Supporting Information for "Ontogenetic Shifts of Heart Position in Snakes"

Harvey Lillywhite Steven Lillywhite

Contents

1 Introduction

We present the findings of a statistical analysis of the data described in the main paper. These data consist of heart-to-head and total body lengths for four species of snakes. The data and the biological theory suggested segmented regressions, and we find that a regression fitting a piecewise linear function with one vertex (a V shape) is preferred for several species of snake, while a simple line regression is chosen for the only climbing species.

We use non-parametric tests for model selection, as well as bootstrapped likelihood ratio tests. We employ bootstrap methods for parameter inference as the sampling distributions of segmented regression parameters can be highly skewed.

The following is an outline of what we cover. Section 2 describes the data. Section 3 discusses a small list of proposed model specifications that we investigate, which are all regression models. Section 4 discusses segmented regression and provides a brief, mostly self-contained mathematical explication of the estimation process. Section 5 considers model specification tests for the proposed models. Section 6 shows parameter estimates for the fits of the chosen models. Section 7 shows hypothesis testing results for some parameter restrictions of interest in the segmented regression models. Section 8 shows non-parametric kernel regression fits to the data. In the Appendices, we show plots of the functions involved in the segmented regression calculations of Section 6 as well as a summary of the software used in the study.

2 Data

We examine data for four species of snakes:

Name	Common Name	Sample Size	Notes
Agkistrodon conanti	Florida cottonmouth	192	terrestrial, no climbing
Acrochordus granulatus	Little file snake	62	marine
Hydrophis platurus	Yellow-bellied sea snake	190	marine
Pantherophis obsoleta	Yellow rat snake	43	terrestrial, climbs

It is important to note that the rat snake is the only species among these four which climbs.

The data consists of measurements (cm) of

- total body length
- head-to-heart length

We assume that length measurement error is negligible and we ignore it.

For *Agkistrodon conanti*, we also have trunk data consisting of analogous measurements, but with the head removed, and the tail removed at the vent.

3 Model Specification

Based on biological theory and in the interests of parsimony, we shall examine the following model specifications. All are regression models. We consider linear, quadratic, log-transformed, and segmented regressions. Segmented regression models are similar to ordinary least squares models, but with a piecewise-linear functional form. We describe them in more detail below.

It is natural and intuitive to think of the relative heart position as a variable, so we shall examine models where the dependent variable is the ratio of the head-heart length divided by the body length. The independent regression variables shall be functions of the body length.

For the remainder of this report, let us denote the head-heart length by the variable y and the total body length by the variable x. We shall consider the following models

(1) $y_i = a + bx_i + \varepsilon_i$

(2) $log(y_i) = a + b\ log(x_i) + \varepsilon_i$

(3) $y_i = a + bx_i + cx_i^2 + \varepsilon_i$

(4) $y_i = bx_i + cx_i^2 + \varepsilon_i$

(5) $\frac{y_i}{x_i} = b + cx_i + \varepsilon_i$

(6) segmented linear regression with one vertex, or breakpoint (described in Section 4)

In these equations, ϵ_i denotes residuals, and is not necessarily the same for different models.

Note 1
In practice, we shall scale $\frac{y}{x}$ *by 100 (ie: use percentage). But we do not denote this in the formulas below.*

The line model (1) and the log-transformed line model (2) are the simplest and perhaps most commonly used specifications in allometric studies. The quadratic models (3) and (4) are motivated by the desire to consider the ratio regression (5), as discussed in Section 3.1.

In accord with the physiological discussion in the main paper, we expect a relative migration of the heart during ontogeny that moves the heart relatively closer to the head in gravity-challenged species. However, a more central position tends to minimize work of the heart, and a reversal of heart migration from anterior to posterior direction might be expected in non-climbing snakes as they get larger. A segmented regression (6) is a very simple model to consider for such data.

3.1 Ratio Regression Models

There is no problem using ratios in regressions, as has been shown in very great detail in [5] as well as numerous other articles. Assuming that the data x_i are far from zero, we note that a model such as

$$\frac{y_i}{x_i} = b + cx_i + \eta_i \tag{3.1}$$

is the same as the model

$$y_i = bx_i + cx_i^2 + \eta_i x_i \tag{3.2}$$

where η is an error term. The question of which form to use is only relevant in the sense of which form will allow a better estimation. For example, one or the other forms could have homoscedastic errors. A related question is whether the omission of an intercept term in (3.2) is reasonable or not. If included, the corresponding equivalent models would be

$$y_i = a + bx_i + cx_i^2 + \eta_i x_i \tag{3.3}$$

$$\frac{y_i}{x_i} = \frac{a}{x_i} + b + cx_i + \eta_i \tag{3.4}$$

3.1.1 Our Preferred Specification

We discard the specification (3.4) (and hence (3.3)) for the following reasons.

- First, this model tends to the model (3.1) as $x \to \infty$. For our data, when we fit (3.4) (not shown), we find that most of the data are in the region where the function is highly linear (ie: $\frac{a}{x}$ is very small). The term $\frac{a}{x}$ only has an impact on the extreme short end of the data.
- Second, it makes sense for the intercept a in (3.3) to be zero since y goes to zero as x does.

3.1.2 Whether to Choose Ratio Variables

For all the snakes, the residuals from regressions of the form (3.2) have a clear pattern (not shown) of increasing variance with increasing x. Dividing by x largely normalizes these variances. Thus we choose the specification (3.1).

Note. For all the segmented regression models that we consider henceforth, we shall use $\frac{y}{x}$ as the dependent variable, and x as the independent variable.

Finally, the above considerations notwithstanding, we test all six model specifications listed above in Section 5 for completeness.

4 Segmented Regression

We describe briefly the segmented regression used in our study. Segmented regression refers to regression functions which are defined piecewise. Splines are an example. The least-squares regression of such functions becomes complicated when the break-points, or knots, of the functions are unknown parameters which must be estimated. Segmented regression problems have been studied since at least the 1960s. For an introduction to the subject, we mention [3], [9] and the references contained therein. We shall focus on our specific segmented regression problem and give details of a closed-form algorithm for the estimation, as in [9].

Let use begin by defining the function we wish to fit to the data. We call this function informally the *V-fit* function.

$$f(x) = \begin{cases} v + m_1(x-u) & x \le u \\ v + m_2(x-u) & x > u \end{cases} \tag{4.1}$$

Here we have parametrized two half-lines with slopes m_1, m_2 which meet at a common vertex (u, v). See Section 6 for some plots to fitted data.

Our statistical model shall be univariate and have the form

$$y = f(x) + \epsilon$$

where ϵ has mean zero and variance σ^2. We shall assume we have an i.i.d. sample $\{(y_i, x_i)\}_{i=1}^{N}$.

We'll estimate parameters by means of regression. To this end, we define the residual sum of squares as

$$G(u, v, m_1, m_2) \overset{\text{def}}{=} \sum_{i=1}^{N} (y_i - f(x_i; u, v, m_1, m_2))^2 \tag{4.2}$$

When solving for parameters which minimize G, it turns out that we may concentrate out v, m_1, m_2. That is, conditional on a value for u, the remaining parameters are all determined. This is easily seen by writing (4.1) as

$$f(x) = v + m_1(x - u)\mathbb{1}_{x \le u} + m_2(x - u)\mathbb{1}_{x > u} \tag{4.3}$$

Then if we fix a value for u, we may solve for the remaining parameters by ordinary least squares (ols). We may explicitly write this solution in closed form as

$$v = \frac{\displaystyle\sum_{i=1}^{N} y_i - \frac{\displaystyle\sum_{x_i \le u} y_i(x_i - u) \sum_{x_i \le u} (x_i - u)}{\displaystyle\sum_{x_i \le u} (x_i - u)^2} - \frac{\displaystyle\sum_{x_i > u} y_i(x_i - u) \sum_{x_i > u} (x_i - u)}{\displaystyle\sum_{x_i > u} (x_i - u)^2}}{N - \frac{\displaystyle\sum_{x_i \le u} (x_i - u) \sum_{x_i \le u} (x_i - u)}{\displaystyle\sum_{x_i \le u} (x_i - u)^2} - \frac{\displaystyle\sum_{x_i > u} (x_i - u) \sum_{x_i > u} (x_i - u)}{\displaystyle\sum_{x_i > u} (x_i - u)^2}} \tag{4.4}$$

$$m_1 = \frac{\displaystyle\sum_{x_i \le u} (y_i - v)(x_i - u)}{\displaystyle\sum_{x_i \le u} (x_i - u)^2} \tag{4.5}$$

$$m_2 = \frac{\displaystyle\sum_{x_i > u} (y_i - v)(x_i - u)}{\displaystyle\sum_{x_i > u} (x_i - u)^2} \tag{4.6}$$

So given u, these formulas determine the remaining parameters v, m_1, m_2. Let us denote the function of u which gives the sum of squares as

$$g(u) \overset{\text{def}}{=} \sum_{i=1}^{N} (y_i - f(x_i; u))^2 \tag{4.7}$$

This function is continuous, but unfortunately not differentiable at the data points $u = x_i$. Furthermore, it can possess multiple local minima, even for moderate sample sizes (see plots in Appendix A). Nonetheless, it is possible to give a closed-form algorithm which solves for the minimum of the sum of squares $g(u)$. This was first described in [7], but see also [12] and [9] for more description and generalizations of the method.

Let us order the data $x_1 < x_2 < \cdots < x_N$. It is easy to see that for $x_1 < u < x_2$ the minimizer of the V-fit is given by taking the the right-hand-side line to be that determined by ols for the data $\{(x_2, y_2), \ldots, (x_N, y_N)\}$, and taking the left-hand-side line to be the straight line from (x_1, y_1) to the right-hand-side line. A similar result holds for $x_{N-1} < u < x_N$. So $g(u)$ is constant on the intervals $[x_1, x_2]$ and $[x_{N-1}, x_N]$.

Next we consider the intervals $x_i < u < x_{i+1}$ for $i = 2, 3, \ldots, N-2$. On such an interval, the function G is smooth. (Simple calculation shows that $\frac{\partial G}{\partial u}$ is discontinous at $u = x_i$). To check for critical points of the function G on the interval $x_j < u < x_{j+1}$, we solve

$$\frac{\partial G}{\partial u} = \frac{\partial G}{\partial v} = \frac{\partial G}{\partial m_1} = \frac{\partial G}{\partial m_2} = 0 \tag{4.8}$$

$\frac{\partial G}{\partial u} = 0$ implies

$$m_1 \sum_{x_i \le x_j} [y_i - v - m_1(x_i - u)] + m_2 \sum_{x_i \ge x_{j+1}} [y_i - v - m_2(x_i - u)] = 0 \tag{4.9}$$

$\frac{\partial G}{\partial v} = 0$ implies

$$\sum_{x_i \le x_j} [y_i - v - m_1(x_i - u)] + \sum_{x_i \ge x_{j+1}} [y_i - v - m_2(x_i - u)] = 0 \tag{4.10}$$

$\frac{\partial G}{\partial m_1} = 0$ implies

$$\sum_{x_i \le x_j} [y_i - v - m_1(x_i - u)]\,(x_i - u) = 0 \tag{4.11}$$

$\frac{\partial G}{\partial m_2} = 0$ implies

$$\sum_{x_i \ge x_{j+1}} [y_i - v - m_2(x_i - u)]\,(x_i - u) = 0 \tag{4.12}$$

Now, (4.9) and (4.10) together imply

$$\left(1 - \frac{m_2}{m_1}\right) \sum_{x_i \ge x_{j+1}} [y_i - v - m_2(x_i - u)] = 0 \tag{4.13}$$

There are two cases to consider. First assume $m_1 \neq m_2$. Then

$$\sum_{x_i \geq x_{j+1}} [y_i - v - m_2(x_i - u)] = 0$$

and (4.10) implies

$$\sum_{x_i \leq x_j} [y_i - v - m_1(x_i - u)] = 0$$

So the system (4.8) becomes

$$\left\{ \begin{array}{l} \sum_{x_i \leq x_j} [y_i - v - m_1(x_i - u)] = 0 \\ \sum_{x_i \leq x_j} [y_i - v - m_1(x_i - u)] (x_i - u) = 0 \end{array} \right. \tag{4.14}$$

$$\left\{ \begin{array}{l} \sum_{x_i \geq x_{j+1}} [y_i - v - m_2(x_i - u)] = 0 \\ \sum_{x_i \geq x_{j+1}} [y_i - v - m_2(x_i - u)] (x_i - u) = 0 \end{array} \right. \tag{4.15}$$

But (4.14) are just the equations which solve for ols for the data set $\{(x_1, y_1), \ldots (x_j, y_j)\}$ and (4.15) are the equations which solve for ols for the data set $\{(x_{j+1}, y_{j+1}), \ldots (x_N, y_N)\}$. We can easily see this by parameterizing the lines using the point-slope formula with the point being (u, v) and the slopes m_1 and m_2 respectively. Setting the derivatives with respect to the parameters equal to zero gives these equations (the *normal equations*).

We compute the closed-form ols solutions for these two data sets. The slopes of the resulting lines and the intersection of the two lines gives the critical points of G. If the u coordinate of the intersection of the two ols lines lies in the interval (x_j, x_{j+1}), then we have found a critical point for G on this interval. Otherwise there are no critical points for G on this interval.

If $m_1 = m_2$, we are in the case of a line fit (ie: ols) to the entire data set. Then every point value for u on the interval (x_j, x_{j+1}) gives a minimum for G, and consequently the function $g(u)$ is constant on the interval.

4.1 Algorithm For V-fit Minimizer

We thus have the following algorithm to find the argmin of G (4.2). We compare the values of g at all the critical points.

1. For each interval $x_i < u < x_{i+1}$ for $i = 2, 3, \ldots, N - 2$ solve ols for the left-hand data set and the right hand data set. If the intersection of the two lines lies in the interval (x_i, x_{i+1}), then record the resulting value for g (4.7).

2. Record the values for g for $u = x_1, x_2, \ldots, x_N$.
3. Find the value of u which gives the minimum value for g among the recorded values in the previous steps. This is the solution.

4.2 Restriction of Fixed Right-Hand-Side Slope

We also consider the model restriction obtained by fixing the right-hand-side slope m_2 to some constant value, $\widetilde{m_2}$. In this case, the formula (4.5) for m_1 remains the same, and (4.4) becomes

$$v = \frac{\displaystyle\sum_{i=1}^{N} y_i - \frac{\displaystyle\sum_{x_i \le u} y_i(x_i - u) \sum_{x_i \le u}(x_i - u)}{\displaystyle\sum_{x_i \le u}(x_i - u)^2} - \widetilde{m_2} \sum_{x_i > u}(x_i - u)}{N - \frac{\displaystyle\sum_{x_i \le u}(x_i - u) \sum_{x_i \le u}(x_i - u)}{\displaystyle\sum_{x_i \le u}(x_i - u)^2}}$$

Then with this change, the algorithm of Section 4.1 remains the same with the exception of replacing Step 1 with the following:

1. For each interval $x_i < u < x_{i+1}$ for $i = 2, 3, \ldots, N-1$ solve ols for the left-hand data set and solve ols-with-restricted-slope for the right hand data set. If the intersection of the two lines lies in the interval (x_i, x_{i+1}), then record the resulting value for g (4.7).

The formula for ols with restricted slope is the following

$$v = \frac{1}{N_r} \sum_{r} [y_i - \widetilde{m_2} x_i] + \widetilde{m_2} u$$

where r denotes the set of (right-hand-side) data points and N_r is the cardinality of this set.

We note that in practice, we add the restriction that there must be at least four distinct data points at the ends. This avoids pathological fits and overfitting that we would not wish to consider valid. It does not affect any estimates, and only seems to have a material effect on the confidence intervals for the variable u.

5 Model Specification Tests

We examine the models of Section 3 for goodness of fit. We do not implement an exhaustive procedure to select only one model. Rather, we eliminate as many specifications as we can, and leave somewhat open the choice between the models that remain. Our intention here is more towards data exploration than deciding on one "best" model. We have chosen to use non-parametric techniques here, but note that we obtained essentially identical results with k-fold cross validation (not shown).

Let y denote the heart-to-head length, and x the total body length. The first set of model specification tests test parametric regression models using bootstrapped comparisons with non-parametric smoothers. We employ two types of these tests. The first is the test `npcmstest` from the R package `np`. The reference for this is [6].

The second test is similar in spirit, and was implemented in python. It is described in the forthcoming book by Shalizi, [11], Section 10.1. An online version of the book is available at: `http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/`. In the implementation of this test, we used python StatsModels kernel regression, with least-squares cross-validation for bandwidth selection.

We have found these two tests to give consistent results. As such, we omit results for the latter test for the first set of regression models. We use the latter test specifically for testing the V-fit model, since our segmented regression code is implemented in python, not R. In what follows, we only report the R `np` tests for all standard regression models. We only report the second test for the V-fit models.

We note that here we are testing six possible models. The results are essentially unchanged if we apply the Bonferroni correction, which would divide the significance levels by six.

The third test is a bootstrapped likelihood ratio test. Specifically, we test the restriction of the V-fit to the straight line ordinary least squares regression model. In this case, under the null hypothesis, we can consider the restricted parameters to lie on a non-identifiable region of the parameter space, so that standard asymptotic properties of the likelihood ratio statistic do not apply, see [2] for more elaboration. However, as in [4] or [10], we claim that we may use the bootstrapped likelihood ratio test in this case. In fact, we ran monte carlo tests (not shown) assessing the power of this method using the model estimates as the true models, and the results support the method. See also [8], and [12]. We also ran permutation tests as in those papers, and the results were essentially the same.

The procedure we follow is to compute the log-likelihood ratio statistic $-2(l_0 - l_1)$, where l_0 is the log-likelihood of the restricted model, and l_1 is the log-likelihood of the V-fit, both evaluated at the respective maximum likelihood estimates. We then simulate bootstrap samples under the null of the restricted model by resampling residuals, as described in Section 6, computing the log-likelihood ratio statistic for each sample. This bootstrapped distribution is then used to compute a p-value corresponding the log-likelihood ratio statistic for the original sample. For reference, this procedure is described in [10], as well as numerous other articles.

