S. Li (2015), “Migration externalities in Chinese cities”, [20] *European Economic Review*, Vol. 76, pp. 152-167, https://doi.org/10.1016/j.euroecorev.2015.02.004.

Combes, P. et al. (2019), “Unequal migration and urbanisation gains in China”, *Journal of Development Economics*, https://doi.org/10.1016/j.jdeveco.2019.01.009. [44]

Combes, P., G. Duranton and L. Gobillon (2008), “Spatial wage disparities: Sorting matters!”, *Journal of Urban Economics*, Vol. 63/2, pp. 723-742, https://doi.org/10.1016/j.jue.2007.04.004. [5]

Combes, P., G. Duranton and H. Overman (2005), “Agglomeration and the adjustment of the spatial economy[§]”, *Papers in Regional Science*, Vol. 84/3, pp. 311-349, https://doi.org/10.1111/j.1435-5957.2005.00038.x. [50]

Combes, P. and L. Gobillon (2015), *The Empirics of Agglomeration Economies*, Elsevier B.V., https://doi.org/10.1016/B978-0-444-59517-1.00005-2. [28]

De La Roca, J. and D. Puga (2016), “Learning by Working in Big Cities”, *The Review of Economic Studies*, Vol. 84/1, pp. 106-142, https://doi.org/10.1093/restud/rdw031. [53]

Docquier, F. et al. (2020), “Birthplace diversity and economic growth: evidence from the US states in the Post-World War II period”, *Journal of Economic Geography*, Vol. 20/2, pp. 321-354. [14]

Duranton, G. and D. Puga (2001), “Nursery cities: Urban diversity, process innovation, and the life cycle of products”, *American Economic Review*, Vol. 91/5, pp. 1454-1477, https://doi.org/10.1257/aer.91.5.1454. [30]

Dustmann, C., F. Fabbri and I. Preston (2005), “The Impact of Immigration on the British Labour Market”, *The Economic Journal*, Vol. 115/507, pp. F324-F341, https://doi.org/10.1111/j.1468-0297.2005.01038.x. [42]

Elias, A. and Y. Paradies (2016), “The regional impact of cultural diversity on wages: evidence from Australia”, *IZA Journal of Migration 2016 5:1*, Vol. 5/1, pp. 1-24, https://doi.org/10.1186/S40176-016-0060-4. [24]

Epstein, G. and I. Gang (2010), “Migration and Culture”, *Bar-Ilan University Department of Economics Research Paper* 17, https://doi.org/10.2139/ssrn.1868830 (accessed on 24 September 2021). [39]

Fassio, C., S. Kalantaryan and A. Venturini (2015), “Human Resources and Innovation: Total Factor Productivity and Foreign Human Capital”, *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2631086. [7]

Fassio, C., F. Montobbio and A. Venturini (2019), “Skilled migration and innovation in European industries”, *Research Policy*, Vol. 48/3, pp. 706-718, https://doi.org/10.1016/J.RESPOL.2018.11.002. [18]

Felbermayr, G. and F. Toubal (2012), “Revisiting the trade-migration nexus: Evidence from new OECD data”, *World Development*, Vol. 40/5, pp. 928-937, https://doi.org/10.1016/j.worlddev.2011.11.016. [19]

Gal, P. (2013), “Measuring Total Factor Productivity at the Firm Level using OECD-ORBIS”, *OECD Economics Department Working Papers*, No. 1049, OECD Publishing, Paris, [26] https://doi.org/10.1787/5k46dsb25ls6-en.

Goldsmith-Pinkham, P., I. Sorkin and H. Swift (2020), “Bartik Instruments: What, When, Why, and How”, *American Economic Review*, Vol. 110/8, pp. 2586-2624, https://doi.org/10.1257/aer.20181047. [41]

Gross, D. and N. Schmitt (2003), “The role of cultural clustering in attracting new immigrants”, *Journal of Regional Science*, Vol. 43/2, pp. 295-318, https://doi.org/10.1111/1467-9787.00300. [38]

Hering, L. and S. Poncet (2010), “Market access and individual wages: Evidence from China”, *Review of Economics and Statistics*, Vol. 92/1, pp. 145-159, https://doi.org/10.1162/rest.2009.11418. [31]

Hjort, J. (2014), “Ethnic Divisions and Production in Firms *”, *The Quarterly Journal of Economics*, Vol. 129/4, pp. 1899-1946, https://doi.org/10.1093/qje/qju028. [23]

Jaeger, D., J. Ruist and J. Stuhler (2018), *Shift-share instruments and the impact of immigration*, National Bureau of Economic Research, https://doi.org/10.3386/w24285. [43]

Kemeny, T. and A. Cooke (2018), “Spillovers from immigrant diversity in cities”, *Journal of Economic Geography*, Vol. 18/1, pp. 213-245, https://doi.org/10.1093/jeg/lbx012. [47]

Kerr, S. and W. Kerr (2017), “Immigrant Entrepreneurship”, https://www.nber.org/books-and-chapters/measuring-entrepreneurial-businesses-current-knowledge-and-challenges/immigrant-entrepreneurship (accessed on 25 November 2022). [16]

Krugman, P. and A. Venables (2006), “Globalization and the Inequality of Nations”, *The Quarterly Journal of Economics*, Vol. 110/4, pp. 857-880, https://doi.org/10.2307/2946642. [32]

