Scaling and Hierarchy in Urban Economies

Cosma Rohilla Shalizi * †

*Statistics Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 1513 USA, and †Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501 USA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

In several recent publications, Bettencourt, West and collaborators claim that properties of cities such as gross economic production, personal income, numbers of patents filed, number of crimes committed, etc., show super-linear power-scaling with total population, while measures of resource use show sub-linear power-law scaling. Re-analysis of the gross economic production and personal income for cities in the United States, however, shows that the data cannot distinguish between power laws and other functional forms, including logarithmic growth, and that size predicts relatively little of the variation between cities. The striking appearance of scaling in previous work is largely artifact of using extensive quantities (city-wide totals) rather than intensive ones (per-capita rates). The remaining dependence of productivity on city size is explained by concentration of specialist service industries, with high value-added per worker, in larger cities, in accordance with the long-standing economic notion of the "hierarchy of central places".

Applied statistics | Model comparison | Urban economics | Urban scaling | Central place hierarchy | Non-parametric smoothing

Abbreviations: MSA, metropolitan statistical area; GMP, gross metropolitan product; BEA, Bureau of Economic Analysis; RMS, root mean square

Recent dramatic advances in explaining metabolic scaling relations in biology by the properties of optimal transport networks [1, 2] suggest the possibility of examining social assemblages, especially cities, in similar terms. In a well-known series of papers, Bettencourt, West and collaborators [3, 4] claim that many social and economic properties of cities — gross economic production, total personal income, number of patents filed, number of people employed in "supercreative" [5] occupations, number of crimes committed, etc. — grow as super-linear powers of population size, while measures of total resource use grow as sub-linear powers. These two claims imply that per capita output grows as a positive power of population, while per capita resource use shrinks as a negative power. If reliable and precise scaling laws of this type exist, they would be of considerable importance for both science and policy [6][1].

Reasonable arguments from long-standing principles of economic geography would lead one to expect that larger cities would have higher economic output per capita, through a combination of the benefits to firms in related industries clustering together ("agglomeration economies"), and the tendency of firms and specialists with large increasing returns to scale to be located high in the "hierarchy of central places". (For reviews of these concepts, including historical notes, mathematical models and empirical evidence, see Refs. [7, 8, 9].) These arguments would carry over to producing technologically useful knowledge and to "supercreative" services as well. However, these economic considerations do not point to either a particular functional form for the growth of per-capita output with population, or suggest that it should be very strong. Moreover, these theories do not look at individual cities as isolated monads, as scaling arguments do, but rather rely on there being assemblages of multiple cities (and rural areas), coupled by common economic processes, and assuming distinct roles in those processes through a history of mutual interaction and combined and uneven development.

The purpose of this note is to argue that, at least for the United States, while there is indeed a tendency for per-capita economic output to rise with population, power-law scaling predicts the data no better than many other functional forms, and worse than some others. Furthermore, the impressive appearance of scaling displayed in Refs. [3, 4] is largely an aggregation artifact, arising from looking at extensive (city-wide) variables rather than intensive (per-capita) ones. The actual ability of city size to predict economic output, no matter what functional form is used, is quite modest. These conclusions hold whether economic output is measured by gross metropolitan product or by total personal income. If we control for metropolitan areas' varying concentration of industrial sectors, we find that the remaining scaling with population is negligible, and much of the variance across cities is predicted by the extent to which they host specialist service providers with strongly increasing returns, as predicted by the idea of the hierarchy of central places.

I begin by re-analyzing the gross metropolitan product data, showing that scaling is far weaker than it seemed in Refs. [3, 4]. I then re-analyzes the data on walking speed, originating in Ref. [10] and presented in Ref. [3], which makes the problems with the scaling analysis very clear. Rather, per-capita productivity is better predicted by how much a city depends on industrial sectors which indicate a high position in the hierarchy of specialist service provision. This actually eliminates any significant role for scaling with size. The conclusions summarize the scientific import of the data analyses.

Appendices show that (i) scaling is also weak for personal income, (ii) the hypothesis of power-law scaling cannot be saved by positing a mixture of distinct scaling relations, and that (iii) *contra* Ref. [4], neither a Gaussian nor a Laplace distribution is a good fit to the deviations from the power-law scaling relations.

All calculations were done using R [11], version 2.12. Code for reproducing figures and analyses is available at `http://www.stat.cmu.edu/~cshalizi/urban-scaling`.

Reserved for Publication Footnotes

[1] Bettencourt and West summarize their claims regarding this "unified theory of urban living" [6] thus: "We have recently shown that these general trends [to cities] can be expressed as simple mathematical laws"; "Our work shows that, despite appearances, cities are approximately scaled versions of one another . . . : New York and Tokyo are, to a surprising and predictable degree, nonlinearly scaled-up versions of San Francisco in California or Nagoya in Japan. These extraordinary regularities open a window on underlying mechanism, dynamics and structure common to all cities"; "Surprisingly, size is the major determinant of most characteristics of a city; history, geography and design have secondary roles".

Weakness of Scaling in Gross Metropolitan Products

Ref. [3] reported a power-law scaling between the population of cities in the United States and their economic output. To be precise, the units of analysis are "metropolitan statistical areas" (MSAs) as defined by the official statistical agencies[2]. The measure of economic output is the gross domestic product for each metropolitan area ("gross metropolitan product" or GMP), as calculated by the U.S. Bureau of Economic Analysis (`http://www.bea.gov/regional/gdpmetro/`), which is supposed to be the sum of all "incomes earned by labor and capital and the costs incurred in the production of goods and services" in the metropolitan area [13][3]. Ref. [3] analyzed data for 2006, deflated to constant 2001 dollars, and I will do likewise; the 2008 and 2004 data are not much different.

Ref. [3] propose that output scales as a power of population, $Y \propto N^b$. This is connected to the data via the linear regression model

$$\ln Y = \ln c + b \ln N + \epsilon, \qquad [1]$$

with ϵ being a mean-zero noise term. For later comparisons, it will be convenient to denote this by $Y \sim cN^b$. Ref. [3] estimated b by least squares [15], i.e., by minimizing $n^{-1}\sum_{i=1}^{n}(\ln Y_i - \ln c - b\ln N_i)^2$, where the index i runs over metropolitan areas, of which there are $n = 366$. Repeating this analysis[4], I estimate the best-fitting scaling exponent $\hat{b} = 1.12$, with 95% bootstrap confidence interval of $(1.10, 1.15)$ [15]. Figure 1 shows the data and the fitted trend, with both axis plotted on a logarithmic scale, so that a power law relationship appears as a straight line. The root-mean-squared (RMS) error for predicting $\ln Y$ is 0.23, and the "coefficient of determination" R^2 is 0.96, i.e., the fitted values retain 96% of the variance in the actual data.

Visually, this looks like reasonable data collapse. But there is a simple test of the model which has not, so far as I know, been applied before. If production does scale as some power of population, $Y \sim cN^b$, then per-capita production should also scale, $Y/N \equiv y \sim cN^{b-1}$, and vice versa. Figure 2 accordingly plots per-capita output y as a function of population N.

Figure 2 shows a trend curve for the the power-law scaling implied by Ref. [3]. (The exponent estimated for y is 0.12, matching that estimated for Y, as it must.) The figure also shows a logarithmic scaling relationship, i.e., $y \sim r \ln N/k$ (estimated by nonlinear least-squares), which is extremely close to the power law over the range of the data. It also shows an attempt to find a scaling relationship without requiring any particular function form by fitting a smoothing spline [16, 17] to the logged data[5], corresponding to the relationship $y \sim e^{s(\ln N)}$. Note that the fitted curve is not even monotonically increasing in N.

While the three curves in Figure 2 correspond to very different modeling assumptions — the differences between the implications of power-law and logarithmic growth are perhaps especially striking — they all account for the data about equally well, or rather, equally poorly, because there is substantial variation in per-capita production which is unrelated to population. (Note that the vertical axis is plotted on a linear and not a logarithmic scale.) The RMS error of the power law is, on the natural log scale, 0.23, while that of the spline is 0.22. They would predict y, for a randomly chosen city, to within ± 26, ± 26 and ± 25 percent, respectively. Predicting the same value of y for all cities, however, has an RMS error of 0.27, a margin of $\pm 30\%$, and the R^2 values are, respectively, 0.24, 0.23 and 0.29. On the linear scale, i.e., in terms of dollars per person-year, the RMS errors of the power law, logarithmic and spline curves are, respectively, 7.9×10^3, 7.9×10^3 and 7.7×10^3, as compared to 9.2×10^3 for predicting the mean for all cities.[6] In other words, even allowing for quite arbitrary functional forms, city size does not predict economic output very well.