The results are divided in the following way. First, we group by snake species. Then we present results for standard regression models using the first test, followed by results for the V-fit using the second test, followed by results of the likelihood ratio test.

5.1 Summary

1. *Agkistrodon conanti*: All models rejected except the V-fit.
2. *Agkistrodon conanti trunk*: All models rejected except the V-fit.
3. *Acrochordus granulatus*: All models rejected except the V-fit.
4. *Hydrophis platurus*: Line model is borderline acceptance; we cannot reject quadratic models; we cannot reject V-fit. We choose the V-fit to compare with the other snakes and since the line is nearly rejected. However, the line as the restriction of the V-fit is borderline not rejected by our bootstrapped likelihood ratio test. This corresponds to a line model with the ratio $\frac{y}{x}$ as the dependent variable, ie: model (3.1). This ratio line model is also not rejected by the specification test 5.5.3. We consider both the V-fit and (3.1).
5. *Pantherophis obsoleta*: We cannot reject any model. In interest of parsimony, we choose the line model. We also will look at results for the V-fit by way of comparison with the other snakes. The line restriction of the V-fit is not rejected by the bootstrapped likelihood ratio test. The ratio line model is also not rejected by the specification test 5.6.3.

5.2 *Agkistrodon conanti*

5.2.1 $y \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 6.494182        P Value: < 2.22e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.2.2 $log(y) \sim a + b\ log(x)$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 5.079437        P Value: < 2.22e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.2.3 $y \sim bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 8.085069        P Value: < 2.22e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.2.4 $y \sim a + bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 7.57716       P Value: < 2.22e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.2.5 $\frac{y}{x} \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 5.942697       P Value: < 2.22e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.2.6 V-fit

Next we show results for the second specification test described above for the V-fit. Note that the null hypothesis is that the parametric model is true.

```
bootstrap num replications:  10000
p value:  0.1485
```

5.2.7 V-fit Bootstrapped Likelihood Ratio Test for Line Restriction

Next we show results for the third specification test described above for the V-fit. Recall that the null hypothesis is that the restricted line model is true.

```
bootstrap num replications: 1000000
bootstrapped p value: 1e-06
```

5.3 *Agkistrodon conanti trunk*

5.3.1 $y \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 5.24261       P Value: < 2.22e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.3.2 $log(y) \sim a + b\ log(x)$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 4.514148        P Value: 1e-04 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.3.3 $y \sim bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 5.538031        P Value: < 2.22e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.3.4 $y \sim a + bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 5.800038        P Value: < 2.22e-16 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.3.5 $\frac{y}{x} \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 4.004986        P Value: 2e-04 ***
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 0.1% level
```

5.3.6 V-fit

Next we show results for the second specification test described above for the V-fit. Note that the null hypothesis is that the parametric model is true.

```
bootstrap num replications:  10000
p value:  0.0650
```

5.3.7 V-fit Bootstrapped Likelihood Ratio Test for Line Restriction

Next we show results for the third specification test described above for the V-fit. Recall that the null hypothesis is that the restricted line model is true.

```
bootstrap num replications:  1000000
bootstrapped p value:  0.0002
```

5.4 *Acrochordus granulatus*

5.4.1 $y \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 2.706697        P Value: 0.0033 **
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 1% level
```

5.4.2 $log(y) \sim a + b\ log(x)$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 2.529444        P Value: 0.0055 **
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 1% level
```

5.4.3 $y \sim bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 2.655755        P Value: 0.0038 **
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 1% level
```

5.4.4 $y \sim a + bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 2.063245        P Value: 0.009 **
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Null of correct specification is rejected at the 1% level
```

5.4.5 $\frac{y}{x} \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 2.754004        P Value: 0.0026 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null of correct specification is rejected at the 1% level
```

5.4.6 V-fit

Next we show results for the second specification test described above for the V-fit. Note that the null hypothesis is that the parametric model is true.

```
bootstrap num replications:  10000
p value: 0.1409
```

5.4.7 V-fit Bootstrapped Likelihood Ratio Test for Line Restriction

Next we show results for the third specification test described above for the V-fit. Recall that the null hypothesis is that the restricted line model is true.

```
bootstrap num replications: 1000000
bootstrapped p value: 0.0001
```

5.5 *Hydrophis platurus*

5.5.1 $y \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 1.069891        P Value: 0.0523 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null of correct specification is rejected at the 10% level
```

5.5.2 $log(y) \sim a + b\ log(x)$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 1.183549        P Value: 0.0433 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null of correct specification is rejected at the 5% level
```

5.5.3 $y \sim bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: -0.0159585       P Value: 0.2522
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Fail to reject the null of correct specification at the 10% level
```

5.5.4 $y \sim a + bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: -0.2153411       P Value: 0.2575
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Fail to reject the null of correct specification at the 10% level
```

5.5.5 $\frac{y}{x} \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: 0.4091021       P Value: 0.1164
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Fail to reject the null of correct specification at the 10% level
```

5.5.6 V-fit

Next we show results for the second specification test described above for the V-fit. Note that the null hypothesis is that the parametric model is true.

```
bootstrap num replications:  10000
p value:  0.7701
```

5.5.7 V-fit Bootstrapped Likelihood Ratio Test for Line Restriction

Next we show results for the third specification test described above for the V-fit. Recall that the null hypothesis is that the restricted line model is true.

```
bootstrap num replications: 1000000
bootstrapped p value: 0.0481
```

5.6 *Pantherophis obsoleta*

5.6.1 $y \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: -1.111558       P Value: 0.7903
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Fail to reject the null of correct specification at the 10% level
```

5.6.2 $log(y) \sim a + b\ log(x)$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: -1.21196       P Value: 0.8593
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Fail to reject the null of correct specification at the 10% level
```

5.6.3 $y \sim bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: -1.016752       P Value: 0.7198
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Fail to reject the null of correct specification at the 10% level
```

5.6.4 $y \sim a + bx + cx^2$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: -1.187718       P Value: 0.7618
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Fail to reject the null of correct specification at the 10% level
```

5.6.5 $\frac{y}{x} \sim a + bx$

```
IID Bootstrap (10000 replications)
Test Statistic Jn: -0.8425649       P Value: 0.4931
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
Fail to reject the null of correct specification at the 10% level
```

5.6.6 V-fit

Next we show results for the second specification test described above for the V-fit. Note that the null hypothesis is that the parametric model is true.

```
bootstrap num replications:  10000
p value: 0.8215
```

5.6.7 V-fit Bootstrapped Likelihood Ratio Test for Line Restriction

Next we show results for the third specification test described above for the V-fit. Recall that the null hypothesis is that the restricted line model is true.

