On Allometric Equations for Predicting Body Mass of Dinosaurs

Gavin C. Cawley* and Gareth J. Janacek
School of Computing Sciences
University of East Anglia
Norwich NR4 7TJ, U.K.

July 7, 2009

Abstract

Packard et al. (2009) investigate the prediction of the body mass of dinosaurs, using allometric models, advocating parameter estimation via direct optimisation of a least-squares criterion on arithmetic axes rather than the conventional approach based on linear least squares regression on logarithmic axes. In this letter, we examine the statistical assumptions underpinning each approach, and find the method of Packard to be conceptually unsatisfactory as it assumes absolute rather than relative variability in body mass for a given long bone circumference, which is biologically implausible. The proposed approach is thus unduly sensitive to small relative errors for large mammals; as the largest (the elephant) is comparatively light for its large-bone circumference, the resulting model grossly over-estimates the body mass of small mammals and is likely to substantially underestimate the body mass of dinosaurs. It is important to note, however, that the error bars for the conventional model already indicate substantial uncertainty in body mass, such that for example, the body mass of *Apatasaurus lousiae* may be as high as 63 metric tons, or as low as 23 metric tons, with a mean value of 38 metric tons.

---

*Corresponding author, email `gcc@cmp.uea.ac.uk`

1 Introduction

Packard et al. (2009) suggest that conventional allometric modelling practices substantially overestimate the body mass of dinosaurs (e.g. Anderson et al., 1985) based on measurements of long-bone circumference, because the logarithmic transformation involved imposes a misleading bias on the resulting model. Instead, Packard et al. advocate fitting the equivalent two parameter power function via direct minimisation of the least-squares error on the untransformed measurements. The direct minimisation of the least-squares error assumes that the variability in body mass can be expressed in absolute terms, however this is patently not the case in the prediction of body mass. A natural variation of 5 *kg* in the body mass of mammals with long bone circumferences similar to those of the yellow baboon would seem plausible, perhaps due to evolutionary adaption to different environment or food sources. A variation in body mass of 5 *kg* for mammals with long bone circumference measurements similar to those of a meadow mouse, however, is obviously absurd. Similarly a natural variation of only 5 *kg* in mammals with long bone circumferences like those of an elephant is also clearly unrealistic, as individuals within a single species are likely to vary more in body mass than 5 *kg*; indeed the seasonal and diurnal variability in the body mass of an individual elephant might well be greater than 5 *kg*. Thus there is a clear argument, based on biological plausibility, that the non-linear regression model, with its assumption of common absolute variability, is unsatisfactory *a-priori*. On the other hand, the conventional model assumes that the natural variability in body mass is greater for larger mammals than for smaller mammals, which is more closely in accord with our intuition.

Even though the non-linear regression approach may seem conceptually unsatisfactory, it may still be of value if it provides a good fit to the calibration data, within the limits specified by its underlying statistical assumptions (i.e. the data argue in favour of the non-linear regression model) and there is evidence to suggest that it might give more reliable predictions. In the next section, we briefly review these statistical assumptions underpinning each approach, and critically reappraise the model fits and the reliability of predictions.

2 Statistical Assumptions of Allometric Models

The underlying generative model for the conventional approach to allometric modelling, based on linear least-squares regression on log-transformed data, can be written as,

$$\log_{10} y_i = \beta_1 \log_{10} x_i + \beta_0 + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{1}$$

where x_i and y_i are the long bone circumference and body mass of the i^{th} observation, β_1 and β_0 are regression coefficients and ϵ_i represents an error term. This corresponds to the assumption that the logarithm of body mass can be modelled by a linear function of log-long bone circumference with additive zero mean Gaussian noise representing the uncertainties due to natural variation in body plan. The maximum likelihood estimate of the variance of the noise process, $\hat{\sigma}^2 \approx \sigma^2$, is given by

