

Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling

 Joel E. Cohen^{a,1} and Meng Xu^{a,b}
^aLaboratory of Populations, Rockefeller University and Columbia University, New York, NY 10065; and ^bDepartment of Mathematics and Physics, University of New Haven, West Haven, CT 06516

Contributed by Joel E. Cohen, February 27, 2015 (sent for review November 7, 2014; reviewed by Keiichi Fukaya, Mark Taper, and Ethan P. White)

Taylor's law (TL), a widely verified quantitative pattern in ecology and other sciences, describes the variance in a species' population density (or other nonnegative quantity) as a power-law function of the mean density (or other nonnegative quantity): Approximately, variance = $a(\text{mean})^b$, $a > 0$. Multiple mechanisms have been proposed to explain and interpret TL. Here, we show analytically that observations randomly sampled in blocks from any skewed frequency distribution with four finite moments give rise to TL. We do not claim this is the only way TL arises. We give approximate formulae for the TL parameters and their uncertainty. In computer simulations and an empirical example using basal area densities of red oak trees from Black Rock Forest, our formulae agree with the estimates obtained by least-squares regression. Our results show that the correlated sampling variation of the mean and variance of skewed distributions is statistically sufficient to explain TL under random sampling, without the intervention of any biological or behavioral mechanisms. This finding connects TL with the underlying distribution of population density (or other nonnegative quantity) and provides a baseline against which more complex mechanisms of TL can be compared.

delta method | least-squares regression | skewness | variance function

Taylor's law (TL), named after Taylor (1), relates the variance and the mean of population sizes or population densities of species distributed in space and time by a power-law function:

$$\text{variance} = a(\text{mean})^b \quad [1]$$

or equivalently as a linear function when mean and variance are logarithmically transformed:

$$\log(\text{variance}) = \log(a) + b \times \log(\text{mean}). \quad [2]$$

Eqs. 1 and 2 may be exact if the mean and variance are population moments calculated from certain parametric families of probability distributions (e.g., $a = 1$ and $b = 1$ for a Poisson distribution). Eqs. 1 and 2 may be approximate if the mean and variance are sample moments based on finite random samples of observations. Most empirical tests of TL have not specified the random error associated with Eqs. 1 or 2.

TL has been verified for hundreds of biological species and nonbiological quantities in more than a thousand papers in ecology, epidemiology, biomedical sciences, and other fields (2–4). Recently, examples of TL were found in bacterial microcosms (5, 6), forest trees (7, 8), human populations (9), coral reef fish populations (10), and barnacles (11, 12). TL has been used practically in the design of sampling plans for the control of insect pests of soybeans (13, 14) and cotton (15).

Scientific studies of TL largely focus on the power-law exponent b (or slope b in the linear form), which Taylor believed to contain information about how populations of a species aggregate in space (1). Empirically, b often lies between 1 and 2 (16). Ballantyne and Kerckhoff (17) suggested that individuals' reproductive correlation determines the size of b . Ballantyne (18) proposed that $b = 2$ is a consequence of deterministic population growth. Cohen (19) showed that $b = 2$ arose from exponentially

growing, noninteracting clones. Kilpatrick and Ives (20) proposed that interspecific competition could reduce the value of b . Other models that implied TL were the exponential dispersion model (21–23), models of spatially distributed colonies (24, 25), a stochastic version of logistic population dynamics (16), and the Lewontin–Cohen stochastic multiplicative population model (8). The diversity of empirical confirmations suggests that no specific biological, physical, technological, or behavioral mechanism explains all instances of TL. Such empirical ubiquity suggests that TL could be another of the so-called universal laws (26) like the laws of large numbers (27) and the central limit theorem (28). For example, independently of the present study, Xiao et al. (29) showed numerically (not analytically) that random partitions and compositions of integers led to TL with slopes often between 1 and 2, as commonly observed in empirical examples of TL.

The present work was kindred in spirit and intent, although distinct in technical approach and results. Here we demonstrated that, when independently and identically distributed (iid) observations are sampled in blocks (not necessarily of equal size) from any nonnegative-valued skewed probability distribution with four finite moments, TL arises. We do not assert this is the only way TL arises. If these conditions are not satisfied in an empirical confirmation of TL, other mechanisms are likely to be in play. Under these assumptions (iid sampling in blocks from a skewed probability distribution with four finite moments), we derived analytically the explicit approximate formulae for the TL slope (b in Eq. 2), intercept [$\log(a)$ in Eq. 2], and standard error (SE) of the slope estimator [$s(b)$, see *Theorem in Results*]. In simulated random samples from probability distributions, these theoretical formulae approximated well the TL parameters. An empirical example using basal area densities of red oak trees in a temperate forest showed that our theory explained some published

Significance

One of the most widely confirmed empirical patterns in ecology is Taylor's law (TL): The variance of population density is approximately a power-law function of the mean population density. We showed analytically that, when observations are randomly sampled in blocks from a single frequency distribution, the sample variance will be related to the sample mean by TL, and the parameters of TL can be predicted from the first four moments of the frequency distribution. The estimate of the exponent of TL is proportional to the skewness of the distribution. Random sampling suffices to explain the existence and predict the parameters of TL in well-defined circumstances relevant to some, but not all, published empirical examples of TL.

Author contributions: J.E.C. and M.X. designed research; J.E.C. and M.X. performed research; J.E.C. contributed new reagents/analytic tools; J.E.C. and M.X. analyzed data; and J.E.C. and M.X. wrote the paper.

Reviewers: K.F., Institute of Statistical Mathematics; M.T., Montana State University; and E.P.W., Utah State University.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed. Email: cohen@rockefeller.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1503824112/-DCSupplemental.

estimates of the TL slope when the assumptions of the theory were satisfied, and also successfully predicted the TL slope when the assumptions of the theory were shown to be mildly violated. Our results showed that TL may arise without any complicated mechanisms, and provided simple assumptions (skewed and iid observations) against which empirical applications of TL can be tested.

Results

Analytical Results. Suppose X is a nonnegative real-valued random variable with cumulative distribution function F , finite mean $E(X) = M > 0$, finite variance (or second central moment) $var(X) \equiv E(X^2) - [E(X)]^2 = V > 0$, and finite third and fourth central moments $E[(X - M)^h] = \mu_h$, $h = 3, 4$. Consider $N > 2$ “blocks” or sets of iid observations (random samples) of X . Let x_{ij} denote observation i of block j , $i = 1, \dots, n_j$, assuming the number of observations in block j satisfies $n_j > 3$, $j = 1, \dots, N$. The total number of observations is $n_1 + n_2 + \dots + n_N$. For block j the sample mean of observations and the expectation and variance of the sample mean are, respectively, $m_j = (x_{1j} + \dots + x_{n_j j})/n_j$, $E(m_j) = M$, $var(m_j) = V/n_j$. The unbiased sample variance of observations in block j and its expectation and variance are, respectively,

$$v_j = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} x_{ij}^2 - \frac{n_j}{n_j - 1} m_j^2,$$

$$E(v_j) = V, \quad var(v_j) = \frac{1}{n_j} \left(\mu_4 - \frac{n_j - 3}{n_j - 1} V^2 \right).$$

The formula for $var(v_j)$ is from Neter et al. (30). As $n_j \rightarrow \infty$, $Prob\{m_j = 0\} \rightarrow 0$ and $Prob\{v_j = 0\} \rightarrow 0$ by Chebyshev’s tail inequality (31). We assume that n_j is large enough that $m_j > 0$ and $v_j > 0$.

In this theory, the variation between blocks in the sample mean and the variation between blocks in the sample variance are small because every block is randomly (iid) sampled from the same distribution. Because any two smoothly varying functions can be locally linearly related, the logarithm of the sample variance of a block can be approximated as a linear function of the logarithm of the sample mean of that block. The theorem below quantifies this qualitative observation.

In empirical examples, if the variation of sample means among blocks is too large to arise from random sampling alone (e.g., if ANOVA rejects homogeneity of block means), then the assumptions of the theorem do not apply, and it remains to be determined empirically whether the theorem’s conclusions apply, as if the theorem’s assumptions were close enough to reality. We give empirical examples, based on published data, where the theorem’s assumptions do, and do not, hold and the conclusions do hold in both situations.

