6-23-2015

Random Sampling of Skewed Distributions Implies Taylor's Power Law of Fluctuation Scaling

Joel E. Cohen
Rockefeller University

Meng Xu
University of New Haven, mxu@newhaven.edu

Publisher Citation

Cohen JE, Xu M (2015). Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling. Proceedings of the National Academy of Sciences, U.S.A. 112 (25): 7749-7754, June 23, 2015.doi: 10.1073/pnas.1503824112.

Classification: BIOLOGICAL SCIENCES, Ecology; PHYSICAL SCIENCES, Applied Mathematics

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

Short Title: Skewed distributions lead to Taylor's power law

Joel E. Cohen[1,*] and Meng Xu[1,2]

[1]Laboratory of Populations, Rockefeller University & Columbia University, 1230 York Avenue, Box 20, New York, NY 10065 United States

[2]Department of Mathematics and Physics, University of New Haven, 300 Boston Post Road, West Haven, CT 06516 United States

[*]Corresponding author: Joel E. Cohen, Laboratory of Populations, Rockefeller University & Columbia University, 1230 York Avenue, Box 20, New York, NY 10065 United States. Telephone: (+1) 212-327-8883. E-mail: cohen@rockefeller.edu

Keywords: delta method; least-squares regression; skewness; variance function

Abstract

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 of the species' population density (or other nonnegative quantity): approximately, variance = $a(\text{mean})^b$, $a > 0$. In the past half-century, multiple mechanisms have been proposed to explain and interpret TL. Here we show analytically that TL arises when data are randomly sampled in blocks from any skewed frequency distribution with four finite moments. We give approximate formulas 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.

Significance Statement (limited to 120 words)

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 of population data 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.

Introduction

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:

$$variance = a(mean)^b, \quad a > 0, \qquad \text{(Eqn 1)}$$

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

$$\log(variance) = \log a + b \times \log(mean). \qquad \text{(Eqn 2)}$$

Eqns 1 and 2 may be exact if the mean and variance are population moments calculated from certain parametric families of probability distributions. Eqns 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 Eqns 1 or 2.

TL has been verified for hundreds of biological species and non-biological 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 individuals of a species aggregate in space (1). Empirically, b often lies between 1 and 2 (16). Ballantyne and Kerkhoff

(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, non-interacting clones. From an ecological community perspective, 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 of varying sizes (24, 25), a stochastic version of logistic population dynamics (16), and Lewontin-Cohen stochastic multiplicative population model (8).The substantive diversity of empirical confirmations has suggested that no narrowly specific mechanism, biological, physical, technological, or behavioral, 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 observed in empirical examples of TL.

The present work was kindred in spirit and intent, though distinct in technical approach and results. Here we demonstrated that TL arises 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. Under these assumptions, we derived analytically the explicit approximate formulae for the TL slope (b in Eqn 2), intercept ($\log(a)$ in Eqn 2), and standard error of the slope estimator ($s(\hat{b})$, 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 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 ecological or statistical mechanisms, and provided a null hypothesis 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, mean $E(X) = M > 0$, variance $var(X) = V > 0$, and finite 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, $i = 1, \ldots, n_j$, assuming the number of observations in block j satisfies $n_j > 3$, $j = 1, \ldots, N$. The total number of observations is $n_1 + n_2 + \cdots + n_N$. For block j the sample mean of observations and the expectation and variance of the sample mean are, respectively, $m_j = \left(x_{1j} + \cdots + x_{n_j j}\right)/n_j$, $E(m_j) = M$, $var(m_j) = V/n_j$. The unbiased sample variance of block j and its expectation and variance are, respectively,

$$v_j = \frac{1}{n_j - 1}\sum_{i=1}^{n_j} x_{ij}^2 - \frac{n_j}{n_j - 1}m_j^2, E(v_j) = V, var(v_j) = \frac{1}{n_j}\left(\mu_4 - \frac{n_j - 3}{n_j - 1}V^2\right).$$

The formula for $var(v_j)$ is from Neter, Wasserman and Kutner (30). As $n_j \to \infty$, Prob$\{m_j = 0\} \to 0$ and Prob$\{v_j = 0\} \to 0$ 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 is small because it arises only from differences due to random sampling of the same distribution for every block. In empirical examples, if the variation of sample means among blocks is too large to arise from random sampling alone, e.g., if analysis of variance rejects homogeneity of block means, then the theory of TL here is inapplicable.

Variation between blocks in the sample variance is also small for the same reason, under the assumptions of this theory. Since 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 following result interprets this observation analytically.

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 follow this practice here.

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 $p > 0,\ E(Y) = K + o(n^{-p})$.

Theorem. Suppose the nonnegative real-valued random variable X has finite first four moments, with strictly positive mean and strictly positive variance. Suppose that $n_j > 1$ observations x_{ij} ($i = 1, \ldots, n_j$) of X are randomly assigned to block j ($j = 1, \ldots, N$), $N > 2$, and all the observations, which number $\sum_{j=1}^{N} n_j$ in total, are independently and identically distributed. 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 of b and $\log(a)$ in TL, $\log(v_j) = \log(a) + b \times \log(m_j)$, $j = 1, \dots, N$ (Eqn 2) respectively. Let $s(\hat{b})$ denote the standard error of the least-squares slope estimator $\hat{b}$. Then, in the limit of large N and large n_j,

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

$$\widehat{\log(a)} \approx \log V - \frac{\gamma_1}{CV} \cdot \log M \quad \text{(Eqn 4)}$$

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

Proof of this Theorem is given in the Supporting Information (SI). Since $CV > 0$, Eqn 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 Eqn 5 yields the estimated variance of $\hat{b}$. Since any variance is nonnegative by Cauchy's inequality (31), the numerator of the variance estimate ($\kappa - 1 - \gamma_1^2$) is nonnegative. Eqn 5 thus provides an alternative proof and adds a new interpretation of the inequality $\kappa - 1 - \gamma_1^2 \geq 0$ which 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 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 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, $j = 1, \ldots, N$. Fig. 1 visualizes the relationship between population distributions and TL.

For each of the five positively skewed distributions, an approximately linear relationship with positive slopes was observed (Fig. 1 a-e), but the lognormal slope 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. 1 f).

To illustrate our Theorem numerically, we applied the theoretical formulae (Eqns 3-5) to each of the six probability distributions and analytically computed the predicted values of the slope and intercept in Eqn 2, and standard error 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 to bootstrap medians and 95% confidence intervals (CIs) (2.5% and 97.5% quantiles) of TL parameters from the corresponding regression point estimates, and median and 95% CIs of the quadratic coefficient from the corresponding quadratic regression. 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 (Eqns 3-5) to evaluate the theoretical TL slope, intercept, and standard error of the slope estimator. Their medians and 95% CIs were similarly bootstrapped from the 10,000 random copies of the matrix. Estimates from the regression were compared with the corresponding theoretical predictions computed from the formulae analytically and numerically (Table 1).

The mean, variance, third and fourth central moments, computed analytically using the given parameters, are respectively 1, 2, 5, and 15 for Poisson ($\lambda = 1$), 7.5, 75, 142.5, and 9553.125 for negative binomial ($r = 5$, $p = 0.4$), 1, 2, 6, and 24 for exponential ($\lambda = 1$), 4, 20, 120, and 840 for gamma ($\alpha = 4$, $\beta = 1$), 4.4817, 54.5982, 1808.0400, and 162754.7914 for lognormal ($\mu = 1$, $\sigma = 1$), and 5, 26, 140, and 778 for shifted normal ($5 + \mathcal{N}(0,1)$). Except for the shifted normal distribution, a positive slope estimate $\hat{b}$ 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% bootstrapped CI of b under regression was on the right side of zero. The 95% bootstrapped 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% bootstrapped CI of the quadratic coefficient 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 for the five skewed probability distributions.

