Life Course Centre

The Australian Research Council Centre of Excellence for Children and Families over the Life Course

Phone +61 7 3346 7477 Email lcc@uq.edu.au lifecoursecentre.org.au

WORKING PAPER SERIES

AGGLOMERATION ECONOMIES AND THE URBAN WAGE PREMIUM IN AUSTRALIA

Jordy Meekes

Melbourne Institute: Applied Economic & Social Research

No. 2021-18

September 2021

Australian Government Australian Research Council

The University Of Queensland Australia

The University of Melbourne

The University of Western Australia

The University of Sydney

NON-TECHNICAL SUMMARY

This study is the first to quantify the economic impact of urban density on individual wages in Australia. The international literature shows a positive urban density effect on wages, referred to as the urban wage premium. This result is explained by agglomeration mechanisms that affect the benefits and costs of the spatial concentration of economic activity. The benefits of density lead to improved matching of employers to employees, better sharing of resources and risk among companies; and improved learning of knowledge by employees. The costs of density include crowding and frictions such as increased congestion in urban areas. This paper contributes to the literature by analysing the urban wage premium in Australia, a country which has a wide range of urban areas of different sizes that are spatially separate from one another.

Ultimately, the size and net effect of agglomeration is an empirical question. Individual microdata are used from the longitudinal survey: Household Income and Labour Dynamics in Australia Survey 2001-2019. The paper assesses costs and benefits of agglomeration by estimating the impact of population relative to spatial area size on individual wages. It also uses commuting flows to define spatial aggregation to estimate the urban wage premium. This methodology enables a comparison of the urban wage premium across different spatial scales, including Statistical Area Level 2 and Local Government Area. This is a novel methodology for studying the significance of place for socio-economic disadvantage.

It was found that the Ordinary Least Squares estimate of the urban wage premium peaks at 2.7 per cent, whereas the estimate peaks at 1.6 per cent after controlling for individual fixed effects. This evidence suggests that wages increase by 1.6 to 2.7 per cent if the population density doubles: in other words, density has a significant positive impact on wages. This paper adds to our understanding of the costs and benefits of density and the variables that enable workforce participation and employment outcomes. The paper provides a strong baseline from which to consider changes underway in the COVID-19 era where hybrid models of office and home-based employment are becoming more widespread. The winners and losers of the changing geographies of work will have implications for socio-economic advantage and disadvantage.

ABOUT THE AUTHORS

Jordy Meekes is a Research Fellow (Assistant Professor) in the Melbourne Institute: Applied Economic & Social Research at the University of Melbourne, a research affiliate of the IZA, and a research fellow of the ARC Centre of Excellence for Children and Families over the Life Course. Jordy obtained his PhD in economics at the Utrecht University School of Economics, Utrecht University. His research is in applied microeconomics at the intersection of labour economics with spatial economics. Email: jordy.meekes@unimelb.edu.au

Acknowledgements

The author thanks participants at the Australian Conference of Economists 2021 and the special symposium on Recent Advances in Spatial Econometrics hosted by the University of Melbourne. The author also thanks, among others, Guyonne Kalb, Kate Gilkison, Sangeetha Chandrashekeran, Bram van Besouw and Vincent Schippers for helpful comments. This paper uses unit record data from Restricted Release 19 of the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Survey is conducted by the Melbourne Institute: Applied Economic & Social Research on behalf of the Australian Government Department of Social Services (DSS). The findings and views reported in this paper are those of the author and should not be attributed to the Australian Government, DSS or the Melbourne Institute. This research was supported by the Australian Research Council Centre of Excellence for Children and Families over the Life Course (CE140100027).

ABSTRACT

Understanding the benefits of dense agglomerations is important for decisions on where to live and for our understanding of deep economic disadvantage. This paper is the first to quantify the economic impact of urban density on individual wages, referred to as the urban wage premium, in Australia. By combining Household Income and Labour Dynamics in Australia Survey microdata on 13,112 employed individuals and regional-level population data, population density effects on individual hourly wages are studied over the period 2001 to 2019. A unique feature of this paper is to apply a flow-based clustering algorithm that uses commuting flows to define spatial structures, which are compared with the Australian Statistical Geography Standard spatial structures. The Ordinary Least Squares estimate of the urban wage premium peaks at 2.7 per cent. Controlling for individual fixed effects, the estimate peaks at 1.6 per cent. This evidence suggests that wages increase by 1.6 to 2.7 per cent if local density doubles.

Keywords: urban wage premium, agglomeration, population density, wages, Australia, HILDA Survey

Suggested citation: Meekes, J. (2021). 'Agglomeration Economies and the Urban Wage Premium in Australia', Life Course Centre Working Paper Series, 2021-18. Institute for Social Science Research, The University of Queensland.

1. Introduction

Australia is not only one of the largest countries geographically, but also one of the most urbanized countries in the world. In 2019, 86 per cent of the 25 million people in Australia lived in urban areas, slightly higher than the 82 per cent of the population for the United States and 84 per cent for the United Kingdom (The World Bank, 2021). Although the population density across Australia is relatively low, because of people's tendency to live in cities the density of Australian urban areas is relatively high. Australia is also one of the Western countries with the highest relative increase in population, mostly through immigration (Moallemi and Melser, 2020), which has increased the population density in urban areas over the last decades. Interestingly, evidence on the economic effects of density for Australia is very limited. In the era of COVID-19, where a hybrid model of working at home and in the office is considered, it is of paramount importance to have a better understanding of the benefits of living and working in close proximity.

This study fills this gap by quantifying the economic impact of urban density on individual wages in Australia. The literature on spatial economics studies the economic effects of density on labour productivity (Ahlfeldt and Pietrostefani, 2019). The evidence base shows a positive urban density effect on wages, referred to as the urban wage premium. This result is explained by agglomeration mechanisms that affect the benefits and costs of the spatial concentration of economic activity (Duranton and Puga, 2004, 2020; Glaeser and Gottlieb, 2009). The benefits of density lead to improved matching of employers to employees, better sharing of resources and risk among companies and improved learning of knowledge by employees. The costs of density include crowding and frictions such as increased congestion in urban areas. Ultimately, the size and direction of the net density effect is an empirical question.

This paper contributes to the literature by providing the first estimate of the urban wage premium in Australia. In addition, it extends the literature in an important way by analysing the role of spatial unit sizes in the estimate of the urban wage premium. Previous research shows that the aggregation of spatial units into larger units is important for results of empirical analyses.[1] Given Australia's geographical size

[1] For example, see Burger, Van Oort, and Van der Knaap (2010) and Meekes and Hassink (2019) for the Netherlands and Briant, Combes, and Lafourcade (2010) for France.

and high number of spatially separate clusters of urban areas, it is a relevant country to study to what extent the geographical definition of spatial units matters for estimates of the urban wage premium.

To estimate the urban wage premium for Australia, individual microdata on workers are used from the Household Income and Labour Dynamics in Australia (HILDA) Survey over the period 2001 to 2019. The paper approximates agglomeration spillover effects by using variables that represent the population density (constructed by taking the annual estimated resident population relative to the area size in square kilometres) and area size at the spatial unit level (Australian Bureau of Statistics, 2020). The urban wage premium is estimated by the regression of individual log hourly wages on log population density and log area size, while controlling for a rich set of covariates. Several different empirical specifications are used to examine the sensitivity of the urban wage premium.

A methodological contribution of this paper is to define regional aggregations on a continuous scale. Specifically, I define self-contained regional areas of residence and work activity using a hierarchical flow-based cluster algorithm applied on commuting flows across spatial units as a measure of relative interaction. The empirical analysis of the urban wage premium is repeated separately for various Australian Statistical Geography Standard structures, defined by the Australian Bureau of Statistics (ABS), and a continuum of regional aggregations ranging between 2,164 disaggregated and 10 highly aggregated unique spatial units. These different sets of spatial structures are used in the multivariate panel analyses. By using the different sets of spatial structures, the densities and area sizes are different holding all else constant in the regression.