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where y_i and $\hat{y}_i$ are the true and predicted responses for the i^{th} observation respectively. The variance provides only the most basic indication of the level of the inherent uncertainty in the predictions. In practice there is also uncertainty in estimating the parameters of the model, however this added complication would only obscure the substantive issue, so here we construct error bars based solely on the variance of the noise process. Figure 1(a) depicts the conventional allometric model on logarithmic axes, showing the ±2 standard deviation error bars. The conventional model can also be expressed as a two-parameter power function, of the form

$$y_i = a x_i^b \times 10^{\epsilon_i} \qquad \epsilon_i \sim \mathcal{N}(0, \sigma), \tag{2}$$

where $a = 10^{\beta_0}$ and $b = \beta_1$. Expressed in this form, it is clear that the noise process is *heteroscedastic*, the variance of the noise increasing linearly with predicted body mass. This power function model is shown in Figure 1(b), along with the transformed ±2 standard deviation error bars; it is readily apparent that the uncertainty in the predicted body mass increases with the size of the creature, as illustrated by the error bars, which is in accord with our intuition. Inspection of (2) reveals that the model assumes that the *relative* uncertainty is constant as the error term is scaled linearly by predicted body mass. Note also that the statistical assumptions for this model cannot accommodate the idea of a mammal with a negative body mass through natural variability, as even the lower error bar is necessarily non-negative, a comforting feature of the model!

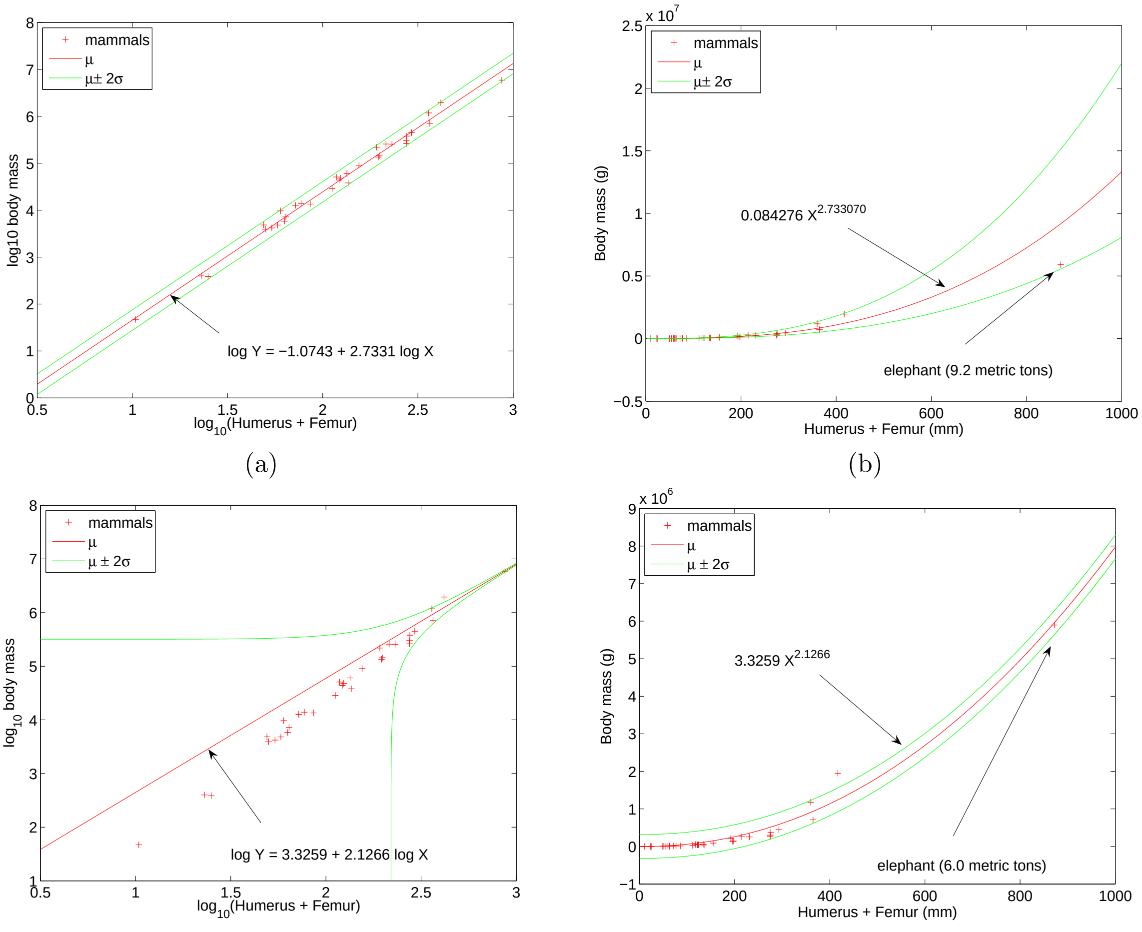
The underlying generative model for the approach proposed by Packard *et al.* assumes that body mass can similarly be modelled as a power function of long bone circumference, but this time with zero mean constant variance additive Gaussian noise, representing the effects of natural variability,

$$y_i = a x_i^b + \epsilon_i \qquad \epsilon_i \sim \mathcal{N}(0, \sigma), \tag{3}$$

