Line-Fitting by rotation: A nonparametric method for bivariate allometric analysis

Isler, K ; Barbour, A D ; Martin, R D

DOI: https://doi.org/10.1002/1521-4036(200204)44:3<289::AID-BIMJ289>3.0.CO;2-C

Line-Fitting by Rotation: A Nonparametric Method for Bivariate Allometric Analysis

KARIN ISLER†, A. D. BARBOUR‡ AND ROBERT D. MARTIN†

†Anthropological Institute and Museum and ‡Mathematical Institute

University of Zürich-Irchel

Switzerland

Summary

The choice of an appropriate method for fitting a straight line to data is one of the major procedural problems in bivariate allometric analysis. Commonly used techniques, such as ordinary least-squares regression and major axis, are derived from the general structural relationship model and thus require some knowledge about the distribution of the data. In this paper, we explore a nonparametric alternative, referred to as "rotation method", which involves no assumptions about error distributions. It is symmetrical in the two variables and highly resistant against the influence of outlying data points. In this initial study, eight alternative versions of line-fitting by rotation are compared, using simulated data. The versions with the best overall performance are applied to empirical data from selected examples in biological anthropology. A comparison with conventional parametric methods reveals some marked advantages of the rotation method for descriptive allometric studies, indicating that further investigation into procedures of this kind is clearly warranted.

Key words: Allometry; major axis regression; nonparametric line-fitting; robust methods

Short running title: Line-fitting by rotation

1. Introduction

Allometric analysis is a standard tool for comparison of quantitative features between individuals or species of different body size. Many empirical relationships are well described by the standard allometric formula

$$U = aV^b$$

where U is an individual biological parameter and V a measure of body size, e.g. total body weight. Converted into logarithmic form, this power equation becomes linear:

$$\log(U) = b\log(V) + \log(a)$$

One of the fundamental problems of allometric analyses is the need to choose one of the various methods for estimating the exponent b of the power equation or, equivalently, fitting a straight line to the logarithmically transformed data. Three techniques are commonly used for estimating the parameters of the allometric equation: ordinary least-squares regression (OLS) (Model I regression), major axis (MA) and standardized or reduced major axis (RMA) (both Model II regression).

There has been extensive discussion about these different regression procedures (e.g. HARVEY, 1982; SEIM and SAETHER, 1983; RAYNER, 1985; MCARDLE, 1988; MARTIN and BARBOUR, 1989; JOLICOEUR, 1990; HARVEY and PAGEL, 1991; RISKA, 1991; AIELLO, 1992; SMITH, 1994 contra HARTWIG-SCHERER and MARTIN, 1992). Each of the three methods is derived from the general structural relations model (SPRENT, 1969; RAYNER, 1985) by making certain assumptions about the distribution of error in the data. The general structural relations model requires some knowledge about the distribution of the data, i.e. the true and error variances of both variables should be known, as well as the correlation of the errors (HARVEY and PAGEL, 1991). Especially in interspecific datasets, estimation of the error variances is problematic because the scatter in the data is a mixture between sampling error and biological variation, and the latter cannot be eliminated (RISKA, 1991).

The line estimated by any of the parametric methods can be greatly influenced by outlying data points, i.e. these approaches are poor in the statistical property of "resistance" (SMITH,

1994). Robust regression techniques, such as WALD's grouping method or BARTLETT's three-group method, are not symmetrical in the two variables and thus do not give one relationship, but rather two, depending on which variable is being grouped (MCARDLE, 1988).

Instead of accumulating more arguments in favour of one or another method, it would seem to be more profitable to seek a nonparametric method for estimating the best-fit line. Such a method should make few assumptions about the distribution of the data and be resistant against outliers. Like model II regressions, but unlike most robust regression methods (overview in EMERSON and HOAGLIN, 1983), it should be symmetrical in the two variables. Such a method was suggested by MARTIN and BARBOUR (1989). In this paper, it is referred to as the "rotation method".

We compared eight different versions of the rotation method applied to simulated datasets. The versions with the best overall performance were then applied to empirical data from biological anthropology. A comparison with conventional regression techniques revealed some marked advantages of the nonparametric rotation method.

2. The Method of Line-Fitting by Rotation

The underlying model is very general: It is assumed that the bivariate data arise as realizations of a random vector (X, Y), whose distribution, after a suitable rotation θ, can be expressed as a joint distribution F with independent marginals. In this case, $\tan(\theta)$ is interpreted as the slope of the approximate linear relationship between X and Y; if X and Y are themselves independent, this slope is zero. Of course, if F has independent marginals after rotation through θ, the same is true after rotation through $\theta + 90°$, $\theta + 180°$ and $\theta + 270°$; but in the allometric model it is clear which of the two values of the tangent corresponds to b. To find the slope of the line for a given dataset, the data points are rotated stepwise until the marginals are "as independent as possible". The tangent of the resulting angle of rotation is then interpreted as the slope of the "best-fit" line. For the allometric

equation, with $y = \log(U)$ and $x = \log(V)$, the true line has the form $y = bx + \log(a)$, where then $\tan(\theta) = b$.

Let F_n be the empirical distribution generated by the data points $(x_1, y_1), \ldots, (x_n, y_n)$ and $F_n(\theta)$ the empirical distribution after a rotation of the data points through an angle θ. The easiest way to determine a value of θ most consistent with an assumption of independent marginals is to choose a measure of dependence $D = D(F)$, whose minimum value 0 is attained when F has independent marginals, and to minimize the value of $D(F_n(\theta))$ with respect to θ. Then

$$\theta^* := \arg\min_{\theta} D\big(F_n(\theta)\big)$$

is the estimated angle of rotation, or equivalently, $b^* = \tan(\theta^*)$ is the estimated slope of the best-fit line.

