Supplementary Material for:
On the Meaning of Urban Scaling

Ulysse Marquis[1,2*] and Marc Barthelemy[3,4,5†]
[1] *Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy*
[2] *Department of Mathematics, University of Trento, Via Sommarive 14, 38123 Povo (TN), Italy*
[3] *Université Paris-Saclay, CNRS, CEA, Institut de Physique Théorique, 91191, Gif-sur-Yvette, France*
[4] *Centre d'Analyse et de Mathématique Sociales (CNRS/EHESS) Paris, France and*
[5] *Complexity Science Hub, Vienna, Austria*

I. METHODS

A. OLS estimate of the transversal exponent

At a given date, we write the transversal relation in log–log form as

$$y_i = \alpha_T + \beta_T x_i + \varepsilon_i, \tag{1}$$

with $x_i = \ln P_i$ and $y_i = \ln Y_i$. The OLS estimate minimizes the total deviation given by

$$S(\alpha_T, \beta_T) = \sum_{i=1}^{N} (y_i - \alpha_T - \beta_T x_i)^2. \tag{2}$$

The minimization equations $\partial S/\partial\alpha_T = \partial S/\partial\beta_T = 0$ imply

$$\alpha_T = \bar{y} - \beta_T \bar{x} \tag{3}$$

and

$$\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) = \beta_T \sum_{i=1}^{N} (x_i - \bar{x})^2. \tag{4}$$

Hence

$$\beta_T = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}. \tag{5}$$

Thus, the transversal exponent estimated by OLS is the ratio between the cross-sectional covariance of $x = \ln P$ and $y = \ln Y$, and the cross-sectional variance of x.

B. Goodness of fit

In addition to the expression of the transversal exponent, the quality of the fit can also be written in a simple form. For the OLS regression

$$y_i = \alpha + \beta x_i + \varepsilon_i, \tag{6}$$

the coefficient of determination is given by

$$R^2 = \frac{\mathrm{Cov}(x, y)^2}{\mathrm{Var}(x)\,\mathrm{Var}(y)} = \rho_{xy}^2, \tag{7}$$

*Electronic address: ulyssepierre.marquis@unitn.it
†Electronic address: marc.barthelemy@ipht.fr

where ρ_{xy} is the Pearson correlation coefficient computed over the cross-section. Using the expression $\beta_T = \mathrm{Cov}(x, y)/\mathrm{Var}(x)$, this can also be written as

$$R^2 = \beta_T^2 \frac{\mathrm{Var}(x)}{\mathrm{Var}(y)}. \tag{8}$$

This relation shows that, while the exponent β is determined by the covariance between x and y, the goodness of fit depends on how this covariance compares to the total variance of y. In particular, a well-defined exponent can coexist with a low value of R^2 if the variance of y is large. In the context of transversal scaling, this highlights that the existence of a scaling exponent does not necessarily imply a tight functional relationship, but rather reflects the underlying covariance structure of the ensemble of cities.

C. Two slopes case

We consider here the case when the individual cities have a piecewise linear behaviour with slope a_1 for $x < x^*$ and slope a_2 for $x > x^*$. We will compute the transversal exponent using the general OLS equation discussed in the text

$$\beta_T(t) = \frac{\mathrm{Cov}(x, \alpha)}{\mathrm{Var}(x)} + \frac{\mathrm{Cov}(x, \beta x)}{\mathrm{Var}(x)} \tag{9}$$

We consider that the logarithm x of the population is distributed according to distribution $P(x)$. The area can thus be written in this case as

$$A = \begin{cases} a_1 P & P < P^* \\ a_2 P + b & P > P^* \end{cases} \tag{10}$$

where $b = (a_1 - a_2)P^*$ ensures the continuity of A. We will assume that we are interested in regimes where P is large enough so that we can neglect b. In all cases we have an elasticity (local slope) equal to 1, possibly with $1/P$ corrections. We write $\alpha_i = \ln a_i$ and we can write for city i

$$\alpha_i = \alpha_1 + (\alpha_2 - \alpha_1)\,\theta(x_i - x_i^*) \tag{11}$$

where x_i^* corresponds to P_i^*, the slope-changing population for city i ($\theta(x)$ is the Heaviside function). We assume that this threshold is distributed according to $\rho(x^*)$, independently of the current population x.

