

# Estimation Issues in Allometric Scaling Models

M. Bulmer and K. Fleetwood

Department of Mathematics, The University of Queensland, Brisbane QLD 4072, Australia.  
m.bulmer@uq.edu.au

**Abstract** Allometric scaling models are a common tool for describing and quantifying relationships between biological measurements. Recent years have seen a lively debate on the parameter values for certain relationships. For example, allometric scaling of basal metabolic rate with body mass has historically been found to follow Kleiber's law, a  $3/4$  power relationship, in contrast to the  $2/3$  power expected from simple geometry. In this paper we examine some of the issues involved in estimating parameter values for power law models. We will illustrate some simple techniques that can be used to enrich this area of biological modelling.

**Keywords:** *Allometric scaling; Estimation; Bootstrap.*

## 1 INTRODUCTION

Recent years have seen a surge of interest in biological power laws. A prime example is the modelling of metabolic rate with body mass. Allometric scaling of basal metabolic rate with body mass has generally been found to follow Kleiber's law, a  $3/4$  power relationship (Kleiber, 1961). This is in contrast to the  $2/3$  power expected from simple geometry, and its wide applicability has been difficult to explain. West et al. (1999) have shown that a fractal model of nutrient supply networks can lead to the  $3/4$  power, while Darveau et al. (2002) use a distributed control model to explain this and the different powers that arise for maximum metabolic rates. Dodds et al. (2001) take an opposing view, suggesting that there is in fact little evidence to reject the  $2/3$  power.

The estimation of the parameters in these "laws", particularly the exponent, is thus an important topic. The aim of this paper is to use a case study to compare some simple methods for parameter estimation and discuss the issues associated with them. Section 2 describes the data set and how its observations were obtained. Section 3 shows estimates using standard linear regression, while Section 4 presents the obvious alternative of nonlinear fitting. In Section 5 we use the bootstrap to provide confidence intervals for parameter estimates in both methods, and this approach is extended to multiple nonlinear regression in Section 6.

## 2 DATA

The data used in this paper came from experimental work by Wilson and Franklin (2000) on the burst swimming performance of tadpoles. The aim of the work was to quantify the relationships between various indicators of swimming performance, such as maximum velocity and acceleration, with body characteristics, such as total body length. The experiments were conducted at two different temperatures to see whether water viscosity had an effect on these relationships.

Tadpoles of the Striped Marsh Frog (*Limnodynastes peronii*) were collected in Brisbane and kept in aquaria at  $24^{\circ}\text{C}$  for 6 weeks. Individuals with a range of body lengths were selected for the experiment but all came from a similar developmental stage, using a standard method for classification (Gosner, 1960). This helped eliminate the developmental stage as a confounding variable and should thus provide clearer patterns of association between the attributes of interest.

Each tadpole's swimming performance was measured from 10 swimming sequences. These sequences were recorded using a high-speed video camera and then analyzed frame by frame to determine maximum velocity and acceleration. The fastest of the 10 sequences was taken as the measure

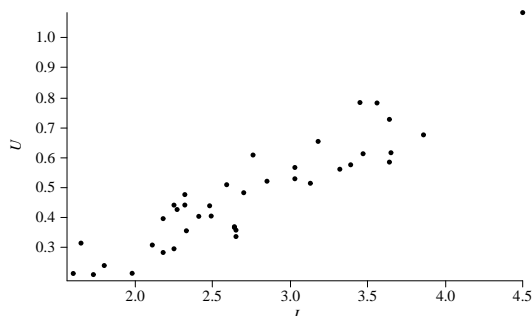
of maximum performance. The distance moved in this sequence between the first full tail beat was also recorded to calculate stride length. The time taken for the first full tail beat was used to calculate tail-beat frequency.

This data set is useful for our purposes since observations were made on a number of variables which exhibit a range of allometric relationships. The authors Wilson and Franklin (2000) report power relationships involving a number of different exponents, some significant and some not. We will look at the effect of total body length,  $L$ , on the maximum velocity,  $U$ , the maximum acceleration,  $A$ , and the tail-beat frequency,  $F$ . The experiment was repeated with the water cooled to 10°C. Researchers are also interested in how the performance measurements might change as temperature,  $T$ , changes.