There are many possible measures of dependence D that could be chosen. One classical possibility would be the absolute value of the product moment correlation coefficient. This yields the major axis (not the OLS regression) as best-fit line (MARTIN and BARBOUR, 1989). However, the product moment correlation coefficient is very sensitive to outliers in the data, and is therefore not to be recommended for practical use; it also has the disadvantage that it would not necessarily detect departures from the underlying model, since it only detects linear relationships. Instead, a measure of dependence D is to be preferred which is robust against outliers, and which only attains the minimal value zero at joint distributions F with independent marginals.

For a distribution F with independent marginals, the product law holds true; that is, if the pair of random variables (X, Y) has joint distribution F, then

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B]$$

for any intervals A and B. Hence we let

$$\delta(A, B) := \big|P[X \in A, Y \in B] - P[X \in A]P[Y \in B]\big|$$

be a measure of the dependence of the marginals with respect to a pair of intervals A and B, and then define

$$D := \sum_{i,j} \delta_{ij} = \sum_{i,j} \delta(A_i, B_j)$$

for a suitable choice of intervals A_i, B_j, $1 \le i,j \le m = m(n)$. Alternatively, the maximum of the δ_{ij} could be used instead of the sum. With regard to resistance, percentiles are used as interval boundaries. Altogether, eight different empirical measures of dependence were defined, including four based on quantities

$$\hat{\delta}_{ij} := \left| E[h_i(X)k_j(Y)] - E[h_i(X)]E[k_j(Y)] \right|,$$

for logistic functions h_i and k_j, which are smooth approximations to the indicators of the sets A_i and B_j, respectively (Table 2.1). Both sum and maximum were explored: three choices of intervals with sum, only one with maximum; and either with indicators or with their smooth approximations. Each of the measures involving the sum can be thought of as a discretized approximation to the measure of dependence

$$D(F) := \int \left| f(x,y) - f^1(x) f^2(y) \right| dx\, dy,$$

where F has density f and f^1 and f^2 denote the x- and y-marginal densities; those involving the smooth functions h and k vary more smoothly as the data are rotated, but take longer to compute. The measure D has the required properties, and is robust against outliers: some asymptotic properties are discussed in the Appendix. Since the maximum performed worse in our simulations, we have not examined its theoretical properties any further.

Minimization of $D(F_n(\theta))$ with regard to θ yields the slope of the line. Unlike the classical parametric methods, the line of the rotation method does not pass through the means, but rather through the medians of the two variables. Since the median is not invariant under rotation, the line in fact passes through the back-rotated median of the rotated data points.

Confidence intervals can be obtained by the following procedure: From the rotated data, we construct K new datasets by permuting the y-values randomly. Then the rotation method

is applied to each of these datasets, resulting in K values $\pi^{(1)}, \ldots, \pi^{(K)}$, estimating a "true" slope of 0. If the underlying model is true, randomly associating the x- and y-values of the rotated data gives data with distribution close to that of a random sample from the true product distribution of the correctly rotated data. The values $\pi^{(1)}, \ldots, \pi^{(K)}$ are then representative of the typical departures of the sample estimates of slope from the true value, so that θ^* can be viewed as a realization of $\theta + \Pi$, where θ is the true value and Π is a random error having a distribution which is approximated by the empirical distribution of $\pi^{(1)}, \ldots, \pi^{(K)}$. Thus, if $p_K(\alpha/2)$ and $p_K(100 - \alpha/2)$ are the $\alpha/2\%$ and $100 - \alpha/2\%$ quantiles of the empirical distribution of $\pi^{(1)}, \ldots, \pi^{(K)}$, then θ lies in the interval $(\theta^* - p_K(100 - \alpha/2), \theta^* + p_K(\alpha/2))$ with approximate confidence $100 - \alpha\%$. The corresponding minimizing values $D(F_n^{(k)}(\pi^{(k)}))$ can also be compared with $D(F_n(\theta^*))$; if the latter is much larger, it suggests that the data may not fit the underlying straight line model very well.

3. Application to Simulated Data: Comparison of Eight Versions of Line-Fitting by Rotation

3.1 Methods

For the comparison of the eight versions of the method as defined above (Table 2.1), 272 simulated datasets ($N = 50$, 100, 200 and 400) were generated using the uniform pseudorandom numbers produced by Mathematica™ 2.2.2 (WOLFRAM, 1993). The data came from either a bivariate normal or an "extruded normal" distribution (MARTIN and BARBOUR, 1989) with a theoretical slope of 0.842 (angle of rotation 0.7 rad) and differing degrees of scatter (r^2 between 0.505 and 0.996). The programs were implemented in Mathematica™ 2.2.2. Subsequent statistical analyses were performed in Statview™ 4.02.

The minimization process was executed with three different sets of starting points near the theoretical value of 0.7 rad: {0.5, 0.9}, {0.6, 0.8} and {α, α +0.1}, tan(α) being the slope of the major axis for the respective dataset. Since the derivative of the measure of

dependence does not necessarily exist, the secant method has to be used for minimization, which requires two starting points (SCHWARZ, 1993).

The eight versions were compared with regard to the quality of the minimization process (number of global minima found) and the number of unique global minima found in a total of 816 minimizations. If at least one of the three minimizations per dataset yielded a unique global minimum, the result was labeled "useful". The bias of the results was compared using *t* tests. Other points of interest were time consumption and resistance to outliers. As a measure of resistance, the breakdown point is the percentage of the data points that can become arbitrarily wild without destructively affecting the line (HAMPEL, 1971). Theoretical considerations of the breakdown point under ideal conditions are elaborated in the Appendix. Resistance to outliers was investigated by applying the eight versions of the rotation method, OLS and MA regression to 272 simulated datasets ($N = 50$, 100, 200 and 400) derived from bivariate normal and extruded normal distributions, that were contaminated with an increasing percentage of wild data points (0, 1, 10 and 20%) from a circular normal distribution centered on a very distant point.