Except for the lognormal distribution, the theoretical values of *b* (Fig. 2) and log(*a*) (Fig. 3) predicted analytically from Eqns 3 and 4, and the standard error of the slope estimator (Fig. 4) calculated from Eqn 5 fell within the corresponding 95% CI from linear regression. In the lognormal distribution, the analytical predictions of the slope *b* and the standard error of its estimator were on the right side of the corresponding 95% CI from regression, meaning that the theoretically predicted values were significantly larger 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 bootstrapped 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).

BRF is a 1550-hectare forest preserve in Cornwall, NY (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, and each sampling point contributed one observation of basal area density for each tree species, with no repeated measurements at any sampling point (Friday and Friday, 1985 unpublished MS available from Black Rock Forest Consortium, Cornwall, NY, USA, courtesy of Dr. William S. F. Schuster, Executive Director). Each of the 218 sampling points is also geographically separated from the others so that the oak tree growth surrounding any two sampling points is not likely to be correlated due to geophysical or biological conditions (e.g. slope, soil moisture, topography). Hence the 218 measurements of basal area density could reasonably be interpreted as representing an iid sample of each tree species' basal area density in the whole BRF preserve 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. 5 e). Taylor and colleagues (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. 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 standard error of the slope estimator. Repeatedly randomizing the assignment of observations into blocks 10,000 times, we bootstrapped the median and 95% percentile CI of the slope, intercept and standard error of the slope estimator respectively from the corresponding 10,000 regression point estimates (Fig. 5 a-c). 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 bootstrapped the median and 95% CI of the quadratic coefficient.

Eqn 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). 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). The median of the standard error of the slope estimator was 0.4045 with 95% CI (0.2257, 0.7272). Thus TL held for RO basal area density with positive slope and positive intercept under random assignment of observations to blocks. The finding that the intercept was positive excluded the possibility that the basal area density of RO was Poisson distributed with different means in different blocks, because in that case the intercept would have been 0. 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 (Eqns 3-5), the predicted slope, predicted intercept, and standard error 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 bootstrapped from point estimates under linear regression (Fig. 5 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 (33, their Supplementary Tables S1, S2, S3, and S4). In summary, all four point estimates of the slope of TL under the four biological groupings fell within the 95% bootstrapped CI of the slope under random assignment of sampling points to blocks, and all four CIs of the slope under the biological groupings estimated from normal theory heavily overlapped the 95% bootstrapped 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 assignment of sampling points to blocks.

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

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 the random grouping and the 95% CI, (-0.8830, 1.4037), again almost contained the 95% CI under random assignment of sampling points to blocks.

The random sampling model of TL would account for 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 analysis of variance of the mean basal area density by block, for each method (Fig. 6). For Friday's, Schuster's, and watershed groupings, the null hypothesis that all blocks had equal means was rejected ($P = 0.014$, $P < 0.001$, $P = 0.009$, 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. How robust the predictions are with respect to violations of the assumptions is a question for future theoretical and empirical research.

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 standard error of the slope estimator. No biological or physical mechanisms need be invoked to explain TL under this form of sampling. Our examples show that this model has relevance to some, but not all, published empirical examples of TL.

Our null hypothesis 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 7 or more orders of magnitude (36), the differences in mean and variance of population density probably cannot be attributed to random sampling variation from a single underlying distribution. On the other hand, when the mean population densities range over little more than one order of magnitude ((11), p.12, their Fig. 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. 4 e). A possible reason is that $s(\hat{b})$ for the lognormal distribution (namely, 0.6660 in Table 1) was twice as large as $s(\hat{b})$ for any of the other four skewed distributions (the maximum being 0.3194 for the gamma distribution in Table 1), whereas the sample sizes for all of the distributions were the same n=100. In addition, since 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. Among tested distributions, the fourth moment of the lognormal distribution was at least 17 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 bootstrapped from 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 formulae imply that TL slope $b > 0$ arises from random sampling of observations in blocks of any right-skewed distribution, and $b < 0$ arises from random sampling of observations in blocks of any left-skewed distribution. These results connect TL with the underlying probability distribution but do not explain why the distribution of observations (e.g. Fig. 5 e) 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 deducing biological information about population aggregations is a subject of continuing scientific debate. Alternative mean-variance relationships have been proposed as competitors of TL (25, 37, 38). It has been argued that sampling error and sampling coverage may lead to TL-like patterns as statistical artifacts (39) and to substantially biased TL parameters (40). Our results offer another statistical mechanism that leads to TL.

Methods

Traditionally, when tested against empirical data, TL has been taken to be confirmed if the fitted linear regression Eqn 2 had statistically significantly non-zero linear coefficient (with P-value < α, where α is the significance level; here $\alpha = 0.05$), and if a least-squares quadratic regression between the independent variable log(mean) and dependent variable log(variance) did not yield a statistically significant quadratic term (quadratic coefficient P-value > α). 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, 41-43) and defended (44, 45).

Our numerical examples combined the ordinary least-squares regression approach with parameter bootstrapping. Specifically, in multiple realizations, we sampled from a single probability distribution, organized each sample into a block, calculated the mean and the variance of sample observations per block, recorded the parameters and quadratic coefficient estimates from the corresponding linear and quadratic regressions (46, p. 155), respectively, for each realization, and constructed CIs of the parameters using percentiles of the regression point estimates from all bootstrap realizations.

Similarly, in the empirical example of red oak trees, we randomly grouped observations into blocks. We adopted the bootstrapping method instead of using the standard *P*-value approach because the bootstrap CI does not assume normality of the parameter distribution (47, 48). Linear and quadratic regressions were performed using the MATLAB function “regress” (49).

The analytical formulae for the TL parameter estimators and the standard error of the slope estimator were derived using the delta method (50, 51). 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 (52) and the consistency of sample estimators (see SI). The delta method is increasingly accurate as the variation around the point of expansion becomes smaller. Since the variation in sample means and sample variances is small when sufficiently large random samples are blocked, it is not surprising that the delta method yields a quite accurate approximation to TL parameters estimated from linear regression.

Acknowledgments

This work was partially supported by United States National Science Foundation grants EF-1038337 and DMS-1225529. The authors thank William S. F. Schuster for the data from Black Rock Forest used here, Richard Chandler, Russell Millar, William S. F. Schuster, Nadav Shnerb, Ethan White, and Andrew Wood for helpful comments, and Priscilla K. Rogerson for assistance. The authors declare that they have no conflict of interest.

References

1. Taylor LR (1961) Aggregation, variance and the mean. *Nature* 189:732-735.

2. Taylor LR (1984) Assessing and interpreting the spatial distributions of insect populations. *Annu Rev Entomol* 29:321-357.

3. Kendal WS (2002) A frequency distribution for the number of hematogenous organ metastases. *J Theor Biol* 217(2):203-218.

4. Eisler Z, Bartos I, Kertész J (2008) Fluctuation scaling in complex systems: Taylor’s law and beyond. *Adv Phys* 57:89-142.

5. 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.

6. 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. *Ecological Processes* 1:5.

7. 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 U S A* 109(39):15829-15834.

8. Cohen JE, Xu M, Schuster WSF (2013a) Stochastic multiplicative population growth predicts and interprets Taylor's power law of fluctuation scaling. *Proc Biol Sci* 280:20122955.

9. Cohen JE, Xu M, Brunborg H (2013b) Taylor's law applies to spatial variation in a human population. *Genus* 69(1):25-60.

10. Mellin C, Huchery C, Caley MJ, Meekan MG, Bradshaw CJA (2010) Reef size and isolation determine the temporal stability of coral reef fish populations. *Ecology* 91(11):3138-3145.

