

ORIGINAL ARTICLE

Open Access



# The regional impact of cultural diversity on wages: evidence from Australia

Amanuel Elias\* and Yin Paradies

\* Correspondence:  
amanuel.h@deakin.edu.au  
Alfred Deakin Institute for  
Citizenship & Globalisation, Faculty  
of Arts and Education, Deakin  
University, Burwood, Australia

## Abstract

This paper investigates the impact of cultural diversity on labour market outcomes, particularly on wages across regions using a large longitudinal data. We apply an instrumental variable approach and account for individual and time fixed effects. Our findings indicate that the current level of cultural diversity positively affected current regional weekly wages; however, the positive effect holds only partially when the diversity is lagged. The results appear to be robust in all estimations controlling for heterogeneity factors and accounting for the self-selection of individuals into places with better economic opportunities. Our findings concerning the effect of lagging on the effect of diversity may explain the variation in the literature where some studies report that cultural diversity increases wages across time while others do not.

**JEL Classification:** J610, R23, Z190

**Keywords:** Cultural diversity, Wages, Instrumental variable, Shift-share, Australia

## 1 Introduction

Economic theory indicates that cultural diversity is related to economic performance.<sup>1</sup> Large cities with culturally diverse population are usually likely to be the centres of rapid economic growth and employment. But they can also be the centres of attraction for more labour and diversity. Therefore, endogeneity and reverse causality have been the focus of substantial research in the economics of diversity (Longhi 2013; Ottaviano and Peri 2006). In addition, whether the net effect of diversity is good or bad for the economy in general, and the labour market in particular, continues to stir debate among researchers (Alesina and La Ferrara 2004; Herring 2009; Longhi 2013; Ottaviano and Peri 2006). Generally, the literature in this area has focused on four outcomes of interest: labour market, innovation, social capital/tolerance, and economic growth. In this paper, we focus on the labour market, examining the effects of cultural diversity on regional wages. The pathway through which this relationship plays out depends on both the demographic composition and the cultural distance that underlie the diversity. Competing theories have suggested that cultural diversity is beneficial for long-term economic growth but can reduce trust in the short-term (Putnam 2007). In a situation where a culturally diverse climate contributes to a variety of skills in the workforce, diversity has a positive impact on economic growth. However, the impact becomes negative if it leads to conflict and polarisation (Alesina and La Ferrara 2005).

Previous cross-sectional studies suggest that cultural diversity increases employment and wages at the regional level, thereby leading to economic growth (Bellini et al. 2013; Kohler 2012; Ottaviano and Peri 2006). A range of reasons have been proposed for this, including factors that affect the labour market, businesses, and industry. For example, cultural diversity has been linked with greater employee commitment and improved productivity as well as greater creativity, innovation, and problem solving arising from a wider pool of skills and with the diffusion of these capabilities (Damelang and Haas 2012; Herring 2009; Perotin et al. 2003; Putnam 2007; Richard 2000; Suedekum et al. 2014). Diversity has also been associated with an increased variety of preferences, better customer satisfaction, larger market share, increased sales revenue, and greater relative profits (Bertone and Leahy 2001; Herring 2009; Page 2008). In relation to human and social capital, it has been linked with improved student wellbeing in schools (Juvonen et al. 2006) as well as augmented social capital (Putnam 2007).

In contrast, the ethnic fractionalisation literature indicates that cultural diversity can lead to interracial conflict or racism, at least in the short-term, followed by a decline in economic performance due to reduced investment and public spending (Alesina and La Ferrara 2005; Fearon 2003; Montalvo and Reynal-Querol 2005; Stahl et al. 2010). In a dynamic model with a lagged measure of diversity, Campos et al. (2011) found that diversity has a significant negative impact on economic growth. A range of reasons can be posited for this negative economic impact. Some studies, for example, have linked ethnic diversity with reduced social cohesion leading to conflicts (Kochan et al. 2003; Lieber 2009; Roberson and Kulik 2007). At the organisational level, diversity that is poorly managed can reduce staff morale and productivity, provoke conflict between employees and managers, and harm social cohesion (Kochan et al. 2003; Roberson and Kulik 2007; Wrench 2005). Diversity may also result in the perpetration of, and exposure to, prejudice and racism, marginalisation of minorities, deterioration of social capital, and political conflict (Stahl et al. 2010), with any potential benefits offset by the costs of such phenomena (Campos et al. 2011; Montalvo and Reynal-Querol 2005; Triana et al. 2015).

This study extends the empirical evidence in this literature in two ways. First, most studies have used cross-sectional data in analysing the economic impact of diversity. We use a large longitudinal data that allows us to investigate this relationship accounting for variation over time. Second, Australia has unique migrant characteristics due to its geographic isolation from source countries which allows it to control the flow of migrants through specific gateways. In addition, per capita, it is one of the largest migrant-receiving countries in the world with 26 % of the population born overseas and an additional 20 % having at least one parent born overseas (Australian Bureau of Statistics, ABS 2014). In the capital cities, the average is even higher with the migrant population accounting for 28.9 % of the urban residents. More than 19 % of the overseas-born population, aged 5 years and over, speak a language other than English at home (ABS 2012b). Yet, the economic impact of this diversity has yet to be investigated. Therefore, in this study, we analyse the causal impact of diversity on wages by examining the regional variation in the cultural composition of the labour force.

We use the Household, Income and Labour Dynamics in Australia (HILDA) data which clusters individuals based on their postcodes. Using the country of birth variable from census data, we create a local government area (LGA)-level measure of diversity (fractionalisation index). Following Pischke and Velling (1997), Card (2001), Ottaviano

and Peri (2006), and Bellini et al. (2013), we specify an instrumental variable estimation to overcome the endogeneity of diversity in our model. Following Longhi (2013), we address time and individual effects by specifying fixed effects in addition to an ordinary least squares (OLS) model. We also account for heterogeneity in the effects of diversity using sub-sample analysis based on the ancestry, mobility, skill, and residency of respondents in our data.

The current evidence in relation to diversity varies depending on the type and context under which diversity is studied. Studies such as Alesina and La Ferrara (2004), Fearon (2003), and Montalvo and Reynal-Querol (2005) conceptualise diversity in terms of ethnic composition, which they link with interethnic conflicts. Earlier studies focused mainly on the diversity of immigrants in a country and indicate a zero or negative (but small) correlation between the influx of immigrants and native wages and no association between the proportion of immigrants and native rates of employment (Altonji and Blank 1999; Borjas et al. 1997; Friedberg and Hunt 1995). However, assuming a perfect substitutability between migrants and native workers, Borjas et al. (2008) found a negative effect of immigrant share on men and women's wages. Pischke and Velling (1997), on the other hand, found the national impact of immigration to be minimal in a German regional study which accounted for the self-selection of migrants into local labour markets. A more recent cross-national study by Sanderson (2013) showed that immigration raises the overall living standards in host countries in the long-term (although this is attenuated in high fertility contexts).

Other studies have focused on diversity in terms of country of birth. Ottaviano and Peri (2006) who investigated the impact of immigrants on 160 US cities found that "on average, cultural diversity has a net positive effect on the productivity of U.S.-born citizens because it is positively correlated with both the average wage received and the average rent paid by U.S.-born individuals" (p. 11). They concluded that US-born urban residents living in areas where the share of foreign-born residents increased (in 1970–1990) had a substantial rise in their wages and the rental prices they pay. Using similar approaches, Bellini et al. (2013) found a positive wage effect of diversity in 12 European regions and Nathan (2011) found a positive impact of diversity on wages for a range of British studies. Another study by D'Amuri et al. (2010) found that the flow of new immigrants depressed the employment levels and wages of old immigrants while having no meaningful effect on the employment and wages of natives. This is contradicted by Longhi (2013) who analysed the labour market effects from seven international studies. After accounting for individual and time heterogeneity, she found that the lagged measure of diversity was negatively associated with wages and employment. Similarly, Angrist and Kugler (2003) found diversity to be weakly but negatively associated with the level of employment in European data. Borjas et al. (2008) further report that even accounting for long-run adjustments, an increased supply of immigrants lowers native wages. In this study, we test some of the previous findings using longitudinal data, over a 10 year period, to account for individual and time effects and the time lag in measuring the effect of diversity.

The rest of this paper is organised as follows: Sections 2 and 3 describe the data and analytical methods, respectively, detailing the measurement issues related to diversity. Section 4 presents the results while Section 5 concludes and discusses the implications of this research.

## 2 Description of data

Three main datasets are used in this paper. The primary data source is the 2001–2011 HILDA Survey. The second is demographic data from the 2001 and 2011 censuses while the third data includes a set of annual population estimates compiled from the ABS creative commons release for the period 2001–2011.

### 2.1 The HILDA Survey

This paper uses the un-confidentialised version of the HILDA Survey (Release 11) which has postcode data that allows for area-level analysis (see Watson 2012, for detailed description of the HILDA Survey). This postal area data is exploited to measure an index of diversity as the key variable in measuring the distribution of foreign nationals and to merge with an instrument and an alternative index of diversity computed from census data. The initial sample interviewed in the first wave (wave 1) consisted of 7682 households while the corresponding sample of enumerated persons was 19,914 people. Out of these, 24 % are children aged below 15 years. Of those who are eligible, the response rate was 92.3 % ( $n = 13,969$ ). In subsequent waves, new household members and children increased while overseas emigration and deaths decreased the sample size. In addition, attrition rates of 3.7–13.2 % across waves contributed to the decline of the sample size in subsequent waves.<sup>2</sup> On average, 13,438 respondents were interviewed every year of which 7228 individuals continued to participate in the survey each year without missing any wave. However, the combined sample size for the 11 waves (waves 1–11) including those who were added in subsequent waves is 26,028. Out of this, a long panel was constructed for the 11 waves amassing an overall sample of 286,308 observations.

