

Life-history invariants with bounded variables cannot be distinguished from data generated by random processes using standard analyses

R. CIPRIANI* & R. COLLIN†

*Departamento de Estudios Ambientales, Universidad Simón Bolívar, Venezuela

†Smithsonian Tropical Research Institute, Panama

Keywords:

hypothesis testing;
linear regression;
 $L(x_m)/L_{inf}$;
Protandry;
sex change.

Abstract

A dimensionless approach to the study of life-history evolution has been applied to a wide variety of variables in the search for life-history invariants. This approach usually employs ordinary least squares (OLS) regressions of log-transformed data. In several well-studied combinations of variables the range of values of one parameter is bounded or limited by the value of the other. In this situation, the null hypothesis normally applied to regression analysis is not appropriate. We generate the null expectations and confidence intervals (CI) for OLS and reduced major axis (RMA) regressions using random variables that are bounded in this way. Comparisons of these CI show that, for log-transformed data, the patterns generated by random data and those predicted by life history invariant theory often could not be distinguished because both predict a slope of 1. We recommend that tests based on the putative invariant ratios and not the correlations between the two variables be used in the exploration of life-history invariants using bounded data. Because empirical data are often not normally distributed randomization test may be more appropriate than standard statistical tests.

Introduction

A dimensionless approach to the study of life-history evolution has been applied to a wide variety of relationships among reproductive and demographic variables (e.g. Juergens & Prothero, 1987; Charnov & Berrigan, 1990; Shine & Charnov, 1992; Charnov, 1993; Prothero, 1993; Morand, 1996; Gemmill *et al.*, 1999; Charnov, 2000; He & Stewart, 2001; Jones & MacLarnon, 2001; Allsop & West, 2003a,b; Williams & Shertzer, 2003). The idea is that certain relationships between life-history characteristics are invariant with respect to transformations across a variety of taxa and that such patterns of invariants and deviations from them can tell us something interesting about the organizing principles of life-

history evolution (Charnov, 1993). As most commonly applied, two traits of interest are examined for the shape of their scaling relationship across species. If the relationship is linear then the ratio of the two is invariant to transformations of the other features that may also vary across the same set of taxa. Charnov (1993) gives the example that the regression of clutch size on mortality rate gives a linear relationship across 14 bird species. This implies that the dimensionless ratio of the two changes little despite considerable variation in body size and other biological features of the species sampled.

Studies looking for life-history invariants have used ordinary least squares (OLS) [and sometimes reduced major axis (RMA) or major axis] linear regressions of log-transformed data to discover invariant relationships (Charnov, 1993; Jones & MacLarnon, 2001; Allsop & West, 2003a,b). An estimated slope of 1 on the log-transformed data demonstrates a proportional relationship between the variables while slopes different from 1 show nonlinear relationship. The level of correlation of the variables (r^2) is

Correspondence: Rachel Collin, Smithsonian Tropical Research Institute, Unit 0948, APO AA 43002, USA.
Tel.: +1-507-2128766; fax: +1-507-2128790;
e-mail: collinr@naos.si.edu

considered to reflect the strength or robustness of the hypothesized invariance with respect to variation in other demographic parameters. However guidelines on how much variation is acceptable in 'invariant' has not been explicitly addressed to our knowledge (see Charnov, 1993, p. 5). Those studies interested in the actual value of the ratio between the variables consider either the average of the observed ratios or the slope of the regression as the estimate of the value (e.g. Charnov & Berrigan, 1991; Allsop & West, 2003a).

This regression approach is useful when the assumptions of linear regression analyses are met and when the appropriate null hypothesis is tested. It has been widely applied to a variety of questions in evolutionary ecology and fisheries science (e.g. Juergens & Prothero, 1987; Shine & Charnov, 1992; Charnov, 1993; Morand, 1996; Gemmill *et al.*, 1999; Charnov, 2000; He & Stewart, 2001). However, there are several cases of proposed life-history invariants where the range of one variable is bounded by the value of the other and therefore the null hypothesis normally used for regressions of two unbounded variables is no longer appropriate. Two proposed invariant relationships relating to sex change (1) Age at maturity vs. age at sex change (Age_{mat}/Age_{50}) and (2) size at sex change vs. maximum size (L_{50}/L_{max}) and two other invariant relationships (3) age (or size) at maturity vs. maximum age (or size) (Age_{mat}/Age_{max}) and (4) weaning weight vs. maternal weight are bounded in this way. In all of these examples one of the variables cannot exceed the other (e.g. age at maturity cannot exceed maximum age), so $Y < X$. It is intuitively reasonable to expect that regressions on randomly distributed data subject to the constraint $Y < X$ would result in a nonzero slope. Therefore, the conclusions that can be drawn on the basis of hypothesis testing using results of regression analysis, as normally applied to slopes (as implemented in Shine & Charnov, 1992; Charnov, 1993; Charnov & Skúladóttir, 2000; Allsop & West, 2003a,b) are limited.