11. Fukaya K, Okuda T, Hori M, Yamamoto T, Nakaoka M, Noda T (2013) Variable processes that determine population growth and an invariant mean-variance relationship of intertidal barnacles. *Ecosphere* 4(4):1-20.

12. 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. *Journal of Animal Ecology* 83(6):1334-1343.

13. 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.

14. 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.

15. 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 (John Wiley & Sons, Inc., New York), pp 85-119.

16. Maurer BA, Taper ML (2002) Connecting geographical distributions with population processes. *Ecology Letters* 5:223–231.

17. Ballantyne F, Kerkhoff AJ (2007) The observed range for temporal mean-variance scaling exponents can be explained by reproductive correlation. *Oikos* 116:174-180.

18. 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.

19. Cohen JE (2013) Taylor's power law of fluctuation scaling and the growth-rate theorem. *Theor Popul Biol* 88:94-100.

20. Kilpatrick AM, Ives AR (2003) Species interactions can explain Taylor’s power law for ecological time series. *Nature* 422:65-68.

21. Jørgensen B (1987) Exponential dispersion models. *J R Stat Soc Ser B* 49:127-162.

22. Jørgensen B (1997) *The theory of dispersion models* (Chapman & Hall, London).

23. Kendal WS, Jørgensen B (2011) Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. *Phys Rev E* 83:066115.

24. Yamamura K (1990) Sampling scale dependence of Taylor's power law. *Oikos* 59(1):121-125.

25. Yamamura K (2000) Colony expansion model for describing the spatial distribution of population. *Population Ecology* 42:160-169.

26. Tao T (2012) *E pluribus unum*: from complexity, universality. *Daedalus* (American Academy of Arts and Sciences) 141(3):23-34.

27. Révész P (1968) *The Laws of Large Numbers* (Academic Press, New York).

28. Fischer H (2011) *A History of the Central Limit Theorem from Classical to Modern Probability Theory* (Springer, New York).

29. 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], [v1] Mon, 27 Oct 2014.

30. Neter J, Wasserman W, Kutner MH (1990) *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs* (Richard D. Irwin, Inc. Homewood, IL), 3rd Ed.

31. Steele JM (2004) *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities* (Cambridge University Press, Cambridge, UK).

32. Rohatgi VK, Székely GJ (1989) Sharp inequalities between skewness and kurtosis. *Stat Probab Lett* 8**:**297-299.

33. Xu M, Schuster WSF, Cohen JE (2015) Robustness of Taylor's law under spatial hierarchical groupings of forest tree samples. *Popul. Ecol.* In press.

34. Schuster WSF, Griffin KL, Roth H, Turnbull MH, Whitehead D, Tissue DT (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 Physiology* 28:537–549.

35. 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.

36. Lagrue 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. PNAS in press 2015.

37. Routledge RD, Swartz TB (1991) Taylor's power law re-examined. Oikos 60(1):107-112.

38. Tokeshi M (1995) On the mathematical basis of the variance-mean power relationship. *Res Popul Ecol* 37(1):43-48.

39. Kalyuzhny M, et al. (2014) Temporal fluctuation scaling in populations and communities. *Ecology* 95(6):1701-1709.

40. Downing JA (1986) Spatial heterogeneity: evolved behaviour or mathematical artefact? *Nature* 323:255-257.

41. Packard GC (2009) On the use of logarithmic transformations in allometric analyses. *J Theor Biol* 257:515-518.

42. Packard GC, Birchard GF (2008) Traditional allometric analysis fails to provide a valid predictive model for mammalian metabolic rates. *J Exp Biol* 211:3581-3587.

43. Packard GC, Birchard GF, Boardman TJ (2011) Fitting statistical models in bivariate allometry. *Biol Rev Camb Philos Soc* 86:549-563.

44. 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.

45. 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.

46. Snedecor GW, Cochran WG (1980) *Statistical Methods* (Iowa State University Press, Ames, IA), 7th Ed.

47. Efron B (1987) Bootstrap confidence intervals. *J Am Stat Assoc* 82(397):171-185.

48. DiCiccio TJ, Efron B (1996) Bootstrap confidence intervals. *Stat Sci* 11(3):189-228.

49. MATLAB 2011a (2011) The MathWorks, Inc. (Natick, Massachusetts, United States).

50. Cramér H (1946) *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ).

51. Oehlert GW (1992) A note on the delta method. *Am Stat* 46(1):27-29.

52. Loève M (1977) *Probability Theory I* (Springer-Verlag, New York), 4th Ed.

Figure Legends

Fig. 1. Taylor's law with positive slope arises from random samples from a single (a) Poisson (λ = 1), (b) negative binomial (r = 5, p = 0.4), (c) exponential (λ = 1), (d) gamma (α = 4, β = 1), and (e) lognormal (μ = 1, σ = 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, $j = 1, \ldots, N$. The solid grey line is the least-squares linear regression $\log_{10} v_j = \log_{10} a + b \log_{10} m_j$. Slope and intercept of the dashed black line were computed analytically from Eqns 3 and 4 respectively (see Table 1). Population skewness in each distribution is 1 (Poisson), 0.9238 (negative binomial), 2 (exponential), 1 (gamma), 6.1849 (lognormal), and 0 (shifted normal).

Fig. 2. Comparison of TL slope estimator $\hat{b}$ predicted from theory and computed using linear regression for (a) Poisson (λ = 1), (b) negative binomial (r = 5, p = 0.4), (c) exponential (λ = 1), (d) gamma (α = 4, β = 1), (e) lognormal (μ = 1, σ = 1), and (f) shifted normal (5 + (0,1)) distributions. Grey 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 bootstrapped from 10,000 random copies of $n \times N$ iid samples using the theoretical formula (Eqn 3).