By definition, the coefficient of variation of X is $CV = V^{1/2}/M$, the skewness is $\gamma_1 = \mu_3/V^{3/2}$, and the kurtosis is $\kappa = \mu_4/V^2$. Most empirical tests of TL estimated the intercept $\log(a)$ and the slope b of TL using ordinary least-squares regression of $\log(v_j)$ as the dependent variable and $\log(m_j)$ as the independent variable, and we analyze this practice. We do not assume that the sample mean is an error-free estimate of the population mean: the theorem deals with a relation among sample moments.

Definition: Suppose a random variable Y is a function of a random sample of size n from a distribution F , and suppose the expectation $E(Y)$ exists. Then the expression $Y \approx K$, where K is a constant independent of the random sample, is defined to mean that, for some $p > 0$, $E(Y) = K + o(n^{-p})$, where $o(n^{-p})$ refers (in Landau’s little- o notation) to any function $f(n)$ such that $\lim_{n \rightarrow \infty} f(n)/n^{-p} = 0$ (e.g., see ref. 31, p. 121).

Theorem. Suppose the nonnegative real-valued random variable X has finite first four moments, with strictly positive mean and strictly positive variance. Suppose that $n_j > 3$ observations x_{ij} ($i = 1, \dots, n_j$) of X are randomly assigned to block j ($j = 1, \dots, N$), $N > 2$, and all

of the observations, which number $\sum_{j=1}^N n_j$ in total, are iid. Let m_j, v_j be the sample mean and the sample variance, respectively, of the n_j observations in block j , and suppose n_j is large enough that m_j and v_j are strictly positive. Let \hat{b} and $\widehat{\log(a)}$ denote the least-squares estimators, respectively, of b and $\log(a)$ in TL, namely, $\log(v_j) = \log(a) + b \times \log(m_j)$, $j = 1, \dots, N$ (Eq. 2). Let $s(\hat{b})$ denote the SE of the least-squares slope estimator \hat{b} . Then, in the limit of large N and large n_j ,

$$\hat{b} \approx \frac{cov(m_j, v_j)}{MV} \bigg/ \frac{var(m_j)}{M^2} = \mu_3 M / V^2 = \gamma_1 / CV, \quad [3]$$

$$\widehat{\log(a)} \approx \log V - \frac{\gamma_1}{CV} \cdot \log M, \quad [4]$$

$$s(\hat{b}) \approx \sqrt{\frac{M^2(\mu_4 V - V^3 - \mu_3^2)}{(N-2)V^4}} = \sqrt{\frac{\kappa - 1 - \gamma_1^2}{(N-2)(CV)^2}}. \quad [5]$$

Proof of this Theorem is given in SI Text. Because $CV > 0$, Eq. 3 shows that random sampling in blocks of any right-skewed distribution (one with $\gamma_1 > 0$) generates a positive TL slope.

Squaring both sides of Eq. 5 yields the estimated variance of \hat{b} . Because any variance is nonnegative, the numerator of the variance estimate ($\kappa - 1 - \gamma_1^2$) is nonnegative. Eq. 5 thus provides an alternative proof and adds a new interpretation of the inequality $\kappa - 1 - \gamma_1^2 \geq 0$ that was obtained by Rohatgi and Székely (32).

Numerical Simulations. We illustrate our theory of TL using six probability distributions, five of which are positively skewed. We created six square matrices (here $N = n$) to mimic the blocks commonly found in ecological field data. Each column can be viewed as a block containing n observations (rows). For each matrix, we plotted (Fig. 1) the log of the sample variance v_j of each column j on the ordinate against the log of the sample mean m_j on the abscissa, $j = 1, \dots, N$.

For each of the five positively skewed distributions, an approximately linear relationship with positive slopes was observed (Fig. 1A–E). The lognormal slope (Fig. 1E) was larger than most estimates observed in ecological applications. For the shifted normal distribution, which had zero skewness, no relationship between the log sample variance and the log sample mean was observed (i.e., analytically $b = 0$ and numerically and by regression $\hat{b} = 0$) (Fig. 1F).

To illustrate our Theorem numerically, we applied the theoretical formulae (Eqs. 3–5) to each of the six probability distributions and analytically computed the predicted values of the slope and intercept in Eq. 2, and the SE of the slope estimator. The first four moments used in the formulae are standard results for these distributions. For each distribution, we also generated 10,000 random copies of the n ($= 100$) by N ($= 100$) matrix and fitted a linear regression and a quadratic regression to the $\log(m_j)$ and $\log(v_j)$, $j = 1, \dots, N$ from each random copy. We obtained an approximate sampling distribution for each parameter of TL and for the quadratic coefficient c in the hypothetical quadratic relationship $\log(\text{variance}) = \log(a) + b \times \log(\text{mean}) + c \times [\log(\text{mean})]^2$. Medians and 95% confidence intervals (CIs) of TL parameters, SE of the slope estimator, and the quadratic coefficient were calculated respectively from the 50, 2.5, and 97.5% quantiles of the sampling distribution of the corresponding linear and quadratic regression point estimates. To test the robustness of our theory, the $n \times N$ observations in each matrix were used to calculate sample estimates of the first four moments of the corresponding probability distribution, as if the first four moments were not known a priori but were based on a sample. These estimates were then plugged into the formulae (Eqs. 3–5) to evaluate the theoretical TL slope, intercept, and SE of the slope estimator. Their medians and 95% CIs were similarly

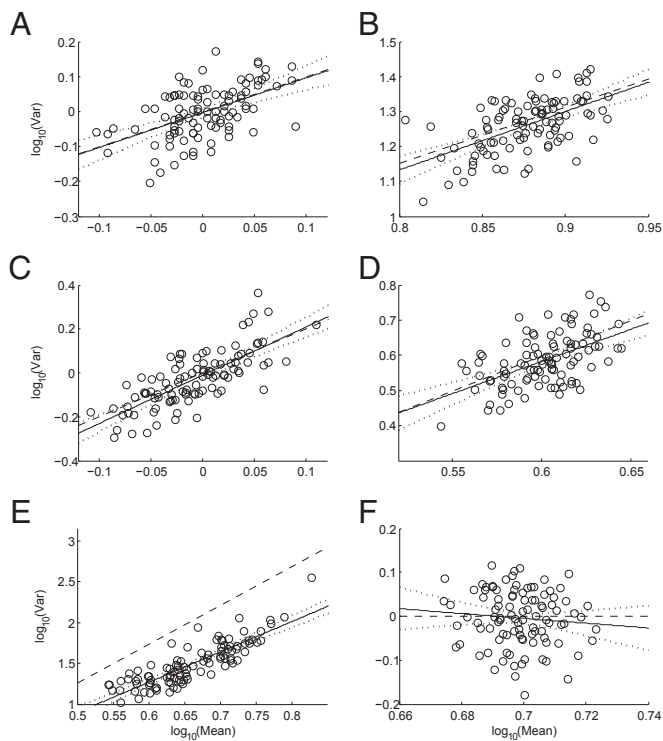


Fig. 1. Taylor's law with positive slope arises from random samples from a single (A) Poisson ($\lambda = 1$), (B) negative binomial ($r = 5, p = 0.4$), (C) exponential ($\lambda = 1$), (D) gamma ($\alpha = 4, \beta = 1$), and (E) lognormal ($\mu = 1, \sigma = 1$) distribution, but not from a (F) shifted normal [$5 + \mathcal{N}(0,1)$] distribution [i.e., a $\mathcal{N}(0,1)$ distribution with 5 added to each value to make each block's mean positive with high probability]. For each panel, 10,000 iid observations from the selected distribution were arranged randomly in a square matrix with $n = 100$ rows and $N = 100$ columns. For each column j , the sample mean m_j and the sample variance v_j were calculated and plotted on log-log coordinates using open circles, $j = 1, \dots, N$. The solid black line is the least-squares linear regression $\log_{10} v_j = \log_{10} a + b \log_{10} m_j$. The dotted curved lines above and below the solid black line give the 95% confidence interval of the regression line. Slope and intercept of the dashed black line were computed analytically from Eqs. 3 and 4, respectively (Table 1), using the population moments used to generate the observations, not the sample moments of the observations. Population skewness in each distribution is 1 (Poisson), 0.9238 (negative binomial), 2 (exponential), 1 (gamma), 6.1849 (lognormal), and 0 (shifted normal).

calculated from the 10,000 random copies of the matrix drawn from the true distribution. Estimates from the regression were compared with the corresponding theoretical predictions computed from the formulae analytically and numerically (Table 1 and Figs. S1–S3).