Simulations and empirical data

We took a simulation approach to demonstrate the random expectations of regression analysis of data subject

to the constraint $Y < X$. We used a program written in R (Version 1.9.0) to generate 5000 replicates each of 15, 25, 50, 75 and 100 observations of X and Y randomly sampled from a normal distribution but subject to the constraint of $Y < X$. We calculated two-tailed 95% confidence intervals (CI) for the slope and r^2 from ordinary least squares and reduced major axis regression analyses of raw data and on the log-transformed data (Fig. 1).

We re-analysed seven datasets from the literature (Table 1; Loehle, 1988; Shine & Charnov, 1992; Charnov & Skúladóttir, 2000; Allsop & West, 2003a,b) to compare the results from OLS and RMA regressions of raw and log-transformed data to the CI generated from random data. Because X and Y are constrained, the slope of regressions of bounded data is dependent on both the mean and standard deviation (SD) of both variables (Fig. 2). This effect is particularly pronounced when the range of the Y variable covers only a small part of the range of the X values, or when the ranges of the two variables do not overlap (e.g. the data from Loehle, 1988). Therefore, we used re-sampling techniques individually tailored for each dataset to test explicit hypotheses.

To determine if the slopes are different from those generated by random data we generated a random expectation for each empirical dataset, conforming to the assumptions of regression analysis by (1) Randomly drawing a number within the range of the empirical X data, (2) randomly drawing a number from a normal distribution between the minimum Y value of the empirical data and the selected X value, (3) repeating this to generate a dataset the size of the empirical dataset and (4) repeating this 10 000 times, calculating the regression statistics each time.

To test the significance of regression slopes (i.e. null hypothesis slope = 0) and to test for a slope of 1, we used a re-sampling method outline by Manly (1997). One-tailed t -tests were also used to test for a slope of 1 (following McArdle, 1988 for RMA regressions) for comparison with previous studies. Finally, CI of the slopes was generated using standard bootstrapping.

We tested for 'invariance' of the Y/X ratio by determining if the empirical values are less variable than expected at random. The mean and SD were calculated and compared

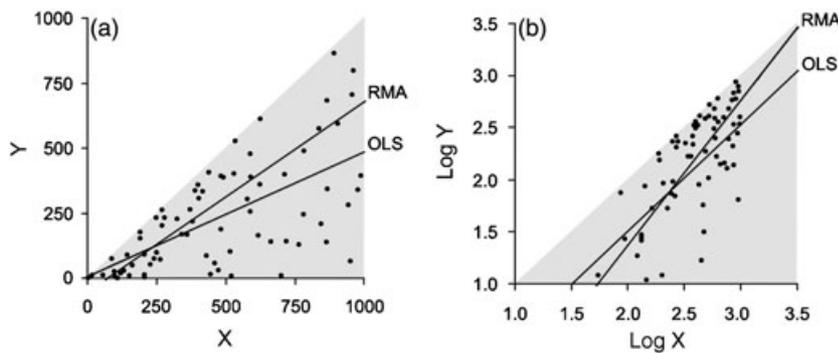


Fig. 1 Scatter plots showing one randomly generated dataset bounded by $Y < X$ (the grey area). (a) The raw data and (b) the same data after log-transformation. The OLS and RMA estimates for these data are drawn through the points. Raw data OLS (slope = 0.47, intercept = 10.9, $r^2 = 0.42$) and RMA (slope = 0.73). Log-transformed data OLS (slope = 1.02, intercept = -0.53, $r^2 = 0.53$) and RMA (slope = 1.40).

Table 1 Regression statistics from analysis of random data under the constraint $Y < X$ and from published studies of empirical data with the same constraint.