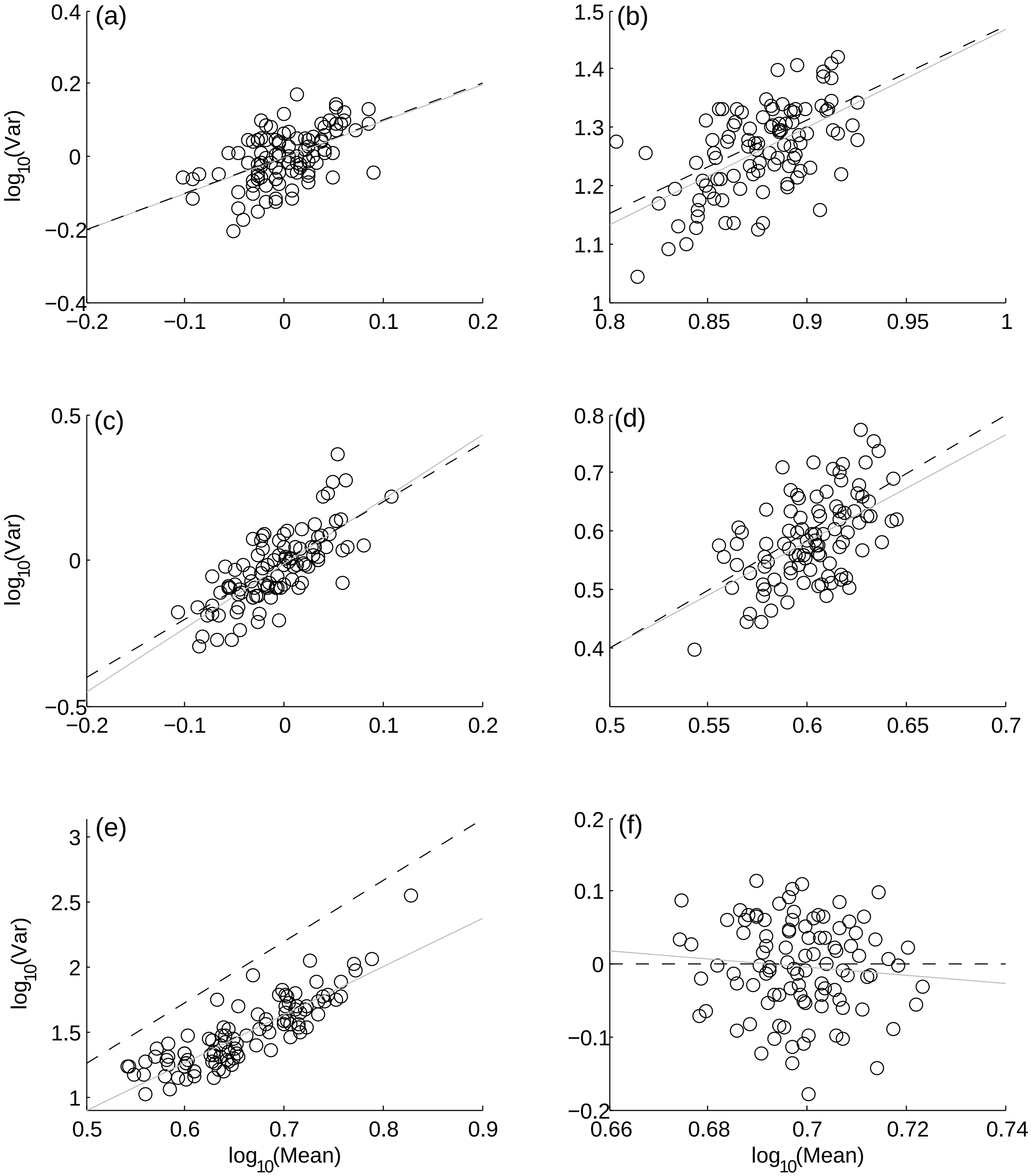
Fig. 3. Comparison of TL intercept estimator $\widehat{\log(a)}$ predicted from theory and computed using linear regression for (a) Poisson ($\lambda = 1$), (b) negative binomial ($r = 5$, $p = 0.4$), (c) exponential ($\lambda = 1$), (d) gamma ($\alpha = 4$, $\beta = 1$), (e) lognormal ($\mu = 1$, $\sigma = 1$), and (f) shifted normal ($5 + (0,1)$) distributions. Grey histogram shows the distribution of point estimates of $\log(a)$ from 10,000 linear regressions. For each distribution, the black solid line and dashed lines gave respectively the median and 95% CI of $\log(a)$ bootstrapped from 10,000 random copies of $n \times N$ iid samples using the theoretical formula (Eqn 4).

Fig. 4. Comparison of standard error 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. Grey histogram shows the distribution of point estimates of the standard error 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 standard error of $\hat{b}$ bootstrapped from 10,000 random copies of $n \times N$ iid samples using the theoretical formula (Eqn 5).

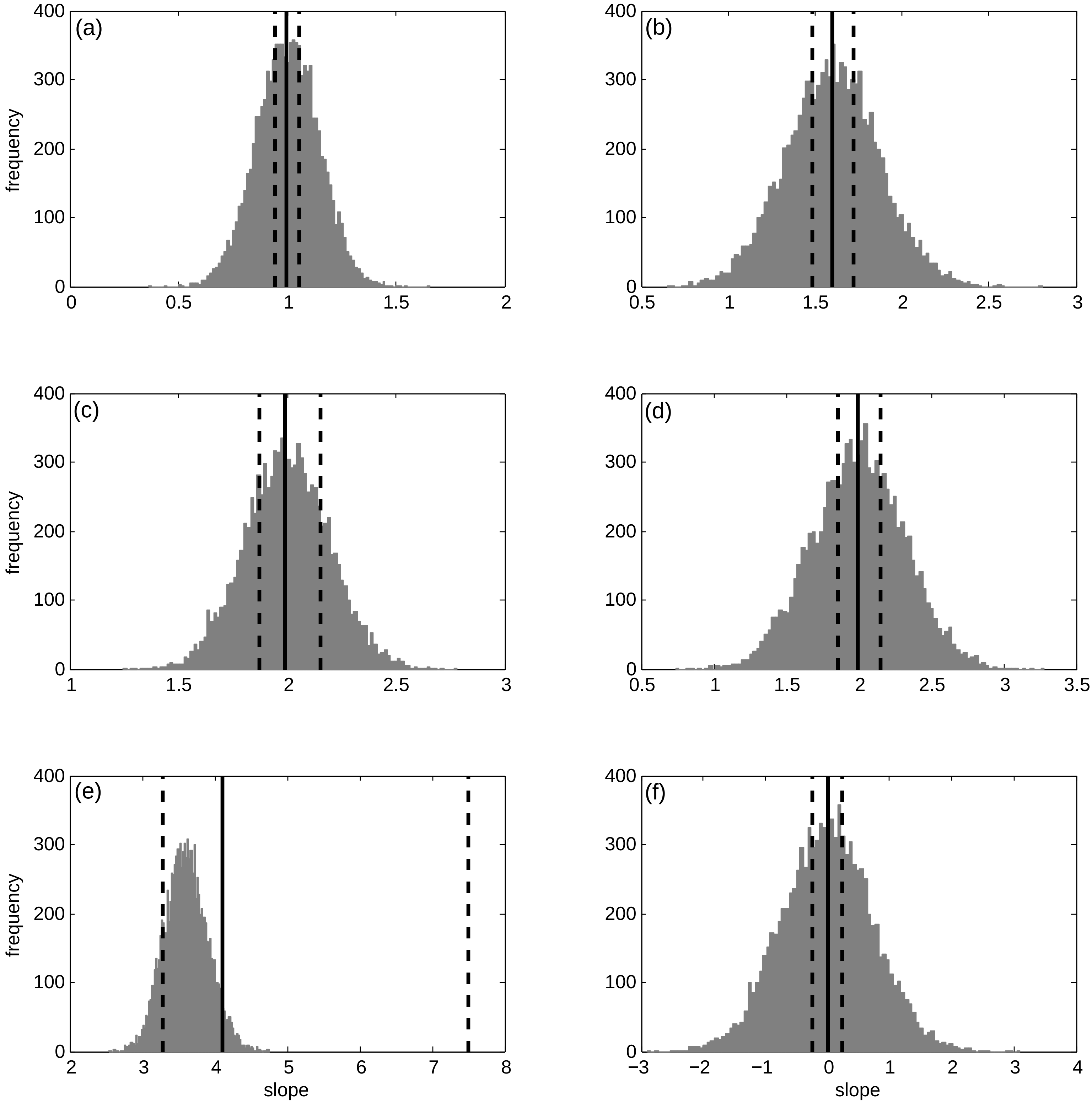
Fig. 5. Testing TL using basal area density of red oak in Black Rock Forest. (a-c) Histograms of the slope, intercept, and standard error of the slope estimator, respectively, estimated by regression from 10,000 random assignments of observations into blocks, with the theoretically predicted values marked by the solid vertical lines. (d) A bivariate fit between the independent

variable log(mean) and dependent variable log(variance) under one realization of random groupings. Each open circle represents a mean and a variance calculated over observations within a single block. The grey line is the least-squares linear regression line. (e) Histogram of basal area density of red oaks at 218 sampling points is right-skewed.