The elasticity $\beta_i = 1$ (with possible $1/P$ corrections) and we therefore have $\mathrm{Cov}(x, \beta x)/\mathrm{Var}(x) = 1$. We therefore get

$$\beta_T(t) = 1 + \frac{\mathrm{Cov}(x, \alpha)}{\mathrm{Var}(x)} \tag{12}$$

The covariance can be written as

$$\begin{aligned} \mathrm{Cov}(x, \alpha) &= \langle x\alpha\rangle - \langle x\rangle\langle\alpha\rangle \\ &= (\alpha_2 - \alpha_1) \int dx^*\, \rho(x^*) \int_{x^*}^{\infty} dx\, P(x)\,(x - \langle x\rangle) \end{aligned} \tag{13}$$

We denote by

$$\mathcal{I} = \int dx^*\, \rho(x^*) \int_{x^*}^{\infty} dx\, P(x)\,(x - \langle x\rangle) \tag{14}$$

and it is easy to show that $\mathcal{I} \geq 0$. Indeed, the inner integral $\int_{x^*}^{\infty} dx\, P(x)\,(x - \langle x\rangle)$ is non-negative for all $x^* > 0$: since $\int_0^{\infty} dx\, P(x)\,(x - \langle x\rangle) = 0$ by definition of the mean, it equals $-\int_0^{x^*} dx\, P(x)\,(x - \langle x\rangle)$, which is non-negative whether x^* is below or above $\langle x\rangle$. Since $\rho(x^*) \geq 0$, the outer integral $\mathcal{I}$ is non-negative as well.

The equation for β_T thus gives us

$$\beta_T = 1 + C(\alpha_2 - \alpha_1) \tag{15}$$

where $C = \mathcal{I}/\mathrm{Var}(x) > 0$. We thus obtain the result discussed in the text: when $a_2 > a_1$ (i.e. $\alpha_2 > \alpha_1$) we have $\beta_T > 1$—a superlinear behaviour, and conversely.