Since the sample of interest in this paper is the labour market performance of those who are potentially active in the labour force, the sample was restricted to the working age group. The total sample size for waves 1–11 in the long panel of those aged 16–45 years is 79,636 (27.8 %). Of these, those who are employed and earning wages account for 64.7 % ( $n = 51,538$ ). Finally, in the multivariate analyses, the sample was further restricted to allow for a longer time period (3-year lag), with the final sample of 44,634 (56 %).

### 2.2 Country of birth data in ABS census 2001 and 2011

Census data was used to generate an index of diversity at the local government area (LGA) level as well as an instrument based on the projected population. This data, which includes the regional distribution of Australians by country of birth, was obtained from the ABS via TableBuilder using the 2001 and 2011 census data (ABS 2012a). A total of 293 countries classified under the four-digit Standard Australian Classification of Countries (SACC) are available in the census (ABS 2011a), along with a total of 2516 postcodes.<sup>3</sup> These data are also available at the LGA level.

The LGA is a geographically contiguous classification which divides Australia into 676 regional categories. Each LGA, also known as a local council, handles community-related tasks and town planning within its jurisdiction. Diverse local entities including cities, towns, suburbs, shires, and villages make up these local councils. We originally obtained the census data at the postcode level. To make the analysis relevant to regional labour market and community characteristics, we decided to broaden the classification to the LGA level. The LGA-level data also accounts for socio-demographic and regional policy differences across Australia. Using the 2006 ABS Postal Area Concordances that map LGAs and postcodes,

the postcode data was merged and then collapsed into 676 LGAs across Australia. The LGA-level data was used to generate the regional distribution of diversity and the main instrument. Then the LGA and postal area data were used to merge these census-based data with the HILDA data. Although the merging variable is the postcode, the diversity index used for analysis is at the LGA level. The diversity measure from the census is used for spatial analysis, visually portraying changes in diversity over the 2001–2011 period.

### 2.3 Annual population estimates 2001–2011

Australia's annual population estimates for the period 2001 to 2011 were obtained from the ABS website. These datasets include an aggregate distribution of the population estimates by country of birth at the national level.<sup>4</sup> A total of 255 countries of origin were represented in these datasets. From this distribution, the annual growth rate of the population was estimated and was utilised along with the 2001 census data in the calculation of time-variant shift-share instruments for each wave (see Section 3.7 for a detailed discussion).

## 3 Measures

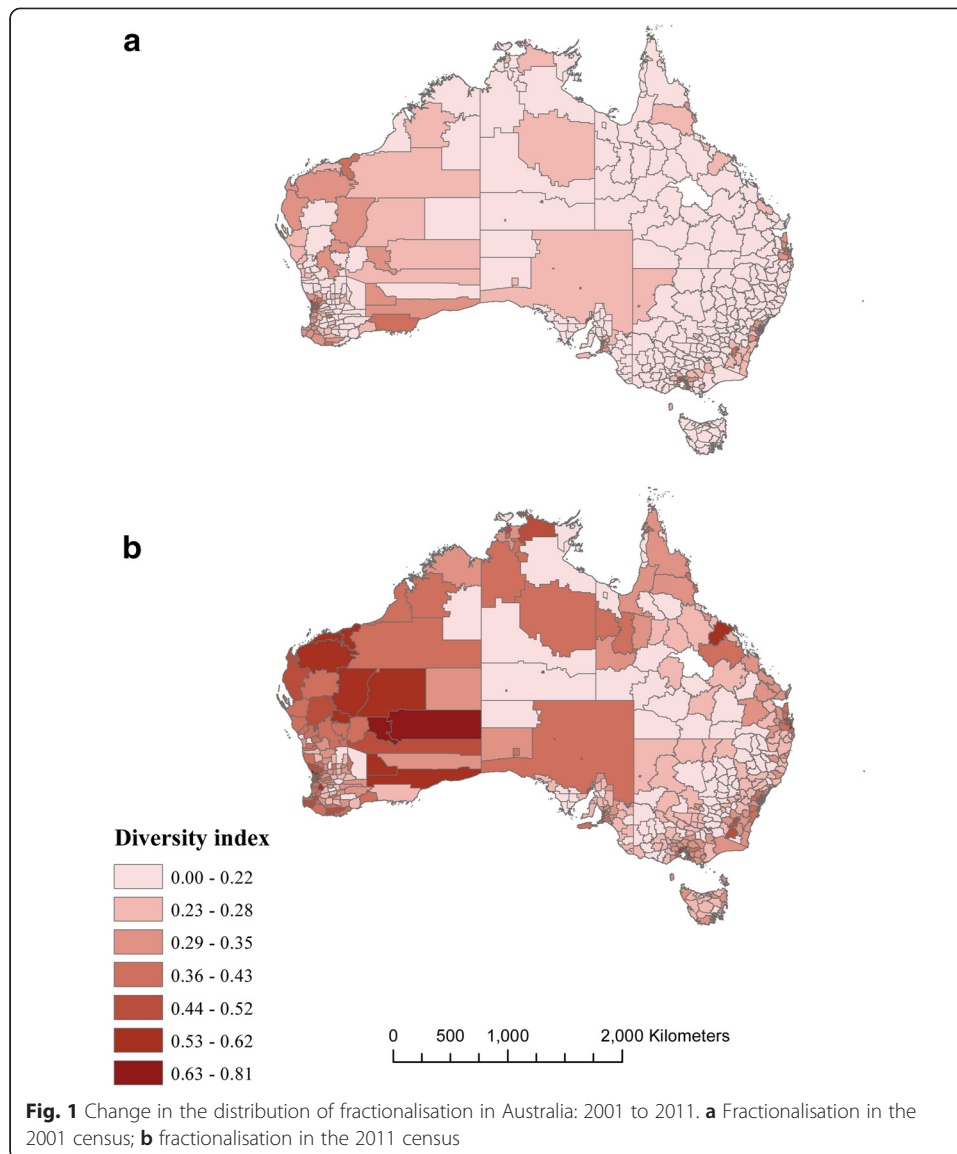
### 3.1 Diversity index

The cultural diversity literature uses the fractionalisation index in measuring the impact of diversity on economic outcomes. Ethno-linguistic fractionalisation (ELF) is defined as the “probability that two randomly chosen citizens in a country belong to a different ethnic group where(in) group belonging is attributed by language” (Neumann and Graeff 2013). Vigdor (2008) uses a slightly different approach, the probability that a randomly selected individual is an immigrant, to estimate an assimilation index in the USA. Others have used the country of birth data to measure cultural diversity (Alesina et al. 2013; Bellini et al. 2013; Damelang and Haas 2012; Longhi 2013; Ottaviano and Peri 2006). Given the availability of data in the Australian context, this study uses country of birth/nationality instead of ethnicity/language.<sup>5</sup> Specifically, the proportion of the nationals of each country of origin (birth) in each LGA in Australia is used to compute a fractionalisation index. This index (hereafter diversity index) has a similar theoretical interpretation to the *Herfindahl Index* which is widely used in marketing research to measure the market/monopoly power of firms located in specific areas (Gomez-Mejia and Palich 1997) and is given by

$$DI_{rt} = 1 - \sum_{i=1}^I C_{irt}^2 \quad \forall i = 1, 2, \dots, N, t = 1, 2, \dots, T \quad (1)$$

where  $C_{irt}$  represents the proportion of the nationals of country  $i$  in region (LGA)  $r$  in a given year  $t$ . The values fall in the range  $[0, 1]$  with “zero” indicating perfect homogeneity and “1” indicating perfect heterogeneity. We use the HILDA panel to construct the index of diversity. For the sake of visual comparison, we also estimated the indices for the 2001 and 2011 censuses (see Fig. 1). However, given the annual time series nature of our data, our main analyses is based on the indices constructed from HILDA.