3.2 Results

Results of the comparison of the eight versions are summarized in Table 3.1. The b-versions yield more unique global minima and more useful results than the a-versions. Between the versions 1b, 2b and 3b there are no significant differences regarding global minima found, unique global minima found or useful results (χ^2 test, DF = 4, $p = 0.98$). For testing the bias of the results, only the useful results were included. The hypothesis of no bias cannot be rejected for any of the versions of the rotation method (*t* test, *p*-values between 0.06 and 0.58). Computation time increases with the number of quantile intervals used. The b-versions are about five times slower than the a-versions. Resistance to outliers, tested in simulations of contaminated data ($N = 50$, 100, 200 and 400), was very high in all the versions of the rotation method (Fig. 1). In versions 1a, 1b, 2a, 2b, 3a and 3b, the breakdown point is higher than 20%. Only versions 4a and 4b yielded unpredictable results for datasets containing more than 10% outliers.

3.3 Discussion

Generally, the quality of the minimization process increases with the sample size. The versions 4a and 4b, which use the maximum of the δ_{ij}, are less robust and yield the least useful results, as the function $D(F_n(\theta))$ often does not have a unique global minimum. Of the other versions, the b-versions do better than the a-versions regarding the quality of the minimization process. This general advantage is due to the fact that the indicator function used in the a-versions, as compared to the logistic function used in the b-versions, causes considerably more local minima.

Versions 1b, 2b and 3b show an equally high resistance to outliers. They diverge only in the number of percentiles used, and there is no significant difference between the percentages of unique global minima found or the percentages of useful results obtained for a given sample size. On the basis of theoretical considerations concerning the consistency of the estimates (see Appendix), use of version 1b can be recommended for a sample size of $32 \leq N < 243$, version 2b for $243 \leq N < 1024$ and version 3b for $1024 \leq N < 3125$.

Compared with the parametric methods OLS and MA regression, the robustness of the rotation method is much higher. However, with an increasing percentage of outliers, the slightly decreasing mean value of the estimated slope shows that the outliers are not completely neglected, but have an appropriate influence on the estimated line (Fig. 1).

4. Application of the Rotation Method to Empirical Data

As the canonical correlation coefficient r^2 is not invariant under rotation, we use another measure for variation, denoted here r^2_{max}. r^2_{max} is the maximal r^2 that can be achieved through a rotation of the data points. For a given dataset Z, the maximum of $r^2(Z_\theta)$ is obtained with the slope of the major axis of the θ-rotated dataset Z_θ being 1 (see Appendix). Therefore we define

$$r^2{}_{max}(Z) := r^2{}_{max}\left(Z_{\theta_1}\right) \; with \; \theta_1 := \arctan\left(a_{MA}\right) - \frac{\pi}{4},$$

where a_{MA} is the slope of the major axis of Z.

For a large sample of highly correlated data, the function of dependence $D(F_n(\theta))$ is smooth with a marked and unique global minimum. In such a case, the choice of the method is not crucial and the fitted lines are quite similar (e.g brain weight vs. body weight in 477 mammalian species (BRAIN, Table 4.1, Figs. 2 and 3)). If the dataset contains only a few, but highly correlated data points, the global minimum is still clearly marked, although the function $D(F_n(\theta))$ is less smooth and may exhibit some local minima.

For data with some amount of scatter, the slopes of the least-squares regression and the major axis differ. Generally, the difference between the rotation method and the major axis is smaller than the difference between the rotation method and least-squares regression, as in the data of HARRIS and BENEDICT (1919) on basal metabolic rate vs. body weight in 136 human males (BMR, Table 4.1, Figs. 2 and 3).

If the slope of the rotation method is different from both least-squares regression and the major axis, this indicates either poorly correlated data or the influence of outliers on the parametric methods. An example of the latter is provided by data on the intermembral index vs. body weight in 148 primate species (IMI, Table 4.1, Figs. 2 and 3. Data from ROWE (1996)). The intermembral index is defined as the ratio of forelimb length to hindlimb length. Gibbons have unusually long arms for their body weight, due to a special locomotor adaptation for ricochetal brachiation. This group influences the slope of both OLS and MA, but not the slope of the rotation method, which is hence somewhat lower.

Poorly correlated data are reflected by a relatively flat function $D(F_n(\theta))$, as in the data on body weight vs. body height in human males (BW, Table 4.1, Figs. 2 and 3). Such a case must be analysed carefully. Even our rather weak formulation of "points distributed around a line", which supposes that, for some rotation θ of the coordinate axes, the x- and y-components of the data are independent, is still too strong for many data appearing in practice. A check is given by comparing the global minimum $D(\theta^*)$ obtained from our algorithm with the values of D obtained with by permuting the y-coordinates of the θ^*-rotated data. If $D(\theta^*)$ is not typical of the D-values from the permuted data, this indicates that the idealized model is not true; for instance, because the data are actually a superposition of two "linear" datasets with differing central lines or because the variances of the x-values

are not equal for differing y-values. This seems to be the case for the data on brain weight vs. body weight in mammals (BRAIN) and on the intermebralindex vs. body weight in primates (IMI, Fig. 3), the two datasets for which the hypothesis of independent marginals for a θ can be rejected at a significance level of 95% in a test as described above (Table 4.2). For the data on body height vs. body weight in human males (BW), the hypothesis of independent marginals for a θ cannot be rejected. Thus, the flatness of the function $D(F_n(\theta))$ is probably caused only by a large amount of scatter around the line. If the three outlying points with extremely low body height are removed, the slope of the major axis increases, becoming much closer to the slope of the rotation method, which remains unaffected.

95% confidence intervals for the slope of the line fitted by rotation were obtained with the method defined above using $K = 60$ permutations for each dataset (Table 4.3). These intervals are quite similar in size to the confidence intervals for the slope of the least-squares regression line.

