



# Towards a theory of sustainable city sizes



Liton Md Kamruzzaman

This study proposes a new theory of sustainable city size: the condition where a city's actual population aligns with its theoretical capacity to function effectively. Drawing from four foundational theories (locational fundamentals, increasing returns, central place, and central flow), the study develops a conceptual framework to estimate theoretical city size using data from 655 Urban Centres and Localities (UCLs) in Australia. It examines how deviations of actual size from this theoretical size influence sustainability outcomes, using rent, walking-to-work rates, and multi-vehicle household share as proxies for economic, environmental, and socio-environmental sustainability. Results show that UCLs within  $\pm 4\%$  of their theoretical size achieve optimal outcomes across all indicators. Nationally, achieving equilibrium could save \$5.3 billion in annual rent, generate 44,000 additional daily walking trips, and reduce multi-vehicle dependency in 275,000 households. These findings support the proposed theory and offer a practical tool to align cities with their systemic capacity.

Hundreds of new cities are built globally each decade<sup>1–4</sup>, yet few have population size targets grounded in systematic theory. Most rely instead on political aspirations or speculative investment, raising critical planning questions: (a) How can city sizes be determined before they are built or exceed their limits? and (b) Would systematically predetermined sizes enhance urban sustainability? Determining city size is foundational to urban planning<sup>5</sup>, shaping land use, infrastructure, and service provision<sup>6</sup>. Without a priori estimates, urban planning becomes reactive and fragmented. Yet, the literature offers little guidance on how to determine the size of new cities. Existing approaches focus on current size distributions (e.g., rank-size rule) or project growth based on past trends<sup>7</sup>, offering little strategic guidance or sensitivity to planning interventions. As a result, many cities face unsustainable growth or population decline<sup>2</sup>.

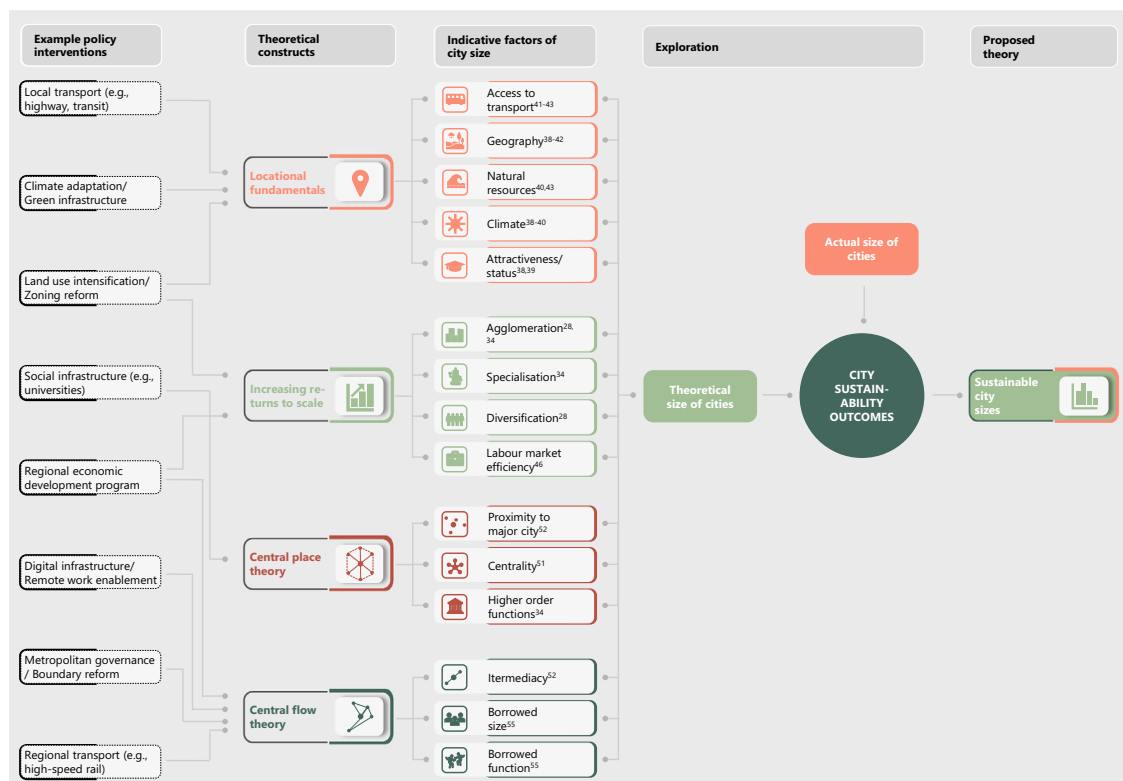
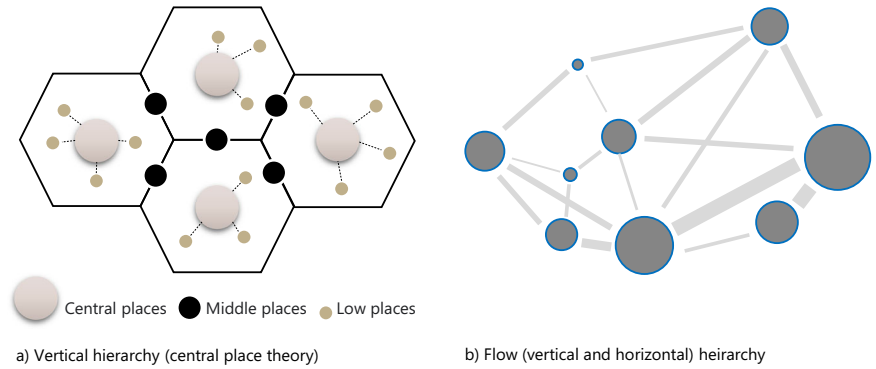
The case for a priori city size estimates is further complicated by conflicting evidence on the relationship between city size and sustainability. While larger cities are often seen as growth-inducing<sup>8</sup>, others find that productivity is independent of size<sup>9</sup>, or highly context-dependent<sup>10,11</sup>. Large cities also face challenges such as crime<sup>12</sup>, emissions<sup>13</sup>, heat islands<sup>14</sup>, unaffordable housing<sup>15</sup>, social segregation<sup>16</sup>, and declining civic engagement<sup>17</sup>—yet remain attractive for their amenities<sup>18</sup> and innovation<sup>19</sup>. Some studies suggest medium-sized cities perform better on life satisfaction<sup>20</sup>, and environmental sustainability<sup>21</sup>. Although commuting distances typically increase with city size<sup>22</sup>, the effect is often small<sup>23,24</sup>, or even absent<sup>25,26</sup>. This study proposes that such inconsistencies stem from a reliance on *actual* population size, without considering a city's *theoretical* or systemic capacity. This reflects the absence of a robust framework for estimating *theoretical city size*—that is, the population a city can sustainably support given its unique opportunities and constraints. Importantly, theoretical city size is not fixed but can be altered through planning interventions. The study further

hypothesises that cities perform better on sustainability outcomes when their actual sizes approximate their theoretical sizes, a condition referred to here as *sustainable city size*.

To support this hypothesis, it is essential to revisit the theoretical foundations that explain how city sizes form and evolve. A large body of literature offers insights into the drivers of urban growth, structure, and interaction, though often in fragmented strands<sup>27,28</sup>. Early approaches aimed to identify a universal optimum<sup>29</sup>, but this notion was dismissed as overly simplistic. Zipf's law, a widely cited empirical regularity, proposes that city size follows a rank-size distribution<sup>30,31</sup>; yet it lacks predictive capacity<sup>28,32</sup>, especially for forward-looking planning. Cities differ in their specificity (e.g., topographic or climatic conditions) and unicity (e.g., capital city status)<sup>33</sup>, resulting in varying theoretical sizes<sup>34</sup> that can change over time through interventions such as infrastructure investment, zoning reforms, or institutional change.

Three dominant theories explain how individual cities grow: random growth, locational fundamentals, and increasing returns to scale<sup>35,36</sup>. Random growth theory posits that cities grow randomly, eventually conforming to Zipf's law<sup>37</sup>. Locational fundamentals theory suggests that geographic, physical, and climatic attributes (e.g., elevation, port access, and rainfall) are randomly distributed and shape a city's potential<sup>38–43</sup>. Although traditionally seen as fixed, recent studies suggest that these fundamentals can shift in response to technological or social change<sup>44</sup>. Increasing returns to scale theory highlights how economic agglomeration drives growth through specialisation, diversity, and productivity<sup>8,45</sup>. These effects are dynamic and policy-sensitive: labour productivity, connectivity, and spatial clustering can be shaped by zoning reform, regional investment, or new transport infrastructure<sup>44,46,47</sup>. However, these theories focus on individual cities and overlook their interactions within broader urban

**Fig. 1 | Schematic representation of the hierarchical distribution of cities.** This figure compares conceptual frameworks for understanding city systems. **a** shows a vertical hierarchy based on central place theory, with central, middle, and low-order places. **b** illustrates a combined vertical and horizontal flow hierarchy, reflecting the interaction of centrality with functional linkages in the urban network.



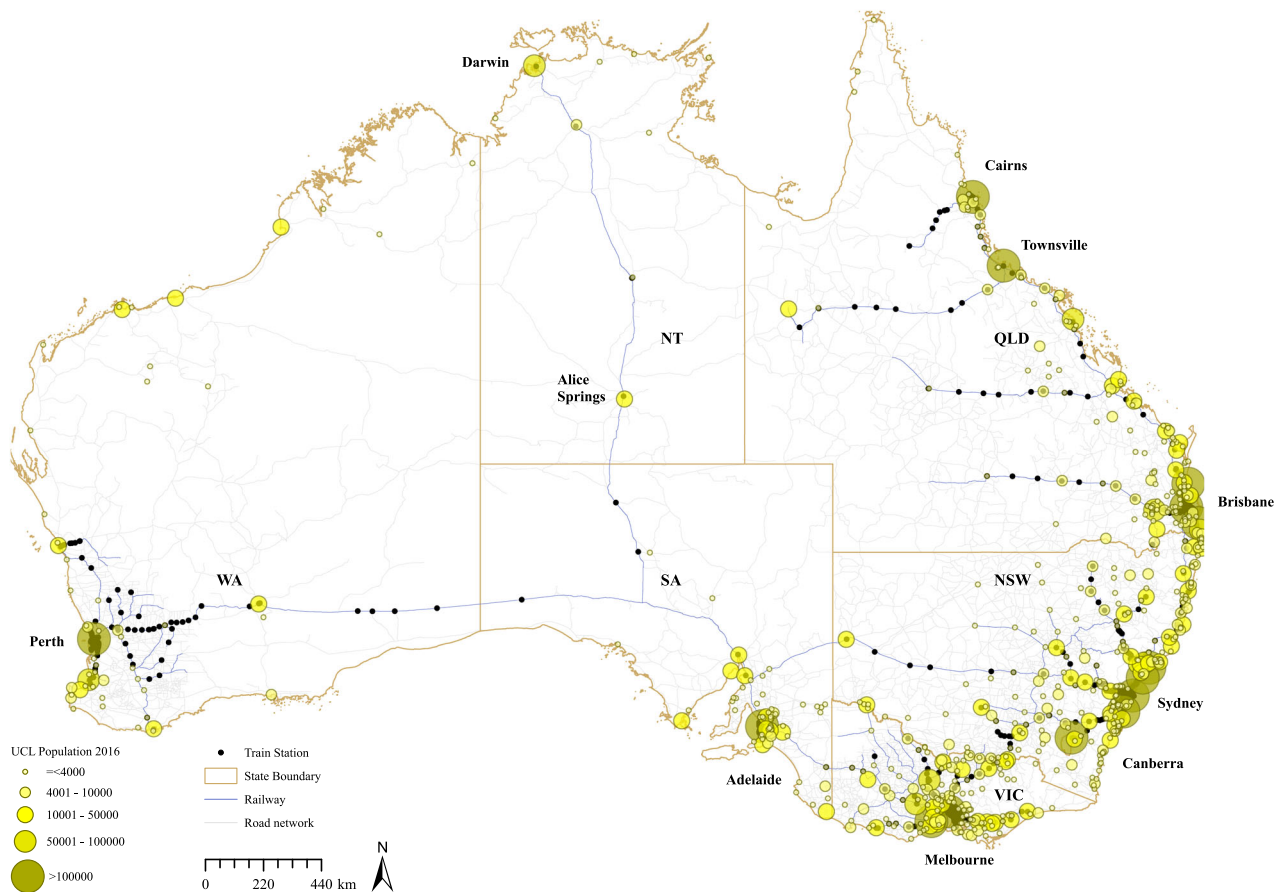
**Fig. 2 | Conceptual framework for theorising sustainable city size.** This conceptual diagram presents an integrated framework connecting policy levers to urban theories and the factors that influence theoretical city size. On the far left, key policy levers (e.g., governance, transport infrastructure, digital connectivity, and land-use planning) are shown as inputs that influence four foundational urban theories: central place theory, central flow theory, agglomeration economics, and locational

fundamentals. Each theory is associated with specific city size factors, such as accessibility, inter-city connectivity, productivity, and environmental constraints. These factors contribute to the conceptualisation of a city’s theoretical size, which is compared against actual population to identify deviations with implications for sustainability outcomes. The superscript numbers in the figure refer to references cited in the main text.

systems. Central place theory explains vertical hierarchies based on functional differentiation and service range (Fig. 1a)<sup>48</sup>. Central flow theory, by contrast, highlights horizontal flows (e.g., labour, capital, freight, information), emphasising inter-city connectivity (Fig. 1b)<sup>49–52</sup>. Concepts such as ‘borrowed size’ and ‘borrowed function’ illustrate how cities can benefit from their neighbours—e.g., smaller cities borrowing function, and larger cities borrowing population<sup>27,53–55</sup>.