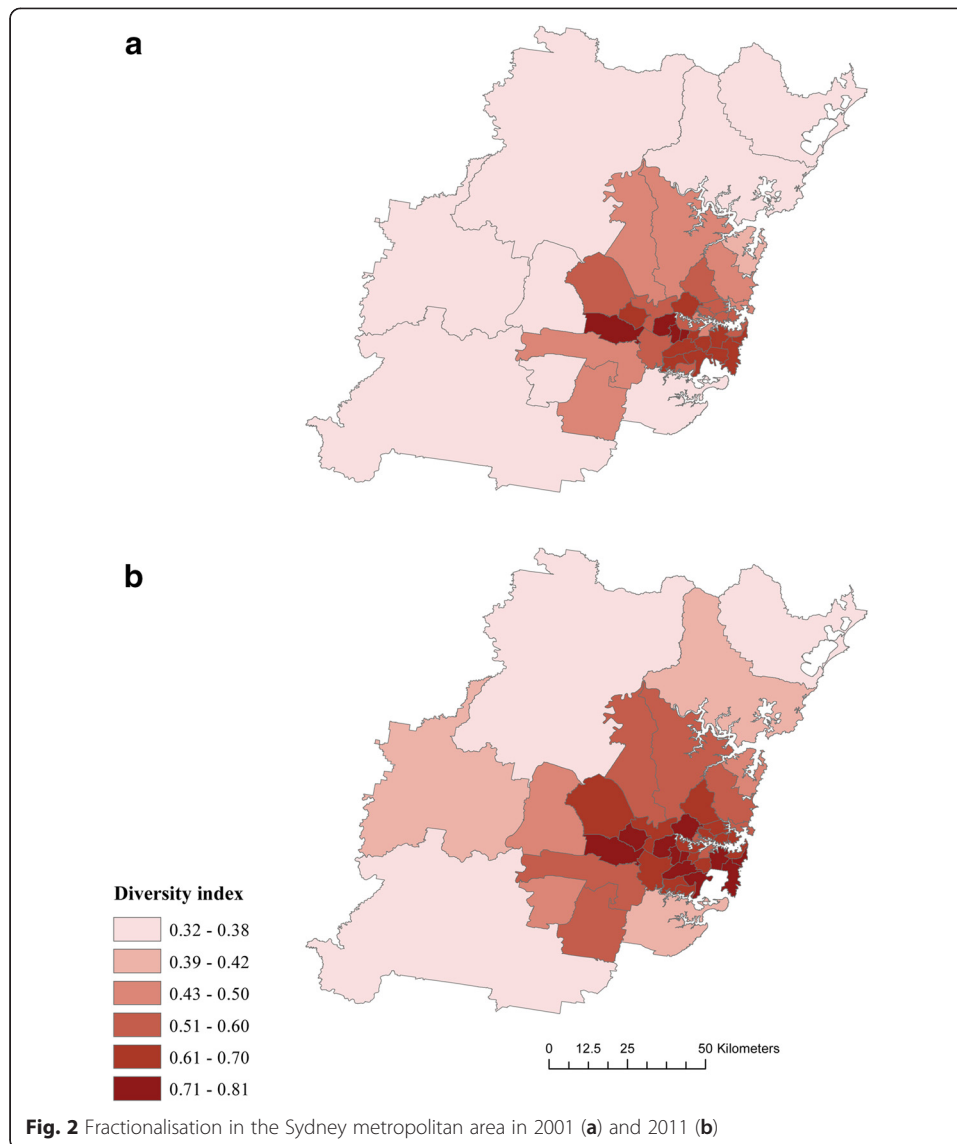
These maps are constructed based on ABS census data. Although we had diversity data available for 676 LGAs, the 2011 Australian Standard Geographical Classification (ASGC, ABS 2011b) digital boundaries allow for only 560 LGAs upon which the maps reported in Fig. 1 are based. The first figure, Fig. 1a, is the distribution of the diversity



index based on the 2001 census while Fig. 1b is based on the 2011 census. A comparison of the two figures indicates that cultural diversity increased in several regions over the decade. This is particularly visible in the metropolitan areas including Sydney and Melbourne (see Figs. 2 and 3, respectively). Both interregional mobility and international migration have contributed to this demographic change (Hugo and Harris 2011). Therefore, the analysis of diversity in this study accounts for both factors, by using a predicted instead of actual diversity index.

### 3.2 Share of migrants

The “share of migrants” is an alternative measure to assess whether the proportion of immigrants in a region per se has any effect on weekly wages. In addition, a diversity index is estimated for migrants excluding the Australian-born population. This is then included to see whether diversity among migrants (as opposed to diversity in general) contributes to labour market outcomes.

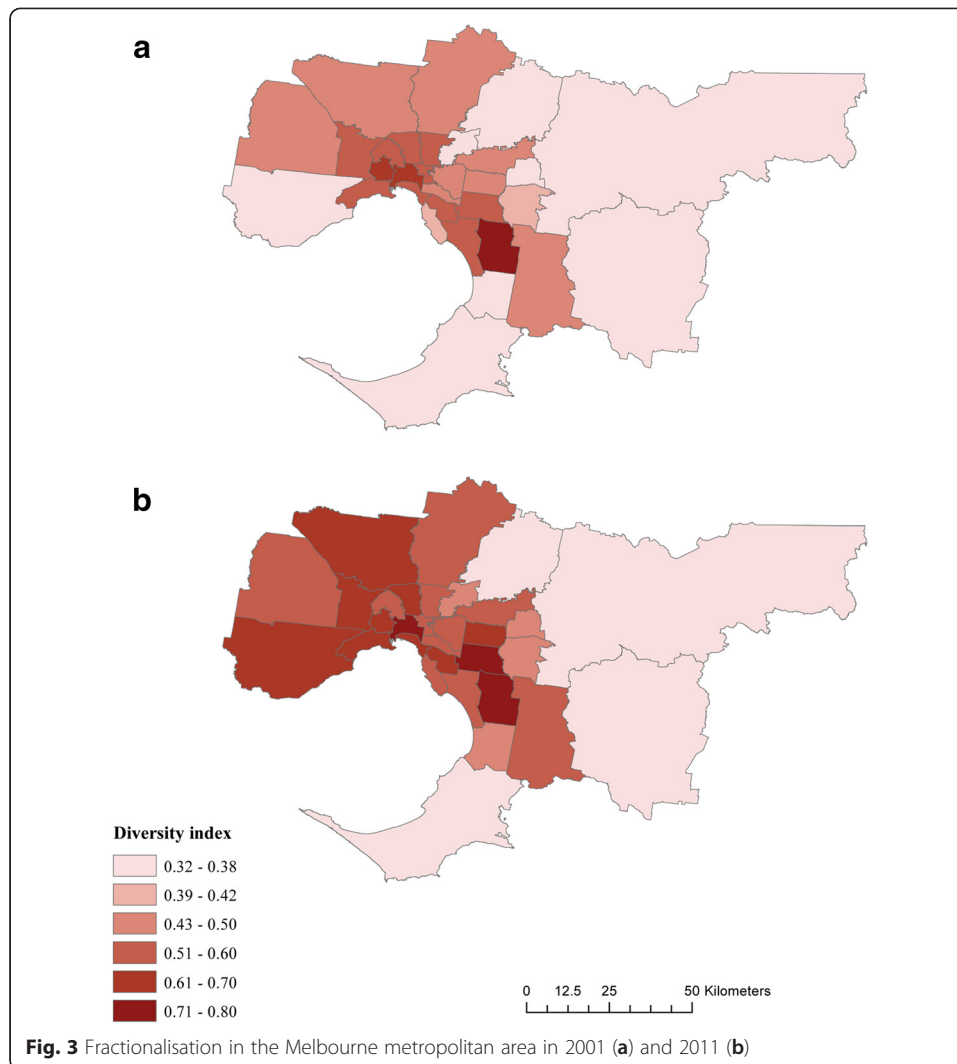


### 3.3 Weekly wages

The main dependent variable in this analysis is the log of weekly wages. Originally, HILDA respondents were asked a series of questions such as “For your [job/main job] what was the total gross amount of your most recent gross pay before tax or anything else was taken out?” Responses were recorded as “gross weekly wages and salaries” for the responding persons. For the complete panel, the mean weekly wage was \$651.6 (SE = \$254.1).

### 3.4 Other control variables

In addition to diversity (fractionalisation) and the share of migrants, standard demographic variables (age, age squared, gender, marital status) are included as control variables. English language fluency is also included, as the ability to speak English well is usually associated with labour market outcomes for migrants (Dustmann and Fabbri 2003). Foreign-born HILDA respondents were asked how well they spoke English with four response options ranging from “very well” to “not at all”. The third and fourth



options (“not well” and “not at all”) are collapsed because those who responded with the fourth option were negligible (0.02 %).

### 3.5 Analytic framework

In analysing the HILDA data, this study aims to test the hypothesis that cultural diversity can have a positive impact on labour market outcomes by boosting regional economic growth. The labour market channel involves a dynamic interaction between employment and wages. However, in this study, the main focus is the impact of diversity on wages, taking into consideration regional variations. The effect of diversity on wages can be estimated via panel data analysis that accounts for individual and regional effects over different time periods. The first model estimated is a simple OLS model of the log of weekly wages ( $\ln(w_{irt})$ ) for each employed respondent aged 16–45 years.

$$\ln(w_{irt}) = \alpha_{1i} + \beta_1 \text{div}_r + \delta_1 X_{irt} + \varepsilon_{1irt} \quad (2)$$

where the main variable of interest is the diversity index  $\text{div}_r$ . As suggested in the literature, further explanatory variables ( $X_{irt}$ ) are included such as weekly number of hours

worked and job tenure as well as time indicator variables. In addition, age and its square, dummies for female, marital status, and region as well as English language skill and education indicators are included where appropriate.

Apart from the observable characteristics, there can be individual heterogeneity that can affect the relationship between diversity and labour market outcomes. Longhi (2013) shows that the positive wage effects of diversity reported in cross-sectional studies (Nathan 2011) can be explained by individual differences. A fixed effect (FE) model is therefore estimated in this study capturing the unobserved individual characteristics among HILDA respondents. All the explanatory time-variant variables included in OLS are also included in the FE models.

### 3.6 Endogeneity of cultural diversity

The impact of cultural diversity on an economy is confounded due to the possibility of reverse causality, whereby Eq. (2) results in a spurious correlation (Friedberg and Hunt 1995). Our purpose is to determine the effect of diversity on wages, but a two-way causality between diversity and wages is possible. While diversity can directly affect economic performance, it is also possible that people from diverse backgrounds can self-select to live in places with economic opportunities.