Taxon	Sample size	OLS slope (95% CI)†	RMA slope (95% CI)†	r^2 (95% CI)	References
Raw data					
Random data	45	0.50 (0.39, 0.61)	0.60 (0.50, 0.72)	0.69 (0.50, 0.84)	Shine & Charnov (1992)
Age _{maturity} /Age _{max} Squamates		0.64	0.63	0.96	
Random data	77	0.50 (0.42, 0.58)	0.60 (0.52, 0.68)	0.70 (0.56, 0.81)	Allsop & West (2003a))
L_{50}/L_{max} Animals		0.81	0.82	0.98	
Random data	52	0.50 (0.40, 0.60)	0.60 (0.51, 0.70)	0.70 (0.53, 0.83)	Allsop & West (2003b))
L_{50}/L_{max} Fishes		0.85	0.86	0.99	
Random data	14	0.50 (0.29, 0.71)	0.61 (0.42, 0.84)	0.69 (0.31, 0.92)	Allsop & West (2003b))
Age _{maturity} /Age ₅₀ Fishes		0.34	0.36	0.82	
Random data	35	0.50 (0.35, 0.64)	0.64 (0.51, 0.79)	0.62 (0.38, 0.80)	Loehle (1988)
Age _{max} /Age _{maturity} Angiosperm trees		0.06	0.11	0.29	
Random data	43	0.50 (0.38, 0.62)	0.61 (0.50, 0.73)	0.67 (0.47, 0.82)	Loehle (1988)
Age _{max} /Age _{maturity} Gymnosperm trees		0.018	0.04	0.27	
Random data	21	0.50 (0.34, 0.66)	0.60 (0.45, 0.77)	0.70 (0.41, 0.89)	Charnov & Skúladóttir (2000)
L_{50}/L_{max} Shrimp		0.83	0.86	0.92	
Log-transformed					
Random data	45	0.85 (0.74, 0.97)	0.97 (0.85, 1.13)	0.79 (0.56, 0.90)	Shine & Charnov (1992)
Age _{maturity} /Age _{max} Squamates		0.94 (0.91, 0.98)*	0.95 (0.91, 0.98)*, ‡	0.99	
Random data	77	0.99 (0.90, 1.09)	1.08 (0.98, 1.22)	0.84 (0.69, 0.93)	Allsop & West (2003a))
L_{50}/L_{max} Animals		1.05 (1.02, 1.08)*, ‡	1.06 (1.03, 1.09)*, ‡	0.98	
Random data	52	0.92 (0.81, 1.03)	1.02 (0.91, 1.19)	0.81 (0.61, 0.92)	Allsop & West (2003b))
L_{50}/L_{max} Fishes		1.06 (1.01, 1.12)*, ‡	1.08 (1.02, 1.14)*, ‡	0.97	
Random data	14	0.88 (0.61, 1.16)	1.03 (0.80, 1.44)	0.76 (0.35, 0.94)	Allsop & West (2003b))
Age _{maturity} /Age ₅₀ Fishes		0.78 (0.59, 0.91)*, ‡	0.85 (0.71, 0.99)*, ‡	0.87	
Random data	35	0.95 (0.71, 1.18)	1.19 (0.96, 1.52)	0.66 (0.35, 0.84)	Loehle (1988)
Age _{maturity} /Age _{max} Angiosperm trees		0.43 (0.18, 0.71)*	0.89 (0.66, 1.17)*, ‡, §	0.25	
Random data	43	0.98 (0.81, 1.16)	1.13 (0.97, 1.39)	0.76 (0.5, 0.89)	Loehle (1988)
Age _{maturity} /Age _{max} Gymnosperm trees		0.29 (0.08, 0.48)*	0.71 (0.56, 0.95)*	0.18	
Random data	21	0.56 (0.39, 0.71)	0.66 (0.51, 0.84)	0.71 (0.44, 0.90)	Charnov & Skúladóttir (2000)
L_{50}/L_{max} Shrimp		1.02 (0.80, 1.22)*, ‡, §	1.08 (0.86, 1.26)*, §	0.90	

Bootstrap test of significance of the regression obtained for testing $\beta = 0$ by randomly combining the observed X and Y values from each empirical dataset as outlined by Manly (1997, p. 149).

Slopes that are significantly different from 0 are indicated in bold.

*Slope of log-transformed data not significantly different from 1 using bootstrap test for slope as outlined by Manly (1997, p. 149).

†Random data from a normal distribution.

‡Slope of log-transformed data not significantly different from 1 using a one-tailed t -test.

§Bootstrap CI on the slope includes 1.

to the CI of the mean and SD generated from both random data generated as described above and from randomized Y/X datasets generated by re-sampling the empirical X and Y values independently, subject to $Y < X$.