5. Conclusions

The method presented here shows several advantages over common line-fitting techniques such as least-squares regression or major axis regression: (1) Information about the distribution of the data is generally difficult to obtain, as a test of normality for biological variables typically requires a sample size of 1000 or more (GINGERICH, 1995). As a nonparametric method, line-fitting by rotation requires neither an assumption about the distribution of the data nor a questionable estimation of the error variances. (2) As it is highly resistant against outliers and symmetrical in the two variables, the nonparametric rotation method combines the major advantages of robust regression techniques and the major axis. Outliers are mostly attributable not to measurement errors but to meaningful biological variation. Therefore, they should ideally not be excluded from an analysis, but neither should they be allowed to have an overdue influence on the line. This is guaranteed by the rotation method (see Fig. 1). (3) Logarithmic transformation can introduce bias into statistical analysis of biological measurements, because it skews normality (SMITH, 1993).

This is not the case with the rotation method, as a logarithmic transformation does not change the rank order of data.

The rotation method is more costly in terms of computation effort than parametric methods. The minimum found by the minimization algorithm must be checked visually. If it is not global or not unique, this indicates a doubtful linear trend or too much variation in the data. Like the major axis, the rotation method is not invariant under unequal changes of scale, although this could be achieved by a generalization to non-orthogonal co-ordinate axes (MARTIN and BARBOUR, 1989). The minimization of the measure of dependence $D(F_n(\theta))$ would then have to be executed in a three-dimensional surface, which can lead to difficulties in practice. However, the requirement of scale invariance is not usually a problem with logarithmically transposed data.

As this is a first exploration of the possibilities of line-fitting by rotation, the emphasis is laid on the practical applicability of the method. Further investigations into this kind of methods are clearly warranted.

The algorithm is implemented in a Mathematica™ 3.0 package that is available on the internet (http://www.anthro.unizh.ch/Main/Who/Karin/rotation.html).

Appendix

1. Consistency

Let $f_\theta(x,y)$ denote the true joint probability density function underlying the data, after rotation through an angle of θ. The idea behind our method is that

$$D(F_\theta) := \int \left| f_\theta(x,y) - f_\theta^{\,1}(x) f_\theta^{\,2}(y) \right| dx\, dy \geq 0,$$

where $f_\theta^{\,1}(x)$ and $f_\theta^{\,2}(y)$ denote the x- and y-marginals of $f_\theta(x,y)$, is zero only if the θ-rotated joint distribution has independent x- and y-components. Indeed, it follows from a theorem of DARMOIS (1951) that, if $D(F_\theta) = 0$ for two values of θ which do not differ by an integer multiple of $\pi/2$, then F has independent normally distributed components with the same

variance, in which context the concept of "slope" has no meaning. In any other case in which, after some rotation θ_0, the x- and y-components are independent, the equation $D(F_\theta) = 0$ has θ_0 as a unique solution (modulo $\pi/2$). If $D(F_\theta)$ is a continuous function of θ, minimizing any uniformly consistent estimators $D_n(\theta)$ of $D(F_\theta)$ leads to a consistent estimator $\hat{\theta}_n$ of θ_0 (up to multiples of $\pi/2$) satisfying $D_n(\hat{\theta}_n) \to 0$. If there is no rotation θ which makes the x- and y-components independent, then $\min_\theta D(F_\theta) > 0$, and $D_n(\hat{\theta}_n)$ does not converge to zero. Thus, if we can construct a uniformly consistent sequence of estimators $D_n(\theta)$ of $D(F_\theta)$, we can estimate the slope by minimizing $D_n(\theta)$ with respect to θ, and also obtain some indication of the support for the hypothesis of independent components for the rotated data.

Our estimate $D_n(\theta)$ of $D(F_\theta)$ is simple. We partition the x- and y-axes into intervals $A_1, \ldots, A_{m(n)}$ and $B_1, \ldots, B_{m(n)}$ respectively, chosen so that the empirical marginal probabilities $F_n(\theta)\{A_i \times R\}$ and $F_n(\theta)\{R \times B_j\}$ of the rotated data are close to $1/m(n)$. We then compute the sum

$$D_n(\theta) := \sum_{i,j=1}^{m(n)} \left| F_n(\theta)\{A_i \times B_j\} - F_n(\theta)\{A_i \times R\} F_n(\theta)\{R \times B_j\} \right|$$

$$= \sum_{i,j=1}^{m(n)} \left| n^{-1} \sum_{l=1}^{n} ind[x_l, A_i]\, ind[y_l, B_j] - n^{-1} \sum_{l=1}^{n} ind[x_l, A_i]\, n^{-1} \sum_{l=1}^{n} ind[y_l, B_j] \right|,$$

where, for an interval I,

$$ind[x, I] := 1 \quad if \quad x \in I; \qquad ind[x, I] := 0 \quad if \quad x \notin I.$$

This is a uniformly consistent estimator of $D(F_\theta)$ if $m^2(n) n^{-1/2} \sqrt{\log(n)} \to 0$ (see, for example, Pollard (1984, Theorem II.37)). A choice of $m(n)$ of order $n^{1/5}$ would therefore be reasonable; in practice, for the sample sizes that we consider in our examples, we use the values $m = 3$, $m = 4$ and $m = 5$. We also obtain a smoother objective function $D'_n(\theta)$ by replacing $ind[x, I]$ in the definition of $D_n(\theta)$ by a smooth approximation,

$$lf[x, I] := \frac{1}{1 + e^{20(b-a)(a-x)}} \left(1 - \frac{1}{1 + e^{20(b-a)(b-x)}} \right).$$

2. Theoretical Breakdown Point

For this calculation, we assume that our data give us accurate estimates of $D(F_\theta)$ and of $\theta^* = \arg\min_\theta D(F_\theta)$. We now suppose that, while attempting to sample from a true distribution F, a proportion p of our sample comes from an arbitrary distribution G, so that we are really sampling from the mixture $(1\text{-}p)F + pG$. Then

$$\begin{aligned}
&\left|D(F) - D\big((1-p)F + pG\big)\right| \\
&\quad \le \int \left| \begin{array}{l} pf(x,y) - pg(x,y) - f^1(x)f^2(y) + \\ \big((1-p)f^1(x) + pg^1(x)\big)\big((1-p)f^2(y) + pg^2(y)\big) \end{array} \right| dx\,dy \\
&\quad \le \int \left(\begin{array}{l} pf(x,y) + pg(x,y) + p(2-p)f^1(x)f^2(y) + \\ p(1-p)\big(f^1(x)g^2(y) + g^1(x)f^2(y)\big) + p^2 g^1(x)g^2(y) \end{array} \right) dx\,dy \\
&\quad = 2p(3-p) \\
&\quad \le 6p
\end{aligned}$$