The results show that the pooled Ordinary Least Squares (OLS) estimate of the urban wage premium peaks at 2.7 per cent, whereas the individual Fixed Effects (FE) estimate peaks at 1.6 per cent. This evidence suggests that wages increase by 1.6 to 2.7 per cent if the population density doubles. Internationally, the mean and median effect of density on wages is found to be 4 per cent (Ahlfeldt and Pietrostefani, 2019). Compared to this international evidence, the urban wage premium of 1.6 to 2.7 per cent in Australia is relatively low.

2. Background

In their seminal study, Ciccone and Hall (1996) argue that density effects explain a large share of variation in labour productivity across U.S. states. They find that a doubling of employment density increases labour productivity by 6 per cent. Glaeser and Maré (2001) document that workers in metropolitan areas earn

much higher wages than workers in regional areas of the U.S., a finding that is commonly referred to as the "urban wage premium". Glaeser and Maré examined whether the urban wage premium is a wage growth or a wage level effect, concluding that it is a combination of both. That is, on average, workers not only experience an instantaneous increase in wages after becoming employed in an urban area, but they also experience higher wage growth which increases with the amount of time spent in urban areas. They find an urban wage premium of around 4 to 11 per cent after including individual fixed effects, depending on the empirical specification. Based on a meta-analysis of density effects, Ahlfeldt and Pietrostefani (2019) document a mean and median effect of density on wages of 4 per cent.

The theoretical benefits of micro-foundations of productivity for firms and workers in dense agglomerations are based on three mechanisms: sharing, matching and learning (Duranton and Puga, 2004). Sharing refers to improved sharing of risk and better access to resources and services, matching refers to the improved matching rate and matching quality because of a larger pool of employers and employees; and learning refers to faster learning by workers because of more knowledge spillovers and improved generation and diffusion of knowledge. For a comprehensive overview of the literature on agglomeration economies and density effects, see Duranton and Puga (2004, 2020), Ahlfeldt and Pietrostefani (2019) and Proost and Thisse (2019). In sum, local interactions among workers and firms, which disproportionally take place in dense areas characterised by a high spatial concentration of economic processes, increase productivity.

So far, research on agglomeration benefits in Australia is mainly done from the perspective of the housing market. For example, see Maclennan, Ong, and Wood (2015) for a study of the role of housing in agglomeration benefits, or see Gurran et al. (2021) for a report on the importance of private rental housing for urban productivity. From a labour market perspective, a recent report by Leishman et al. (2021) provides a comprehensive overview of city population effects. For a panel of regional areas from 2011 to 2016, they find a city population size effect on regional income of about 8 per cent. Overall, however, the evidence base on agglomeration benefits for individuals in Australia is limited.

2.1 The empirics of density effects

There are several reasons why identification of density effects is challenging. From an empirical viewpoint, Combes and Gobillon (2015) provide a comprehensive overview and discuss several confounding mechanisms for why wages could be higher in denser areas, including individual-level endogeneity and local-level endogeneity. Individual-level endogeneity is caused by sorting of more productive workers into

denser areas. Specifically, the extent of sorting of more productive workers into denser areas is non-random, and it depends on unobservables such as ability. The estimate of density effects could then be biased by the difference in the composition of individuals' unobserved ability across areas, with high-ability workers making location decisions that direct them to denser areas. A common strategy to limit the individual-level endogeneity is including individual fixed effects, controlling for time-constant unobservable individual attributes.[2]

Local-level endogeneity refers to endogeneity at the spatial unit level caused by missing aggregate variables, for example local amenities or local productivity, which may make firms and workers more likely to move to a specific area. The issue of reverse causality becomes transparent if higher population causes more positive density effects, which in turn increases population. The literature uses an instrumental variable estimator to limit the issue of missing aggregate variables and reverse causality (for example, see Combes, Duranton, and Gobillon (2008) and Mion and Naticchioni (2009)). However, as is well known, finding a convincing instrument that satisfies the exogeneity restriction and relevance condition is difficult. One of the most commonly used instruments in analyses of the urban wage premium is based on historical values of population or population density (Graham, Melo, Jiwattanakulpaisarn, and Noland, 2010). It is argued that the exogeneity assumption holds because of changes in economic activity over a long period of time, which ensures past population does not affect current wages. In addition, it is argued that the relevance condition holds as past population is correlated to current population.

2.2 The modifiable areal unit problem

In spatial economics and economic geography, empirical work focuses on the modifiable areal unit problem, which covers the issue of empirical results being sensitive to the measurement of geographic space (Openshaw and Taylor, 1979; Fotheringham and Wong, 1991). For example, in the context of this paper, the Australian geographic space could be measured in many different ways, for example using the hierarchical Statistical Area (SA) Level 2, SA3 or SA4. The modifiable areal unit problem poses issues

[2] A more convincing strategy may involve a natural or quasi-experimental design that exploits exogenous variation in an aggregate agglomeration variable or individuals' location to identify causal effects of density. However, for a design exploiting exogenous regional-level variation, it can be challenging to find a valid counterfactual for "treated" areas given that the set of observationally equivalent "control" areas is limited. Alternatively, a structural model could be used to jointly model wages and location choices of workers and firms (for example, see Gould (2007) and Baum-Snow and Pavan (2012)).

because of scale effects and zonation effects. The scale effects relate to the size of spatial units that may affect empirical results. The zonation effects are caused by borders of spatial units which are inherently arbitrary.

Several studies analysed the modifiable areal unit problem in the context of identifying agglomeration economies. Work by Briant, Combes, and Lafourcade (2010) using 1976-1996 data for France concludes that scale effects are important but cause less issues than endogeneity issues such as model misspecification because of unobserved heterogeneity. The paper examined scale effects and zonation effects by using six different definitions of geographic space: three based on pre-defined administrative regional classifications (341 “Employment areas”, 94 “Départements” and 21 “Régions”) and three based on grid zoning systems (341 small squares, 91 medium squares and 22 large squares).

For the Netherlands, Burger, Van Oort, and Van der Knaap (2010) confirm that scale effects matter for agglomeration estimates, using spatial units based on three pre-defined administrative classifications (municipality, district and region). More recent work by Meekes and Hassink (2019) is the first to estimate the urban wage premium for a continuum of regional aggregations. They use Dutch data from 2006 to 2014 and repeatedly estimate the urban wage premium for different sets of aggregation, ranging between 398 disaggregated unique spatial units (municipalities) and 13 highly aggregated unique spatial units, which the authors consider to be representative of so-called local labour markets that are characterized by strong commuting connectivity within each labour market and weak connectivity to outside labour markets.

I contribute to this literature by analyzing the urban wage premium for Australia. I apply the same approach as Meekes and Hassink (2019), estimating the urban wage premium for a continuum of regional aggregations. A key limitation of Meekes and Hassink, however, is to apply the approach to the Netherlands, which arguably has only very few local labour markets, given its land area size of 33,671 km^2. In contrast, Australia, with a land area size of 7.692 million km^2, has relatively large variation in how the country can be divided into non-overlapping spatial units while maintaining a sufficiently high number of unique spatial clusters of economic activity.

3. Data, Sample, and Descriptive Statistics

Individual microdata are used from the HILDA Survey, covering 19 waves of data over the period 2001 to 2019. The HILDA Survey is a nationally representative survey that follows households that are interviewed every year.[3] I create an annual panel for workers, pooling the waves from years 2001 to 2019, which contains 13,112 unique workers and 95,670 worker-year observations. The key dependent variable is the hourly wage, which is computed for the main job of the worker by taking the wage relative to the number of working hours. I analyse the natural logarithm of hourly wage in the empirical analysis.

3.1 Sample selection

Individual-year observations are included in the sample of analysis for individuals aged 21 years or older and younger than 65 years. I only keep observations of workers if they work full time. Full-time workers, working more than 35 hours, are selected to ensure changes are captured in hourly wages conditional on being in a full-time job, removing any part-time wage effect on hourly wages caused by transitions between full-time employment and part-time employment. Several sample selections are implemented to limit the incidence and variation in hourly wages because of outliers. Employee-year observations are excluded if the employee works over 80 hours, earns below A$200 a week or earns over A$10,000 a week. In addition, I exclude the top 0.1% and bottom 0.1% of observations of the hourly wage distribution and of the growth of hourly wages to limit error from outliers, removing 295 observations. Finally, 753 observations are excluded because of missing observations for the control variables. The sample of analysis contains 13,112 unique employed individuals and 95,760 observations.