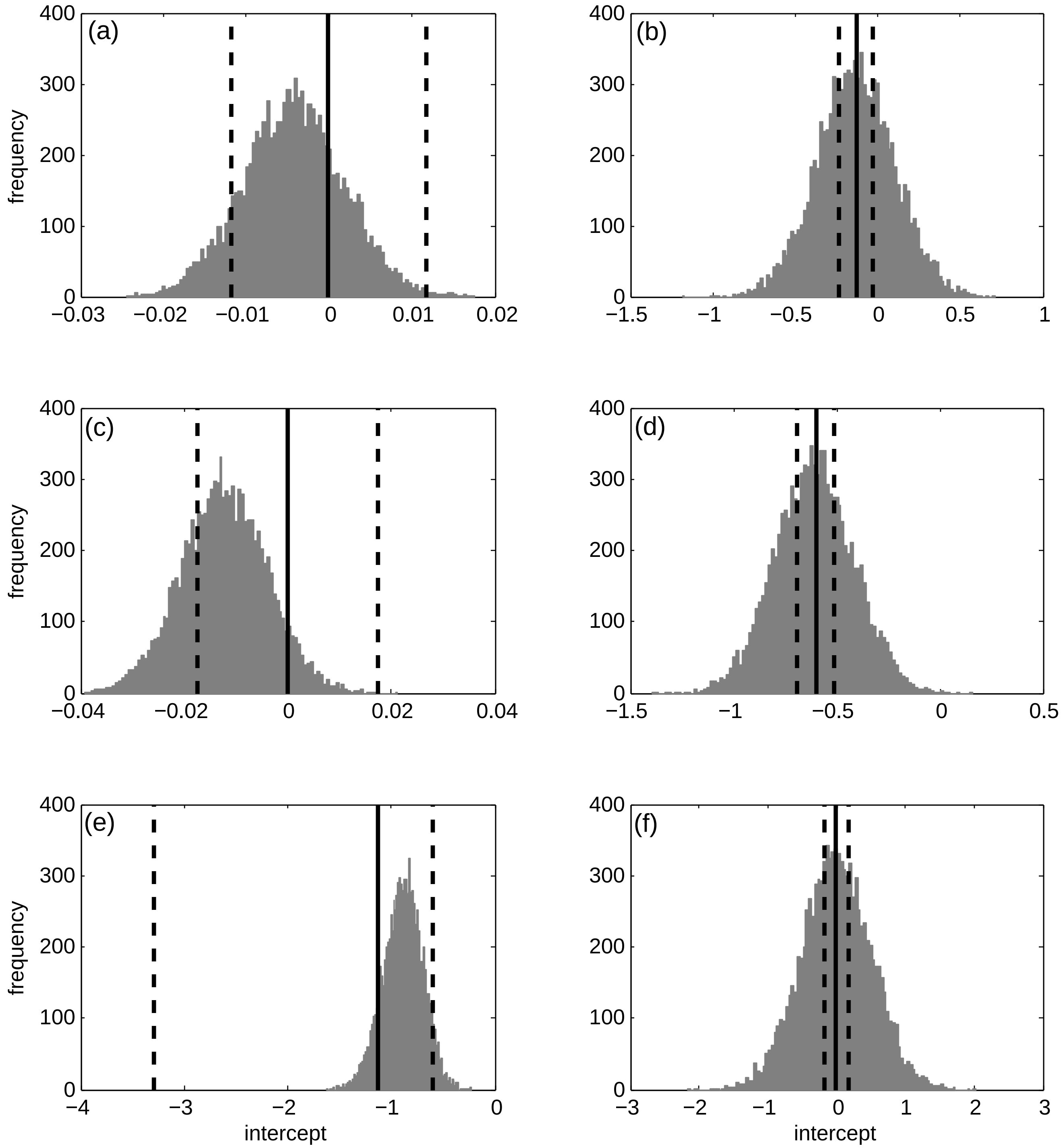
Fig. 6. Analysis of variance (ANOVA) of basal area density of red oak in Black Rock Forest, according to four biological methods of assigning plots to blocks. In each boxplot, the median is the bold black bar, the box covers the interquartile range, and the whiskers cover the entire range of basal area 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.



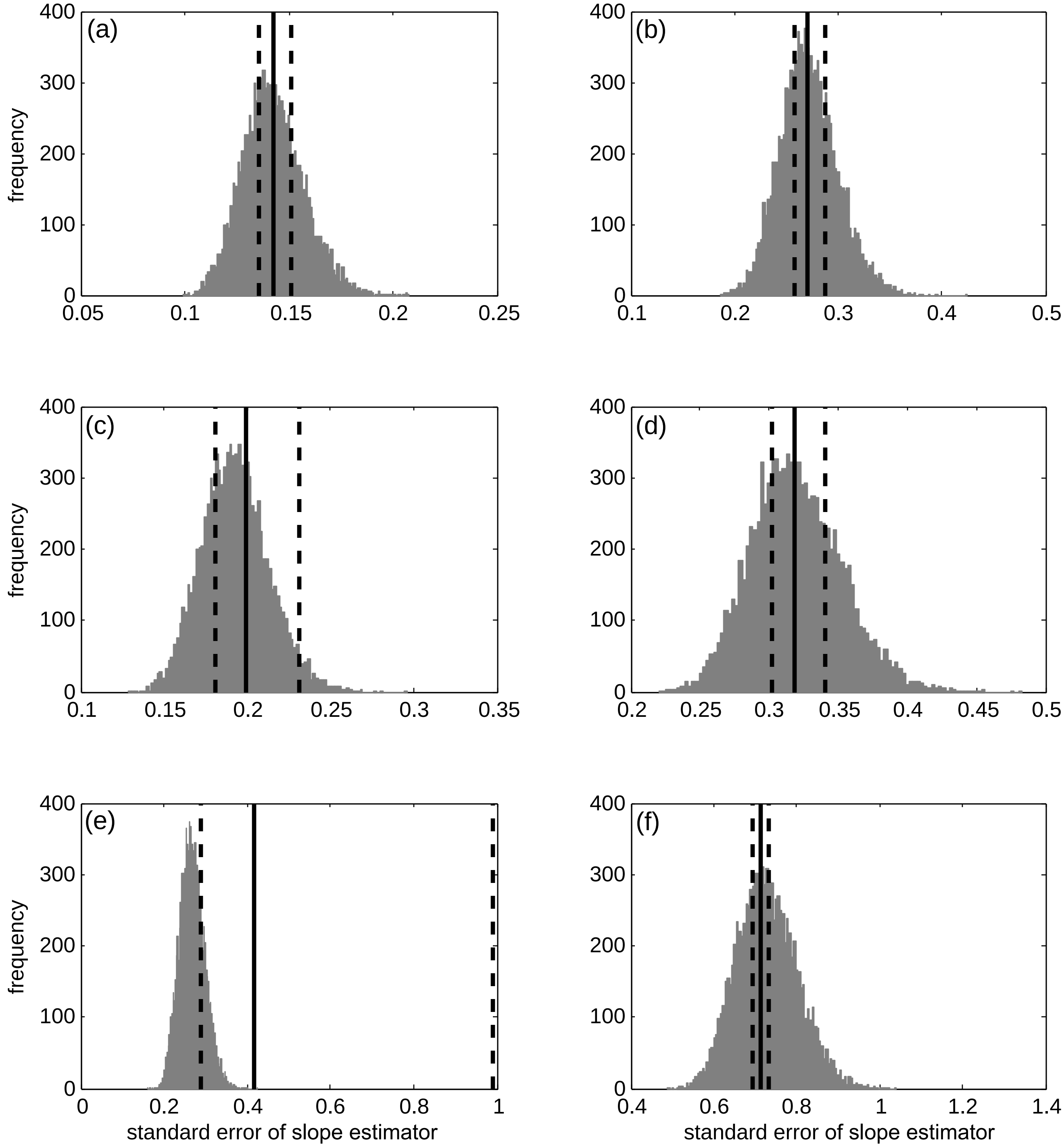
(a)
(b)
(c)
(d)
(e)
(f)
log10(Var)
log10(Mean)