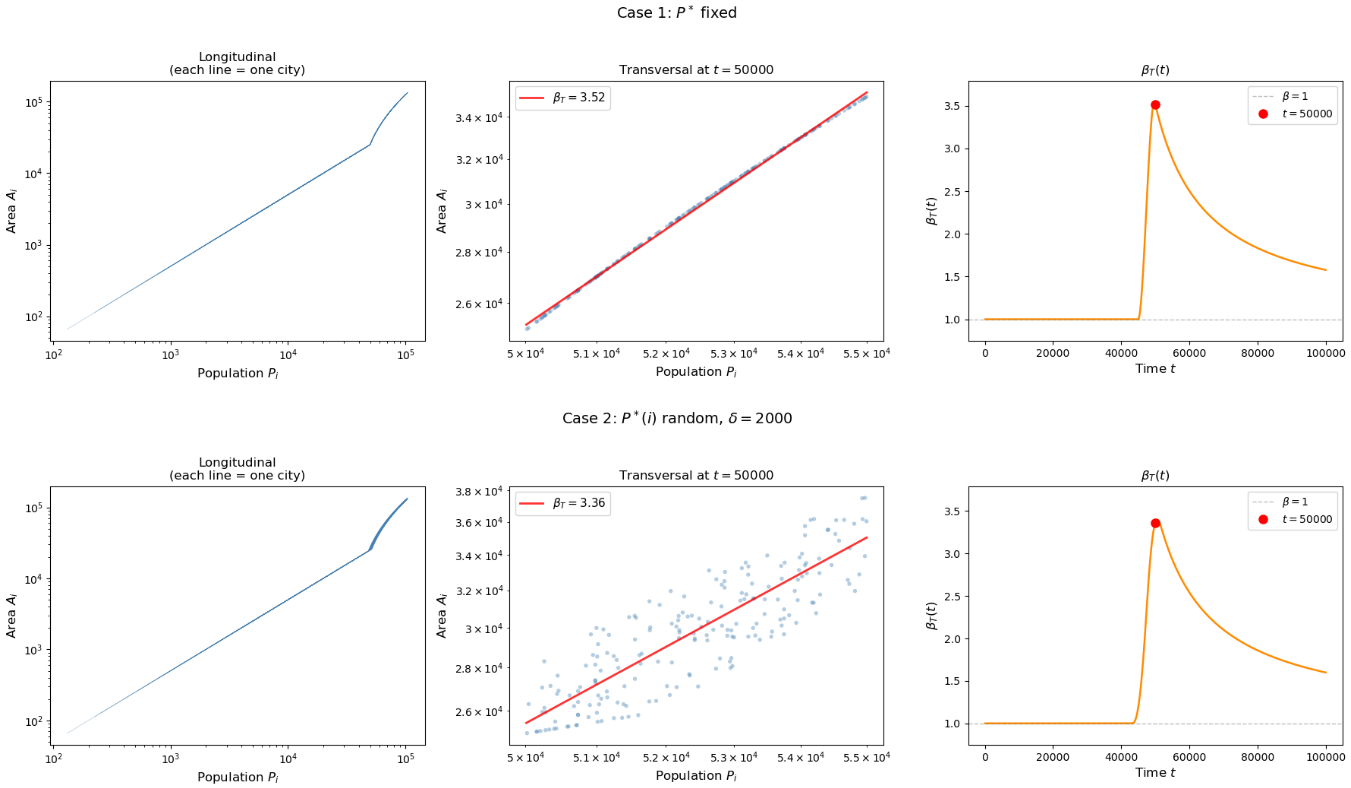

FIG. S1: Left: longitudinal trajectories of individual cities (A_i vs. P_i as t increases). Middle: transversal snapshot at $t = 50{,}000$ with OLS fit (red line). Right: $\beta_T(t)$; the red dot marks the snapshot time. Top: homogeneous case ($P_i^* = P_{\max}/2$ for all cities); $\beta_T \simeq 1$ except during a narrow transition. Bottom: heterogeneous case (P_i^* random with half-width $\delta = 2000$); β_T departs broadly from the true elasticity $\beta = 1$.

II. SIMULATION: THE ROLE OF DYNAMICS HETEROGENEITY

We perform here a simple simulation in order to show the importance of heterogeneity in the dynamics of cities and how it impacts the value of the transversal exponent, even when the local dynamics are described by the same exponent (here $\beta_i = 1$).

We consider N cities with populations $P_i(t) = P_i^{(0)} + t$, where the initial populations $P_i^{(0)}$ are drawn uniformly at random in the interval $[P_{\rm min}^{(0)}, P_{\rm max}^{(0)}]$. The area of city i is given by $A_i(t) = a_i(t)\, P_i(t)$, where the prefactor switches at a threshold: $a_i = a_1$ if $P_i < P_i^*$ and $a_i = a_2$ otherwise (we also impose the continuity of the area at $P = P^*$ which gives an affine function in the second regime). The local elasticity is $\beta_i = 1$ for every city at all times. A cross-sectional OLS regression of $\ln A_i$ on $x_i = \ln P_i$ at fixed t yields the transversal exponent

$$\beta_T(t) = 1 + \frac{\mathrm{Cov}(\alpha,\, x)}{\mathrm{Var}(x)}\,, \tag{16}$$

where $\alpha_i = \ln a_i(t)$. This expression shows that any deviation from the true elasticity is controlled entirely by the cross-sectional correlation between log-prefactors and log-populations.

When all cities share a common threshold $P_i^* = P^*$, then at most times t either all cities satisfy $P_i(t) < P^*$ or all satisfy $P_i(t) \geq P^*$; the prefactor is uniform across cities, $\mathrm{Cov}(\alpha, x) = 0$, and $\beta_T = 1$ exactly (Fig. S1, top row). However, because the initial populations $P_i^{(0)}$ are heterogeneous (spread over the interval $[P_{\rm min}^{(0)}, P_{\rm max}^{(0)}]$), cities do not all cross P^* at the same time: larger cities reach the threshold first. During the narrow transition window of width $\Delta t = P_{\rm max}^{(0)} - P_{\rm min}^{(0)}$, α_i is a step function of x_i and β_T spikes sharply. If all cities started with the same initial population $P_i^{(0)} = P_0$, they would cross P^* simultaneously and β_T would remain equal to 1 at all times. Even in this simple case, heterogeneity—here in initial conditions—is responsible for the departure from the true exponent.