3.2 Independent variables

The empirical analysis includes a rich set of independent variables. The independent variable of interest is population density, which is computed by dividing the annual estimated resident population by the area size in square kilometres of the spatial unit (Australian Bureau of Statistics, 2020). I use different measurements of geographic space in the analysis, which affects the borders of the spatial units and the values of population density and area size. For example, the SA2, SA3, SA4; and State and Territory (hereafter: state), which are based on the Australian Statistical Geography Standard structures of the ABS,

[3] Survey non-response is a main limitation of using household surveys for empirical analysis. The HILDA Survey initiated a sample top-up in 2011 to increase the representativeness of the Australian population.

are used in the empirical analysis. In addition, the empirical analysis is repeated separately for the administrative local government areas (LGAs)[4], and the sets of spatial units that are defined based on the hierarchical flow-based clustering algorithm. Spatial units coded as 'Other Territory', 'Migratory – Offshore – Shipping' or 'No usual address' were excluded from the sample of analysis. I also removed several islands and spatial units with an estimated resident population below 100 residents. This leaves a total of 2,164 SA2s and 9,711,627 commuting flows.

The individual's characteristics include zero-one indicator variables for gender (2 categories), Indigenous origin (2), being born abroad (2), age (9 categories in increments of five years), education (Year 11, Year 12, Certificates III and IV, Diploma; and Bachelor or higher educational attainment), number of household members (1, 2, 3-4; and 5 or more members), marital status (6 categories), number of own resident children (0, 1, 2; and 3 or more children), type of contract (permanent, fixed-term, casual, other; and missing contract), job occupation (8), job industry (20), the private sector (2) and year (19).

Table 1 provides summary statistics by population density quintile of individuals' residential location. The population density quintiles are defined annually at the SA2 level based on the data on estimated resident population and area sizes, which ensures the quintiles are not affected by the sample of analysis observed in the HILDA Survey. Table 1 shows that high-density areas are characterised by higher hourly wages and higher hourly wage growth. In addition, it is clear from Table 1 that the workforce in denser areas is more female, younger, more likely born abroad, more highly educated, and with fewer of Indigenous origin. From the number of observations by quintile, it can be observed that the majority of the sample (51%) lives in relatively dense area as part of the top two quintiles of population density.

4. Methodology

To study the urban wage premium, I assess the impact of population density on workers' hourly wage. A Mincer-style empirical model is specified:

$$y_{irt} = \beta_1 \ln dens_{rt} + \beta_2 \ln area_{rt} + \delta' X_{it} + \alpha_i + D_t + \varepsilon_{irt} \qquad (1)$$

[4] The analysis based on LGAs should be interpreted more carefully because of amalgamations of LGAs over time.

The subscripts i, r and t refer to individual, spatial unit and year, respectively. y is the natural logarithm of nominal hourly wage. The model contains the variables log population density and log area size. The individual's characteristics are represented by X, which include zero-one indicator variables for gender, Indigenous origin, being born abroad, age, education, number of household members, marital status, number of own resident children, type of contract, job occupation, job industry and the private sector. α_i represents the individual fixed effects, D_t represents the calendar year fixed effects and ε_{irt} represents the idiosyncratic error term.

Identification of the impact of density comes from regional-level changes in population over time and individual-level changes in the spatial unit, where individual-level changes in the spatial unit are caused by changes in individuals' residential location. As Combes and Gobillon (2015) emphasize, agglomeration benefits exist if the impact of density or area is significantly positive. I focus on estimating β_1, which represents the effect of local population density on hourly wages from increases in density or local population while controlling for local area size. A positive effect of area, holding density constant, represents the agglomeration impacts from increasing the area size and population proportionally. Note that using population size instead of population density would result in the same estimate of β_1 but a different estimate of β_2.

The coefficients in the empirical analysis are based on two sets of regressions, estimated as the effect of log population density on log hourly wage using the OLS estimator and the FE estimator, respectively. The individual fixed effects specification is used to limit the potential of time-constant unobserved heterogeneity. The time-constant variables gender, Indigenous origin and being born abroad, are absorbed by α_i for the FE estimator. Changes in hourly wages because of business cycle effects and inflation are absorbed by the year fixed effects. The standard errors and the 95% confidence intervals are computed using clustered standard errors by spatial unit, where the number of unique spatial units depends on the regional classification ranging between 2,041 (SA2 areas) and 8 (states) unique units.

Table 1
Sample means by population density quintile (proportions unless otherwise noted)

	SA2 population density year-specific quintiles				
	First quintile	Second quintile	Third quintile	Fourth quintile	Fifth quintile
Real hourly wage (A$, deflated by CPI)	27.81	29.06	30.30	32.16	34.02
Log real hourly wage (log of A$)	3.229	3.282	3.324	3.374	3.419
Real hourly wage growth (Average of individual year-to-year changes, %)	5.786	5.973	6.044	6.085	6.778
Density (population/km^2, based on SA2)	6.394	140.2	759.2	1,696	3,664
Log density (based on SA2)	1.147	4.715	6.559	7.421	8.114
Area size (km^2, based on SA2)	6,757	129.8	21.34	8.300	5.212
Log area size (based on SA2)	7.639	4.558	2.836	1.997	1.536
Female	0.318	0.345	0.373	0.370	0.404
Age					
$21 \leq$ age < 25	0.086	0.093	0.095	0.088	0.090
$25 \leq$ age < 30	0.116	0.129	0.139	0.145	0.182
$30 \leq$ age < 35	0.110	0.125	0.125	0.132	0.164
$35 \leq$ age < 40	0.122	0.126	0.125	0.125	0.122
$40 \leq$ age < 45	0.134	0.139	0.138	0.130	0.115
$45 \leq$ age < 50	0.152	0.137	0.133	0.135	0.114
$50 \leq$ age < 55	0.135	0.122	0.120	0.119	0.100
$55 \leq$ age < 60	0.096	0.091	0.086	0.087	0.076
$60 \leq$ age < 65	0.049	0.039	0.038	0.040	0.036
Indigenous origin	0.023	0.026	0.019	0.013	0.012
Born abroad	0.087	0.133	0.183	0.246	0.287
Education					
Year 11	0.235	0.201	0.191	0.152	0.095
Year 12	0.117	0.133	0.142	0.149	0.140
Cert III and IV	0.354	0.331	0.293	0.231	0.170
Diploma and adv. diploma	0.095	0.108	0.111	0.112	0.100
Bachelor, grad and postgrad	0.199	0.228	0.263	0.356	0.493
Partnered	0.779	0.762	0.731	0.722	0.673
Own resident children	0.477	0.488	0.470	0.477	0.382
Type of contract					
Permanent contract	0.736	0.763	0.789	0.793	0.769
Fixed contract	0.081	0.095	0.097	0.094	0.113
Casual contract	0.086	0.067	0.061	0.050	0.048
Other contract	0.002	0.001	0.001	0.002	0.002
Missing contract	0.096	0.074	0.052	0.061	0.069
Private sector	0.729	0.745	0.738	0.729	0.753
Number of observations	9,517	16,723	20,214	22,551	26,755

Notes: Sample means for individual characteristics are provided by quintile of population density. The quintiles are defined by year at the regional level based on the SA2 spatial structure. The sample of analysis includes 95,760 individual-year observations and 13,112 unique employed individuals. For the variable hourly wage growth, which is computed for individuals with at least two consecutive waves, the total number of observations equals 74,152. The time period under observation is 2001 to 2019.