In this case, the noise process is *homoscedastic* with the level of noise expected to be exactly the same, regardless of body mass. Figure 1(c) shows the fit for this model, including the ±2 standard deviation error bars, assuming additive zero mean homoscedastic Gaussian noise. The width of the error bars represents a uniform uncertainty of approximately ±318 kg across the scale; this represents a very low degree of uncertainty for an elephant with an observed body mass of 5897 kg, but extremely high degree of uncertainty for a meadow mouse with an observed body mass of 47 g! Note that unlike the conventional model, the non-linear regresssion model predicts the possibility of mammals with a negative body mass for a combined long bone circumference of $\approx 220mm$ or less (about the size of a blue wildebeest). While this does not represent a fatal flaw in the model, it is strongly suggestive that the underlying statistical assumptions are inappropriate for a strictly positive response variable, such as body mass.

Figure 1(d) shows the non-linear regression model transformed into logarithmic axes, showing very clearly that the non-linear regression model exhibits a very strong bias, consistently over-estimating the body mass of small animals. This occurs because penalising the absolute error, rather than the relative error, means that the model is very sensitive to the fit to large species at the expense of the fit to smaller species. Table 3 shows the predicted body mass using both the non-linear regression and back transformation models. The non-linear regression approach grossly exagerates the mass of small creatures, for example the meadow mouse is predicted to weigh 480 *g* instead of 47 *g*! Only the body mass of the bison and hippopotamus is under-estimated by this approach. On the other hand, for the back transformation approach, there is no such bias, over-estimating the mass of 16 creatures and under-estimating 17, with no obvious pattern. Again, the existence of a systematic bias is indicative of a poor choice of statistical assumptions.

Figure 2 shows the relative error as a function of log long-bone circumference. If a relative error of greater than 50% is regarded as a gross relative error, then the conventional approach exhibits only one gross relative error, substantially overestimating the body mass of the elephant. The non-linear regression approach on the other hand grossly overestimates the body mass of 24 of the 33 mammals,



(a) (b)

(c) (d)

Figure 1: Comparison of conventional back transformation (a and b) and non-linear regression (c and d) based allometric models of 33 mammals. The models are shown in logarithmically transformed (a and c) and original spaces (b and d).

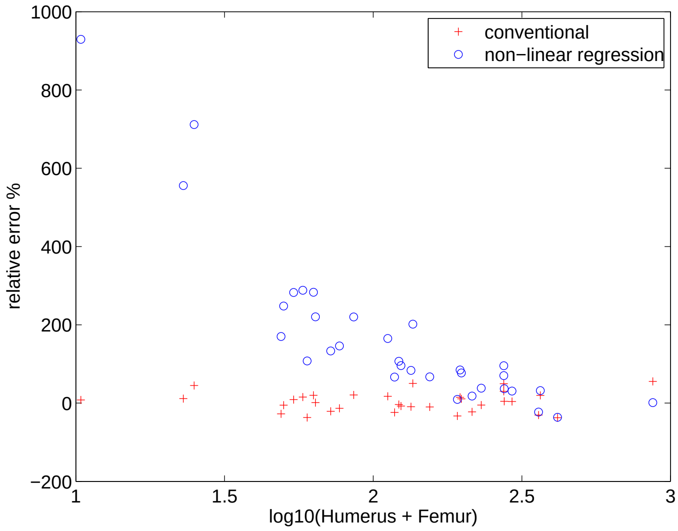


Figure 2: Relative errors for the conventional and non-linear regression based models for 33 mammalian species.

which clearly represents an unsatisfactory model fit. The smallest relative error for this model ocurrs for the elephant (≈ 1%), providing further evidence that the non-linear regression approach weights the error on large mammals much too highly. If relative, rather than absolute error is considered important, the conventional model clearly out-performs the non-linear regression model.