The impact of diversity on economic outcome can be positive or negative. On the positive side, diversity can augment economic performance as it can stimulate creativity and problem solving. Diversity can also boost economic growth by drawing labour from a pool of immigrants. On the negative side, it can deplete trust and social capital due to ethnic/racial fragmentation. This can in turn weaken economic performance. Whether the positive effects of diversity on economic performance outweigh the negative ones, at one level, is a simple empirical question. However, when economic outcomes directly or indirectly affect diversity rather than the reverse, there arises an econometric issue.

In this study, the issue of reverse causality arises when variations in regional weekly wages resulted in the concentration of people from diverse cultural backgrounds in specific regions. For example, in Australia, there is no restriction in the mobility of immigrants within the country, and potentially, immigrants can move to places with more perceived economic opportunities (see Hugo and Harris 2011). The HILDA data, for example, shows substantial internal migration across waves among the respondents. Therefore, reverse causality cannot be ruled out from a regression of economic performance on cultural diversity. Instead of diversity causing variation in regional labour market outcomes, the economic conditions such as prospects of employment may be driving the regional distribution of diversity. This poses an econometric issue, endogeneity, in estimating the causal effect of cultural diversity on employment outcomes. The effect of the explanatory variable, diversity, as measured by the share of foreign country citizens in a region (LGA) is confounded by the possibility of migrants' concentration in response to economic incentives. Therefore, the coefficient of diversity cannot be consistently estimated due to correlation with the error term in the wage regression where the share of migrants is endogenous. This entails the violation of the Gauss-Markov (zero conditional mean) assumption in OLS (Wooldridge 2010).

A suitable procedure to correct the endogeneity problem is to apply instrumental variable (IV) estimation (see Baltagi 2008; Wooldridge 2010). The main challenge in

applying this procedure is the identification of a valid instrument. If such an instrument can be found, the confounding, for example, between diversity and economic performance can be disentangled and causal relationship between these two variables established. In this study, the shift-share method is used to instrument for the index of cultural diversity and the share of migrants. Following Bellini et al. (2013), Ottaviano and Peri (2006), Longhi (2013), and, recently, Alesina et al. (2013), two-stage least squares (2SLS) estimation is applied to OLS and FE models.

### 3.7 Identification strategy

For an instrumental variable estimation to be specified for Eq. (2), two assumptions should be satisfied. First, the instrument chosen should be correlated with cultural diversity, the key explanatory variable, and second, it should not be correlated with economic performance. In addition, a correctly specified model should not omit relevant variables. Several instruments have been developed in the literature to solve the endogeneity issue in relation to cultural diversity. Altonji and Card (1991) use the 1970 immigrant stock in the USA while Hunt (1992) uses regional temperature and French repatriates of 1962 in a French-Algerian migration study. Ottaviano and Peri (2006) use distance from gateway cities in the USA while Longhi (2013) uses “the proportion of minorities joining the ‘New Deal Program’” in the UK. As detailed in the introductory section, mixed (positive and negative) results were obtained by these studies regarding the impact of diversity on economic outcomes.

An instrument suitable for the data used in this paper is the shift-share variable first utilised by Card (2001) in assessing the local labour market impact of immigrant flow in the USA. This instrument was later used in modelling the causal effect of cultural diversity on wages and rental prices for US cities (Ottaviano and Peri 2006) and European regions (Bellini et al. 2013). The shift-share analysis assumes that the regional migrant distribution can be used to generate an exogenous variable using two-time-period data. For example, the 2001 Australian census datasets have regional distribution of Australians based on their country of birth which along with ABS annual population estimates can be used to construct a measure of diversity.<sup>6</sup> The latter is composed of nationally aggregated distribution and has annual estimates by country of birth for the period 1992–2014. In this study, we use the period 2001–2011. These datasets offer two variables that are relevant here. One is the total number of population in each region by country of birth, and the other is the total number of population in each region. From these variables, it is possible to calculate the annual population growth in Australia by country of origin. Then these annualised estimates can be used along with the baseline (2001) regional population data to estimate the predicted population for each year up to 2011. Since these predicted figures are based on historical (year 2001) regional distribution rather than actual regional distribution, they are not confounded by population growth that could have resulted from economically driven mobility. Therefore, they are assumed to be exogenous to regional economic shocks.

The primary purpose is to estimate the predicted version of the share of migrants in Australia. First, the overall growth rate in the Australian population between time  $t$  (which is 2001) and time  $t + 1$  is required. Formally, this rate  $g_i$  is given by

$$g_i = \frac{(p_i^{t+1} - p_i^t)}{p_i^t} \quad (3)$$

where  $p_i^{t+1}$  and  $p_i^t$  represent the total number of the Australian population born in country  $i$  in the years  $t$  and  $t + 1$ , respectively. The next step is to generate the predicted number of Australian residents born in country  $i$  and residing in region  $r$  based on Eq. (3). This is given by the formula

$$p_{ir}^{*t+1} = p_{ir}^t (1 + g_i) \quad (4)$$

where  $*$  indicates that the value is predicted for the year  $t$ . Summing this value ( $p_{ir}^{*t+1}$ ) across all countries of birth provides the predicted total population for each region (LGA) in the next year.

$$P_r^{*t+1} = \sum_i p_{ir}^{*t+1} \quad (5)$$

where  $P_r^{*t+1}$  indicates the predicted total number of all residents in each region in  $t + 1$ . This value which differs from the actual population in that year is used to calculate the predicted diversity index ( $DI_{st}$  as in Eq. (1)). Furthermore, this analysis is repeated to estimate the predicted migrant share in each region. The value is then used to calculate the predicted diversity index among migrants. Finally, the two instruments, namely the predicted diversity index and predicted share of migrants, are merged into the individual-level HILDA data based on the postal area variable.<sup>7</sup>

Generally, the indices of diversity and the instruments generated using the 2001 census data and population estimates are correlated, satisfying the relevance criteria. However, the correlation coefficients are larger for diversity measure based on annual population estimates with  $r = 0.40$  compared to diversity based on a 3-year lag where  $r = 0.30$ . On the other hand, the exogeneity criteria are also met, as the instruments are not correlated with the error term. The correlation coefficients between the residual and the two instruments (predicted diversity index and predicted share of migrants) are  $r = 0.06$  and  $r = 0.06$ , respectively.

## 4 Results

### 4.1 Descriptive statistics

As indicated in Section 2.1, this analysis is based on the age-restricted sample of HILDA respondents. The overall age of the whole sample (waves 1–11) ranges 15–99 years. Excluding those below 16 years as well as those aged 46 years and over (27.8 %), the final sample size is  $n = 79,636$ . However, in all estimations, the reported sample sizes differ from the original due to the application of sampling weights and lagging (dropping waves 1–3). Weighted descriptive statistics summarising the characteristics of this sample utilised in this study are presented in Tables 1 and 2.<sup>8</sup>

All the variables reported in Table 1 are strictly time-invariant except education. The table shows that HILDA participants on the average are roughly evenly distributed by gender, with females making 52 % of the sample. More than three quarters (78.9 %, and 78.2 % for all waves) are born in Australia, with 2.2 % of these identifying as Indigenous. The biggest source of the migrant sample is made up of those from Asia-Pacific countries (10.9 %). Married respondents are more than those who were never married in both the whole (52.8 to 35.2 %) and restricted samples (53 to 36.2 %). Overall, 14.5 % of the total sample and 14.2 % of the restricted sample speak English well or above. These indicate that 61 and

**Table 1** Weighted mean and standard errors of demographic and socio-economic variables

Variables	Waves 1–11		Wave 4–11	
	Mean	Jackknife std. err.	Mean	Jackknife std. err.
Gender (female = 1; male = 0)	0.523	0.043	0.523	0.053
Ancestry (Australian-born = 1; foreign-born = 0)	0.782	0.031	0.789	0.035
Geographic origin				
Australian	0.759	0.026	0.767	0.032
Indigenous	0.022	0.014	0.022	0.019
Asia-Pacific	0.111	0.044	0.109	0.046
Europeans and North Americans	0.068	0.029	0.063	0.030
East Europeans	0.005	0.009	0.005	0.010
Middle East, Africans, and Latin Americans	0.034	0.013	0.033	0.014
Father's ancestry (Australian-born = 1; foreign-born = 0)	0.614	0.044	0.618	0.055
Mother's ancestry (Australian-born = 1; foreign-born = 0)	0.635	0.052	0.638	0.071
Marital status				
Married	0.528	0.288	0.530	0.380
Separated/divorced/widowed	0.120	0.138	0.108	0.151
Never married	0.352	0.423	0.362	0.527
How well speaks English				
Not asked	0.874	0.102	0.877	0.122
Very well	0.085	0.071	0.087	0.088
Well	0.060	0.041	0.055	0.038
Not well	0.032	0.068	0.026	0.082
Highest year of school achieved				
Postgrad—masters or doctorate	0.040	0.035	0.043	0.043
Grad diploma, grad certificate	0.059	0.070	0.063	0.091
Bachelor or honours	0.165	0.066	0.170	0.085
Adv. diploma, diploma	0.096	0.052	0.095	0.060
Cert III or IV	0.210	0.080	0.219	0.113
Cert I or II	0.015	0.007	0.016	0.008
Cert not defined	0.004	0.008	0.004	0.009
Year 12	0.184	0.151	0.182	0.221
Year 11 and below	0.227	0.150	0.208	0.169
State				
New South Wales	0.304	0.086	0.300	0.094
Victoria	0.245	0.058	0.245	0.063
Queensland	0.217	0.030	0.220	0.032
South Australia	0.078	0.020	0.078	0.020
Western Australia	0.094	0.041	0.094	0.039
Tasmania	0.029	0.022	0.029	0.023
Northern Territory	0.011	0.013	0.011	0.015
Australia Capital Territory	0.022	0.020	0.023	0.025
Section of state				

**Table 1** Weighted mean and standard errors of demographic and socio-economic variables  
(Continued)

Major urban	0.661	0.032	0.655	0.050
Other urban	0.220	0.038	0.222	0.050
Bounded locality	0.022	0.013	0.022	0.011
Rural balance	0.097	0.011	0.102	0.011
Observations (weighted)	37,455		25,231	
df	54		54	

Means and standard errors of categorical variables are presented in this table for the combined panel of the HILDA Individual Person Respondent sample. The sample is restricted to those aged 16–45 years. Survey jackknife method was used to estimate standard errors for weighted mean values. Source: Author's calculation based on HILDA Release 11

69.5 % of those who are foreign-born speak English well or above, respectively. In addition, 22.7 and 20.8 % of each sample did not complete high school, while 26.4 and 27.6 % have a college degree or above. Finally, 66.1 % of the whole and 65.5 % of the restricted sample reside in major urban areas.