```
bootstrap num replications: 1000000
bootstrapped p value: 0.3893
```

---

6 Estimation Results

We present our findings by species. For all species, we show the V-fits. For *Hydrophis platurus* we show in addition the ratio line fit (3.1). For *Pantherophis obsoleta* we show in addition the line fit and the ratio line fit (3.1).

To estimate the V-fit, we use segmented regression as described in Section 4. Since we have an exact algorithm for estimating the segmented regression, we may employ bootstrapping of sampling parameters with no issues such as might be had when using a numerical optimization routine. We note in passing the R package `segmented` which estimates segmented regressions. It uses a numerical optimization routine, so we prefer our exact numerical algorithm, which we implemented in python. However, we cross checked our results and the `segmented` package gives nearly identical parameter estimates. We show graphs of the sum-of-squares functions we minimize in Appendix A.

Regarding the V-fit of the ratio regression data, small to medium sample sizes can have parameter sampling distributions which are quite far from (asymptotic) normal, so we present bootstrap statistics for the parameters.

For the bootstrap replications, we simulate in two standard ways. We first sample with replacement from the fitted initial V-fit residuals to generate error terms. We keep the data for the body lengths (the x values) constant. Then we generate a bootstrap sample using the estimated V-fit function together with the simulated error terms. We'll refer to this below as *resampling residuals*.

The second form of bootstrapping regards the original dataset as random draws from some joint distribution of $\{(X, Y)\}$. The bootstrapping resampling takes draws with replacement from the original data pairs $\{(x_i, y_i)\}$. We'll refer to this bootstrap scheme as *resampling cases*, following terminology in [1], Chapter 6, to which we refer the reader for more information.

We present plots of the bootstrap sampling distributions for the right-hand-side slope parameter under the resampling residuals scheme. In all cases, it is highly skewed.

We note that the residuals for the selected models in this section evidence less heteroscedasticity than the models rejected in Section 5, although we have not shown this.

We also note that in the standard OLS regressions below, we checked the bootstrapped standard errors (not shown), and they were all quite close to those reported for the OLS.

6.1 Summary

We summarize the findings as follows.

- *Agkistrodon conanti*: V-fit right-hand-side slope is positive. The bootstrap confidence intervals for resampling residuals and cases are similar.
- *Agkistrodon conanti trunk*: Results are broadly similar to those for *Agkistrodon conanti*. However, the bootstrap confidence intervals for resampling residuals are different from those for cases, being much wider. This is the only dataset with this behavior.
- *Acrochordus granulatus*: Similar results as for *Agkistrodon conanti*. However, the data here is quite sparse for shorter snakes, so we take the results with a greater degree of caution.
- *Hydrophis platurus*: The V-fit pattern is inverted relative to *Agkistrodon conanti* and *Acrochordus granulatus*. However, two separate model specification tests do not reject the ratio line model (3.1). Graphically, the V-fit is close to the non-parametric regression. But the line fit for the ratio data would be more in line with the biological theory.
- *Pantherophis obsoleta*: The line fits well to the original data (no ratios), and agrees with the non-parametric fit. The regression R^2 is high and the bootstrapped standard errors are very close to the asymptotic standard errors. This is our preferred specification.

  As regards the V-fit, the right-hand-side slope is nearly zero, which does not agree with the non-parametric regression which is nearly a straight line with the ratio variable as dependent variable. The line restriction of the V-fit is not rejected. Based on this evidence, we prefer the line model with the ratio variable (3.1) when using the ratio data.

6.2 *Agkistrodon conanti*