This study integrates four theoretical traditions (locational fundamental, increasing returns, central place, and central flow) into a unified framework for estimating theoretical city size (Fig. 2). These theories have typically been applied in isolation due to perceived contradictions or assumptions that each applies to a different stage of urbanization<sup>36,38</sup>. For example, Meijers<sup>56</sup>, p.246 states that central flow theory “is essentially

opposite to the central place model”, emphasising horizontal rather than vertical hierarchies. In contrast, this study considers the theories as complementary and jointly explanatory of city size and growth. The framework identifies theory-specific factors and shows how policy interventions can influence them. It also recognises feedback loops: for example, population growth can enhance accessibility and specialisation, which in turn attract further growth<sup>57,58</sup>, highlighting the co-evolutionary nature of city size. This framework informs a method to assess whether cities operate above, below, or near their theoretical size. Cities near this threshold are defined as having *sustainable city size*. The paper tests whether such cities perform better across three sustainability outcomes—economic (median rent), environmental (% of commuters walk to work), and socio-environmental (share of households with 2+ vehicles)—using data from 655 Urban Centres and



**Fig. 3 | Location and population distribution of Urban Centres and Localities (UCLs) across Australia.** The map displays urban centres across Australia, classified by their 2016 population size (UCL) into five categories. Major transport

infrastructure, state boundaries, and key cities are overlaid. It shows that Australia’s larger cities are all located along the coastline, with a greater concentration in the southeast.

Localities (UCLs) in Australia (Fig. 3). Two population thresholds are examined: (a) UCLs with  $\geq 4000$  residents ( $n = 205$ ) and those with  $\geq 10,000$  residents ( $n = 100$ ), with and without inclusion of the seven mainland capital cities. To ensure robustness, the analysis addresses potential reverse causality between city size and its determinants using instrumental variables.

## Results

By testing the research hypotheses, the study addresses the following three research questions: (1) How can the theoretical size of cities be robustly determined? (2) How does a city’s actual size affect its ability to achieve its theoretical size? and (3) How does the alignment (or misalignment) between actual and theoretical size influence sustainability outcomes?

Table 1 presents descriptive statistics for the variables used to estimate theoretical size of Australian cities. On average, each of the 655 UCLs had approximately 25 K people in 2016. As expected, the average population size increases to 78 K and 152 K when the analytical samples are reduced to UCLs with  $\geq 4$  K and  $\geq 10$  K populations, respectively. However, this reduction in sample size increases variability, suggesting that the excluded UCLs have more homogenous (albeit smaller in size) populations.

The increase in population density with reduced sample sizes indicates that the excluded UCLs are more rural, occupying larger lot sizes. Similarly, job accessibility levels increased as the sample was restricted, suggesting that accessible employment opportunities are more concentrated in larger settlements. Average driving time to the nearest train station was lower for the  $\geq 4$  K sample (40 min) compared to the full sample (56 min) and the  $\geq 10$  K sample (46 min), implying that many mid-sized UCLs (4 K–10 K) had a train station nearby. On average, each UCL had 57 K population in nearby settlements (borrowed size), a figure that decreased slightly (to 46–50 K) as

the sample was restricted. This suggests that larger UCLs had lower borrowed sizes due to their greater internal population mass. However, the slightly higher borrowed size for the  $\geq 10$  K UCLs compared to the  $\geq 4$  K UCLs suggests that mid-sized UCLs ( $\geq 4$  K and  $< 10$  K) are more isolated. Borrowed function indicators showed similar patterns. A gradual decrease in per capital availability of amenities (per 10 K population) with reduced sample sizes indicates that smaller settlements enjoyed greater access to amenities relative to their size. Little variation was observed in travel time to the nearest capital city or across climatic variables (rainfall, sunshine, temperature), indicating that excluded and included UCLs are similarly distributed in geographic terms.

Larger UCLs had a higher proportion of specialist jobs compared to smaller UCLs. However, occupational diversity levels were similar across settlement groups, suggesting that smaller UCLs trade off specialist roles with a broader base of non-specialist employment. Little variation was found in closeness centrality across the groups, while betweenness centrality scores were notably higher for larger UCLs, indicating that they more frequently fall on the shortest paths connecting other cities. However, higher standard deviations for the  $\geq 10$  K group imply greater internal variation in network dominance. Only 7 UCLs are designated as state or territory capitals (Sydney, Melbourne, Brisbane, Adelaide, Perth, Canberra and Darwin) (Fig. 3), and all are included in the  $\geq 10$  K sample group.

Smaller UCLs performed better on two of the three sustainability outcomes examined in this study (economic and environmental), but exhibited the opposite pattern for the socio-environmental sustainability outcome. As shown in Table 2, smaller settlements had lower median rents and a higher share of people walking to work, indicating stronger economic and environmental sustainability. However, they also exhibited a slightly

**Table 1 | Descriptive summary of the variables**

Variables	Population ≥1 K		Population ≥4 K		Population ≥10 K	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
<i>Dependent variable</i>						
Population size	25,495.12	215,419.40	77,845.85	381,447.80	152,380.50	536,044.20
<i>Explanatory variables</i>						
Population density (person/km2)	407.11	265.02	477.76	250.86	550.62	259.35
Job accessibility (driving)	30,249.71	56,643.00	38,877.96	82,848.54	58,303.40	111,615.40
Average driving time to train station (min)	55.76	112.62	40.09	100.83	45.76	139.84
Borrowed sizes (population)	57,180.15	86,291.29	45,959.69	70,419.31	49,782.31	73,439.49
Borrowed functions (professional job)	5,347.63	9000.54	4412.60	7849.89	4912.03	8695.78
Point of Interests per 10,000 population	117.99	111.34	82.65	60.75	72.04	51.90
Occupational diversity index	0.86	0.01	0.86	0.01	0.86	0.01
Specialist jobs (% of professional job)	15.60	5.17	16.90	4.62	18.43	4.74
Closeness centrality*	1.18	0.31	1.19	0.31	1.18	0.32
Betweenness centrality	4170.30	7525.83	4378.20	7791.32	4681.06	8518.96
Capital city status (1 – Yes, 0 – No)	0.01	0.10	0.03	0.18	0.07	0.26
Travel time to capital city	219.56	244.17	219.87	243.86	209.14	244.69
Rainfall	851.56	434.14	857.19	428.76	829.13	398.95
Sunshine	6.74	0.94	6.71	0.97	6.71	1.05
Temperature	17.20	3.38	17.31	3.39	17.40	3.53
<i>City outcome variable</i>						
Median rent (AU\$) paid per week (economic)	261.56	80.70	274.15	63.33	282.59	56.91
% of people walked to work (environmental)	5.66	6.15	4.08	1.68	3.52	1.23
% of households with 2+ vehicles (socio-environmental)	54.67	11.58	52.40	7.49	52.50	6.28
N	655		205		100	

\* Original values are multiplied by 1000,000,000.

**Table 2 | Multiple linear regression model results showing the factors associated with optimum city sizes**

Explanatory factors	Dependent variable: natural log of population sizes in UCLs			
	Population ≥10 K		Population ≥4 K	
	(1) Population ≥10 K	(2) Population ≥10 K without capital cities	(3) Population ≥4 K	(4) Population ≥4 K without capital cities
	B	B	B	B
Capital city status (1 – Yes, 0 – No)	2.53**	-	3.26**	-
Log of job accessibility	0.43**	0.37**	0.36**	0.33**
Log of % employed in the professional category (high-level urban function)	1.55**	1.51**	1.24**	1.23**
Log of urban diversity	27.13**	27.48**	20.59**	20.38**
Log of betweenness centrality	-0.04**	-0.05**	NS	NS
Constant	5.44**	6.35**	5.54**	5.74**
R <sup>2</sup>	0.72	0.38	0.60	0.29
F	48.96**	13.23**	74.80**	25.80**
N	100	93	204	197

- Variables are not included in the model, \*\*Coefficients are significant at the 0.05 level, NS Not statistically significant.

higher proportion of households owning two or more vehicles, suggesting reduced socio-environmental sustainability due to increased reliance on private transport.

Overall, the descriptive trends observed across different indicators align with the literature as reported in the Introduction and are commensurate with the general understanding of the topic.

**Determinants of theoretical city sizes**

Table 2 presents results from four multiple linear regressions models estimating how theoretically derived factors explain city sizes in Australia (Model 1: ≥10 K model with capital cities; Model 2: ≥10 K model without capital cities; Model 3: ≥4 K model with capital cities; and Model 4: ≥4 K model without capital cities).

**Table 3 | Instrumental variable estimates**

Explanatory factors	Dependent variable: natural log of population sizes in UCLs			
	UCLs with Population $\geq 10$ K	UCLs with Population $\geq 4$ K		
	(5) Baseline structural model	(6) Baseline structural model		
	(6) Instrument: Spatial lag of job accessibility	(7) Instrument: Spatial lag of urban diversity		
	B	B		
	(9) Instrument: Spatial lag of job accessibility	(10) Instrument: Spatial lag of urban diversity		
	B	B		
Capital city status (1 - Yes, 0 - No)	2.73**	2.62**	3.63**	3.33**
Log of job accessibility	0.43**	0.47**	0.36**	0.37**
Log of % employed in the professional category (high-level urban function)	1.55**	1.60**	1.24**	1.44**
Log of urban diversity	27.13**	26.45**	20.59**	31.67**
Log of betweenness centrality	-0.04**	-0.33**	NS	-
Constant	5.44**	6.18**	5.54**	6.50**
R <sup>2</sup>	0.72	0.71	0.60	0.59
F (Wald chi <sup>2</sup> for the instrument models)	48.96**	233.06**	209.75**	289.26**
N	100	100	204	204
Durbin (score) chi <sup>2</sup>	1.54	2.43	12.95**	0.59
Wu-Hausman F test	1.46	2.31	13.42**	0.57
Partial R <sup>2</sup>	0.50	0.11	0.56	0.12

- Variables are not included in the model, \*\*Coefficients are significant at the 0.05 level, NS Not statistically significant.