4.1 A clustering algorithm to define a continuum of regional aggregations

A key identification challenge in research on regional differences is the modifiable areal unit problem, which involves the issue that empirical results and conclusions depend on the measurement of geographic space. In the context of this paper, Australia can be divided into distinct spatial units in many different ways (e.g., see Watts (2013)), which may affect the estimate of the effect of local population density on hourly wages. To assess this issue, the empirical analysis is repeated to examine the role of spatial unit sizes in the estimate of the urban wage premium. That is, I estimate the urban wage premium separately for several pre-defined administrative spatial structures as well as for a continuum of regional aggregations, using different spatial unit sizes and thereby different densities and area sizes holding all else constant in the regression of individual hourly wages on population density.

I define regional aggregations by applying a flow-based cluster algorithm, entitled flowbca, introduced by Meekes and Hassink (2018).[5] The continuum of regional aggregations contains sets of spatial units that range from 2,164 to 10 unique spatial units, which are aggregated one-by-one using a hierarchical clustering algorithm. The regional aggregations are computed based on a starting set of commuting flows from the ABS Census of Population and Housing 2016, through TableBuilder Pro 2016 (Australian Bureau of Statistics, 2016). The main input for the algorithm is relational data of 9,711,627 commuting flows from place of current residence to place of work observed at the SA2 level, as this information is not available at a finer spatial unit level. Alternative sets of regional clusters of economic activity are defined over a continuous level of regional aggregation to assess the role of regional aggregation in the estimation of the urban wage premium. After integrating the regional-level data with the individual HILDA Survey microdata, an additional 123 spatial units are excluded from the sample of analysis because of zero employed individuals in these units. In the empirical analysis, I repeat the analysis 2,155 times for all regional aggregations from 2,164 to 10 unique spatial units in increments of one defined by the algorithm.

The steps of the algorithm can be described as follows. From the starting set of commuting flows across SA2s, the algorithm selects the maximum of the single directed relative commuting flow from one source unit to a different destination unit. Relative flows are computed by taking the source unit's outgoing flow relative to the source unit's total of outgoing flows. The algorithm then aggregates the source unit to the

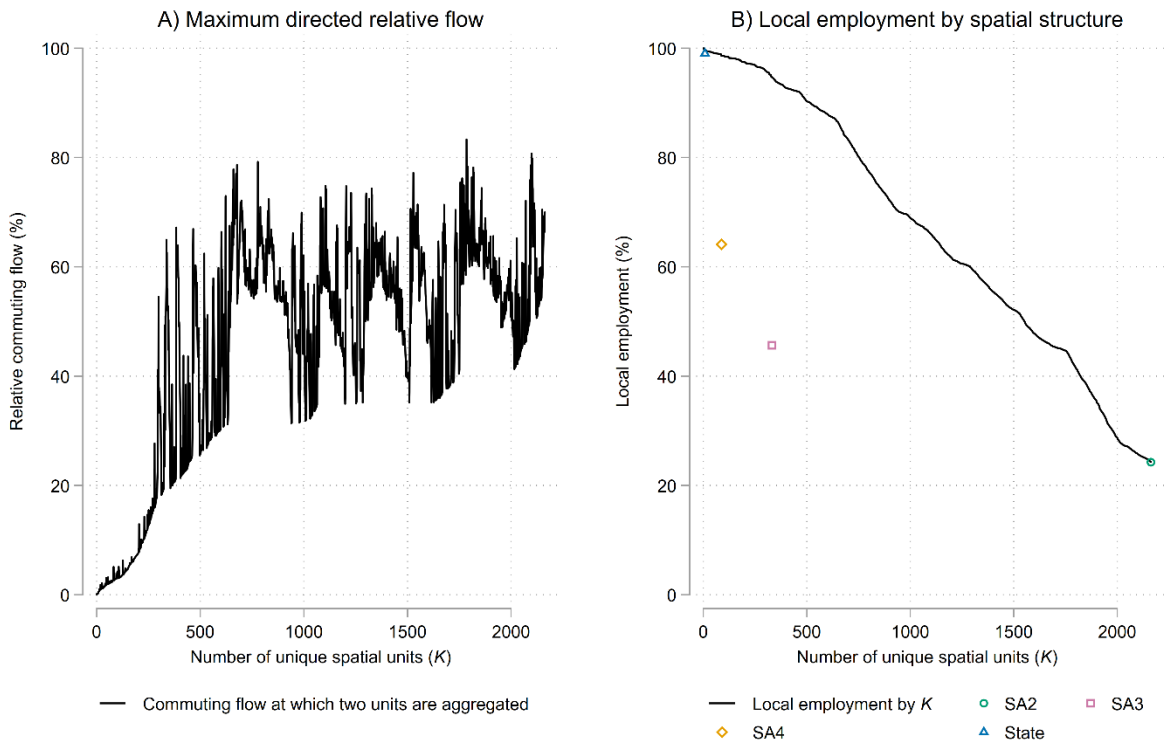
[5] See Coelli, Maccarrone, and Borland (2021) for an application of flowbca on SA3 spatial units to define local labour markets in Australia, studying the impact of increased Chinese imports on regional labour market outcomes.

destination unit, adding the absolute flows from the source unit and destination unit, where the core of the new spatial unit is defined as the initial destination unit's core. Following, the algorithm repeats these steps and starts again by selecting the maximum of the single directed relative commuting flow.

Figure 1A shows the maximum of the single relative commuting flow from one source unit to a destination unit at each iteration of the algorithm. As the algorithm starts with 2,164 unique spatial units, the relative commuting flow at which two units are aggregated has a decreasing pattern given the two spatial units with the strongest commuting connectivity are aggregated first. However, Figure 1A shows a nonmonotonically decreasing pattern. That is, although there is a downward trend in the maximum relative flow at which two units are aggregated, the maximum relative flow in the current iteration can be higher than that of the previous iteration. For example, after two spatial units are aggregated, another third spatial unit might have a higher relative commuting flow to the newly aggregated spatial unit that was computed by combining two units in the previous iteration.

Figure 1B shows the local employment rate by spatial structure. Local employment is defined as the number of individuals who live and work in the same spatial unit relative to the total number of employed individuals. The local employment rate is increasing if the number of unique spatial units becomes lower. For the ABS spatial structures SA3 and SA4, it is clear that local employment is relatively low, and much lower compared to the spatial units defined by using flowbca, conditional on the number of unique spatial units. This finding is important as it highlights that using SA3s, characterised by a local employment rate of 46 per cent, to define self-contained areas of residence and work activity is not very accurate.

Figure 1 – Defining regional areas using flowbca



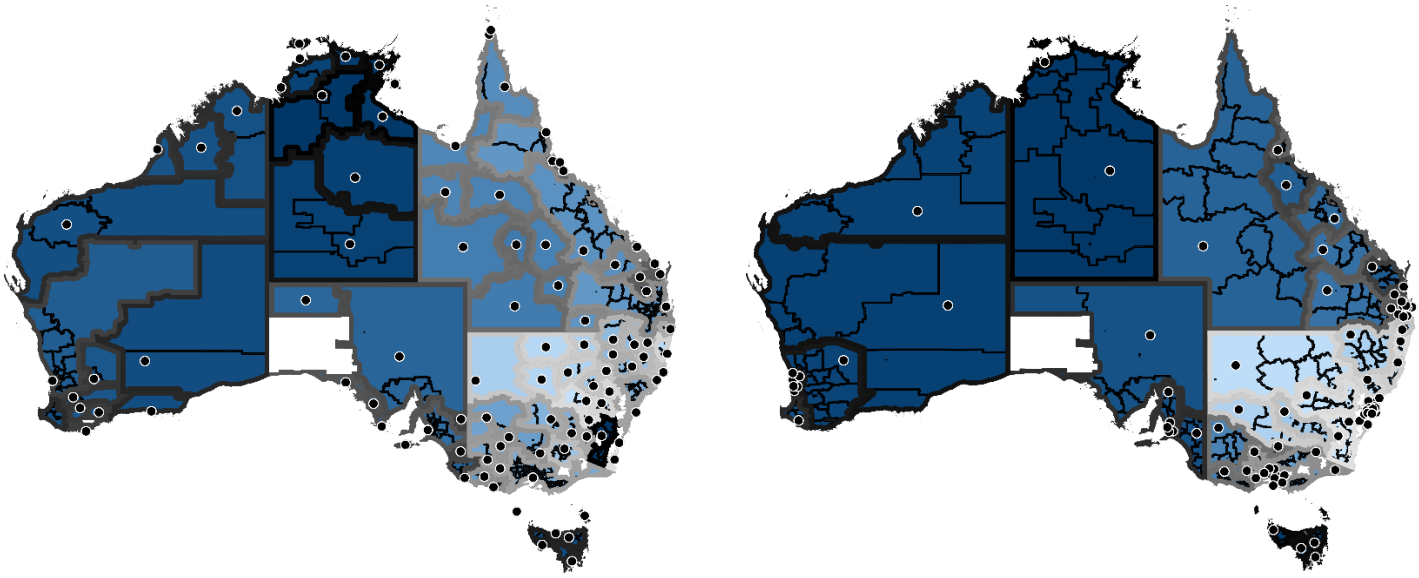
Notes: The number of unique spatial units is represented by *K*. The number of commuting flows used for defining spatial units equals 9,711,627, retrieved from the ABS Census of Population and Housing 2016. The process of clustering starts with 2,164 unique spatial units. In each iteration, one spatial unit is aggregated to another unit, aggregating the source unit to a destination unit characterized by the highest directed relative flow. The maximum directed relative flow is defined as the highest value observed of an outgoing flow relative to the unit's total of outgoing flows. Local employment is defined as the relative share of individuals living and working in the same spatial unit.