McCann, P. (2023), *How Have Place-Based Policies Evolved to Date and What Are They For Now?*, https://www.oecd.org/regional/how-have-place-based-policies-evolved-to-date-and-what-are-they-for-now.pdf. [46]

Mitaritonna, C., G. Orefice and G. Peri (2017), “Immigrants and firms’ outcomes: Evidence from France”, *European Economic Review*, Vol. 96, pp. 62-82, https://doi.org/10.1016/j.euroecorev.2017.05.001. [13]

Moretti, E. (2011), “Local Labor Markets”, in *Handbook of Labor Economics*, Elsevier, https://doi.org/10.1016/s0169-7218(11)02412-9. [52]

Moretti, E. (2004), “Chapter 51 Human capital externalities in cities”, in *Handbook of Regional and Urban Economics, Cities and Geography*, Elsevier, https://doi.org/10.1016/s1574-0080(04)80008-7. [51]

Moulton, B. (1990), “An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units”, *The Review of Economics and Statistics*, Vol. 72/2, p. 334, https://doi.org/10.2307/2109724. [33]

Nguyen, T. and C. Parsons (2018), “The Labour Market Impact of a High Skilled Migration Wave”, *Working Paper*. [48]

OECD (2024), “Migration and regional innovation in Australia”, *OECD Regional Development Papers*, OECD Publishing, Paris, https://doi.org/10.1787/267a6231-en. [3]

OECD (2023), *International Migration Outlook 2023*, OECD Publishing, Paris, https://doi.org/10.1787/b0f40584-en. [4]

OECD (2023), *Mining Regions and Cities Case of the Pilbara, Australia*, OECD Rural Studies, OECD Publishing, Paris, https://doi.org/10.1787/a1d2d486-en. [45]

OECD (2023), "Regional productivity, local labour markets, and migration in Australia", *OECD Regional Development Papers*, No. 39, OECD Publishing, Paris, https://doi.org/10.1787/3cc8f669-en. [1]

OECD (2023), "The impact of migration on regional labour markets in Australia", *OECD Regional Development Papers*, OECD Publishing, Paris, https://doi.org/10.1787/267a6231-en. [2]

OECD (2022), *The Contribution of Migration to Regional Development*, OECD Regional Development Studies, OECD Publishing, Paris, https://doi.org/10.1787/57046df4-en. [10]

OECD (2020), *Enhancing Productivity in UK Core Cities: Connecting Local and Regional Growth*, OECD Urban Policy Reviews, OECD Publishing, Paris, https://doi.org/10.1787/9ef55ff7-en. [57]

OECD (2001), *Measuring Productivity - OECD Manual: Measurement of Aggregate and Industry-level Productivity Growth*, OECD Publishing, Paris, https://doi.org/10.1787/9789264194519-en. [25]

Ortega, F. and G. Peri (2011), "The Aggregate Effects of Trade and Migration: Evidence from OECD Countries". [21]

Ottaviano, G. and G. Peri (2005), "The economic value of cultural diversity: evidence from US cities", *Journal of Economic Geography*, Vol. 6/1, pp. 9-44, https://doi.org/10.1093/jeg/lbi002. [55]

Özgüzel, C. (2022), "Agglomeration effects in a developing economy: evidence from Turkey", *Journal of Economic Geography*, https://doi.org/10.1093/JEG/LBAC035. [29]

Özgüzel, C. (2020), "Agglomeration economies in Great Britain", *OECD Regional Development Working Papers*, No. 2020/04, OECD Publishing, Paris, https://doi.org/10.1787/3aa63b9a-en. [49]

Parrotta, P., D. Pozzoli and M. Pytlikova (2014), "Labor diversity and firm productivity", *European Economic Review*, Vol. 66, pp. 144-179, https://doi.org/10.1016/j.euroecorev.2013.12.002. [22]

Peri, G. (2016), "Immigrants, Productivity, and Labor Markets", *Journal of Economic Perspectives*, Vol. 30/4, pp. 3-30, https://doi.org/10.1257/jep.30.4.3. [35]

Peri, G. (2012), "The effect of immigration on productivity: Evidence from US states", *Review of Economics and Statistics*, Vol. 94/1, pp. 348-358. [12]

Peri, G. and C. Sparber (2009), "Task Specialization, Immigration, and Wages", *American Economic Journal: Applied Economics*, Vol. 1/3, pp. 135-69, https://doi.org/10.1257/APP.1.3.135. [11]

Quintero, L. and M. Roberts (2023), "Cities and productivity: Evidence from 16 Latin American and Caribbean countries", *Journal of Urban Economics*, Vol. 136, p. 103573, [54] https://doi.org/10.1016/j.jue.2023.103573.

Annex A. Data sources

Multi-Agency Data Integration Project (MADIP)

The Multi-Agency Data Integration Project (MADIP) dataset by the Australian Bureau of Statistics (ABS) is an individual-level panel dataset that provides longitudinal information for more than 27 million individual records between 2011 and 2020. MADIP combines administrative information from different departments, such as the Australian Taxation Office (ATO), the Department of Education, the Department of Health and Aged Care, the Department of Social Services, Services Australia, and the Department of Home Affairs. The availability of the dataset is subject to the agreement of the data custodians of the individual agencies and depends on the individual research question. In addition to administrative data, the MADIP includes one of the quinquennial Australian Census of Housing and Population. Besides the MADIP core data, this analysis relies on tax data by ATO, migration data by the Department of Home Affairs, and the Census 2018. The following subsections describe the individual components of the dataset.