The similarity of the RMS errors, and indeed of the curves, arises in part from the limited range of y. The difference between the largest and smallest per-capita products (6.3×10^4 dollars/person-year) is "only" a factor of 5.2, i.e., not even one order of magnitude. This is too small, with only 366 observations, to distinguish among competing functional forms for the trend, while still being quite consequential in human and economic terms. Per-capita production is simply not very strongly related to population.

Taking any per-capita (intensive) quantity which is statistically independent of population, and looking at the corresponding aggregate (extensive) quantities will yield a scaling exponent close to one. The overwhelming majority of the apparent fit of the scaling relationship in Figure 1 is just such an artifact of aggregation. This can be shown in three different ways: by extrapolating the different per-capita functional forms back to city-wide totals, by constrained regression, and by permutation.

Figure 3 shows the same data and scaling curve as Figure 1, but with three additional trend lines. Two of these come from taking the logarithmic and spline fit to the per-capita data, and plotting the implied aggregates, i.e., these are the regressions $Y \sim \hat{r}N \ln N/\hat{k}$ and $Y \sim Ne^{\hat{s}(\ln N)}$. These are, visually, almost indistinguishable from the power law $Y \sim \hat{c}N^{\hat{b}}$. The figure also shows a second power law scaling curve, constrained to have exponent 1. (This was obtained by a linear regression with the slope fixed at 1 but an adjustable intercept.) This curve corresponds to exactly linear scaling. It is a bit low, systematically, at large N, but it still has an R^2 of 0.82, as opposed to the common 0.96 for the unconstrained power law, per-capita logarithmic growth, and the per-capita spline. (Examples like this are why regression textbooks advise against using R^2 to check goodness of fit [18, 19, 20].)

Figure 4 demonstrates in a different way that the data do not support the idea of power-law scaling. The circles in the figure show the actual data values. The stars, by contrast, are surrogate data simulated from the fitted logarithmic growth model, with the actual population sizes. The surrogate per-capita output values $\tilde{y}$ were set equal to $\hat{r}\log N/\hat{k}$, and then randomly perturbed according to the empirical distribution of deviations from that model. The figure plots the surrogate aggregate products $\tilde{y}N$, which look very much like the data.

If a power-law scaling relation is fit to the surrogate data from the logarithmic-growth regression, then, averaging over many simulations, the median scaling exponent is 1.12, with 95% of the estimates falling between 1.10 and 1.14, and the median R^2 of the power-law was 0.96. Recall that the es-

[2] To quote Ref. [12], MSAs are "standardized county-based areas that have at least one urbanized area with a population of 50,000 or more plus adjacent territory that has a high degree of social and economic integration with the core, as measured by commuting ties."

[3] A word on the BEA's procedure is in order [14]. The BEA estimates gross products for each industry for each state, and conducts surveys to estimate what fraction of each industry's state-wide earnings is located in each metropolitan area. Multiplying these ratios by the state-wide gross products, and summing over industries, gives the gross metropolitan product. The BEA provides no estimates of measurement uncertainty for these numbers.

[4] Which is easily shown to yield reliable estimates of c and b, unlike least-squares regression of log-transformed values in the superficially parallel situation of fitting a power-law distribution.

[5] That is, the estimated spline is the function s minimizing $n^{-1}\sum_i(\ln y_i - s(\ln N_i))^2 + \lambda\int(s''(x))^2dx$, with the smoothness penalty $\lambda > 0$ chosen by cross-validation. Smoothing splines of this type are universal approximating functions, and picking the penalty by cross-validation controls the risk of over-fitting non-generalizing aspects of the data — see Ref. [17] for details. A smoothing spline fit to the un-transformed data was similar, but visually somewhat more jagged.

[6] All of these measures of error are calculated on the same data used to fit the models, which of course makes them over-optimistic estimates of the models' predictive powers. However, using six-fold cross-validation to estimate the out-of-sample risk gives RMS errors of 0.23 for all three models.

timate for the actual data was 1.12, with a 95% confidence interval of $(1.10, 1.15)$, and $R^2 = 0.96$.

It is true that the RMS error for $\ln y$ on the real data is very slightly lower for the power law (0.2322) than for the logarithmic model (0.2336), but such a minute difference can easily arise by chance. Repeating both fits for the surrogate data, in fact, the power law is a better fit to $\ln \tilde{y}$ by at least the empirically-observed margin on 45% of the simulations from the other model. Reliably discriminating between the two models simply requires more information (in the sense of [21]) than the data provides: either much smaller fluctuations of y around the regression curve, or many more data points.

To sum up, the appearance of a strong, super-linear relationship between gross production Y and population N is mostly driven by production growing in proportion to population — that is, linearly. Per-capita production y does not have a strong scaling relationship with N, and the data are unable to distinguish between different functional forms for such trends as there are. Lacking ready access to the data sets on patents, crime, infrastructure and resource consumption used in Ref. [3], I cannot say whether the reported scaling relations for those aggregate variables suffer from the same problem. I return to the question of why there is a weak and noisy tendency for per-capita output to rise with population in Sections and below.

"The Pace of Life". A further claim of Ref. [3] is that the speed at which people walk grows as a positive power of the number of people in a city. The source given for this is Ref. [10], a two-page letter to *Nature* in 1976. The authors of Ref. [10] went to 15 cities, towns and villages, picked locations and individuals which seemed to them to be comparable, and timed how long it took them to walk fifty feet (15.2 meters). Such unsystematic data, however intriguing, is too weak to support substantial scientific conclusions. Nonetheless, it is instructive to examine it, as in Figure 5.

The original plot (Figure 1 in Ref. [10]) showed population on a log scale, and speed on a linear scale, as in Figure 5. The linear regression, for this transformation of the data, corresponds to assuming that speed grows logarithmically with population, $v \sim r \ln N/k$. Figure 2a in Ref. [3] re-plots the same data, but with the vertical axis on a logarithmic scale, so the linear regression assumes speed grows as a power of population, $v \sim cN^b$. (Neither figure included error bars, though Bornstein and Bornstein give the standard deviations in their caption.) As can be seen from Figure 5, the two regressions are very similar in this data, while they embody very different assumptions, and at most one can be right.

The explanation for this apparent paradox is that the range of reported walking speeds is small, from 0.7 m/s to 1.8 m/s, and if $|x| \ll 1$, then $\ln 1 + x \approx x$. Observed over a narrow range, then, logarithmic and power law scaling simply are very similar, and hard to distinguish. This is also why the the power-law and logarithmic fits to per-capita production in Figure 2 were so close.

Hierarchy as an Alternative to Scaling

The idea of the hierarchy of central places, introduced by Lösch and Christaller in the 1930s, has become a corner-stone of urban economic geography. In outline, the idea is that developed economies contain many specialized goods, and especially services, that the mass of consumers need only rarely (such as the services of a surgeon), or indirectly (such as the services of a professor of surgery, or a maker of surgical instruments). The provision of such services has comparatively high fixed costs (the time needed to train a surgeon) but low marginal costs (the time needed to perform an operation), leading to increasing returns to scale. It thus becomes economically efficient for these specialists to locate in central places, where their fixed costs can be distributed over large consumer bases, and the more specialized they are, the more centrally located they need to be, and the larger the customer base they require. This logic leads to the formation of a hierarchy of market centers and cities, in which increasingly specialized skills, with (as it were) increasingly increasing returns, can be had, and so predicts positive associations between the population of urban centers, the concentration of specialist skills within them, and (owing to increasing returns) their per-capita economic output. Good reviews of the theory, including historical citations and connections to modern economic models of increasing returns, may be found in Refs. [7, 9].

This is relevant to the problem at hand because the BEA also makes available estimates of the shares of gross metropolitan products which are attributable different industrial sectors, some of which correspond to the specializations emphasized in central place theory. I specifically consider "Information, Communication, and Technology (ICT)", "Financial activities", "Professional and technical services" and "Management of companies and enterprises" (industry codes 106, 102, 58 and 62, respectively)[7]. Writing the proportions of gross metropolitan product deriving from each of these sectors as x_1 through x_4, the level of per-capita production can be predicted by a log-additive model which incorporates power-law scaling with city size:

$$\ln y = \ln c + b \ln N + \sum_{j=1}^{4} f_j(x_j) + \epsilon, \qquad [2]$$

where each of the "partial response" functions f_j summarizes the contribution of the j^{th} industrial sector, and is estimated by non-parametric spline smoothing, through an iterative "back-fitting" procedure [16]. Doing so adjusts for the correlations between the industrial sectors and each other, and all of them with city size N, and will result in a different value of the scaling exponent b than in the pure power law model of Eq. 1.

Fitting Eq. 2 to the data yields the partial response functions shown in Figure 6. (The fitting was done using the `mgcv` library [22].) As expected from the urban-hierarchy argument, all four of the partial response functions are monotonically increasing, so that rising shares of those industries predict increasing per capita production. Very notably, however, the estimated power-law scaling exponent is actually negative, -2.6×10^{-3}, but statistically indistinguishable from zero (standard error 2.8×10^{-2}). That is, in the log-additive model, controlling for these four industrial sectors makes population effectively irrelevant for predicting urban productivity. Indeed, dropping population from the model altogether produces no appreciable difference in the fit. At least at the level of expectation values, controlling for these four industrial sectors "screens off" the effects of city size on per-capita production.