Table 2 reports the weighted descriptive statistics of the metric variables which are utilised in further analysis. The standard errors are estimated using jackknife replication. The mean age of the sample in the panel is roughly 34 years. Of those who are in the labour force, about 4 % were unemployed at the time they were interviewed. For those who are employed (employment status = 1), the average number of hours worked is 38.5, with a weekly average wage (in logs) of 6.67. In monetary terms, weekly wages and salaries ranged between \$2 and \$10,070, with more than 32.1 % of the sample earning below the 2001–2011 average annual minimum wage of \$502.3 a week. Respondents who were unemployed at the time of interview indicated an average reservation wage (the lowest wage per hour they considered acceptable) of 2.82 (in log, or \$16.8), above the average annual (2001–2011) minimum wage of \$13.22 per hour.

**Table 2** Weighted descriptive statistics of metric variables in the HILDA individual person respondent sample aged 15–45 years (waves 4–11)

Variable	Mean	Jackknife std. err.	Min	Max	Weighted observations	Sample restriction
Age (years)	34.32	10.247	16	45	25,345	Whole sample
Number of living children	1.35	0.552	0	12	25,345	Whole sample
Diversity index	0.36	0.032	0	1	25,345	Whole sample
Diversity index (among migrants)	0.72	0.009	0	1	25,345	Whole sample
Share of foreigners	0.23	0.035	0	1	25,345	Whole sample
Employment status (1 = employed; 0 = otherwise)	0.96	0.056	0	1	25,983	Persons in labour force
Years of experience (in years)	6.55	4.302	0	33	25,983	Persons in labour force
Hours per week worked (all jobs)	38.51	7.037	0	120	20,538	Employed persons
Log of wages and salaries (weekly)	6.67	0.634	0.69	9.22	20,538	Employed persons
Job tenure (in years)	4.93	2.751	0	33	20,538	Employed persons
Log reservation wage	2.82	0.451	0	7.31	950	Unemployed persons

Means and standard errors of metric variables are reported in this table. The sample is restricted to those aged 16–45 years. Observations vary due to missing values in the variables in addition to sample restriction. Survey jackknife method was used to estimate standard errors for weighted mean values. Source: Author's calculation based on HILDA Release 11

The mean values of the diversity indices in the sample are 0.36, indicating a fairly large concentration of foreign nationals in the HILDA data. On average, the LGA level share of foreign nationals in Australia was as high as 23 %, roughly double the 2011 OECD average of 12.5 % (OECD 2013).

Additional descriptive statistics to compare Table 1 with the 2001 and 2011 censuses are reported in Table 8 in the Appendix. For brevity, we will not discuss the table in detail. However, we note that the proportions reported for some of the variables are roughly equivalent to those reported for the HILDA data. The distribution of ancestry and geographic origin in the 2001 census are equivalent to the unrestricted HILDA sub-sample while the distribution of the HILDA variables gender, state, and section of state are equivalent to both censuses. In case of marital status, the categories “separated” and “never married” in the HILDA differ from the 2001 census by 5 and 3.6 %, respectively.

#### 4.2 Multivariate analysis

Table 3 reports the wage effects of diversity and other covariates estimated using ordinary least squares (OLS) and fixed effect (FE) models. The OLS results indicate that cultural diversity has a positive impact on weekly wages while the FE results indicate no impact on wages. Although we found results that are robust to estimator type in annualised diversity measure (not reported here), introducing a 3-year lag to diversity appears to eliminate the positive results in FE estimation. In the OLS estimates, column 1 indicates that cultural diversity has strong positive effect on weekly wages for the whole sample controlling for age, age-squared, gender, and marital status ( $\beta = 0.20$ ,  $p < 0.01$ ). Column 2 introduces human capital variables including English language skill, education,

**Table 3** Weighted OLS and FE estimates showing the impact of diversity on wages. Dependent variable: log of weekly wages

	OLS				Fixed effect		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Diversity index	0.199*** (0.029)	0.091*** (0.021)	0.357*** (0.063)	0.279*** (0.064)	−0.012 (0.017)	−0.020 (0.014)	−0.059 (0.045)
Diversity index—squared	–	–	−0.318*** (0.069)	−0.262*** (0.070)	–	–	0.045 (0.049)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Human capital controls		Yes	Yes	Yes		Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual effects					Yes	Yes	Yes
Movers (0 = not moved; 1 = moved)				Yes			
Ancestry (0 = foreign-born; 1 = Australian-born)				Yes			
Residence (0 = rural; 1 = urban)				Yes			
Observations	20,238	20,156	20,156	20,055	20,241	20,159	20,159
F-statistic	164.65	250.39	243.07	240.1	182.60	229.56	219.37
R-squared	0.222	0.594	0.595	0.607	0.053	0.327	0.329

Values are estimated for the HILDA wave 4–11 respondents aged 16–45 years. Robust standard errors are in parentheses and are adjusted for clusters by individuals. The reported sample sizes differ from the original ( $n = 44,634$ ) due to sample weighting and lagging. Demographic controls include age and age-squared in all models, plus gender and marital status in the OLS models. Human capital controls include education, weekly hours worked, and job tenure, in all models, plus English language skill in OLS

Significance values indicate \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

weekly hours worked, and job tenure with the result still indicating a strong positive relationship but smaller coefficient ( $\beta = 0.09$ ,  $p < 0.01$ ). Further introduction of a non-linear specification in column 3 indicates the persistence of a strong positive relationship with a larger coefficient size, with the quadratic coefficient indicating an upper bound for the positive effect of diversity.

Column 4 introduces additional controls including a dummy variable for movers, ancestry, and a person's residence to account for heterogeneity in the impact of diversity. The dummy variable "movers" (=1 if a person ever moved to another LGA) controls for self-selection bias that can arise due to the movement of individuals to places with better paying regions. "Ancestry" refers to whether a respondent is Australian-born (ancestry = 1) or a migrant (ancestry = 0). The residency dummy variable classifies respondents into rural (residence = 0) and urban residents (residence = 1). Controlling for these variables appears to have minimal effect on the results, with just a slight decline in the coefficient size.

The saturated OLS model (column 4,  $F = 240.1$ ) explained 60.7 % of the variance with diversity having a statistically significant effect. The corresponding FE models (columns 5–7) all indicated no relationship between cultural diversity and weekly wages. This indicates that controlling for individual differences eliminated the positive effects of diversity. The result corroborates previous OLS and FE results obtained using UK data (Longhi 2013). Although contemporaneous diversity appears to have strong effect on weekly wages, it is not robust to the estimator type when the dependent variable is lagged. In an alternative analysis where contemporaneous diversity is categorised (results not shown), the authors found diversity to be statistically significant in the OLS but not in the FE estimation. This contradicts the expectation that higher levels of diversity yield negative labour market outcomes due to communication issues and polarisation that can possibly arise in the workplace (Zhan et al. 2015).

Furthermore, we run alternative regressions using the share of migrants instead of the diversity index. Table 4 which reports similar results indicates a strong positive relationship between migrant share and log weekly wages. As in Table 3, the share of migrants in HILDA yields strongly significant coefficients in the OLS models ( $\beta = 0.33$ ,  $p < 0.1$ ,  $F = 238.1$  for the saturated model and  $\beta = 0.44$ ,  $p < 0.1$ ,  $F = 243.8$  for the non-linear model). The saturated model fit (column 4) explains 60.7 % of the variance in the wage regression. However, none of the FE models show a significant association between migrant share and weekly wages.

Overall, both Tables 3 and 4 indicate that the relationship between cultural diversity or the share of migrants and weekly wages among HILDA respondents aged 16–46 years varies depending on the estimator type used. We find consistent relationship in the OLS but not in the FE models. The effect in the OLS models are stronger when non-linearity is accounted for, indicating a decline in the effect of diversity beyond a certain point, and slightly smaller when heterogeneity is controlled for. The coefficients for movers, ancestry, and residence are statistically significant indicating variation in the effect of diversity for different groups.