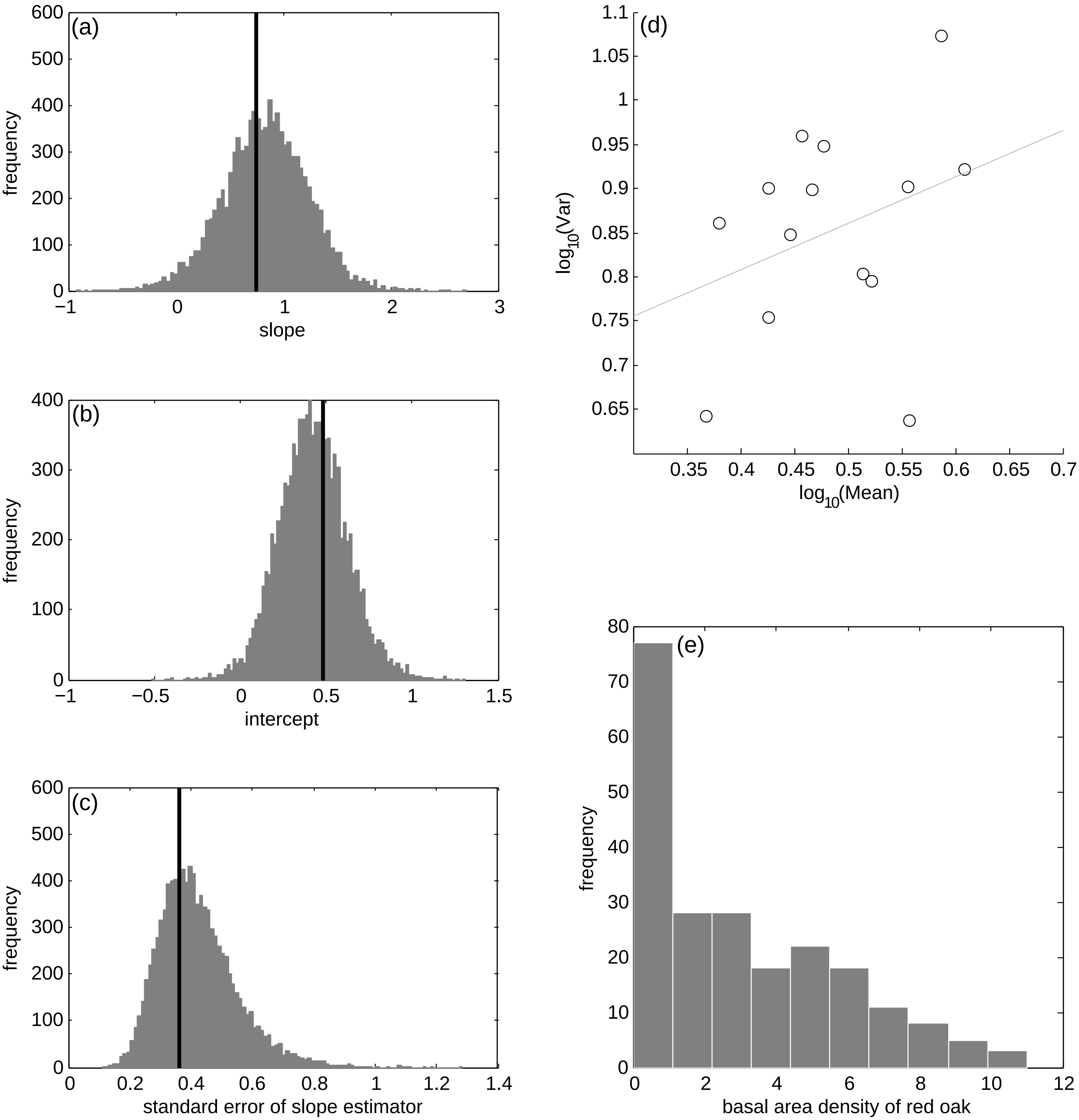
(a)
(b)
(c)
(d)
(e)
(f)
frequency
slope



(a)
(b)
(c)
(d)
(e)
(f)
frequency
intercept
0
100
200
300
400
−0.03
−0.02
−0.01
0
0.01
0.02
−1.5
−1
−0.5
0
0.5
1
−0.04
−0.02
0
0.02
0.04
−4
−3
−2
−1
−3
2
3



(a)
(b)
(c)
(d)
(e)
(f)
frequency
standard error of slope estimator
0
100
200
300
400
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.6
0.8
1
1.2
1.4



(a)
frequency
slope
(b)
frequency
intercept
(c)
frequency
standard error of slope estimator
(d)
log10(Var)
log10(Mean)
(e)
frequency
basal area density of red oak