Statistically, there is no question that the log-additive model predicts better than the simple scaling model. The RMS error of the former, on the log scale, is 0.218, corresponding to an R^2 of 38.8%, and an accuracy of $\pm 24\%$ or $\$6.8 \times 10^3$,

[7] The BEA withholds the GMP-contribution figures for some industry-MSA combinations, when the sector is so concentrated in that city that releasing the number would provide consequential business information about specific firms. I have fit the model discussed below for the 133 cities with complete data in the four selected sectors. Experimenting with various forms of imputation for the missing data did not materially change the results.

better than any model based on size alone. The log-additive model is a more flexible specification, and so over-fitting to the data is an issue, but this can be addressed by cross-validation, which directly measures the ability of a model to extrapolate from one part of the population to another [16]. The cross-validated mean squared error of the log-additive model is 0.053, while that of the pure power law is 0.067, clearly showing that the extra complexity of the former is being used to capture genuinely predictive patterns, and not merely to memorize the training data[8].

The simple log-additive model is unlikely to be a fully adequate predictor of systematic differences in urban productivity. If nothing else, these four coarse-grained industrial sectors were selected merely for convenience, as approximate indicators of position in the urban hierarchy, and presumably one could do better. Moreover, it does not even try to represent the interactive processes which lead cities to have the industrial mixes that they do. In reality, these industries can be so concentrated towards the largest cities, at the top of the hierarchy (e.g., New York), and away from lower-rank cities (e.g., San Francisco, Peoria), only because all these cities are part of a single national, and even international, division of labor [9].

Conclusion

Neither gross metropolitan product nor personal income scales with population size for U.S. metropolitan areas. The appearance of scaling in Refs. [3, 4] is an artifact of inappropriately looking at extensive variables (city-wide totals) rather than intensive ones (per-capita values). Scaling is also unpersuasive for walking speed. I was not able to examine the other variables claimed to show scaling in Refs. [3, 4], but, as they were all extensive variables, the analyses reported there could be subject to the same aggregation artifacts. It is also possible that cities in the United States are anomalous, and that scaling of income and economic output holds elsewhere.

It is evident from Figures 2 (and Supplemental Figure S1) that there is a weak tendency for per-capita output and income to rise with population, though the relationship is simply too loose to qualify as a scaling law. (Arguably, the real trend in those figures is for the *minimum* per-capita output to rise with population, though I would not want to press this point.) Qualitatively, this is what one would expect from well-established findings of economic geography [8]. The data do not really support any stronger *quantitative* statement. In particular, asserting any specific functional form, such as a power law, goes far beyond the what the data can support. Accordingly, extrapolations based on such claims (e.g., the finite-time singularity in the model for city growth in [3]) are speculative at best. The amplitude of fluctuations around the trend lines are, in any case, so large that predictions based on size alone can have very little utility.

By taking account of the shares of just a few industries in the gross metropolitan product, we can obtain much better predictions of the level of per-capita production. In this statistical model, elaborated in Section , population plays no significant direct role in predicting per capita economic output. Rather, the industrial sectors used are chosen as signs of where metropolitan areas stand in the urban hierarchy, which is also related, of course, to size. One could interpret this as the mechanism by which size scaling happens (to the limited extent that it does), but this would imply that an exogenous increase in a city's population would automatically shift its industrial pattern, which is implausible. Indeed, the whole scaling picture for cities seems to rest on an oddly monadic, interaction-free view of metropolitan areas. The logic of central place theory, in contrast, relies on cities being part of an interactive assemblage, coupled by processes of production, distribution and exchange. This not only seems more plausible, but also better matches the evidence at hand.

As Refs. [3, 4, 6] have stressed, developing a sound scientific understanding of cities should be a priority for an increasingly urban species. In seeking such understanding, it is a sound strategy to begin with simple hypotheses, and to reject them in favor of more complicated ones only as they prove unable to explain the data. This is not because the truth is more likely to be simple, in some metaphysical sense, but because this strategy leads us to the truth faster and more reliably than ones which invoke needless complexities [23]. The elegant hypothesis of power-law scaling marked a step forward in our understanding of cities, but it is now time to leave it behind.

Appendix

Personal Income. The BEA also makes available estimates of personal income by metropolitan area, a variable closely related to, but not quite the same as, the gross metropolitan product. (See `http://www.bea.gov/regional/reis/` for definitions, estimation techniques, and data.) Ref. [4] reports that total personal income L also scales as a power of population, implying per capita personal income $L/N \equiv l$ should scale likewise. Figure 7 plots l versus N, with the best-fitting power law, logarithmic relationship, and spline.

Once again, the appearance of power-law scaling in the aggregate variable is not supported by examination of the per-capita values. The RMS error, on the log scale, of predicting a constant per capita income over all cities is 0.18, while the RMS errors of the power-law and logarithmic scaling relations are both 0.16 (indeed they match to three significant digits), and that of the spline 0.15. Repeating the procedures of Figures 3 and 4 from the main text yields similar results. Thus, personal income also fails to display non-trivial scaling with population.

Mixtures of Scaling Relations. Recall that the posited scaling relation is $y \sim cN^b$. As shown above, this does not fit the data, at least not assuming, following Ref. [3], that both parameters, the scaling exponent b and the pre-factor c, are the same for all cities. A natural way to try to reconcile the data with the model would be to modify the latter, allowing c to depend on the *type* of the city. The rationale for such a regression would be that there are several different kinds of cities, and that city type shifts the over-all level of production up or down, but, once that is factored out, all cities scale with size in the same way. This common scaling exponent would not, naturally, be the same as the one estimated from the pooled data.

Formally, we introduce a latent variable Z for each city, treated as a discrete random variable independent of N, and consider the statistical model $y \sim c_Z N^b$. This leads to a "mixture-of-regressions" or "latent-class regression" model, which can be fit by the expectation-maximization algorithm [24]. Such fitting would lead not only to estimates of b and the pre-factors c_z, but also to the probability that each city belonged to each of the different city types or mixture components, categorizing cities inductively from the data.

[8] Dropping population size N from the log-additive model altogether does however lower the cross-validation score very slightly, to 0.052.

To investigate this, I fit mixture-of-regression models to the data from Figure 2 in the main text, varying the number of mixture components from 1 to 10, using the software of Ref. [24].[9] To determine the correct number of mixture components, I used both Schwarz's "Bayesian" information criterion and cross-validation, which are both known to be consistent for such mixture problems, unlike the Akaike information criterion, which over-fits [25]. Both BIC and cross-validation strongly favored *one* mixture component, meaning that the fit to the data is not actually improved by allowing for multiple scaling curves.

This does not completely rule out the $y \sim c_Z N^b$ model, as only 366 observations may not have enough information to simultaneously induce appropriate categories and fit scaling relations. An alternative would be to expand the information available, by defining the categorical variable Z in terms of measurable attributes of cities other than N and y, such as geographic location or the mix of industries. (See Ref. [26] on such variable-intercept, constant-slope regressions with known categories.) Success with such models hinges on selecting categories to represent important features of the data-generating process, a task I must leave to other inquirers.

Assuming that such a statistical model works, there would still be the question of its interpretation. Whether one would judge such a model to really show scaling in urban assemblages would depend on how much importance one gives, on the one hand, to a common scaling exponent, and on the other to most of the fit coming from the un-modeled differences across city types.

Residuals. Ref. [4] proposes ranking cities not by their per capita values of quantities like economic production or patents or crime, but by the deviation, positive or negative, from the scaling relationship, i.e., by the residuals of the trend lines. (It does not compare this to ranking by per capita values. The Spearman rank correlation between the two variables is 0.87 for GMP and 0.83 for personal income.) They consider both a Gaussian distribution for the residuals, i.e., a probability density $f(x) \propto e^{-x^2/2\sigma^2}$, and a Laplace distribution, $f(x) \propto e^{-\lambda|x|}$, and claim that both fit very well.

Figure 8 shows the situation for GMP. Visually, neither distribution matches the residuals well. Quantitatively, goodness-of-fit can be checked by "data-driven smooth tests" [27], which transform their inputs so that they will be uniform if and only if the postulated distribution holds, and then measure departures from uniformity (coefficients from expanding the transformed empirical distribution in a series of orthogonal polynomials). Such tests reject both the Gaussian and the Laplace distribution with high confidence (p-values of 1×10^{-3} and 8×10^{-3}, respectively, calculated using code provided by Ref. [28]).