Hence, with p-contaminated data, the contaminated estimate of $D(F_\theta)$ is for each θ within $\pm 6p$ of the true value. If, therefore, for some θ_0, $f_{\theta_0}(x,y) = f_{\theta_0}(x) f_{\theta_0}(y)$ for all x,y, so that $D(F_{\theta_0}) = 0$, a p-contaminated sample gives an estimate θ^* such that $D(F_{\theta^*}) \le 12p$, restricting the possible choices to the set

$$\big\{\theta : D(F_\theta) \le 12p\big\}.$$

Hence the breakdown point at a true underlying distribution F is at least

$$p^* = \tfrac{1}{12} \max_\theta D(F_\theta) > 0.$$

This is a considerable advantage over a measure of dependence derived from a chi-squared statistic, which would have zero breakdown point.

3. Measure of Variation

For a given dataset Z, the maximum of $r^2(Z_\theta)$ is obtained with the slope of the major axis of the θ-rotated dataset Z_θ being 1. The proof of this is straightforward: Let $s_x^2 := \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$, $s_y^2 := \frac{1}{n-1}\sum_i (y_i - \bar{y})^2$ and $s_{xy} := \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y})$. We show that $\frac{d}{d\theta} r^2(Z_\theta) = 0$ for $\theta = \theta_0 := \frac{1}{2}\arctan\left(\frac{2s_{xy}}{s_y^2 - s_x^2}\right)$, which is the angle of the slope of the major axis of Z and for $\theta = \theta_1 := \frac{1}{2}\arctan\left(-\frac{s_y^2 - s_x^2}{2s_{xy}}\right) = \theta_0 - \frac{\pi}{4}$. As $\frac{d^2}{d\theta^2} r^2(Z_{\theta_1}) < 0$ and $\frac{d^2}{d\theta^2} r^2(Z_{\theta_0}) > 0$, the minimum of $r^2(Z_\theta)$ is obtained with $\theta = \theta_0$ and the maximum with $\theta = \theta_1 = \theta_0 - \frac{\pi}{4}$, so that the angle of the slope of the major axis at the maximum is $\theta_0 - \left(\theta_0 - \frac{\pi}{4}\right) = \frac{\pi}{4}$, completing the proof.

Acknowledgements

We thank a referee for the many helpful comments on earlier versions of this manuscript.

References

AIELLO, L. C., 1992: Allometry and the analysis of size and shape in human evolution. *Journal of Human Evolution* **22**, 127-147.

DARMOIS, G., 1951: Sur une propriété caractéristique de la loi de probabilité de Laplace. *Comptes Rendues de l'Académie de Sciences, Paris* **232**, 1999-2000.

EMERSON, J. D. and HOAGLIN, D. C., 1983: Resistant lines for y versus x. In: D. C. HOAGLIN, F. MOSTELLER and J. W. TUKEY (eds): *Understanding Robust and Explanatory Data Analysis.* John Wiley and Sons, New York, 129-165.

GINGERICH, P. D., 1995: Statistical power of EDF tests of normality and the sample size required to distinguish geometric-normal (lognormal) from arithmetic-normal distributions of low variability. *Journal of Theoretical Biology* **173**, 125-136.

HAMPEL, F. R., 1971: A general qualitative definition of robustness. *Annals of Mathematical Statistics* **42**, 1887-1896.

HARRIS, J. A. and BENEDICT, F. G., 1919: A biometric study of basal metabolism in man. *Carnegie Institute of Washington Publications* **279**, 1-266.

HARTWIG-SCHERER, S. and MARTIN, R. D., 1992: Allometry and prediction in hominoids: a solution to the problem of intervening variables. *American Journal of Physical Anthropology* **88**, 37-57.

HARVEY, P. H., 1982: On rethinking allometry. *Journal of Theoretical Biology* **95**, 37-41.

HARVEY, P. H. and PAGEL, M. D., 1991: *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.

JOLICOEUR, P., 1990: Bivariate allometry interval estimation of the slopes of the ordinary and standardized normal major axes and structural relationship. *Journal of Theoretical Biology* **144**, 275-285.

MARTIN, R. D. and BARBOUR, A. D., 1989: Aspects of line-fitting in bivariate allometric analyses. *Folia Primatologica* **53**, 65-81.

MCARDLE, B. H., 1988: The structural relationship: regression in biology. *Canadian Journal of Zoology* **66**, 2329-2339.

POLLARD, D., 1984: *Convergence of stochastic processes.* Springer, New York.

RAYNER, J. M. V., 1985: Linear relations in biomechanics: the statistics of scaling functions. *Journal of Zoology London (A)* **206**, 415-439.

RISKA, B., 1991: Regression models in evolutionary allometry. *American Naturalist* **138**, 283-299.

ROWE, N., 1996: *The Pictorial Guide to the Living Primates*. Pogonias Press, New York.

SCHWARZ, H. R., 1993: *Numerische Mathematik*. Teubner, Stuttgart.

SEIM, E. and SAETHER, B.-E., 1983: On rethinking allometry: which regression model to use? *Journal of Theoretical Biology* **104**, 161-168.

SMITH, R. J., 1993: Logarithmic transformation bias in allometry. *American Journal of Physical Anthropology* **90**, 215-228.

SMITH, R. J., 1994: Regression models for prediction equations. *Journal of Human Evolution* **26**, 239-244.

SPRENT, P., 1969: *Models in Regression and Related Topics*. Methuen, London.