To give an idea of the scales involved here, Table 1 shows summary statistics for these four variables of interest. It is clear from these values that temperature does have an effect on burst swimming performance. The question is whether the nature of the relationship with body length is the same at the different temperatures. This can be answered by comparing estimates of the allometric exponent  $b$ .

**Table 1:** Summary statistics for total length ( $L$ ), maximum velocity ( $U$ ), maximum acceleration ( $A$ ) and tail-beat frequency ( $F$ )

	Units	$T$	Mean	St Dev
$L$	cm	-	2.76	0.6864
$U$	m/s	24°C	0.48	0.1839
		10°C	0.24	0.0804
$A$	m/s <sup>2</sup>	24°C	1.96	0.7414
		10°C	0.54	0.2414
$F$	Hz	24°C	22	4.116
		10°C	7.8	1.479



**Figure 1:** Maximum tadpole velocity  $U$  (m/s) at 24°C and total length  $L$  (cm)

Figure 1 shows a plot of maximum velocity against total length for the trials at 24°C. The bulk of the

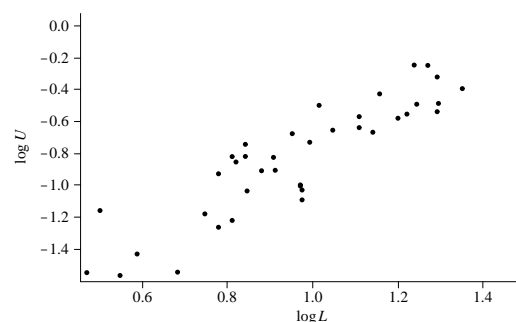
relationship appears quite linear but we will see that there is evidence of an exponent greater than 1.

### 3 LINEAR REGRESSION

The task of allometric modelling is to estimate the parameters  $a$  and  $b$  in the function  $y = ax^b$ . The standard approach to this problem is to start by taking logarithms of both sides, giving

$$\log y = \log a + b \log x.$$

This is then a linear relationship between  $\log y$  and  $\log x$ , and  $\log a$  and  $b$  can be estimated using simple linear regression. Figure 2 shows a log-log plot of the relationship between maximum velocity and total length for the 24°C trials.



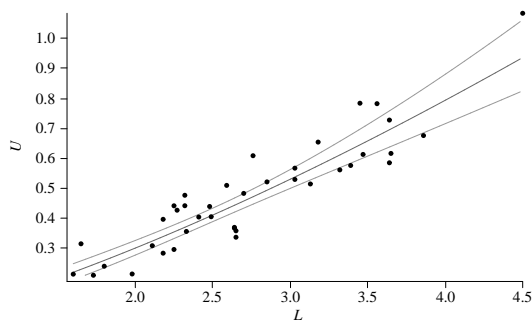
**Figure 2:** Log-log plot maximum velocity ( $U$ ) by total length ( $L$ )

The least-squares fit to the log-transformed data estimates  $a = 0.114$  and  $b = 1.40$ , giving the power law  $U = .114L^{1.40}$ . This equation is plotted with the original data in Figure 3. The tadpole that was 4.5 cm long is an influential point in this data. However, its maximum velocity of 1.09 m/s fits the general pattern well; removing it only changes the exponent  $b$  from 1.40 to 1.37. Table 2 gives the log-based estimates of the other parameters of interest for the two trial temperatures.

**Table 2:** Log-based estimates of allometric equations for maximum velocity ( $U$ ), maximum acceleration ( $A$ ) and tail-beat frequency ( $T$ )

	$T$	$\hat{a}$	$\hat{b}$
$U$	24°C	0.114	1.40
	10°C	0.061	1.33
$A$	24°C	0.524	1.21
	10°C	0.171	1.08
$F$	24°C	26.0	-0.200
	10°C	9.59	-0.228

Regression analysis can be used to construct a 95% confidence interval for the underlying linear relationship in the log-transformed data. This range can be inverse transformed to obtain a 95% interval for the allometric estimates, as shown in Figure 3.