MADIP core data

The MADIP core dataset is at the centre of every analysis using MADIP data. It contains demographic information like date of birth, gender, and date of death, as well as location information, on all residents in Australia. Moreover, the dataset includes a *spine* ID integral to merging the individual datasets from different agencies. By default, the dataset covers every Australian resident recorded in either *Social Security and Related Information*, *Personal Income Tax* data, or *Medicare Benefits Schedule* data between 2006 and 2020, resulting in a total of 27.1 million individual records. However, not every recorded person is listed in every individual dataset. For instance, income tax data are not available if the person has never reported taxes (e.g., children).

The geographical information is available at different granularity levels, including SA4, SA3, and SA2. Given the overwhelming coverage of the Australian population, the data are expected to be representative at every geographical level. Location information is distinguished by residential and mail address. For migrants, the business address is also reported. In the analysis, the individual location information is based on the residential address or mail address, depending on data availability.

MADIP is expected to cover the vast majority of Australian citizens and residents due to the combination of medicare, social benefits, and income tax records. According to the ABS, the following groups are potentially underrepresented: i) recently arrived migrants without Medicare, ii) non-earning partners and family members of working visa holders, iii) non-earning foreign students, iv) military personnel, v) prisoners, vi) recently born individuals, not yet included in the Medicare Benefits Schedule.

Census of Population and Housing 2016

The Australian Census of Population and Housing is conducted every five years and includes, among others, information on educational attainment, employment and work, family, and personal characteristics. This paper uses data from the Census wave of 2016, the latest available for research, linked to the MADIP universe. [17] Due to Australian data confidentiality rules, only one Census wave at a time can be used in the MADIP environment. The Census data refers to the data collected on the 9th of August, 2016.

In Australia, participating in the census is mandatory for Australian residents, with very few exceptions. The ABS linked 20.7 million records of the 2016 Census to the MADIP data, which corresponds to 88% of all collected Census records in 2016. According to the ABS, the following groups are not within the scope of the Census: i) Australians overseas, ii) residents for less than six months, iii) visitors, and iv) diplomatic personnel and their families. The paper retrieves information on age, occupation, industry, and country of birth from the Census.

Australian Taxation Office

The Australian Taxation Office provides administrative information on all employed individuals in Australia based on official tax returns. The dataset covers around 16.7 million individual records, including everyone with a tax return in Australia in at least one year from 2010/2011 – 2017/2018. The Australian financial tax year ranges from July until June of the following year. However, in order to combine the data with other datasets, the tax records are assumed to follow the calendar year (January-December rather than July-June).[18] Data span from wages, total income, and insurance payments to job sector information. The variables of interest to this analysis are age, individual wage/salary, the main salary or wage occupation code, and industry. Employed individuals with an income below the threshold imposed by the ATO and, hence, without a tax record, are not considered in the data. This also includes most migrants on a working holiday maker (WHM) visa.

Department of Home Affairs

The *Department of Home Affairs* provides administrative data on the native and migrant populations. The data includes every individual (native- or foreign-born) who crossed the border of Australia between 1990 and 2020. The dataset is used to retrieve information on the country of birth, date (month and year) of birth, and gender. Visa information is not available for all migrants. Moreover, due to changes in the visa status after arriving in Australia, the visa information might not be reliable for all migrants.

Historic Census

As discussed in Section 4, the identification strategy of the paper requires the use of a historical instrument based on the settlement patterns of migrants in the past. The information on the past settlement patterns is obtained from Census data from 1981, 1986, 1991, and 2001. The historical data have been adjusted to 2016 borders by the ABS and provide data on the total employed population of the respective year disaggregated by country of birth, the industry of employment, and highest post-school qualification. The country of birth consists of 60 national groupings.

[17] At the time of the analysis, the 2021 Census (published in summer and fall 2022) was not yet available to a sufficient extent. Further due to the disruptive effect of COVID-19 and the substantial travel restrictions for travel and immigration to Australia, the results of the analysis would not necessarily represent the situation in Australia.

[18] This means that the tax return for the financial year 2011/12 is treated as the tax return for the year 2012. The age retrieved from the ATO, is adjusted accordingly.

Table A. 1 Migrant decomposition

Share of employed migrants among the employed migrant population, 1981

Country groupings	Share of total foreign-born employed population
Albania, Bulgaria & Romania	0.32%
Argentina & Uruguay	0.51%
Austria	0.87%
Bangladesh	0.03%
Belgium	0.15%
Brazil	0.05%
Cambodia, Laos & Myanmar	0.42%
Canada	0.46%
Chile	0.40%
China	0.99%
Colombia, Ecuador & Peru	0.13%
Cyprus	0.82%
Denmark, Finland, Norway & Sweden	0.87%
Egypt	1.20%
Fiji	0.27%
Former Czechoslovakia	0.63%
Former USSR	1.67%
Former Yugoslavia	5.63%
France	0.42%
Germany	4.52%
Greece	6.00%
Hong Kong & Macau	0.41%
Hungary	1.11%
India	1.44%
Indonesia & Timor-Leste	0.54%
Iran	0.11%
Iraq	0.11%
Ireland	1.62%
Israel	0.21%
Italy	10.97%
Japan	0.27%
Kenya, Malawi, Zimbabwe, Tanzania, Uganda & Zambia	0.27%
Korea	0.13%
Lebanon	1.48%
Malaysia & Brunei	0.76%
Malta	2.32%
Mauritius	0.33%
Mexico	0.01%
Netherlands	3.92%
New Zealand	4.63%
Other Africa	0.37%
Other Middle East	0.05%
Pakistan	0.08%
Papua New Guinea	0.25%
Philippines	0.51%
Poland	2.18%
Portugal	0.40%
Singapore	0.30%