#### 4.3 Instrumental variable estimates

Diversity based on country of birth data is considered to be endogenous as economic opportunities can attract people from different countries. This can then

**Table 4** Weighted OLS and FE estimates showing the impact of share of migrants on wages.  
Dependent variable: log of weekly wages

Explanatory variables	OLS				Fixed effect		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Migrant share	0.311*** (0.046)	0.166*** (0.032)	0.440*** (0.074)	0.330*** (0.074)	-0.005 (0.027)	-0.014 (0.022)	-0.063 (0.053)
Migrant share-squared	–	–	-0.397*** (0.086)	-0.307*** (0.084)	–	–	0.076 (0.063)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Human capital controls		Yes	Yes	Yes		Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual effects					Yes	Yes	Yes
Movers (0 = not moved; 1 = moved)				Yes			
Ancestry (0 = foreign-born; 1 = Australian-born)				Yes			
Residence (0 = rural; 1 = urban)				Yes			
Observations	19,876	19,795	19,795	19,794	19,878	19,797	10,797
F-statistic	158.54	249.80	243.77	238.09	178.20	223.15	212.04
R-squared	0.222	0.594	0.595	0.607	0.053	0.325	0.326

Values are estimated for the HILDA wave 4–11 respondents aged 16–45 years. Robust standard errors are in parentheses and are adjusted for clusters by individuals. The reported sample sizes differ from the original ( $n = 44,634$ ) due to sample weighting and lagging. Demographic controls include age and age-squared in all models, plus gender and marital status in the OLS models. Human capital controls include education, weekly hours worked, and job tenure, in all models, plus English language skill in OLS

Significance values indicate \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

result in more prosperous regions with higher weekly wages becoming culturally more diverse rather than diversity causing higher weekly wages. Our findings (Table 3) show that the models based on the diversity index generated from the HILDA data appear to indicate the existence of endogeneity. An endogeneity test of the wage model with the diversity index as an independent variable yields a Durbin score chi-squared of 164.6 ( $p < 0.01$ ) which implies that the diversity index is endogenous. This is also the case with the share of migrants which yields a Durbin score chi-squared of 36.6 ( $p < 0.01$ ). Instrumental variable estimation is therefore specified for both the OLS and FE models. Table 5 reports the IV results.

Models 1 and 3 report results estimated using the predicted diversity index (shift-share) instrument based on census data and population estimates with a 3-year lag. In all the models, the results in Tables 3 and 4 are replicated consistently. The index of diversity is strongly significant indicating a positive impact of diversity on weekly wages in the OLS model ( $\beta = 0.63$ ,  $p < 0.01$ ). For the FE model, we detect no relationship. In comparison to the corresponding models in Tables 3 and 4, instrumenting has increased the coefficient size in the IV-OLS model. Yet, the explanatory power is not affected. From the overidentification restriction tests of the instrument reported, it can be seen that both models are overidentified and the null hypotheses ( $H_0$ : the models are underidentified) can be rejected at less than the 1 % level. Further, in both models, the presence of weak instruments can be rejected at the 1 % level, implying that the instruments are relevant.

**Table 5** Instrumental variable estimation showing the impact of diversity on wages. Dependent variable: log of weekly wages

Explanatory variables	Whole sample			
	IV-OLS		IV-fixed effects	
	(1)	(2)	(3)	(4)
Diversity index	0.625*** (0.074)		0.203 (0.168)	
Share of foreigners		0.693*** (0.084)		0.159 (0.123)
Demographic controls	Yes	Yes	Yes	Yes
Human capital controls	Yes	Yes	Yes	Yes
Time dummies <sup>a</sup>	Yes	Yes	Yes	Yes
Individual effects	No	No	Yes	Yes
Observations	20,156	19,795	19,706	19,335
R-squared	0.565	0.581	0.347	0.359
First stage for diversity index				
Predicted diversity index	0.614*** (0.016)		0.245*** (0.030)	
First stage for share of foreigners				
Predicted share of foreigners		0.835*** (0.021)		0.594*** (0.039)
Instrument tests				
Underidentification test <sup>b</sup>	622.98	626.47	52.32	140.65
p value	0.000	0.000	0.000	0.000
Weak identification test <sup>c</sup>	1454.11	1637.39	66.10	229.06

Values are estimated for the HILDA wave 4–11 respondents aged 16–45 years. Robust standard errors are in parentheses and are adjusted for clusters by individuals. The reported sample sizes differ from the original ( $n = 44,634$ ) due to sample weighting and lagging. Demographic controls include age and age-squared in all models, plus gender and marital status in the OLS models. Human capital controls include education, weekly hours worked, and job tenure, in all models, plus English language skill in OLS

Significance values indicate \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

<sup>a</sup>Time dummies are partialled out in the FE models

<sup>b</sup>Based on the Kleibergen-Paap rk LM statistic

<sup>c</sup>Based on the Kleibergen-Paap rk Wald F statistic: Stock-Yogo weak ID test critical value maximal IV sizes range as follows: 16.38 (10 %), 8.96 (15 %), 6.66 (20 %), and 5.53 (25 %)

Columns 2 and 4 report IV models estimated using the predicted share of migrants. The results are similar to those for the index of diversity replicating the findings in Table 4. While no relationship is evident in the FE model, coefficients in the IV-OLS model ( $\beta = 0.69$ ,  $p < 0.01$ ) indicate that the share of migrants is positively related to the log of weekly wages. Relevance is also maintained, with the null that the models are underidentified and that the instrument is weak satisfactorily rejected at less than the 1 % level.

Generally, instrumenting yields statistically significant OLS coefficients and has corrected the endogeneity with no effect on the FE models. Overall, our results indicate that there is causal relationship between cultural diversity and weekly wages at the regional level, after correcting for the possibility of endogeneity through the shift-share method using predicted population from the baseline data rather than using actual population data. However, the results are not robust to the type of estimator we use.

#### 4.4 Heterogeneity tests

The effect of diversity can vary for individuals depending on their individual and group characteristics, such as ancestry, skill level, self-selection due to mobility, and residency. To account for these heterogeneities, we re-estimate the wage model for different sub-samples. Tables 6 and 7 report the results for the re-estimated models.

Columns 1–4 of Table 6 report findings comparing the Australian-born and migrant samples. The OLS models (columns 1 and 2) indicate that the index of diversity is a strong positive predictor of weekly wages for both samples with the effect on wages larger among those born in Australia ( $\beta = 0.65$ ,  $p < 0.01$ ) than among migrants ( $\beta = 0.44$ ,  $p < 0.01$ ). The FE model is marginally significant for the Australian-born sample while not statistically significant for the migrant sample.

Columns 5–8 compare the findings for those who moved between LGAs in at least 1 year and those who did not move throughout the HILDA Survey. Again, the OLS models indicate strong positive relationship with the effect of diversity on weekly wages slightly larger among movers ( $\beta = 0.64$ ,  $p < 0.01$ ) than non-movers ( $\beta = 0.61$ ,  $p < 0.01$ ). However, the FE models are not statistically significant for both groups.

**Table 6** Weighted OLS and FE estimates comparing the impact of diversity on wages for different groups. Dependent variable: log of weekly wages

Explanatory variables	Australian-born sample		Migrant sample		Movers		Non-movers	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Diversity index	0.650*** (0.087)	0.340* (0.180)	0.436*** (0.127)	−0.318 (0.320)	0.639*** (0.086)	0.180 (0.138)	0.607*** (0.127)	−0.089 (0.510)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	16,756	16,410	3400	3296	12,284	12,044	7872	7662
Adjusted <i>R</i> -squared	0.565	0.318	0.569	0.376	0.561	0.366	0.574	0.339
First stage for diversity index								
Predicted diversity index	0.599*** (0.018)	0.260*** (0.030)	0.664*** (0.038)	0.292*** (0.098)	0.573*** (0.020)	0.338*** (0.035)	0.679*** (0.027)	−0.162*** (0.039)
Instrument tests								
Underidentification test <sup>b</sup>	512.43	56.74	99.50	6.98	350.60	68.57	290.56	16.87
<i>p</i> value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Weak identification test <sup>c</sup>	1113.17	75.02	312.33	8.92	846.59	95.97	628.93	17.24

Values are estimated for the HILDA wave 4–11 respondents aged 16–45 years. Robust standard errors are in parentheses and are adjusted for clusters by individuals. The reported sample sizes differ from the original ( $n = 44,634$ ) due to sample weighting and lagging. Demographic controls include age and age-squared in all models, plus gender and marital status in the OLS models. Human capital controls include education, weekly hours worked, and job tenure, in all models, plus English language skill in OLS

Significance values indicate \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

<sup>a</sup>Time dummies are partialled out in the FE models

<sup>b</sup>Based on the Kleibergen-Paap rk LM statistic

<sup>c</sup>Based on the Kleibergen-Paap rk Wald *F* statistic: Stock-Yogo weak ID test critical value maximal IV sizes range as follows: 16.38 (10 %), 8.96 (15 %), 6.66 (20 %), and 5.53 (25 %)

**Table 7** Weighted OLS and FE estimates comparing the impact of diversity on wages for different groups. Dependent variable: log of weekly wages

Explanatory variables	Skilled		Unskilled		Urban residents		Rural residents	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Diversity index	0.722*** (0.111)	−0.046 (0.234)	0.437*** (0.089)	0.222 (0.245)	0.428*** (0.078)	0.036 (0.164)	0.648** (0.276)	−0.027 (0.309)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Human capital controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies <sup>a</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	10,345	9759	9779	9170	17,758	17,321	2397	2143
Adjusted <i>R</i> -squared	0.460	0.237	0.612	0.337	0.623	0.389	0.545	0.281
First stage for diversity index								
Predicted diversity index	0.599*** (0.021)	0.240*** (0.044)	0.625*** (0.024)	0.241*** (0.046)	0.629*** (0.016)	0.264*** (0.033)	0.499*** (0.050)	−0.399*** (0.110)
Instrument tests								
Underidentification test <sup>b</sup>	398.33	24.01	344.38	24.17	586.48	52.60	41.89	11.49
<i>p</i> value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Weak identification test <sup>c</sup>	813.93	29.61	706.87	27.76	1453.03	64.92	97.93	13.15