Results for personal income are similar (Figure 9). The Gaussian distribution can be rejected with high confidence ($p < 10^{-4}$). While the data do not rule out the Laplace distribution in the same way ($p = 0.27$), the limited power of the test at the comparatively small sample size means that there is not strong evidence in its favor either. (See Ref. [29] on the evidential interpretation of significance tests.)

ACKNOWLEDGMENTS. Thanks to Jordan Ellenberg for valuable questions, and to grants from the NIH (# 2 R01 NS047493) and the Institute for New Economic Thinking for partial support.

References

1. G. B. West, J. H. Brown, and B. J. Enquist. A general model for the origin of allometric scaling laws in biology. *Science*, 276:122–126, 1997. URL `http://www.santafe.edu/media/workingpapers/97-03-019.pdf`.
2. J. H. Brown and G. B. West, editors. *Scaling in Biology*, Oxford, 2000. Oxford University Press.
3. L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Künhert, and G. B. West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences (USA)*, 104:7301–7306, 2007. doi: 10.1073/pnas.0610172104.
4. L. M. A. Bettencourt, J. Lobo, D. Strumsky, and G. B. West. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS One*, 5:e13541, 2010. doi: 10.1371/journal.pone.0013541.
5. R. L. Florida. *The Rise of the Creative Class*. Basic Books, New York, 2002.
6. L. M. A. Bettencourt and G. B. West. A unified theory of urban living. *Nature*, 467:912–913, 2010. doi: 10.1038/467912a.
7. P. R. Krugman. *The Self-Organizing Economy*. Blackwell, Oxford, 1996.
8. J. V. Henderson, Z. Shalizi, and A. J. Venables. Geography and development. *Journal of Economic Geography*, 1:81–105, 2001. doi: 10.1093/jeg/1.1.81. URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=632526`.
9. M. Fujita, P. R. Krugman, and A. J. Venables. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge, Massachusetts, 1999.
10. M. H. Bornstein and H. G. Bornstein. The pace of life. *Nature*, 259:557–559, 1976. doi: 10.1038/259557a0.
11. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL `http://www.R-project.org`. ISBN 3-900051-07-0.
12. M. J. McCormick, S. D. Panek, and R. M. Rodriguez. Gross domestic product by metropolitan area: Accelerated statistics for 2008, new statistics for 2007, and revised statistics for 2005–2006. *Survey of Current Business*, pages 100–128, October 2009. URL `http://www.bea.gov/scb/pdf/2009/10%20October/1009_gdpmetro.p`
13. S. Assenova, J. Broda, and S. Taylor. Gross domestic product by state: Advance statistics for 2009 and revised statistics for 1963–2008. *Survey of Current Business*, pages 72–77, December 2010. URL `http://www.bea.gov/scb/pdf/2010/12%20December/1210_gdp_state`
14. S. D. Panek, F. T. Baumgardner, and M. J. McCormick. Introducing new measures of the metropolitan economy: Prototype gdp-by-metropolitan-area estimates for 2001–2005. *Survey of Current Business*, pages 79–114, November 2007. URL `http://www.bea.gov/scb/pdf/2007/11%20November/1107_gdpmetro.`
15. L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer-Verlag, Berlin, 2003.
16. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, Berlin, second edition, 2009. URL `http://www-stat.stanford.edu/~tibs/ElemStatLearn/`.
17. G. Wahba. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, Philadelphia, 1990.

[9] For computational reasons, it is easier to fit the more general specification in which the scaling exponent is also allowed to vary, $y \sim c_Z N^{b_Z}$. (Sharing a parameter across the regressions complicates the maximization step of the expectation-maximization algorithm.) If the constant-exponent model is right, the estimated exponents for each mixture component should agree to within statistical precision.

18. M. H. Birnbaum. The devil rides again: Correlation as an index of fit. *Psychological Bulletin*, 79:239–242, 1973. doi: http://dx.doi.org/10.1037/h0033853.
19. N. H. Anderson and J. Shanteau. Weak inference with linear models. *Psychological Bulletin*, 84:1155–1170, 1977. doi: 10.1037/0033-2909.84.6.1155.
20. R. A. Berk. *Regression Analysis: A Constructive Critique*. Sage, Thousand Oaks, California, 2004.
21. S. Kullback. *Information Theory and Statistics*. Dover Books, New York, 2nd edition, 1968.
22. S. N. Wood. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99:673–686, 2004. URL `http://www.maths.bath.ac.uk/~sw283/simon/papers/magic.pdf`.
23. K. T. Kelly. Simplicity, truth, and probability. In Prasanta Bandyopadhyay and Malcolm Forster, editors, *Handbook on the Philosophy of Statistics*. Elsevier, Dordrecht, 2010. URL `http://www.andrew.cmu.edu/user/kk3n/ockham/prasanta-submitted-final.pdf`.
24. Tatiana Benaglia, Didier Chauveau, David R. Hunter, and Derek S. Young. mixtools: An R package for analyzing mixture models. *Journal of Statistical Software*, 32, 2009. URL `http://www.jstatsoft.org/v32/i06`.
25. Gerda Claeskens and Nils Lid Hjort. *Model Selection and Model Averaging*. Cambridge University Press, Cambridge, England, 2008.
26. Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, England, 2006.
27. Wilbert C. M. Kallenberg and Teresa Ledwina. Data-driven smooth tests when the hypothesis is composite. *Journal of the American Statistical Association*, 92:1094–1104, 1997. URL `http://doc.utwente.nl/62408/`.
28. Przemyslaw Biecek and Teresa Ledwina. *ddst: Data driven smooth test*, 2010. URL `http://CRAN.R-project.org/package=ddst`. R package, version 1.02.
29. Deborah G. Mayo and D. R. Cox. Frequentist statistics as a theory of inductive inference. In Javier Rojo, editor, *Optimality: The Second Erich L. Lehmann Symposium*, pages 77–97, Bethesda, Maryland, 2006. Institute of Mathematical Statistics. URL `http://arxiv.org/abs/math.ST/0610846`.
30. Jeffrey S. Simonoff. *Smoothing Methods in Statistics*. Springer-Verlag, Berlin, 1996.

