An alternative approach to estimating agglomeration and productivity using geography, demography and evidence from satellite imagery

Chris Leishman and Weidong Liang

ABSTRACT

In this article we combine novel geographical, demographic and satellite data with an instrumental variables (IVs) estimation approach to tackle a thorny set of data and econometric issues. Studies of agglomeration effects on economic productivity at the city scale have been growing in popularity in both the literature and the interests of urban policymakers in recent years. Despite growing sophistication in econometric approaches, underlying weaknesses in the quality and availability of data at the city scale have hindered efforts to unravel the endogeneity issues, or circularity between city scale, amenity and productivity. We set out both to extend the agglomeration literature to the case of Australia – which hitherto has been lightly studied in this field – and to take advantage of Australia’s unique geography and reliance on migration to drive population change. We also draw on satellite data measuring night-time luminosity as a proxy for economic productivity. We demonstrate the construction of convincing and robust IVs that proxy for population at the city scale. Our econometric estimates indicate that a 1% increase in city populations causes productivity to be enhanced by 0.24–1.70% on average.

ARTICLE HISTORY

Received 26 April 2021; Accepted 20 October 2021

KEYWORDS

climate; migration; agglomeration; productivity; cities; satellite imagery

JEL CLASSIFICATIONS

J30; O12; O18; O15; O25; R23; R32

INTRODUCTION

In recent years there has been a resurgence of interest in the concept of agglomeration economies, and in empirical testing strategies designed to uncover the strength and direction of such effects, together with the circumstances in which they may and may not hold. The advantages to economic productivity that arise due to city scale and/or density are argued to be potentially wide-ranging, including competitive advantages that arise when firms are able to access larger markets, internal advantages to the firm such as access to more highly trained labour supply and lowering costs through accessing established supply chains (Puga, 2010). Positive agglomeration
effects also occur over space, such that concentrations of activity affect outcomes in nearby urban
neighbourhoods, cities and/or regions (Combes & Gobillon, 2015). It has also been argued that
agglomeration effects are unlikely to be linear and may be subject to threshold effects. This gives
rise to the possibility that negative agglomeration effects arise at high concentrations of econ-
omic activity (Wheeler, 2003).

Questions about agglomeration effects generate interest further afield than the economics
literature. They are pertinent to urban and regional growth strategies because decisions about
where to encourage and concentrate economic activity impact on productivity, hence the overall
economic output of a nation. Agglomeration economies are arguably more important than ever
to the urban planning debate in the wake of the Covid-19 pandemic given the imperative to
rebuild economic output rapidly in order to repay the record sovereign debts incurred to pay
for economic stimulus measures. In some countries, as in Australia where economic growth
has been predicated on significant overseas migration for decades (McDonald & Temple,
2010), the debate is particularly important.

As we shall see in the next section, the empirical evidence on agglomeration economies that
exists is heavily based on the North American (particularly US) experience and, to a lesser extent,
European cities. Very little is known about the applicability, existence or scale of agglomeration
effects outwith these geographical contexts. The case of Australia is particularly interesting given
its enormous scale of 7.692 million km², modest population of 25 million in 2019 (Australian
Bureau of Statistics (ABS), 2016), and heavy concentration of economic activity in a small num-
ber of major capital cities (Kelly & Donegan, 2014). Distances between concentrations of econ-
omic activity in Australia tend to be very large, and it is not clear if or how this affects the
relationship between city scale and economic productivity. By extending the empirical analysis
to the case of Australia cities, we make a unique and novel contribution to the literature.

In this article we deliberately depart from the econometric-led approaches to untangling
endogeneity that have become dominant in the economics literature. Australia’s geographical
scale has an analytically attractive property because numerous climactic zones are present in a
single country. Climate conditions are an important determinant of individuals’ decisions to
migrate both to and within a country (Jordan, 2007). People particularly prefer to reside in cities
with more liveable climates such as mild, warm and cool climates (Albouy & Lue, 2015; Cragg
& Kahn, 1997). Climate conditions are pre-determined and exogenous (Dell et al., 2014).
Additionally, Australia’s population expansion after 1995 is due to overseas migrant intakes
and the number of overseas migrants who arrived in Australia is positively in
fluenced by migrant
quotas and the number of visas issued (see the third section below for more details). As only the
federal government of Australia has authority to set migrant quotas or issue visas, changes in
quotas and number of visas issued can create exogenous variations on Australian city popu-
lations. Therefore, we combine city climates with quotas or visa issuance to construct new
instrumental variables (IVs) for city population to overcome the endogeneity and reverse caus-
ality estimation issues. To the best of our knowledge, this has not been attempted elsewhere in
the literature.

A further empirical innovation is our use of information on satellite-recorded night lumin-
osity data to reveal the concentration of human activity in urban areas, and to act as a proxy for
urban productivity. As productivity is positively related to (real) gross domestic product (GDP)
and income levels that are shown to be well approximated by night-time luminosity (Pinkovskiy
& Martin, 2016), there is a logical expectation that a more economically productive city will be
more lit. It has also been shown that cities with larger populations of skilled people are more
brightly illuminated at night (Dingel et al., 2021). Many previous empirical studies (see
Table A1 in Appendix A in the supplemental data online) have employed micro- (or survey)
and macro-data to measure productivity. However, such approaches suffer from errors induced
through measurement and/or self-reporting (Kjellsson et al., 2014). To address this issue, our
study follows Henderson et al. (2012) in the use of impersonally collected night-time luminosity data. Unlike micro- (survey) or macro-data as used in many previous studies, satellites collect data impersonally and this helps to mitigate the problem of measurement errors in productivity data.