WOLFRAM, S., 1993: *Mathematica: A System for Doing Mathematics by Computer.* Addison-Wesley, New York.

Full addresses:

Karin Isler

Anthropologisches Institut und Museum

Universität Zürich-Irchel

Winterthurerstr. 190

CH-8057 Zürich

Switzerland.

Robert D. Martin

Anthropologisches Institut und Museum

Universität Zürich-Irchel

Winterthurerstr. 190

CH-8057 Zürich

Switzerland.

A. D. Barbour

Mathematisches Institut

Universität Zürich-Irchel

Winterthurerstr. 190

CH-8057 Zürich

Switzerland.

Figure legends

Fig. 1.

Boxplot showing the resistance to outliers of the rotation method, compared to least-squares regression and major axis. Versions of the rotation method as defined in Table 2.1. Each method was applied to 272 datasets ($N = 50$, 100, 200 and 400) from bivariate normal and extruded normal distributions with a theoretic slope of 0.842 containing varying percentages of outliers (0, 1, 10 and 20%).

Fig. 2.

The function $D(F_n(\theta))$, a measure of dependence of the marginal distributions for four datasets from biological anthropology (see Table 4.1); IMI: intermembral index vs. body weight in primates ($N = 148$), BMR: basal metabolic rate vs. body weight in human males ($N = 136$), BRAIN: brain weight vs. body weight in mammals ($N = 477$), BW: body weight vs. body height in human males ($N = 79$).

Fig. 3.

The rotation method (solid line), ordinary least-squares regression (dotted line) and the major axis (dashed line) for the same datasets as in Fig. 2.

Tables

Table 2.1

The eight versions of the rotation method

version	description
1a	33%-percentiles, sum of δ_{ij}, indicator function
1b	33%-percentiles, sum of δ_{ij}, logistic function
2a	25%-percentiles, sum of δ_{ij}, indicator function
2b	25%-percentiles, sum of δ_{ij}, logistic function
3a	20%-percentiles, sum of δ_{ij}, indicator function
3b	20%-percentiles, sum of δ_{ij}, logistic function
4a	25%-percentiles, maximum of δ_{ij}, indicator function
4b	25%-percentiles, maximum of δ_{ij}, logistic function

Table 3.1

Results of the simulation experiments for the 8 versions of the rotation method, applied to 272 simulated datasets from bivariate normal and extruded normal distributions; the total number of minimizations for each version is 816; computation time is normalized to version 1a.

	N	1a	1b	2a	2b	3a	3b	4a	4b
% global minima	50	49.0	52.5	35.3	41.2	25.0	42.6	59.8	32.4
found	100	46.1	59.3	35.8	53.9	37.3	61.3	74.0	43.6
	200	50.0	69.6	44.6	65.2	45.6	73.0	67.6	42.2
	400	52.0	78.4	58.3	76.0	46.6	71.6	64.7	48.0
% unique global	50	19.6	52.5	25.5	41.2	17.2	42.6	11.8	30.4
minima found	100	34.8	59.3	27.5	53.9	24.0	60.8	10.8	43.6
	200	44.1	69.6	37.7	65.2	38.7	72.5	23.0	42.2
	400	45.1	78.4	45.1	76.0	42.2	70.1	28.9	48.0
% useful results	50	36.8	91.2	57.4	79.4	44.1	75.0	25.0	73.5
	100	79.4	94.1	60.3	97.1	54.4	92.6	22.1	79.4
	200	88.2	100	76.5	100	75.0	94.1	39.7	83.8
	400	91.2	100	82.4	100	88.2	97.1	54.4	95.6
computation time		1.0	4.8	1.6	8.4	2.5	13.0	1.6	8.4

Table 4.1

Ordinary least-squares regression (OLS), major axis (MA) and the rotation method (ROT) for four data sets from biological anthropology. IMI: intermembral index vs. body weight in primates, BMR: basal metabolic rate vs. body weight in human males, BRAIN: brain weight vs. body weight in mammals, BW: body weight vs. body height in human males.

	N	r^2	r^2_{max}	OLS	MA	ROT	version
IMI	148	0.413	0.962	$0.083x + 3.76$	$0.084x + 3.75$	$0.071x + 3.84$	1b
BMR	136	0.640	0.660	$0.63x + 4.77$	$0.74x + 4.30$	$0.75x + 4.28$	1b
BRAIN	477	0.969	0.971	$0.76x + 3.90$	$0.77x + 3.85$	$0.77x + 3.82$	2b
BW	79	0.598	0.901	$2.92x - 10.83$	$4.74x - 20.13$	$5.76x - 25.37$	1b