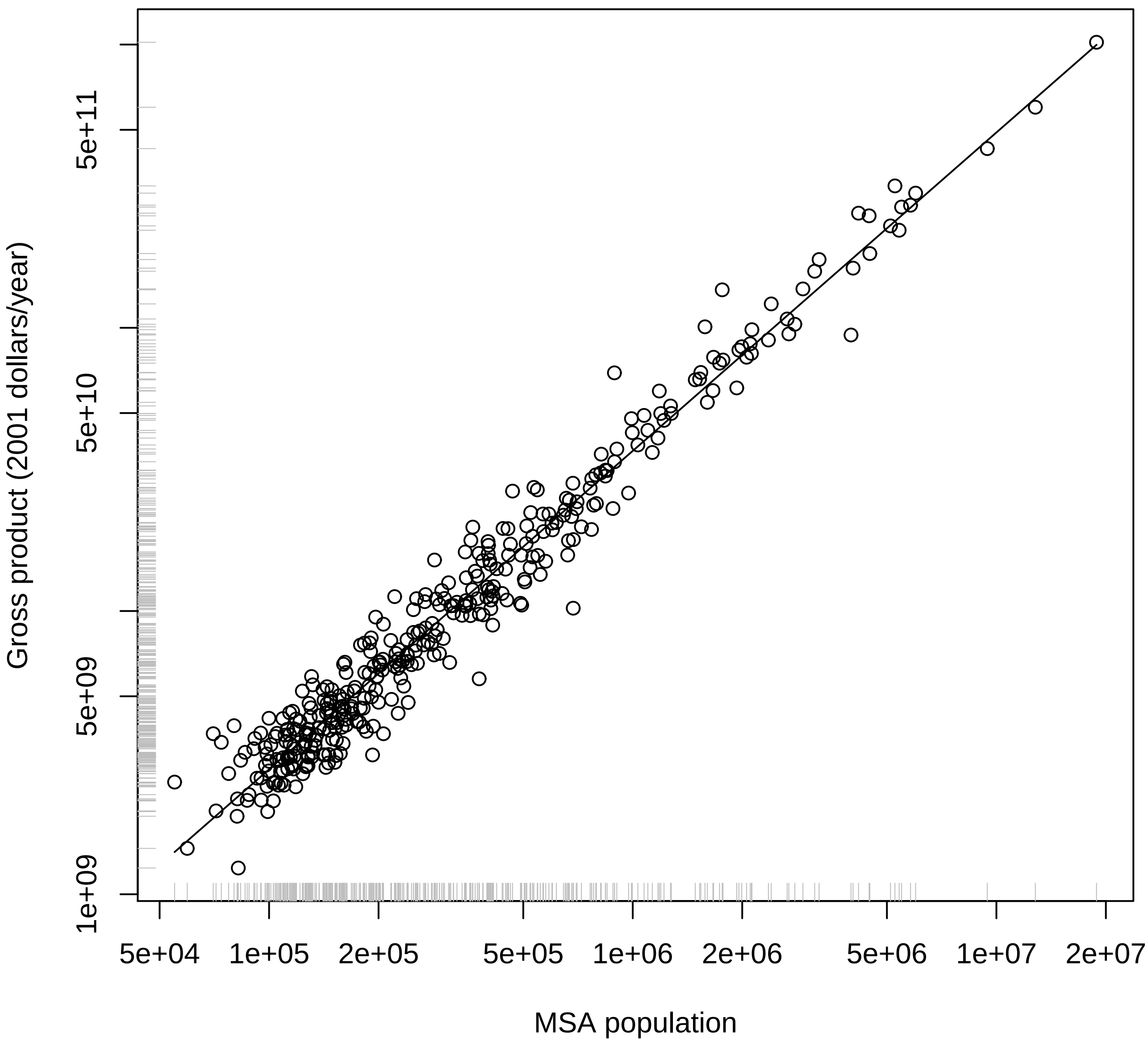

Fig. 1. Horizontal axis: population of the 366 US metropolitan statistical areas in 2006, log scale; vertical axis, 2006 gross product of each MSA, in constant 2001 dollars, log scale. (In all figures, grey inner ticks on axes mark observed values.) Solid line: ordinary least squares regression of log gross metropolitan product on log population, i.e., the regression $Y \sim cN^b$, with estimated exponent $\hat{b} = 1.12$.

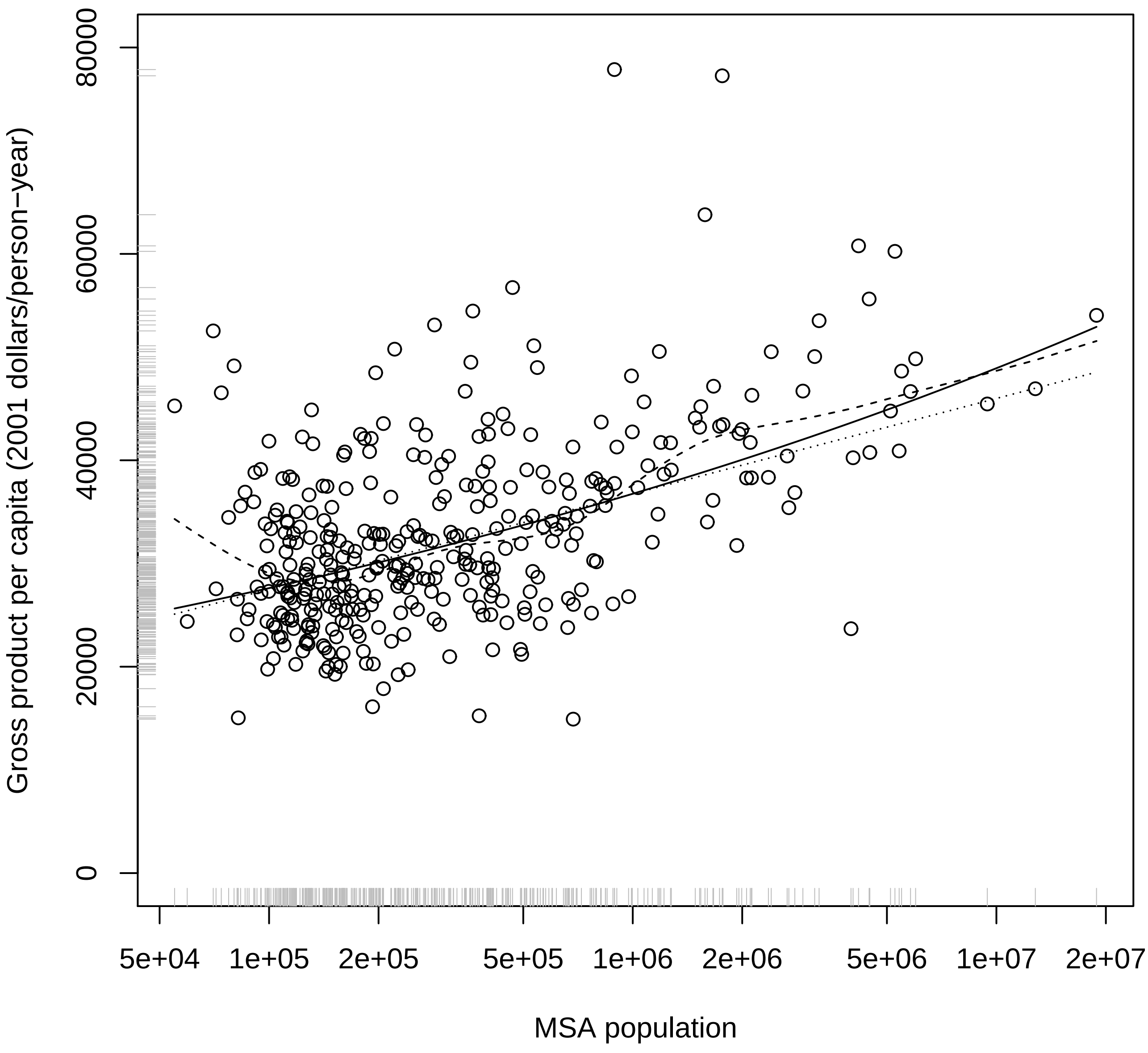

Fig. 2. Horizontal axis: population, as in Figure 1, log scale. Vertical axis: gross product per capita, but on a linear and not a logarithmic scale. The two largest values are 7.8×10^4 dollars/person-year (in Bridgeport-Stamford-Norwalk, CT, a center for hedge funds and other financial firms) and 7.7×10^4 dollars/person-year (in San Jose-Sunnyvale-Santa Clara, CA, i.e., Silicon Valley), and the smallest are 1.5×10^4 dollars/person-year (in McAllen-Edinburg-Mission, TX and Palm Coast, FL). Solid line: fitted power-law scaling relation. Dotted line: fitted logarithmic scaling relationship. Dashed curve: smoothing spline fitted to the logged data.