2.1 Presentation of Model Predictions

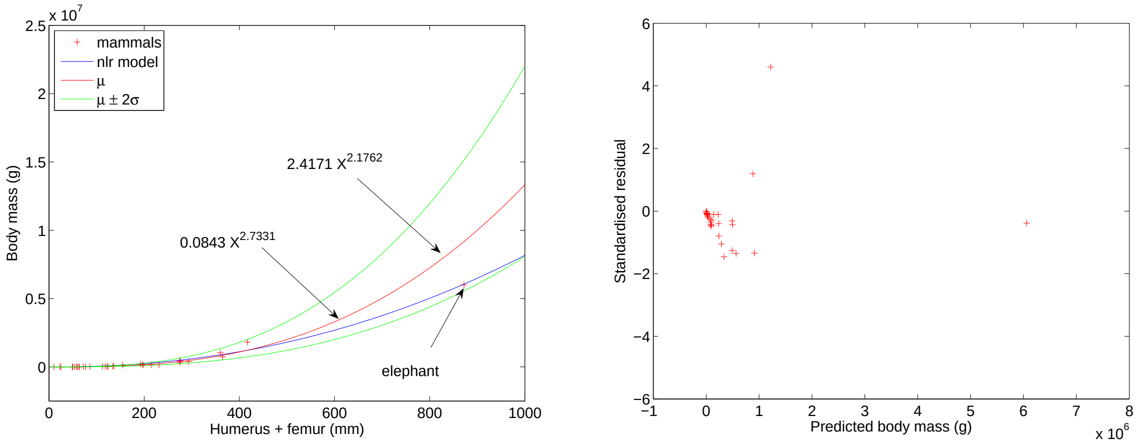
Packard et al. (2009) show plots of predicted body mass as a function of long-bone circumference, however these do not give a reliable indication of goodness of fit. As the responses differ by five orders of magnitudes the plots only provide an indication of the relative error for the larger mammals, the resolution being too small to reveal errors of tens of kilos, which would represent very substantial errors for most of the mammals described in the data. Had the numeric values of the predictions been tabulated, as in Table 3, it would be immediately apparent that the non-linear regression model provides a very poor subjective fit for most mammals.

2.2 Diagnostic Tests

Packard et al. (2009) performed tests for normality and homoscedasticity for both the conventional and non-linear regression approaches. However, while the conventional model passed on both counts (suggesting that the generative model was appropriate), they still decide in favour of the non-linear regression model that failed both tests (suggesting that the underlying statistical assumptions were invalid). This seems a somewhat strange practice; the conventional model passed the diagnostic tests because it embodies reasonable assumptions regarding the natural variability of body mass and is conceptually the superior model. As the diagnostic tests for the non-linear regression models revealed the statistical assumptions to be invalid, this also casts doubt on the existence of outliers, as the identification of outliers is dependent on the statistical assumptions regarding the underlying distribution from which the sample was drawn. If those assumptions were invalid, the identification of outliers will be unreliable.

2.3 Use of Standardised Residuals to Detect Heteroscedasticty

Packard et al. (2009) use a plot of the standardised residuals to detect signs of heteroscedasticity in the data, commenting that *“Whereas the display of residuals clearly points to a problem with the distribution of the data, it does not reveal the funnel-shaped pattern that would be expected of data exhibiting multiplicative error”*, (referring to Figure 1(b) in their paper). However, it is not clear that this approach is reliable for non-linear models; as a test, we generate a representative synthetic dataset from the conventional allometric model (1), with the optimal parameters for the dataset of Anderson et al. (1985) (the parameter settings are as follows: $a = 0.0843, b = 2.7331$ and $\sigma^2 = 0.0118$). A non-linear regression model was then fitted to the resulting data, as shown in Figure 3a. The standardised residuals are plotted in Figure 3b, note that again there is little sign of the funnel-shaped pattern, even though in this case we know *by construction* that the data have a heteroscedastic multiplicative error structure.



(a) (b)

Figure 3: Synthetic data generated from an allometric model with multiplicative errors, with non-linear regression model (a) and corresponding standardised residual plot (b) (c.f. Packard et al., 2009, Figure 1b).

Figure 1b of Packard et al. (2009) however does clearly reveal a significant bias in the non-linear model as the standardised residuals are negative for all but two of the observations (and hence it consistently over-estimates the body mass of the remaining mammals) and the main cluster of residuals has a relatively narrow downward sloping linear structure. The presence of visible structure in the residuals is an indication that the model is at best questionable. In our experiments, a residual structure of this nature is easily reproduced by selecting a sample of synthetic data with a large body mass for the hippopotamus and low body mass for the elephant, as shown in Figure 3. This suggests further evidence that the non-linear regression model over-fits the observations representing the very largest mammals, in the sense that the fit is closer than is warranted considering the likely extent of natural variability for mammals of that size.