Figure 2 shows that urban areas are larger in the case of the 100 spatial units defined by flowbca than the 88 pre-defined ABS spatial structure SA4. By allowing for larger urban areas as can been observed in Figure 2, justified by the relatively strong commuting connectivity, local employment is higher for a given number of unique spatial units as is shown by Figure 1.

Figure 2 – Maps of Australia: A comparison between 100 defined areas and the SA4 units

A) 100 spatial units defined by flowbca

B) 88 SA4 spatial units



Notes: Figure 2A shows the map of 100 spatial units defined by flowbca. Figure 2B displays the SA4 spatial units as defined by the ABS. The core of a spatial unit is visible as a black dot with a white circle. Each unique spatial unit is surrounded by a thick border and highlighted by a different shade of blue. Several islands and spatial units with an estimated resident population below 100 residents were excluded. See Figure 1 for additional notes.

5. Empirical Analysis

This section documents the urban wage premium in Australia based on an empirical analysis of population density effects on hourly wages. The first part of the empirical analysis investigates the density effects based on various pre-defined regional classifications. Results are provided based on the OLS estimator and FE estimator, respectively. The second part of the analysis uses area fixed effects and the instrumental variable estimator to limit biases from local-level endogeneity. The third part of the analysis examines the role of the measurement of spatial units in estimates of the urban wage premium, estimating the population density effects on hourly wages for a continuum of regional aggregations ranging from 2,164 to 10 unique spatial units.

5.1 Density effects based on pre-defined regional classifications

Table 2 displays the impact of population density on hourly wages, estimated separately for five spatial structures: SA2, SA3, SA4, state and LGA. For the OLS estimator, columns (1) to (3) show how the estimate of the urban wage premium changes after including additional covariates.[6] Similar results are provided for the FE estimator in columns (4) to (6), based on regressions controlling for individual fixed effects to limit individual-level endogeneity.

Table 2 shows that for the SA2 spatial structure, the density effect on wages – the urban wage premium – equals about 1 per cent and is statistically significant at the 10 per cent significance level for the OLS estimator after including all covariates (see column (3)). For the FE estimator, the urban wage premium is also equal to about 1 per cent and is significant at the 5 per cent significance level irrespective of the set of covariates. This evidence suggests that wages increase by about 1 per cent if the population density doubles.

Based on SA3s, a weakly significant positive effect of population density on hourly wages is found for the FE specification only (columns (4) and (6)). In addition, for SA4s and states, no statistically significant effects are found. Conversely, using LGAs to measure geographic space, the impact of population density on hourly wages is statistically significant for all six specifications, ranging between 0.59 per cent and 2.09 per cent. Controlling for individual fixed effects slightly reduces the urban wage premium estimates found for LGAs, making these estimates comparable in magnitude as those obtained based on SA2s and SA3s.

Table 2 shows that the empirical evidence of positive density effects on hourly wages is mixed, as the size of the urban wage premium estimate as well as the statistical significance strongly depends on the spatial structure and the model specification. Compared to the literature (see, for example, D'Costa and Overman (2014) and Meekes and Hassink (2019)), it is surprising that the density effect on wages based on SA2s is higher after including individual fixed effects, as it controls for the endogenous sorting of more able workers to denser areas. By contrast, the estimates of the urban wage premium based on LGAs is indeed smaller for specifications that include individual fixed effects. In a meta-analysis on density effects,

[6] Tables A1 and A2 in Appendix A show the results for a sample of full-time and part-time employees together and a sample of part-time employees, respectively. Similar results are found for the sample of full-time and part-time employed workers together, whereas less significant results are found for the sample of part-time employees.

Ahlfeldt and Pietrostefani (2019) show that, internationally, the mean and median effect of density on wages is around 4 per cent. As such, the urban wage premium in Australia appears relatively low.

Table 2
Urban wage premium, based on pre-defined Australian spatial structures (Equation (1))

	(1)	(2)	(3)	(4)	(5)	(6)
		OLS			FE	
Log population density, based on:						
SA2 (2041 unique spatial units)	0.0065	0.0083	0.0096*	0.0114***	0.0089**	0.0095**
	(0.0088)	(0.0067)	(0.0053)	(0.0043)	(0.0041)	(0.0040)
SA3 (332 unique spatial units)	0.0054	0.0070	0.0068	0.0117**	0.0083	0.0095*
	(0.0139)	(0.0105)	(0.0080)	(0.0053)	(0.0052)	(0.0049)
SA4 (88 unique spatial units)	-0.0067	-0.0041	0.0012	0.0052	0.0039	0.0058
	(0.0201)	(0.0159)	(0.0116)	(0.0093)	(0.0083)	(0.0079)
State (8 unique spatial units)	0.0034	-0.0051	0.0011	0.0004	0.0019	0.0030
	(0.0279)	(0.0226)	(0.0158)	(0.0097)	(0.0096)	(0.0097)
LGA (426 unique spatial units)	0.0209**	0.0141**	0.0118***	0.0069*	0.0059*	0.0071**
	(0.0098)	(0.0063)	(0.0041)	(0.0036)	(0.0032)	(0.0030)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	Yes	No	Yes	Yes
Occupation & industry dummies	No	No	Yes	No	No	Yes
Individual fixed effects	No	No	No	Yes	Yes	Yes

Notes: The outcome variable is log hourly wages. Each row represents a different regression for a different regional classification. Each column represents a different set of control variables and fixed effects. Standard errors, clustered by spatial unit, are in parentheses. The number of unique spatial units for each regional classification in the sample of analysis is provided in parentheses after the relevant spatial structure. ***, **, *, correspond to the significance level of 1%, 5%, 10%, respectively. The set of covariates contains the individual's characteristics, which include zero-one indicator variables for gender (1 estimated parameter), Indigenous origin (1), being born abroad (1), age (8), education (4), number of household members (3), marital status (5), number of own resident children (3), type of contract (4) and the private sector (1). The occupation and industry dummies consist of job occupation (7) and job industry (19). The time-constant variables gender, Indigenous origin and being born abroad, are absorbed by α_i for the FE estimates. The R-squared is between 0.17 and 0.46 depending on the specification. The sample of analysis contains 95,760 individual-year observations and 13,112 unique employed individuals. The time period under observation is 2001 to 2019.

5.2 Density effects using area fixed effects and the instrumental variable estimator

One limitation of the benchmark results shown in the previous section is that these results could be affected by local-level endogeneity. That is, local amenities or local productivity may make a selective group of firms and workers more likely to move to a dense area, which in turn increases density, a mechanism also referred to as reverse causality. An approach to limit the potential of local-level endogeneity is to include local area fixed effects. The area fixed effects would capture time-constant unobserved heterogeneity at the area level.