Values are estimated for the HILDA wave 4–11 respondents aged 16–45 years. Robust standard errors are in parentheses and are adjusted for clusters by individuals. The reported sample sizes differ from the original ( $n = 44,634$ ) due to sample weighting and lagging. Demographic controls include age and age-squared in all models, plus gender and marital status in the OLS models. Human capital controls include education, weekly hours worked, and job tenure, in all models, plus English language skill in OLS

Significance values indicate \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

<sup>a</sup>Time dummies are partialled out in the FE models

<sup>b</sup>Based on the Kleibergen-Paap rk LM statistic

<sup>c</sup>Based on the Kleibergen-Paap rk Wald *F* statistic: Stock-Yogo weak ID test critical value maximal IV sizes range as follows: 16.38 (10 %), 8.96 (15 %), 6.66 (20 %), and 5.53 (25 %)

A similar analysis in Table 7 compares two sets of groups based on skill (columns 1–4) and residency (columns 5–8). For the first set of groups, i.e. skilled vs. unskilled groups, we found positive effect of cultural diversity on weekly wages in the OLS model. The effect of diversity is larger among those who are skilled ( $\beta = 0.72$ ,  $p < 0.01$ ) by a factor of 1.7 compared to the unskilled ( $\beta = 0.42$ ,  $p < 0.01$ ). Similarly, the second set, urban vs. rural residents indicate a positive effect of diversity in OLS, with effects larger among rural ( $\beta = 0.65$ ,  $p < 0.01$ ) compared to urban residents ( $\beta = 0.41$ ,  $p < 0.01$ ). However, in the IV-FE models, although the instruments are strongly significant, the coefficients of both measures are not statistically significant at the 5 % level.

## 5 Conclusions

Diversity is a complex concept as it varies depending on the context in which it is studied. It can be expressed in the form of differences in race, linguistic background, national origin, ethnic background, or culture. This fluid notion of what constitutes diversity has been addressed by researchers using a range of conceptualisations. Vigdor (2008) uses culture to define diversity, focusing on the latter as a “measure of cultural dissimilarity” between groups or individuals. Alesina and La Ferrara (2004) use ethnicity in measuring diversity as a fractionalisation based on ethnic origin. Alesina

et al. (2013), Bellini et al. (2013), Longhi (2013), and Ottaviano and Peri (2006) take the country of birth to construct an index of diversity. In all of these studies, diversity is conceptualised as a continuum of dissimilarity which ranges from perfect homogeneity to perfect heterogeneity.

This study applied the Ottaviano and Peri (2006) version of diversity (based on country of birth) to Australian data. This has some limitations in terms of capturing detailed differences in race, ethnicity, and language existing due to the heterogeneity of migrants from any particular country of origin. However, country of birth is an important source of diversity in Australia where race and ethnicity categories are not regularly used in labour market analyses. This is also the case in the HILDA Survey which collects the most comprehensive Australian labour-market-related household data.

Although there is underrepresentation of migrants in HILDA (21.3 % migrants), at least two patterns emerge from this study. First, on the average, there is substantial diversity in Australian regions, as high as 0.36. However, there is high degree of variability in the index of diversity at the LGA level within the  $[0, 1]$  range. Second, the effect of diversity, as measured by the degree of concentration of foreign nationals at the LGA level, on wages varies depending on the model specification. All OLS regressions indicate a strong positive impact of diversity on weekly wages among HILDA respondents. On average, respondents who reside in more diverse environment tend to earn better weekly wages than those who live in more homogenous regions. This result shows that more than half of the variation is explained by the specified models. However, although these findings appear to be robust in non-linear specifications with time and a range of other controls, they are not robust when a fixed effect specification is specified.

Generally, the results of this longitudinal study replicate two aspects of previous, mostly cross-sectional findings, using US and European data. First, our main results partially corroborate Ottaviano and Peri (2006) who found that a substantial rise in wages was experienced by US citizens living in areas where the share of migrants rose between 1970 and 1990. Similar findings were obtained by Bellini et al. (2013) for European regions constituting 12 countries in a study that used regional GDP per capita as a proxy for regional wages. Second, the findings that indicate a positive effect of diversity tend to disappear after controlling for individual fixed effects. This robustness issue supports the findings reported by Longhi (2013).

The main concerns with this kind of analysis, as discussed in the literature, are endogeneity and selection bias. We addressed the problem of endogeneity by applying instrumental variable estimation to 10 years of longitudinal data using the shift-share method. We used the predicted rather than actual fractionalisation data to instrument for diversity index, with results that are consistent after instrumenting. Using a spatial analysis of cultural diversity, employed in only a few studies to date (Longhi 2013; Ottaviano and Peri 2006; Pischke and Velling 1997), we found the positive effect of diversity on weekly wages based on country of birth as reported in studies to be limited by the estimator type used. This is particularly the case with the introduction of a 3-year lag to the index of diversity which yields valid instruments and statistically significant OLS coefficients. In this estimation,

people living in relatively more diverse regions tend to earn better wages than those who live in less diverse regions. To account for selection bias, we controlled for those who moved LGAs. Further, we controlled for heterogeneity issues by including ancestry, skill differences, and residency in addition to other demographic and human capital factors. Our results appear to be robust to self-selection and heterogeneity issues although the models were robust only in OLS while the FE models were not robust.

Finally, our findings also indicate that although contemporaneous levels of cultural diversity strongly and positively affect weekly wages, the impact of previous levels of diversity on weekly wages is not conclusive. This may explain the variation in the literature where some studies report that cultural diversity increases wages across time (Bellini et al. 2013; Ottaviano and Peri 2006) while others do not (Longhi 2013). Future research should explore this further to examine whether the type of data used to construct diversity and the time lag employed has an effect on the outcome of diversity in studies.

## Endnotes

<sup>1</sup>The phrase cultural diversity is used throughout this paper in reference to a heterogeneous group of people making a society. We use the country of birth as a measure although ethnicity, race, religion, language, and/or nationality can also be used to assess this form of diversity.

<sup>2</sup>Attrition is relatively higher among single persons, unemployed, younger (15–24 year olds), low skilled, and Indigenous people.

<sup>3</sup>Although 2156 postcodes (including “offshore” and “no usual address”) are available in the 2006 and 2011 censuses, seven postcodes have missing values. The ABS’s TableBuilder instrument enables researchers to build population-level tables of diverse demographic and socio-economic issues based on the 2006 and 2011 censuses. The ABS provided us with a country of origin by postal area data for the 2001 census. However, between the 2001 and 2011 data, there is a discrepancy of 206 postcodes, with 63 postcodes excluded in 2011 while 143 additional postcodes included.

<sup>4</sup>These datasets are not disaggregated by region of residence. Therefore, they only indicate the total residents and new arrivals irrespective of their residence or mobility inside Australia.

<sup>5</sup>The country of birth data is arguably a crude indicator of cultural diversity in that Australian-born respondents have a range of different ethnicities (and to a lesser extent languages) that are not accounted for in this study. In addition, the contribution of migrants to diversity will be overestimated to some extent in that some of them share ethnicity and language with Australian-born respondents.

<sup>6</sup>The annual population estimates were obtained from annual population estimates from the ABS creative commons web page.

<sup>7</sup>NB: the index measures diversity at the LGA level although, originally, the individual identifier in HILDA is the postcode.

<sup>8</sup>The observations vary from the restricted final sample due to missing values and population weighting.

## Appendix

### 5.1 Descriptive analysis of two Australian censuses

**Table 8** Demographic characteristics

Variables	Census 2001 N or %	Census 2011 N or %
Total population <sup>b</sup>	18,769,249	21,507,717
Gender (Female) <sup>a,b</sup>	50.6	50.6
Ancestry (Australian-born) <sup>b</sup>	78.1	69.8
Geographic origin <sup>c</sup>	17,578,932	20,258,695
Australian	74.4	71.3
Indigenous	2.3	2.7
Asia-Pacific	8.2	11.6
Europeans and North Americans	10.1	9.1
East Europeans	2.4	1.9
Middle East, Africans, and Latin Americans	2.6	3.5
Age <sup>b</sup>	18,769,249	21,507,719
Under 15 years	20.8	19.3
15–24 years	13.7	13.3
25–54 years	43.5	41.8
55–64 years	9.4	11.6
65 years and over	12.6	14
Parents' ancestry	17,735,144	20,030,163
At least one parent born overseas	78.4	46.2
Both parents born in Australia	19.6	53.7
Not stated	2.0	-
Marital status <sup>a,b</sup>	14,856,774	17,363,696
Married	51.4	48.7
Separated/divorced/widowed	17.0	16.9
Never married	31.6	34.3
Employment status <sup>a,b</sup>	14,224,741	17,363,696
Employed (full time)	37.7	36.7
Employed (part time)	18.9	17.6
Unemployed	4.6	3.5
Not in labour force	37.0	38.6
State <sup>d</sup>	18,755,867	21,504,382
New South Wales	33.6	32.1
Victoria	24.6	24.9
Queensland	19.1	20.1
South Australia	7.8	7.4
Western Australia	9.8	10.4
Tasmania	2.4	2.3
Northern Territory	1.1	1.0
Australia Capital Territory	1.6	1.9

**Table 8** Demographic characteristics (*Continued*)

Section of state	18,755,867	21,504,382
Major urban	65.1	69.4
Other urban	22.1	19.8
Bounded locality	2.6	2.7
Rural balance	10.3	8.2

This table presents proportions for a range of demographic variables in the 2001 and 2011 censuses

<sup>a</sup>Persons are aged 15 and above

<sup>b</sup>Source: ABS (2001)

<sup>c</sup>Authors' calculation from census data. Values are calculated excluding overseas visitors, observations with inadequately described values, and those who did not state their countries of birth

<sup>d</sup>Source: ABS (2003)

### Competing interests

The IZA Journal of Migration is committed to the IZA Guiding Principles of Research Integrity. The authors declare that they have observed these principles.