V-Fit

Bootstrap BCA Confidence Intervals

num bootstrap resamples: 1,000,000

Resample Residuals

Resample Cases

confidence level: 90%

	estimate	lower	upper
u	107.000	98.761	125.910
v	31.393	30.893	31.826
m_1	-0.049	-0.055	-0.042
m_2	0.023	0.001	0.054
σ	1.045	0.974	1.153

	estimate	lower	upper
u	107.000	106.500	131.500
v	31.393	31.080	31.697
m_1	-0.049	-0.054	-0.043
m_2	0.023	0.007	0.040
σ	1.045	0.970	1.149

confidence level: 95%

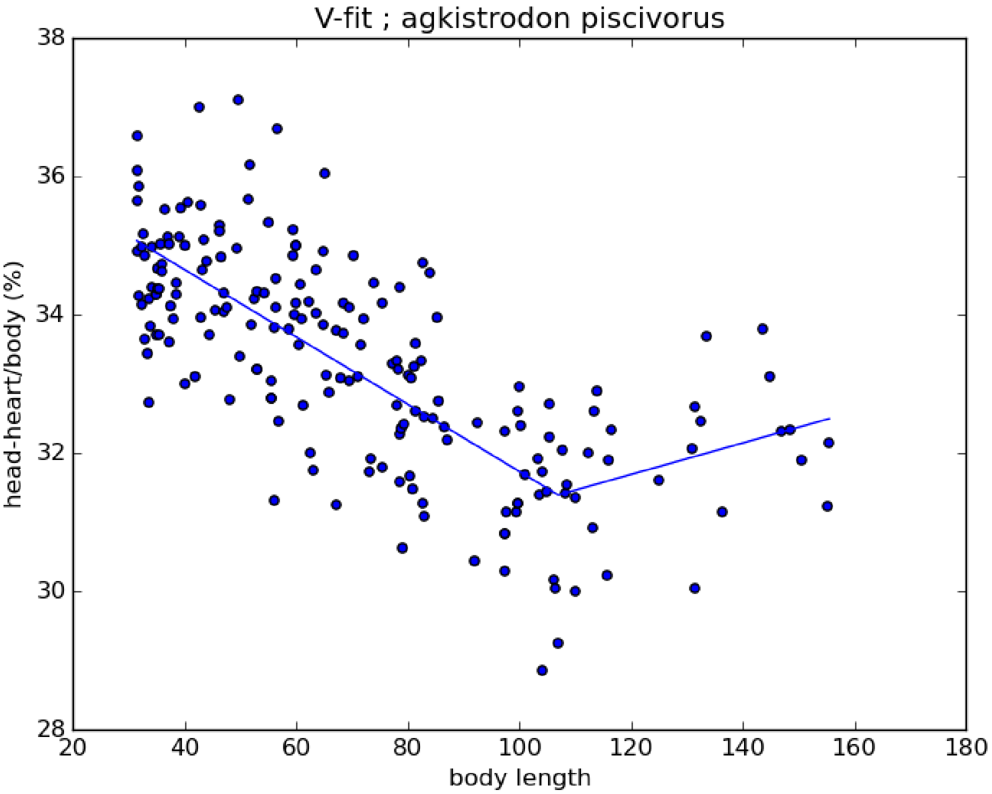
	estimate	lower	upper
u	107.000	97.300	129.955
v	31.393	30.753	31.915
m_1	-0.049	-0.057	-0.041
m_2	0.023	-0.003	0.068
σ	1.045	0.958	1.172

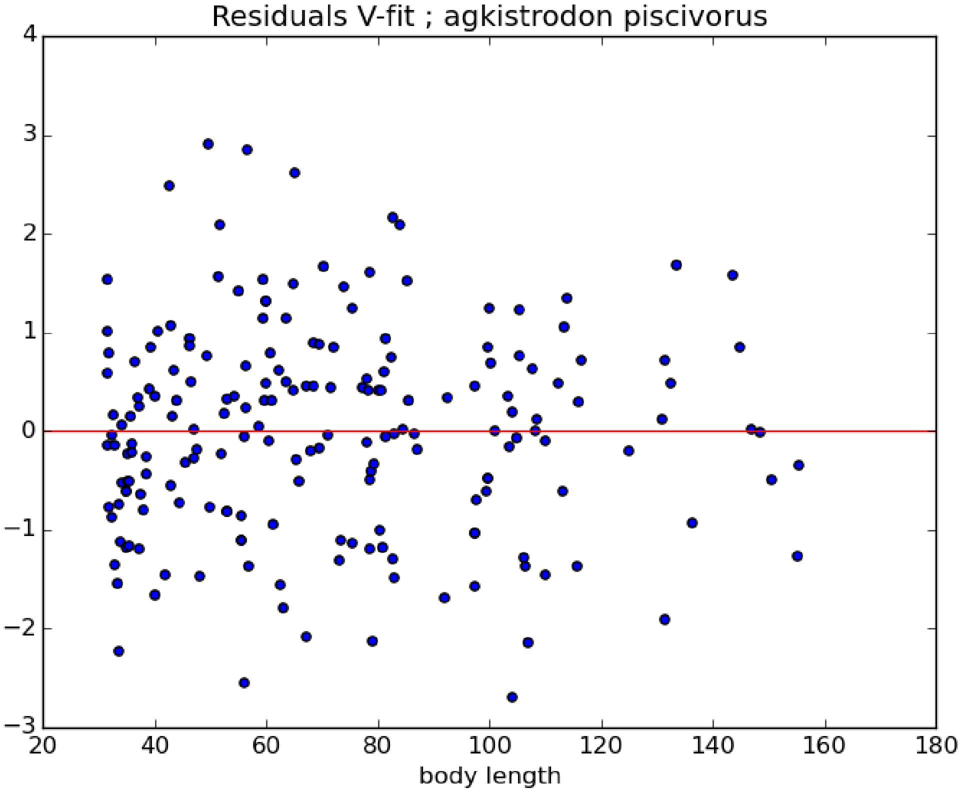
	estimate	lower	upper
u	107.000	106.100	131.500
v	31.393	30.978	31.757
m_1	-0.049	-0.055	-0.042
m_2	0.023	0.005	0.047
σ	1.045	0.955	1.168

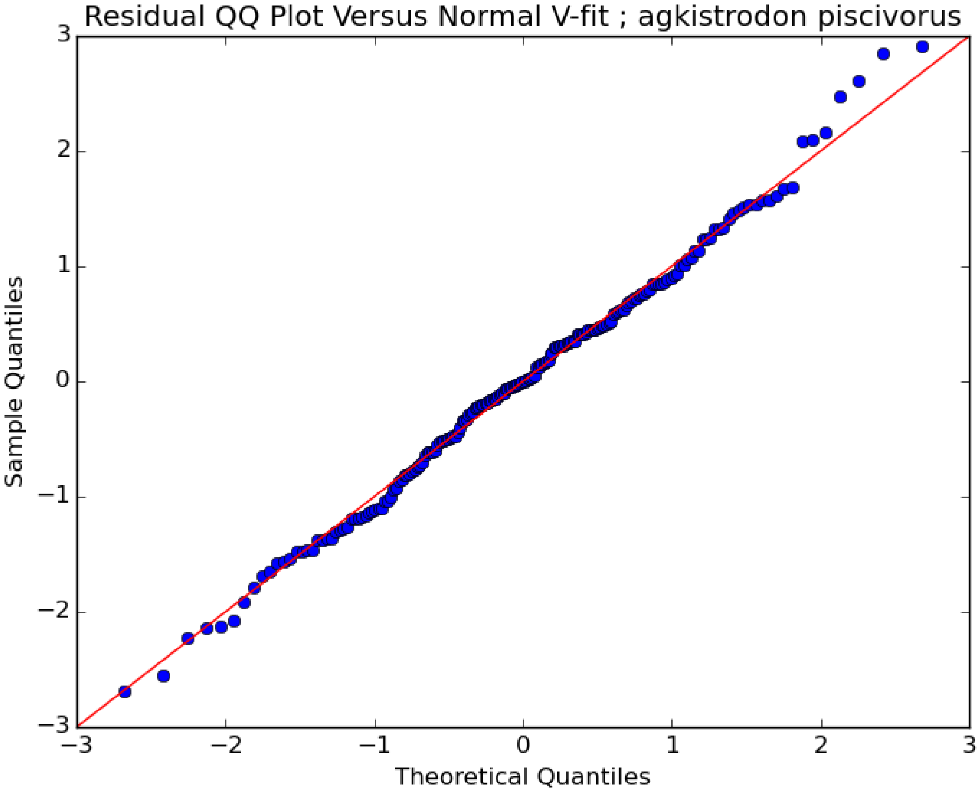
confidence level: 99%

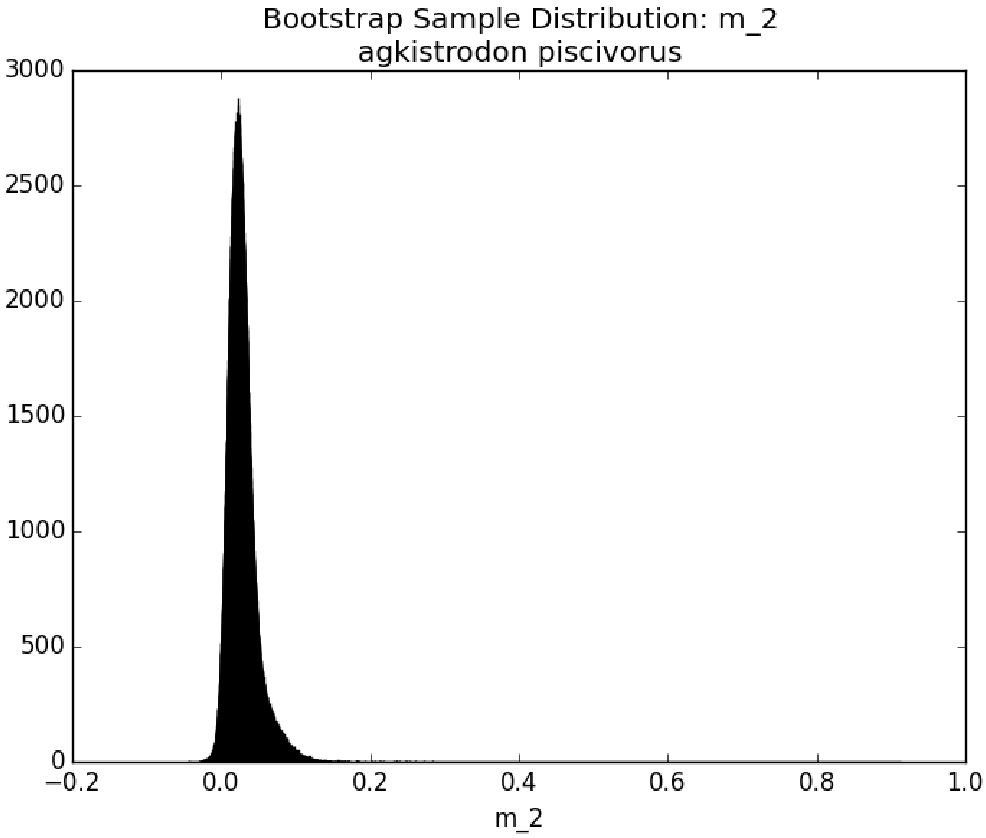
	estimate	lower	upper
u	107.000	93.508	146.663
v	31.393	30.502	32.159
m_1	-0.049	-0.061	-0.039
m_2	0.023	-0.012	0.095
σ	1.045	0.929	1.210

	estimate	lower	upper
u	107.000	104.000	133.840
v	31.393	30.608	31.868
m_1	-0.049	-0.057	-0.040
m_2	0.023	-0.001	0.076
σ	1.045	0.925	1.207

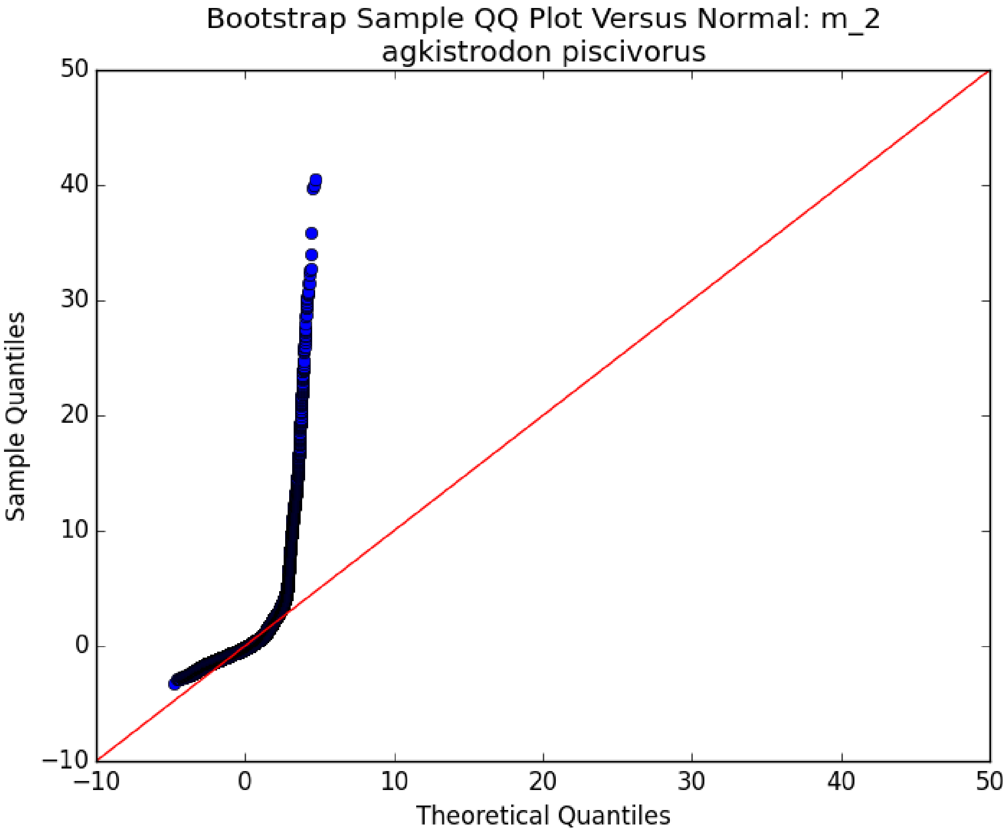

V-fit ; agkistrodon piscivorus
head-heart/body (%)
body length
38
36
34
32
30
28
20
40
60
80
100
120
140
160
180


Residuals V-fit ; agkistrodon piscivorus
4
3
2
1
0
−1
−2
−3
20
40
60
80
100
120
140
160
180
body length


Residual QQ Plot Versus Normal V-fit ; agkistrodon piscivorus
Sample Quantiles
Theoretical Quantiles
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3


Bootstrap Sample Distribution: m_2
agkistrodon piscivorus
3000
2500
2000
1500
1000
500
0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
m_2

V-Fit



6.3 *Agkistrodon conanti trunk*

V-Fit

Bootstrap BCA Confidence Intervals

num bootstrap resamples: 1,000,000

Resample Residuals

Resample Cases

confidence level: 90%

	estimate	lower	upper
u	85.590	62.790	100.069
v	33.811	33.134	34.629
m_1	-0.056	-0.068	-0.045
m_2	0.030	-0.012	0.094
σ	1.477	1.368	1.643

	estimate	lower	upper
u	85.590	82.700	90.210
v	33.811	33.382	34.224
m_1	-0.056	-0.065	-0.047
m_2	0.030	0.007	0.058
σ	1.477	1.363	1.636

confidence level: 95%

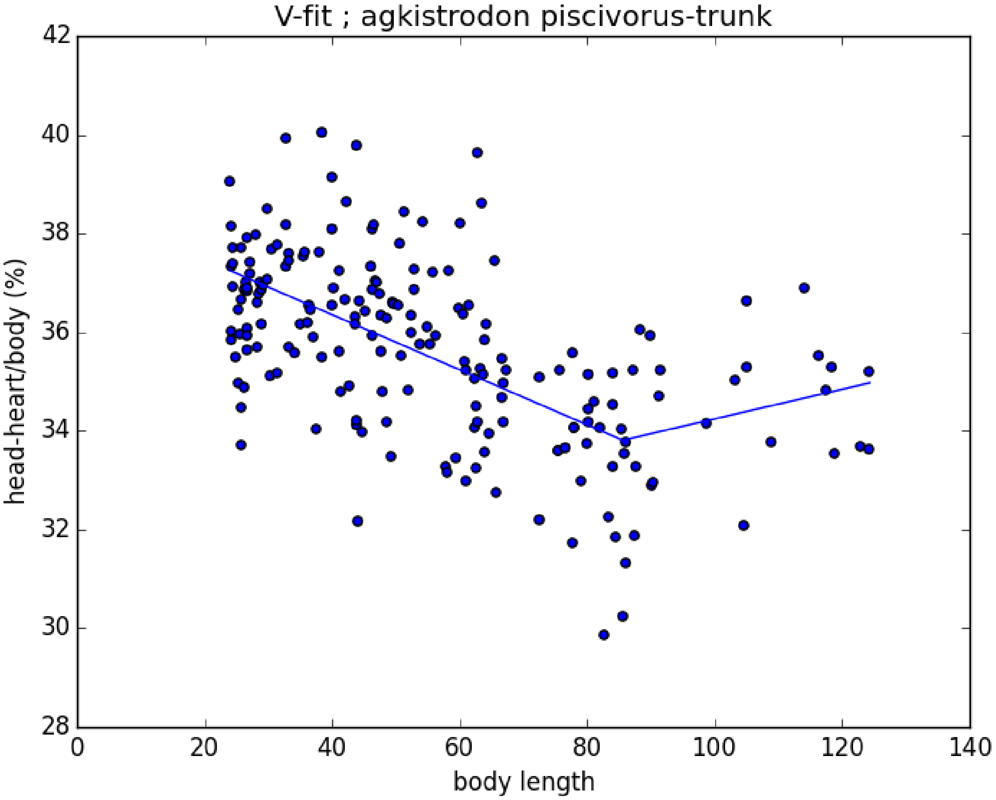
	estimate	lower	upper
u	85.590	42.186	103.130
v	33.811	32.952	35.094
m_1	-0.056	-0.071	-0.043
m_2	0.030	-0.022	0.125
σ	1.477	1.345	1.673

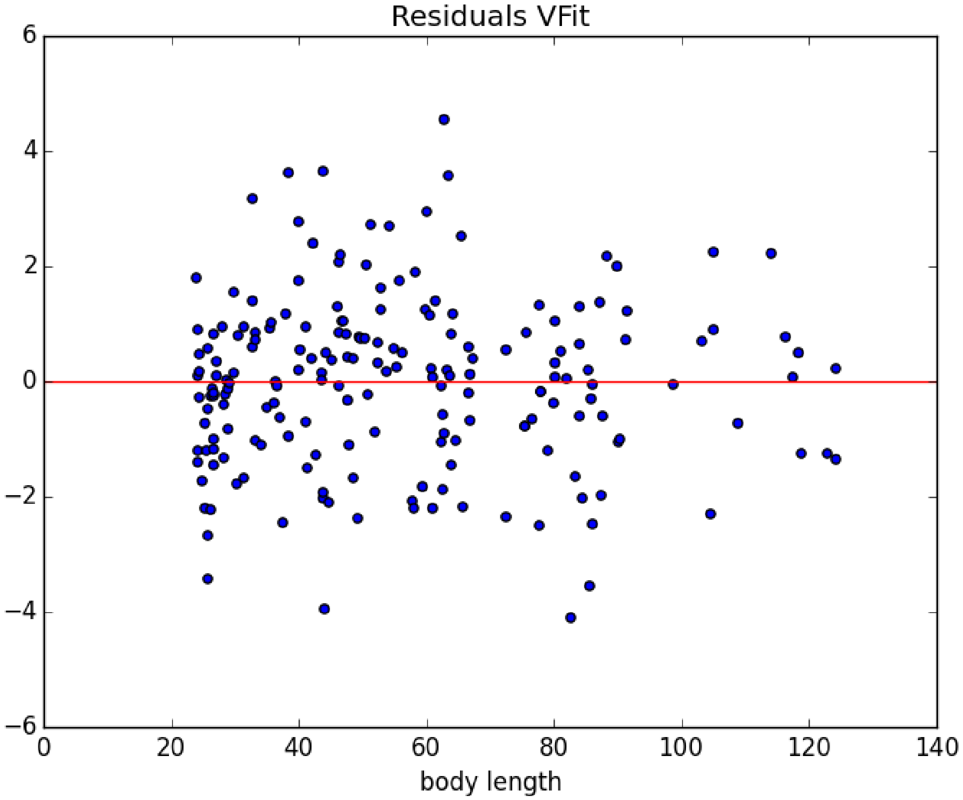
	estimate	lower	upper
u	85.590	82.700	96.919
v	33.811	33.263	34.301
m_1	-0.056	-0.067	-0.045
m_2	0.030	0.003	0.068
σ	1.477	1.340	1.666

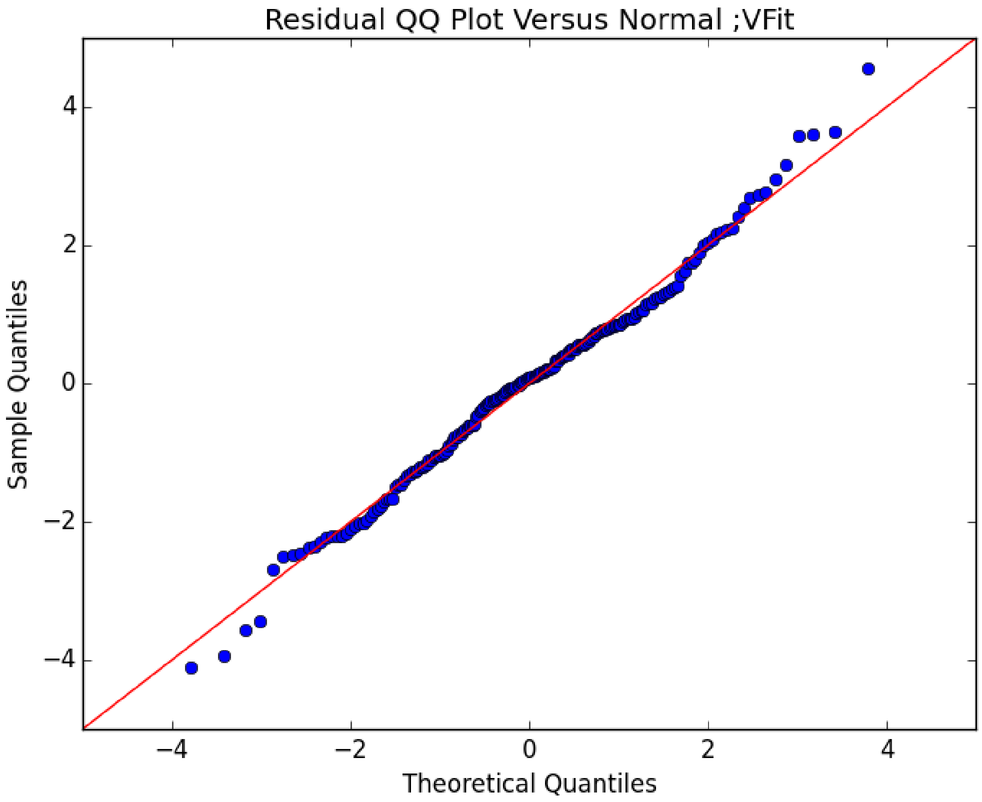
confidence level: 99%

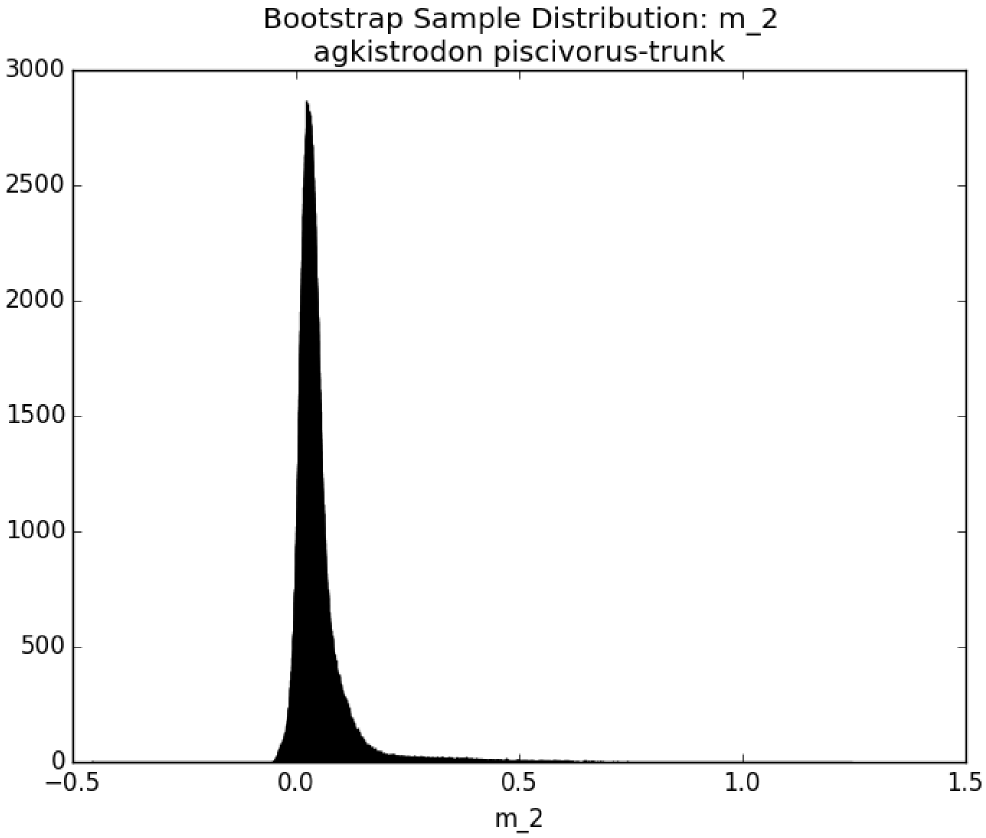
	estimate	lower	upper
u	85.590	24.230	111.104
v	33.811	32.613	36.557
m_1	-0.056	-0.089	-0.039
m_2	0.030	-0.038	0.294
σ	1.477	1.302	1.735

	estimate	lower	upper
u	85.590	77.650	104.550
v	33.811	32.901	34.459
m_1	-0.056	-0.071	-0.041
m_2	0.030	-0.005	0.113
σ	1.477	1.296	1.726

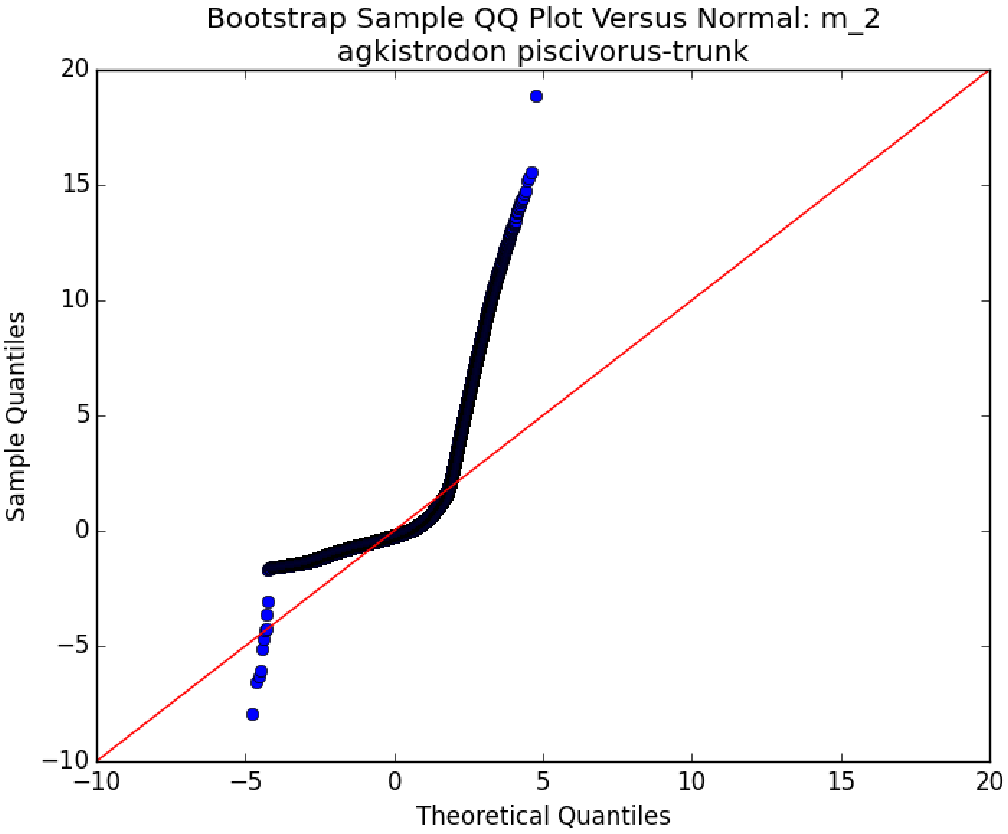

V-fit ; agkistrodon piscivorus-trunk
head-heart/body (%)
42
40
38
36
34
32
30
28
0
20
40
60
80
100
120
140
body length


Residuals VFit
6
4
2
0
−2
−4
−6
0
20
40
60
80
100
120
140
body length


Residual QQ Plot Versus Normal ;VFit
Sample Quantiles
Theoretical Quantiles
4
2
0
−2
−4
−4
−2
0
2
4


Bootstrap Sample Distribution: m_2
agkistrodon piscivorus-trunk
3000
2500
2000
1500
1000
500
0
−0.5
0.0
0.5
1.0
1.5
m_2

V-Fit



6.4 *Acrochordus granulatus*

V-Fit

Bootstrap BCA Confidence Intervals

num bootstrap resamples: 1,000,000

Resample Residuals

Resample Cases

confidence level: 90%

Resample Residuals

	estimate	lower	upper
u	64.000	58.100	67.500
v	44.329	44.019	44.733
m_1	-0.087	-0.125	-0.051
m_2	0.202	0.090	0.369
σ	1.212	1.061	1.566

Resample Cases

	estimate	lower	upper
u	64.000	62.600	69.768
v	44.329	44.008	44.756
m_1	-0.087	-0.133	-0.066
m_2	0.202	0.098	0.321
σ	1.212	1.051	1.544

confidence level: 95%

Resample Residuals

	estimate	lower	upper
u	64.000	50.620	68.723
v	44.329	43.954	44.809
m_1	-0.087	-0.135	-0.044
m_2	0.202	0.070	0.470
σ	1.212	1.030	1.633

Resample Cases

	estimate	lower	upper
u	64.000	61.316	70.077
v	44.329	43.939	44.839
m_1	-0.087	-0.151	-0.060