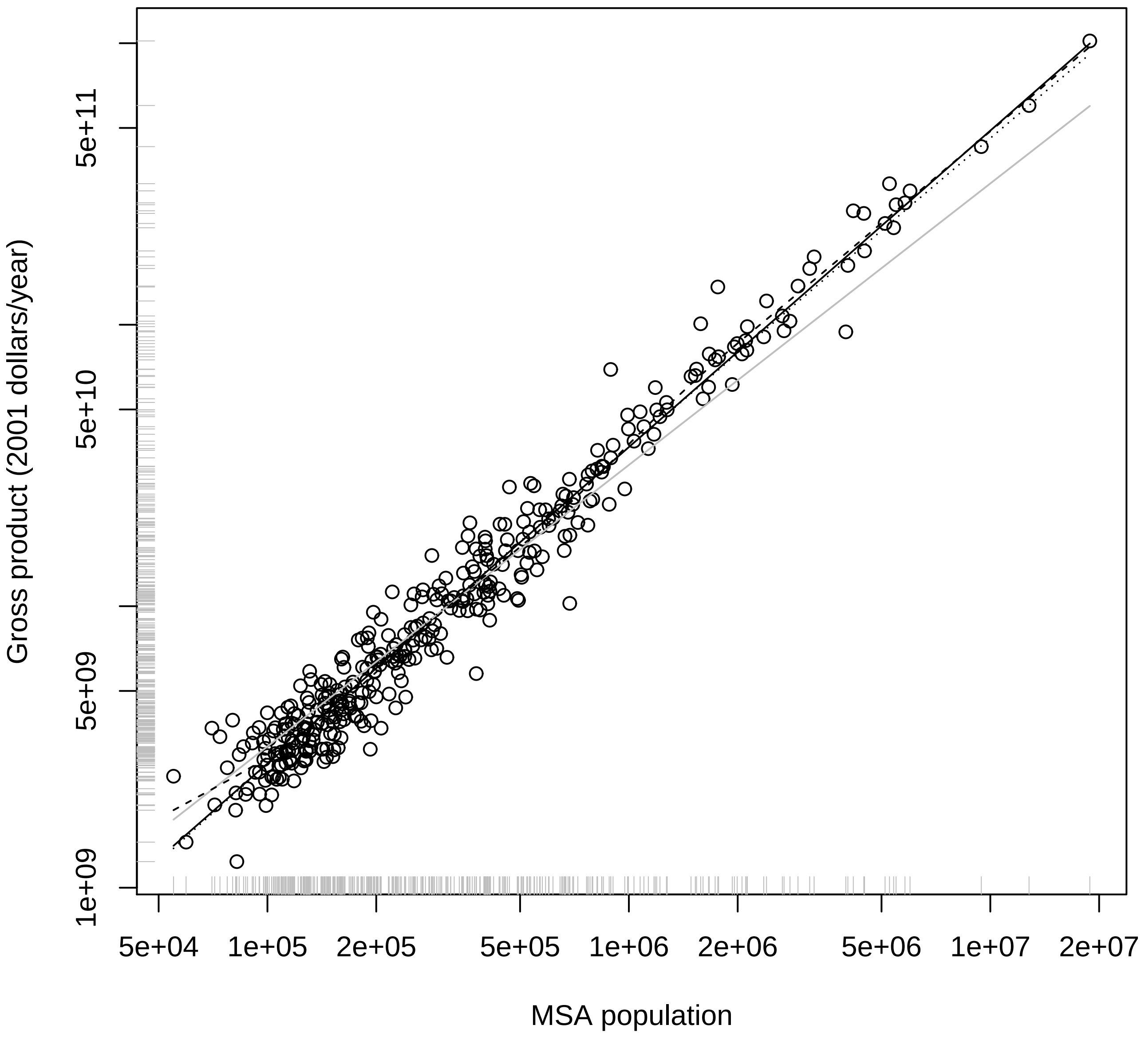

Fig. 3. As in Figure 1, but with the addition of (i) dotted line, showing logarithmic growth of per-capita income, (ii) dashed line, showing spline fit to per-capita income (both as in Figure 2), and (iii) solid grey line, showing a fitted linear (not super-linear) scaling relationship. Note that (i) and (ii) are extremely close to the solid power-law line.

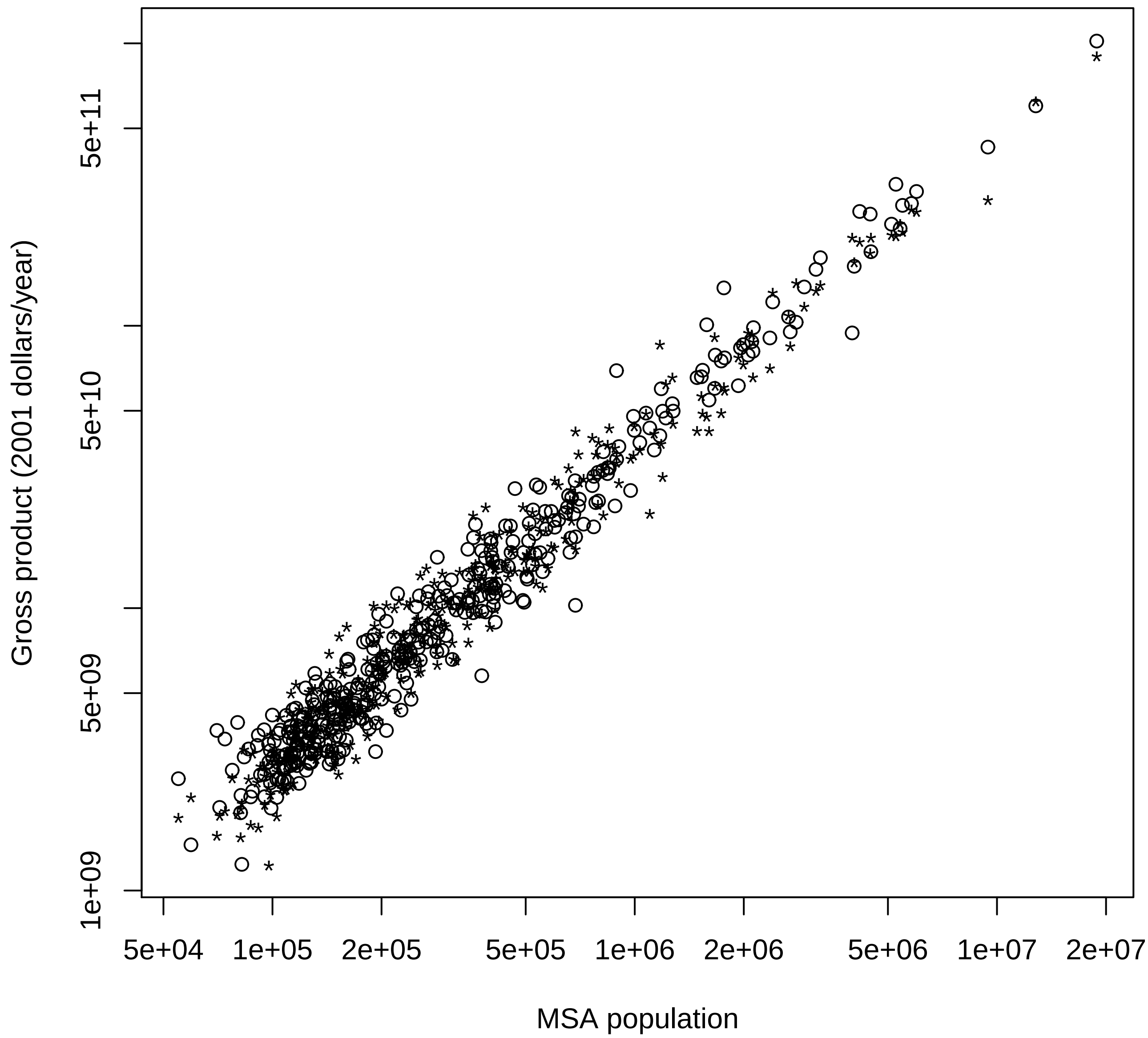

Fig. 4. Axes: as in Figure 1 and 3. Circles: Actual values. Stars: simulated values, with per-capita production figures drawn from the logarithmic (not power-law) scaling model.

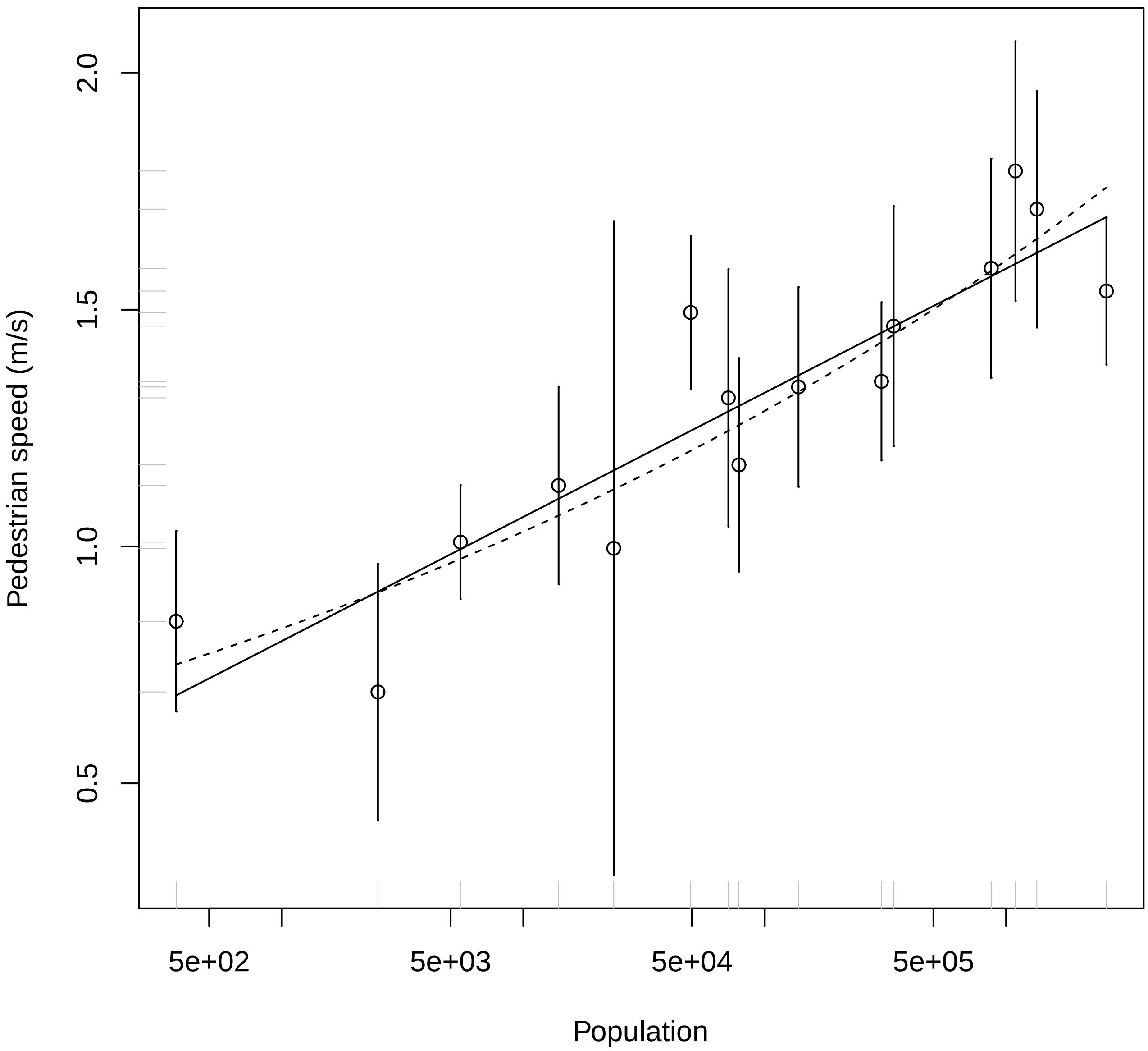

Fig. 5. Horizontal axis: city population, logarithmic scale. Vertical axis: estimated pedestrian speed in meters/second, plus or minus one standard deviation, linear scale. Solid line: the regression $v \sim r \ln N/k$, as proposed by Ref. [10]. Dashed line: the regression $v \sim cN^b$, as proposed by Ref. [3]. (Data from Ref. [10], who report the mean and standard deviation of the time taken to walk 50 feet = 15.2 meters; I calculated standard deviations by propagation of error.)