We obtain similar results when each city draws its own threshold $P_i^* \sim \mathcal{U}(P^* - \delta,\, P^* + \delta)$. Cities with larger P_i are more likely to have crossed their individual P_i^*, inducing a positive $\mathrm{Cov}(\alpha, x)$ and pushing β_T above 1 during the

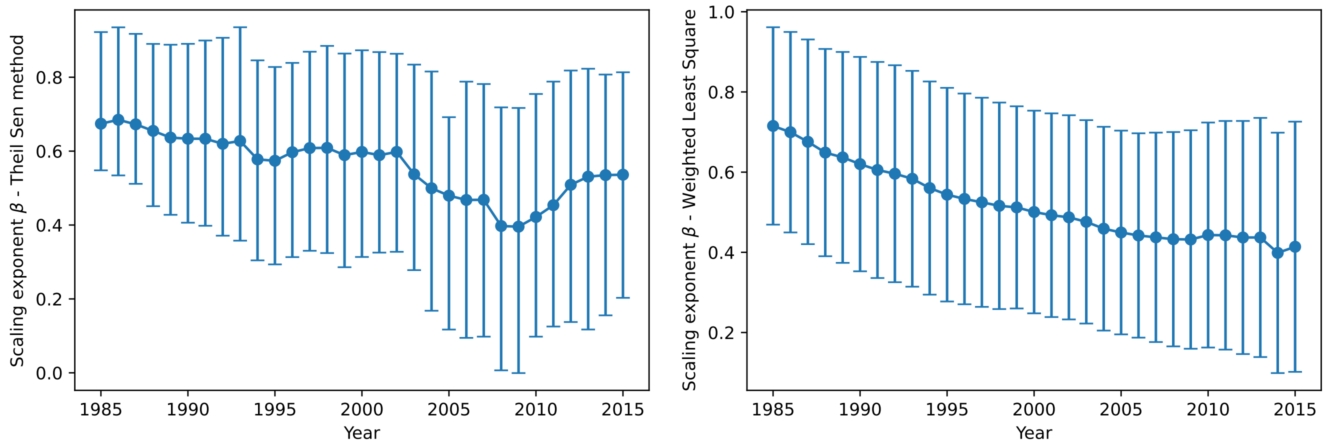

FIG. S2: Comparison of transversal exponent estimates using two robust fitting methods. Left panel: the Theil–Sen regression computes the slope of a dataset as the median of all pairwise slopes between points, which reduces its sensitivity to outliers. Right panel: Weighted Least Squares (WLS), which assigns weights to account for heteroskedasticity, giving more reliable estimates when variance varies across data points.

transition period (Fig. S1, bottom row).

The message resulting from this simple simulation is clear: β_T recovers the true elasticity $\beta_i = 1$ only when all cities follow identical dynamics *and* share the same initial conditions. Any source of heterogeneity—whether in the initial populations, in the transition thresholds, or more generally in the dynamical parameters—can induce cross-sectional correlations between α_i and x_i, causing β_T to depart from the (true) longitudinal exponent. The decomposition (16) serves as a diagnostic: a nonzero covariance term signals that the measured transversal exponent is contaminated by heterogeneous dynamics rather than reflecting a genuine scaling law.

III. CHANGING FITTING PROCEDURE

While the Ordinary Least Squares (OLS) method in log-log space provides an analytical expression for the transversal exponent, it also has certain limitations: it is sensitive to outliers and relies on the assumption of homoskedasticity. Alternative approaches are more robust against these issues. For instance, the Theil–Sen regression is a linear model that is resilient to extreme values, whereas weighted least squares (WLS) account for heteroskedasticity by assigning different weights to data points. Using the area–population data as an example, Fig. S2 illustrates the exponents and associated uncertainty ranges estimated with these methods.

While the exact exponents value differ, the results are qualitatively similar to those found using the OLS – both in terms of exponent values and confidence interval widths.

IV. EMPIRICAL RESULTS

A. The fundamental allometry

In [1], Batty and Ferguson compile estimates of the transversal exponent β_T for the area–population relation reported across a range of studies. These results are summarized in Fig. S3, where the horizontal axis indicates the reference date of each dataset (i.e., the time at which the cross-sectional measurements are performed).

Several features are immediately apparent. First, the reported values of β_T display substantial variability across datasets. While most estimates satisfy $\beta_T < 1$, indicating sublinear scaling, a significant number of cases exhibit $\beta_T > 1$. Overall, no clear universal value or systematic trend emerges from these measurements.

These results do not point to a universal exponent, but rather highlight the strong variability of transversal estimates, consistent with their sensitivity to heterogeneity and correlations across cities.