**ADDING GEOGRAPHICAL APPROACHES TO THE STUDY OF AGGLOMERATION ECONOMIES**

The theoretical basis for expecting interactions between urban scale (or density) and economic productivity are well rehearsed elsewhere in the literature. Briefly, Marshall (1920) set out three broad categories of agglomeration economies: labour pooling, shared input markets and technological spillovers. Although these principal categories have since been refined, with additional arguments added, they form the basis for much of the empirical work that has been published in this vein. More recently, Glaeser’s (2008) general equilibrium model has been widely used and replicated in modelling the impact of urban economic development policy. Elsewhere, Glaeser and Gottlieb (2009) have argued that attracting workers to an area will boost local productivity through human capital spillovers, but they also note that low-income workers are attracted to urban areas with greater employment opportunities. Indeed, they make numerous cautionary points about the adoption of agglomeration economics in policy, including the fact that such effects are likely to be non-linear, difficult to quantify and subject to threshold effects. In addition, it is very difficult to determine whether gains through agglomeration benefits arising from shifting economic activity to one location more than offset losses experienced elsewhere.

Several recent studies have examined the relationship between employment density and labour productivity (e.g., Brulhart & Mathys, 2008; Ciccone & Hall, 1996; Morikawa, 2011), while others have examined the impact on other proxies for productivity, such as patent intensity (Carlino et al., 2007). Wheeler (2001) focused on the first of Marshall’s categories (labour pooling) and found that the effects of larger markets on productivity are greater for higher skilled workers. Rosenthal and Strange (2003, 2008) examine information spillovers and find that the benefits of spatial concentration fall quickly with distance, but that labour market pooling and shared input benefits dissipate more slowly. Baldwin et al. (2010), based on Canadian data, also found that knowledge spillovers are bounded at a distance of 10 km from a plant. Significantly, this study also found that spillover effects are non-uniform between industrial sectors, adding to the complexity.

A summary of past empirical studies is shown in Table A1 in Appendix A in the supplemental data online. In our survey of the literature, we found a heavy reliance on survey-based and/or macro-dataset to measure (or proxy) productivity. However, we note that such data are subject to measurement errors (e.g., Holland & King, 1979; Johnson et al., 2013; Kjellsson et al., 2014). A second limitation is that many of the studies surveyed focused either on a subset of economic sectors or on a single economic sector. While this simplifies the analysis, it nevertheless introduces the risk that important interactions between economic sectors are missed, giving rise to misleading results.

In order to address the endogeneity issue of estimating the causal effect of city populations on productivity, some studies have followed Ciccone and Hall (1996) and Combes et al. (2010) to use historical population levels and soil fertility respectively as IVs for city populations (e.g., Combes et al., 2011; Mion & Naticchioni, 2009; see also Echeverri-Carroll & Ayala, 2011; and Combes, 2011). As these instruments are time invariant, only cross-sectional regression approaches may be used. Consequently, time-invariant city characteristics such as locations and land sizes that are correlated with population and productivity are omitted in the estimations. The result is that the estimates of the effect of city population on productivity may remain biased when such instruments are chosen.
In this study we investigate the link between city population size and productivity in 293 Australian ‘cities’ (defined as local government areas (LGAs) with populations over 10,000) between 2001 and 2013. It should be noted that the Australian definition of ‘city’ is more akin to the concept of an urban neighbourhood. By focusing on LGAs with 10,000 or greater population, our definition is similar to that of significant urban areas (SUAs) in Australia.\(^2\)

Aside from the contributions to the literature outlined above, we make two major empirical innovations to address the data and estimation issues that affect previous studies reliant on survey or macro-data. First, we deploy high-resolution satellite imagery night-time luminosity data from the US Air Force Defense Meteorological Satellite Program (DMSP) to proxy city productivity. As outlined in the introduction, GDP can be approximated by night-time luminosity (Henderson et al., 2012; Pinkovskiy & Martin, 2016). Additionally, Mellander et al. (2015) show that night-time luminosity is a good proxy for economic activities. With over 25 billion grid cells recorded in the DMSP’s global system since 1992, the satellite-recorded light data represent a finely spatially grained, time-varying, dataset that can proxy for the intensity of economic activities from all sectors. This choice of data has the additional advantage that we do not need to restrict the analysis to a subset of economic sectors that are arguably a good proxy for total productivity.

Our second innovation is the development of new IVs for city populations with temporal–spatial exogenous variations that are suitable for the panel data analysis. We combine the following arguments and evidence to construct our instruments. First, we acknowledge that overseas migrant intake – rather than new births – are the most significant component of population change to population expansion in Australia since 1995. Second, we note that annual new migrant intakes are controlled by the federal government of Australia through setting permanent resident quotas and issuing visas annually.\(^3\) Third, we take account of the fact that climactic conditions help to explain differences in population levels between cities. Specifically, cities with more liveable climates (as defined in the third section below) tend to be more populous.

The remainder of the paper is structured as follows. In the next section we examine the role of immigration and climate in annual change to Australian cities’ populations. The fourth section presents the data sources used in the study, and the fifth section introduces our estimation approach. Subsequent sections examine the results, robustness checks and suggest future directions for studies of city productivity that combine geographical and econometric approaches.

OVERSEAS MIGRATION, GEOGRAPHY AND POPULATION AGGLOMERATION IN AUSTRALIA

Australia’s population grew significantly between 1980 and 2019, from 15 million to 25 million (ABS, 2016). The average rate of growth was over 1.5% during this period, which is the second highest among the Organisation for Economic Co-operation and Development (OECD) countries.\(^4\)

There are two contributing factors to population growth in Australia. Figure 1 demonstrates that new births added around 60,000 people each quarter to Australia between 1982 and 2006, and the addition modestly increased to 78,000 in 2017. About 50,000 overseas migrants arrived in Australia each quarter in 1982, and that quarterly arrivals increased significantly to 138,000 in 2017.\(^5\) Figure 2 shows that 54% of population expansion came from new births, and 46% was from new overseas migrant intakes on average between 1982 and 1995.\(^6\) This is different in 2017, when 65% of the new population came from new overseas migrants, whereas 35% was from births. It is therefore clear that overseas migration rather than new births is the most important contributing factor to Australia’s population growth after 1995.