**Figure 3:** Allometric fits of maximum velocity ( $U$ ) by total length ( $L$ ) using log-based method with 95% confidence intervals

#### 4 NONLINEAR LEAST-SQUARES

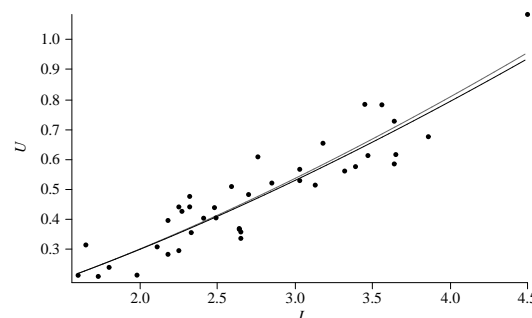
Using log-transformed data is an easy approach to parameter estimation and also brings with it the theory of linear models. The calculations required can all be performed by hand, particularly for small data sets, and so this method has been historically useful. However, with computational assistance it is straightforward to fit a power function directly. In Section 5 we will then use the bootstrap to give confidence limits for these estimates.

Here we can use Newton's method to find the values of  $a$  and  $b$  which minimize the sum of the squared deviations between the curve  $y = ax^b$  and the observed data. This provides a good comparison with the linear least-squares method, but other algorithms can also be used to fit non-differentiable criteria such as minimizing the maximum deviation or the sum of absolute deviations (Bulmer and Eccleston, 1998). Table 3 gives the resulting estimates for the variables of interest.

**Table 3:** Nonlinear estimates of allometric parameters

	$T$	$\hat{a}$	$\hat{b}$
$U$	24°C	0.112	1.43
	10°C	0.072	1.18
$A$	24°C	0.548	1.28
	10°C	0.178	1.09
$F$	24°C	26.8	-0.216
	10°C	10.3	-0.284

For example, this approach fits the maximum velocity data using the allometric equation  $U = 0.112L^{1.426}$ . Figure 4 shows a plot of this fit to the original data, along with the log-based fit. The main difference between these two fits is that the nonlinear estimates are somewhat affected by the unusually long tadpole. Whereas dropping this observation for the log-based method resulted in only a small change in  $b$ , here its removal changes  $b$  from 1.43 to 1.32. Looking at Figure 2 we see that the log transform reduces the vertical influence of this point, whereas the nonlinear fit has to deal with the increasing scale of variability as  $L$  increases.



**Figure 4:** Allometric fits of maximum velocity ( $U$ ) by total length ( $L$ ) using log-based method (black) and nonlinear method (grey)

#### 5 BOOTSTRAP ESTIMATION

The advantage of linear regression is that its associated model can be used routinely to give confidence bounds for response estimates. However, these are based on certain assumptions, such as the normality and constant scale of variability in the response. In this section we use the bootstrap (Efron and Tibshirani, 1993) to provide nonparametric bounds for the parameters  $a$  and  $b$ , and for estimates based on the resulting allometric equation.

Bootstrap estimates were computed by taking a random sample, with replacement, of observations from the data set. The allometric parameters were then estimated, using either the log-based or nonlinear method. This process was then repeated 200 times to give 200 estimates for each parameter. The middle 95% of the range of estimates was then taken as a 95% bootstrap confidence interval. Table 4 shows the results for the variables of interest using log-based fitting, while Table 5 shows the results using the nonlinear method.

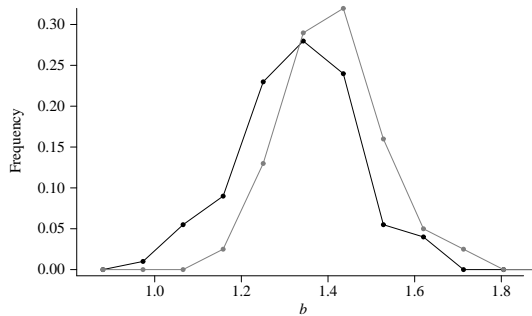
**Table 4:** 95% bootstrap intervals based on 200 estimates using log-based fitting