Results

Linear regression analysis of randomly generated data shows that the expected OLS slope is 0.50 for the raw data and is close to 1 (0.99) for the log-transformed data. The expected RMA slopes are 0.76 for raw data and 1.4 for log-transformed data. The 95% CI of the OLS slope of log-transformed random data and the 95% CI for RMA slopes with small sample sizes both include 1. Because both processes are expected to produce log-transformed data with a slope of 1 under the constraint of $Y < X$,

standard regression analysis cannot be used to distinguish between patterns of data produced by underlying organizing principles of life history evolution and those generated by random processes. In addition, the average randomly generated r^2 values were 0.43 and 0.5 but could be as high as 0.7 or 0.8. Results from uniformly, normally, log-normally and exponentially distributed data were similar.

Hypothesis testing using the empirical data reported in previously published papers are presented in Tables 1 and 2. The 95% CI of the slope of the log-transformed dataset-specific random data include 1 for five of the seven OLS regressions and seven of the seven RMA regressions (Table 1), again supporting the idea that the predictions of life history invariants cannot be distinguished from randomly generated data. The same five

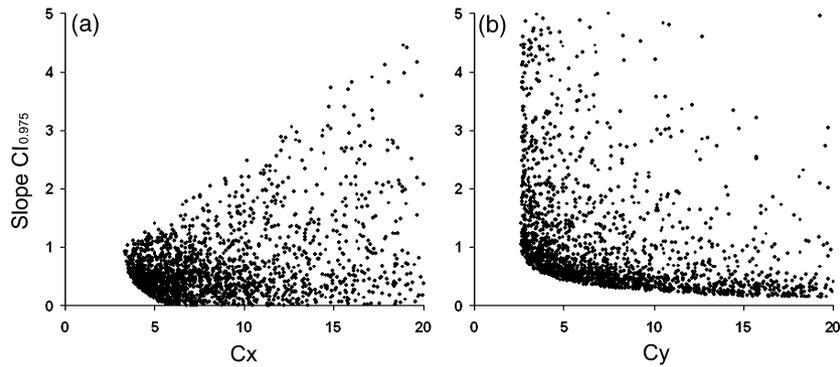


Fig. 2 Scatter plots showing the dependence of the 97.5% CI of the slopes of OLS regressions fitted to the log-transformed data on the mean and SD of the X and Y variables when Y is bounded by X . (a) CI vs. the co-efficient of variation (mean of X /SD of X) of variable X (C_x) and (b) CI vs. the co-efficient of variation of variable Y (C_y). Each of the 3000 points was calculated from 1000 simulations of 100 constrained $Y < X$ data values, drawn from normal distributions. The mean value of X was defined as the upper limit of the values of Y , in order to simulate partial overlapping of both distributions.

Table 2 Means and SD of ratios of bounded data from bootstrap re-sampling and empirical data.

Data set	Sample size	L_{50}/L_{max}		References
		Mean (CI)	Mean SD (CI)	
Random data	45	0.53 (0.47, 0.59)	0.16 (0.13, 0.19)	This study
Random bootstrapped data		0.44 (0.33, 0.56)	0.30 (0.26, 0.34)	This study
$Age_{maturity}/Age_{max}$		0.72* , †	0.090* , †	Shine & Charnov (1992)
Random data	77	0.50 (0.45, 0.55)	0.16 (0.14, 0.19)	This study
Random bootstrapped data		0.32 (0.24, 0.39)	0.28 (0.23, 0.32)	This study
L_{50}/L_{max}		0.74* , †	0.13* , †	Allsop & West (2003a)
Random data	52	0.51 (0.45, 0.57)	0.16 (0.13, 0.19)	This study
Random bootstrapped data		0.44 (0.35, 0.53)	0.27 (0.23, 0.31)	This study
L_{50}/L_{max}		0.78* , †	0.11* , †	Allsop & West (2003b)
Random data	14	0.53 (0.42, 0.63)	0.15 (0.10, 0.22)	This study
Random bootstrapped data		0.34 (0.21, 0.48)	0.20 (0.11, 0.29)	This study
$Age_{maturity}/Age_{50}$		0.42†	0.13	Allsop & West (2003b)
Random data	35	0.51 (0.45, 0.58)	0.16 (0.12, 0.20)	This study
Random bootstrapped data		0.16 (0.11, 0.22)	0.11 (0.06, 0.17)	This study
$Age_{maturity}/Age_{max}$		0.14†	0.071†	Loehle (1988)
Random data	43	0.50 (0.44, 0.57)	0.16 (0.13, 0.20)	This study
Random bootstrapped data		0.11 (0.07, 0.15)	0.09 (0.05, 0.14)	This study
$Age_{maturity}/Age_{max}$		0.09†	0.062†	Loehle (1988)
Random data	21	0.87 (0.84, 0.90)	0.07 (0.05, 0.09)	This study
Random bootstrapped data		0.87 (0.82, 0.91)	0.08 (0.05, 0.10)	This study
L_{50}/L_{max}		0.98* , †	0.009* , †	Charnov & Skúladóttir (2000)