South Africa & Namibia	0.73%
Spain	0.54%
Sri Lanka	0.56%
Switzerland	0.26%
Syria	0.11%
Taiwan	0.03%
Thailand	0.09%
Türkiye	0.66%
United Kingdom	33.58%
United States of America	0.96%
Vietnam	0.82%
All other countries	0.54%

Note: Share of the foreign-born employed population in 1981. The countries of origin are aggregated to 60 national groupings. Grouping was conducted by the ABS and refers to the international borders of 1981.
Source: Australian Bureau of Statistics (ABS).

Annex B. Variables and controls

The analysis requires the use of a large set of variables at the individual and regional levels. This section provides details on the construction of all variables used in both steps of the analysis.

First step of the analysis

The analysis requires individual annual wages for the productivity analysis. The sample includes all employed residents with at least one tax return between 2011-2018. In order to gather all relevant information on the employed native population, the individual tax data by the Australian Taxation Office is merged with the MADIP core data. Further, individual-level data from the 2016 Census is merged to reduce the exclusion of observations with incomplete information. The sample of the employed native population contains information on personal wages, age, industry and occupation codes (all ATO), SA4 location, gender, and age (MADIP core). The Census data are used to complement missing information on age, gender and occupation.

Table A. 2 Variables of the first step

Variable name	Formula	Definition
Wages		Individual yearly wages of employed natives
Male		Indicator function for gender (Male =1, Female = 0)
Age		Individual age
Age[2]	Age*age	Individual age squared
Occupation		Occupational level according to ANZSCO Major Group classifications (1220.0 - ANZSCO -- Australian and New Zealand Standard Classification of Occupations, 2013, Version 1.3 (abs.gov.au))Occupational level according to ANZSCO Major Group classifications (1220.0 - ANZSCO -- Australian and New Zealand Standard Classification of Occupations, 2013, Version 1.3 (abs.gov.au))
Industrial specialisation	$\frac{workers_{s,j,t}}{workers_{j,t}}$	Share of employed workers (native- and foreign-born) in sector *s*, region *j*, and year *t* over all employed workers in region *j* and year *t*.[19]

Second step of the analysis

The second step of the analysis requires the construction of variables that capture the regional share of migrants, their skill composition, and other regional characteristics.

[19] Only employed workers in the sample are considered.

Table A. 3 Variables of the second step

Variable name	Formula	Definition
Regional labour productivity differences of native workers		Region-year fixed-effects from first-step regression
Migrant share	$\log \frac{Migrants_{j,t}}{Migrants_{j,t} + Natives_{j,t}}$	Share of foreign-born in region *j* and year *t* over the foreign- and native-born population in region *j* and year *t*.
Share of Higher-Skilled Migrants	$\log \frac{Migrants_{j,t}^{high}}{Migrants_{j,t}^{high} + Migrants_{j,t}^{low}}$	Share of higher-skilled foreign-born in region *j* and year *t* over the higher- and lower-skilled foreign-born population in region *j* and year *t*.
Share of Higher-Skilled Natives	$\log \frac{Natives_{j,t}^{high}}{Natives_{j,t}^{high} + Natives_{j,t}^{low}}$	Share of higher-skilled native-born in region *j* and year *t* over the higher- and lower-skilled native-born population in region *j* and year *t*.
Density of Employed Workers	$\frac{workers_{j,t}}{area\ size_j}$	The total of workers (foreign- and native-born) in region *j* and year *t* over the area (in km^2) in region *j*.
Domestic Market Potential	$DMP_{rt} = \sum_{k \neq r}^{n} \frac{Density_{kt}}{Distance_{kr}}$	The sum of the employment density of other regions *k* at time *t* over the road distance of the regions k to region r and represents the inland market potential.
Foreign Market Potential		Travel duration (by car) between the centroid of a region *r* and the closest major mixed cargo naval port.

Migrant share: The regional migration share is calculated using administrative data from the Department of Home Affairs and the 2016 Census. As the country of birth is not part of the MADIP core data, and the Department of Home Affairs data only records this information for residents who cross the national border within the observation period, the variable needs to be imputed. Hence, the analysis combines information from the Department of Home Affairs data and the 2016 Census. Since the Department of Home Affairs provides administrative data, this source is prioritized over complementary information from the 2016 Census. Records without sufficient country of birth information are excluded. The nationality information is further used to define the country of birth of employed workers and to restrict the sample to native-born.

Naitves' and Migrants' Human Capital: The share of higher-skilled native- and foreign-born is calculated using the sample of employed Australian residents. In the absence of sufficient educational information, the analysis uses the occupation level following the ANZSCO. Workers employed in the Major Groups "Managers" and "Professionals" are considered higher-skilled. This follows the ABS, which assigns the highest "Predominant Skill Level" to "Managers" and "Professionals".