Among the 15 explanatory factors considered in this study (Table 1), Model 1 shows that only five of them had a statistically significant association with theoretical size of cities (capital city status, job accessibility, high-level urban function, urban diversity, and betweenness centrality). These five factors explained 72% variations in the population size among the 100 UCLs considered for this sample. The signs of the coefficients were also found to be intuitive. Model 1 shows that the expected percent increase in the geometric mean of population sizes from non-capital cities to capital cities is about 1155%  $(\frac{e^{2.53}-1}{e-1} * 100) \approx (12.55-1) * 100 \approx 1155$ . Or, simply, it can be said that population sizes were nearly 12 times higher for the UCLs designated as capital cities than the other UCLs considered in this group. This finding indicates the strong influence of the locational fundamentals theory in determining city sizes in Australia. It is also to be noted here that this variable was strongly correlated with the travel time to the capital city variable, suggesting the strong influence of the central place theory as well. The influence of central place dynamics is also reflected by the high-level urban function variable in Model 1. It shows that a one percent increase in the high-level urban function increases the theoretical size of cities by 1.5%.

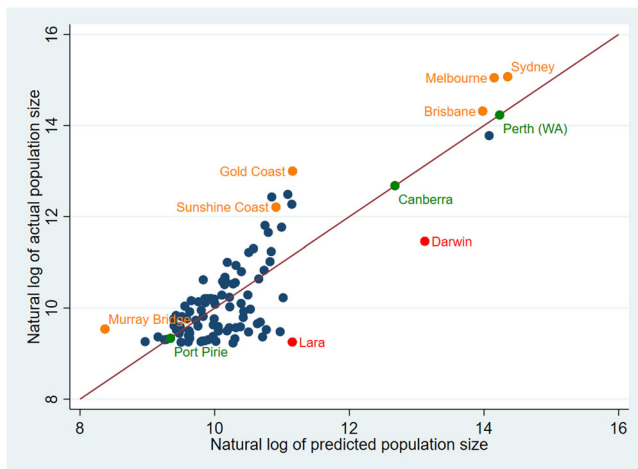
The increasing returns to scale theory is represented by the job accessibility indicator in Table 2, the effect of which is also verified in this study. Model 1 shows that for a one percent increase in job accessibility level, theoretical sizes increased by 0.43%. The effect of the increasing returns to scale theory is further reinforced by the evidence that a one percent increase in urban diversity level, theoretical sizes increased by 27%.

In contrast to the theoretical underpinning of the central flow theory that cities receiving an increasing flow will have a larger theoretical size, the betweenness centrality factor in Model 1 shows that the UCLs that frequently fell on the shortest paths between other UCL had a lower theoretical population size. To be specific, the theoretical size of a UCL decreased by 0.04% with a one percent increase in their betweenness centrality. This finding is not unexpected in the Australian context, where major cities are historically located on the coastline (Fig. 3), reducing their likelihood of falling frequently along the shortest paths between other settlements.

Model 2 broadly corroborates the results of Model 1, suggesting that the underlying drivers of theoretical city sizes are similar for both capital and non-capital cities. Since Model 2 represents non-capital cities exclusively, the capital city variable logically does not appear in this model.

The modelling results from the  $\geq 4$  K sample group (Model 3) follow the general trend observed in the  $\geq 10$  K group (Model 1), with the exception of four significant factors identified in Model 3 instead of five factors in Model 1. While the betweenness centrality factor was statistically significant in Model 1, this factor became insignificant in Model 3. However, the remaining four significant factors in Model 3 were also appeared in Model 1, and the signs of the coefficients remained consistent, suggesting the stability of these factors across different sample groups. The magnitude of the coefficients was slightly modified, which is expected due to the inclusion of relatively smaller-sized UCLs in Model 3. For example, Model 3 indicates that capital cities theoretically contained 25 times more people than other UCLs in this group. Overall, Model 3 explains 60% of the variation in city size. Additionally, Model 4 in Table 2 supports the conclusions drawn from Model 3, indicating that the determinants of theoretical city sizes are consistent across both capital and non-capital cities.

Table 3 presents findings generated from a two-stage least squares (2SLS) model applying instrument variable approaches to isolate the effects of reverse causality between city size and its determinants. Models 5 and 8 replicate Models 1 and 3, respectively; and are included in Table 3 to facilitate comparisons. Four 2SLS models were estimated. In Models 6 and 7, job accessibility and urban diversity are respectively instrumented based on the specification of Model 5 for the  $\geq 10$  K group, while Models 9 and 10 represent the same for the  $\geq 4$  K group. Identifying suitable instruments is widely recognised as a methodological challenge<sup>59</sup>. Nevertheless, the strength test results of the applied instruments in this study (spatial lag of job accessibility and spatial lag of urban diversity) indicate that they were strong, as reflected by their partial R<sup>2</sup> values. The overall explanatory powers of the 2SLS models (Models 6, 7, 9, and 10) remained similar to their respective



**Fig. 4 | Theoretical vs. actual population sizes of cities with  $\geq 10,000$  residents.** This scatter plot compares the actual population size of Australian cities ( $\geq 10,000$  residents) with their theoretical size, as estimated from the integrated urban framework. The 1:1 reference line (diagonal line) indicates perfect alignment between theoretical and actual sizes. Cities plotted above the reference line exceed their theoretical population, while those below fall short. To provide contextual understanding, selected cities are labelled based on their categories. Cities labelled in orange represent those whose current population exceeds what can be theoretically supported. Cities labelled in red have the capacity to accommodate further growth, while those in green are considered to be in population equilibrium.

structural models (Models 5 and 8). Despite the strength of the instruments, both the Wu-Hausman and Durbin test results show that they are not statistically significant in Models 6 and 7 for the  $\geq 10$  K group, and in Model 10 for the  $\geq 4$  K group. This means that the urban diversity variable was not subjected to endogeneity bias for any of the sample groups whereas the presence of endogeneity is confirmed for the job accessibility variable for the  $\geq 4$  K group. Consequently, the linear regression coefficient for job accessibility in Model 8 is likely biased, and Model 9 provides the preferred, unbiased estimate for  $\geq 4$  K group. Irrespective of the presence/absence of endogeneity, the directions of the coefficients are identical, and the magnitudes of the coefficients are in an acceptable level of agreement between the models. As expected, the result from the 2SLS model (Model 9) shows a relatively weaker coefficient estimate for the job accessibility variable.

These results establish a robust basis for estimating theoretical city sizes, enabling the subsequent comparison between cities' actual and theoretical population sizes.

### Actual vs. theoretical city sizes in Australia

This section presents results based on Model 1, as key findings remain consistent across alternative model specifications (Models 1–10). Figure 4 plots the theoretical population sizes (log-transformed) as estimated in Model 1 against the actual population sizes (log-transformed) for the 100 UCLs in the  $\geq 10$  K group, with selected UCLs labelled for contextual illustration.

It is evident that several UCLs, such as Perth, and Canberra, operated near equilibrium conditions in 2016, lying close to the diagonal line. Their actual population sizes closely matched their theoretical sizes. Importantly, actual population size itself was not a determinant of whether a UCL operated in equilibrium. For example, Perth had a large actual population (14.23 in log scale), while Port Pirie had a much smaller population (9.34 in log scale), yet both aligned closely with their estimated theoretical sizes.

Conversely, many cities exhibited stress—where actual population exceeded theoretical capacity—as indicated by points located above the diagonal (e.g., Melbourne, Gold Coast, Murray Bridge). Despite large differences in actual population, both Melbourne and Murray Bridge showed

signs of stress, illustrating that stress was not simply a function of absolute population size.

Cities falling below the diagonal line represent cases where the actual population was lower than the theoretical size. These cities could be deemed unproductive because they have the available infrastructure and services to accommodate more people, yet experience population shortfalls. As with stress, underpopulation was not driven by size alone. For example, both Darwin (11.46 in log scale) and Lara (9.25 in log scale) had actual populations below their theoretical expectations, despite Darwin being substantially larger.

Having established how actual and theoretical city sizes align or diverge, the next section examines how these deviations relate to cities' sustainability outcomes.

### Actual vs. theoretical sizes and sustainability outcomes

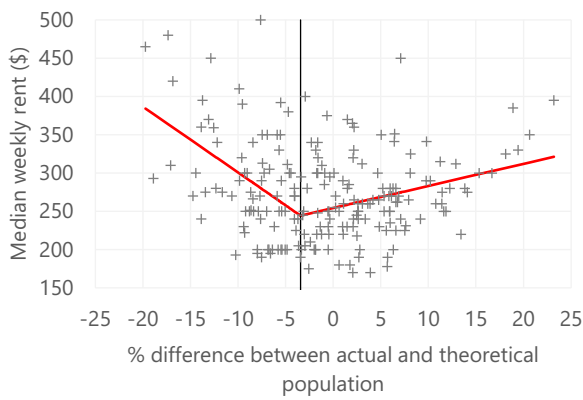
A key hypothesis of this study is that cities perform more sustainably when their actual population sizes align closely with their theoretical sizes. This condition is referred to as having a 'sustainable city size'. Figure 5 summarises findings generated from piecewise linear regression models showing how the deviations between theoretical and actual population sizes are linked to the three sustainability outcomes examined in this study (weekly median rent, % of households with two or more vehicles, and % of people walking to work) across the two groups of UCLs ( $\geq 4$  K and  $\geq 10$  K). Further statistical details generated from these models are shown in Table 4. These models were estimated separately with and without the two covariates considered (population density and median household income), yielding 12 models in total (Models 11–22). Models without and with the covariates are referred to as the basic and expanded models, respectively. For theoretical size estimation, Model 1 was used for the  $\geq 10$  K group and Model 9 for the  $\geq 4$  K group.

In Fig. 5, the x-axis shows the percentage deviation of each UCL from its theoretical population size (on a log scale), while the y-axis presents the real values of the three sustainability outcome variables. A zero value on the x-axis represents perfect population equilibrium; positive values indicate actual population exceeding theoretical size, and negative values indicate population shortfall. Vertical lines in Fig. 5 indicate breakpoints – i.e., the optimal point for each outcome, estimated using Stata's *nl* (nonlinear regression) command.

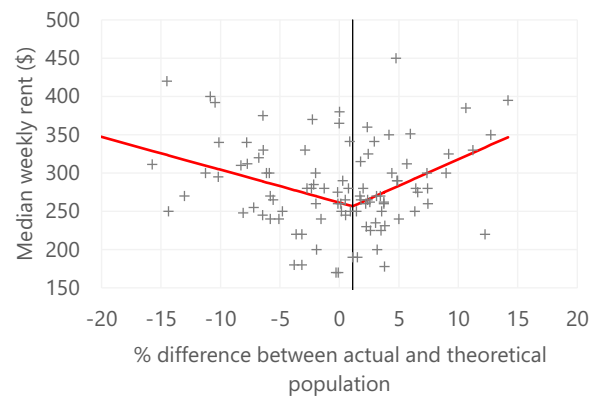
Table 4 confirms that all basic models (without covariates) achieve statistical significance at the 95% confidence level, suggesting that deviations in city size meaningfully predict sustainability outcomes regardless of sample groups. Including covariates substantially increased explanatory power (expanded models), suggesting that sustainability outcomes are also influenced by other factors. Nevertheless, the primary intent here is to test the robustness of the deviation effect. Table 4 demonstrates that the sign and magnitude of the deviation coefficients remain stable across basic and expanded models, confirming that city size deviation is a robust predictor of sustainability. Figure 5 is based on the basic models, offering a consistent and interpretable visual representation of the core relationships without the influence of covariates.

In Fig. 5a, UCLs with actual populations below their theoretical size experienced higher rental burdens. Prior to the breakpoint, a 1% reduction in the deviation (i.e., actual size approaching theoretical) was associated with \$9 ( $\beta_1$ ) decrease in the rent, reaching the lowest predicted rent of \$244 (as indicated by the model constant) when the deviation was approximately -3% ( $\alpha$ ). The figure is \$30 lower than the overall average rent (\$274) for this group, translating to annual savings of \$1560 per household. After the breakpoint, rent increased by \$3 ( $\beta_2$ ) for every 1% increase in actual population compared to theoretical size. A similar pattern is evident for the  $\geq 10$  K group (Fig. 5b), although the breakpoint shifts to 1.12% with different rates of decline and increase ( $\beta_1 = -4.29$ ;  $\beta_2 = 6.90$ ). The minimum rent at the breakpoint is \$256, which is \$27 less than the group's average of \$283, implying annual savings of \$1404 per household.