The federal government of Australia promoted the immigration-led population growth in Australia after 1995. This is because that the federal government sets permanent resident quotas
and issue visas annually to determine the number of overseas migrants arriving in Australia each year to fulfil the country’s economic and labour market needs (Productivity Commission, 2016). For example, the federal government increased quotas and issued more visas to address a perceived labour supply shortage after 1995 (Spinks, 2010). These changes are shown in Figures 3 and 4. Subsequently, Australia’s population began to grow more significantly after the federal government decided to increase permanent resident quotas and to issue more visas annually. The positive association between Australia’s population and quotas or visas issued is reflected in the upward trends in \( \log(\text{population}) \), \( \log(\text{quotas}) \) and \( \log(\text{visa issuance}) \) in Figures 3 and 4. The correlation between Australia’s population and quotas or visas issued is over 0.82.
The positive relationship between overseas migrant intakes and population can also be shown at the city level. For instance, we selected four Australian capital cities (i.e., Sydney, Melbourne, Brisbane and Adelaide) and plotted the natural logarithm of their population levels against the log of permanent resident quotas and visa issuance from 2001 to 2016 in Figures 5 and 6. We can observe the upward trends in these cities, implying that there is a positive link between city-level population and the permanent resident quotas or visa issuance. The correlation indices are over 0.8, which suggests that the positive link between selected city population and quotas or visas issued is strong.
City populations and overseas migrants in Australia are distributed unevenly in different climate conditions. Cities with more liveable climate conditions tend to be more highly populated (Albouy & Stuart, 2014). In Australia, the Bureau of Meteorology (BoM) developed seven climate zones.8 The zones we include are: Zone 1: Hot humid summer and warm winter climate; Zone 2: Warm humid summer and mild winter climate; Zone 3: Hot dry summer and warm winter climate; Zone 4: Hot dry summer and cool winter climate; Zone 5: Warm temperate climate; Zone 6: Mild temperate climate; and Zone 7: Cool temperate climate. Table 1 shows that the cities with a warm, mild or cool climate are much more populous than those with a hot or hot, dry summer climate.

Figure 5. Australian city populations and the permanent resident quota.

Figure 6. Australian city populations and visas issued.
In this section we set out our methodology for estimating the influence of city population on economic productivity using our novel approach. We begin with ordinary least squares (OLS) regression, then go on to discuss the estimation issues in this approach. To establish a causal relationship between population and productivity, and that the relationship is robust, we propose our IVs for Australian city populations and employ them in two-stage least squares (2SLS) and reduced-form regressions.

**OLS regression**

Equation (1) shows the OLS regression that relates the natural log of light intensity for city $i$, state $s$ at year $t$ as the proxy of annual city productivity level.

$$\log \log \left( \text{light} \right)_{i,s,t} = c + \beta_i \log \log \left( \text{population} \right)_{i,s,t} - k + \delta \text{greatercapital}_i + \gamma \log \log \left( \text{bachelorp} \right)_{i,s,t} + \mu_i + \mu_s + \mu_t \times v_{i,s,t}$$

To control for the probable impact of educational attainment on productivity, we include the log number of people with a bachelor’s degree or above ($\text{bachelorp}$). The variable $\text{greatercapital}$ is a dummy variable where city $i$ is equal to 1 (or 0) if it is in the (non-)greater capital area. Including this variable aims to control the pattern in which population and economic activities are geographically clustering at the greater capital areas. $\mu$ subscripts $i$ and $s$ are the city and year fixed effects, respectively, that absorb all time invariant city-specific characteristics and all time-varying (observed and unobserved) macroeconomic shocks that affect productivity respectively. $\mu$ subscript $s$ is the state fixed effect, and we interact it with the year fixed effect (i.e., $\mu_i \times \mu_t$) to control for any factors that vary across states and years (e.g., annual gross state product (GSP) or state-level unemployment). Lastly, to control the potential spatial clustering, we choose our idiosyncratic error term ($v_{i,s,t}$) to be clustered at the city level.

**Estimation issues in OLS regression**

Even though we include control variables to reflect educational attainment, greater capital city and fixed effects, OLS regression cannot identify $\beta_i$ – the elasticity of productivity with respect to city population – as a result of the following issues. First, other determinants of productivity that are correlated with population are contained in $v_{i,s,t}$. These factors could be city housing prices and amenities (Albouy, 2016). Hence, the OLS regression is subject to omitted variable bias. To further examine the problem of endogeneity in the OLS regression we conduct Durbin–Wu–Hausman tests and present the results in Tables A2 and A3 in Appendix B in the supplemental data online. The results suggest that city population is endogenous to the error term.
in equation (1). Second, people prefer to work in cities with higher wage rates. This means that the OLS regression is susceptible to self-selection bias or reverse causality issues.

**Two-stage least squares (2SLS) regression**

To overcome the estimation issues in OLS regression, we employ an IV strategy for the 2SLS regression. Specifically, we interact two exogenous variables that are correlated with city populations. The first variable is climate, which varies across cities, and can be considered to be both predetermined and exogenous (Dell et al., 2014; Roos, 2005). Jordan (2007) and Albouy and Lue (2015) argued that areas with a mild, warm or cool climate are more favourable for habitation than those with hot or hot, dry summers (see also Cragg & Kahn, 1997; Albouy et al., 2013). As we have shown, Australian cities with one of these ‘favourable’ climates for habitation are more populous than those with a hot or hot, dry summer climate (see Table 1 for details). Thus, we construct a dummy variable (\( \text{favourableclimate}_i \)) where the city \( i \) at state \( s \) would be equal to 1 if it has a mild, warm or cool climate (or the city is in Zone 2, 6, 5 or 7), otherwise it is set to 0 for cities that have a hot or hot, dry summer climate (i.e., the city is in Zone 1, 4 or 3).