Employed Density: The employed population information is sourced from the sample, while the land area size is provided by the ABS. Due to changes in the employed regional population over time, density is time-variant.

Domestic Market Potential: DMP aggregates the market potential between the considered region and every other region in Australia. The road distance is calculated from the centroid of the region using Mapbox Directions API.

Foreign Market Potential: FMP is the travel duration (by car) between the centroid and the closest major port in Australia out of the 11 nationally significant major cargo naval ports identified by the Australian Department of Infrastructure, Transport, Regional Development, Communications and the Arts.

Annex C. Validity of the instrument variable

Past settlement patterns are not correlated with past regional characteristics

This test, in support of the exogeneity of the shares, checks whether baseline migrant shares in 1981 are associated with regional characteristics, as they, in turn, might be correlated with current levels of labour productivity (Goldsmith-Pinkham, Sorkin and Swift, 2020[41]). Implementing this test requires checking that the initial shares of the top origin countries that explain most of the variation during the 2011-2018 period, i.e., India, China, Philippines, and Korea[20], are not associated with regional characteristics in 1981.

Origin country shares are mostly not associated with regional characteristics in 1981. The following table shows the results of regressions of top origin-specific shares on regional variables and industry composition in 1981. These include shares of highly educated workers, the distribution of workers across sectors, and the logarithm of wages and employment. Columns 1 to 4 highlight that out of 20 coefficients, only two are statistically significant. Additionally, as shown by Column 5, origin shares of these four top nationalities combined together are not correlated with regional characteristics. In consequence, the instrument is not correlated with regional characteristics in 1981. Columns 6 to 8 assess the association of regional variables in 1981 with the predicted migrant increase, i.e., the instrument. Neither the share of highly educated individuals nor the sectoral shares or wage or employment levels are correlated with the instrument. Taken together, these results provide further support to the assumption that the instrument is affecting productivity levels only through its effect on migration flows.

Table A. 4 Explanatory variables in 1981

	India	China	Philippines	Korea	Top 4	Δ Predicted Migrant		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share high-educated	0.038	-0.116*	-0.127**	-0.120	-0.205	0.008	0.117	-0.758
	(0.067)	(0.063)	(0.056)	(0.144)	(0.135)	(0.286)	(0.367)	(0.731)
Share primary sector	0.023	0.010	0.027	0.047	0.059		-0.036	-0.307
	(0.028)	(0.026)	(0.023)	(0.060)	(0.056)		(0.233)	(0.276)
Share secondary sector	0.042	-0.021	0.028	0.043	0.050		0.380	0.352
	(0.035)	(0.033)	(0.029)	(0.075)	(0.071)		(0.299)	(0.299)
Wage (logarithm)	0.100	-0.084	-0.066	-0.015	-0.050			0.256
	(0.064)	(0.061)	(0.054)	(0.139)	(0.130)			(0.395)
Employment (logarithm)	-0.025	0.053	0.048	0.009	0.076			0.568
	(0.037)	(0.035)	(0.032)	(0.081)	(0.076)			(0.612)

Note: The table presents results based on estimating each dependent variable on a set of regional characteristics, including the share of highly-educated workers, the shares of workers in the primary or secondary sector (the tertiary sector is omitted to avoid perfect multicollinearity), and wages and employment in logarithm, across regions in 1981. Columns 1 to 4 use as dependent variables the share of Indian, Chinese, Filipino or Korean individuals in each region in 1981, respectively. Column 5 sums the shares of these four nationalities within each region. Columns 6 to 8 use the predicted change in migrant population over the total baseline population from 2011 to 2016. Standard errors in parenthesis. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% levels, respectively.
Source: OECD calculations based on MADIP and IPGOD (accessed July 2023).

[20] These are the four nationalities that contribute the most to the increase in migration during the 2011-2018 period.

Annex D. Instrumental variable for population density

The endogeneity concerns stemming from population density are addressed using an instrumental variable approach. As discussed in Section 4, the relationship between population density and productivity might suffer from reverse causality as people tend to settle in more productive regions. This settlement behaviour results in a higher population density in more productive places, which biases the analysis. A well-established approach to eliminate such spurious variation is to use past settlement behaviours, which are unrelated to current regional productivity levels. Therefore, the population density is calculated using the headcounts of employed workers in 1981. This instrumental variable is time-invariant. Hence, variation over time comes from changes in the endogenous variable containing the actual population density in the years of analysis.

$$\widehat{Population\ Density}_j = \frac{Workers_j^{1981}}{Area\ size_j} \tag{8}$$

Annex E. Explanatory power with individual fixed-effects

Table A. 5 The explanatory power of different specifications with individual fixed-effects

R^2 for individual wages regressed on different specifications, including individual fixed-effects, 2011-2018

Row	Specification	R^2
(1)	Individual Characteristics (incl. Individual FE)	73.7%
(2)	Industry Characteristics	6.6%
(3)	Regional Effects	1.4%
(4)	Regional Effects + Individual Characteristics (incl. Individual FE)	73.7%
(5)	Regional Effects + Industry Characteristics	7.8%
(6)	Individual Characteristics + Industry Characteristics (incl. Individual FE)	73.7%
(7)	All three sets of variables (incl. Individual FE)	73.7%
	Number of observations	25 845 298

Note: The table presents the adjusted R^2 from an OLS estimation where the dependent variable is the natural logarithm of individual yearly wage. Different rows include different sets of independent variables. In row 1, individual observable characteristics (sex, age, age², occupation) and unobservable characteristics (individual FEs) are included. In row 2, industrial sector dummies (ANZSIC rev. 1 classification) are included. In row 3, industrial specialisation and regional dummies are considered. Rows 4-7 combine the independent variables from rows 1-3. The analysis is based on 25 845 298 observations. Standard errors are clustered at the regional level in all rows.
Source: OECD calculations based on MADIP (accessed June 2023).