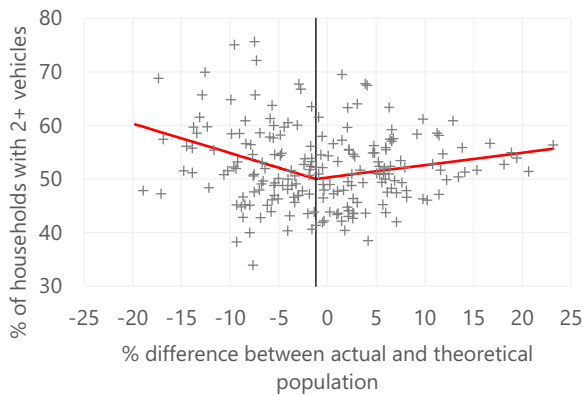
Figure 5c, d show that equilibrium city sizes are associated with lower household vehicle ownership. In the  $\geq 4$  K sample, two-plus vehicle



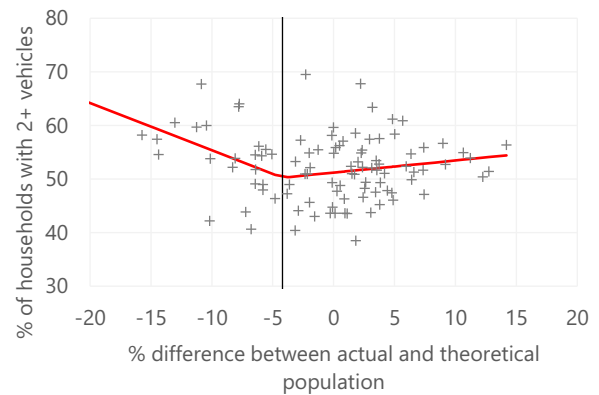
a) Economic sustainability (≥4K sample)



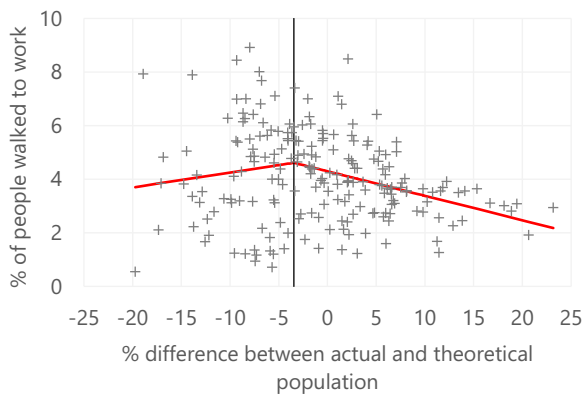
b) Economic sustainability (≥10K sample)



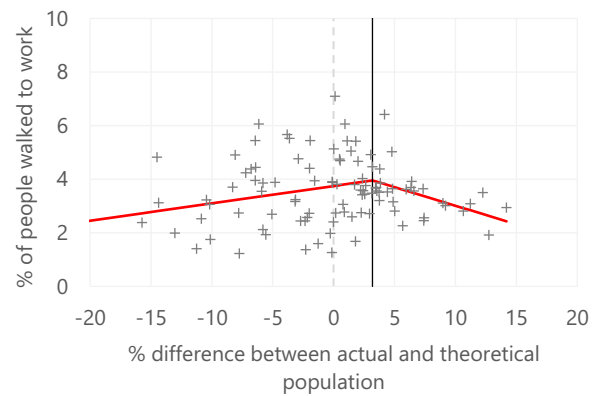
c) Social/environmental sustainability (≥4K sample)



d) Social/environmental sustainability (≥10K sample)



e) Environmental sustainability (≥4K sample)



f) Environmental sustainability (≥10K sample)

**Fig. 5 | Piecewise regression plots showing relationship between deviations from theoretical city size and sustainability outcomes.** This figure presents six scatter plots illustrating how deviations from a city’s theoretical size are associated with three sustainability outcomes: median weekly rent (top row), percentage of households with two or more vehicles (middle row), and percentage of people walking to work (bottom row). The left column includes Urban Centres and Localities (UCLs) with populations ≥ 4000, and the right column includes those with populations ≥ 10,000. Each red line represents a fitted piecewise regression line (based on basic models without covariates), identifying potential breakpoints where the relationship between size deviation and the sustainability outcome shifts. The x-axis in each panel shows the deviation between actual and theoretical city size, while the y-axis shows

the corresponding value for each sustainability measure. The individual panels are described below for clarity: **a** Median weekly rent vs. size deviation for UCLs with populations ≥ 4000. **b** Median weekly rent vs. size deviation for UCLs with populations ≥ 10,000. **c** Percentage of households with two or more vehicles vs. size deviation for UCLs with populations ≥ 4000. **d** Percentage of households with two or more vehicles vs. size deviation for UCLs with populations ≥ 10,000. **e** Percentage of people walking to work vs. size deviation for UCLs with populations ≥ 4000. **f** Percentage of people walking to work vs. size deviation for UCLs with populations ≥ 10,000. These plots support the evaluation of whether city over- or under-sizing is associated with better or worse sustainability performance.

**Table 4 | Piecewise regression analysis results**

Variables	Weekly rent (AU\$)—4 K sample		Weekly rent (AU\$)—10 K sample		2+ vehicles—4 K sample		2+ vehicles—10 K sample		% walked to work—4 K sample		% walked to work—10 K sample	
	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
Basic	-3.44	-3.44	1.12	1.12	-1.20	-1.20	-4.2	-4.21	-3.44	-3.44	3.19	3.19
Expanded	-6.06**	-6.06**	-4.29**	-3.13**	-0.55**	-0.30**	-0.88**	-0.66**	0.06	0.01	0.06**	0.06**
Coefficient before breakpoint ( $\beta_1$ )	2.89**	1.11	6.90**	4.77**	0.23**	0.02	0.23	0.11	-0.09**	-0.06**	-0.14**	-0.12**
Coefficient after breakpoint ( $\beta_2$ )												
Covariates												
Density (person/km <sup>2</sup> )	-	0.06**	-	0.07**	-	-0.01	-	-0.01*	-	-0.01**	-	-0.01
Median household income (\$/week)	-	0.07**	-	0.07**	-	0.02**	-	0.01**	-	-0.01**	-	-0.01
Constant	244.24**	139.99**	256.46**	139.01**	50.00**	32.32**	50.25**	37.82**	4.61**	6.51**	3.95**	4.37**
R <sup>2</sup>	0.17	0.40	0.16	0.41	0.07	0.62	0.13	0.59	0.08	0.19	0.09	0.11
N	205	205	100	100	205	205	100	100	205	205	100	100

\*\*\* Breakpoint was estimated individually for each model using nonlinear (nl) regression.

\*\* Coefficients are statistically significant at the 0.05 level.

\* Coefficients are statistically significant at the 0.1 level.

- Related explanatory factors were not included in the model.

households decline to 50% at the breakpoint ( $\alpha = -1.20$ ), then increased thereafter (Model 15). For the  $\geq 10$  K sample (Fig. 5d and Model 17), two-plus vehicle ownership declined significantly up to  $\alpha = -4.21$ , after which the pattern flattened. Across both groups, equilibrium cities had about 2.5% fewer households with multiple vehicles (50% vs. 52.5%).

Sustainable travel outcomes are shown in Fig. 5e, f. In the  $\geq 10$  K group (Fig. 5f), walking rates rose by 0.06% for every 1% shift towards equilibrium, peaking at 4% when the deviation reached +3%. Beyond this point, walking declined by 0.14% per 1% increase in actual population size. On average, 4% of commuters walked to work in UCLs with equilibrium sizes, higher than the group mean of 3.52%. Similar trends are observed for the  $\geq 4$  K group.

These findings consistently support the theoretical proposition that deviations from sustainable city size, in either direction, are associated with measurable reductions in sustainability performance, reinforcing the value of aligning actual city sizes with their theoretical capacities.

### Discussion

The optimum city size has long been debated in the literature, yet findings linking city size to sustainability outcomes remain inconsistent. This study posits that such inconsistency arises from relying on actual population measures, rather than assessing cities against their theoretical capacity to function within an urban system. To test this, a unified conceptual framework was developed by integrating four foundational theories—location fundamentals, increasing returns to scale, central place, and central flow. The findings demonstrate the utility of this integrated approach for estimating theoretical city size and confirm that these theories collectively explain variation across urban contexts, challenging the view that they are only applicable in specific spatial or temporal conditions<sup>36,38,56</sup>.

This theoretical contribution is supported by empirical evidence. The study found that the determinants of theoretical city size—particularly capital city status, job accessibility, urban diversity, and high-level urban functions—are consistent across different sample groups ( $\geq 10$  K and  $\geq 4$  K UCLs) and remain stable when capital cities are excluded. While betweenness centrality exhibited a negative association with city size among larger UCLs, this finding aligns with Australia’s unique coastal urban morphology. Instrumental variable analyses further validated the model, identifying endogeneity in the job accessibility measure for the  $\geq 4$  K group but confirming the stability of the core results. Together, these findings provide strong empirical support for factors determining theoretical city size and set the stage for testing the main hypothesis: that cities perform better when their actual populations closely align with their theoretical size—a condition defined in this study as sustainable city size. This hypothesis is confirmed by piecewise regression models, which show that cities operating near their theoretical sizes—within a  $\pm 4\%$  deviation—consistently achieve better outcomes in rental affordability, multi-vehicle ownership, and active commuting. Based on the findings, three key conclusions are drawn: (1) Theoretical city size can be determined a priori; (2) Actual city size does not determine whether it operates near theoretical size; and (3) A city’s sustainability hinges on its equilibrium size, not its actual size.

Urban planning typically relies on projected population growth<sup>5,7</sup>. For existing cities, this is relatively straightforward to derive due to known demographic trends. However, determining the population for new cities is more complex, particularly given their diversity in type and function—ranging from capital cities to smart cities, eco-cities, and airport hubs<sup>1,60,61</sup>. Examples include Canberra, Brasilia, Putrajaya, Islamabad, Abu Dhabi, Dholera, Songdo, Shenzhen, Kwinana, Yanbu, Waterfall City, and Lanseria Airport City<sup>4,61-66</sup>. Many planned cities, such as Canberra, Yanbu, Putrajaya have used arbitrary or politically motivated population targets, lacking theoretical justification<sup>63,67,68</sup>. This study addresses that gap by demonstrating that theoretical population size can be derived from known or assignable characteristics—such as capital city status, planned job accessibility, urban function diversity, and network centrality. These factors are either designed explicitly by planners (e.g., land use zoning, transit infrastructure) or known in advance (e.g., whether the city is a capital). The framework is equally useful for recalibrating the size of existing cities where

traditional trend-based projections fall short. Such projects often fail to account for inter-city relationships or the impact of new infrastructure on regional connectivity<sup>7</sup>. By incorporating central place and central flow dynamics, the model captures how investments or spatial shifts influence population potential—something absent in conventional methods.

The findings presented in this study confirm that cities of any actual size can operate at or near their theoretical size. For example, both Perth (a large capital city) and Port Pirie (a small regional centre) aligned closely with their theoretical size in 2016. This supports the study's hypothesis: that each city has a unique theoretical size, shaped by a specific combination of factors—capital status, urban function diversity, job accessibility, professional job share, and network centrality. These determinants are dynamic and can be modified through targeted planning interventions, suggesting that a city's theoretical size is not static. For example, cities operating above their theoretical size (e.g., Melbourne or the Gold Coast) could increase job accessibility or functional diversity through new transport links or decentralised employment strategies, helping align actual and theoretical sizes. Conversely, underperforming cities like Darwin or Lara already have the capacity to absorb growth without further investment—offering latent potential for spatially distributed growth. These findings suggest a new way of planning for growth—not by reinforcing expansion in already large cities, but by identifying and leveraging untapped capacity in smaller, well-positioned settlements. Unlike traditional population distribution models, this approach embeds strategy, infrastructure alignment, and systemic capacity into city sizing decisions<sup>69</sup>.