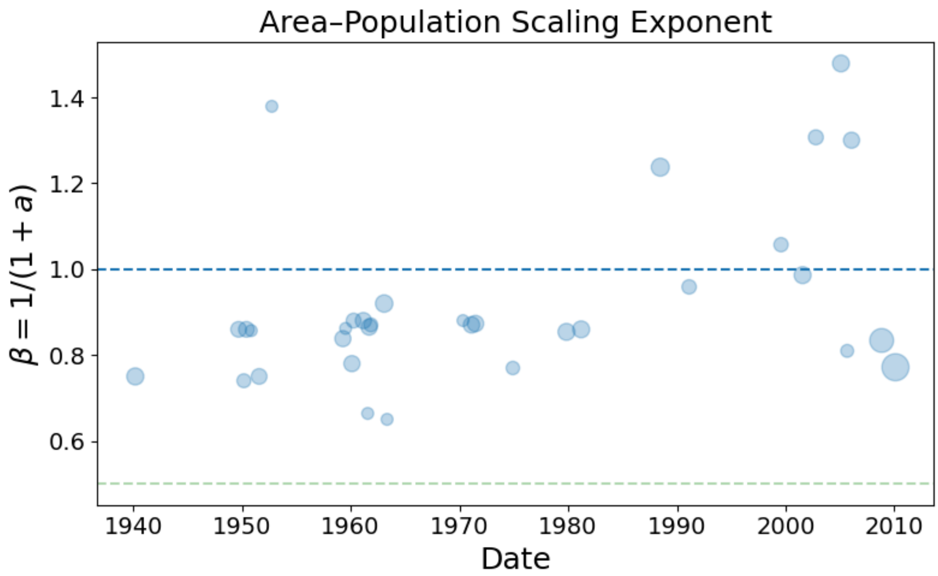

FIG. S3: Compilation of transversal exponents β_T for the area–population relation across different datasets (from [1]). The horizontal axis indicates the reference date of each dataset (the size of symbols refer to the number of cities in the dataset).

B. Wages: longitudinal exponent

We show in Fig. S4 the distribution of the individual longitudinal exponents β_i, obtained by fitting a power-law relation for each city. The distribution exhibits a broad dispersion around its mean value, revealing a significant heterogeneity in the longitudinal dynamics across cities.

For comparison, we also indicate on the same figure the mean and the theoretical value of the transversal exponent β_T (vertical line). Strikingly, this theoretical value does not correspond to any typical individual behavior, but rather lies within a wide distribution of exponents, further emphasizing that the transversal exponent does not reflect the dynamics of a representative city.

C. Congestion delay

We consider here the case of congestion-induced delays, as analyzed in [2], using a dataset of 101 US cities over the period 1982–2014 (for details on the dataset, see [3]). This dataset is particularly relevant, as it is known to exhibit strong heterogeneity in the temporal evolution of individual cities, together with a robust superlinear scaling at the cross-sectional level.

In Fig. S5(inset), we compare the transversal exponent β_T directly measured from the data with the value predicted by

$$\beta_T = \langle\beta\rangle + \frac{\mathrm{Cov}(x,\alpha)}{\mathrm{Var}(x)} + \frac{\mathrm{Cov}(x,(\beta-\langle\beta\rangle)x)}{\mathrm{Var}(x)}\,. \tag{17}$$

As for wages, we observe an excellent agreement between the two, confirming that the covariance-based decomposition provides an accurate description of the origin of the transversal scaling.

In Fig. S5, we display the contribution of the different terms entering Eq. (17). We observe a behavior similar to that found for the area–population relation. In particular, the average longitudinal exponent $\langle\beta\rangle$ differs significantly from the measured transversal exponent β_T, which varies only weakly in the range $[1.31, 1.39]$.

This discrepancy is explained by the covariance terms, whose contributions are large but partially compensate each other, resulting in a relatively small net correction. In this case, the combined effect of these terms is negative, leading to a transversal exponent that is smaller than the average longitudinal exponent and remains remarkably stable over time.

At the level of individual cities, the dynamics of congestion delays are highly heterogeneous, as already emphasized in [2], with large temporal fluctuations and, in some cases, very large effective exponents when fitted independently.