Annex F. First stage results and additional tests

Table A. 6 First stage of the baseline 2SLS regression

	(1)
Dependent variable:	Log Mig. Share
IV Log Migrant Share	0.77068***
	(0.048)
Log DMP	0.08973***
	(0.027)
Log FMP	0.00997
	(0.025)
Share Higher-Skilled Migrants	0.60763***
	(0.214)
Share Higher-Skilled Natives	-0.60280***
	(0.187)
Total Emp. Density	-0.00002
	(0.000)
Constant	-1.20135***
	(0.369)
Time FE	Yes
Pop. weights	Yes
N	704
R^2	0.919

Note: The table presents the first stage from the 2SLS estimation of the second-step regression for regional productivity differences. The endogenous variable, the logarithm of the migrant share, is the dependent variable. The instrumental variable of the logarithm of the migrant share (see Section 4), the Domestic Market Potential (DMP), the Foreign Market Potential (FMP), the share of higher-skilled migrants, the share of higher-skilled natives and population density are the independent variables. The regression is weighted by the number of employed natives in the region. Time fixed-effects are applied to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively.
Source: OECD calculations based on MADIP (accessed June 2023).

Annex G. Robustness of the baseline results

Presenting alternative definitions of key variables and the empirical strategy validates the robustness of the estimation. Depending on how migration is measured, the estimated impact of migrants on regional productivity might vary. Similarly, different identification strategies address other sources of potential endogeneity and may lead to different conclusions. To ensure that the findings are not sensitive to such empirical choices, Table A. 7 and Table A. 8 present alternative measures and IV strategies, respectively.

The findings are robust to the use of alternative measures

The definition for measuring the regional migrant presence or the population density may affect the results. To ensure that the results do not depend on the definitions, alternative definitions of the migrant share used in the literature are presented as robustness tests. Table A. 7 presents the 2SLS regression results using the preferred specification (Table 3, Panel B, Column 5) with alternative measures.

The first column reproduces the baseline result, which measures the presence of migrants as the number of migrants relative to the total regional population (in logarithm) for comparison. Column 2 follows the same definition but only considers the working-age population of migrants and natives (15-64 years old). Column 3 follows the baseline definition but uses levels of the migrant share rather than the natural logarithm. Column 4 follows Combes et al. (2015[20]) and defines the migrant share as the logarithm of 1/(1-Migrant Share), where the migrant share is defined as the number of migrants over the total population. Column 5 uses the same computation as Column 4 yet uses only the working-age population.

The impact of migration on regional productivity is considerable and significant, regardless of the migration measure. Using the working-age population instead of the total population reduces the estimated effect slightly. While a 10% increase in the migrant share, following the baseline definition, increases productivity by 1.3% (Column 1; 0.14), a 10% increase in the migrant share across the working-age population boosts productivity by 1.2% (Column 2; 0.12). Moreover, the estimate remains significantly positive when using levels instead of the logarithm. In this specification, a one percentage point increase in the migrant share raises labour productivity by 0.55 (Column 3). An alternative measure proposed by Combes et al. (2019[44]) and used in Columns 4 and 5 also find a positive and significant effect on productivity. In this analysis, the estimated impact of migrations on productivity is positive for either specification, total (Column 4) or working age population (Column 5).[21]

It is also possible to measure the presence of migrants relative to the subset of the native population that migrants are expected to affect in the labour markets. This approach proposed by Combes et al. (2019[44]) measures the presence of migrants as the number of migrants relative to the lower-skilled native population (Column 6). In addition, this study also considers the ratio of migrants to higher-skilled natives (Column 7), given the high educational attainment of migrants in Australia. Both measures confirm the positive effects of migrants on labour productivity in Australia.

Overall, regardless of the definition or sample used in measuring the regional migrant presence, the positive effect of migration on labour productivity remains statistically significant and robust. These results confirm that the findings in this analysis do not depend on the specific definitions used in the analysis.

[21] This analysis uses levels of density, instead the natural logarithm of density as in previous research (Özgüzel, 2020[49]). Nonetheless, the impact of migration remains positive if the log population density is used instead of level.