m_2	0.202	0.073	0.416
σ	1.212	1.021	1.611

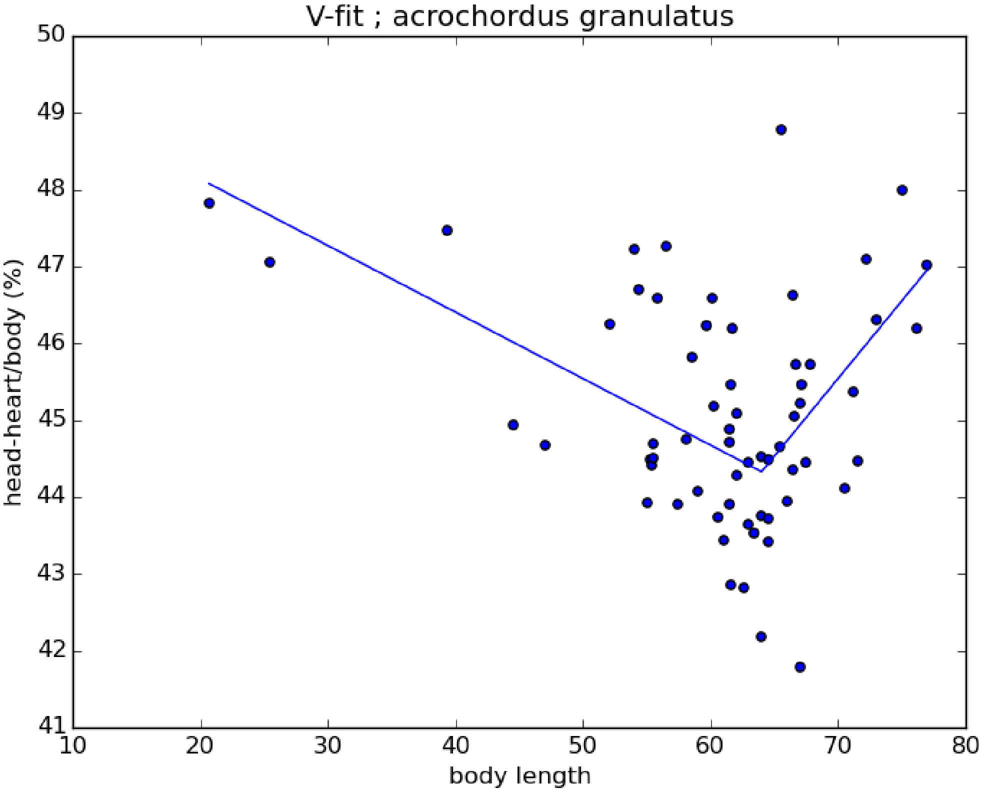
confidence level: 99%

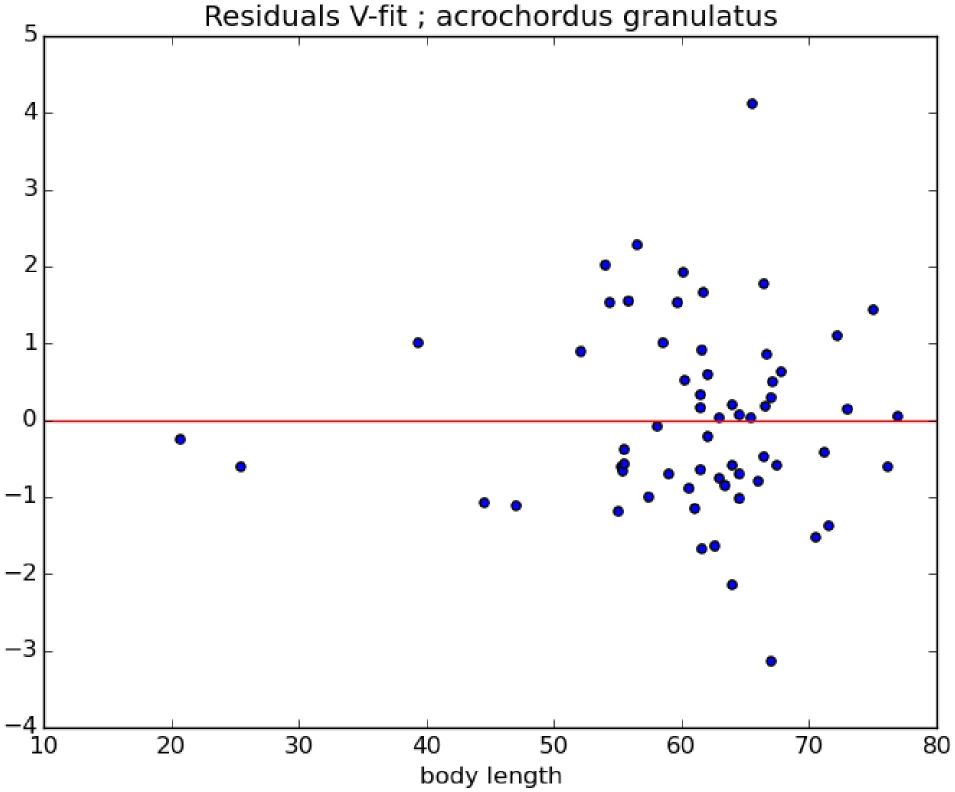
Resample Residuals

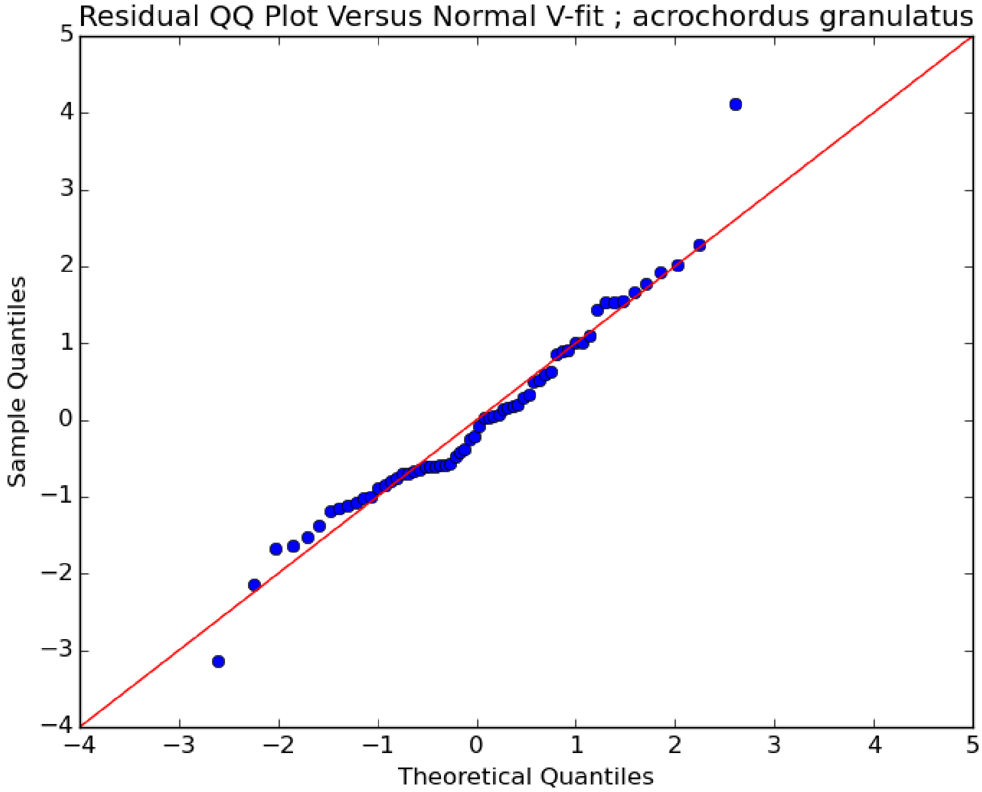
	estimate	lower	upper
u	64.000	44.500	69.961
v	44.329	43.827	44.966
m_1	-0.087	-0.162	-0.031
m_2	0.202	0.037	0.720
σ	1.212	0.976	1.765

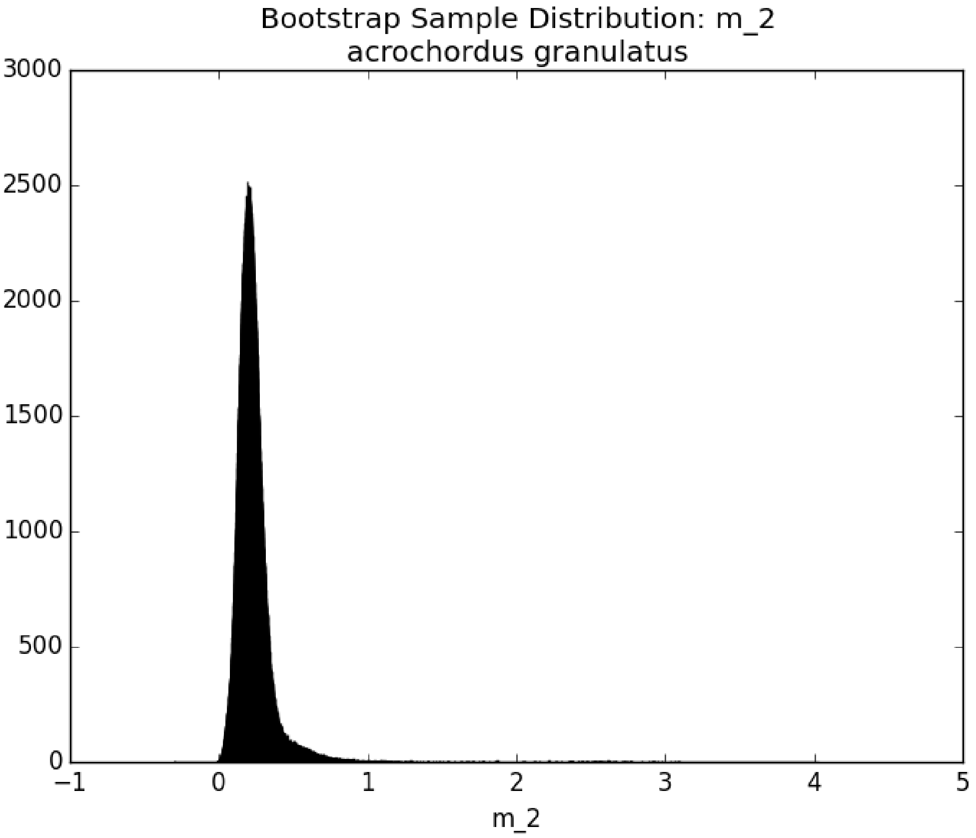
Resample Cases

	estimate	lower	upper
u	64.000	47.000	70.421
v	44.329	43.801	45.038
m_1	-0.087	-0.193	-0.043
m_2	0.202	0.005	0.746
σ	1.212	0.970	1.761

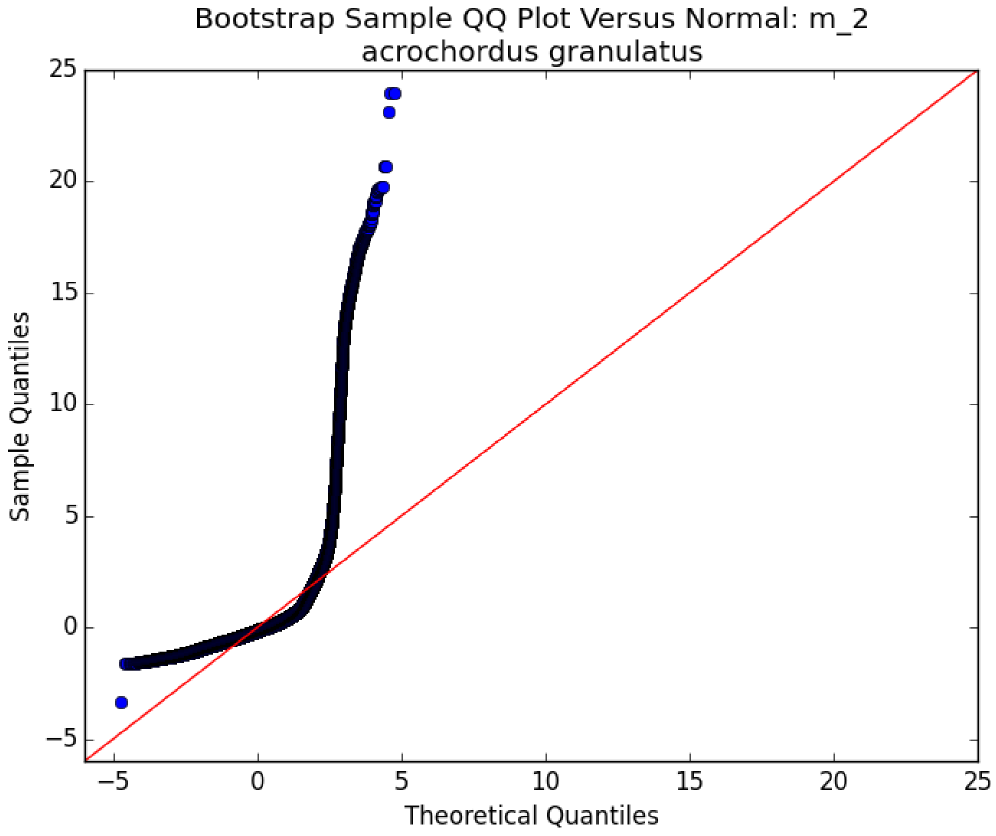

V-fit ; acrochordus granulatus
head-heart/body (%)
body length
41
42
43
44
45
46
47
48
49
50
10
20
30
40
50
60
70
80


Residuals V-fit ; acrochordus granulatus
5
4
3
2
1
0
−1
−2
−3
−4
10
20
30
40
50
60
70
80
body length


Residual QQ Plot Versus Normal V-fit ; acrochordus granulatus
Sample Quantiles
Theoretical Quantiles
5
4
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3
4
5


Bootstrap Sample Distribution: m_2
acrochordus granulatus
3000
2500
2000
1500
1000
500
0
−1
0
1
2
3
4
5
m_2

V-Fit



6.5 *Hydrophis platurus*

V-Fit

Bootstrap BCA Confidence Intervals

num bootstrap resamples: 1,000,000

Resample Residuals

Resample Cases

confidence level: 90%

	estimate	lower	upper
u	44.324	32.100	61.000
v	25.474	24.182	25.829
m_1	0.022	-0.042	0.124
m_2	-0.060	-0.085	-0.034
σ	1.039	0.972	1.141

	estimate	lower	upper
u	44.324	32.700	70.000
v	25.474	24.268	25.929
m_1	0.022	-0.035	0.157
m_2	-0.060	-0.079	-0.013
σ	1.039	0.973	1.142

confidence level: 95%

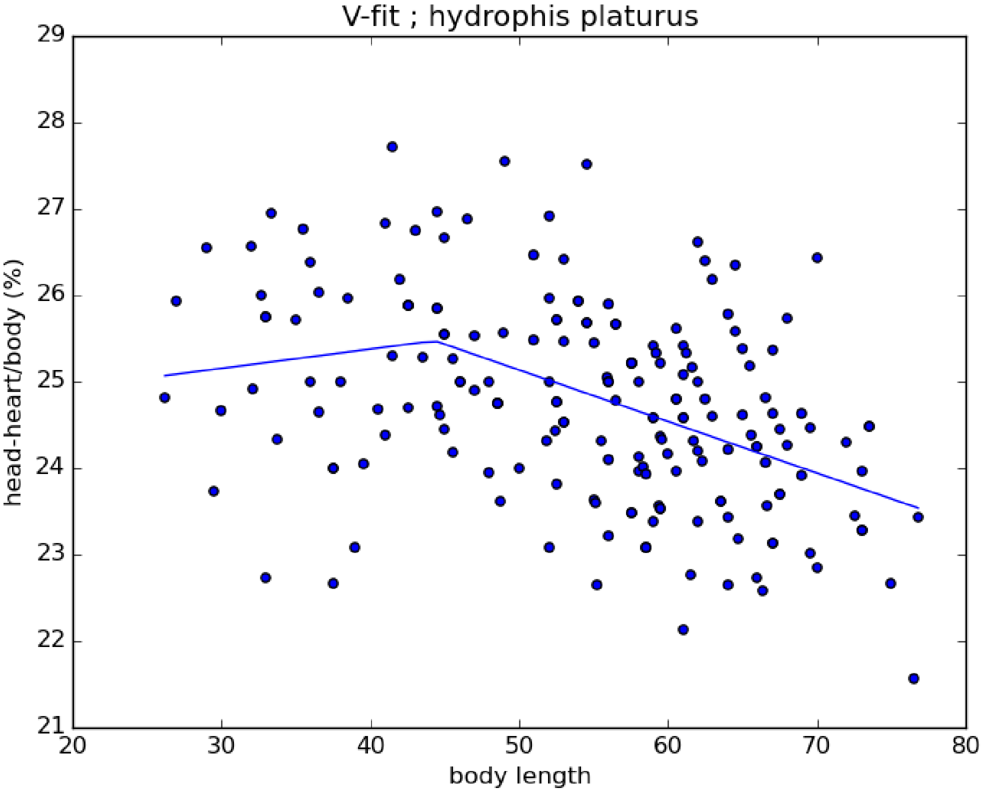
	estimate	lower	upper
u	44.324	30.000	69.000
v	25.474	24.001	25.897
m_1	0.022	-0.050	0.185
m_2	-0.060	-0.101	0.080
σ	1.039	0.958	1.159

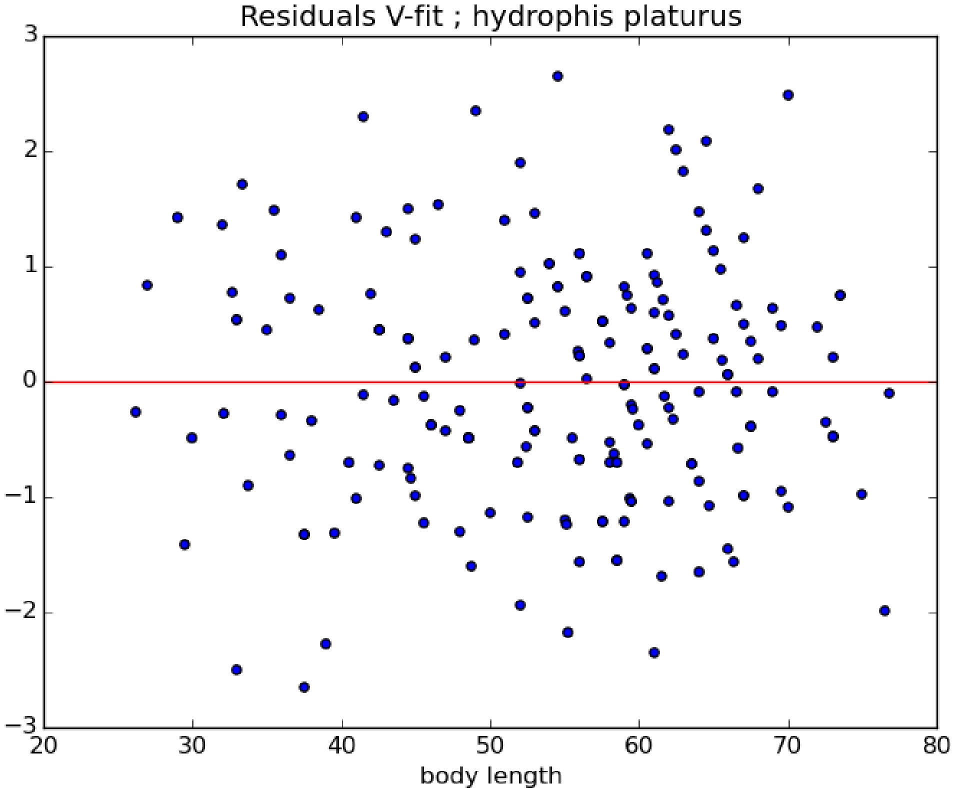
	estimate	lower	upper
u	44.324	32.000	71.907
v	25.474	24.146	26.004
m_1	0.022	-0.040	0.252
m_2	-0.060	-0.115	0.011
σ	1.039	0.958	1.159

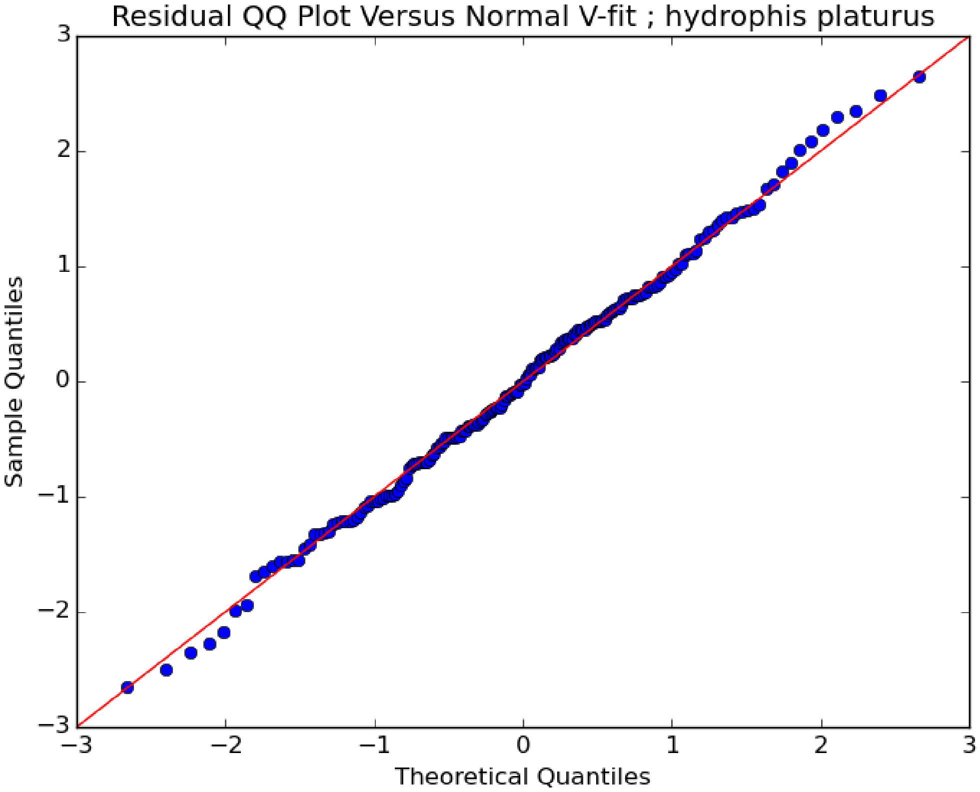
confidence level: 99%

	estimate	lower	upper
u	44.324	29.500	73.000
v	25.474	23.764	26.024
m_1	0.022	-0.464	0.454
m_2	-0.060	-0.297	0.368
σ	1.039	0.930	1.194

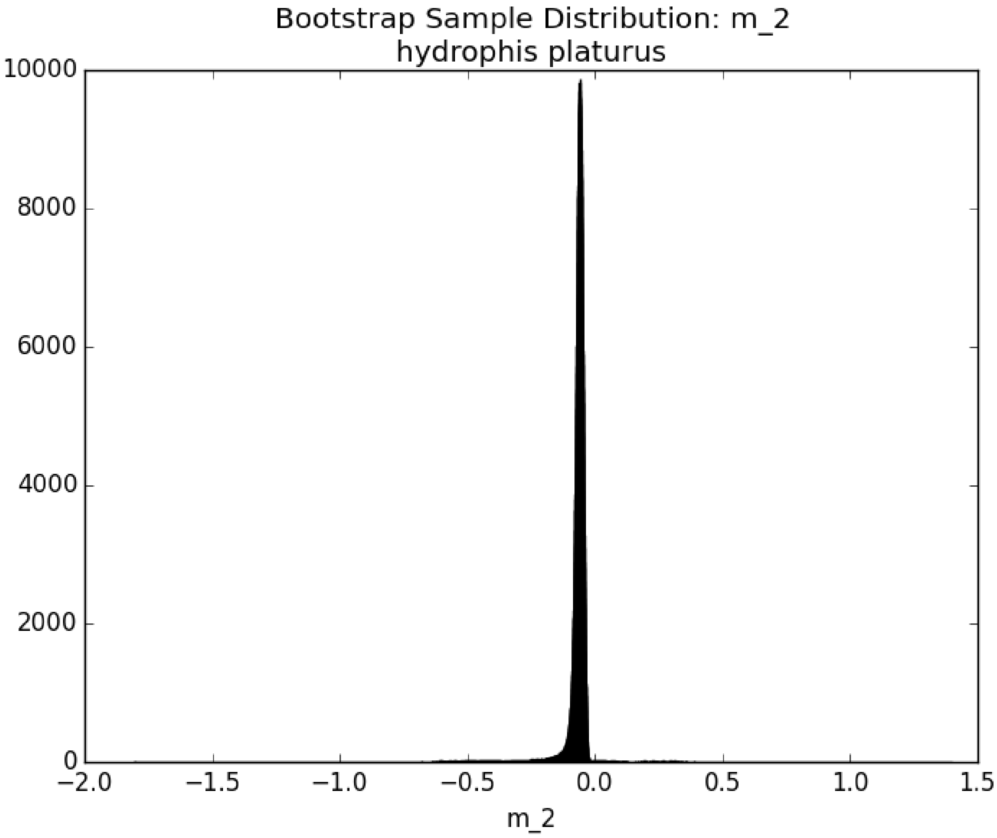
	estimate	lower	upper
u	44.324	32.000	72.810
v	25.474	23.975	26.148
m_1	0.022	-0.048	0.562
m_2	-0.060	-0.442	0.093
σ	1.039	0.930	1.195

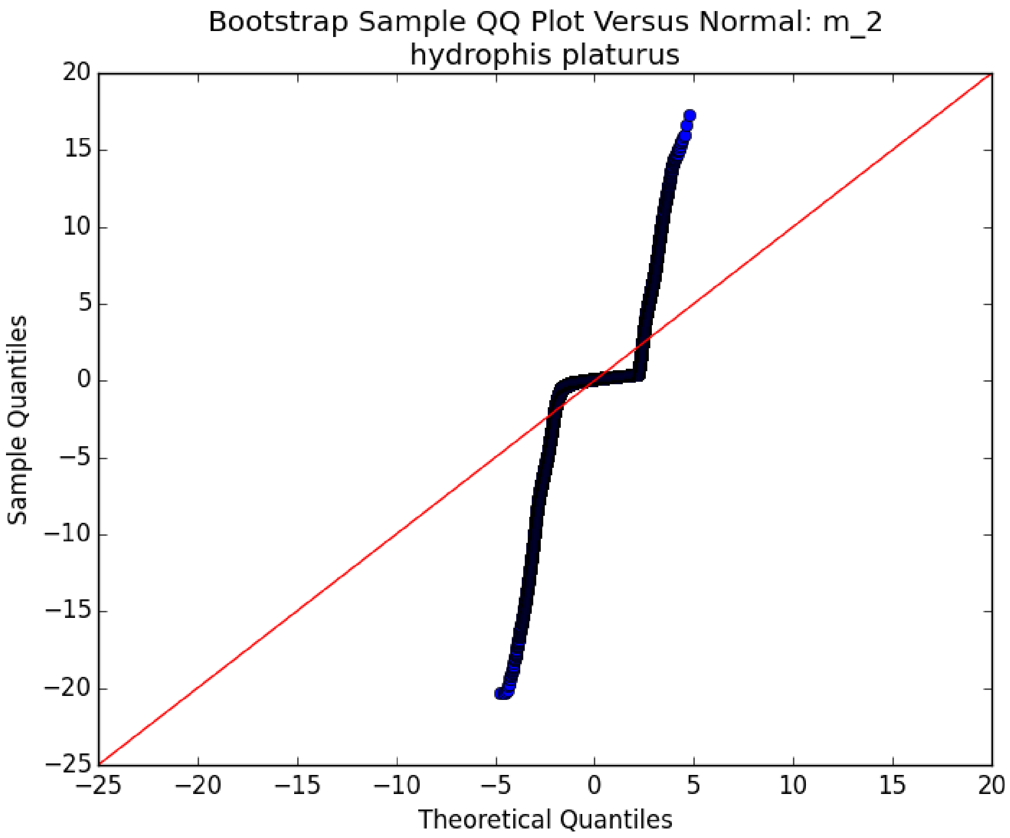

V-fit ; hydrophis platurus
head-heart/body (%)
body length
21
22
23
24
25
26
27
28
29
20
30
40
50
60
70
80


Residuals V-fit ; hydrophis platurus
3
2
1
0
−1
−2
−3
20
30
40
50
60
70
80
body length



Residual QQ Plot Versus Normal V-fit ; hydrophis platurus
Sample Quantiles
Theoretical Quantiles
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

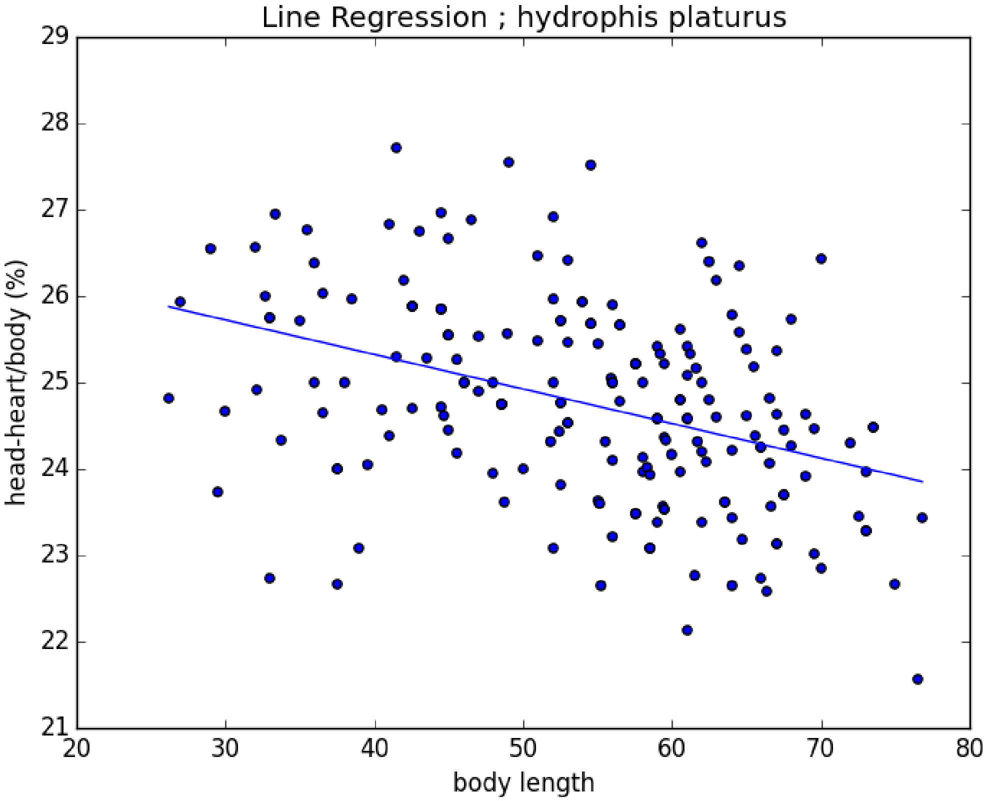

Bootstrap Sample Distribution: m_2
hydrophis platurus
10000
8000
6000
4000
2000
0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
m_2

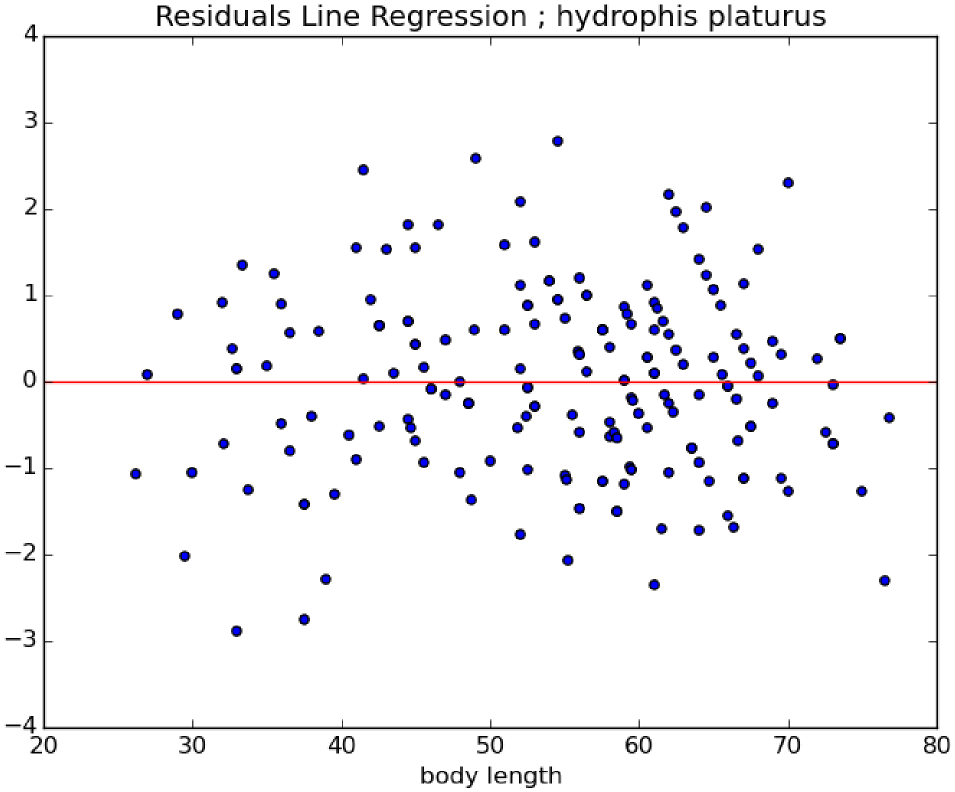


OLS Ratio Line Fit (3.1)

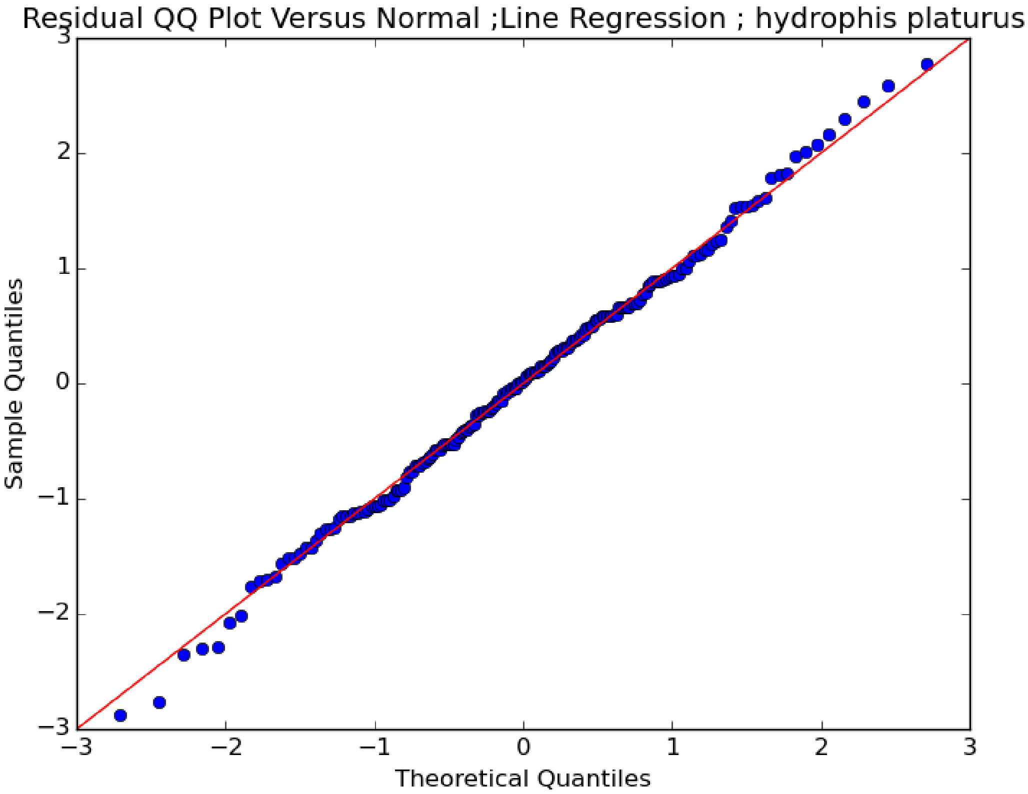
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.159
Model:                            OLS   Adj. R-squared:                  0.154
Method:                 Least Squares   F-statistic:                     35.49
Date:                Sun, 15 Mar 2015   Prob (F-statistic):           1.24e-08
Time:                        18:13:15   Log-Likelihood:                -280.52
No. Observations:                 190   AIC:                             565.0
Df Residuals:                     188   BIC:                             571.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
const         26.9233      0.373     72.258      0.000        26.188    27.658
x1            -0.0400      0.007     -5.958      0.000        -0.053    -0.027
==============================================================================
Omnibus:                        0.008   Durbin-Watson:                   2.012
Prob(Omnibus):                  0.996   Jarque-Bera (JB):                0.041
Skew:                          -0.013   Prob(JB):                        0.980
Kurtosis:                       2.933   Cond. No.                         268.
```

```
==============================================================================
```