The second variable is the time-varying permanent resident quotas and number of visas issued. The rationale is that overseas migrant intakes rather than new births are the major contributor to population expansion after 1995 in Australia. The annual number of overseas migrants to arrive in Australia is controlled by the Australian federal government by setting permanent resident quotas and issuing visas annually. Therefore, we can employ annual quotas and visas issued to approximate yearly city population well (see the third section for more details). More importantly, as Australian cities do not have the authority to set quotas or to issue visas, the (log of) lagged quotas and number of visas issued (i.e., log(\( quota_{t-k} \)) and log(\( visasissued_{t-k} \) \( k=1 \) or 2) are exogenous to a city’s productivity (as proxied in this article by its night-time luminosity). Our IVs combine climate and quotas or visa issuance, that is:

\[
\begin{align*}
\text{favourableclimate}_i \times \log (\text{quota}_{t-k}) \\
\text{favourableclimate}_i \times \log (\text{visasissued}_{t-k})
\end{align*}
\]

in the first-stage regression as equation (2) can effectively predict the yearly–city population; and \( \rho_i \) and \( \tau_i \) are the elasticities of city population with respect to climate and quotas/visas issued.

\[
\log \log (\text{population}_{i,t}) = c + \{ \rho \text{favourableclimate}_i \times \log (\text{quota}_{t-k}) \} + \{ \tau \text{favourableclimate}_i \times \log (\text{visasissued}_{t-k}) \} + \varphi \text{greatercapital}_i + \hat{\text{loglog (bachelorp)}_{i,t}} + \mu_i + \mu_t + \mu_x \times \psi_{i,t}
\]

(2)

Equation (3) is the second-stage regression in which the predicted yearly–city population (i.e., the fitted values for log(\( \text{population}_{i,t-k} \)) as obtained in the first-stage regression) can act as exogenous variations on productivity. To summarize the effect of city population instrumented by climate and quotas/visas issued on productivity:

\[
\log \log (\text{lights}_{i,t}) = c + \alpha \log \log (\text{population}_{i,t-k}) + \lambda \text{greatercapital}_i + (\text{bachelorp})_{i,t} + \mu_i + \mu_t + \mu_x \times \mu_x + \epsilon_{i,t}
\]

(3)
Reduced-form regression

Equation (4) is the reduced-form regression. This model aims to directly estimate the effect of our proposed IVs (i.e., climate and quotas/visa issuance) on city productivity (as proxied by night-time luminosity).

\[
\begin{align*}
\text{loglog} (\text{lights}_{i,t}) &= \epsilon + \{ \pi \text{favourable climate}_i \times \text{loglog} (\text{quota}_{i,t}) \} \eta \text{favourable climate}_i \\
&\times \text{loglog (visas issued}_{i,t} \} \varphi \text{greater capital}_i + \chi \text{loglog (bachelorp)}_{i,t} + \mu_i + \mu_t + \mu_{i,t} + \psi_{i,t} \\
\end{align*}
\]

(4)

DATA

Our dataset is a panel of Australia cities defined as LGAs with an annual population over 10,000 between 2001 and 2013, and we can understand them as SUAs in Australia. This is a slightly different definition to that used by the ABS, which is any Statistical Area Level 2 (SA2) with a population over 10,000. The data for city population and the number of people with a bachelor’s degree or more are from the ABS (2001, 2006 and 2011 censuses, with extrapolation for intervening years). The city boundaries are defined by the Australian Statistical Geography Standard (ASGS) (2016). The data for annual permanent resident quotas and visas issued are from the Department of Home Affairs, and the visas issued include four types of applicants: permanent residents, working holiday, students and long-term business holders. The summary statistics are shown in Table 2.

Climate zone data

To account for the variation in city climate in Australia, we use climate zone data developed by the BoM that categorizes all Australian cities into one of seven climate zones on the basis of a city’s historical climate conditions (i.e., humidity, temperature and precipitation levels between 1961 and 1990). As noted above, we consider seven instead of eight zones here. The BoM determines that cities belong to Zone 2 (warm, humid summer and mild winter climate) if their average annual temperature, humidity and precipitation levels are 18–24°C, 50–70% and 1000–2000 mm individually between 1961 and 1990. If cities’ average annual temperature, humidity and precipitation levels in this period are 30–36°C, 30–40% and 300–400 mm individually, they belong to Zone 3 (hot, dry summer and warm winter). To save space, we do not illustrate here the full details for humidity, temperature and precipitation levels for the rest of these zones. However, they can be found at BoM.

Satellite recorded city light data

The raw data for night-time luminosity have been recorded by satellites from the DMSP using Operational Linescan System sensors. These satellites detect luminosity to the space at every location on Earth between 20:30 and 22:00 hours local time. Researchers at the National Oceanic and Atmospheric Administration’s (NOAA) and National Geophysical Data Centre (NGDC) then process these raw data (i.e., the removal of observations for locations with natural lights,

| Table 2. Descriptive statistics. |
|---------------------------------|
| Variable          | (1) Observations | (2) Mean | (3) SD | (4) Minimum | (5) Maximum |
| log(population)   | 3516             | 11.09    | 1.01   | 7.76        | 12.80       |
| log(quota)        | 3516             | 10.63    | 0.31   | 12.43       | 13.62       |
| log(visas issued) | 3516             | 11.32    | 0.24   | 10.25       | 12.15       |
| log(bachelorp)    | 3516             | 8.07     | 1.07   | 1.62        | 9.63        |
| log(lights)       | 3516             | 2.21     | 1.34   | 0.68        | 6.33        |
forest fires and obscurity due to cloud cover over the Earth’s surface) and produce yearly light intensity valued between 0 (i.e., the dimmest) and 63 (i.e., the brightest) for each 30 arc-second grid cell (or 1 × 1 km at the equator) of the Earth. The light data are available from 1992 to 2013.