The mean, variance, third and fourth central moments, computed analytically using the given parameters, are respectively 1, 1, 1, and 4 for Poisson ($\lambda = 1$), 7.5, 18.75, 75, and 1,495.3125 for negative binomial ($r = 5, p = 0.4$), 1, 1, 2, and 9 for exponential ($\lambda = 1$), 4, 4, 8, and 72 for gamma ($\alpha = 4, \beta = 1$), ~ 4.4817 , 34.5126, 1,254.0009, and 135,711.9683 for lognormal ($\mu = 1, \sigma = 1$), and 5, 1, 0, and 3 for shifted normal [$5 + \mathcal{N}(0,1)$] distributions. Except for the shifted normal distribution, a positive slope estimate b was observed when a linear regression was fitted to the independent variable log mean and dependent variable log variance. In all cases except the shifted normal distribution, the 95% CI of b under regression was on the right side of zero. The 95% CI of b under regression for the shifted normal contained zero and therefore a linear relationship between log mean and log variance was not observed. These findings were consistent with Fig. 1. The 95% CI of the quadratic coefficient c from quadratic regression contained zero in all six distributions,

so there was no statistically significant evidence that quadratic regression provided a better model than linear regression when describing the relationship between log variance and log mean. Therefore, TL was confirmed for each of the five skewed probability distributions.

Except for the lognormal distribution, the theoretical values of b (Fig. S1) and $\log(a)$ (Fig. S2) predicted analytically from Eqs. 3 and 4, and the SE of the slope estimator (Fig. S3) calculated from Eq. 5, fell within the corresponding 95% CI from linear regression. In the lognormal distribution, the analytical predictions of the slope b and the SE of its estimator were on the right side of the corresponding 95% CI from regression, meaning that the theoretically predicted values were significantly larger and more variable than those estimated from linear regression. Under the more robust calculations using random copies of $n \times N$ iid samples, for each combination of probability distribution and parameter, the 95% CI of the parameter from the theoretical formulae and from the regression overlapped.

Empirical Data. The basal area density of red oaks (*Quercus rubra*, abbreviated as RO) in Black Rock Forest (BRF) illustrates empirically that random sampling of iid data can generate TL, and that the TL parameters and their CIs calculated from least-squares linear regression using random samples agree with the corresponding values predicted analytically using our formulae. Moreover, four empirical methods of grouping observations into blocks give estimates of the TL slope that are not statistically distinguishable from the estimates of TL given by our random-sampling theory. The complete data on which this example is based were published and analyzed for other purposes (33). Measuring population density by basal area density, rather than stems per unit area, is widespread in forestry and is explained in ref. 33.

BRF is a 1,550-ha forest preserve in Cornwall, New York (34). In a 1985 forest-wide survey, 218 sampling points were randomly designated to sample the basal area density of tree species. Each forest location was equally likely to be selected as a sampling point with no repeated measurements at any sampling point (33). The 218 measurements of basal area density could reasonably be interpreted as representing an iid sample of each tree species' basal area density in BRF in 1985.

We tested TL using the basal area density data of RO because RO was the most dominant tree species in the 1985 survey (32.72% of all 2,078 stems sampled) and served as a biological indicator of the forest composition and timber production (Fig. 2E). Taylor et al. (35) argued that when testing TL, the number of blocks should be at least 5 and the number of observations per block should be at least 15. Following this practice, we randomly assigned the 218 observations into 14 blocks (15 observations in each of the first 13 blocks and 23 observations in the 14th block) and computed the means and variances of RO basal area density across the observations within each block. The sample size per block here, namely, 15 observations except in one block with 23 observations, is smaller than the sample size per block in the numerical simulations, namely, 100, so we expected greater variability in the sample mean and sample variance of each block for the tree data. We then fitted an ordinary least-squares regression of log variance of each block as a linear function of the log mean of the block and obtained point estimates for the slope and the intercept, and the SE of the slope estimator. Repeatedly randomizing the assignment of observations into blocks 10,000 times, we calculated the median and 95% percentile CI of the slope, intercept, and SE of the slope estimator, respectively, from the corresponding 10,000 regression point estimates (Fig. 2A–C). In this empirical application, the true underlying distribution was unknown, so we randomized the sample of observations. To check for nonlinearity between log mean and log variance, we also fitted a quadratic regression under each random assignment of observations to blocks and calculated the median and 95% CI of the quadratic coefficient from the corresponding 10,000 quadratic regression point estimates.

Table 1. Estimates of the slope (*b*), intercept [log(*a*)], and SE of the slope estimator in Taylor's law using the theoretical formulae Eqs. 3–5 and linear regression for six probability distributions

Probability distribution	<i>b̂</i>			log(<i>â</i>)			<i>s</i> (<i>b̂</i>)			Quadratic coefficient
	Formula (analytic)	Formula (numeric)	Regression	Formula (analytic)	Formula (numeric)	Regression	Formula (analytic)	Formula (numeric)	Regression	
Poisson ($\lambda = 1$)	1.0000	0.9976 (0.9458, 1.0551)	1.0027 (0.7211, 1.2775)	0.0000	-0.0001 (-0.0119, 0.0118)	-0.0043 (-0.0164, 0.0076)	0.1429	0.1424 (0.1357, 0.1508)	0.1416 (0.1157, 0.1738)	0.0550 (-4.9482, 4.9072)
Negative binomial ($r = 5, p = 0.4$)	1.6000	1.5972 (1.4860, 1.7213)	1.6017 (1.0729, 2.1367)	-0.1271	-0.1250 (-0.2340, -0.0263)	-0.1351 (-0.6023, 0.3312)	0.2711	0.2701 (0.2573, 0.2882)	0.2703 (0.2214, 0.3322)	0.2370 (-16.0949, 16.6441)
Exponential ($\lambda = 1$)	2.0000	1.9929 (1.8709, 2.1518)	1.9972 (1.6235, 2.3849)	0.0000	-0.0001 (-0.0174, 0.0174)	-0.0123 (-0.0288, 0.0042)	0.2020	0.1990 (0.1812, 0.2313)	0.1920 (0.1560, 0.2352)	0.0332 (-6.4607, 7.0247)
Gamma ($\alpha = 4, \beta = 1$)	2.0000	1.9957 (1.8562, 2.1496)	2.0011 (1.3760, 2.6237)	-0.6021	-0.5995 (-0.6928, -0.5140)	-0.6096 (-0.9848, -0.2312)	0.3194	0.3180 (0.3019, 0.3411)	0.3178 (0.2607, 0.3900)	-0.0815 (-22.7731, 22.7344)
Lognormal ($\mu = 1, \sigma = 1$)	4.7183	4.0982 (3.2918, 7.4927)	3.5991 (3.0485, 4.2296)	-1.0970	-1.1320 (-3.2884, -0.6054)	-0.8815 (-1.2848, -0.5294)	0.6660	0.4155 (0.2880, 0.9895)	0.2662 (0.2132, 0.3305)	3.6911 (-3.7832, 12.3419)
Shifted normal [$5 + \mathcal{N}(0,1)$]	0.0000	-0.0009 (-0.2407, 0.2386)	0.0011 (-1.4290, 1.4273)	0.0000	-0.0006 (-0.1659, 0.1694)	-0.0062 (-1.0024, 0.9936)	0.7143	0.7140 (0.6946, 0.7345)	0.7249 (0.5933, 0.8843)	-0.1759 (-128.0845, 124.9325)

Each parameter was first predicted analytically from the corresponding formula using the given distribution parameters [Formula (analytic)], then approximated using the $n \times N$ random observations of each distribution from the formulae [Formula (numeric)] and from the regression (Regression) separately. For the last two methods, median and 95% CI of each parameter were calculated from 10,000 random copies of the $n \times N$ iid observations (95% CI is given below the associated median value). For each distribution, the median and 95% CI of the quadratic coefficient from the least-squares quadratic regression were similarly calculated from the 10,000 random copies of the $n \times N$ iid observations.