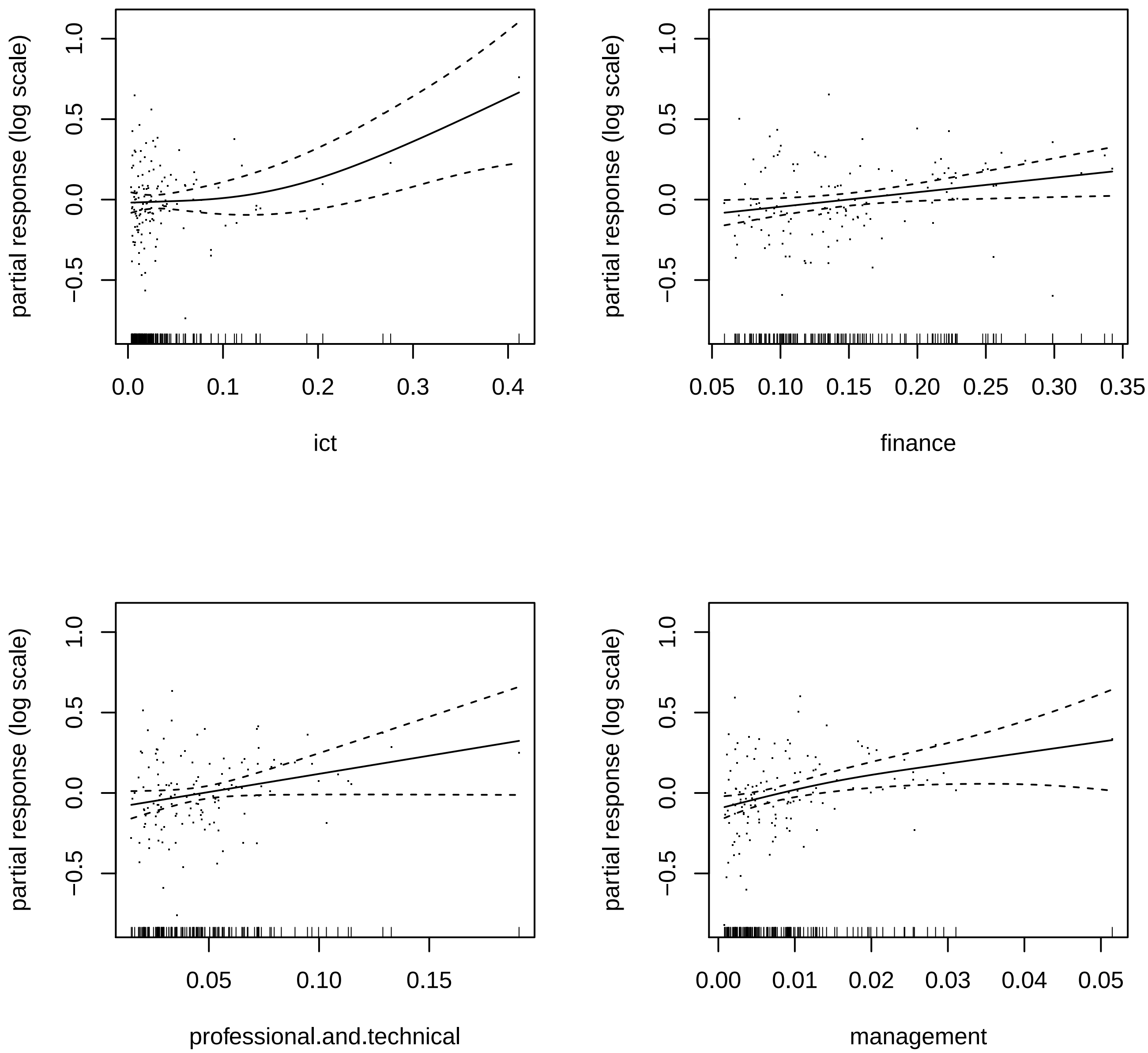

Fig. 6. Partial response functions for the log-additive model (Eq. 2). Horizontal axes indicate the fraction of each metropolitan area's gross product derived from each industry, while the vertical axis shows the predicted logarithmic increase, or decrease, to per capita output, relative to the baseline of the mean over all cities. Solid curves are the main estimate, with dashed curves at ± 2 standard errors in the partial response function. Dots show "partial residuals", the difference between actual $\log y$ values and those predicted by the model including all the other variables.

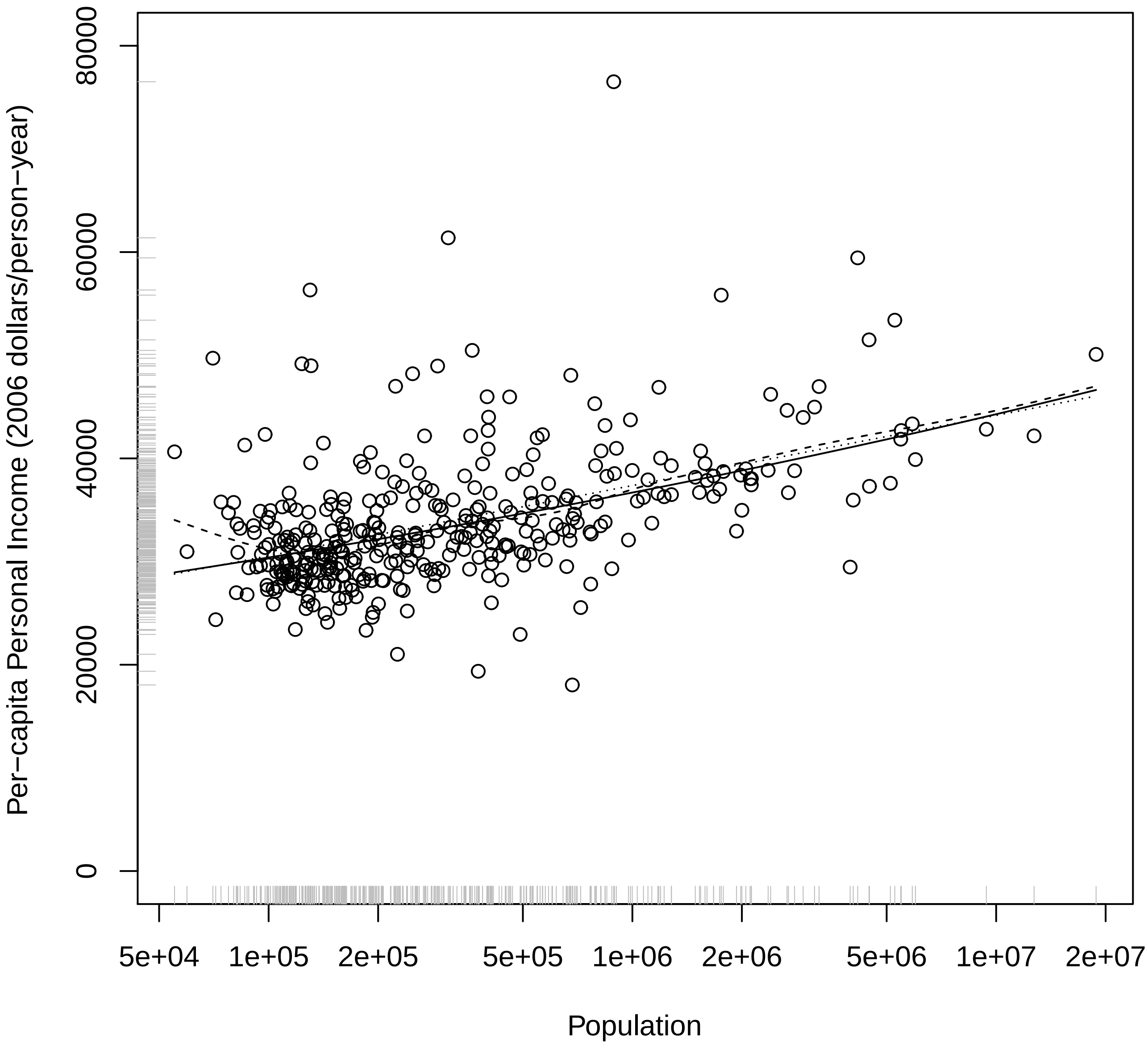

Fig. 7. Personal income per capita versus population, 2006. Horizontal axis: population of MSAs (log scale). Vertical axis: personal income per capita, in nominal 2006 dollars (linear scale). Solid line: power-law scaling curve (estimated exponent 0.082). Dotted line: logarithmic growth curve. Dashed line: spline fit to logged data.