Line Regression ; hydrophis platurus
head-heart/body (%)
21
22
23
24
25
26
27
28
29
body length
20
30
40
50
60
70
80


Residuals Line Regression ; hydrophis platurus
4
3
2
1
0
−1
−2
−3
−4
20
30
40
50
60
70
80
body length

V-Fit



6.6 *Pantherophis obsoleta*

V-Fit

Bootstrap BCA Confidence Intervals

num bootstrap resamples: 1,000,000

Resample Residuals

Resample Cases

confidence level: 90%

	estimate	lower	upper
u	152.000	122.500	166.943
v	16.856	16.273	17.819
m_1	-0.024	-0.031	0.039
m_2	0.001	-0.023	0.054
σ	0.634	0.581	0.758

	estimate	lower	upper
u	152.000	152.000	167.000
v	16.856	16.295	17.264
m_1	-0.024	-0.029	0.017
m_2	0.001	-0.018	0.041
σ	0.634	0.576	0.753

confidence level: 95%

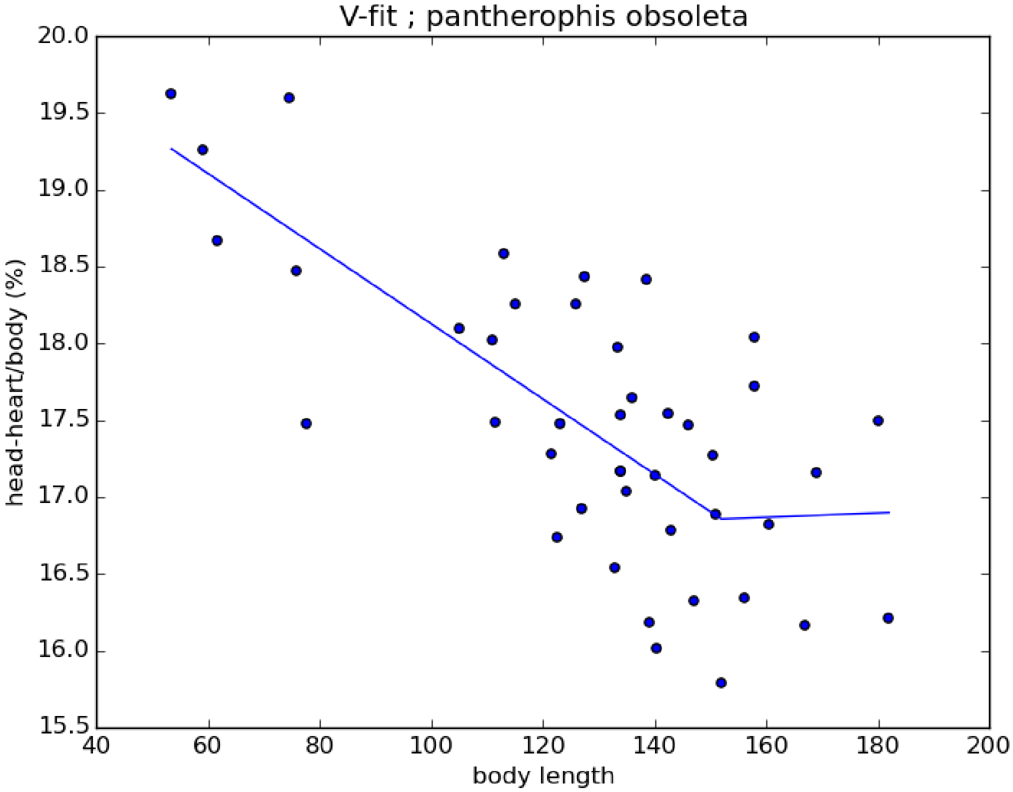
	estimate	lower	upper
u	152.000	105.000	166.992
v	16.856	16.167	18.160
m_1	-0.024	-0.034	0.052
m_2	0.001	-0.026	0.067
σ	0.634	0.565	0.775

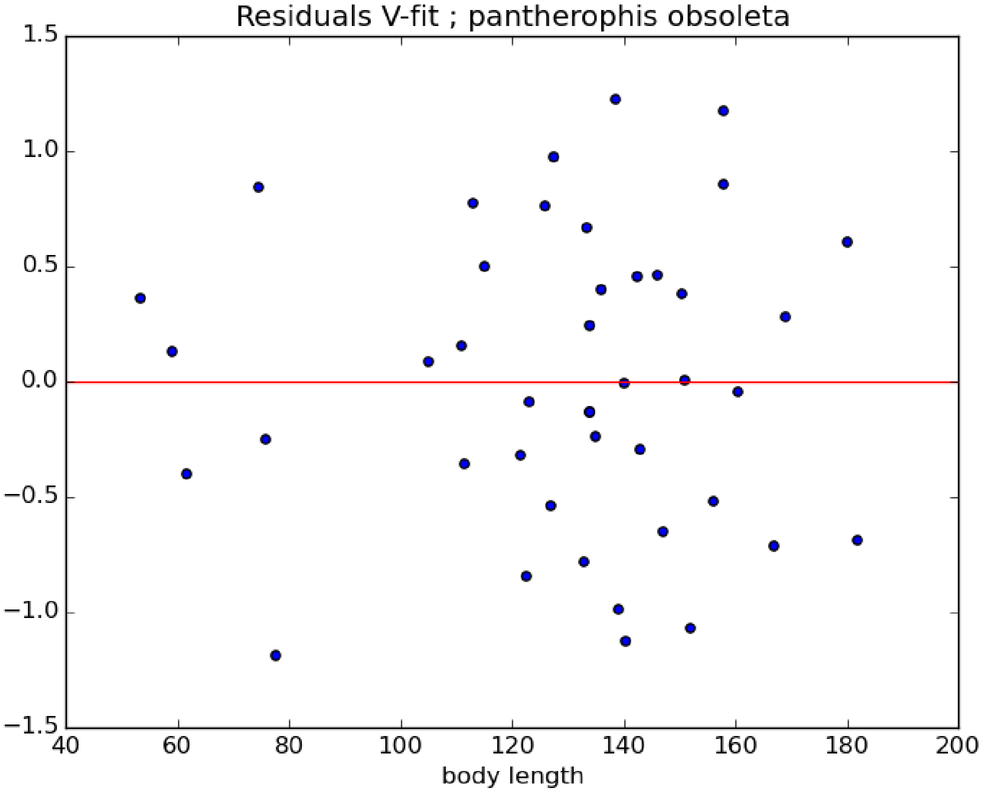
	estimate	lower	upper
u	152.000	152.000	167.000
v	16.856	16.203	17.611
m_1	-0.024	-0.030	0.025
m_2	0.001	-0.022	0.051
σ	0.634	0.560	0.771

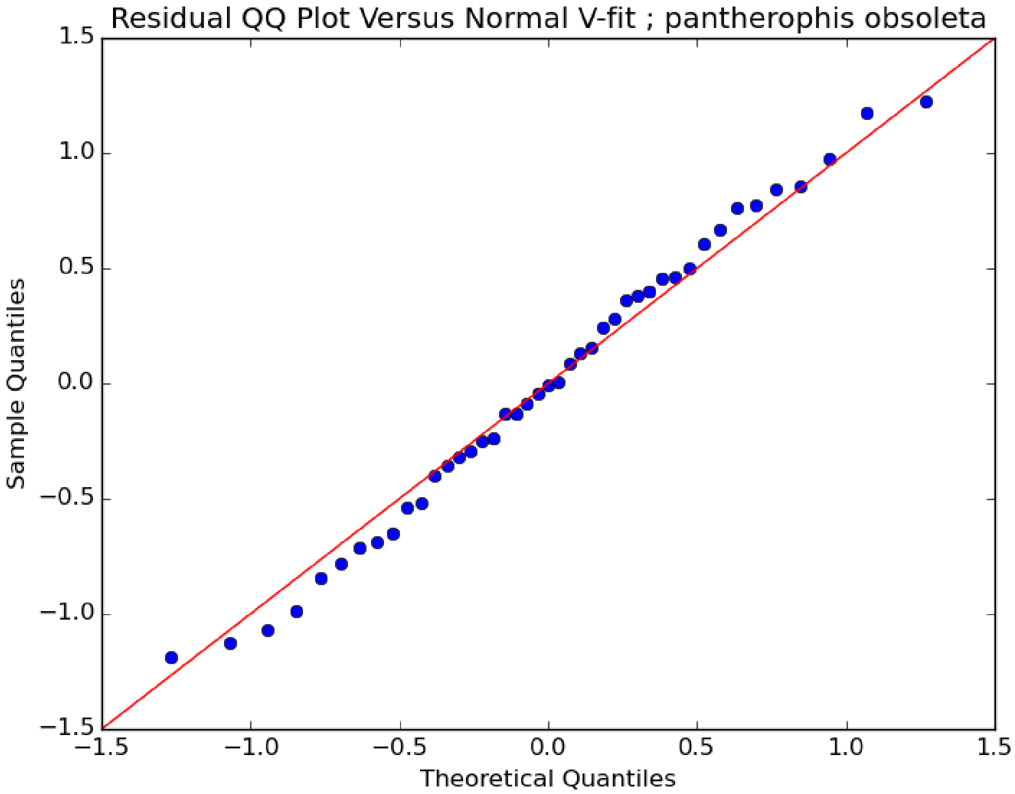
confidence level: 99%

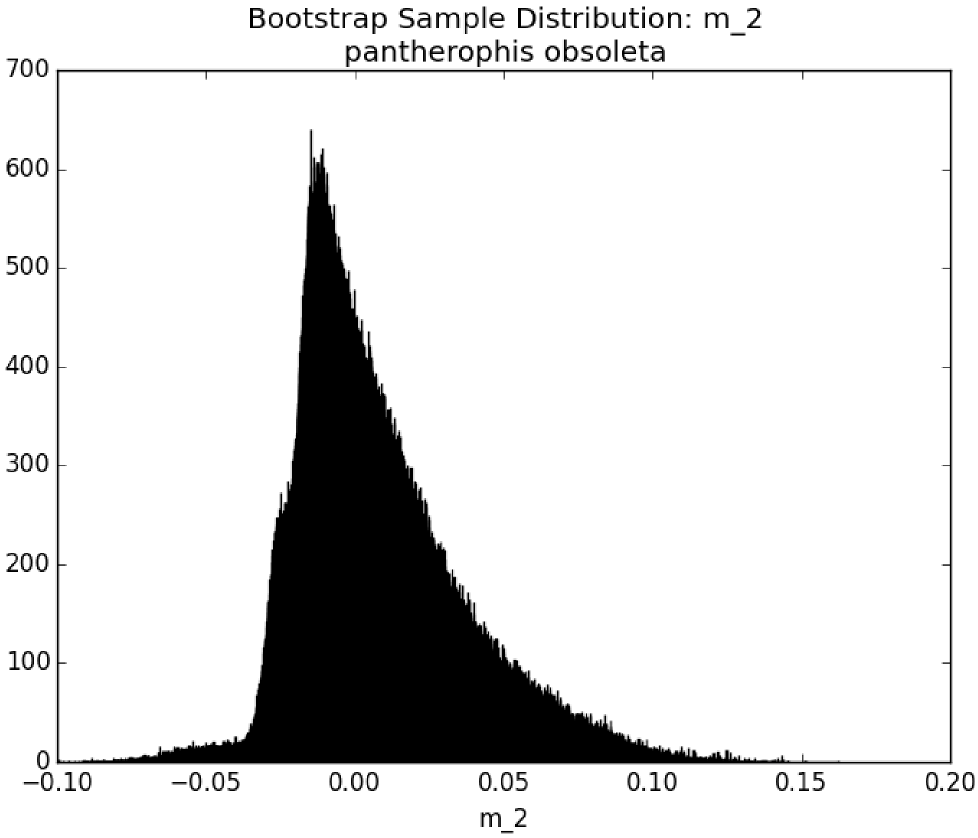
	estimate	lower	upper
u	152.000	77.800	167.000
v	16.856	15.916	18.573
m_1	-0.024	-0.048	0.076
m_2	0.001	-0.032	0.093
σ	0.634	0.534	0.810