To construct the yearly light intensity for Australian cities, we consider the following steps. First, we need to address the top-coding (or truncation) issue in the light data. This is because cities from developed countries such as Australia are likely to have a light intensity over 63 (Stor- eygard, 2016). To overcome this issue, we use the data that are free from the top-coding issue provided by Bluhm and Krause (2018). Second, we sum the yearly light intensity from all the grid cells within a city and compute its average to represent that city’s yearly light intensity level. For example, if there are 10 grid cells in the city of Sydney and the light intensity is 1 for each cell in 2010, the sum of night-time light intensity for Sydney in 2010 is 10 and the light intensity (or the average from all cells) in 2010 would be 1. We follow this method to compute the annual night-time luminosity for the rest of Australia’s cities.

Night-time luminosity in Australia rises significantly over time and varies dramatically between cities. To illustrate, we can compare the night-time light intensity in Australia in 1992 (Figure 7a) with that in 2013 (Figure 7b). In addition, we can see that the increase in night light during this period mainly comes from the coast and nearby inland areas. We can also see that luminosity is quite different between the inland (or Outback) and coastal cities within a given year. To illustrate, compare Figure 8a with Figure 8b to see that Alice Springs (an inland city) is much dimmer than Newcastle (a coastal city) in 2013.

RESULTS

In this paper we note that many previous studies (23 of 27 summarized in Table A1 in Appendix A in the supplemental data online) have used survey/micro- or macro-data to measure economic productivity in cities. In our above discussion we noted that there are two principal problems with this approach. First, such data are affected by measurement and self-reporting errors (e.g., Bollinger & Hirsch, 2013; Bound et al., 2001; Holland & King, 1979; Johnson et al., 2013; Kjellsson et al., 2014). Second, the time series for variables such as wages, value added...
and income that are used to measure productivity tend to be very short. Taking Australian cities as an example, wage and income data are only available from 2013 to 2017 from the ABS. We build on recent studies that have employed satellite-recorded night-time luminosity data to circumvent these issues (e.g., Chen & Nordhaus, 2011; Henderson et al., 2012; Storeygard, 2016). They argued that luminosity data are collected impersonally, which avoids measurement and self-reporting error. Most importantly, Dingel et al. (2021) argue that the brightness of cities at night can effectively approximate their productivity (including accounting for educational attainment and skill levels).

To illustrate the utility of luminosity data as a proxy for productivity in Australia, we regress the log of value added, GSP and wage on the log of night-time luminosity at the state and city level.
levels in different time periods and present their results in Table 3, columns (1–5). The results suggest that brighter states (or cities) at night are more productive in terms of having greater value added and GSP and higher wages.

Having established that night-time luminosity is a good proxy for productivity, the remainder of this section now steps through the three estimation stages, beginning with the results obtained from the OLS regression.

**OLS estimates**

Columns (I) and (II) of Table 4 show that the OLS estimates of city population on productivity, approximated by night-time luminosity, are between 0.48 and 0.56 when \( \log(population_{i,t-1}) \) and \( \log(population_{i,t-2}) \) are considered separately in equation (1) without any control variables (i.e., \( \text{greatercapital}_i \) and \( \log(bachelor_{i,t}) \)). After adding the control variables, the elasticity changes slightly to between 0.37 and 0.46, as shown in columns (III) and (IV) of Table 4. This could be interpreted to mean that adding the control variables in the OLS regression did not affect the estimated effects of population on productivity.

If city scale (higher levels of population) increases economic productivity through agglomeration effects, as many empirical studies claim, our OLS estimates of \( \log(population_{i,t-1}) \) and \( \log(population_{i,t-2}) \) in the different model specifications summarized in Table 4 should be positive. Our mixed results here could be due to the estimation issues of endogeneity and reverse causality, as discussed above.

**2SLS estimates**

As we noted in the last section, the OLS results cannot be assumed robust because the estimation approach does not address the endogeneity or reverse causality that almost certainly exists between city scale, urban amenity and productivity. In our review of the literature, it is clear that migration both to and within a country is driven by a range of factors including individuals’ pursuit of higher wages and greater liveability. It has also been shown that higher levels of amenity and more liveable climate conditions are associated with higher levels of population in cities. Without controlling for these influences on individuals’ migration and relocation decisions, it is not possible to robustly estimate the relationship between city scale (population) and economic productivity.
For this reason, we introduce two IVs to represent the endogenous variable (annually observed city population). These IVs enter the 2SLS regression. The first IV combines climate conditions with permanent resident quotas, and their 2SLS estimates are presented in Table 5. The first-stage estimates at the lower panel of Table 5 suggest that city population is positively related to the favourable climate conditions for residence and permanent resident quotas. Specifically, column (I) suggests that if the federal government increases the quotas by 1% at year $t - 1$, the population of cities with a favourable climate for residence (i.e., the cities with a mild, warm or cool climate) would be enhanced by 0.16% more than the cities without it (i.e., the cities with a hot or hot, dry summer climate). The estimated impact becomes 0.12% after the control variables are included in column (III). If the 1% increase occurs at year $t - 2$, our first-stage estimation without control variables suggests that the population would be expanded by 0.14% more on average for the cities with a favourable climate than the one without it in column (II). Without the control variable, the estimated impact would be 0.10% in column (IV). Our IVs are plausibly strong. This is because (1) all the first-stage estimates from columns (I–IV) in Table 5 are statistically significant at the 1%

Table 3. Relationship between lights and productivity in Australia.

| Variable          | (1) log(value added) | (2) log (GSP) | (3) log (wage) | (4) log (wage) | (5) log (wage) |
|-------------------|----------------------|---------------|----------------|----------------|----------------|
| log(lights)       | 0.21** (0.09)        | 0.49** (0.23) | 0.28** (0.10)  | 0.10** (0.005) | 0.07*** (0.0007) |
| State level       | Yes                  | Yes           | Yes            | No             | No             |
| City level        | No                   | No            | No             | Yes            | Yes            |
| State fixed effect| Yes                  | Yes           | Yes            | Yes            | Yes            |
| Year fixed effect | Yes                  | Yes           | Yes            | n.a.           | n.a.           |
| Sample year       | 1992–2013            | 1994–2013     | 1994–2013      | 2013           | 2013           |
| Observations      | 176                  | 160           | 160            | 525            | 293            |

Note: Robust standard errors clustered at the city level are shown in parentheses. *Significant at 10%; **significant at 5%; and ***significant at 1%. Value added refers to gross value added by all industries at current prices. Gross state product (GSP) is measured by chain volume. Wages are weekly means for adults working full time. Column (4) is for all Australian cities; and column (5) is for cities with populations of 10,000 or greater.