Eq. 2 held with median slope 0.8391 and 95% CI (0.0146, 1.5975) and median intercept 0.4196 and 95% CI (0.0469, 0.8335). The median of the SE of the slope estimator was 0.4045 with 95% CI (0.2257, 0.7272). Quadratic fitting did not indicate statistically significant nonlinearity in the relationship between log mean and log variance: the median quadratic coefficient was -1.0665 and 95% CI was (-11.0598, 8.4996). Thus, TL held for RO basal area density with positive slope and positive intercept under random assignment of observations to blocks. In linear regressions fitted to repeatedly randomly assigned observations, 50% of the coefficients of determination (R^2) fell below 0.3019 and 19.81% of the R^2 fell above 0.5. This example of our theory did not account for the large R^2 of TL observed in some empirical data (36). Whether the observed positive intercept is due to measurement error, sampling scale, environmental variation in habitat suitability, or biological interactions of RO with conspecifics or other species remains to be determined.

We computed the sample estimates of the mean (3.1193), variance (7.0917), skewness (0.6435), and kurtosis (2.5550) of RO density from the 218 observations. From the theoretical formulae (Eqs. 3–5), the predicted slope, predicted intercept, and SE of the slope estimator were, respectively, 0.7537, 0.4784, and 0.3230, all of which were comparable with the corresponding median values and fell within the corresponding 95% CI calculated from point estimates under linear regression (Fig. 2 A–C). Our theory provided a reasonable estimate of the TL parameters for skewed biological field observations randomly grouped into blocks.

We also compared the TL slope estimated from random grouping in blocks with the published TL slopes estimated from four biological methods of grouping (ref. 33, tables S1–S4). In summary, all four point estimates of the slope of TL under the four biological groupings fell within the 95% CI of the slope under random assignment of sampling points to blocks, and all four 95% CIs of the slope under the biological groupings estimated from normal theory heavily overlapped the 95% CI of the slope under random assignment of sampling points to blocks.

In detail, for Friday's grouping, the point estimate of the slope, 0.9854, fell within the 95% CI (0.0146, 1.5975) from the random grouping of sampling points into blocks, and the 95% confidence interval of the slope of TL under Friday's grouping, (0.0552, 1.9156), heavily overlapped the 95% CI under random grouping.

Under Schuster's grouping, the point estimate of the slope, 0.9316, again fell within the 95% CI (0.0146, 1.5975) from random grouping and the 95% CI, (0.6940, 1.1692), of the slope of TL from Schuster's method fell entirely within that of random grouping.

Under the watershed grouping, the point estimate of the TL slope, 0.6234, again fell within the 95% CI (0.0146, 1.5975) from random grouping, and the 95% CI of the slope of TL under the watershed grouping, (-0.2666, 1.5133), almost contained the 95% CI under random grouping.

Finally, under the topography grouping, the point estimate of the slope of TL, 0.2603, again fell within the 95% CI (0.0146, 1.5975) from random grouping and the 95% CI, (-0.8830, 1.4037), again almost contained the 95% CI under random grouping.

The random sampling model of TL would explain the agreement between the slope from random grouping and the slopes from the four biological groupings if the model's assumption of iid sampling within and across all blocks were valid. To test that assumption, we did an ANOVA of the mean basal area density by block, for each method (Fig. 3). For Friday's, Schuster's, and watershed groupings, the null hypothesis that all blocks had equal means was rejected ($P = 0.014$, $P < 0.001$, and $P = 0.009$, respectively), contrary to the random sampling model. Under the topography grouping, the mean basal area density did not differ significantly from one block to another ($P = 0.115$).

This example shows that the random sampling model can predict the exponent of TL even when some of its assumptions are violated or some other mechanisms are in play. How robust the predictions are with respect to violations of the assumptions is a question for future theoretical and empirical research.

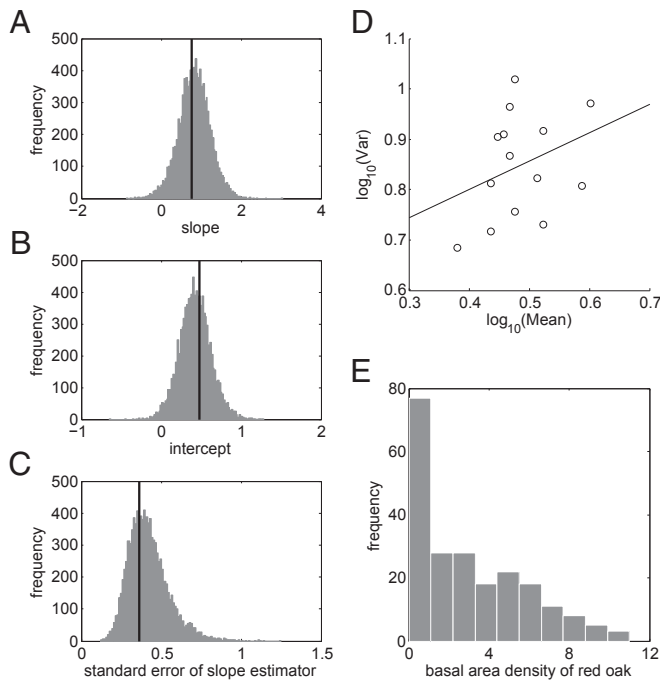


Fig. 2. Testing TL using basal area density of RO in BRF. (A–C) Histograms of the slope, intercept, and SE of the slope estimator, respectively, estimated by regression from 10,000 random assignments of observations into blocks, with the theoretically predicted values marked by the solid vertical lines. (D) A least-squares regression (solid line) of the dependent variable $\log(\text{variance})$ as a linear function of the independent variable $\log(\text{mean})$ under one realization of random groupings. Each open circle represents a mean and a variance calculated over observations within a single block. (E) The histogram of basal area density of RO at 218 sampling points is right-skewed.

Discussion

Our results show that random sampling of a distribution in blocks leads to TL. Moreover, the first four moments of the distribution and the number of blocks predict the TL parameters and the SE of the slope estimator. No biological or physical mechanisms need be invoked to explain TL under this form of sampling. In the empirical example of RO, the 95% CI of the TL slope calculated from our theory (0.0146, 1.5975) overlapped the range of 1–2 commonly but not universally observed in ecological data (36). Our examples show that this model has relevance to some, but not all, published empirical examples of TL.

Our random sampling model does not purport to be a universal explanation of TL in all or most circumstances. For example, when the mean population densities in large samples of different species of widely different body masses range over seven or more orders of magnitude (37), the differences in mean and variance of population density probably cannot be attributed to random sampling variation from a single underlying distribution. However, when the mean population densities range over little more than one order of magnitude (ref. 11, p. 12, figure 7), the invariance of TL parameters under different regimes of population dynamics might be accounted for by our sampling model.

In our numerical examples, the discrepancy between the theoretical prediction and the regression estimate of TL slope b under random sampling was largest for the lognormal distribution, which also had the least realistic values of \hat{b} (Fig. S1E). A possible reason is that $s(\hat{b}) = 0.6660$ for the lognormal distribution was twice as large as $s(\hat{b})$ for any of the other four skewed distributions, the second largest being 0.3194 for the gamma distribution (Table 1), whereas the sample sizes for all of the distributions were the same $n = 100$. In addition, because the fourth moment of lognormal distribution grows exponentially as a function of the parameter σ^2 , our estimates of the variance for

the lognormal distribution were likely to be least reliable among the estimates for the skewed distributions [see formula for $\text{var}(v_j)$ in Results, Analytical Results]. Among tested distributions, the fourth central moment of the lognormal distribution was at least 90 times the fourth moment of any other distribution. Evidently, in the lognormal example, we did not simulate enough linear regressions to sample adequately the full range of variation of the parameters. Nevertheless, when applied to the 10,000 random copies of $n \times N$ lognormal observations, our formula provided a robust theoretical estimate of b compatible with that from the regression (Table 1).