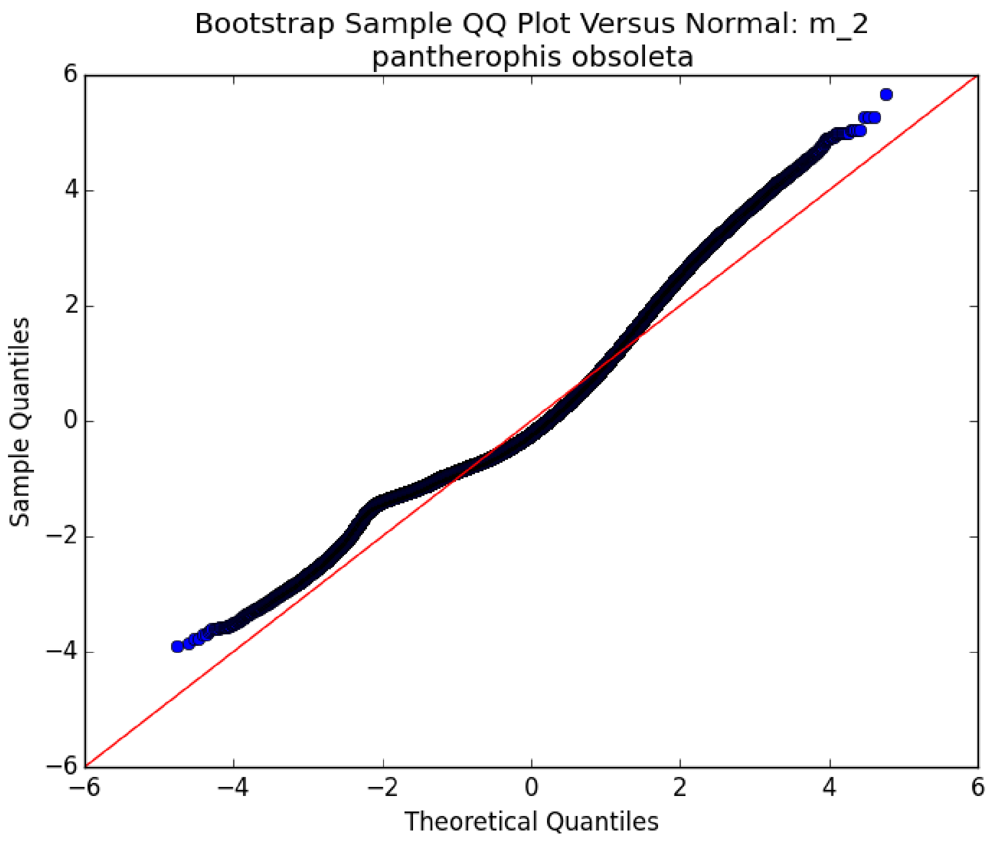
	estimate	lower	upper
u	152.000	147.000	167.000
v	16.856	16.035	18.087
m_1	-0.024	-0.040	0.234
m_2	0.001	-0.031	0.095
σ	0.634	0.528	0.805


V-fit ; pantherophis obsoleta
head-heart/body (%)
body length
20.0
19.5
19.0
18.5
18.0
17.5
17.0
16.5
16.0
15.5
40
60
80
100
120
140
160
180
200


Residuals V-fit ; pantherophis obsoleta
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
40
60
80
100
120
140
160
180
200
body length

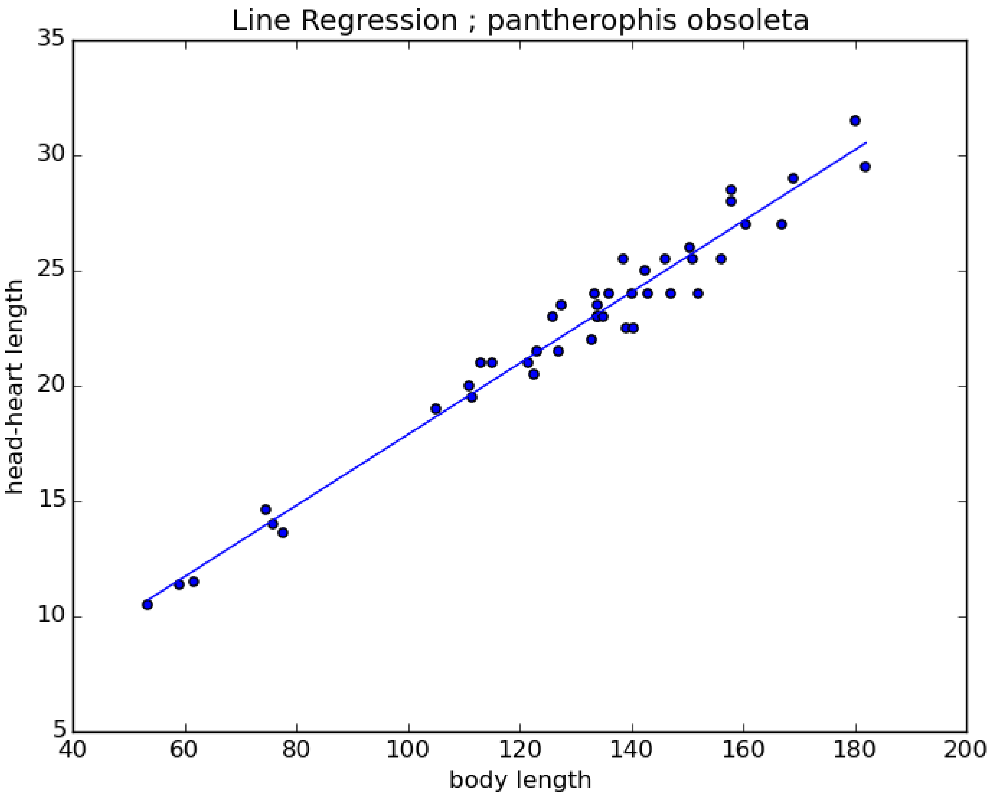

Residual QQ Plot Versus Normal V-fit ; pantherophis obsoleta
Sample Quantiles
Theoretical Quantiles
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

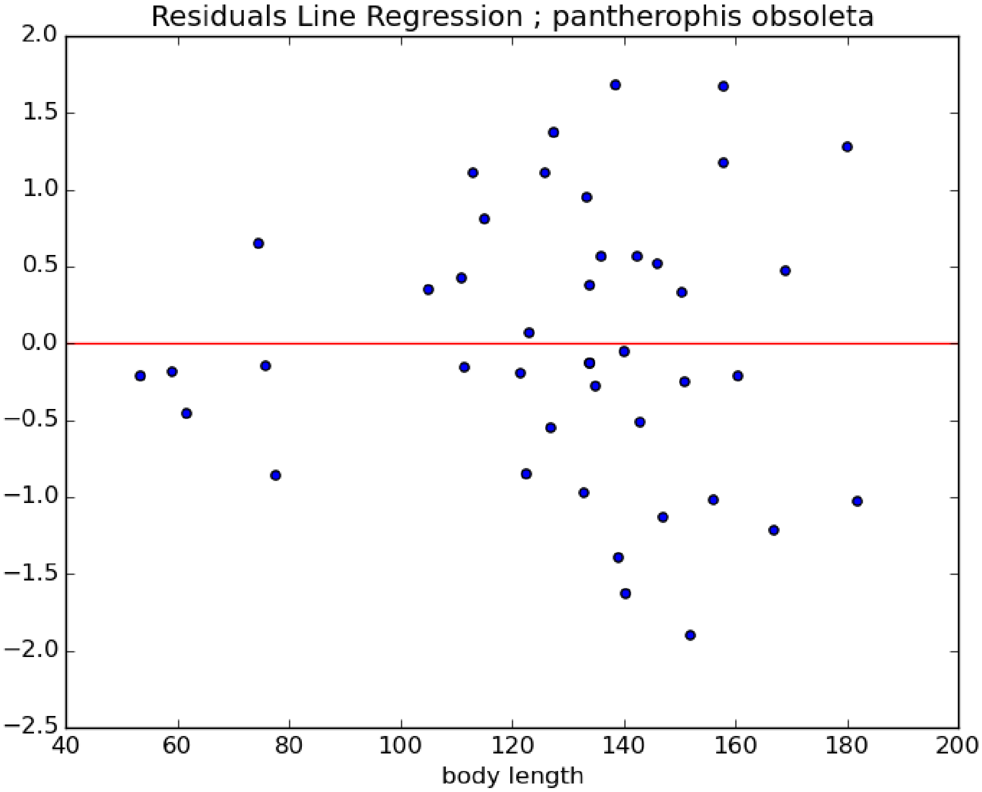

Bootstrap Sample Distribution: m_2
pantherophis obsoleta
700
600
500
400
300
200
100
0
−0.10
−0.05
0.00
0.05
0.10
0.15
0.20
m_2

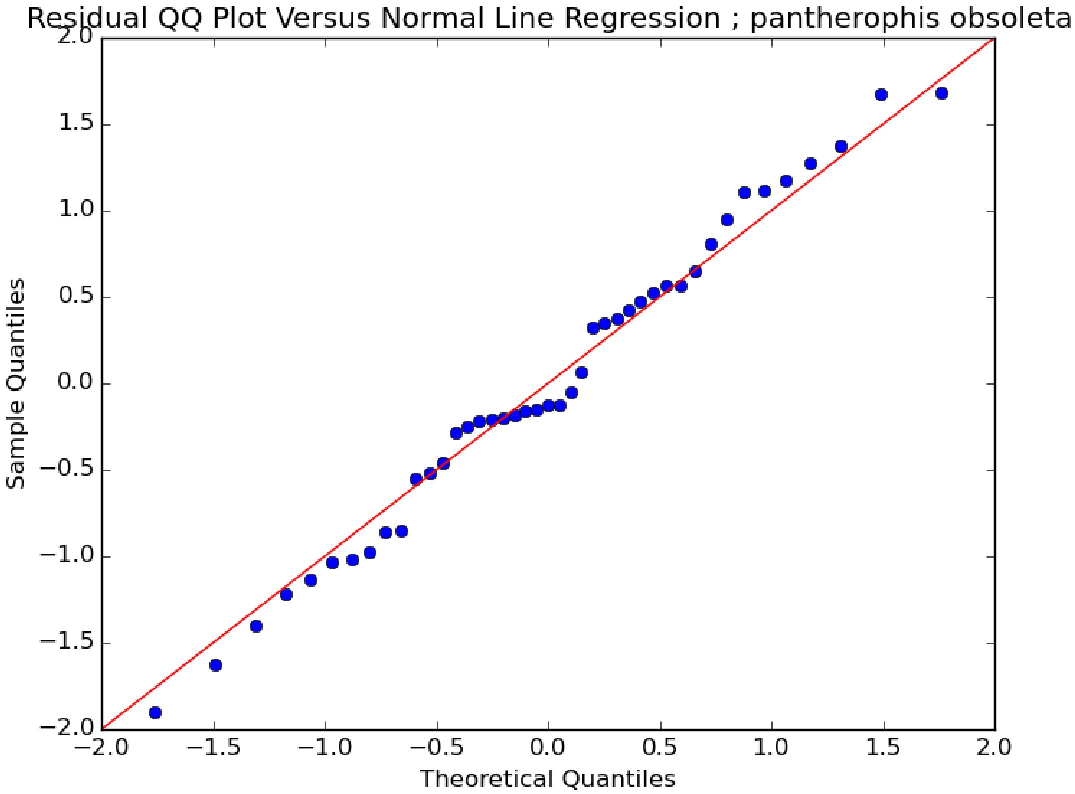


OLS Line Fit

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.966
Model:                            OLS   Adj. R-squared:                  0.966
Method:                 Least Squares   F-statistic:                     1181.
Date:                Thu, 12 Mar 2015   Prob (F-statistic):           7.56e-32
Time:                        21:10:15   Log-Likelihood:                -55.517
No. Observations:                  43   AIC:                             115.0
Df Residuals:                      41   BIC:                             118.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
const          2.4552      0.597      4.112      0.000         1.249     3.661
x1             0.1543      0.004     34.363      0.000         0.145     0.163
==============================================================================
Omnibus:                        0.723   Durbin-Watson:                   2.304
Prob(Omnibus):                  0.696   Jarque-Bera (JB):                0.734
Skew:                          -0.037   Prob(JB):                        0.693
Kurtosis:                       2.364   Cond. No.                         578.
==============================================================================
```

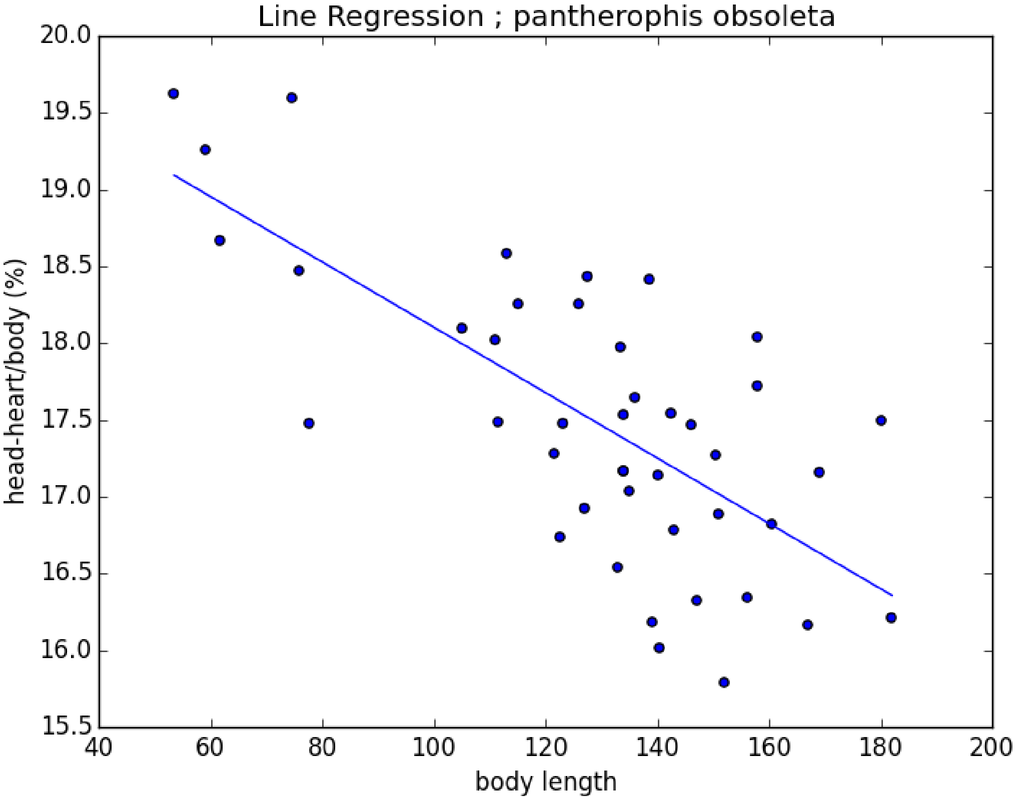

Line Regression ; pantherophis obsoleta
head-heart length
body length
5
10
15
20
25
30
35
40
60
80
100
120
140
160
180
200

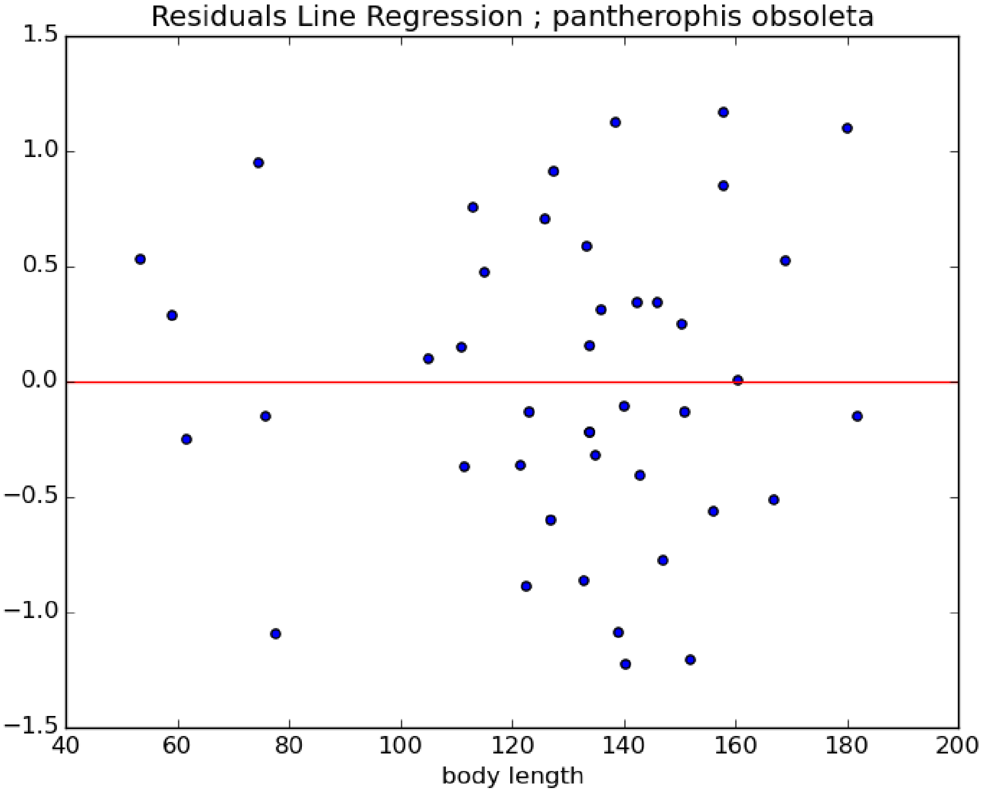

Residuals Line Regression ; pantherophis obsoleta
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
40
60
80
100
120
140
160
180
200
body length

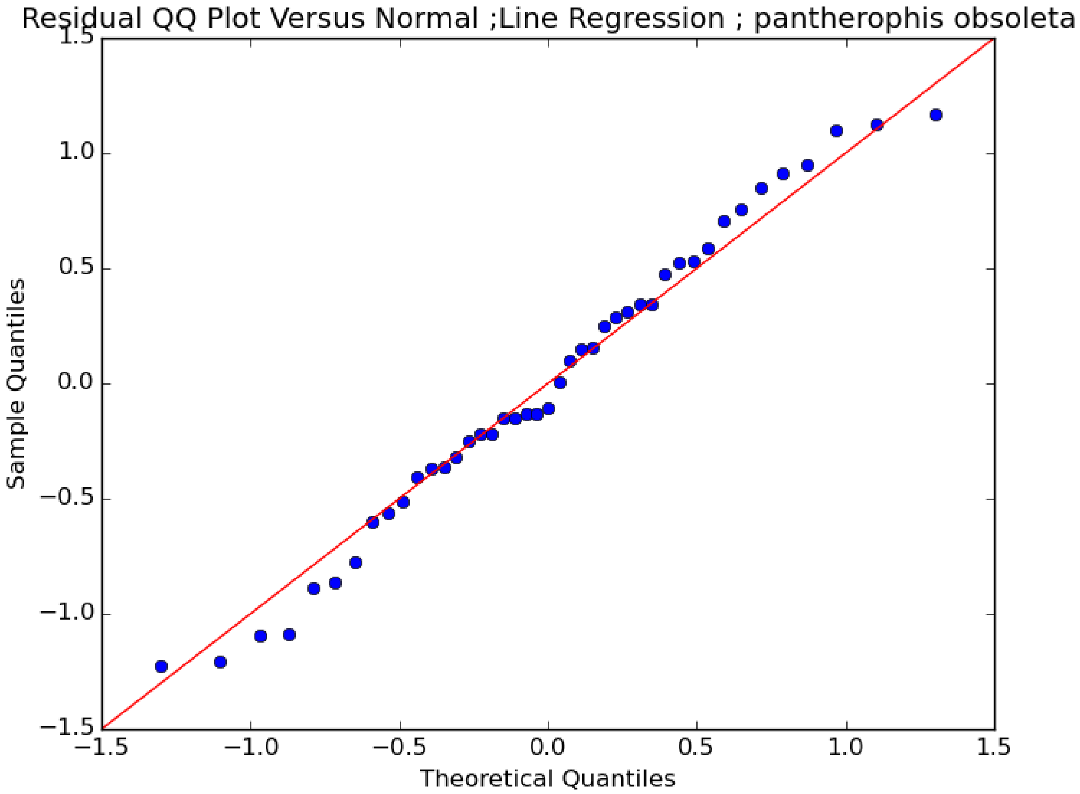


OLS Ratio Line Fit (3.1)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.500
Model:                            OLS   Adj. R-squared:                  0.488
Method:                 Least Squares   F-statistic:                     41.02
Date:                Sun, 15 Mar 2015   Prob (F-statistic):           1.15e-07
Time:                        18:09:04   Log-Likelihood:                -42.579
No. Observations:                  43   AIC:                             89.16
Df Residuals:                      41   BIC:                             92.68
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
const         20.2316      0.442     45.782      0.000        19.339    21.124
x1            -0.0213      0.003     -6.405      0.000        -0.028    -0.015
==============================================================================
Omnibus:                        1.725   Durbin-Watson:                   2.297
Prob(Omnibus):                  0.422   Jarque-Bera (JB):                1.148
Skew:                          -0.054   Prob(JB):                        0.563
Kurtosis:                       2.207   Cond. No.                         578.
==============================================================================
```