Table 4. Ordinary least squares (OLS) regression of night-time luminosity on city population.

| Variable | I | II | III | IV |
|----------|---|----|-----|----|
| Dependent: log(lights) | log(population<sub>n,i,t-1</sub>) 0.48** (0.09) | log(population<sub>n,i,t-2</sub>) 0.56*** (0.10) | greatercapital<sub>n</sub> 4.04*** (0.08) | bachelorp<sub>n,i</sub> 0.09** (0.04) |
|          | Adjusted $R^2$ 0.52 | 0.56 | 0.53 | 0.56 |
| City fixed effect | Yes | Yes | Yes | Yes |
| Year | Yes | Yes | Yes | Yes |
| Year × State | Yes | Yes | Yes | Yes |
| Observations | 3516 | 3223 | 3495 | 3206 |

Note: Robust standard errors clustered at the city level are shown in parentheses. *Significant at 10%; **significant at 5%; and ***significant at 1%.
(2) all the Chi-squared p-values for Kleibergen–Paap LM statistics are 0.00, suggesting that our IVs are highly correlated with city population; and (3) all the Kleibergen–Paap Wald F-statistics are greater than the Stock and Yogo critical value, which allows us to reject the null hypothesis of weak IVs.

The second-stage estimates in Table 5 show that the elasticity of night-time luminosity with respect to city population instrumented by climate conditions and permanent resident quotas is between 1.15 and 1.73 through columns (I–IV). This suggests that city population agglomeration increases productivity. In addition, the estimates suggest that this effect is stronger for cities in greater capital areas, and for cities with a higher concentration of human capital (measured by people with a bachelor’s degree or above).

Table 6 sets out the 2SLS estimation results when city population is instrumented by the combined effects of favourable climate conditions for residence and visa issuance. The lower panel presents the first-stage estimates where the combined effects on population are 0.13–0.19 in five different model specifications (columns I–V). The statistically significant levels of all the IVs estimates, and their Kleibergen–Paap LM and Wald F-statistics indicate that our IVs are ‘valid’ and ‘strong’ in the first-stage regression, following the same approach applied and explained for Table 5. The second-stage estimates in Table 6 show that the estimated effects of city population on night-time luminosity are 1.10–1.79 in five model specifications, which are similar to the results in Table 5.

### Reduced-form estimates

Finally, to estimate the direct impact of our IVs on productivity (as proxied by night-time light intensity), we estimate the reduced-form regression shown as equation (4) and document their estimates in Tables 7 and 8. Our reduced-form estimates indicate that cities with a
favourable climate would be more economically productive if the federal government were to increase migrant intakes by increasing migrant quotas or issuing more visas. Table 7 shows that the estimated effects of favourable climate condition and quotas on productivity are between 0.12 and 0.23 based on the different model specifications. In the case of favourable climate conditions and visa issuance, Table 8 shows that their combined effects on productivity are between 0.13 and 0.30.
Conclusions

In recent years there has been a resurgence of interest in the study of agglomeration economic effects centred on the economics literature, but with offshoots of interest in the regional science, urban geography, applied economics and housing economics literature. The study of agglomeration is important from a number of policy perspectives, and that importance has risen steeply as governments grapple with questions about how to rebuild economies and rebalance urban and regional economies in the aftermath of the Covid-19 pandemic.

This study examined the relationship between population and productivity in Australian cities between 2001 and 2013. We tackled several empirical issues and set out to make three new contributions to the literature. First, we have extended the agglomeration economics analytical framework to Australia. We would argue that this is a welcome development given the heavy concentration of previous studies on the North American context. Australia has a unique set of geographical circumstances which lead us to question whether the concept of agglomeration economies is even applicable in the Australian context. The geographical extent, sparsity, small population and large distances between cities make it very different contextually to the ambit of most of the rehearsed literature.

Second, we have adopted an entirely different empirical strategy that the econometrics-led approaches that have come to dominate the economics literature in the study of agglomeration effects. To do this, we exploited Australia’s unique and diverse geography, proxied by climate zone, to account for the endogeneity between city scale, population change and economic productivity. We also capitalized on Australia’s unique reliance on overseas migration as the driver of annual population expansion. By combining information from geographical and demographic processes in this way, we were able to construct convincing and robust IVs that proxy for yearly-city populations. This breaks the endogeneity and allows us to arrive at robust estimates of agglomeration effects for Australian cities.

The analysis demonstrates that the relationship between city scale and productivity is volatile and unreliable when estimated using OLS. Our first-stage estimation results show that the instruments are valid and strong in predicting yearly-city populations. Our second-stage estimates indicate that an 1% increase in Australian city populations causes Australian cities’ productivity to increase by 0.97–1.78% on average. The positive link between city population and productivity is reinforced in the reduced-form estimates. Our 2SLS and reduced-form estimation results remain robust after conducting a series of robustness checks. Thus, we would argue that the analysis in this article has shown that a combination of geographical and

### Table 8. Reduced-form regression of log(lights) on climate and visas issued.

|                      | I     | II    | III   | IV    |
|----------------------|-------|-------|-------|-------|
| Dependent: log(lights) |       |       |       |       |
| favourable climate  | 0.16*** (0.02) | 0.13*** (0.03) |       |       |
| favourable climate  | 0.20*** (0.03) | 0.17*** (0.03) |       |       |
| greater capital | 4.02*** (0.11) | 4.00*** (0.11) |       |       |
| log(visas issued)   | 0.10*** (0.02) | 0.16*** (0.04) |       |       |
| Adjusted R²          | 0.50  | 0.54  | 0.52  | 0.55  |
| City fixed effect    | Yes   | Yes   | Yes   | Yes   |
| Year                 | Yes   | Yes   | Yes   | Yes   |
| Year × State         | Yes   | Yes   | Yes   | Yes   |
| Observations         | 3516  | 3223  | 3495  | 3206  |

Note: Robust standard errors clustered in the city level are shown in parentheses. *Significant at 10%; **significant at 5%; and ***significant at 1%.
demographic data, set within an IVs approach, is effective in tackling the econometric estimation issues that arise from short time series and productivity data paucity that have plagued previous studies.