2.4 Likelihood Ratio Test

The likelihood ratio test provides a simple means to evaluate the relative validity of the underlying statistical assumptions of competing models. In this case as the functional form of the models are identical, only differing in the assumptions regarding the multiplicative or additive Gaussian noise processes, the ratio of the maximum likelihood of the calibration set provides a meaningful criterion. In this case, the likelihood for the conventional model is approximately $4.75 \times 10^{+11}$, and for the non-linear regression model approximately 1.02×10^{-192}, and so the likelihood ratio test finds emphatically in favour of the conventional model. The reason for the failure of the non-linear regression model is clearly evident in Figure 1d, where the error bars are much broader than necessary to capture the variability of small mammals, and the model is penalised for giving unduly vague predictions of their body mass.

2.5 On Outliers

The definition of an outlier is somewhat fraught, often the most appropriate definition depends on the nature of the analysis, however a good working definition for the purposes of this study might be:

> *An outlier is an observation that cannot be adequately reconciled with a model that otherwise provides a good fit to the data.*

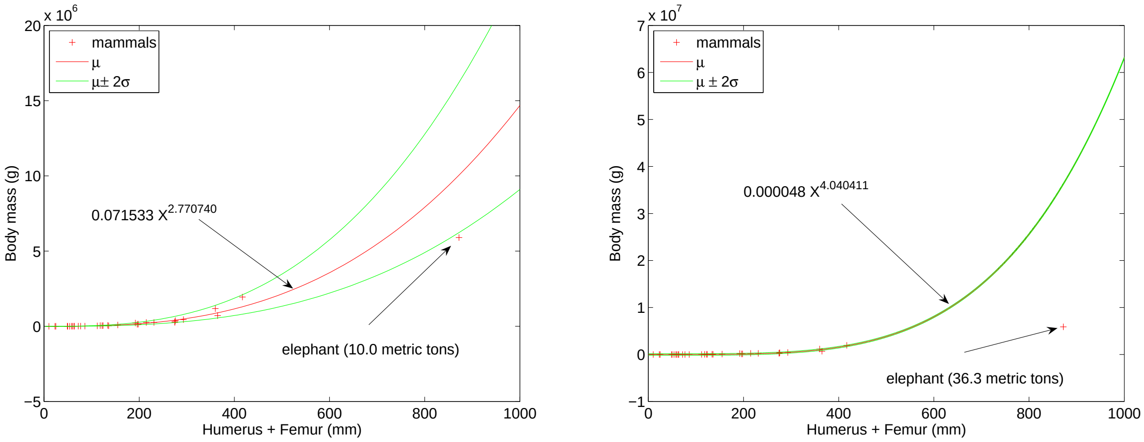
Note that whether an observation is an outlier can only be defined in terms of the model, as that defines the distribution from which the data are considered to have been drawn. If an observation lies a distance from the regression, but is within the error bars, then it is still adequately explained by the model as being within expected variation. Similarly, it may be the case that a model with additional information, such as the length of the long bones might be able to give a more accurate prediction of the weight of an elephant, as it would have a more realistic picture of the range of basic body plans seen in mammals, and so would no longer be an outlier. It would not be reasonable therefore to discard an observation purely because the model provides a bad fit for that particular observation. It may be that the underlying statistical assumptions are wrong, or simply that the model is not sufficiently complex to capture the structure of the data. In the particular cases of the hippopotamus and elephant, there may be evolutionary explanations for their departure from the norm, and should be retained in the modelling process as the dinosaurs involved may also exhibit a similar range of adaptions. In that case, deleting outliers would result in error bars giving an unduly confident prediction of the body mass of dinosaurs.

If the model were confined to predicting the body mass of large land-dwelling dinosaurs, there may be a case for deleting semi-aquatic mammals, such as the hippopotamus as being biologically unrepresentative of the dinosaurs considered, but there seems little *statistical* reason to discard any of them, based on the evidence presented by Packard et al. (2009).

The reason that the conventional model does not identify the elephant, hippopotamus or bison as outliers is not a deficiency of the model, quite the opposite in fact. The reason the conventional model does not flag any of these creatures as possible outliers is because under the assumption of approximately constant relative, rather than absolute natural varibility, they are unusual, but by no means exceptional as they lie close to, but not outside the error bars of the model.