Previous works have analyzed TL in relation to frequency distributions. For example, Taylor (2) observed that insect populations at progressively higher densities conformed to different frequency distributions (e.g., Poisson, negative binomial, and lognormal) with identical slope parameter b , but he did not explain why TL arises from these distributions. Our results connect TL with the underlying probability distribution but do not explain why the distribution of observations (e.g., Fig. 2E) was right-skewed. Future studies on TL and other general empirical scaling patterns should give attention to the role of population distributions in understanding these patterns.

The usefulness of TL in inferring biological information about population aggregations is a subject of continuing scientific debate. Alternative mean–variance relationships have been proposed as competitors of TL (25, 38, 39). It has been argued that sampling error and sampling coverage may lead to TL-like patterns as statistical artifacts (40) and to substantially biased TL parameters (41). Our results offer another statistical mechanism that leads to TL.

Methods

Throughout, $\log = \log_{10}$. P represents the P value and α is the significance level of any hypothesis testing (except when α is used as a parameter of the gamma distribution).

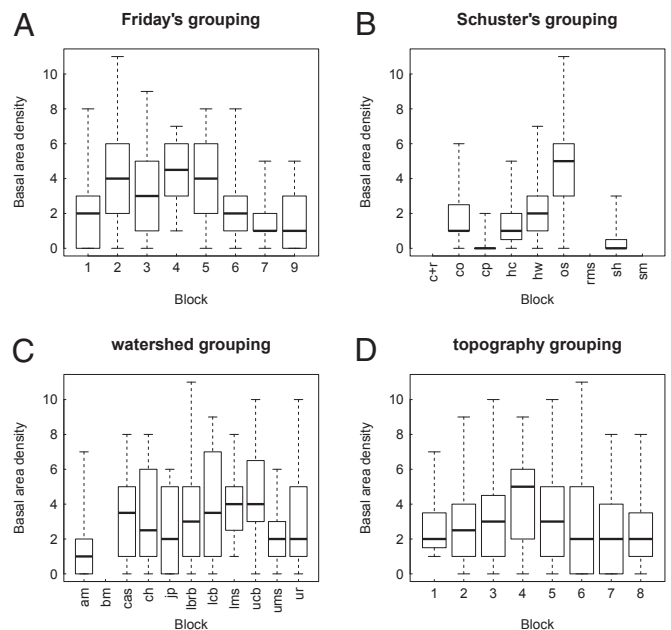


Fig. 3. Boxplots of basal area density of RO in BRF, according to four biological methods of assigning plots to blocks. In each boxplot, the median is the bold black bar, the box covers the interquartile range, and the whiskers cover the entire range of basal area density within a block. One-way unbalanced ANOVA tests of the null hypothesis of no difference between blocks in mean basal area density rejected the null hypothesis ($P < 0.05$) for all grouping methods except for the topography grouping. (A) Friday's grouping. (B) Schuster's grouping. (C) Watershed grouping. (D) Topography grouping.

Traditionally, when tested against empirical data, TL has been taken to be confirmed if the fitted linear regression Eq. 2 had statistically significantly nonzero linear coefficient b (with $P < \alpha$; here $\alpha = 0.05$), and if a least-squares quadratic regression between the independent variable $\log(\text{mean})$ and dependent variable $\log(\text{variance})$ did not yield a statistically significant quadratic coefficient c ($P > \alpha$). The use of the doubly logarithmic scale in the testing of TL and other bivariate allometric relationships (e.g., scaling of metabolic rate with body mass) has been questioned (39, 42–44) and defended (45, 46).

Our numerical examples combined the ordinary least-squares regression approach with random sampling. Specifically, in multiple realizations, we sampled from a single probability distribution, grouped observations into blocks, calculated the mean and the variance of observations per block, recorded the parameters and quadratic coefficient estimates from the corresponding linear and quadratic regressions (47, p. 155), respectively, for each realization, and constructed CIs of the parameters using percentiles of the corresponding approximate sampling distributions obtained from all realizations.

In the empirical example of RO trees, we randomly grouped observations into blocks and obtained TL parameters from linear regression fitted to repeatedly randomly grouped samples. We also analyzed biologically based groupings (33) of these trees that gave rise to TL. Linear and quadratic

regressions were performed using the MATLAB (The MathWorks, Inc.) function “regress.”

The analytical formulae for the TL parameter estimators and the SE of the slope estimator were derived using the delta method (48, 49). The delta method, which is commonly used by statisticians, relies on Taylor series expansions (not the same Taylor as in Taylor’s law) for moments of functions of random variables. To implement the delta method we relied on a moment estimate of the difference between population mean and sample mean by Loève (50) and the consistency of sample estimators (*SI Text*). The delta method is increasingly accurate as the variation around the point of expansion becomes smaller. Because the variation in sample means and sample variances is small when sufficiently large random samples are blocked, the delta method yields quite accurate approximations to TL parameters estimated from linear regression.

ACKNOWLEDGMENTS. The authors thank William S. F. Schuster for data from Black Rock Forest; Richard Chandler, Russell Millar, William S. F. Schuster, Nadav Shnerb, Andrew Wood, and three referees for helpful comments; and Priscilla K. Rogerson for assistance. This work was partially supported by National Science Foundation Grants EF-1038337 and DMS-1225529.