Finally, our study used high-resolution satellite imagery – specifically night-time luminosity – to proxy for city-scale economic productivity. This approach avoids measurement errors such as frame deficiency and response/non-response bias that are available in the survey data as used in previous studies. We have shown that areas with higher levels of night-time light are more economically productive in Australia. In other words, night-time luminosity is an effective proxy for city-scale economic productivity, and it accounts for change over time, spatial differences at the city level and differences arising from regional effects (particularly coastal versus inland differences).

NOTES

1 Measurement errors include frame deficiency and response and non-response bias. For more examples, see Bound et al. (2001) and Bollinger and Hirsch (2013).
2 The definition of SUA here is slightly different to that defined by the ABS as any Statistical Area Level 2 (SA2) with a population over 10,000.
3 These visas are issued for permanent and temporary overseas migrants. For more information, see the data section of this article.
4 See https://www.oecd-ilibrary.org/economics/oecd-factbook-2013/population-growth-rates_factbook-2013-table4-en/.
5 Overseas migrants in this study refer to the sum of temporary and permanent residents of Australia. Temporary residents include international students and business long-stay residents; but they exclude tourists.
6 To compute the percentage of new births (or overseas migration) in Australian population expansion each year, we divide the annual new births (or new overseas migrants) by the annual total new populations (i.e., the sum of new births and overseas migrants.) For example, if new births are 10 and new overseas migrants are 20 in 2002, the total new population is 30. The percentage of new births in new population is about 33% in 2002.
7 Supply shortages are mainly in the fields of medical, information technology (IT), engineering and construction trade workers. The federal government temporarily slashed the quotas and visas issued to respond to the economic slowdown in the Global Financial Crisis in 2008–09. For more details, see Productivity Commission (2016).
8 These climate zone data are developed by BoM to assist the Australian Building Codes Board (ABCB) to regulate the building and construction industry (see https://www.abcb.gov.au/Resources/Tools-Calculators/Climate-Zone-Map-Australia-Wide). The BoM developed eight zones for the ABCB. One of these (Alpine) is not included in this study as there are very few areas with this climate in Australia. As all the cities with this climate have a predominantly cool temperate climate, we reassign them to the cool temperate climate category.
9 For more information about these data, see https://www.homeaffairs.gov.au/research-and-statistics/statistics/visa-statistics/live/migration-program/.
10 The BoM developed these data to assist the ABCB to regulate the building and construction industry; for these data, see https://www.abcb.gov.au/Resources/Tools-Calculators/Climate-Zone-Map-Australia-Wide/.
11 See http://www.bom.gov.au/jsp/ncc/climate_averages/climate-classifications/index.jsp?maptype=tmp_zones#maps/.
12 For these data, see https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html/.
13 Bluhm and Krause (2018) provided evidence to show that the distribution of lights data from the DMSP is a Pareto distribution, and had corrected the top-coding issue based on
this evidence. For the DMSP’s light data free from the top-coding issue, see https://www.lightinequality.com/.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

The research reported in this article was supported by funding from The Australian Housing and Urban Research Institute (AHURI), project number 31222.

ORCID

Chris Leishman  http://orcid.org/0000-0002-7853-5035
Weidong Liang  http://orcid.org/0000-0001-8291-0697

REFERENCES

Albouy, D. (2016). What are cities worth? Land rents, local productivity, and the total value of amenities. *Review of Economics and Statistics*, 98(3), 477–487. https://doi.org/10.1162/REST_a_00550

Albouy, D., Leibovici, F., & Warman, C. (2013). Quality of life, firm productivity, and the value of amenities across Canadian cities. *Canadian Journal of Economics/Revue Canadienne d’Economique*, 46(2), 379–411. https://doi.org/10.10111/caje.12017

Albouy, D., & Lue, B. (2015). Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life. *Journal of Urban Economics*, 89, 74–92. https://doi.org/10.1016/j.jue.2015.03.003

Albouy, D., & Stuart, B. (2014). Urban population and amenities: The neoclassical model of location (Technical Report National Bureau of Economic Research, Working Paper 19919). https://www.nber.org/system/files/working_papers/w19919/revisions/w19919.rev1.pdf

Australian Bureau of Statistics (ABS). (2016). *Australian historical population statistics*. https://www.abs.gov.au/

Baldwin, J. R., Brown, W. M., & Rigby, D. L. (2010). Agglomeration economies: Microdata panel estimates from Canadian manufacturing. *Journal of Regional Science*, 50(5), 915–934. https://doi.org/10.1111/j.1467-9787.2010.00675.x

Bluhm, R., & Krause, M. (2018). Top lights-bright cities and their contribution to economic development.

Bollinger, C. R., & Hirsch, B. T. (2013). Is earnings nonresponse ignorable? *Review of Economics and Statistics*, 95(2), 407–416. https://doi.org/10.1162/REST_a_00264

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J.J. Heckman and E. Leamer (Eds.), *Handbook of econometrics* (Vol. 5, pp. 3705–3843). Elsevier.