2.6 Reliability of Predictions

The reliability of predictions for the unobservable body mass of large dinosaurs can be assessed by considering the effect of leaving the largest observed mammal, the elephant, out of the calibration set to see if the models can still give a credible predictions of its body mass. Figure 4 shows the conventional and non-linear regression model fits, excluding the elephant from the calibration set. For the conventional model, the parameter estimates are very similar, and as a result the predicted body mass for the elephant is also quite stable. Using the entire dataset, the conventional model predicts a body mass of 9.2 metric tons (error bars cover 5.6 metric tons to 15.1 metric tons), without the elephant in the calibration set, the predicted body mass is 10.0 metric tons (with error bars from 6.2 to 16.2 metric tons), a change of only $800kg$ or $\approx 9\%$. Note that the true body mass of the elephant is only slightly below the -2 s.d. error bar. For the non-linear regression model on the other hand, the parameter estimates change markedly, largely due to the influence of the hippopotamus, and so the predicted body mass for the elephant also changes very substantially, from 6.0 metric tons ($\pm 318kg$) to 36.3 metric tons ($\pm 171kg$), a difference of 30.3 metric tons or $\approx 500\%$. The observed body mass of the elephant then lies approximately 700 standard deviations from the predicted value! This suggests that the conventional model provides much more reliable predictions for the body mass of large dinosaurs as the model is far less sensitive to the natural variability observed in large mammals forming the calibration data.



(a) (b)

Figure 4: Comparison of conventional back transformation (a) and non-linear regression (b) based allometric models trained on all mammals with the exception of the elephant.

The high sensitivity of the non-linear regression model to the body mass of the elephant suggests that the predicted weight of the large dinosaurs will be similarly unreliable. The conventional model, with the full calibration set gives a body mass of 38.5 metric tons for *Apatosaurus louisiae* (with error bars from 23.3 to 63.5 metric tons), if the elephant is excluded, the predicted body mass rises to 43.0 tons (error bars: 26.7 to 69.4 tons), a difference of only $\approx 12\%$. Using the non-linear model, the predicted body mass rises from 18.2 metric tons ($\pm 318kg$) to 302 metric tons ($\pm 171kg$), a difference of $\approx 1500\%$. This extreme sensitivity to the presence or absence of a particular observation in the calibration set suggests that the non-linear model is unable to provide reliable predictions of the body mass of dinosaurs, unlike the conventional model which appears to be reasonably stable.

The elephant exhibits a lower body mass than might be expected for an animal with a long bone circumference of such magnitude, and hence is likely to result in under-prediction of the body mass of large dinosaurs using a model that is highly sensitive to its body mass. As Packard (2009) suggests that the elephant is possibly an outlier, it seems unreasonable to assert a lower body mass for large dinosaurs that is largely predicated on a potential outlier.

3 Summary

Packard *el al.* propose a non-linear regression approach to allometric estimation of the body mass of dinosaurs, however this approach has many disadvantages in this particular application not shared by the conventional approach:

- The model assumes constant absolute natural variability regardless of long bone circumference, which is biologically implausible.
- The non-linear regression fails statistical tests of two of its underpinning assumptions, namely normality and homoscedasticity of the residuals.
- The non-linear regression model predicts the possibility of natural variation producing mammals with negative body mass.
- The error bars are far too broad for small mammals and too narrow for large mammals and dinosaurs for biological plausibility.
- The model exhibits a consistent bias, over-estimating the body mass of small and medium sized mammals.
- The model is extremely sensitive to the natural variability observed in the body mass of large mammals.
- The model is unable to provide a credible prediction of the body mass of an elephant unless it is included in the calibration data.
- The model has a very low likelihood, suggesting that the data could not be plausibly regarded as being an i.i.d. sample from the implied distribution.
- The predictions for the large dinosaurs are essentially predicated on the observed body mass of the elephant, an observation close to being regarded as an outlier by the model.

The conventional model appears to have a single disadvantage, namely that it has a higher least-squares error on the untransformed data, however as the data are clearly heteroscedastic, an unweighted least-squares error is a poor criterion on which to judge goodness of fit. The non-linear regression model is found to be strongly biased due to the inappropriate statistical assumption of uniform uncertainty in the absolute body mass, rather than uniform uncertainty in relative body mass[1]. As a result, while the proposed approach gives rise to lower estimates of the body mass of dinosaurs, it systematically over estimates the observed body mass of current mammalian species, in one case by more than an order of magnitude. It seems likely then that the original back-transformation approach provides a more reliable predictor, and dinosaurs are likely to have been as large as previously thought. Note however the error bars of that model indicate that the body mass of large dinosaurs remains highly uncertain.