- Taylor LR (1961) Aggregation, variance and the mean. *Nature* 189:732–735.
- Taylor LR (1984) Assessing and interpreting the spatial distributions of insect populations. *Annu Rev Entomol* 29:321–357.
- Kendal WS (2002) A frequency distribution for the number of hematogenous organ metastases. *J Theor Biol* 217(2):203–218.
- Eisler Z, Bartos I, Kertész J (2008) Fluctuation scaling in complex systems: Taylor’s law and beyond. *Adv Phys* 57:89–142.
- Ramsayer J, Fellous S, Cohen JE, Hochberg ME (2012) Taylor’s Law holds in experimental bacterial populations but competition does not influence the slope. *Biol Lett* 8(2):316–319.
- Kaltz O, Escobar-Páramo P, Hochberg ME, Cohen JE (2012) Bacterial microcosms obey Taylor’s law: Effects of abiotic and biotic stress and genetics on mean and variance of population density. *Ecol Process* 1:5.
- Cohen JE, Xu M, Schuster WSF (2012) Allometric scaling of population variance with mean body size is predicted from Taylor’s law and density-mass allometry. *Proc Natl Acad Sci USA* 109(39):15829–15834.
- Cohen JE, Xu M, Schuster WSF (2013a) Stochastic multiplicative population growth predicts and interprets Taylor’s power law of fluctuation scaling. *Proc Biol Sci* 280(1757):20122955.
- Cohen JE, Xu M, Brunborg H (2013b) Taylor’s law applies to spatial variation in a human population. *Genus* 69(1):25–60.
- Mellin C, Huchery C, Caley MJ, Meehan MG, Bradshaw CJA (2010) Reef size and isolation determine the temporal stability of coral reef fish populations. *Ecology* 91(11):3138–3145.
- Fukaya K, et al. (2013) Variable processes that determine population growth and an invariant mean-variance relationship of intertidal barnacles. *Ecosphere* 4(4):1–20.
- Fukaya K, Okuda T, Nakaoka M, Noda T (2014) Effects of spatial structure of population size on the population dynamics of barnacles across their elevational range. *J Anim Ecol* 83(6):1334–1343.
- Kogan M, Ruesink WG, McDowell K (1974) Spatial and temporal distribution patterns of the bean leaf beetle, *Cerotoma trifurcata* (Forster), on soybeans in Illinois. *Environ Entomol* 3(4):607–617.
- Bechinski EJ, Pedigo LP (1981) Population dispersion and development of sampling plans for *Orius insidiosus* and *Nabis* spp. in soybeans. *Environ Entomol* 10(6):956–959.
- Wilson LT, Sterling WL, Rummel DR, DeVay JE (1989) Quantitative sampling principles in cotton. *Integrated Pest Management Systems and Cotton Production*, eds Frisbie RE, El-Zik KM, Wilson LT (Wiley, New York), pp 85–119.
- Maurer BA, Taper ML (2002) Connecting geographical distributions with population processes. *Ecol Lett* 5:223–231.
- Ballantyne F, Kerkhoff AJ (2007) The observed range for temporal mean-variance scaling exponents can be explained by reproductive correlation. *Oikos* 116:174–180.
- Ballantyne F (2005) The upper limit for the exponent of Taylor’s power law is a consequence of deterministic population growth. *Evol Ecol Res* 7(8):1213–1220.
- Cohen JE (2013) Taylor’s power law of fluctuation scaling and the growth-rate theorem. *Theor Popul Biol* 88:94–100.
- Kilpatrick AM, Ives AR (2003) Species interactions can explain Taylor’s power law for ecological time series. *Nature* 422(6927):65–68.
- Jørgensen B (1987) Exponential dispersion models. *J R Stat Soc, B* 49:127–162.
- Jørgensen B (1997) *The Theory of Dispersion Models* (Chapman & Hall, London).
- Kendal WS, Jørgensen B (2011) Taylor’s power law and fluctuation scaling explained by a central-limit-like convergence. *Phys Rev E Stat Nonlin Soft Matter Phys* 83(6 Pt 2):066115.
- Yamamura K (1990) Sampling scale dependence of Taylor’s power law. *Oikos* 59(1):121–125.
- Yamamura K (2000) Colony expansion model for describing the spatial distribution of population. *Popul Ecol* 42:160–169.
- Tao T (2012) *E pluribus unum: From complexity, universality*. *Daedalus* 141(3):23–34.
- Révész P (1968) *The Laws of Large Numbers* (Academic, New York).
- Fischer H (2011) *A History of the Central Limit Theorem from Classical to Modern Probability Theory* (Springer, New York).
- Xiao X, Kenneth JL, Ethan PW (2014) A process-independent explanation for the general form of Taylor’s Law. arXiv:1410.7283 [q-bio.PE].
- Neter J, Wasserman W, Kutner MH (1990) *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs* (Irwin, Homewood, IL), 3rd Ed.
- Steele JM (2004) *Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities* (Cambridge Univ Press, Cambridge, UK).
- Rohatgi VK, Székely GJ (1989) Sharp inequalities between skewness and kurtosis. *Stat Probab Lett* 8:297–299.
- Xu M, Schuster WSF, Cohen JE (2015) Robustness of Taylor’s law under spatial hierarchical groupings of forest tree samples. *Popul Ecol*, 10.1007/s10144-014-0463-0.
- Schuster WSF, et al. (2008) Changes in composition, structure and aboveground biomass over seventy-six years (1930–2006) in the Black Rock Forest, Hudson Highlands, southeastern New York State. *Tree Physiol* 28(4):537–549.
- Taylor LR, Perry JN, Woiwod IP, Taylor RAJ (1988) Specificity of the spatial power-law exponent in ecology and agriculture. *Nature* 332(21):721–722.
- Keil P, Herben T, Rosindell J, Storch D (2010) Predictions of Taylor’s power law, density dependence and pink noise from a neutrally modeled time series. *J Theor Biol* 265(1):78–86.
- Laguerre C, Poulin R, Cohen JE (2015) Parasitism alters three power laws of scaling in a metazoan community: Taylor’s law, density-mass allometry, and variance-mass allometry. *Proc Natl Acad Sci USA* 112(6):1791–1796.
- Routledge RD, Swartz TB (1991) Taylor’s power law re-examined. *Oikos* 60(1):107–112.
- Tokeshi M (1995) On the mathematical basis of the variance-mean power relationship. *Res Popul Ecol (Kyoto)* 37(1):43–48.
- Kalyuzhny M, et al. (2014) Temporal fluctuation scaling in populations and communities. *Ecology* 95(6):1701–1709.
- Downing JA (1986) Spatial heterogeneity: Evolved behaviour or mathematical artefact? *Nature* 323:255–257.
- Packard GC (2009) On the use of logarithmic transformations in allometric analyses. *J Theor Biol* 257(3):515–518, discussion 519–521.
- Packard GC, Birchard GF (2008) Traditional allometric analysis fails to provide a valid predictive model for mammalian metabolic rates. *J Exp Biol* 211(Pt 22):3581–3587.
- Packard GC, Birchard GF, Boardman TJ (2011) Fitting statistical models in bivariate allometry. *Biol Rev Camb Philos Soc* 86(3):549–563.
- Xiao X, White EP, Hooten MB, Durham SL (2011) On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. *Ecology* 92(10):1887–1894.
- Lai J, Yang B, Lin D, Kerkhoff AJ, Ma K (2013) The allometry of coarse root biomass: Log-transformed linear regression or nonlinear regression? *PLoS ONE* 8(10):e77007.
- Snedecor GW, Cochran WG (1980) *Statistical Methods* (Iowa State Univ Press, Ames, IA), 7th Ed.
- Cramér H (1946) *Mathematical Methods of Statistics* (Princeton Univ Press, Princeton).
- Oehlert GW (1992) A note on the delta method. *Am Stat* 46(1):27–29.
- Loève M (1977) *Probability Theory I* (Springer, New York), 4th Ed.

Supporting Information

Cohen and Xu 10.1073/pnas.1503824112

SI Text

If X is a real-valued random variable with finite mean $E(X)$ and finite variance $\text{var}(X)$, and if a real-valued function f of real x is twice differentiable at $E(X)$, then the delta method (ref. 1 and ref. 2, pp. 355–358) gives the approximations

$$\begin{aligned} f(X) &\approx f(E(X)) + (X - E(X)) \left\{ f'(x) \Big|_{x=E(X)} \right\}, \\ E(f(X)) &\approx f(E(X)) + \left\{ \frac{f''(x)}{2} \Big|_{x=E(X)} \right\} \cdot \text{var}(X), \\ \text{var}(f(X)) &\approx \left\{ f'(x) \Big|_{x=E(X)} \right\}^2 \text{var}(X). \end{aligned}$$

In practice, we compute sample moments from observations of X , plug them in to replace the population moments, and accept the result as approximations to the left sides.

Lemma 1. Suppose Y is a nonnegative real-valued random variable with finite mean $E(Y) = M > 0$ and finite variance $\text{var}(Y) = V > 0$. Assume sampled observations are iid and the sample size in block j is n_j ($j = 1, 2, \dots, N$) and N is the number of blocks. If m_j is the sample mean of observations in block j , then the approximations given by the delta method are $\log m_j \approx \log M + (m_j - M)/M$, $E(\log m_j) \approx \log M - V/(2n_j M^2)$, $\text{var}(\log m_j) \approx V/(n_j M^2)$.

Proof: In the approximations from the delta method, we set $X = m_j$, $f(x) = \log(x)$, $x > 0$. Therefore, $f'(x) = 1/x$ and $f''(x) = -1/x^2$. Because $E(m_j) = M$ and $\text{var}(m_j) = V/n_j$,

$$\begin{aligned} \log m_j &\approx f(M) + (m_j - M) \cdot \frac{1}{M} = \log M + (m_j - M)/M, \\ E(\log m_j) &\approx f(M) + \left(-\frac{1}{2M^2} \right) \cdot \frac{V}{n_j} = \log M - V/(2n_j M^2), \\ \text{var}(\log m_j) &\approx \left(\frac{1}{M} \right)^2 \cdot \frac{V}{n_j} = V/(n_j M^2). \end{aligned}$$

The proof is complete.

Lemma 2. Under the assumptions of Lemma 1 also assume the third and fourth central moments of the random variable Y are finite and positive, that is, $\mu_h = E\{(Y - M)^h\} > 0$, $h = 3, 4$. Suppose v_j is the unbiased sample variance of observations in block j and $E(v_j) = V > 0$. Then the approximations given by the delta method are $\log v_j \approx \log V + (v_j - V)/V$, $\text{var}(\log v_j) \approx \left(\mu_4 - \frac{n_j - 3}{n_j - 1} V^2 \right) / (n_j V^2)$, $E(\log v_j) \approx \log V - \frac{1}{2n_j} \left(\frac{\mu_4}{V^2} - \frac{n_j - 3}{n_j - 1} \right)$.