Findings presented in Fig. 5 show that cities whose actual sizes closely align with their theoretical sizes—estimated here as being within  $\pm 4\%$ —consistently perform better across economic, socio-environmental, and environmental sustainability metrics. Economically, households in equilibrium cities save approximately \$1560 per year in rent, equating to \$5.3 billion in national savings across Australia's rental population (30.6% of 11 million households in Australia lived in rental properties in 2021). Socio-environmentally, equilibrium cities have 2.5% fewer households owning two or more vehicles, reducing car-dependence and enhancing social inclusion—equivalent to 275,000 households shifting to fewer vehicles in Australia. Environmentally, these cities support 0.5% more walking trips to work, generating 44,160 additional walking trips daily among the Australian commuters (there were 9.2 million commuters in Australia in 2021). These optimal outcomes occur not at extremes of scale, but at the point where actual and theoretical sizes are aligned—validating the sustainable city size hypothesis. The findings provide an empirically grounded target range (approximately  $-4.2\%$  to  $+3.2\%$ ) that can serve as a benchmark for planning sustainable cities.

A range of actionable insights are drawn from the conceptual framework (Fig. 2) and empirical findings. First, for strategic planning in new cities: the ability to estimate theoretical size using a priori indicators enables more informed decisions on land use, infrastructure investment, and zoning provisions from the outset of city development. Second, for targeting growth in existing cities: underutilised cities (e.g., Lara, Darwin) can be identified as strategic growth nodes, while over-capacity cities (e.g., Melbourne, Gold Coast) may require interventions—such as high-speed rail, employment decentralisation, or land use reforms—to expand their theoretical size and reduce pressure on infrastructure and housing. Third, as an empirical planning benchmark: the  $\pm 4\%$  deviation threshold around theoretical size provides an empirically grounded benchmark for assessing whether a city is operating within a sustainable population range. Fourth, for policy-responsive growth modelling: the framework highlights that theoretical size is shaped by modifiable factors such as connectivity, diversity, and accessibility. This enables policy levers to proactively shift a city's capacity to accommodate growth sustainably. Finally, for scenario planning and investment alignment: by embedding the framework into spatial planning or land use-transport models, agencies can assess how various interventions (e.g., transit upgrades, zoning changes) alter theoretical capacity and associated sustainability outcomes.

While the study presents a novel and empirically supported framework, several limitations and future directions remain. First, city boundary definitions: UCLs provides national consistency but may not fully capture functional urban areas such as labour markets. Future studies could test alternative spatial definitions to assess boundary sensitivity. Second, sustainability indicators: the study relies on three proxies—rent, multi-vehicle ownership, and walking to work—which reflect data availability rather than a comprehensive sustainability framework. Future research should consider broader indicators such as emissions, equity, or health. This is why the study frames it as a *theory* rather than a principle: it is strongly supported but requires further evidence for generalisation. Third, generalisability: the model is developed in an Australian context, and its relevance to other countries or metropolitan regions remains to be tested. Comparative studies could strengthen its applicability. Fourth, governance and implementation: institutional, political, or financial constraints may limit efforts to achieve theoretical city size. Qualitative studies could explore these barriers. Fifth, integration into dynamic models: applying the framework in Land Use-Transport Interaction (LUTI) or simulation models would support long-term, scenario-based planning. Finally, measurement of size deviation: the use of log-residuals converted to percentage terms assumes small linear deviations. While bias is likely minimal here, future studies could test alternative formulations or directly model residuals to assess robustness across broader size distributions.

## Methods

This study adopts a three-step empirical strategy building on the conceptual framework presented in Fig. 2: (a) Estimating the theoretical size of each city; (b) Measuring the level of deviation between actual and theoretical population sizes; and (c) Examining the relationship between these deviations and cities' sustainability outcomes.

### Study context and city boundary definition

Defining the boundary of a city is a long-standing methodological challenge in urban research<sup>70</sup>. Common approaches include administrative boundaries (e.g., municipalities), morphological definitions based on contiguous built-up areas (e.g., density), functional urban areas delineated by labour market integration or commuting flows, and thresholding based on nighttime lights<sup>71–73</sup>. Each approach captures different aspects of urbanity—governance, physical form, or socioeconomic function—and involves trade-offs<sup>74,75</sup>. The study was conducted in Australia and applies the country's statistical definition of urban areas—Urban Centres and Localities (UCLs)—developed by the Australian Bureau of Statistics (ABS)<sup>76</sup>. Only UCLs classified as urban centres are considered, excluding localities. An urban centre is defined as a cluster of contiguous SA1s (Statistical Area Level 1) where each SA1 has a population density of  $\geq 100$  persons/km<sup>2</sup> and a dwelling density of  $\geq 50$  dwellings/km<sup>2</sup>, with an aggregated population of at least 1000 people. This approach offers national consistency and comparability across Australia's urban settlement system. While UCLs may not fully capture functional linkages (e.g., commuting across city edges), they provide a pragmatic spatial unit for testing city size theories. Moreover, the key explanatory variables used in this study—such as amenities, density, and job accessibility—are closely aligned with the spatial extent of UCLs, reinforcing their suitability for the conceptual and empirical framework.

This study used data from the 2016 Census, as the 2021 Census was conducted during COVID-19 lockdowns (in some states and territories) and may not reflect typical behavioural outcomes. In 2016, 1853 UCLs were identified. UCLs located within Mainland Australia and meeting the ABS urban definition (population  $\geq 1000$ ) were initially selected, resulting in a study sample of 655 UCLs (Fig. 3). Subsequently, the hypotheses were tested across two population groups to verify the robustness of findings: (a) UCLs with populations of  $\geq 4000$  ( $n = 205$ ), enabling analysis across a diverse range of urban centres; and (b) UCLs with populations of  $\geq 10,000$  ( $n = 100$ ), focusing on relatively larger cities. Each group was further subdivided by excluding the seven UCLs designated as state or territory capitals in Mainland Australia to explore whether the determinants of theoretical city

size operate differently between capital and non-capital cities. These capital cities also attract the largest share of international migrants in Australia<sup>77</sup>, although recent trends suggest this dominance is diminishing<sup>58</sup>.

**Mathematical formulation of theoretical city size**

The choice of a place to live may be assumed to be driven by utility maximization, which occurs when the benefits of living equal or exceed the costs of living. These benefits and costs can be represented as functions of locational and network advantages and disadvantages, as conceptualised in Fig. 2. However, city size itself can act as both a benefit (e.g., greater amenities) and as a cost (e.g., longer commuting distance), as discussed in the Introduction.

Using job accessibility and distance to a train station as example factors representing benefits (*B*) and costs (*C*) respectively, the functional forms can be expressed as:

$$B = f(\text{size}, \text{job accessibility}) \tag{1}$$

$$C = f(\text{size}, \text{distance to train station}) \tag{2}$$

Assuming a standard Cobb-Douglas specification, these functions can be further specified as:

$$B = \text{size}^\alpha \times \text{job accessibility}^\beta \tag{3}$$

$$C = \text{size}^\gamma \times \text{distance to train station}^\delta \tag{4}$$

In an equilibrium state, the marginal cost equals the marginal benefit, that is:

$$\frac{\delta C}{\delta(\text{size})} = \frac{\delta B}{\delta(\text{size})} \tag{5}$$

Expanding the derivatives:

$$\implies \frac{\delta(\text{size}^\gamma \text{distance to train station}^\delta)}{\delta(\text{size})} = \frac{\delta(\text{size}^\alpha \text{job accessibility}^\beta)}{\delta(\text{size})} \tag{6}$$

$$\implies \gamma \text{size}^{\gamma-1} \text{distance to train station}^\delta = \alpha \text{size}^{\alpha-1} \text{job accessibility}^\beta \tag{7}$$

Rearranging:

$$\implies \text{size}^{\gamma-\alpha} = \frac{\alpha}{\gamma} \frac{\text{job accessibility}^\beta}{\text{distance to train station}^\delta} \tag{8}$$

Equation 8 can be log-linearized to obtain an estimable function:

$$(\gamma - \alpha)\ln(\text{size}) = \ln\left(\frac{\alpha}{\gamma}\right) + \beta \ln(\text{job accessibility}) - \delta \ln(\text{distance to train station}) \tag{9}$$

Generalising using *Y* and *X* as vectors of benefits and costs respectively:

$$\ln(\text{size}) = a + \sum_k b_k \ln(Y_k) - \sum_j c_j \ln(X_j) \tag{10}$$

where, *a* = intercept, *b<sub>k</sub>* = coefficients of benefit factors, and *c<sub>j</sub>* = coefficients of cost factors.

Equation 10 shows that a city’s theoretical size can be systematically derived based on its associated advantages and disadvantages as conceptualised in Fig. 2. This estimated value is referred to as the city’s **theoretical size** throughout the study. To operationalise this model empirically, indicators representing the key factors influencing theoretical city size were constructed from various spatial and census data sources, as described below.

**Operationalising indicators of theoretical city size**

Building on the theoretical model (Eq. 10), this section describes the operationalisation of the framework using empirical data. It details how key factors influencing theoretical city size were measured, the indicators used to represent these factors, and the sources of data, and the methods applied to estimate the theoretical population sizes of UCLs across Australia (Table 5).

The study focuses exclusively on indicators whose values can be either derived or assigned for both new and existing cities. For example, in the case of a new city, an urban planner can exogenously assign a target density, measure distance to planned public transport services, quantify the provision of public amenities (e.g., parks, playgrounds, childcare, schools, hospitals), estimate the number of jobs accessible within a certain travel time, regulate land use diversity through zoning controls, estimate the population or specialist functions in surrounding areas (borrowed size or borrowed function), and calculate centrality measures based on geographic location. This study used geographic centrality rather than flow centrality because the former only requires the geographic position of a development, whereas the latter depends on empirical data about flows of people, goods, or transaction—information unavailable for new developments. Accordingly, some factors known to be associated with city size, such as crime rates, urban sprawl, or rent levels, were excluded from the model<sup>28,34</sup>. These exclusions were made to ensure that the framework remains plannable and applicable to both new and existing cities, enhancing its practical relevance for policy-makers and urban planners. While many factors were directly sourced, some are derived as detailed below.

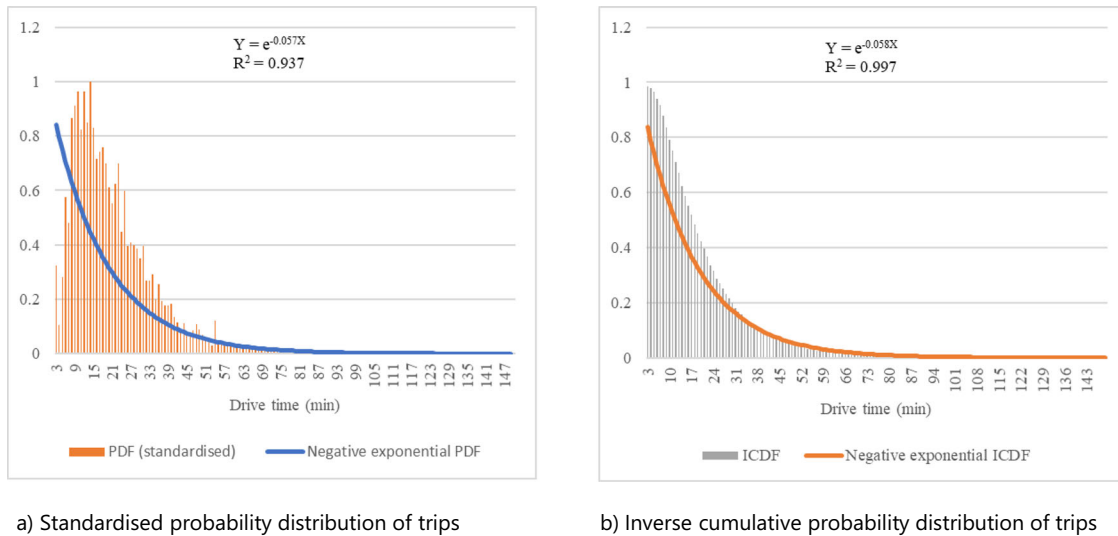
**Time to train station:** The study used OSM (open street map) railway network data featuring the 2021 version of railway lines and stations, downloaded from AURIN (Australian Urban Research Infrastructure Network). The driving time (min) between all the SA2s in Mainland Australia and their closest train station was derived using the Closest Facility function in ArcGIS Pro (v.3.0) based on its built-in transport network. The driving time of all SA2s belonging to a UCL was averaged to derive UCL-level access to train stations.