*Falls outside the two-tailed 95% CI generated by creating ratios with $Y < X$ from subsampling the original X and Y values as outlined by Manly (1997). †Falls outside the two-tailed 95% CI generated by random data with $Y < X$. Bold values are significantly different from random expectations using both tests.

datasets had significant nonzero slopes as assessed by the Manly bootstrap test. The two slopes that are not significant could also not be distinguished from 1, indicating that the bootstrap test had low power in these cases. Four of the datasets had slopes that could not be distinguished from 1 using either the bootstrap CI or t -tests (Table 1). Four of these datasets have r^2 values that are significantly greater than the random expecta-

tion and the two that do not have significant slopes have r^2 values that are smaller than expected at random. Standard linear regressions of these data were originally interpreted as supporting the linear relationship between variables that is predicted by life history invariant theory (Shine & Charnov, 1992; Charnov & Skúladóttir, 2000; Allsop & West, 2003a,b) and they still cannot be distinguished from these predictions.

In the case of the actual ratios, the expected patterns generated from the 'invariant' hypothesis and from random processes are different. The invariant hypothesis predicts that the Y/X values will have less variation than expected at random. Examining the empirical ratios shows that four of the seven datasets have significantly larger average values and significantly smaller SD than expected from randomized data (Table 2). This result is supported by both randomization methods. These are the same four datasets that show slopes indistinguishable from 1. In combination these two results (log-transformed slope = 1 and variance less than expected) support the hypothesis that life-history invariant related processes rather than random processes contributed to these patterns in the data.

Discussion

Researchers looking for life-history invariants interpret regression analyses as follows: (1) the slope of the log-transformed data reflects the shape of the relationship between the two variables and is 1 if the relationship is an invariant, (2) the correlation (r^2) shows the robustness or strength of the invariance if the relationship is linear or the strength of allometry if the relationship is not linear and (3) the slope of the raw data or the average of the ratios give the value of the invariant. Our examination of randomly generated data shows that each of these steps should be taken cautiously when analysing bounded data.

Many previous studies using bounded data have used OLS regression of log-transformed data to test for a slope of 1 (e.g. Shine & Charnov, 1992; Jones & MacLarnon, 2001; Allsop & West, 2003a,b). Because this approach cannot distinguish a slope of 1 from the random expectation (i.e. the random data are correlated), conclusions that slopes that are statistically indistinguishable from 1 provide strong support for predictions of life-history invariant theory have to be viewed with caution. For those studies in Table 1, where the slopes were indistinguishable from 1 the SD of the ratios were also less than expected at random, supporting the original conclusions that the data support life-history invariance (Shine & Charnov, 1992; Charnov, 1993; Allsop & West, 2003a,b) and is further supported by examination of the ratios themselves.

A high r^2 is often cited as reflecting the robustness of the invariant relationship (Charnov, 1993; Allsop & West, 2003a). Our null expectations show that the ability of the independent variable to explain variation in the dependent variable (OLS) or a high correlation between the two (RMA) can be distinguished from random expectations, but that high (e.g. 0.8) r^2 values can be generated by random data. Therefore, although the commonly cited values of $r^2 = 0.8-0.9$ (e.g. Charnov, 1993) reflect a high correlation, they are only somewhat better than the random expectation and do not

necessarily indicate an unusually invariant relationship between the variables. Two of the empirical studies (Loehle, 1988; Table 1) had r^2 values significantly lower than the random expectation. These low values may be due to low slopes, because slope and r^2 values are positively correlated or because the range of the X and Y variables did not overlap. In all cases where r^2 is lower than random, the slope is also significantly lower than random expectations.