Proof: Setting $X = v_j$ and following the same arguments as in the proof of Lemma 1 give the results.

Lemma 3. Under the assumptions of Lemmas 1 and 2, the covariance of the sample mean and sample variance is $\text{cov}(v_j, m_j) = \mu_3/n_j$.

Zhang (3) gives a proof of this classical formula, which has been known at least since 1903 (ref. 4, p. 279, equation xiii; ref. 5, p. 7, equation xxvi; ref. 6, p. 479, equation 67; and ref. 7, p. 402, equations 3 and 4).

Proof of Theorem: When all blocks are weighted equally, the least-squares estimators of slope b and intercept $\log(a)$, and SE of the slope estimator $s(\hat{b})$ are, respectively (8, p. 155),

$$\begin{aligned} \hat{b} &= \text{cov}_+(\log v_j, \log m_j) / \text{var}_+(\log m_j), \\ \widehat{\log(a)} &= \text{mean}_+(\log v_j) - \hat{b} \cdot \text{mean}_+(\log m_j), \\ s(\hat{b}) &= \sqrt{[\text{var}_+(\log v_j) / \text{var}_+(\log m_j) - \{\text{cov}_+(\log v_j, \log m_j)\}^2 / \{\text{var}_+(\log m_j)\}^2] / (N - 2)}. \end{aligned}$$

The notations $\text{mean}_+(\cdot)$, $\text{var}_+(\cdot)$, and $\text{cov}_+(\cdot, \cdot)$ are to be read as the mean, variance, and covariance across all blocks and not as referring to any single block j . Explicitly, the sample estimators are defined by

$$\text{mean}_+(\log m_j) = \frac{1}{N} \sum_{j=1}^N \log m_j,$$

$$\text{mean}_+(\log v_j) = \frac{1}{N} \sum_{j=1}^N \log v_j,$$

$$\text{var}_+(\log m_j) = \frac{1}{N-1} \sum_{j=1}^N (\log m_j)^2 - \frac{1}{N(N-1)} \left(\sum_{j=1}^N \log m_j \right)^2,$$

$$\text{var}_+(\log v_j) = \frac{1}{N-1} \sum_{j=1}^N (\log v_j)^2 - \frac{1}{N(N-1)} \left(\sum_{j=1}^N \log v_j \right)^2,$$

$$\text{cov}_+(\log v_j, \log m_j) = \frac{1}{N-1} \sum_{j=1}^N (\log m_j \cdot \log v_j) - \frac{1}{N(N-1)} \left(\sum_{j=1}^N \log m_j \right) \left(\sum_{j=1}^N \log v_j \right).$$

They are all consistent by the law of large numbers: as $N \rightarrow \infty$, $\text{mean}_+(\log m_j) \rightarrow_P E(\log m_j)$, $\text{mean}_+(\log v_j) \rightarrow_P E(\log v_j)$, $\text{var}_+(\log m_j) \rightarrow_P \text{var}(\log m_j)$, $\text{var}_+(\log v_j) \rightarrow_P \text{var}(\log v_j)$, and $\text{cov}_+(\log v_j, \log m_j) \rightarrow_P \text{cov}(\log v_j, \log m_j)$. Here the symbol “ \rightarrow_P ” means convergence in probability.

To find the limits in probability of \hat{b} and $s(\hat{b})$, we approximate the above estimators by the delta method using *Lemmas 1, 2, and 3*. We first approximate the numerator and the denominator of \hat{b} separately. For the numerator of \hat{b} , namely, $\text{cov}_+(\log v_j, \log m_j)$, the first term is approximately

$$\begin{aligned} \frac{1}{N-1} \sum_{j=1}^N (\log m_j \cdot \log v_j) &\approx \frac{1}{N-1} \sum_{j=1}^N \left\{ \log M + \frac{1}{M} (m_j - M) \right\} \cdot \left\{ \log V + \frac{1}{V} (v_j - V) \right\} \\ &= \frac{N}{N-1} \cdot \log M \cdot \log V + \frac{\log V}{(N-1)M} \sum_{j=1}^N (m_j - M) + \frac{\log M}{(N-1)V} \sum_{j=1}^N (v_j - V) + \frac{1}{(N-1)MV} \sum_{j=1}^N (m_j - M)(v_j - V). \end{aligned}$$

The second term of the numerator of \hat{b} is approximately

$$\begin{aligned} \frac{1}{N(N-1)} \left(\sum_{j=1}^N \log m_j \right) \left(\sum_{j=1}^N \log v_j \right) &\approx \frac{1}{N(N-1)} \sum_{j=1}^N \left\{ \log M + \frac{1}{M} (m_j - M) \right\} \cdot \sum_{j=1}^N \left\{ \log V + \frac{1}{V} (v_j - V) \right\} \\ &= \frac{N}{N-1} \cdot \log M \cdot \log V + \frac{\log V}{(N-1)M} \sum_{j=1}^N (m_j - M) + \frac{\log M}{(N-1)V} \sum_{j=1}^N (v_j - V) \\ &\quad + \frac{1}{N(N-1)MV} \sum_{j=1}^N (m_j - M) \sum_{j=1}^N (v_j - V). \end{aligned}$$

Therefore

$$\begin{aligned} \text{cov}_+(\log v_j, \log m_j) &\approx \frac{1}{(N-1)MV} \sum_{j=1}^N (m_j - M)(v_j - V) - \frac{1}{N(N-1)MV} \sum_{j=1}^N (m_j - M) \sum_{j=1}^N (v_j - V) \\ &= \frac{1}{(N-1)MV} \sum_{j=1}^N m_j v_j - \frac{1}{N(N-1)MV} \sum_{j=1}^N m_j \sum_{j=1}^N v_j = \frac{\text{cov}_+(m_j, v_j)}{MV}. \end{aligned}$$

Similarly, the denominator of \hat{b} is approximately

$$\text{var}_+(\log m_j) \approx \frac{1}{M^2} \left\{ \frac{1}{(N-1)} \sum_{j=1}^N m_j^2 - \frac{1}{N(N-1)} \left(\sum_{j=1}^N m_j \right)^2 \right\} = \text{var}_+(m_j) / M^2.$$

Consequently, for large n_j , $j = 1, 2, \dots, N$, $\hat{b} \approx \frac{\text{cov}_+(m_j, v_j)}{MV} / \frac{\text{var}_+(m_j)}{M^2}$. By consistency, for large N , using *Lemma 3* in the numerator,

$$\hat{b} \approx \frac{\text{cov}(m_j, v_j)}{MV} / \frac{\text{var}(m_j)}{M^2} = \frac{\mu_3}{n_j MV} / \frac{V}{n_j M^2} = \mu_3 M / V^2 = \gamma_1 / CV.$$

Using the consistency of estimator $\text{mean}_+(\cdot)$ and existing expressions for $E(\log m_j)$, $E(\log v_j)$ and \hat{b} , for large N and n_j , $j = 1, 2, \dots, N$,

$$\widehat{\log(a)} \approx E(\log v_j) - \hat{b} \cdot E(\log m_j) \approx \left[\log V - \frac{1}{2n_j} \left(\frac{\mu_4}{V^2} - \frac{n_j - 3}{n_j - 1} \right) \right] - \frac{\gamma_1}{CV} [\log M - V / (2n_j M^2)] \approx \log V - \frac{\gamma_1}{CV} \cdot \log M.$$

The derivation of $\text{var}_+(\log v_j)$ is the same as that of $\text{var}_+(\log m_j)$. Replacing m_j with v_j and M with V yields $\text{var}_+(\log v_j) \approx \text{var}_+(v_j)/V^2$. For large N and $n_j, j=1, 2, \dots, N$, substituting into the formula for $s(\hat{b})$ the estimators corresponding to $\text{var}_+(m_j)$, $\text{var}_+(v_j)$, and \hat{b} yields

$$s(\hat{b}) \approx \sqrt{\frac{1}{N-2} \left[\left(\frac{\mu_4}{V^2} - 1 \right) / \frac{V}{M^2} - (\mu_3 M / V^2)^2 \right]} = \sqrt{\frac{M^2 (\mu_4 V - V^3 - \mu_3^2)}{(N-2)V^4}} = \sqrt{\frac{\kappa - 1 - \gamma_1^2}{(N-2)(CV)^2}},$$

where $\kappa = \mu_4/V^2$ is the kurtosis. This completes the proof.