**Labour market efficiency:** Labour market efficiency is measured by the job accessibility indicator. Research shows that areas with good access to employment exhibit increasing returns to scale due to agglomeration effects<sup>46</sup>. Given that not all jobs are equally attractive (i.e., jobs located close by are more attractive than those located further away), a distance decay parameter was initially estimated to discount the travel time effect to derive the job accessibility level of each UCL. To estimate this parameter, 660 SA2s (statistical area level 2) were randomly selected from 2308 SA2s because of the limit posed by ArcGIS Pro (v.3.0) to derive OD (origin-destination) cost matrix. The SA2s were chosen because it is the smallest statistical unit for which the commuting flow data were available. This selection also enabled a finer level analysis than UCL, given that most UCL contains one or more SA2s (e.g., Melbourne as a UCL contains 292 SA2s). The commuting flow data, people who drove to work, between the SA2s were downloaded using the ABS TableBuilder functionality. The drive time (min) between the centroids of SA2s was derived using the OD (origin-destination) Cost Matrix function in ArcGIS Pro (v.3.0) based on its built-in transport network. The flow data were classified by travel time bins, ranging from 3 to 149 min, with one-minute interval. Commuting flows exceeding 149 min were found to be negligible (0.01%). The flow within each band was transformed into three indicators: (a) natural log; (b) standardised probability density function (PDF); and (c) inverse cumulative density function (ICDF)<sup>78</sup>. A range of curve estimation regression models were estimated in SPSS using these indicators including logarithmic, inverse, power, and exponential. Among them, the negative exponential function was found to better fit the data with a decay coefficient of 0.058 (Fig. 6). As a result, the negative exponential decay parameter was employed in this study, which is also the most widely used function in travel behaviour research<sup>79</sup>.

The derived distance decay parameter was then used to calculate job accessibility of the 655 UCLs based on Eq. 11. The drive times between the centroids of the 655 UCLs and all SA2s in Mainland Australia were derived using the OD Cost Matrix function in ArcGIS Pro based on its built-in

**Table 5 | Operationalisation of theoretical city size determinants: theories, factors, indicators, and data sources**

Theory	Factors	Indicators	Description	Source of data	
Locational fundamentals	City size	Population	Dependent variable: Residential populations within the UCLs	Census 2016	
	Outcome	Time to train station	Average drive time from all SA2s comprising a UCL to their nearest train stations	Derived based on ESRI network, AURIN, ABS, OSM data	
	Access to transport	Amenities	Point of interests per 10,000 residents within a UCL	AURIN, OSM	
	Attractiveness/status	Capital city	Classification of UCLs based on capital city status	Assigned	
		Climate zone	Classification of climate zones (hot humid summer; warm humid summer; hot dry summer, mild winter; hot dry summer, cold winter; warm summer, cool winter; and mild warm summer, cold winter)	Bureau of Meteorology (BoM)	
	Climate	Sunshine	Average annual distribution of sunshine duration	BoM	
		Temperature	Average annual distribution of mean temperature	BoM	
		Rainfall	Average annual rainfall	BoM	
	Increasing return to scale	Labour market efficiency	Job accessibility	Total number of jobs available in all SA2s in Australia from the centroid of an UCL discounted by driving time	Derived based on ESRI network, Census 2016
		Agglomeration	Population density	Population density (person per square kilometre) of UCL	Derived based on Census 2016
Central place theory	Diversification	Job diversity	Diversity of jobs within a UCL	Derived based on Census 2016	
	High level urban functions	Professional jobs	Share of the UCL labour forces belongs to the professional category	Derived based on Census 2016	
	Centrality	Closeness centrality	The inverse of cumulative shortest path distances from a UCL to all other UCLs	Derived based on AURIN, ABS, OSM data	
	Proximity to major city	Time to the capital city	Drive time (min) from the centre of UCLs to the centre of the nearest capital city	Derived based on ESRI network, Census 2016	
	Central flow theory	Intermediacy	Betweenness centrality	The number of times a UCL falls on the shortest paths between other UCLs	Derived based on AURIN, ABS, OSM data
		Borrowed size	Population in surrounding areas	Spatial lags of population i.e., total population located in neighbouring SA2s of a UCL discounted by driving time.	Derived based on ESRI network, Census 2016 data
		Borrowed function	Professional jobs in surrounding areas	Spatial lags of high-functional jobs i.e. total high-functional jobs located in neighbouring SA2s of a UCL discounted by driving time.	Derived based on ESRI network, Census 2016 data
	Adjustment for endogeneity (instrument variables)	Reverse causality	Spatial lag of job accessibility	Average of the job accessibility level of eight neighbouring UCLs of a UCL	Derived from job accessibility
			Spatial lag of urban diversity	Average of the urban diversity level of eight neighbouring UCLs of a UCL	Derived from urban diversity



**Fig. 6 | Travel time decay curve for commuting flows between SA2s.** This figure shows how the likelihood of commuting between areas declines with increasing travel time. Data on car-based commuting flows between Statistical Area Level 2 (SA2) regions were sourced from the Australian Bureau of Statistics and analysed using drive-time estimates between SA2 centroids. Commute durations were grouped into one-

minute intervals up to 149 min. **a** shows the standardised probability distribution of trips in cities. **b** shows the inverse cumulative distribution of trips across the same groups. A negative exponential function was found to best represent the pattern of trip frequency decline, with a decay rate of 0.058. This function, commonly used in transport studies, reflects how people are less likely to commute as travel time increases.

transport network. This derivation was conducted in stages, separating SA2s for different states and territories, so that the maximum number of input features does not exceed the limit imposed by the software. The number of jobs available within the SA2s was derived based on the total number of commuting flows destined there from all SA2s.

$$A_i = \sum_{j=1}^n O_j e^{-\beta t_{ij}} \tag{11}$$

where,  $A_i$  = Jobs accessibility of  $UCL_i$ ,  $O_j$  = Number of jobs available in  $SA2_j$ ,  $\beta$  = estimated decay parameter,  $t_{ij}$  = driving time (min) between  $UCL_i$  and  $SA2_j$ .

**Job diversity:** The urban diversity was calculated following the formula of Simpson’s diversity index (Eq.12). The study used occupation data from the 2016 ABS census for the 655 UCLs and calculated the percentage of people employed in the eight occupational status categories (Managers, Professionals, Technicians and Trades Workers, Community and Personal Service Workers, Clerical and Administrative Workers, Sales Workers, Machinery Operators and Drivers, and Labourers). The diversity index ranges between 0 (homogenous) and 1 (fully diverse) for the UCLs.

$$Diversity = 1 - \left( \frac{\sum n(n-1)}{N(N-1)} \right) \tag{12}$$

where,  $n$  = % of employment in a particular category;  $N$  = % of employment in all categories

**Centrality and intermediacy:** Two centrality measures were employed based on spatial network analysis in this study: betweenness centrality to represent intermediacy and closeness centrality to measure how close a city is located to all other cities. Unlike topological/social network analysis, where the distance between two contacts has little meaning, spatial network analysis takes into account the geographic distance between two contacts to derive centrality indicators. This is particularly important in the case of cities because unlike social contacts in which not everyone is connected with everyone else, all cities are to some extent connected with each other, and as a result, all cities will result in an identical score for some centrality measures if the distances between them are not taken into account. Different types of flows (e.g., passenger, telephone calls) between cities are often used as weights in centrality measures<sup>80</sup>. However, this study has not applied such

weights given that flow data cannot be known for a new city. The two centrality indicators were derived using the Urban Network Analysis Toolbox 1.01 in ArcMap (v.10.8) (<http://cityform.mit.edu/projects/urban-network-analysis>). The tool required a network dataset and building points/polygons. The OSM road network for Australia was downloaded from AURIN and used to derive a network dataset with travel distance as an impedance. Given the national level analytical focus of this study, local and minor roads were excluded, keeping motorways, primary, trunk, secondary, and tertiary roads and giving a hierarchy score of 1, 2, 3, 4, and 5, respectively. For similar reasons, the resulting road networks were simplified through collapsing dual roads and thinning operations. The 655 UCL centroids were used as inputs for ‘Buildings’ in the tool to derive the indicators.

**Borrowed size and borrowed function:** The same distance decay parameter was used to calculate the borrowed size and borrowed function indicators (Table 5). A spatial lag formulation of the UCL boundary was applied, which ensured that the population or high-functional jobs that were located within the boundary were excluded from analyses for the derivation of these two indicators. For this reason, a large UCL receives a low borrowed size score compared to a smaller UCL located next a larger UCL. For example, Melbourne as a UCL had a borrowed population size of 29 K (distance weighted); whereas Lara, a small UCL located next to Melbourne, had a borrowed population size of 250 K. Also, the application of the distance decay function ensured that the population/high-functional jobs that were located in nearby SA2s received a higher weight than those located in distant SA2s.

**Operationalising the theoretical city size model**

All indicators were nature log-transformed to be consistent with the functional form derived in Eq.10, except for the Capital City indicator, which was dummy-coded. The population sizes of the UCLs were regressed on the explanatory factors listed in Table 5 in a linear regression model – i.e., to estimate how much of the observed variation in city sizes can be explained by the identified theoretical factors. A four-step process was employed to estimate the model:

- a. Correlation analysis: A correlation analysis was conducted among the explanatory factors, and variables with strong correlations (>0.75) were excluded. A strong correlation was observed between capital city and

**Table 6 | Correlations among the factors\* of city sizes (≥10 K Sample)**

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Job accessibility	1.00												
2. Population density	0.62	1.00											
3. Borrowed function	0.57	0.44	1.00										
4. Borrowed size	0.56	0.46	<b>0.99</b>	1.00									
5. % professionals	0.35	0.22	0.01	0.00	1.00								
6. Time to train station	-0.33	-0.41	-0.31	-0.34	-0.11	1.00							
7. Point of interests	0.06	0.11	0.03	0.06	0.55	-0.09	1.00						
8. Diversity	-0.32	0.07	-0.02	0.02	-0.66	-0.08	-0.32	1.00					
9. Time to capital city	<b>-0.78</b>	-0.53	-0.40	-0.38	-0.34	0.31	-0.16	0.34	1.00				
10. Betweenness centrality	-0.11	0.00	0.13	0.14	-0.08	-0.14	0.04	0.03	-0.03	1.00			
11. Closeness centrality	0.18	0.32	0.56	0.58	0.11	-0.21	0.16	0.03	-0.16	0.30	1.00		
12. Rainfall	0.28	0.31	0.37	0.39	0.21	-0.16	0.19	-0.02	-0.13	-0.12	0.21	1.00	
13. Capital City	0.56	0.35	-0.02	-0.02	0.37	-0.13	0.14	-0.32	<b>-0.84</b>	-0.05	-0.05	0.07	1.00

\*All factors are in natural logarithmic scale, except for Capital City, which is represented as a binary variable. Strong correlation coefficients are highlighted in bold.

travel time to capital city factors, and between borrowed size and borrowed function factors (Table 6). Consequently, the capital city and borrowed size factors were retained due to their better explanatory power of theoretical city size;

- b. Factor identification: The statistical significance of the remaining factors was tested in a simple (univariable) linear regression model and any factors with insignificant (at the 0.1 level) statistical association with theoretical city size were removed from further analysis;
- c. Multicollinearity analysis: The remaining factors were then entered into a multiple linear regression model to check multicollinearity among the factors based on the variance inflation factor (VIF) and variables with the highest VIF values were gradually removed until all the variables in the model achieve a VIF value of ≤5;
- d. Model refinement: A parsimonious model of theoretical city size was estimated by gradually removing insignificant factors (p > 0.05).