Table A. 7 Alternative measures of migration

Second-step regression for regional productivity differences using different definitions of the presence of migrants, 2011-2018

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Migrant Share	0.13545***						
	(0.017)						
Log Migrant Share (WAP)		0.12292***					
		(0.015)					
Migrant Share (levels)			0.54883***				
			(0.075)				
Migrants				0.36264***			
				(0.051)			
Migrants WAP					0.29596***		
					(0.041)		
Migrant/Low						0.08793***	
						(0.010)	
Migrant/High							0.08288***
							(0.009)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop. weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional controls	All	All	All	All	All	All	All
N	704	704	704	704	704	704	704
R^2	0.901	0.909	0.902	0.901	0.900	0.913	0.915
F-stat	254.251	212.010	173.281	152.935	129.431	176.748	227.865

Note: The table presents IV estimates of the second-step regression for regional productivity differences, with different measures for the migrant share. Region-time fixed-effects estimated in Equation 1 are used as the dependent variable. The migrant share is the independent variable with different definitions in each Column. Additional and undisplayed controls are the share of higher-skilled migrants, the share of higher-skilled natives, Population Density, Domestic Market Potential, and Foreign Market Potential. The endogenous variable “Migrant Share” is instrumented according to its definition (different measures have different instruments). All the specifications are weighted by the number of employed natives in the region. Time fixed-effects are present to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications.
Source: OECD calculations based on MADIP (accessed June 2023)

The estimated impact of migrants holds when instrumenting other regional variables

Other regional characteristics can also suffer from endogeneity. As elaborated in the methodology section, regional variables such as the skill level of the workforce or population density can be driven or impacted by regional native productivity. For instance, a higher share of higher-skilled workers might partially result from economic conditions, i.e., wages and productivity, in the region. Such reverse causality would result in a biased estimate of the impact of human capital on productivity. Similarly, a higher population density might result from the high economic attractiveness of vibrant cities.

Similar to the identification strategy for the migrant share, historical Census data from 1981 are used to instrument the potentially endogenous variables. The approach builds on the assumption that the historical values and settlement patterns determine the current patterns and values yet are unrelated to current economic outcomes if the time lag between the years of analysis and the historical base year is sufficiently large (see Section 4). Building on this argument, the analysis uses instruments based on historical values.

Concretely, the skill levels of natives and migrants are instrumented using their predicted values based on the education decomposition by country of birth in 1981. Similarly, employed density is instrumented using the lagged value of 1981. Annex C provides more details on the construction of these instruments.

Independent of the identification strategy, the estimated impact of the migrant share remains highly significant and robust (Table A. 8). Regardless of whether only the migrant share is instrumented (Column 1) or all variables are considered endogenous and instrumented (Column 4), the positive effect of migration remains unchanged. Moreover, the effect of the skill share of migrants stays positive, although in many cases, it is statistically insignificant. Lastly, population density remains positive, yet the estimates are not statistically significant when addressing endogeneity in the human capital of natives (Columns 4 and 5).

Applying different instrumentation strategies shows that the positive effect of migrants is highly stable and robust. Instrumenting different combinations of regional characteristics does not affect its significance. On the contrary, the positive effect of the skill composition of migrants is less robust. While the direction of the effect does not alternate for different IV strategies, the magnitude and significance vary substantially.

Table A. 8 Alternative identification strategies

Second-step regression for regional productivity differences instrumenting migration and other regional variables, 2011-2018

	(1)	(2)	(3)	(4)	(5)	(6)
Log Migrant Share	0.13545***	0.13533***	0.14978***	0.14887***	0.14869***	0.15012***
	(0.017)	(0.017)	(0.024)	(0.043)	(0.043)	(0.024)
Log Share Higher-Skilled Migrants	0.10413**	0.10276*	0.25047	0.24365	0.24339	0.25404
	(0.052)	(0.053)	(0.159)	(0.327)	(0.328)	(0.156)
Log Share Higher-Skilled Natives	-0.05691	-0.05442	-0.19979	-0.19121	-0.19093	-0.20433
	(0.060)	(0.062)	(0.162)	(0.379)	(0.380)	(0.158)
Total Emp Density	0.00007***	0.00007***	0.00009**	0.00009	0.00009	0.00009***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
IV						
Migrant Share	Yes	Yes	Yes	Yes	Yes	Yes
Higher-Skilled Mig			Yes	Yes	Yes	Yes
Higher-Skilled Natives				Yes	Yes	
Total Emp. Density		Yes	Yes	Yes		
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Population weights	Yes	Yes	Yes	Yes	Yes	Yes
Control	All	All	All	All	All	All
N	704	704	704	704	704	704
R2	0.901	0.901	0.892	0.893	0.893	0.892
Cragg Donald Stat	2241.100	1086.180	69.105	11.958	15.908	116.431

Note: The table presents IV estimates of the second-step regression for regional productivity differences with different instrumented independent variables. Region-time fixed-effects estimated in Equation 1 are used as the dependent variable. The logarithm of the migrant share, the share of higher-skilled migrants, the share of higher-skilled natives and population density are the independent variables. The independent variables are gradually instrumented as indicated in the table. Additional and undisplayed controls are Domestic Market Potential and Foreign Market Potential. All the specifications are weighted by the number of employed natives in the region. Time fixed-effects are present to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications.
Source: OECD calculations based on MADIP (accessed June 2023)

Alternative density measures

The robustness of the results regarding the impact of migration on regional productivity is evident, even when employing different definitions of density. The magnitude of the effect remains considerable and statistically significant across various migration measures. While population density based on the surface area (Column 1) is a well-established measure in the literature, it might be misleading in the unique Australian context. The uneven size and concentrations of residence within regions might not mirror the actual population density in the region. To address potential inaccuracy, the subsection presents regression results using alternative definitions of population density, controlling for the impact of agglomeration economies. Unlike the baseline, alternative measures might use the built-up area (Columns 2 and 3) and the populated area (Column 4) in their denominator. Regardless of the definition, the results do not change.