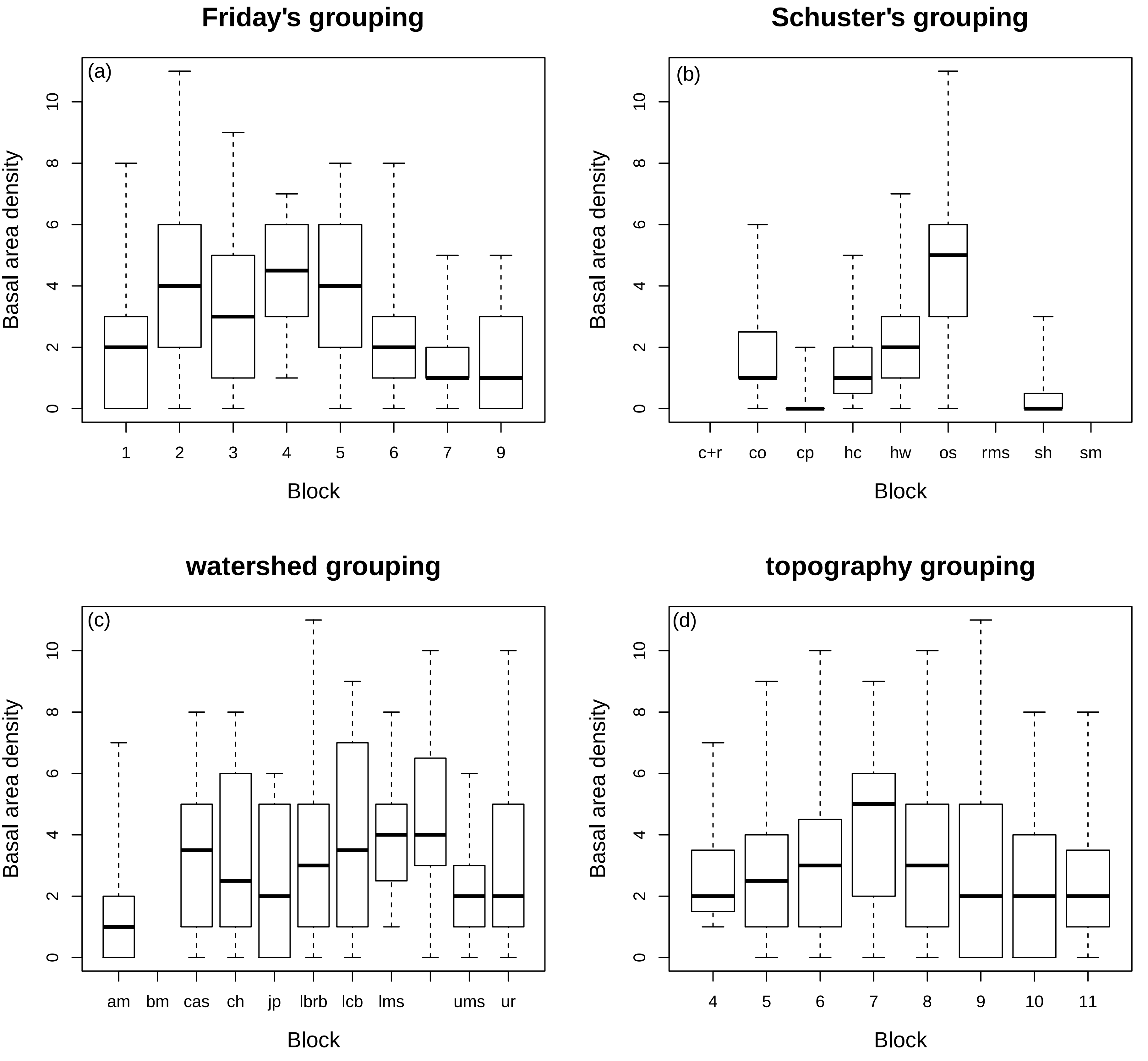

Friday's grouping
(a)
Basal area density
0
2
4
6
8
10
1
2
3
4
5
6
7
9
Block
Schuster's grouping
(b)
Basal area density
c+r
co
cp
hc
hw
os
rms
sh
sm
Block
watershed grouping
(c)
Basal area density
am
bm
cas
ch
jp
lbrb
lcb
lms
ums
ur
Block
topography grouping
(d)
Basal area density
4
5
6
7
8
9
10
11
Block

Table 1. Estimating the slope (b), intercept ($\log(a)$), and standard error of the slope estimator in Taylor's law using the theoretical formulae (Eqn 3)-(Eqn 5) and linear regression for six probability distributions. 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 bootstrapped by repeating the corresponding procedure for 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 bootstrapped from the 10,000 random copies of the $n \times N$ iid observations.

Probability distribution	$\hat{b}$			$\widehat{\log(a)}$			$s(\hat{b})$			Quadratic coefficient
	Formula (analytic)	Formula (numeric)	Regression	Formula (analytic)	Formula (numeric)	Regression	Formula (analytic)	Formula (numeric)	Regression	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)

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

Joel E. Cohen[1,*] and Meng Xu[1,2]

[1]**Laboratory of Populations, Rockefeller University & Columbia University, 1230 York Avenue, Box 20, New York, NY 10065 United States**

[2]**Department of Mathematics and Physics, University of New Haven, 300 Boston Post Road, West Haven, CT 06516 United States**

[*]**Corresponding author: Joel E. Cohen, Laboratory of Populations, Rockefeller University & Columbia University, 1230 York Avenue, Box 20, New York, NY 10065 United States. Telephone: (+1) 212-327-8883. E-mail: cohen@rockefeller.edu**

Supporting Information

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

$$f(X) \approx f\big(E(X)\big) + \big(X - E(X)\big)\big\{\big(f'(x)\big)|_{x=E(X)}\big\},$$

$$E\big(f(X)\big) \approx f\big(E(X)\big) + \left\{\frac{f''(x)}{2}\,|_{\,x=E(X)}\right\} \cdot var(X),$$

$$var\big(f(X)\big) \approx \big\{\big(f'(x)\big)|_{x=E(X)}\big\}^2 var(X).$$

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. If $x > 0$ and $f(x) = \log(x)$, then $f'(x) = 1/x$, $f''(x) = -x^{-2}$. 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. Assume m_j is the sample mean of observations in block j and $E(m_j) = M > 0$. Then the approximations given by the delta method are $\log m_j \approx \log M + (m_j - M)/M$, $var(\log m_j) \approx V/(n_j M^2)$, $E(\log m_j) \approx \log M - V/(2n_j M^2)$.

Proof. In the delta method, we set $X = m_j$, $f(x) = \log(x)$. From Loève (3, p. 276, Exercise 5), Oehlert (1) showed essentially that for $q \geq 0$, $E\left\{\left|m_j - M\right|^{2(q+1)}\right\} = O\left(n_j^{-(q+1)}\right)$. We shall use this bound with $q = 0$, 1/2, and 1 separately. Applying Taylor's expansion to $\log m_j$ yields

$$\log m_j = \log M + \left(m_j - M\right)/M - \left(m_j - M\right)^2/(2M^2) + O\left(\left(m_j - M\right)^3\right).$$

Following Oehlert's notation, we define $g\left(m_j\right) = \log m_j$, and $A_2\left(m_j\right) = \log M + \left(m_j - M\right)/M - \left(m_j - M\right)^2/(2M^2)$. Because $M > 0$ and because the logarithmic function is infinitely differentiable in any open interval that contains M, by Taylor's theorem, there exists a finite constant $C > 0$, such that $\left|g\left(m_j\right) - A_2\left(m_j\right)\right| \leq C\left|\left(m_j - M\right)^3\right|$. From Oehlert (1) with $q = 1/2$, we have $E\left\{C\left|\left(m_j - M\right)^3\right|\right\} = O\left(n_j^{-3/2}\right)$. Therefore, as $n_j \to \infty$, for $1 < \eta < \frac{3}{2}$, $n_j{}^{\eta} \cdot E\left\{\left|g\left(m_j\right) - A_2\left(m_j\right)\right|\right\} = O\left(n_j{}^{\eta - \frac{3}{2}}\right) \to 0$. Here "→" denotes point-wise convergence. By the triangle inequality (4), $E\left(g\left(m_j\right)\right) = E\left(A_2\left(m_j\right)\right) + o\left(n_j{}^{-\eta}\right)$. After substitution, $E\left(\log m_j\right) = \log M + E\left(m_j - M\right)/M - E\left\{\left(m_j - M\right)^2\right\}/(2M^2) + o\left(n_j{}^{-\eta}\right) = \log M - V/(2M^2 n_j) + o\left(n_j{}^{-\eta}\right)$. Hence $E\left(\log m_j\right) \approx \log M - V/(2M^2 n_j)$. As $n_j \to \infty$, this leads to the first-order approximation $E\left(\log m_j\right) \approx \log M$.

Now we estimate $var(\log m_j)$ using the first-order Taylor expansion of $\log m_j$, namely, $\log m_j = \log M + (m_j - M)/M + O\left((m_j - M)^2\right)$. Denote $A_1(m_j) = \log M + (m_j - M)/M$. By Taylor's theorem, there exists a finite constant $C_1 > 0$, such that $|g(m_j) - A_1(m_j)| \leq C_1 \left|(m_j - M)^2\right|$. From Oehlert (1) with q = 0, we have $E\left\{C_1 \left|(m_j - M)^2\right|\right\} = O(n_j^{-1})$. We now approximate $E\left\{(\log m_j)^2\right\}$ using the delta method.

$$\begin{aligned}\{g(m_j)\}^2 &= \{g(m_j) - A_1(m_j) + A_1(m_j)\}^2 \\ &= \{g(m_j) - A_1(m_j)\}^2 + \{A_1(m_j)\}^2 + 2\{A_1(m_j)\} \cdot \{g(m_j) - A_1(m_j)\}.\end{aligned}$$

In other words,

$$\{g(m_j)\}^2 - \{A_1(m_j)\}^2 = \{g(m_j) - A_1(m_j)\}^2 + 2\{A_1(m_j)\} \cdot \{g(m_j) - A_1(m_j)\}.$$