	$T$	$a$	$b$
$U$	24°C	(0.090, 0.145)	(1.18, 1.61)
	10°C	(0.047, 0.088)	(1.03, 1.59)
$A$	24°C	(0.431, 0.722)	(0.92, 1.53)
	10°C	(0.103, 0.257)	(0.62, 1.63)
$F$	24°C	(19.2, 33.2)	(-0.447, 0.064)
	10°C	(6.87, 13.2)	(-0.519, 0.059)

**Table 5:** 95% bootstrap intervals based on 200 estimates using nonlinear fitting

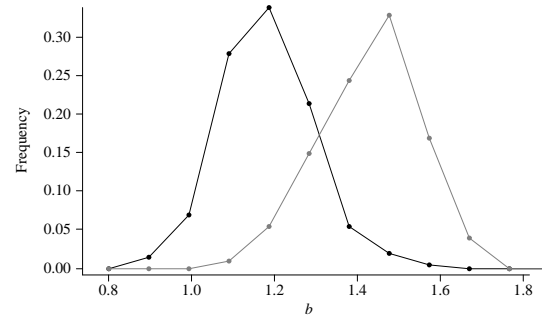
	$T$	$a$	$b$
$U$	24°C	(0.092, 0.153)	(1.11, 1.61)
	10°C	(0.059, 0.089)	(1.00, 1.35)
$A$	24°C	(0.383, 0.717)	(0.99, 1.61)
	10°C	(0.099, 0.261)	(0.69, 1.65)
$F$	24°C	(20.4, 34.7)	(-0.469, 0.025)
	10°C	(7.30, 15.1)	(-0.643, 0.056)

The bootstrap distributions of the maximum velocity exponent at 24°C and 10°C can be seen in Figure 5. It seems that the exponents for the two relationships are quite similar, as noted from standard statistical procedures by Wilson and Franklin (2000).



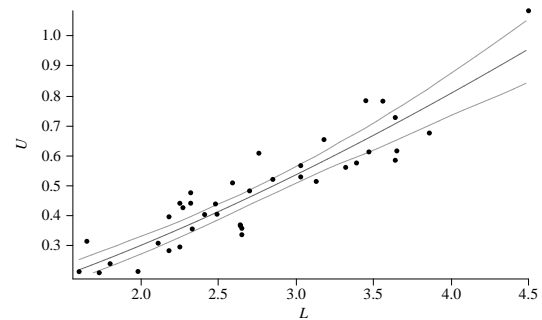
**Figure 5:** Distribution of log-based bootstrap estimates of maximum velocity exponent  $b$  at 10°C (black) and 24°C (grey)

We saw in Figure 4 that the nonlinear method gave quite similar results to the standard approach of fitting a straight line to log-transformed data. However, Figure 6 shows that the bootstrap distributions of the exponent have quite different locations to those seen in Figure 5. The difference between these can be attributed to the influential observation. Excluding this tadpole results in distributions that are very similar.



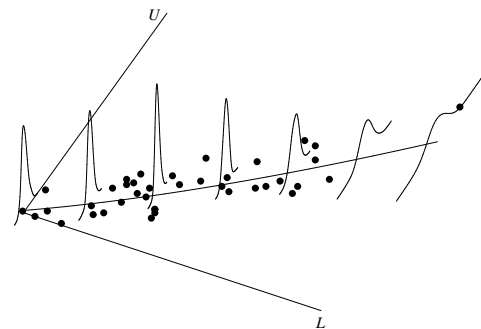
**Figure 6:** Nonlinear bootstrap distributions of maximum velocity exponent at 10°C (black) and 24°C (grey)

Bootstrap resampling can also be used to quantify the variability in the resulting estimates made from the model. This role is analogous to the calculation of 95% confidence interval bands in Figure 3.