Brulhart, M., & Mathys, N. A. (2008). Sectoral agglomeration economies in a panel of European regions. *Regional Science and Urban Economics*, 38(4), 348–362. https://doi.org/10.1016/j.regsciurbeco.2008.03.003

Carlino, G. A., Chatterjee, S., & Hunt, R. M. (2007). Urban density and the rate of invention. *Journal of Urban Economics*, 61(3), 389–419. https://doi.org/10.1016/j.jue.2006.08.003

Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21), 8589–8594. https://doi.org/10.1073/pnas.1017031108

Ciccone, A., & Hall, R. E. (1996). Productivity and the density of economic activity. *The American Economic Review*, 86(1), 54–70. http://www.jstor.org/stable/2118255

Combes, P.-P. (2011). The empirics of economic geography: How to draw policy implications? *Review of World Economics*, 147(3), 567–592. https://doi.org/10.1007/s10290-011-0092-z

Combes, P.-P., Duranton, G., & Gobillon, L. (2011). The identification of agglomeration economies. *Journal of Economic Geography*, 11(2), 253–266. https://doi.org/10.1093/jeg/lbq038
Combes, P. P., Duranton, G., Gobillon, L., & Roux, S. (2010). Estimating agglomeration effects with history, geology, and worker fixed effects. The Economics of Agglomeration, National Bureau of Economic Research.

Combes, P.-P., & Gobillon, L. (2015). The empirics of agglomeration economies. In G. Duranton, J. Vernon Henderson, & W. C. Strange (Eds.), Handbook of regional and urban economics (Vol. 5, pp. 247–348). Elsevier.

Cragg, M., & Kahn, M. (1997). New estimates of climate demand: Evidence from location choice. Journal of Urban Economics, 42(2), 261–284. https://doi.org/10.1006/juec.1996.2027

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate–economy literature. Journal of Economic Literature, 52(3), 740–798. https://doi.org/10.1257/jel.52.3.740

Dingel, J. I., Miscio, A., & Davis, D. R. (2021). Cities, lights, and skills in developing economies. Journal of Urban Economics, 125, 103174. https://doi.org/10.1016/j.jue.2019.05.005

Echeverri-Carroll, E. L., & Ayala, S. G. (2011). Urban wages: Does city size matter? Urban Studies, 48(2), 253–271. https://doi.org/10.1177/0042098010369393

Glaeser, E. L. (2008). Cities, agglomeration, and spatial equilibrium. Oxford University Press.

Glaeser, E. L., & Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. Journal of Economic Literature, 47(4), 983–1028. https://doi.org/10.1257/jel.47.4.983

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. American Economic Review, 102(2), 994–1028. https://doi.org/10.1257/aer.102.2.994

Holland, P., & King, B. (1979). Measurement and interpretation of productivity. National Academies Press chapter Measurement Error in Productivity Statistics.

Johnson, S., Larson, W., Papageorgiou, C., & Subramanian, A. (2013). Is newer better? Penn World Table revisions and their impact on growth estimates. Journal of Monetary Economics, 60(2), 255–274. https://doi.org/10.1016/j.jmoneco.2012.10.022

Jordan, R. (2007). Moving to nice weather. Regional Science and Urban Economics, 37(3), 375–398. https://doi.org/10.1016/j.regsciurbeco.2006.11.004

Kelly, J.-F., & Donegan, P. (2014). Mapping Australia’s economy: Cities as engines of prosperity. Grattan Institute. https://grattan.edu.au/wp-content/uploads/2014/07/814-mapping-australia-economy1.pdf

Kjellsson, G., Clarke, P., & Gerdtham, U.-G. (2014). Forgetting to remember or remembering to forget: A study of the recall period length in health care survey questions. Journal of Health Economics, 35, 34–46. https://doi.org/10.1016/j.jhealeco.2014.01.007

Marshall, A., (1920). Industry and trade.

McDonald, P. F., & Temple, J. (2010). Immigration, labour supply and Per capita Gross domestic product; Australia 2010–2050. Department of Immigration and Citizenship.

Mellander, C., Lobo, J., Stolarick, K., & Matheson, Z. (2015). Night-time light data: A good proxy measure for economic activity? Plos one, 10(10), e0139779. https://doi.org/10.1371/journal.pone.0139779

Mion, G., & Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. Canadian Journal of Economics/Revue Canadienne d’Economique, 42(1), 28–55. https://doi.org/10.1111/j.1540-5982.2008.01498.x

Morikawa, M. (2011). Economies of density and productivity in service industries: An analysis of personal service industries based on establishment–level data. The Review of Economics and Statistics, 93(1), 179–192. https://doi.org/10.1162/REST_a_00065

Pinkovskiy, M., & Martin, X. S.-i. (2016). Lights, camera … income! illuminating the national accounts–household surveys debate. The Quarterly Journal of Economics, 131(2), 579–631. https://doi.org/10.1093/qje/qjw003

Productivity Commission. (2016). Migrant intake into Australia. Inquiry Report, 77, 313–332. https://www.pc.gov.au/inquiries/completed/migrant-intake/report/migrant-intake-report.pdf

Puga, D. (2010). The magnitude and causes of agglomeration economies. Journal of Regional Science, 50(1), 203–219. https://doi.org/10.1111/j.1467-9787.2009.00657.x

Roos, M. W. (2005). How important is geography for agglomeration? Journal of Economic Geography, 5(5), 605–620. https://doi.org/10.1093/jeg/lbi008
Rosenthal, S. S., & Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics, 85*(2), 377–393. https://doi.org/10.1162/003465303765299882

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics, 64*(2), 373–389. https://doi.org/10.1016/j.jue.2008.02.006

Spinks, H. (2010). Australia’s migration program. *Parliamentary Library.* https://parlinfo.aph.gov.au/parlInfo/download/library/prspub/272398/upload_binary/272398.pdf;fileType=application%2Fpdf#search=%

Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth in Sub-Saharan Africa. *The Review of Economic Studies, 83*(3), 1263–1295. https://doi.org/10.1093/restud/rdw020

Wheeler, C. H. (2001). Search, sorting, and urban agglomeration. *Journal of Labor Economics, 19*(4), 879–899. https://doi.org/10.1086/322823

Wheeler, C. H. (2003). Evidence on agglomeration economies, diseconomies, and growth. *Journal of Applied Econometrics, 18*(1), 79–104. https://doi.org/10.1002/jae.678