**Addressing endogeneity in explanatory variables**

The above steps were separately estimated for the two sample groups of UCLs (≥ 10 K, and ≥4 K). The significant factors were evaluated for potential reverse causality based on theoretical reasoning, revealing that two of them – job accessibility and urban diversity - were susceptible to endogeneity bias. Job accessibility serves as an indicator that captures the planning outcomes of investments in both transport and land uses, particularly the job opportunities<sup>57</sup>. Consequently, such investments tend to be made in areas with high population concentration to maximise the benefits<sup>81</sup>. Similarly, urban diversity, represented by the variety of job options available, plays a crucial role in attracting individuals to reside and work in cities, thereby fostering meaningful employment opportunities<sup>58</sup>. Contrary to this argument is that large cities inherently possess a diverse knowledge base, which in turn attracts firms seeking to capitalise on this wealth of expertise through diversifying their operations. Urban planning also responds to this diversity by provisioning alternative goods and services that cater to the needs of the population<sup>82</sup>.

This study applied the instrument variable approach to isolate the reverse causality bias, operationalised within a two-stage least square (2SLS) regression framework – a method widely utilised in the field<sup>36,83</sup>. The spatial lags of job accessibility and urban diversity factors were used as instruments for their respective variables (Table 1). The selection of these instruments was guided by Tobler’s first law of geography<sup>84, p.236</sup>: “everything is related to everything, but near things are more related than distant things”. This implies that cities surrounded by others with higher levels of job accessibility and urban diversity are likely to exhibit higher levels themselves. However, it is unlikely that population size in one city will directly determine job accessibility and urban diversity levels in neighbouring cities. Using spatial

lag as an instrument thus captures spillover effects by leveraging the influence that neighbouring areas’ characteristics exert on a UCL’s own conditions through spatial diffusion or policy interdependence. All models were estimated using Stata software (v.16).

**Measuring deviations from theoretical size**

The parsimonious model, as developed in Section 3.1, was run and the theoretical size of UCLs (unstandardised predicted values) and residuals (unstandardised) were saved in a database. The residuals denote the differences between actual population sizes (natural log transformed) and theoretical (natural log transformed) population sizes of the UCLs. The residuals were transformed as the percentage of actual population to represent percentage differences between the actual and theoretical population sizes.

**Assessing sustainability outcomes of size deviation**

The third research question essentially tests the validity of the inherent hypothesis that cities in the equilibrium of theoretical and actual population sizes would produce more sustainable outcomes for cities. If this hypothesis is to be true, then an optimum sustainability outcome for cities would be achieved when the differences between actual and theoretical population sizes are zero. Empirically, this is operationalised using piecewise/segmented linear regression models. This type of model has been widely applied to test the effects of different policy interventions<sup>85</sup>, including in travel behaviour research<sup>86</sup>. Mathematically, the model can be expressed as:

$$y = \begin{cases} \beta_0 + \beta_1 x + \gamma Z + \epsilon, & x < \alpha \\ \beta_0 + \beta_1 x + \beta_2(x - \alpha) + \gamma Z + \epsilon, & x \geq \alpha \end{cases} \quad (13)$$

where, y = sustainability outcomes of UCLs, x = deviations (% difference between actual and theoretical population sizes), α = breakpoint to be estimated, β<sub>0</sub> = constant, β<sub>1</sub> and β<sub>2</sub> = model coefficients before and after the breakpoint, γ is a vector of coefficients corresponding to covariates Z, and ε = error term. The effect of Z is assumed to be the same on both sides of the breakpoint.

The piecewise regression models were estimated in Stata (v.16). The breakpoint was determined using Stata’s nl (nonlinear regression) command (<https://stats.oarc.ucla.edu/stata/faq/how-can-i-find-where-to-split-a-piecewise-regression/>). To facilitate interpretation, the deviation variable was centred at the estimated breakpoint, allowing the constant term to represent the predicted sustainability outcome at the sustainable city size.

Three types of sustainability outcomes of cities were tested in this study: economic, socio-environmental, and environmental. The economic outcome

is measured by the weekly median rent paid by households. Affordable rent is considered to be directly related to urban productivity as it enables concentration of workers closer to their jobs and an important precondition for digital start-ups<sup>87</sup>. The socio-environmental outcome is measured by the percent of households having two or more vehicles. High rates of multi-vehicle ownership often reflect a lack of accessible, affordable, and inclusive transport options, which can increase household financial vulnerability and social exclusion<sup>88,89</sup>. Moreover, two or more vehicle ownership is associated with increased private vehicle travel and emissions, making it a negative indicator of both social and environmental sustainability<sup>90</sup>. Environmental sustainability is further assessed by the percentage of people (aged 15 and over) who walked to work. Walking is associated with a range of environmental benefits including air quality, reduced noise and reduced carbon emissions through a lower use of car-based fossil-fuel. Studies have shown that if a one percent car trip is substituted by active travel, it would reduce fuel consumption by 2–4%<sup>91</sup>. All three datasets were downloaded from the ABS Census 2016.

The piecewise regression models were estimated twice: first, without any additional covariates; and second, adding two covariates— population density and median household income. The intent was to check the robustness of the sustainable city size effects on sustainability outcomes. These covariates were selected given their established relevance to the sustainability outcomes under consideration: population density has been consistently shown to influence housing costs through agglomeration economies<sup>92</sup>, to reduce vehicle ownership rates by decreasing car dependency<sup>90,93</sup>, and to promote walking behaviour through more compact, walkable environments<sup>94</sup>. Similarly, median household income affects rental affordability<sup>95</sup>, multi-vehicle ownership<sup>96</sup>, and walking participation, as higher-income groups often have more transport choices but are less dependent on walking unless supported by a high-quality urban environment<sup>97</sup>. By including these covariates, the analysis provides a more conservative test of the hypothesis, ensuring that the observed effects of size deviations are not confounded by baseline density or income characteristics.

## Data availability

The sources of data used in this study are listed in Table 5 and are publicly available for download. Derived datasets generated and analysed during the current study are available from the corresponding author upon reasonable request.

## Abbreviations

ABS	Australian Bureau of Statistics
AURIN	Australian Urban Research Infrastructure Network
BoM	Bureau of Meteorology
GIS	Geographic Information System
ICDF	Inverse Cumulative Density Function
OD	Origin–Destination
OSM	OpenStreetMap
PDF	Probability Density Function
SA1	Statistical Area Level 1
SA2	Statistical Area Level 2
UCL	Urban Centres and Localities
VIF	Variance Inflation Factor
2SLS	Two-Stage Least Squares

Received: 16 January 2025; Accepted: 22 July 2025;  
Published online: 13 August 2025

## References

- Moser, S. & Côté-Roy, L. New cities: Power, profit, and prestige. *Geogr. Compass* **15**, e12549 (2021).
- UN HABITAT. World Cities Report 2022: Envisaging the Future of Cities. ([https://unhabitat.org/sites/default/files/2022/06/wcr\\_2022.pdf](https://unhabitat.org/sites/default/files/2022/06/wcr_2022.pdf), 2022).
- Watson, V. African urban fantasies: dreams or nightmares?. *Environ. Urban.* **26**, 215–231 (2014).
- Datta, A. New urban utopias of postcolonial India: ‘Entrepreneurial urbanization’ in Dholera smart city. *Gujarat. Dialogues Hum. Geogr.* **5**, 3–22 (2015).
- Myers, D. Demographic futures as a guide to planning: California’s Latinos and the compact city. *J. Am. Plan. Assoc.* **67**, 383–397 (2001).
- Berke, P. R., Godschalk, D. R., Kaiser, E. J. & Rodriguez, D. A. *Urban land use planning*. (University of Illinois Press Urbana, 2006).
- Wang, X. & Hofe, R. *Research methods in urban and regional planning*. (Springer Science & Business Media, 2008).
- Bettencourt, L. M. A. The origins of scaling in cities. *Science* **340**, 1438–1441 (2013).
- Turok, I. & McGranahan, G. Urbanization and economic growth: the arguments and evidence for Africa and Asia. *Environ. urbanization.* **25**, 465–482 (2013).
- Mendez, P., Atienza, M. & Modrego, F. Urbanization and productivity at a global level: new empirical evidence for the services sector. *Reg. Sci. Policy Pract.* **15**, 1981–1998 (2023).
- Frick, S. A. & Rodríguez-Pose, A. Big or Small Cities? On city size and economic growth. *Growth Change* **49**, 4–32 (2018).
- Glaeser, E. L. & Sacerdote, B. Why is there more crime in cities?. *J. Polit. Econ.* **107**, S225–S258 (1999).
- Mohajeri, N., Gudmundsson, A. & French, J. R. CO<sub>2</sub> emissions in relation to street-network configuration and city size. *Transport. Res. Part D: Transp. Environ.* **35**, 116–129 (2015).
- Zhou, B., Rybski, D. & Kropp, J. P. The role of city size and urban form in the surface urban heat island. *Sci. Rep.* **7**, 4791 (2017).
- Czinkan, N. & Horváth, Á. Determinants of housing prices from an urban economic point of view: evidence from Hungary. *J. Eur. Real. Estate Res.* **12**, 2–31 (2019).
- Krupka, D. J. Are big cities more segregated? Neighbourhood scale and the measurement of segregation. *Urban Stud.* **44**, 187–197 (2007).
- Tavares, A. F. & Carr, J. B. So close, yet so far away? The effects of city size, density and growth on local civic participation. *J. Urban Aff.* **35**, 283–302 (2013).
- Buch, T., Hamann, S., Niebuhr, A. & Rossen, A. What makes cities attractive? The determinants of urban labour migration in Germany. *Urban Stud.* **51**, 1960–1978 (2013).
- Moon, M. J. & Norris, D. F. Does managerial orientation matter? The adoption of reinventing government and e-government at the municipal level. *Inf. Syst. J.* **15**, 43–60 (2005).
- Chen, J., Davis, D. S., Wu, K. & Dai, H. Life satisfaction in urbanizing China: The effect of city size and pathways to urban residency. *Cities* **49**, 88–97 (2015).
- Li, L. et al. Impacts of city size change and industrial structure change on CO<sub>2</sub> emissions in Chinese cities. *J. Clean. Prod.* **195**, 831–838 (2018).
- Levinson, D. Network structure and city size. *PLoS ONE* **7**, e29721 (2012).
- Engelfriet, L. & Koomen, E. The impact of urban form on commuting in large Chinese cities. *Transportation* **45**, 1269–1295 (2018).
- An, Q., Gordon, P. & Moore, J. E. A note on commuting times and city size: Testing variances as well as means. *J. Transp. Land Use* **7**, 105–110 (2014).
- Horner, M. W. Extensions to the concept of excess commuting. *Environ. Plan. A* **34**, 543–566 (2002).
- Gordon, P., Kumar, A. & Richardson, H. W. Congestion, changing metropolitan structure, and city size in the United States. *Int. Regional Sci. Rev.* **12**, 45–56 (1989).
- Alonso, W. The economics of urban size. *Pap. Reg. Sci. Assoc.* **26**, 66–83 (1971).
- Shen, D., Thill, J.-C. & Sun, J. Are Chinese cities oversized?. *Int. Reg. Sci. Rev.* **43**, 632–654 (2020).