Table 4.2

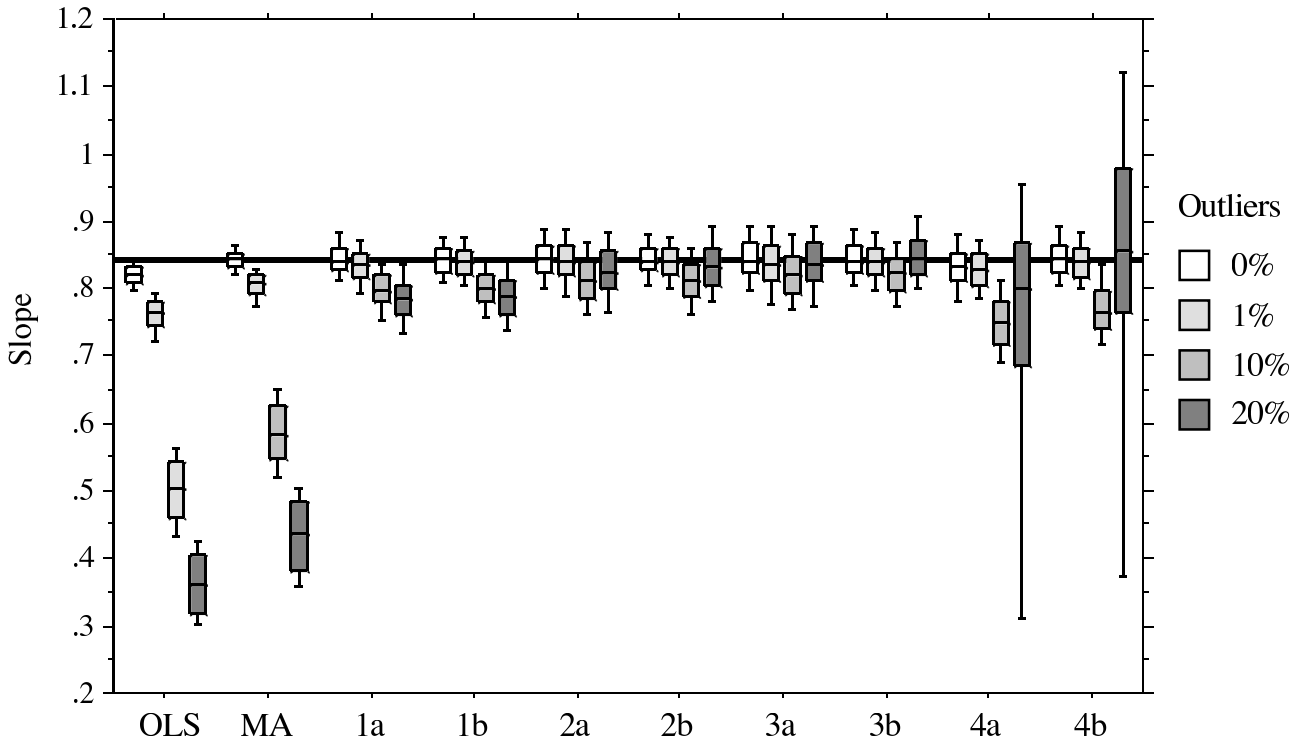
Tests for the hypothesis of independent marginals for a θ. The K=60 test sets were constructed as described in section 2. d_α denotes the α-quantiles of these test sets. The data are the same as in Table 4.1. The two datasets for which the hypothesis of independent marginals can be rejected on a significance level of 95% are intermembralindex vs. body weight in primates (IMI) and brain weight vs. body weight in mammals (BRAIN).

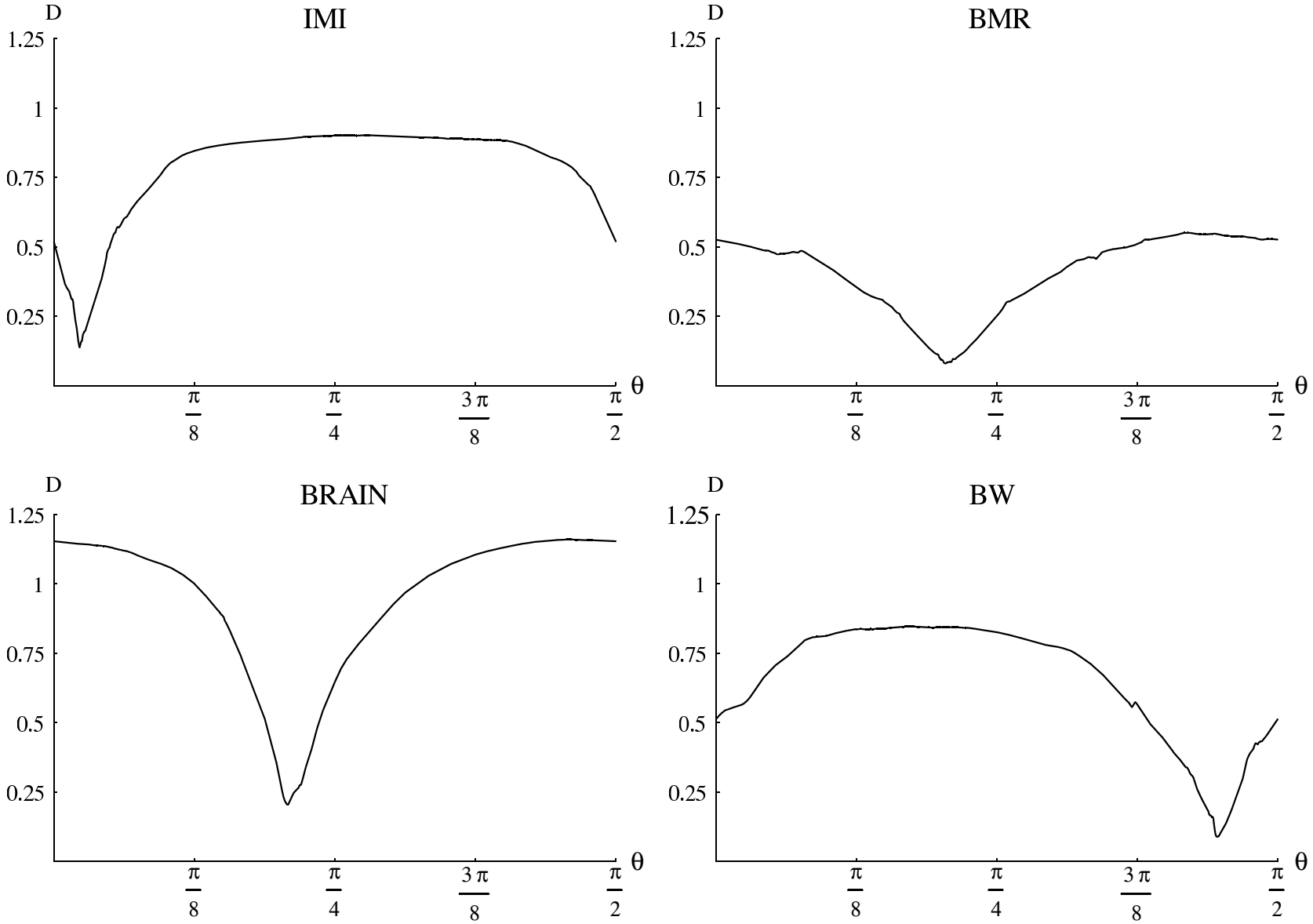
	N	$r^2_{\max}$	version	θ*	$D(\theta^*)$	$d_{95\%}$	$d_{90\%}$
IMI	148	0.962	1b	0.071	0.136	0.131	0.113
BMR	136	0.660	1b	0.644	0.076	0.157	0.131
BRAIN	477	0.971	2b	0.658	0.231	0.147	0.129
BW	79	0.901	1b	1.399	0.085	0.175	0.162