Since $|g(m_j) - A_1(m_j)| \leq C_1 \left|(m_j - M)^2\right|$, $|g(m_j) - A_1(m_j)|^2 \leq C_1^2 \left|(m_j - M)^4\right|$. So

$$\{A_1(m_j)\} \cdot \{g(m_j) - A_1(m_j)\} \leq C_1 \log M \cdot \left|(m_j - M)^2\right| + \frac{C_1}{M}\left|(m_j - M)^3\right|.$$

$$\left|\{g(m_j)\}^2 - \{A_1(m_j)\}^2\right| \leq C_1^2 \left|(m_j - M)^4\right| + 2C_1 \log M \cdot \left|(m_j - M)^2\right| + \frac{2C_1}{M}\left|(m_j - M)^3\right|.$$

From Oehlert (1) using q = 1 for the first term on the right side, q = 0 for the second term, and q = 1/2 for the third term, the expectation of the right side of the above inequality is $O(n_j{}^{-1})$. As $n_j \to +\infty$, for $0 < \gamma < 1$, $n^\gamma E\left|\{g(m_j)\}^2 - \{A_1(m_j)\}^2\right| \leq O(n_j{}^{\gamma-1}) \to 0$. From the triangle inequality, $E\left[\{g(m_j)\}^2\right] = E\left[\{A_1(m_j)\}^2\right] + o(n_j{}^{-\gamma})$. Thus the approximate mean of $(\log m_j)^2$

is $E\left\{\left(\log m_j\right)^2\right\} \approx E\left[\left\{\log M + \left(m_j - M\right)/M\right\}^2\right] = E\left\{(\log M)^2 + 2\,(\log M)\left(m_j - M\right)/M + \left(m_j - M\right)^2/M^2\right\} = (\log M)^2 + V/(M^2 n_j)$.

Overall, the estimated variance of $\log m_j$ from the delta method using the first-order Taylor expansion of $\log m_j$ is $var\left(\log m_j\right) = E\left\{\left(\log m_j\right)^2\right\} - \left\{E\left(\log m_j\right)\right\}^2 \approx (\log M)^2 + V/(M^2 n_j) - (\log M)^2 = V/(M^2 n_j)$. This proves Lemma 1.

Lemma 2. Under the assumptions of Lemma 1, also assume v_j is the sample variance of observations in block j and $E(v_j) = V > 0$. Then the approximations given by the delta method are $\log v_j \approx \log V + (v_j - V)/V$, $var\left(\log v_j\right) \approx \left(\mu_4 - \frac{n_j-3}{n_j-1}V^2\right)/(n_j V^2)$, $E\left(\log v_j\right) \approx \log V - \frac{1}{2n_j}\left(\frac{\mu_4}{V^2} - \frac{n_j-3}{n_j-1}\right)$.

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

Lemma 3. Under the assumptions of Lemmas 1 and 2, the covariance of the sample mean and sample variance is $cov\left(v_j, m_j\right) = \mu_3/n_j$, where μ_3 is the third central moment.

Zhang (5) gives a proof of this classical formula, which has been known at least since 1903 (6, pp. 279, equation (xiii), 7, pp. 7, equation (xxvi), 8, pp. 479, equation (67), 9).

Proof of Theorem. When all blocks are weighted equally, the least-squares estimators of slope b and intercept $\log(a)$, and standard error of the slope estimator $s(\hat{b})$ are respectively (10, pp. 155)

$$\hat{b} = cov_+\left(\log v_j, \log m_j\right)/var_+\left(\log m_j\right),$$

$$\widehat{\log(a)} = mean_+(\log v_j) - \hat{b} \cdot mean_+(\log m_j)$$

$$s(\hat{b}) = \sqrt{\left[var_+(\log v_j)/var_+(\log m_j) - \{cov_+(\log v_j, \log m_j)\}^2/\{var_+(\log m_j)\}^2\right]/(N-2)}.$$

The notations $mean_+(\cdot)$, $var_+(\cdot)$, and $cov_+(\cdot,\cdot)$ 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

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

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

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

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

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

They are all consistent by the law of large numbers: as $N \to \infty$, $mean_+(\log m_j) \to_P E(\log m_j)$, $mean_+(\log v_j) \to_P E(\log v_j)$, $var_+(\log m_j) \to_P var(\log m_j)$, $var_+(\log v_j) \to_P var(\log v_j)$, and $cov_+(\log v_j, \log m_j) \to_P cov(\log v_j, \log m_j)$. Here the symbol "$\to_P$" 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, $cov_+(\log v_j, \log m_j)$, the first term is approximately

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

$$= \frac{N}{N-1} \cdot \log M \cdot \log V + \frac{\log V}{(N-1)M}\sum_{j=1}^{N}\left(m_j - M\right) + \frac{\log M}{(N-1)V}\sum_{j=1}^{N}\left(v_j - V\right)$$

$$+ \frac{1}{(N-1)MV}\sum_{j=1}^{N}\left(m_j - M\right)\left(v_j - V\right).$$

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

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

Therefore $cov_+\left(\log v_j, \log m_j\right) \approx \frac{1}{(N-1)MV}\sum_{j=1}^{N}\left(m_j - M\right)\left(v_j - V\right) - \frac{1}{N(N-1)MV}\sum_{j=1}^{N}\left(m_j - M\right)\sum_{j=1}^{N}\left(v_j - V\right) = \frac{1}{(N-1)MV}\sum_{j=1}^{N} m_j v_j - \frac{1}{N(N-1)MV}\sum_{j=1}^{N} m_j \sum_{j=1}^{N} v_j = \frac{cov_+(m_j, v_j)}{MV}$. Similarly, the denominator of $\hat{b}$ is approximately $var_+\left(\log m_j\right) \approx \frac{1}{M^2}\left\{\frac{1}{(N-1)}\sum_{j=1}^{N} m_j^2 - \frac{1}{N(N-1)}\left(\sum_{j=1}^{N} m_j\right)^2\right\} = var_+\left(m_j\right)/M^2$. Consequently, for large n_j, $j = 1, 2, \ldots, N$, $\hat{b} \approx \frac{cov_+(m_j, v_j)}{MV} \Big/ \frac{var_+(m_j)}{M^2}$. By consistency, for large N, using Lemma 3 in the numerator, $\hat{b} \approx \frac{cov(m_j, v_j)}{MV} \Big/ \frac{var(m_j)}{M^2} = \frac{\mu_3}{n_j MV} \Big/ \frac{V}{n_j M^2} = \mu_3 M/V^2 = \gamma_1/CV$.

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

$$\widehat{\log(a)} \approx E(\log v_j) - \hat{b} \cdot E(\log m_j)$$

$$\approx \left[\log V - \frac{1}{2n_j}\left(\frac{\mu_4}{V^2} - \frac{n_j - 3}{n_j - 1}\right)\right] - \frac{\gamma_1}{CV}\left[\log M - V/\left(2n_j M^2\right)\right]$$

$$\approx \log V - \frac{\gamma_1}{CV} \cdot \log M$$

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

References for Supporting Information

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* (John Wiley & Sons, Inc., New York), 2nd Ed.

3. Loève M (1977) *Probability Theory I* (Springer-Verlag, New York), 4th Ed.

4. Steele JM (2004) *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities* (Cambridge University Press, Cambridge, UK).

5. Zhang L (2007) Sample mean and sample variance: their covariance and their (in)dependence. *Am Stat* 61(2):159-160.

6. Editorial [probably K. Pearson, then the editor] (1903) On the probable errors of frequency constants. *Biometrika* 2(3):273-281.

7. Editorial [probably K. Pearson, then the editor] (1913) On the probable errors of frequency constants part II. *Biometrika* 9(1/2):1-10.

8. Neyman J (1925) Contributions to the theory of small samples drawn from a finite population. Biometrika 17(3/4):472-479.

9. 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.

10. Snedecor GW, Cochran WG (1980) *Statistical Methods* (Iowa State University Press, Ames, IA), 7th Ed.