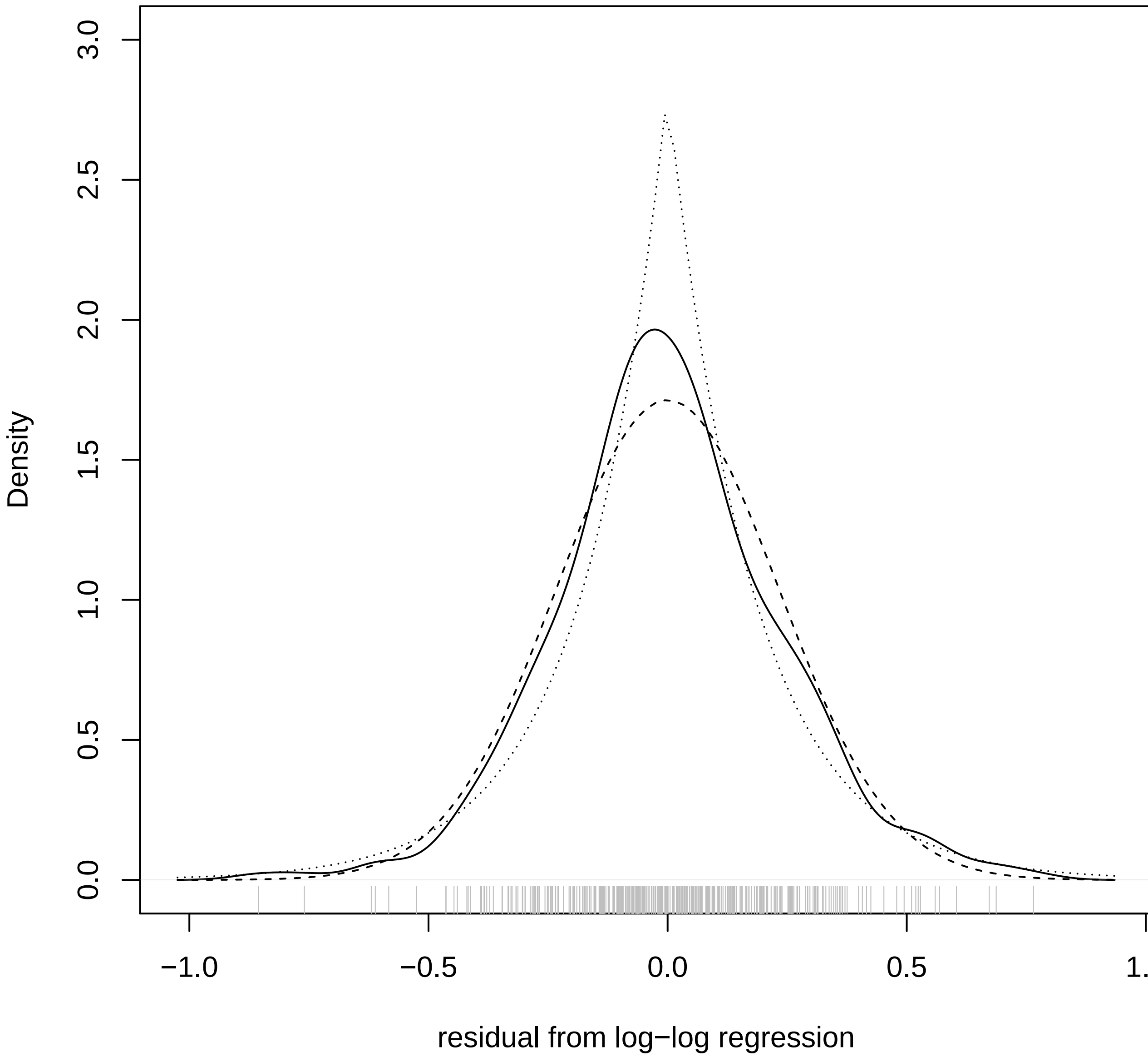

Fig. 8. Horizontal axis: magnitude of residuals from power-law scaling of gross metropolitan product on population, i.e., from regressing $\ln Y$ on $\ln N$. Vertical axis: probability density of the residual distribution. Solid line: Nonparametric kernel density estimate (Gaussian kernel, default bandwidth choice — see Ref. [30]). Dashed line: maximum likelihood Gaussian fit to residuals. Dotted line: maximum likelihood Laplace (doubly-exponential) fit to residuals.

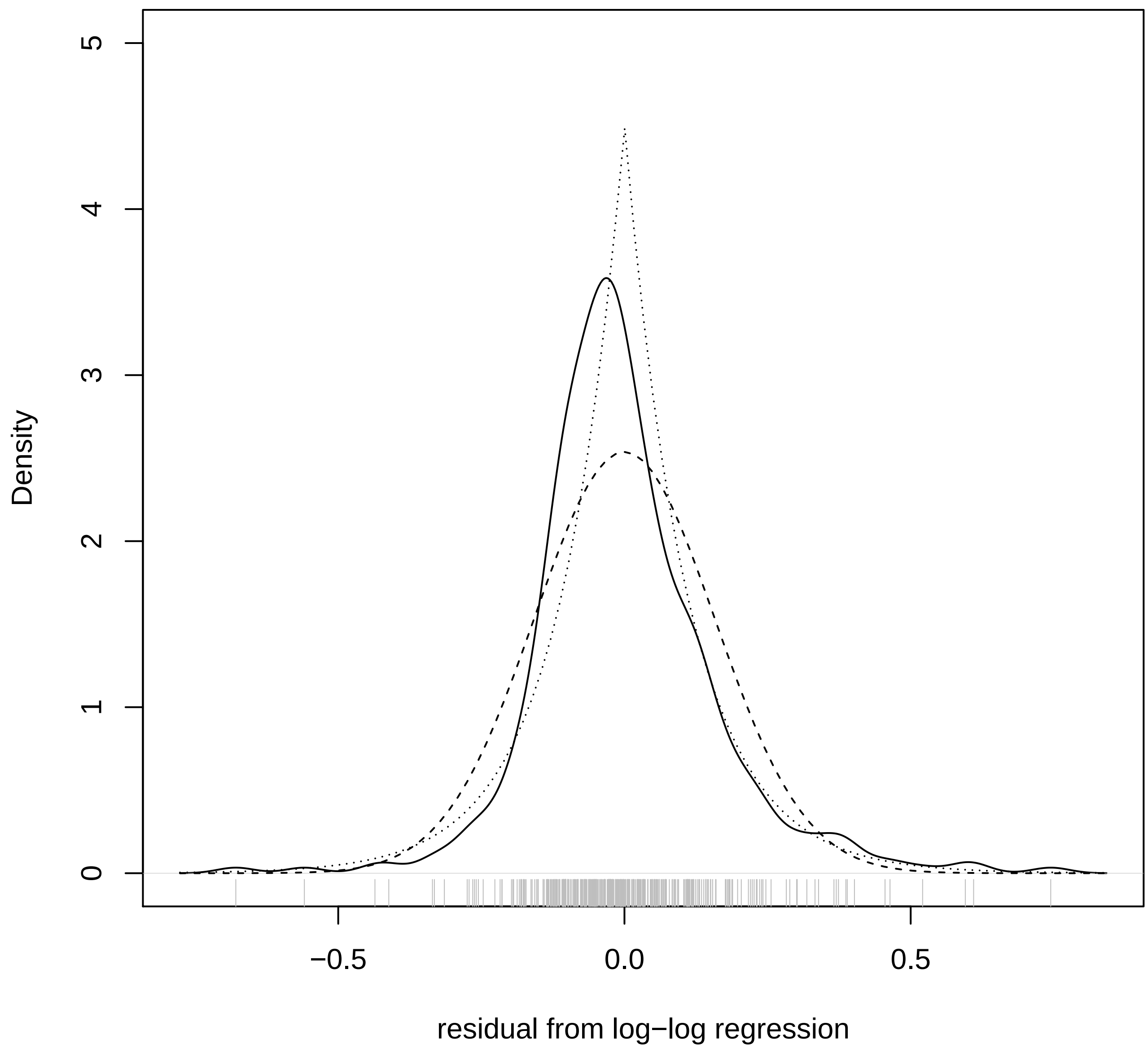

Fig. 9. As in Figure 8, but showing the deviations of personal income from power-law scaling.