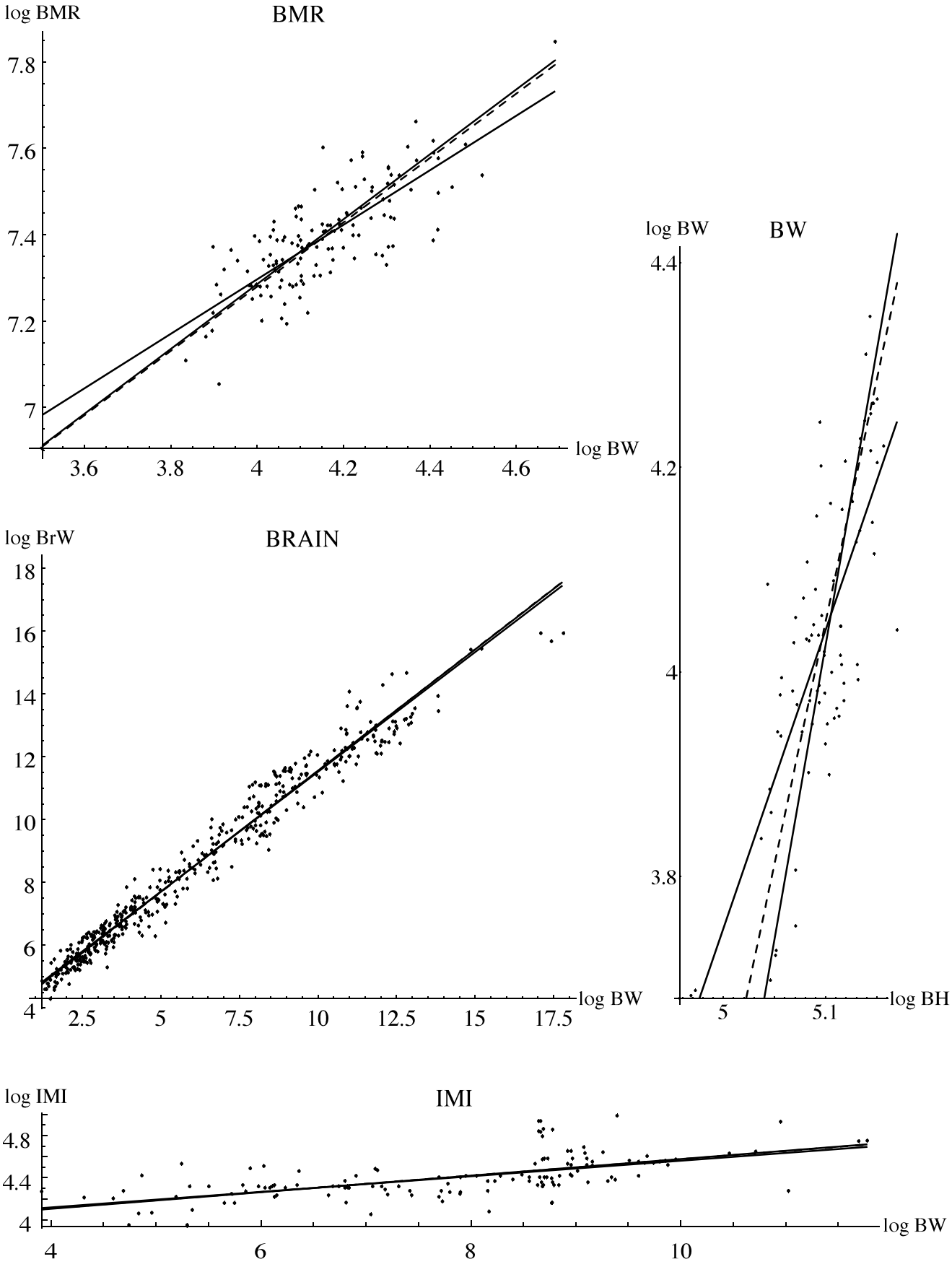
Table 4.3

95% confidence intervals (C.I.) for the slope of the line for the same datasets as in Table 4.1. The confidence bounds for the rotation method (ROT) were obtained with K=60 permutations.

	N	r^2_{max}	slope OLS	95% C.I. for OLS	slope ROT	95% C.I. for ROT	version
IMI	148	0.962	0.083	(0.067, 0.099)	0.071	(0.056, 0.085)	1b
BMR	136	0.660	0.631	(0.550, 0.712)	0.751	(0.615, 0.852)	1b
BRAIN	477	0.971	0.762	(0.749, 0.775)	0.773	(0.762, 0.789)	2b
BW	79	0.901	2.917	(2.375, 3.459)	5.763	(4.507, 7.462)	1b


1.2
1.1
1
.9
.8
.7
.6
.5
.4
.3
.2
Slope
OLS
MA
1a
1b
2a
2b
3a
3b
4a
4b
Outliers
0%
1%
10%
20%


IMI
BMR
BRAIN
BW
D
θ
1.25
1
0.75
0.5
0.25
π/8
π/4
3π/8
π/2


log BMR
BMR
7.8
7.6
7.4
7.2
7
3.6
3.8
4
4.2
4.4
4.6
log BW
log BrW
BRAIN
18
16
14
12
10
8
6
4
2.5
5
7.5
10
12.5
15
17.5
log BW
log BW
BW
4.4
4.2
4
3.8
5
5.1
log BH
log IMI
IMI
4.8
4.4
4
4
6
8
10
log BW