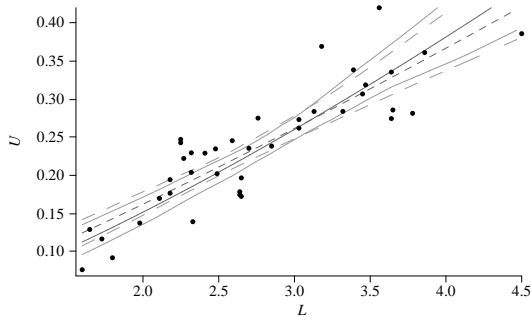
**Figure 7:** Nonlinear fit of maximum velocity with 95% bootstrap intervals

It is again useful to look at the bootstrap distributions in more detail. Figure 7 is essentially showing a series of 95% bootstrap intervals which have had their end points joined together into the two bands. Figure 8 shows the bootstrap distributions behind 7 of these intervals. Nonparametric density curves were computed for each distribution using a FFT with a Gaussian kernel (Silvermann, 1986).

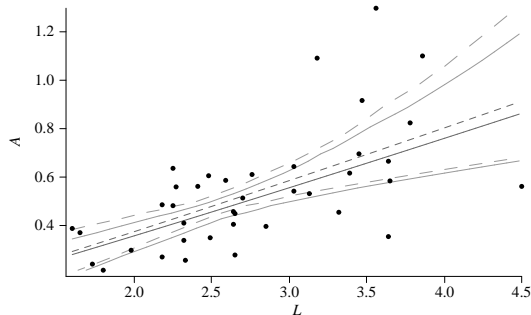


**Figure 8:** Bootstrap distributions of estimated maximum velocity

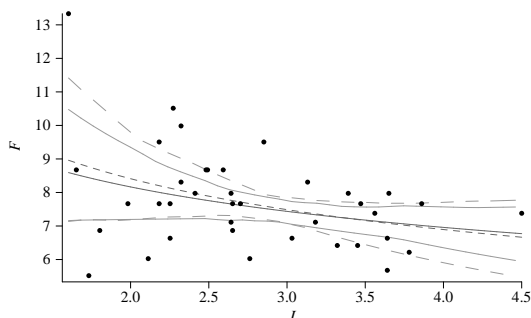
The following figures summarize the two fits to some of the other variables. Each includes 95% bootstrap intervals computed from 200 resamples. Figure 9 gives a comparison for maximum velocity in the trials at 10°C. Figures 10 and 11 gives similar comparisons for maximum acceleration and tail-beat frequency, respectively. Note that the model for tail-beat frequency has a negative exponent, although the results in Tables 4 and 5 suggest that there is no significant association with total length.



**Figure 9:** Allometric fits of maximum velocity ( $U$ ) at 10°C using log-based (solid) and nonlinear (dashed) methods with 95% bootstrap intervals



**Figure 10:** Allometric fits of maximum acceleration ( $A$ ) at 10°C using log-based (solid) and nonlinear (dashed) methods with 95% bootstrap intervals



**Figure 11:** Allometric fits of tail-beat frequency ( $F$ ) at 10°C using log-based (solid) and nonlinear (dashed) methods with 95% bootstrap intervals

## 6 MULTIPLE REGRESSION

The linear fit to the log-transformed data showed no evidence of a difference in the maximum velocity exponent  $b$  between the two temperatures. Assuming a common exponent, Wilson and Franklin (2000) estimated the effect of temperature on maximum velocity by fitting the function

$$\log U = \log a + b \log L + cT,$$

where  $T$  is the temperature. This leads to slightly odd looking power law (see below) but it is linear in the three parameters,  $a$ ,  $b$ , and  $c$ . These can be estimated by least-squares, as for the simple linear case. Here we obtain the fit

$$\log U = -3.34 + 1.36 \log L + 0.05T.$$

When  $T = 10$  this gives  $U = .058L^{1.36}$  and when  $T = 24$  we have  $U = .118L^{1.36}$ .

Regression theory can be used to compute confidence intervals for these three parameter estimates, as given in Table 6.

**Table 6:** 95% confidence intervals for multiple regression parameters in maximum velocity model

Parameter	95% Interval
$a$	(0.030, 0.043)
$b$	(1.21, 1.52)
$c$	(0.045, 0.056)

Inverse transforming the above linear function above gives the power function

$$U = ae^{cT}L^b.$$

As in Section 4, we can estimate  $a$ ,  $b$ , and  $c$  by directly minimizing the sum of squared deviations between this function and the observed values. This is a more complicated optimization problem than before. The Levenberg-Marquardt method (Marquardt, 1963) for nonlinear fitting was found to work well for a range of data sets, as is needed by bootstrap estimation. The parameter estimates obtained give the function

$$U = .034e^{.052T}L^{1.38}.$$

When  $T = 10$  this is  $U = .057L^{1.38}$  and when  $T = 24$  we have  $U = .118L^{1.38}$ . These are essentially the same values as those obtained by standard multiple regression. Table 7 gives 95% bootstrap confidence intervals from 200 resamples, also in good agreement to those in Table 6.