References

Anderson, J. F., Hall-Martin, A., & Russell, D. A. (1985). Long-bone circumference and weight in mammals, birds and dinosaurs. *Journal of Zoology*, *207*(1), 53–61.

Kerkhoff, A. J., & Enquist, B. J. (2009). Multiplicative by nature: Why logarithmic transformation is neccessary in allometry. *Journal of Theoretical Biology*, *257*(3), 519–521.

Packard, G. C. (2009). On the use of logarithmic transformations in allometric analysis. *Journal of Theoretical Biology*, *257*(3), 515–518.

Packard, G. C., Boardman, T. J., & Birchard, G. F. (2009). Allometric equations for predicting body mass of dinosaurs. *Journal of Zoology*.

---

[1] A more general justification for the use of logarithmic transforms in allometry is provided by Kerkhoff and Enquist (2009), in response to other work by Packard (2009).

Table 1: Long bone circumference measurements and true and predicted body mass for 33 mammalian and six dinosaur species, using non-linear regression and traditional back-substitution based approaches to allometric modelling.

	Humerus + Femur (mm)	**Body mass (kg)**				
Species		**Observed**	**non-linear regression**	**back transformation**		
			$\pm 317.7kg$	**lower**	**mode**	**upper**
meadow mouse	10.4	0.047	0.48	0.03	0.05	0.08
guinea pig	25	0.385	3.12	0.34	0.56	0.92
gray squirrel	23	0.399	2.62	0.27	0.44	0.73
opossum	50	3.92	13.6	2.2	3.7	6.1
gray fox	54	4.20	16.1	2.8	4.6	7.6
raccoon	58	4.82	18.7	3.4	5.6	9.2
nutria	49	4.84	13.1	2.1	3.5	5.8
bobcat	63	5.82	22.3	4.2	7.0	11.5
porcupine	64	7.20	23.1	4.4	7.3	12.0
otter	60	9.68	20.1	3.7	6.1	10.1
coyote	72	12.7	29.6	6.1	10.0	16.6
cloud leopard	86	13.5	43.2	9.9	16.3	26.9
duiker	77	13.9	34.2	7.3	12.1	19.9
yellow baboon	112	28.6	75.8	20.4	33.6	55.4
cheetah	136	38.0	114.6	34.6	57.1	94.3
cougar	122	44.0	90.9	25.7	42.4	70.0
wolf	124	48.1	94.1	26.9	44.4	73.2
bushbuck	118	50.9	84.7	23.5	38.8	63.9
impala	134	60.5	111.0	33.2	54.9	90.5
warthog	155	90.5	151.3	49.5	81.7	134.8
nyala	196	135.0	249.2	94.0	155.1	255.9
lion	198	144.0	254.7	96.6	159.5	263.1
black bear	192	218.0	238.6	88.8	146.6	241.9
grizzly bear	231	256.0	353.5	147.3	243.0	401.0
blue wildebeest	215	257.0	303.4	121.0	199.7	329.6
Cape Mountain zebra	275	262.0	512.2	237.2	391.4	645.8
kudu	275	301.0	512.2	237.2	391.4	645.8
Burchells zebra	276	378.0	516.1	239.5	395.3	652.3
polar bear	293	448.0	586.1	282.0	465.4	768.0
giraffe	365	710.0	935.2	514.2	848.5	1400.1
bison	360	1179.0	908.1	495.1	817.1	1348.3
hippopotamus	417	1950.0	1241.4	739.9	1221.0	2014.9
elephant (Jumbo)	872	5897.0	5959.7	5556.6	9169.4	15131.4
Styracosaurus albertensis	658	—	3275	2574	4247	7009
Diplodicus sp.	725	—	4025	3355	5536	9136
Opisthocoelicaudia skarzynskii	1245	—	12710	14706	24267	40046
Apatosaurus alenquerensis	1332	—	14672	17687	29187	48165
Brachiosaurus brancai	1384	—	15917	19639	32408	53479
A. lousiae	1474	—	18199	23329	38497	63528