1. Oehlert GW (1992) A note on the delta method. *Am Stat* 46(1):27–29.
2. Hosmer DW, Lemeshow S, May S (2008) *Applied Survival Analysis: Regression Modeling of Time-to-Event Data* (Wiley, New York), 2nd Ed.
3. Zhang L (2007) Sample mean and sample variance: their covariance and their (in)dependence. *Am Stat* 61(2):159–160.
4. Pearson K (1903) On the probable errors of frequency constants. *Biometrika* 2(3):273–281.
5. Pearson K (1913) On the probable errors of frequency constants part II. *Biometrika* 9(1/2):1–10.
6. Neyman J (1925) Contributions to the theory of small samples drawn from a finite population. *Biometrika* 17(3/4):472–479.
7. Neyman J (1926) On the correlation of the mean and the variance in samples drawn from an “infinite” population. *Biometrika* 18(3/4):401–413.
8. Snedecor GW, Cochran WG (1980) *Statistical Methods* (Iowa State Univ Press, Ames, IA), 7th Ed.

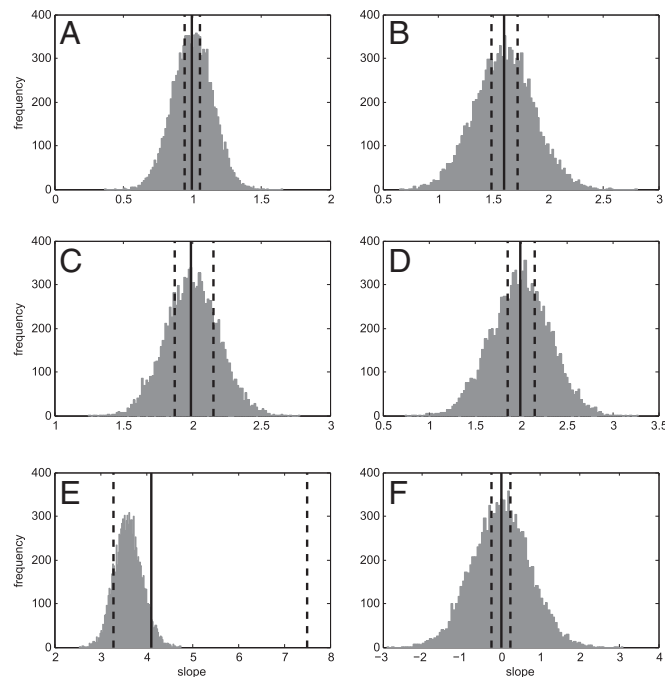


Fig. S1. Comparison of TL slope estimator \hat{b} predicted from theory and computed using linear regression for (A) Poisson ($\lambda = 1$), (B) negative binomial ($r = 5, p = 0.4$), (C) exponential ($\lambda = 1$), (D) gamma ($\alpha = 4, \beta = 1$), (E) lognormal ($\mu = 1, \sigma = 1$), and (F) shifted normal [$5 + \mathcal{N}(0, 1)$] distributions. Gray histogram shows the distribution of point estimates of b from 10,000 linear regressions. For each distribution, the black solid line and dashed lines give, respectively, the median and 95% CI of b calculated from 10,000 random copies of $n \times N$ iid samples using the theoretical formula (Eq. 3).

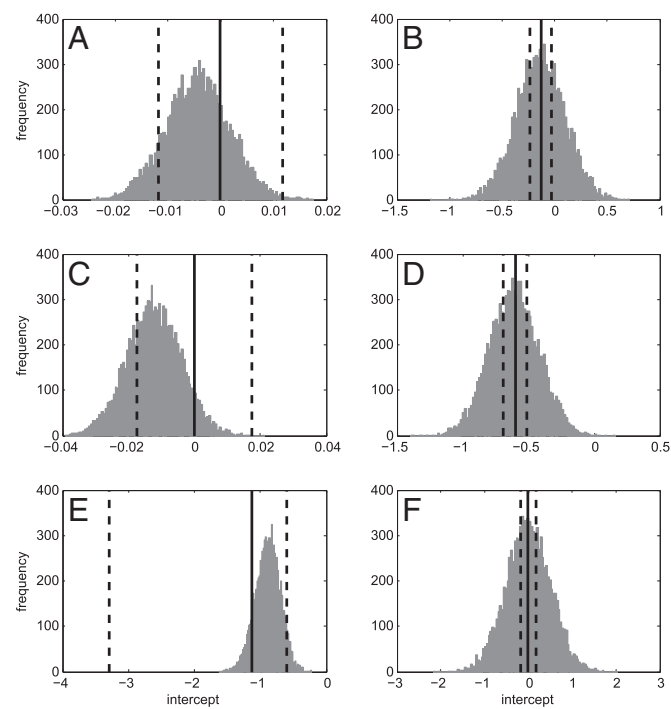


Fig. S2. Comparison of TL intercept estimator $\log(\hat{a})$ predicted from theory and computed using linear regression for (A) Poisson ($\lambda = 1$), (B) negative binomial ($r = 5, p = 0.4$), (C) exponential ($\lambda = 1$), (D) gamma ($\alpha = 4, \beta = 1$), (E) lognormal ($\mu = 1, \sigma = 1$), and (F) shifted normal [$5 + \mathcal{N}(0, 1)$] distributions. Gray histogram shows the distribution of point estimates of $\log(\hat{a})$ from 10,000 linear regressions. For each distribution, the black solid line and dashed lines gave, respectively, the median and 95% CI of $\log(\hat{a})$ calculated from 10,000 random copies of $n \times N$ iid samples using the theoretical formula (Eq. 4).

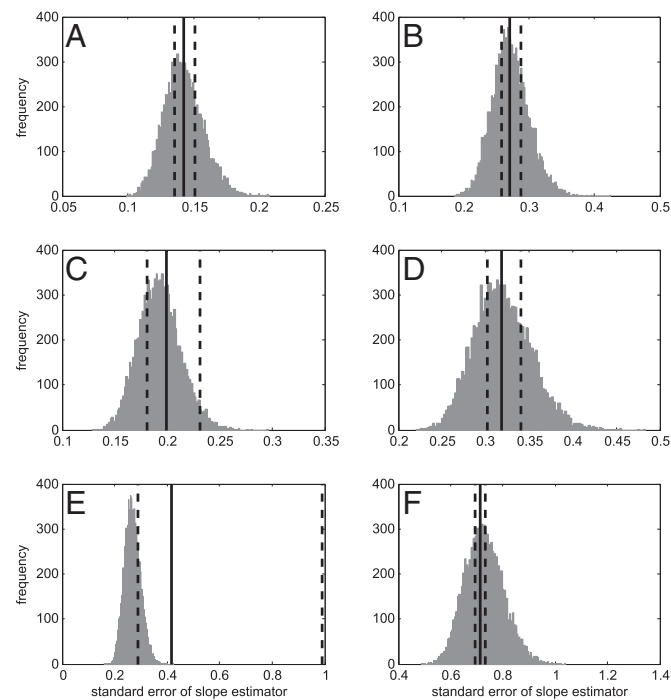


Fig. S3. Comparison of SE of the slope estimator $[s(\hat{b})]$ predicted from theory and computed using linear regression for (A) Poisson ($\lambda = 1$), (B) negative binomial ($r = 5, p = 0.4$), (C) exponential ($\lambda = 1$), (D) gamma ($\alpha = 4, \beta = 1$), (E) lognormal ($\mu = 1, \sigma = 1$), and (F) shifted normal [$5 + \mathcal{N}(0, 1)$] distributions. Gray histogram shows the distribution of point estimates of the SE of \hat{b} from 10,000 linear regressions. For each distribution, the black solid line and dashed lines gave, respectively, the median and 95% CI of the SE of \hat{b} calculated from 10,000 random copies of $n \times N$ iid samples using the theoretical formula (Eq. 5).