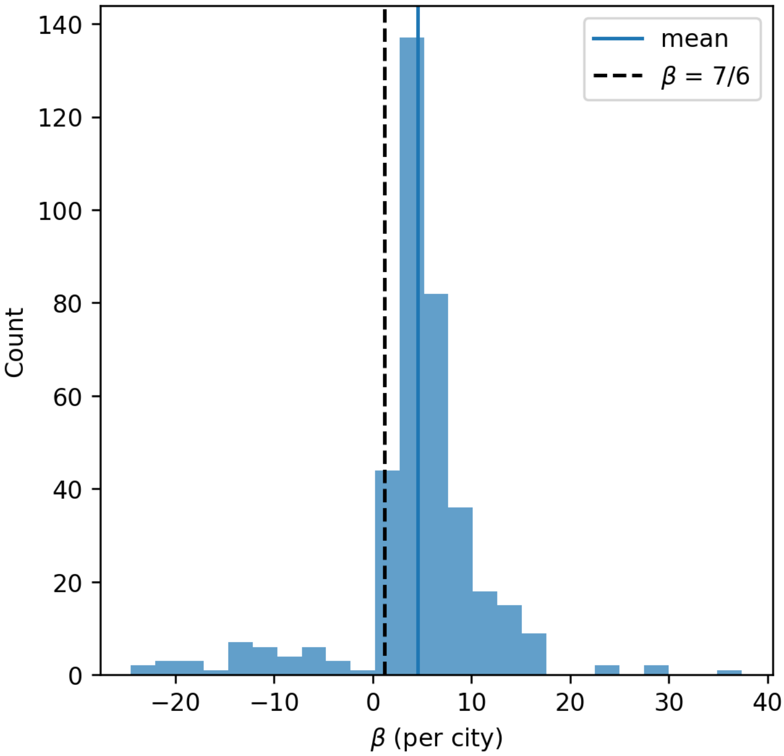

FIG. S4: Distribution of individual longitudinal exponents β_i obtained from power-law fits for each city. The vertical line indicates the transversal exponent β_T estimated from cross-sectional data. The broad distribution highlights the strong heterogeneity of individual city dynamics.

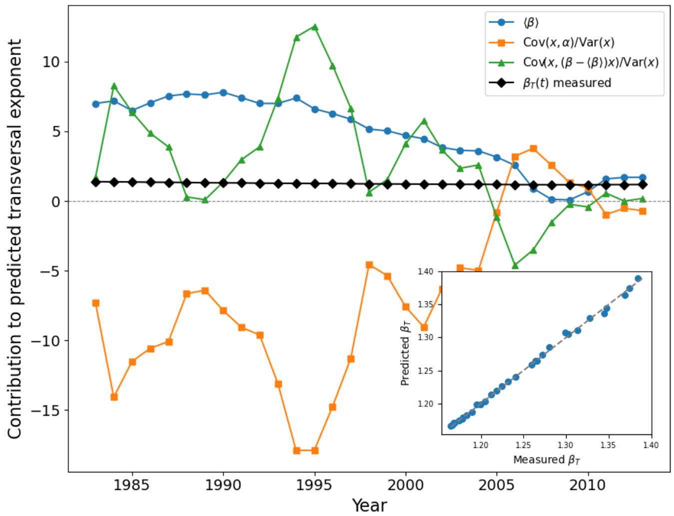

FIG. S5: Decomposition of the transversal exponent into its different contributions, as described by Eq. (17). Inset: Comparison between the transversal exponent measured directly from the data and the value estimated using Eq. (17). Data from [2].

In contrast, the cross-sectional analysis yields a smooth and robust superlinear scaling, reflecting the overall increase of congestion delays with population.

However, this apparent regularity is misleading: the value of the transversal exponent β_T does not directly reflect the intrinsic dynamics of individual cities. Instead, it emerges from the interplay between heterogeneous longitudinal behaviors and statistical correlations across cities. As a consequence, the superlinear scaling observed at the transversal level should not be interpreted as evidence of a universal underlying mechanism governing congestion growth, but rather as a statistical aggregate of diverse and city-specific dynamics.

[1] M. Batty and P. Ferguson, “Defining city size,” *Environment and Planning B: Planning and Design*, vol. 38, no. 5, pp. 753–756, 2011.

[2] J. Depersin and M. Barthelemy, “From global scaling to the dynamics of individual cities,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 10, pp. 2317–2322, 2018.

[3] Texas A&M Transportation Institute and INRIX, “2015 urban mobility scorecard,” Texas A&M Transportation Institute, Technical Report, 2015, accessed: 2026-03-25. [Online]. Available: https://static.tti.tamu.edu/tti.tamu.edu/documents/umr/archive/mobility-scorecard-2015-wappx.pdf