**Table 7:** 95% bootstrap intervals for nonlinear maximum velocity model

Parameter	95% Interval
$a$	(0.027, 0.045)
$b$	(1.15, 1.56)
$c$	(0.047, 0.057)

However, we saw earlier from the bootstrap distributions that the assumption of a common nonlinear exponent might not be plausible in this case. We could extend the model by adding an interaction term to give

$$U = ae^{cT} L^{b+dT},$$

where  $d$  is a new parameter which captures the effect of temperature on the exponent. The nonlinear least-squares fit gives

$$U = .056e^{.028T} L^{.94+.02T}.$$

When  $T = 10$  this simplifies to  $U = .074L^{1.14}$  and when  $T = 24$  we have  $U = .110L^{1.42}$ . Bootstrap intervals for the four parameters are given in Table 8. Since 0 is just within the interval for  $d$ , we would conclude that there is only weak evidence, if any, of a difference in exponent between the temperatures.

**Table 8:** 95% bootstrap intervals for extended nonlinear maximum velocity model

Parameter	95% Interval
$a$	(0.033, 0.085)
$b$	(0.54, 1.41)
$c$	(0.004, 0.055)
$d$	(-0.004, 0.044)

## 7 CONCLUSIONS

The allometric modelling presented in biological studies often relies entirely on the use of least-squares fitting to log-transformed data and associated statistical procedures, such as regression and analysis of variance. In this paper we have taken an example of biological research and given an overview of other simple possibilities for analysis.

Computationally intensive methods, such as nonlinear fitting and bootstrap resampling, can be used to enrich standard statistical modelling. We have seen that these methods can simply validate existing results or can show further details not apparent in the

initial analysis. In particular, robust methods like the bootstrap are useful for validating results when it is not clear that the standard model assumptions hold.

As illustrated by Darveau et al. (2002), there are also important scientific issues in parameter estimation that precede this analysis. Richer modelling tools can feed back into the science by giving greater insight into the results.

## Acknowledgments

This work would not have been possible without the data and background kindly provided by Robbie Wilson and Craig Franklin. Thanks also to John Eccleston for reading a draft of this paper.

## References

- Bulmer, M. and J. A. Eccleston, Automated reliability modelling with a genetic algorithm, in *Proceedings of the Sixth International Applied Statistics in Industry Conference*, 1–8, Melbourne, 1998.
- Darveau, C., R. K. Suarez, R. D. Andrews, and P. W. Hochachka, Allometric cascade as a unifying principle of body mass effects on metabolism, *Nature*, 417, 166–170, 2002.
- Dodds, P. S., D. H. Rothman, and J. S. Weitz, Re-examination of the “3/4-law” of metabolism, *Journal of Theoretical Biology*, 209, 9–27, 2001.
- Efron, B. and R. J. Tibshirani, *An introduction to the bootstrap*, Chapman & Hall, New York, 1993.
- Gosner, K. L., A simplified table for staging anuran embryos and larvae with notes on identification, *Herpetologica*, 16, 183–190, 1960.
- Kleiber, M., *The fire of life : an introduction to animal energetics*, Wiley, New York, 1961.
- Marquardt, D. W., An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431–441, 1963.
- Silvermann, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, 1986.
- West, G. B., J. H. Brown, and B. J. Enquist, The fourth dimension of life: Fractal geometry and allometric scaling of organisms, *Science*, 284, 1677–1679, 1999.
- Wilson, R. S. and C. E. Franklin, Effect of ontogenetic increases in body size on burst swimming performance in tadpoles of the striped march frog, *Limnodynastes peronii*, *Physiological and Biochemical Zoology*, 73, 142–152, 2000.