An alternate approach to the search for invariant relationships is to examine the range of variation of the proposed invariant ratio itself. Estimates of the value of life-history invariants have often been obtained by calculating the mean of the ratios obtained for each species (e.g. Allsop & West, 2003a). This approach produces a mean and a SD but it is not clear how to test for invariance. It is not possible to statistically test a distribution for zero variance. In any case, this approach would not be appropriate, as life-history invariants are predicted to be invariant to a variety of unspecified transformations but are not predicted to have no variance (Charnov, 1993). The SD of the empirical ratios of all but one dataset are significantly smaller than the SD of the distribution of ratios of bootstrapped and randomly generated variables with $Y < X$ (Table 2). Significantly small SD gives support for some type of relationship between the variables. However, as with the interpretation of r^2 , how small a variance is necessary to consider a ratio a 'life-history invariant' remains somewhat subjective (see Charnov, 1993, p. 5). If invariance with respect to a specific feature (or transformation) were to be examined, a test for no correlation between Y/X and that feature would be appropriate.

Conclusions

Patterns predicted by life-history invariant theory cannot be distinguished from data generated by random processes by comparing OLS regression slopes of the log-transformed data, although the slope can be used to reject either of these hypotheses. A completely new type of regression model that takes the bounding explicitly into account is probably necessary to adequately analyse these kinds of data. However such a method is not currently available. There are several ways, using current methodologies, in which data that support life history invariant theory can be distinguished from those generated at random. In the absence of quantitative predictions of life-history data, we suggest that researchers use re-sampling methods to examine the variance of the ratios themselves. Ideally, models of life-history invariants could be used to generate explicit quantitative predictions of invariants like L_{50}/L_{max} . Empirical values could then be compared directly to these predictions either using the ratios directly, or via the slope of the regressions of the untransformed data.

Acknowledgments

We thank Dave Allsop for generously providing us with his data. P. Borges, H. Lessios, H. P. de Vlarar and two anonymous reviewers commented on a previous version of this manuscript. We also thank Decanato de Investigación y Desarrollo of the Universidad Simon Bolivar and the Smithsonian Tropical Research Institute for supporting this work.

References

- Allsop, D.J. & West, S.A. 2003a. Changing sex at the same relative body size. *Nature* **425**: 783–784.
- Allsop, D.J. & West, S.A. 2003b. Constant relative age and size at sex change for sequentially hermaphroditic fish. *J. Evol. Biol.* **16**: 921–929.
- Charnov, E.L. 2000. Evolution of life-history variation among species of altricial birds. *Evol. Ecol. Res.* **2**: 365–373.
- Charnov, E.L. 1993. *Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology*. Oxford University Press, Oxford, 167 pp.
- Charnov, E.L. & Berrigan, D. 1991. Evolution of life history parameters in animals with indeterminate growth, particularly fish. *Evol. Ecol.* **5**: 63–68.
- Charnov, E.L. & Skúladóttir, U. 2000. Dimensionless invariants for the optimal size (age) of sex change. *Evol. Ecol. Res.* **2**: 1067–1071.
- Gemmill, A.W., Skorping, A. & Read, A.F. 1999. Optimal timing of first reproduction in parasitic nematodes. *J. Evol. Biol.* **12**: 1148–1156.
- He, J.X. & Stewart, D.J. 2001. Age and size at first reproduction of fishes: predictive models based only on growth trajectories. *Ecology* **82**: 784–791.
- Jones, K.E. & MacLarnon, A. 2001. Bat life histories: testing models of mammalian life-history evolution. *Evol. Ecol. Res.* **3**: 465–476.
- Juergens, K.D. & Prothero, J. 1987. Scaling of maximal lifespan in bats. *Comp. Biochem. Physiol. A* **88**: 361–368.
- Loehle, C. 1988. Tree life histories: the role of defenses. *Can. J. For. Res.* **18**: 209–222.
- Manly, B.F.J. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd edn. Chapman & Hall, London, p. 399.
- McArdle, B.H. 1988. The structural relationship: regression in biology. *Can. J. Zool.* **66**: 2329–2339.
- Morand, S. 1996. Life-history traits in parasitic nematodes: a comparative approach for the search of invariants. *Funct. Ecol.* **10**: 210–218.
- Prothero, J. 1993. Adult life span as a function of age at maturity. *Exp. Gerontol.* **28**: 529–536.
- Shine, R. & Charnov, E.L. 1992. Patterns of survival, growth and maturation in snakes and lizards. *Am. Nat.* **139**: 1257–1269.
- Williams, E.H. & Shertzer, K.W. 2003. Implications of life-history invariants for biological reference points used in fishery management. *Can. J. Fish Aquat. Sci.* **60**: 710–720.

Received 28 December 2004; revised 10 March 2005; accepted 07 April 2005