29. Richardson, H. W. Optimality in city size, systems of cities and urban policy: A sceptic's view. *Urban Stud.* **9**, 29–48 (1972).
30. Jiang, B., Yin, J. & Liu, Q. Zipf's law for all the natural cities around the world. *Int. J. Geogr. Inf. Sci.* **29**, 498–522 (2015).
31. Gabaix, X. Z. ipf's Law for Cities: An explanation. *Q. J. Econ.* **114**, 739–767 (1999).
32. Soo, K. T. Z. ipf's Law for cities: A cross-country investigation. *Reg. Sci. Urban Econ.* **35**, 239–263 (2005).
33. Schmid, C. Specificity and urbanization: A theoretical outlook. In *The Inevitable Specificity Of Cities: Napoli, Nile Valley, Belgrade, Nairobi, Hong Kong, Canary Islands, Beirut, Casablanca* (Eds. Diener R. et al.) 287–307 (Zurich: Lars Müller Publishers, 2015).
34. Camagni, R., Capello, R. & Caragliu, A. One or infinite optimal city sizes? In search of an equilibrium size for cities. *Ann. Reg. Sci.* **51**, 309–341 (2013).
35. González-Val, R. & Olmo, J. A statistical test of city growth: Location, increasing returns and random growth. Munich Personal RePEc Archive (MPRA) Paper No. 27139. <https://mpra.ub.uni-muenchen.de/27139/> (2010) (accessed 01/08/2025).
36. Davis, D. R. & Weinstein, D. E. Bones, bombs, and break points: the geography of economic activity. *Am. economic Rev.* **92**, 1269–1289 (2002).
37. Chen, Z., Fu, S. & Zhang, D. Searching for the parallel growth of cities in China. *Urban Stud.* **50**, 2118–2135 (2013).
38. Ayuda, M. I., Collantes, F. & Pinilla, V. From locational fundamentals to increasing returns: the spatial concentration of population in Spain, 1787–2000. *J. Geographical Syst.* **12**, 25–50 (2010).
39. Gassebner, M., Schaudt, P. & Wong, M. H. L. Competition and political violence. *J. Dev. Econ* **162**, 103052 (2023).
40. Jedwab, R., Johnson, N. D. & Koyama, M. Pandemics and cities: Evidence from the Black Death and the long-run. *J. Urban Econ* **139**, 103628 (2024).
41. Santos, J. L. & Fernández Fernández, M. T. The spread of urban–rural areas and rural depopulation in central Spain. *Reg. Sci. Policy Pract.* **15**, 863–877 (2023).
42. Esteban-Oliver, G. On the right track? Railways and population dynamics in Spain, 1860–1930. *Eur. Rev. Econ. Hist.* **27**, 606–633 (2023).
43. Gómez Valenzuela, V. & Holl, A. Growth and decline in rural Spain: an exploratory analysis. *Eur. Plan. Stud.* **32**, 430–453 (2024).
44. Nygaard, C. A. & Parkinson, S. Analysing the impact of COVID-19 on urban transitions and urban-regional dynamics in Australia. *Aust. J. Agric. Resour. Econ.* **65**, 878–899 (2021).
45. Krugman, P. Urban concentration: the role of increasing returns and transport costs. *Int. Reg. Sci. Rev.* **19**, 5–30 (1996).
46. Ciccone, A. & Hall, R. E. (National Bureau of Economic Research Cambridge, MA, USA, 1993).
47. Perl, A., Deng, T., Correa, L., Wang, D. & Yan, Y. Understanding the urbanization impacts of high-speed rail in China. *Archives of Transp* **58**, 21–34 (2021).
48. Christaller, W. *Die Zentralen Orte in Suddeutschland*. (Gustav Fischer, 1933).
49. Castells, M. The information age: Economy, society and culture (3 volumes). *Blackwell, Oxf.* **1997**, 1998 (1996).
50. Taylor, P. J., Hoyler, M. & Verbruggen, R. External urban relational process: introducing central flow theory to complement central place theory. *Urban Stud. (Edinb., Scotl.)* **47**, 2803–2818 (2010).
51. Li, X. & Neal, Z. P. Are larger cities more central in urban networks: A meta-analysis. *Global Networks* **24**, e12467 (2024).
52. Prieto-Curiel, R., Schumann, A., Heo, I. & Heinrigs, P. Detecting cities with high intermediacy in the African urban network. *Comput. Environ. Urban Syst.* **98**, 101869 (2022).
53. Phelps, N. A., Fallon, R. J. & Williams, C. L. Small firms, borrowed size and the urban-rural shift. *Reg. Stud.* **35**, 613–624 (2001).
54. Meijers, E. J., Burger, M. J. & Hoogerbrugge, M. M. Borrowing size in networks of cities: City size, network connectivity and metropolitan functions in Europe. *Pap. Regional Sci.* **95**, 181–198 (2016).
55. Yang, T., Zhu, Y. & Du, J. Is there a borrowed size in China's urban agglomerations?. *Prog. Geogr.* **41**, 1156–1167 (2022).
56. Meijers, E. From central place to network model: theory and evidence of a paradigm change. *Tijdschr. voor Econ. Soc. Geogr.* **98**, 245–259 (2007).
57. Lee, J., Arts, J. & Vanclay, F. Investigating institutional barriers and opportunities to an integrated approach for transport and spatial development: Mega urban transport development in a rapidly developing city, Seoul. *J. Urban Aff.* **46**, 40–62 (2024).
58. Kelly, M., Nguyen, M. & Triandafyllidou, A. Why migrants stay in small and mid-sized cities: Analytical and comparative insights. *J. Int. Migr. Integ.* **24**, 1013–1027 (2023).
59. Tyvima, T. & Kamruzzaman, M. The effect of young, single person households on apartment prices: an instrument variable approach. *J. Hous. Built Environ.* **34**, 91–109 (2019).
60. Côté-Roy, L. & Moser, S. 'Does Africa not deserve shiny new cities?' The power of seductive rhetoric around new cities in Africa. *Urban Stud.* **56**, 2391–2407 (2019).
61. Van Noorloos, F. & Kloosterboer, M. Africa's new cities: The contested future of urbanisation. *Urban Stud.* **55**, 1223–1241 (2018).
62. Herbert, C. W. & Murray, M. J. Building from scratch: New cities, privatized urbanism and the spatial restructuring of Johannesburg after apartheid. *Int. J. Urban Reg. Res.* **39**, 471–494 (2015).
63. Berghmans, L., Schoovaerts, P. & Teghem, J. Jr Implementation of health facilities in a new city. *J. Oper. Res. Soc.* **35**, 1047–1054 (1984).
64. Van Leynseele, Y. & Bontje, M. Visionary cities or spaces of uncertainty? Satellite cities and new towns in emerging economies. *Int. Plan. Stud.* **24**, 207–217 (2019).
65. Hashim, A. R. A. A. B. Branding the brand new city: Abu Dhabi, travelers welcome. *Place Brand. Public Dipl.* **8**, 72–82 (2012).
66. Faria, J. R., Ogura, L. M. & Sachsida, A. Crime in a planned city: The case of Brasilia. *Cities* **32**, 80–87 (2013).
67. Jones, M. A. Optimum City Size. *Aust. N.Z. J. Sociol.* **9**, 32–36 (1973).
68. Li, S., Liu, K. & Li, J. Reflections and Enlightenments on the Population Size Determination in Urban Master Planning in China: A Case Study of Xi'an City. *Cheng Shi Gui Hua* **31**, 22–32 (2022).
69. Parham, E., Law, S. & Versluis, L. in *Proceedings-11th International Space Syntax Symposium, SSS 2017*. 103.101–103.117 (SSS).
70. Arribas-Bel, D., Garcia-López, M. À & Viladecans-Marsal, E. Building(s and) cities: Delineating urban areas with a machine learning algorithm. *J. Urban Econ.* **125**, 103217 (2021).
71. Dijkstra, L., Poelman, H. & Veneri, P. The EU-OECD definition of a functional urban area. OECD Regional Development Working Papers 2019/11. <https://doi.org/10.1787/d58cb34d-en> (2019).
72. OECD/European Commission. Cities in the World: A new perspective on urbanisation. OECD Urban Studies, OECD Publishing, Paris, <https://doi.org/10.1787/d0efcbda-en> (2020).
73. Moreno-Monroy, A. I., Schiavina, M. & Veneri, P. Metropolitan areas in the world. *Delineation Popul. Trends J. Urban Econ.* **125**, 103242 (2021).
74. Arcaute, E. et al. Constructing cities, deconstructing scaling laws. *J. R. Soc. Interface* **12**, 20140745 (2015).
75. Bosker, M., Park, J. & Roberts, M. Definition matters. Metropolitan areas and agglomeration economies in a large-developing country. *J. Urban Econ.* **125**, 103275 (2021).
76. Australian Bureau of Statistics (ABS). *Urban Centres and Localities*, ABS Website, accessed 1 August 2025. <https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/significanturban-areas-urban-centres-and-localities-section-state/urban-centres-and-localities>. (2022).

77. Australian Government. Shaping a nation: Population growth and immigration over time. <https://population.gov.au/publications/publications-shaping-nation> (Canberra, 2018) (accessed 01/08/2025).
78. Santana Palacios, M. & El-Geneidy, A. Cumulative versus gravity-based accessibility measures: Which one touse? <https://doi.org/10.32866/001c.32444> (2022).
79. Halás, M., Klapka, P. & Kladiivo, P. Distance-decay functions for daily travel-to-work flows. *J. Transp. Geogr.* **35**, 107–119 (2014).
80. Zhang, F., Ning, Y. & Lou, X. The evolutionary mechanism of China's urban network from 1997 to 2015: An analysis of air passenger flows. *Cities* **109**, 103005 (2021).
81. Lee, J. K. Transport infrastructure investment, accessibility change and firm productivity: Evidence from the Seoul region. *J. Transp. Geogr.* **96**, 103182 (2021).
82. He, X. Energy effect of urban diversity: An empirical study from a land-use perspective. *Energy Econ* **108**, 105892 (2022).
83. Bleakley, H. & Lin, J. Portage and path dependence. *Q. J. Econ.* **127**, 587–644 (2012).
84. Tobler, W. R. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **46**, 234–240 (1970).
85. Vergori, A. S. & Arima, S. Low-cost carriers and tourism in the Italian regions: A segmented regression model. *Ann. Tour. Res* **97**, 103474 (2022).
86. Haider, M. Diminishing Returns To Density And Public Transit. <https://doi.org/10.32866/10679> (2019).
87. Gurrán, N. et al. Urban productivity and affordable rental housing supply in Australian cities and regions. (AHURI Final Report No. 353). <https://www.ahuri.edu.au/research/final-reports/353>. <https://doi.org/10.18408/ahuri7320001> (Australian Housing and Urban Research Institute Limited 2021).
88. Lucas, K. Transport and social exclusion: Where are we now?. *Transp. policy* **20**, 105–113 (2012).
89. Currie, G. & Delbosc, A. Exploring public transport usage trends in an ageing population. *Transportation* **37**, 151–164 (2010).
90. Newman, P. & Kenworthy, J. Sustainability and cities: overcoming automobile dependence. (1998).
91. Litman, T. Evaluating Active Transport Benefits and Costs: Guide to Valuing Walking and Cycling Improvements and Encouragement Programs. (Victoria Transport Policy Institute, <https://www.vtpi.org/nmt-tdm.pdf>, 2023).
92. Glaeser, E. & Gyourko, J. The economic implications of housing supply. *J. Econ. Perspect.* **32**, 3–30 (2018).
93. Cervero, R. & Guerra, E. Urban densities and transit: A multi-dimensional perspective. (2011).
94. Ewing, R. & Cervero, R. Travel and the built environment: A meta-analysis. *J. Am. Plan. Assoc.* **76**, 265–294 (2010).
95. Gan, Q. & Hill, R. J. Measuring housing affordability: Looking beyond the median. *J. Hous. Econ.* **18**, 115–125 (2009).
96. Giuliano, G. & Dargay, J. Car ownership, travel and land use: a comparison of the US and Great Britain. *Transportation Res. Part A: Policy Pract.* **40**, 106–124 (2006).
97. Buehler, R. & Pucher, J. *Walking and Cycling in Western Europe and the United States: trends, policies, and lessons*. T34–4 (TR News, 2012).

### Acknowledgements

The author thanks the two anonymous reviewers and the editor of the journal, Professor Xiaoling Zhang, for their insightful comments and suggestions. This paper forms the foundation for the successful Australian Research Council Discovery Project (DP250101843), which contributed to the development of the research agenda presented herein.

### Author contributions

L.K.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing – Original Draft, and Writing – Review & Editing.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to Liton Md Kamruzzaman.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025