Table A. 9 Second-step regression with different measures of population density

Second-step regression for regional productivity differences, 2011-2018

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Migrant Share	0.11045***	0.13545***	0.09853***	0.11602***	0.09670***	0.11622***	0.10601***	0.13435***
	(0.013)	(0.017)	(0.015)	(0.018)	(0.014)	(0.018)	(0.013)	(0.017)
Log Share Higher-Skilled Migrants		0.10413**		0.03372		0.03576		0.10779**
		(0.052)		(0.053)		(0.049)		(0.052)
Log Share Higher-Skilled Natives		-0.05691		0.00959		-0.00726		-0.07227
		(0.060)		(0.047)		(0.047)		(0.059)
$\left(\frac{\text{Emp.Pop}}{\text{Surface Area (km}^2\text{)}}\right)$	0.00006***	0.00007***						
	(0.000)	(0.000)						
$\left(\frac{\text{Emp.Pop}}{\text{Built-up Area 2010 (km}^2\text{)}}\right)$			0.00003***	0.00003**				
			(0.000)	(0.000)				
$\left(\frac{\text{Emp.Pop}}{\text{Built-up Area 2020 (km}^2\text{)}}\right)$					0.00003***	0.00004***		
					(0.000)	(0.000)		
$\left(\frac{\text{Emp.Pop}}{\text{Populated Area (km}^2\text{)}}\right)$							0.00007***	0.00009***
							(0.000)	(0.000)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control	All	All	All	All	All	All	All	All
N	704	704	704	704	704	704	704	704

R2	0.899	0.901	0.902	0.905	0.904	0.907	0.903	0.904

Note: The table presents IV estimates of the second-step regression for regional productivity differences. Region-time fixed-effects estimated in Equation 1 are used as the dependent variable. The logarithm of the migrant share, the share of higher-skilled migrants and the share of higher-skilled natives are the independent variables. The endogenous variable "Log Migrant Share" is instrumented as described in the empirical strategy. Each Column controls for a different combination of independent and control variables, including different definitions for the population density variable. Columns (1) and (2) use employed population over surface area. Columns (3) and (4) use employed population over built-up area in 2010, while Columns (5) and (6) use employed population over built-up area in 2020. Finally, Columns (7) and (8) use employed population over populated area. Additionally, all even-numbered Columns add the log share of higher-skilled migrants and the log share of higher-skilled natives as additional explanatory variables. All the specifications are weighted by the number of employed natives in the region. Time fixed-effects are applied to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications.
Source: OECD calculations based on MADIP (accessed June 2023).

Annex H. Excluding young and old workers

Table A. 10 Second step regression excluding younger and older workers

Second-step regression for regional productivity differences, 2011-2018

	Baseline	25-64	25-54
	(1)	(2)	(3)
Log Migrant Share	0.13545***	0.12740***	0.11082***
	(0.017)	(0.017)	(0.015)
Log Share Higher-Skilled Migrants	0.10413**	0.10598**	0.09102*
	(0.052)	(0.053)	(0.047)
Log Share Higher-Skilled Natives	-0.05691	-0.06793	-0.06054
	(0.060)	(0.063)	(0.056)
Time FE	Yes	Yes	Yes
Pop. weights	Yes	Yes	Yes
Regional controls	All	All	All
N	704	704	704
R^2	0.901	0.758	0.839
F-stat	254.251	254.251	254.251

Note: The table presents IV estimates of the second-step regression for regional productivity differences. Region-time fixed-effects estimated in Equation 1 are used as the dependent variable. The logarithm of the migrant share, the share of higher-skilled migrants and the share of higher-skilled natives are the independent variables. The endogenous variable "Log Migrant Share" is instrumented as described in the empirical strategy. Column (1) provides the baseline results for comparison, which use population aged 15 to 64. Column (2) excludes from the estimation individuals aged 15 to 24, and Column (3) further excludes individuals aged 55 to 64. All the specifications are weighted by the number of employed natives in the region. Time fixed-effects are applied to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications.
Source: OECD calculations based on MADIP (accessed June 2023).

Annex I. Excluding occupations from the first stage

Table A. 11 Second step regression excluding occupations in the first stage

Second-step regression for regional productivity differences, 2011-2018

	Baseline	No Occ
	(1)	(2)
Log Migrant Share	0.13545***	0.14036***
	(0.017)	(0.018)
Log Share Higher-Skilled Migrants	0.10413**	0.10795**
	(0.052)	(0.054)
Log Share Higher-Skilled Natives	-0.05691	-0.05919
	(0.060)	(0.062)
Time FE	Yes	Yes
Pop. weights	Yes	Yes
Regional controls	All	All
N	704	704
R^2	0.901	0.904
F-stat	254.251	254.251

Note: The table presents IV estimates of the second-step regression for regional productivity differences. Region-time fixed-effects estimated in Equation 1 are used as the dependent variable. The logarithm of the migrant share, the share of higher-skilled migrants and the share of higher-skilled natives are the independent variables. The endogenous variable "Log Migrant Share" is instrumented as described in the empirical strategy. Column (1) provides the baseline results for comparison. Column (2) provides the results using Region-time fixed effects estimated from the first stage but omitting occupations in the estimation. All the specifications are weighted by the number of employed natives in the region. Time fixed-effects are applied to account for time-varying shocks affecting the whole country. Standard errors are clustered at the regional level in all the specifications.
Source: OECD calculations based on MADIP (accessed June 2023).