However, as shown by the results in Table 3, the strategy of including area fixed effects to limit local-level endogeneity does not work well. The area fixed effects are identified based on moves across spatial units

(13,312 residential relocations across SA2s among 6,675 individuals in the sample), which introduces sample selection as it is not random who moves home. Columns (1) and (2) show that including area fixed effects can cause large biases, and the biases are larger the less unique spatial units are used to measure geographic space. This bias, because of limited mobility, is driven by the fact that residential moves do not happen very often in the sample of analysis, especially for spatial structures at high levels of regional aggregation such as SA4s (6,700 residential relocations among 4,051 individuals) and states (1,748 residential relocations among 1,240 individuals). In these instances, area fixed effects are identified based on fewer residential relocations, which may cause the upward bias in estimates of the urban wage premium.

In addition, for models including individual fixed effects and area fixed effects, estimating density effects is challenging because of the very little within-area variation in density over time, and relocations across spatial units are used to identify both the density effect and area fixed effect. Overall, the results in columns (1) and (2) of Table 3 are consistent with the notion that including area fixed effects when studying density effects is not a promising avenue.

An alternative strategy to deal with the local-level endogeneity is to apply the instrumental variable (IV) estimator (for example, see Ciccone and Hall (1996)). The instrumental variable estimator uses local historical population as instruments for local population density and local area size. The underlying assumptions are that historical population by spatial unit is correlated to current population (relevance condition) whereas it is uncorrelated to individuals' current hourly wages and productivity (exogeneity condition). In this paper, this approach is only applied using LGAs to measure geographic space (Australian Bureau of Statistics, 2019), as information on historical population is not available for the ABS spatial structures. Unfortunately, because of changes over time in the composition and existence of LGAs, the sample of analysis is reduced by over half.

The results show that instrumenting population density and area size of LGAs between 2001-2019 by long-lagged population density of LGAs in 1911, 1933 and 1954 increases the density effect on wages for the OLS estimator.[7] Specifically, the effect equals 2.44 per cent (column 3 of Table 3), much higher than 1.18

[7] For column (3) of Table 3, the underidentification test suggests the model is identified, as the null hypothesis is strongly rejected (Kleibergen-Paap rk LM statistic: 17.13; p-value < 0.01). In addition, the Hansen J statistic of 1.19 indicates the model is not overidentified, as the null hypothesis is not rejected (p-value > 0.27).

per cent (column 3 of Table 2).[8] However, using the instrumental variable estimator leads to a null effect after including individual fixed effects (column 4 of Table 3). The results show that including individual fixed effects matters more for density effects than applying the instrumental variable estimator. As such, the results suggest that individual-level endogeneity is more important than local-level endogeneity, corroborating the empirical literature on agglomeration economies (see Combes and Gobillon (2015)).

Table 3
Urban wage premium, strategies to limit local-level endogeneity

	(1)	(2)	(3)	(4)
	Area fixed effects		IV estimator	
Log population density, based on:				
SA2 (2041 unique spatial units)	-0.0255	0.0208	/	/
	(0.0182)	(0.0154)	/	/
SA3 (332 unique spatial units)	-0.0269	0.0425	/	/
	(0.0327)	(0.0419)	/	/
SA4 (88 unique spatial units)	-0.0164	0.1487**	/	/
	(0.0549)	(0.0693)	/	/
State (8 unique spatial units)	0.3801**	0.4452**	/	/
	(0.1216)	(0.1357)	/	/
LGA (426 unique spatial units)	-0.0395	0.0552	0.0244***	0.0002
	(0.0380)	(0.0434)	(0.0076)	(0.0099)
Year dummies	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes
Occupation & industry dummies	Yes	Yes	Yes	Yes
Individual fixed effects	No	Yes	no	Yes
Number of observations	95,760	95,760	40,242	39,604

Notes: The outcome variable is log hourly wages. Each row represents a different regression for a different regional classification. Columns (1) and (2) represent regressions including area fixed effects. Columns (3) and (4) represent regressions based on the instrumental variable estimator, which could only be applied for LGAs as information on historical population is not available for other spatial structures. The two instrumented variables are population density and area size. The three instruments are population density of 1911, population density of 1933 and population density of 1954.

5.3 The role of spatial unit sizes in estimates of density effects

Figure 3 displays the impact of population density on hourly wages for a continuum of regional aggregations, examining the role of the measurement of spatial units in estimates of the urban wage premium. That is, I analyse whether spatial unit sizes (scale effects) and spatial unit borders (zonation

[8] The results in columns (3) and (4) are robust to including a different set of instruments, for example including lagged population of 1976 and 1996, as well as using different combinations of years of lagged population density. Results are available upon request. Replicating the result of column (3) of Table 2 for LGAs, using the smaller sample of 40,242 observations that was used in column (3) of Table 3, produces an estimate of 0.0155 significant at the 1 per cent significance level. This result is also available upon request.

effects) matter for estimates of the impact of population density on wages. For each number of unique spatial units, K, the same regression model is estimated with the only difference being the number and thus the measurement of spatial units. The clustering algorithm that is used to aggregate spatial units is hierarchical, meaning that the building blocks for each set of spatial units are SA2s. Thus, all spatial structures are built from SA2 spatial units. To facilitate a comparison, figures 3A) and 3B) also contain the estimates of the urban wage premium based on the pre-defined spatial structures, as provided in columns 3 and 6 for the OLS estimator and FE estimator in Table 1, respectively.

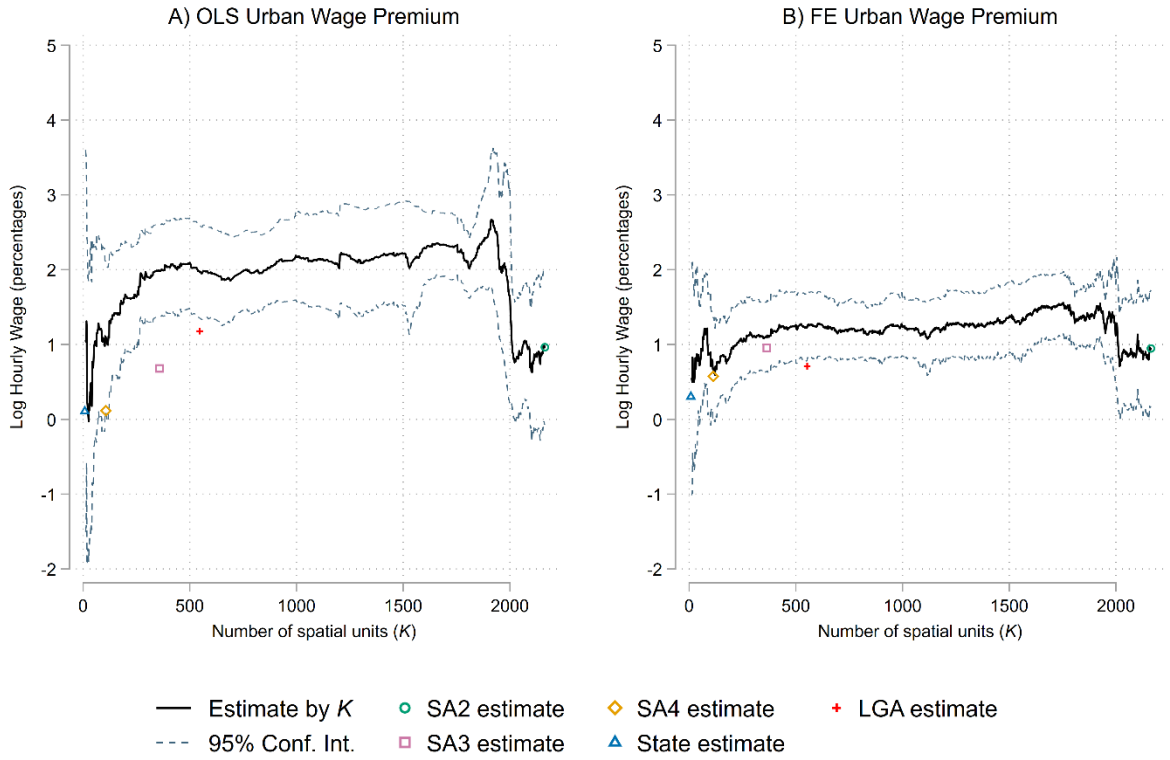
Figure 3 shows that the OLS estimate of the urban wage premium peaks at 2.7 per cent if using around 1,900 unique spatial units, whereas the FE estimate peaks at 1.6 per cent if using around 1,750 spatial units. Importantly, observe that the urban wage premium estimates are remarkably stable based on density effects estimated in the range of 1,800 to 300 unique spatial units. Compared to the international literature (Briant et al., 2010; Burger et al., 2010; Meekes and Hassink, 2019), where it is generally found that estimates of agglomeration externalities on wages are higher when using fewer and larger spatial units, the scale effects of the modifiable areal unit problem in the context of Australia appear relatively small.