### Acknowledgements

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this article, however, are those of the authors and should not be attributed to either FaHCSIA or the Melbourne Institute. The first author was supported by an Australian Postgraduate Award (Industry) as part of linkage project LP100200057 funded by the Australian Research Council (ARC), Victorian Health Promotion Foundation, and the Australian Human Rights Commission. He is also grateful to Mrs. Lidia Abraham for her assistance in the analysis of this research. The second author is supported by an ARC Future Fellowship (FT130101148). We would like to thank an anonymous referee and the editor for the useful remarks. Responsible editor: Denis Fougère

Received: 12 November 2015 Accepted: 31 March 2016

Published online: 28 June 2016

### References

- ABS. Census QuickStats. Canberra: Australian Bureau of Statistics; 2001. [www.abs.gov.au](http://www.abs.gov.au), Accessed 19 Jan 2016.
- ABS. Census of population and housing: population growth and distribution 2001, ABS Catalogue No. 2035.0, Canberra: Australian Bureau of Statistics; 2003.
- ABS. Standard Australian Classification of Countries (SACC), 2011. Canberra: Australian Bureau of Statistics; 2011a.
- ABS. Australian Standard Geographical Classification (ASGC) Digital Boundaries, Australia, July 2011. Canberra: Australian Bureau of Statistics; 2011b.
- ABS. Census of population and housing 2011. Canberra: Australian Bureau of Statistics; 2012a.
- ABS. Reflecting a nation: stories from the 2011 census, 2012–2013. Canberra: Australian Bureau of Statistics; 2012b.
- ABS. Australian social trends. Canberra: Australian Bureau of Statistics; 2014.
- Alesina A, La Ferrara E. Ethnic diversity and economic performance. Cambridge: National Bureau of Economic Research; 2004.
- Alesina A, La Ferrara E. Ethnic diversity and economic performance. *J Econ Lit*. 2005;43:762–800.
- Alesina A, Harnoss J, Rapoport H. Birthplace diversity and economic prosperity. Cambridge: National Bureau of Economic Research; 2013.
- Altonji JG, Card D. The effects of immigration on the labor market outcomes of less-skilled natives. In: Abowd JM, Freeman RB, Editors. *Immigration, trade, and the labor market*. Chicago: University of Chicago Press; 1991. pp. 201–34.
- Altonji JG, Blank RM. Race and gender in the labor market. In: Ashenfelter OC, Editor. *Handbook of labor economics*, vol 3. North-Holland, Amsterdam, the Netherlands: North-Holland; 1999. pp. 3143–259.
- Angrist JD, Kugler AD. Protective or counter-productive? Labour market institutions and the effect of immigration on EU natives. *Econ J*. 2003;113:F302–31.
- Baltagi B. *Econometric analysis of panel data*, Vol. 4, New York: John Wiley & Sons; 2008.
- Bellini E, Ottaviano GI, Pinelli D, Prarolo G. Cultural diversity and economic performance: evidence from European regions. In: *Geography, institutions and regional economic performance*. Berlin: Springer; 2013. p. 121–41.
- Bertone S, Leahy M. Social equity, multiculturalism and the productive diversity paradigm: reflections on their role in corporate Australia. In: Phillips SK, editor. *Everyday diversity: Australian multiculturalism in practice*. Altona: Common Ground Publishing Pty. Ltd; 2001. p. 113–44.
- Borjas GJ, Freeman RB, Katz LF, DiNardo J, Abowd JM. How much do immigration and trade affect labor market outcomes? *Brookings Pap Econ Act*. 1997;1997:1–90.
- Borjas GJ, Grogger J, Hanson GH. Imperfect substitution between immigrants and natives: a reappraisal. Cambridge: National Bureau of Economic Research; 2008.
- Campos NF, Saleh A, Kuzeyev V. Dynamic ethnic fractionalization and economic growth. *J Int Trade Econ Dev*. 2011;20:129–52.
- Card D. Immigrant inflows, native outflows, and the local market impacts of higher immigration. *J Labor Econ*. 2001;19:22–61.

- D'Amuri F, Ottaviano GI, Peri G. The labor market impact of immigration in Western Germany in the 1990s. *Eur Econ Rev.* 2010;54:550–70.
- Damelang A, Haas A. The benefits of migration: cultural diversity and labour market success. *Eur Soc.* 2012;14:362–92.
- Dustmann C, Fabbri F. Language proficiency and labour market performance of immigrants in the UK. *Econ J.* 2003;113:695–717.
- Fearon JD. Ethnic and cultural diversity by country. *J Econ Growth.* 2003;8:195–222.
- Friedberg RM, Hunt J. The impact of immigrants on host country wages, employment and growth. *J Econ Perspect.* 1995;9:23–44.
- Gomez-Mejia LR, Palich LE. Cultural diversity and the performance of multinational firms. *J Int Bus Stud.* 1997;28:309–35.
- Herring C. Does diversity pay? Race, gender, and the business case for diversity. *Am Sociological Rev.* 2009;74:208–24.
- Hugo G, Harris K. Population distribution effects of migration in Australia. Adelaide: University of Adelaide; 2011.
- Hunt J. The impact of the 1962 repatriates from Algeria on the French labor market. *Ind Labor Relations Rev.* 1992;45:556–72.
- OECD. International migration outlook 2013. Paris: OECD Publishing; 2013.
- Juvonen J, Nishina A, Graham S. Ethnic diversity and perceptions of safety in urban middle schools. *Psychological Sci.* 2006;17:393–400.
- Kochan T et al. The effects of diversity on business performance: report of the diversity research network. *Hum Resour Manage.* 2003;42:3–21.
- Kohler P. Economic discrimination and cultural differences as barriers to migrant integration: is reverse causality symmetric? Geneva: Graduate Institute of International and Development Studies Working Paper; 2012.
- Lieber LD. The hidden dangers of implicit bias in the workplace. *Employment Relations Today.* 2009;36:93–8.
- Longhi S. Impact of cultural diversity on wages, evidence from panel data. *Reg Sci Urban Econ.* 2013;43:797–807.
- Montalvo JG, Reynal-Querol M. Ethnic diversity and economic development. *J Dev Econ.* 2005;76:293–323.
- Nathan M. The economics of super-diversity: findings from British cities, 2001–2006. London: Spatial Economics Research Centre, London School of Economics; 2011.
- Neumann R, Graeff P. Method bias in comparative research: problems of construct validity as exemplified by the measurement of ethnic diversity. *J Math Sociology.* 2013;37:85–112.
- Ottaviano GI, Peri G. The economic value of cultural diversity: evidence from US cities. *J Econ Geogr.* 2006;6:9–44.
- Page SE. The difference: how the power of diversity creates better groups, firms, schools, and societies. Princeton: Princeton University Press; 2008.
- Perotin V, Robinson A, Loundes J. Equal opportunities practices and enterprise performance: a comparative investigation on Australian and British data. *Int Labour Rev.* 2003;142:471–505.
- Pischke J-S, Velling J. Employment effects of immigration to Germany: an analysis based on local labor markets. *Rev Econ Stat.* 1997;79:594–604.
- Putnam RD. *E pluribus unum: diversity and community in the twenty-first century* the 2006 Johan Skytte Prize Lecture. *Scand Political Stud.* 2007;30:137–74.
- Richard OC. Racial diversity, business strategy, and firm performance: a resource-based view. *Acad Manage J.* 2000;43:164–77.
- Roberson L, Kulik CT. Stereotype threat at work. *Acad Manage Perspect.* 2007;21:24–40.
- Sanderson MR. Does immigration promote long-term economic development? A global and regional cross-national analysis, 1965–2005. *J Ethnic Migration Stud.* 2013;39:1–30.
- Stahl GK, Maznevski ML, Voigt A, Jonsen K. Unraveling the effects of cultural diversity in teams: a meta-analysis of research on multicultural work groups. *J Int Bus Stud.* 2010;41:690–709.
- Suedekum J, Wolf K, Blien U. Cultural diversity and local labour markets. *Reg Stud.* 2014;48:173–91.
- Triana MC, Jayasinghe M, Pieper JR. Perceived workplace racial discrimination and its correlates: a meta-analysis. *J Organizational Behav.* 2015;36:491–513.
- Vigdor JL. Measuring immigrant assimilation in the United States. New York: Manhattan Institute for Policy Research; 2008.
- Watson N. Longitudinal and cross-sectional weighting methodology for the HILDA Survey. 2012.
- Wooldridge JM. Econometric analysis of cross section and panel data. 2nd ed. Cambridge: MIT Press; 2010.
- Wrench J. Diversity management can be bad for you. *Race Class.* 2005;46:73–84.
- Zhan S, Bendapudi N, Yy H. Re-examining diversity as a double-edged sword for innovation process. *J Organizational Behav.* 2015;36:1026–49.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)