However, Figure 3 also indicates that using very few spatial units or many spatial units, outside the range between 300 and 1,800 unique units, causes less robust results. This finding can be explained by the clustering algorithm that starts with aggregating spatial units with the highest relative commuting flow, which are disproportionally located in urban areas where there are higher levels of commuting connectivity between neighbouring spatial units (see figures 2A and 2B). At a high number of unique spatial units, aggregating units with a relatively high commuting connectivity increases the estimate of the urban wage premium. Conversely, at a low number of unique units, aggregating units with a low connectively reduces the urban wage premium.

Interestingly, the OLS estimates of the urban wage premium for the continuum of regional aggregations appear higher than those based on the ABS spatial structures (SA3 and SA4) and the administrative local government areas. This observation could be attributed to the defined regional aggregations allowing for larger urban areas, compared to the ABS spatial structures, holding the number unique spatial units constant. For example, agglomeration benefits of a central business district do not only occur in the central business district but also in neighbouring spatial units, which are part of the same spatial unit after clustering based on commuting flows as long as this commuting interaction is relatively high. Conversely,

for the ABS spatial structures, which contain more disaggregated and less self-contained urban areas (see Figure 2), there are more commuting flows across spatial units resulting in relatively high spatial autocorrelation, which may result in a measurement bias.

Figure 3 – The Urban Wage Premium in Australia



Notes: The coefficients are based on two sets of regressions, estimated as the effect of log population density on log hourly wage using the Ordinary Least Squares (OLS) estimator and Fixed Effects (FE) estimator, respectively. For each K, an estimate is provided based on a different regression. The 95% confidence intervals are computed using clustered standard errors by spatial unit. The number of unique spatial units is represented by K. The individual's characteristics are represented by X, which include zero-one indicator variables for gender (1), Indigenous origin (1), being born abroad (1), age (8), education (4), number of household members (3), marital status (5), number of own resident children (3), type of contract (4), job occupation (7), job industry (19) and the private sector (1). The time-constant variables gender, Indigenous origin and being born abroad, are absorbed by α_i for the FE estimates. Number of observations: 95,760. Number of individuals: 13,112. The time period under observation is 2001 to 2019.

Consistent with this notion, Stimson, Mitchell, Rohde, and Shyy (2011) and Stimson, Flanagan, Mitchell, Shyy, and Baum (2018) use relational data of commuting flows to define regional areas with a strong commuting connectivity within each area and argue that the issue of spatial autocorrelation is more limited than is the case for using pre-defined administrative spatial structures. Thus, spatial autocorrelation may be more limited for the defined spatial units than the ABS structures if the spatial

units defined by flowbca based on commuting flows are more self-contained. However, it is important to emphasize that differences in the estimate of the urban wage premium caused by zonation effects are statistically insignificant when considering the confidence intervals. In this regard, the zonation effects of the modifiable areal unit problem seem not particularly worrying in the Australian context.

6. Conclusion

As in many other countries, Australia has become more urbanized over time. With increased population in metropolitan areas, mainly because of immigration, more people are living and working in close proximity. A relevant question is how agglomeration economies, which refer to the benefits and costs of the spatial concentration of economic activity, affect local productivity.

This paper is the first to study the economic impact of urban density on individual wages in Australia. By combining HILDA Survey microdata on 13,112 employed individuals and regional-level population data, population density effects on individual hourly wages are studied over the period 2001 to 2019. A unique feature of this paper is to apply a flow-based clustering algorithm that uses commuting flows across SA2s to define spatial units that are characterised by a strong commuting connectivity within each spatial unit. As such, population density effects are analysed over a continuum of regional aggregations, providing more variation in how geographic space could be measured in empirical analyses, as is possible with the pre-defined ABS spatial structures.

Using various empirical strategies, the body of empirical evidence in this paper suggests that the estimate of the urban wage premium in Australia ranges between 0.5 and 2.7 per cent. Compared to international evidence that shows a mean and median effect of density on wages of 4 per cent (Ahlfeldt and Pietrostefani, 2019), the urban wage premium in Australia appears to be low. Clearly, much work remains to be done to develop a better understanding of how policy development options can enhance agglomeration economies in Australia, for example by increasing positive externalities based on the sharing, matching and learning mechanisms or by decreasing congestion and other frictions.

The purpose of this study was to show how agglomeration externalities can be studied in Australia using multiple data sources in a novel way. For other countries, many studies have analysed various research questions related to agglomeration externalities over the last two decades. A key topic has been the heterogeneity in agglomeration benefits among subgroups of the population, for example based on gender, educational attainment and occupation (Adamson, Clark, and Partridge, 2004; Di Addario and

Patacchini, 2008). Other research analyses to what extent agglomeration benefits attenuate with the distance from the core of an urban area (Rosenthal and Strange, 2008). In the context of the COVID-19 pandemic, a study on how a hybrid working environment may alleviate some of the negative effects of congestion on agglomeration economies could lead to important insights. These topics are well beyond the scope of this paper that documents density effects on individual wages in Australia for the first time, but do hint at promising avenues for future research on Australia.

References

Adamson, D. W., Clark, D. E., and Partridge, M. D. (2004). Do urban agglomeration effects and household amenities have a skill bias? *Journal of Regional Science*, *44*, 201–224.

Ahlfeldt, G. M., and Pietrostefani, E. (2019). The economic effects of density: A synthesis. *Journal of Urban Economics*, *111*, 93–107.

Australian Bureau of Statistics. (2016). Place of Work [2016 Census of Population and Housing TableBuilder]. Retrieved September 25, 2019, from https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/2901.0Chapter7702016

Australian Bureau of Statistics. (2019). Australian Historical Population Statistics, 2019 [Catalogue no 3105.0.65.001]. Retrieved April 12, 2021, from https://www.abs.gov.au/statistics/people/population/historical-population/latest-release

Australian Bureau of Statistics. (2020). Regional Population. Retrieved January 16, 2021, from https://www.abs.gov.au/statistics/people/population/regional-population/2018-19#data-download

Baum-Snow, N., and Pavan, R. (2012). Understanding the city size wage gap. *The Review of Economic Studies*, *79*, 88–127.

Briant, A., Combes, P.-P., and Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics*, *67*, 287–302.

Burger, M. J., Van Oort, F. G., and Van der Knaap, G. A. (2010). A treatise on the geographical scale of agglomeration externalities and the Modifiable Areal Unit Problem. *Scienze Regionali*, *9*, 19–40.

Ciccone, A., and Hall, R. E. (1996). Productivity and the density of economic activity. *The American Economic Review*, *86*, 54–70.

Coelli, M. B., Maccarrone, J., and Borland, J. (2021). *The dragon down under: The regional labour market impact of growth in Chinese imports to Australia* (Melbourne Institute Working Paper No. 9/21). Melbourne Institute: Applied Economic & Social Research.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, *63*, 723–742.

Combes, P.-P., and Gobillon, L. (2015). Chapter 5—The empirics of agglomeration economies. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics* (Vol. 5, pp. 247–348). Elsevier.

D'Costa, S., and Overman, H. G. (2014). The urban wage growth premium: Sorting or learning? *Regional Science and Urban Economics*, *48*, 168–179.

Di Addario, S., and Patacchini, E. (2008). Wages and the City. Evidence from Italy. *Labour Economics*, *15*, 1040–1061.

Duranton, G., and Puga, D. (2004). Chapter 48—Micro-foundations of urban agglomeration economies. In J. V. Henderson and J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2063–2117). Elsevier.

Duranton, G., and Puga, D. (2020). The economics of urban density. *Journal of Economic Perspectives*, *34*, 3–26.

Fotheringham, A. S., and Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, *23*, 1025–1044.

Glaeser, E. L., and Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*, *47*, 983–1028.

Glaeser, E. L., and Maré, D. C. (2001). Cities and skills. *Journal of Labor Economics*, *19*, 316–342.

Gould, E. D. (2007). Cities, workers, and wages: A structural analysis of the urban wage premium. *The Review of Economic Studies*, *74*, 477–506.

Graham, D. J., Melo, P. S., Jiwattanakulpaisarn, P., and Noland, R. B. (2010). Testing for causality between productivity and agglomeration economies. *Journal of Regional Science*, *50*, 935–951.

Gurran, N., Hulse, K., Dodson, J., Pill, M., Dowling, R., Reynolds, M., and Maalsen, S. (2021). *Urban productivity and affordable rental housing supply in Australian cities and regions* (AHURI Final Report No. No. 353). Australian Housing and Urban Research Institute.

Leishman, C., Bond-Smith, S., Liang, W., Long, J., Maclennan, D., and Rowley, S. (2021). *Relationships between metropolitan, satellite and regional city size, spatial context and economic productivity* (AHURI Final Report No. 357). Australian Housing and Urban Research Institute.

Maclennan, D., Ong, R., and Wood, G. (2015). *Making connections: Housing, productivity and economic development* (AHURI Final Report No. 251). Australian Housing and Urban Research Institute.

Meekes, J., and Hassink, W. H. J. (2018). flowbca: A flow-based cluster algorithm in Stata. *Stata Journal*, *18*, 564–584.

Meekes, J., and Hassink, W. H. J. (2019). *Endogenous local labour markets, regional aggregation and agglomeration economies* (IZA Discussion Paper No. 12765). Institute of Labor Economics.

Mion, G., and Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics/Revue Canadienne d'économique*, *42*, 28–55.

Moallemi, M., and Melser, D. (2020). The impact of immigration on housing prices in Australia. *Papers in Regional Science*, *99*, 773–786.

Openshaw, S., and Taylor, P. J. (1979). A million or so correlation coefficients, three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial science* (pp. 127–144). London: Pion.

Proost, S. V., and Thisse, J.-F. (2019). What can be learned from spatial economics? *Journal of Economic Literature*, *57*, 575–643.

Rosenthal, S. S., and Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, *64*, 373–389.

Stimson, R. J., Flanagan, M., Mitchell, W., Shyy, T.-K., and Baum, S. (2018). Modelling endogenous employment performance across Australia's functional economic regions over the decade 2001 to 2011. *Australasian Journal of Regional Studies*, *24*, 3–34.

Stimson, R. J., Mitchell, W., Rohde, D., and Shyy, P. (2011). Using functional economic regions to model endogenous regional performance in Australia: Implications for addressing the spatial autocorrelation problem. *Regional Science Policy & Practice*, *3*, 131–144.

The World Bank. (2021). Urban population (% of total population). Retrieved April 11, 2021, from https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?locations=AU

Watts, M. (2013). Assessing different spatial grouping algorithms: An application to the design of Australia's new statistical geography. *Spatial Economic Analysis*, *8*, 92–112.

Appendix A: Additional Robustness Checks

Table A1
Urban wage premium for sample of full-time and part-time employees (Equation (1))

	(1)	(2)	(3)	(4)	(5)	(6)
		OLS			FE	
Log population density, based on:						
SA2	0.0068	0.0094	0.0086*	0.0081**	0.0072*	0.0073**
	(0.0083)	(0.0060)	(0.0048)	(0.0039)	(0.0037)	(0.0036)
SA3	0.0070	0.0091	0.0076	0.0073	0.0049	0.0056
	(0.0125)	(0.0093)	(0.0071)	(0.0048)	(0.0047)	(0.0045)
SA4	-0.0064	-0.0027	0.0003	0.0070	0.0052	0.0051
	(0.0183)	(0.0132)	(0.0099)	(0.0081)	(0.0071)	(0.0066)
State	0.0090	0.0008	0.0045	0.0074	0.0082	0.0089
	(0.0260)	(0.0198)	(0.0148)	(0.0099)	(0.0097)	(0.0097)
LGA	0.0227**	0.0159***	0.0133***	0.0058**	0.0063**	0.0071***
	(0.0088)	(0.0051)	(0.0035)	(0.0029)	(0.0027)	(0.0025)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	Yes	No	Yes	Yes
Occupation & industry dummies	No	No	Yes	No	No	Yes
Individual fixed effects	No	No	No	Yes	Yes	Yes

Notes: The outcome variable is log hourly wages. Each row represents a different regression for a different regional classification. Each column represents a different set of control variables and fixed effects. Standard errors, clustered by spatial unit, are in parentheses. The number of unique spatial units for each regional classification in the sample of analysis is provided in parentheses after the relevant spatial structure. ***, **, *, correspond to the significance level of 1%, 5%, 10%, respectively. The set of covariates contains the individual's characteristics, which include zero-one indicator variables for gender (1 estimated parameter), Indigenous origin (1), being born abroad (1), age (8), education (4), number of household members (3), marital status (5), number of own resident children (3), type of contract (4), and the private sector (1). The occupation and industry dummies consist of job occupation (7), job industry (19) and full-time/part-time contract (1). The time-constant variables gender, Indigenous origin and being born abroad, are absorbed by α_i for the FE estimates. The sample of analysis contains 132,467 individual-year observations and 16,439 unique employed individuals. The time period under observation is 2001 to 2019.

Table A2
Urban wage premium for sample of part-time employees (Equation (1))

	(1)	(2)	(3)	(4)	(5)	(6)
		OLS			FE	
Log population density, based on:						
SA2	0.0076	0.0137*	0.0080	-0.0063	-0.0042	-0.0064
	(0.0104)	(0.0076)	(0.0066)	(0.0092)	(0.0091)	(0.0090)
SA3	0.0083	0.0065	0.0024	-0.0193	-0.0160	-0.0162
	(0.0112)	(0.0086)	(0.0079)	(0.0121)	(0.0120)	(0.0121)
SA4	-0.0142	-0.0127	-0.0118	0.0050	0.0039	-0.0012
	(0.0189)	(0.0130)	(0.0107)	(0.0158)	(0.0155)	(0.0152)
State	0.0221	0.0115	0.0092	0.0049	0.0082	0.0032
	(0.0168)	(0.0152)	(0.0138)	(0.0121)	(0.0133)	(0.0125)
LGA	0.0184***	0.0153***	0.0133***	0.0097	0.0138**	0.0122**
	(0.0058)	(0.0035)	(0.0029)	(0.0063)	(0.0063)	(0.0062)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	Yes	No	Yes	Yes
Occupation & industry dummies	No	No	Yes	No	No	Yes
Individual fixed effects	No	No	No	Yes	Yes	Yes

Notes: The outcome variable is log hourly wages. Each row represents a different regression for a different regional classification. Each column represents a different set of control variables and fixed effects. Standard errors, clustered by spatial unit, are in parentheses. The number of unique spatial units for each regional classification in the sample of analysis is provided in parentheses after the relevant spatial structure. ***, **, *, correspond to the significance level of 1%, 5%, 10%, respectively. The set of covariates contains the individual's characteristics, which include zero-one indicator variables for gender (1 estimated parameter), Indigenous origin (1), being born abroad (1), age (8), education (4), number of household members (3), marital status (5), number of own resident children (3), type of contract (4) and the private sector (1). The occupation and industry dummies consist of job occupation (7) and job industry (19). The time-constant variables gender, Indigenous origin and being born abroad, are absorbed by α_i for the FE estimates. The sample of analysis contains 32,647 individual-year observations and 8,492 unique employed individuals. The time period under observation is 2001 to 2019.