Line Regression ; pantherophis obsoleta
head-heart/body (%)
body length
20.0
19.5
19.0
18.5
18.0
17.5
17.0
16.5
16.0
15.5
40
60
80
100
120
140
160
180
200


Residuals Line Regression ; pantherophis obsoleta
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
40
60
80
100
120
140
160
180
200
body length



7 Hypothesis Testing Parameter Restrictions

Several of the V-fits show a right-hand slope which is positive. If such a slope really were positive, it would mean that the relative head-heart position stops decreasing and begins to increase in longer snakes. In this section we explore whether these positive right-hand slopes may be considered statistically significant.

As a first check, we can see whether the bootstrap confidence intervals of Section 6 omit zero.

We also look at likelihood ratio tests. The standard (log)-likelihood ratio test may be applied asymptotically for segmented regressions when the true regression satisfies certain conditions which are specified in [2]. This is the case here where we test the restriction that the right-hand slope is zero. However, we still prefer to use the bootstrap out of concern for inapplicability of asymptotic results for even moderate sample sizes in segmented regressions.

We are actually interested in testing the restriction that the right-hand slope is less than or equal to zero. We do not have all the tools for this, so we just examine the restriction that the slope is zero, and note that the loglikelihood function for the restricted segmented model with fixed right-hand slope as a function of the right-hand slope decreases for values less than zero (not shown).

The procedure we follow is the same as described in Section 5. We also report the standard asymptotic likelihood ratio test p-value for comparison.

7.1 *Agkistrodon conanti*

7.1.1 V-fit Bootstrapped Likelihood Ratio Test for Right Slope Restriction

Our tests in this section are based upon the segmented regression fit in Section 6.2. There we find that the segmented regression, or V-fit, results in a negative left-hand slope and a positive right-hand slope. It is thus of interest to test whether the positive right-hand slope is statistically significant.

First, we note that the bootstrap confidence intervals of Section 6.2 essentially omit zero. If these confidence intervals are reasonably accurate, then we would expect to miss on each side by 2.5% for the 95% confidence intervals. So we have something close to 97.5% confidence of a non-zero right-hand side slope.

We next show results for the likelihood ratio test. Recall that the null hypothesis is that the restricted V-fit with right-hand-slope equal to zero is true.

```
bootstrap num replications: 1000000
bootstrapped p value: 0.0354
asymptotic p value: 0.0239
```

In sum, we can reject a negative right-hand side slope with fairly high confidence.

7.2 *Agkistrodon conanti trunk*

7.2.1 V-fit Bootstrapped Likelihood Ratio Test for Right Slope Restriction

Our tests in this section are based upon the segmented regression fit in Section 6.3. There we find that the segmented regression, or V-fit, results in a negative left-hand slope and a positive right-hand slope. It is thus of interest to test whether the positive right-hand slope is statistically significant.

First, we note that the resample residuals bootstrap confidence intervals of Section 6.3 do not omit zero. However, the resample cases do omit zero, so this is a mixed bag.

We next show results for the likelihood ratio test. Recall that the null hypothesis is that the restricted V-fit with right-hand-slope equal to zero is true.

```
bootstrap num replications: 1000000
bootstrapped p value: 0.1219
asymptotic p value: 0.0709
```

In sum, we cannot reject a negative right-hand side slope.

7.3 *Acrochordus granulatus*

7.3.1 V-fit Bootstrapped Likelihood Ratio Test for Right Slope Restriction

Our tests in this section are based upon the segmented regression fit in Section 6.4. There we find that the segmented regression, or V-fit, results in a negative left-hand slope and a positive right-hand slope. It is thus of interest to test whether the positive right-hand slope is statistically significant.

First, we note that the bootstrap confidence intervals of Section 6.4 essentially omit zero, even at the 99% level.

We next show results for the likelihood ratio test. Recall that the null hypothesis is that the restricted V-fit with right-hand-slope equal to zero is true.

```
bootstrap num replications: 1000000
bootstrapped p value: 0.0009
asymptotic p value: 0.0002
```

In sum, we can reject a negative right-hand side slope with high confidence.

8 Non-Parametric Fits – Descriptive

In this section we consider some non-parametric fits to the data. These are not models we have chosen to use, but we present the results as complementary evidence to the parametric model choices we have made.

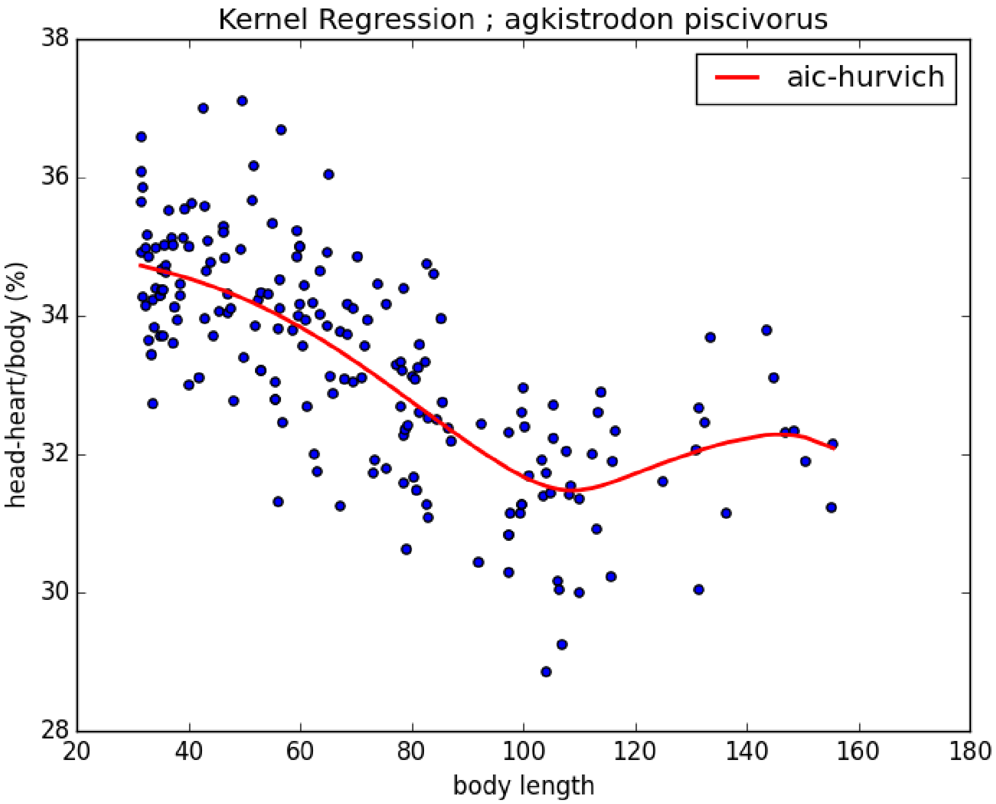
We plot kernel regression fits. For the calculations, we used python StatsModels, with the AIC method of Hurvich et al for bandwidth selection. We also ran these with least-squares cross-validation. In all cases except *Acrochordus granulatus* these fits came out essentially the same. For *Acrochordus granulatus*, the least-squares cross-validation overfit the data at short body lengths where the data is sparse.

8.1 Summary

- The ratio data plot for *Agkistrodon conanti* accords with the V-fit.
- The ratio data plot for *Acrochordus granulatus* accords with the V-fit.
- The ratio data plot for *Hydrophis platurus* is quite close to the V-fit.
- The ratio data plot for *Pantherophis obsoleta* looks very close to a decreasing line. The fit for the original data with head-heart length as the dependent variable looks very much like a line.

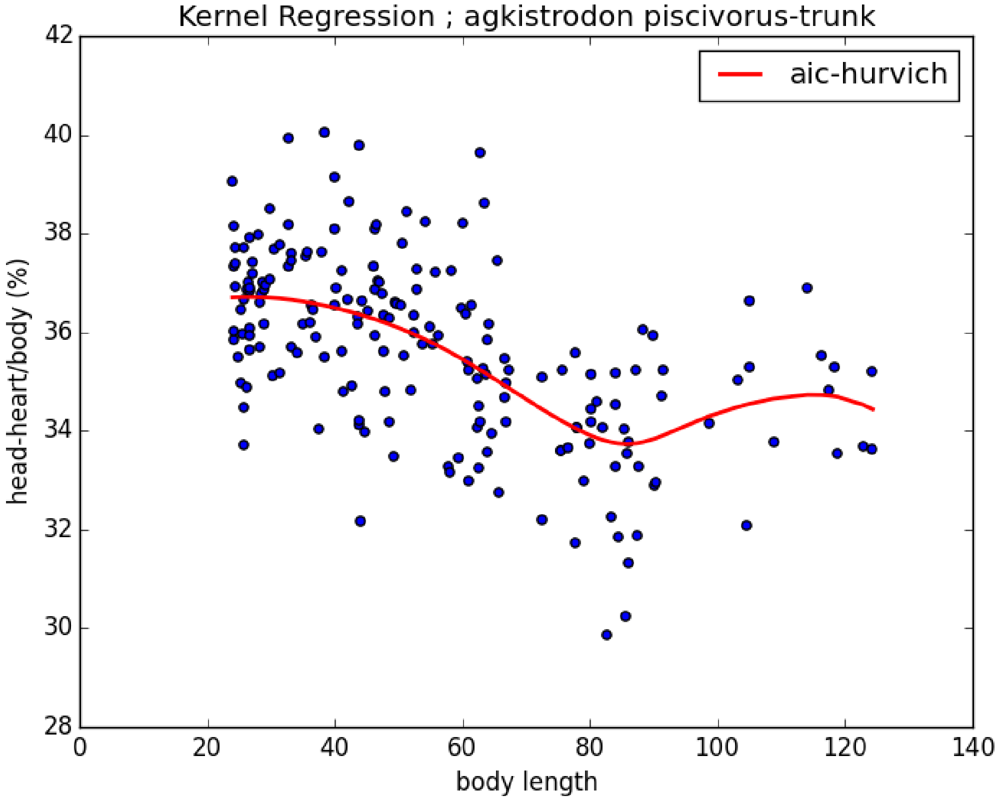
8.2 *Agkistrodon conanti*

We plot the ratio regression (3.1).



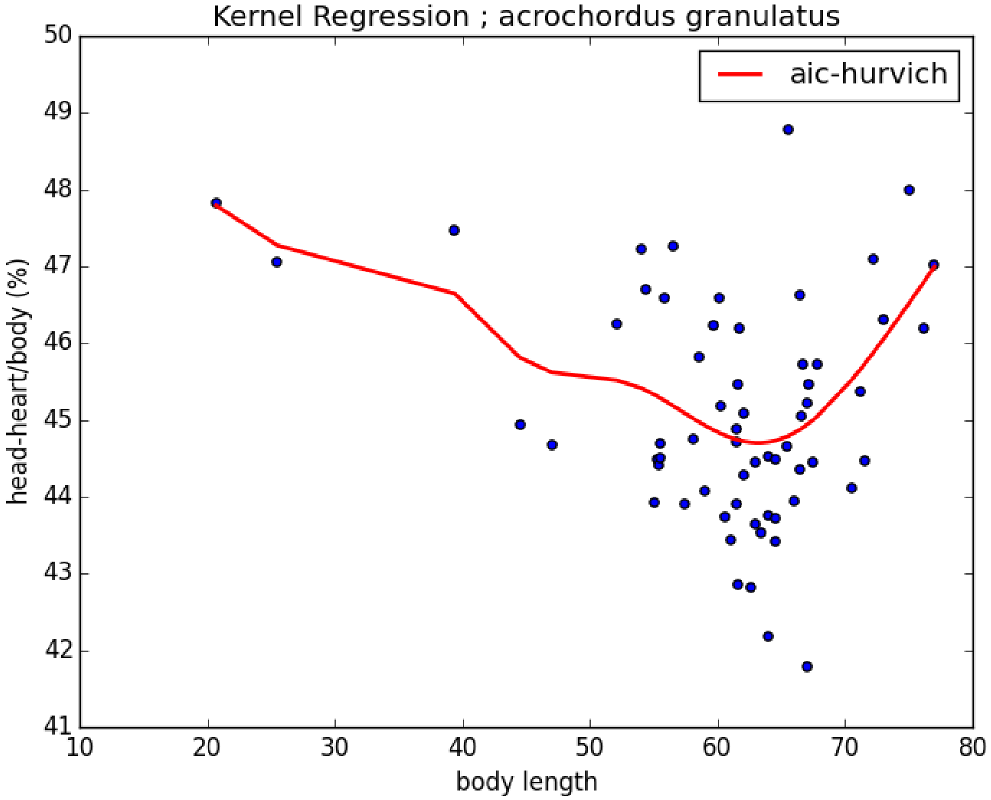
8.3 *Agkistrodon conanti trunk*

We plot the ratio regression (3.1).



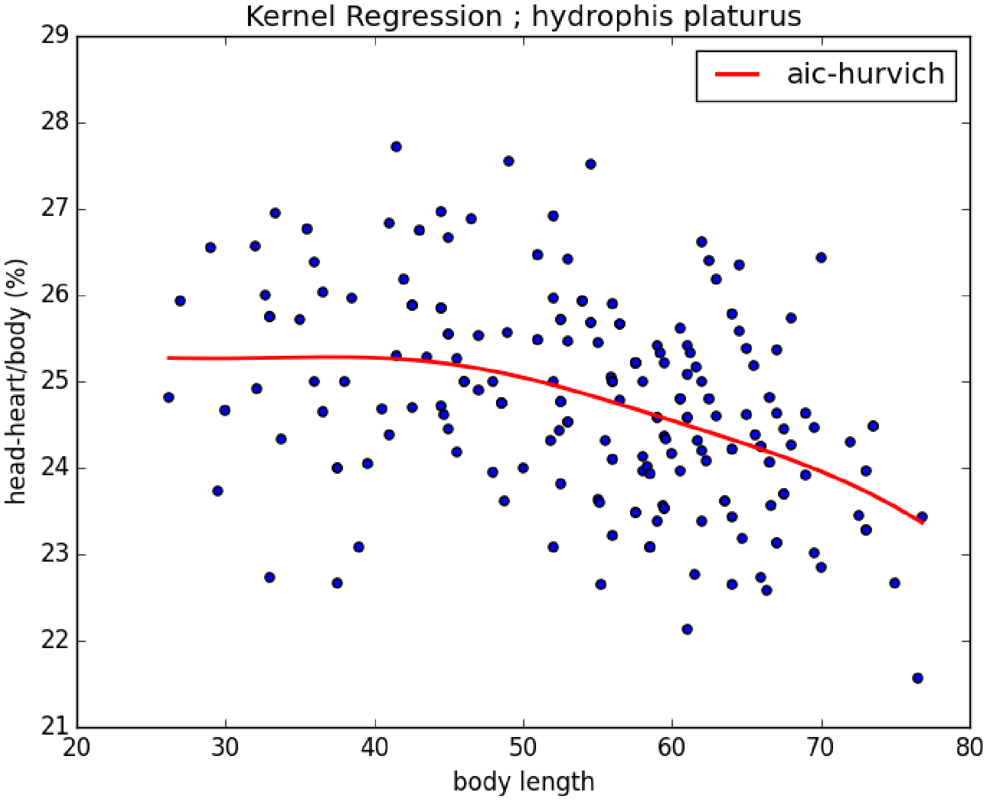
8.4 *Acrochordus granulatus*

We plot the ratio regression (3.1).



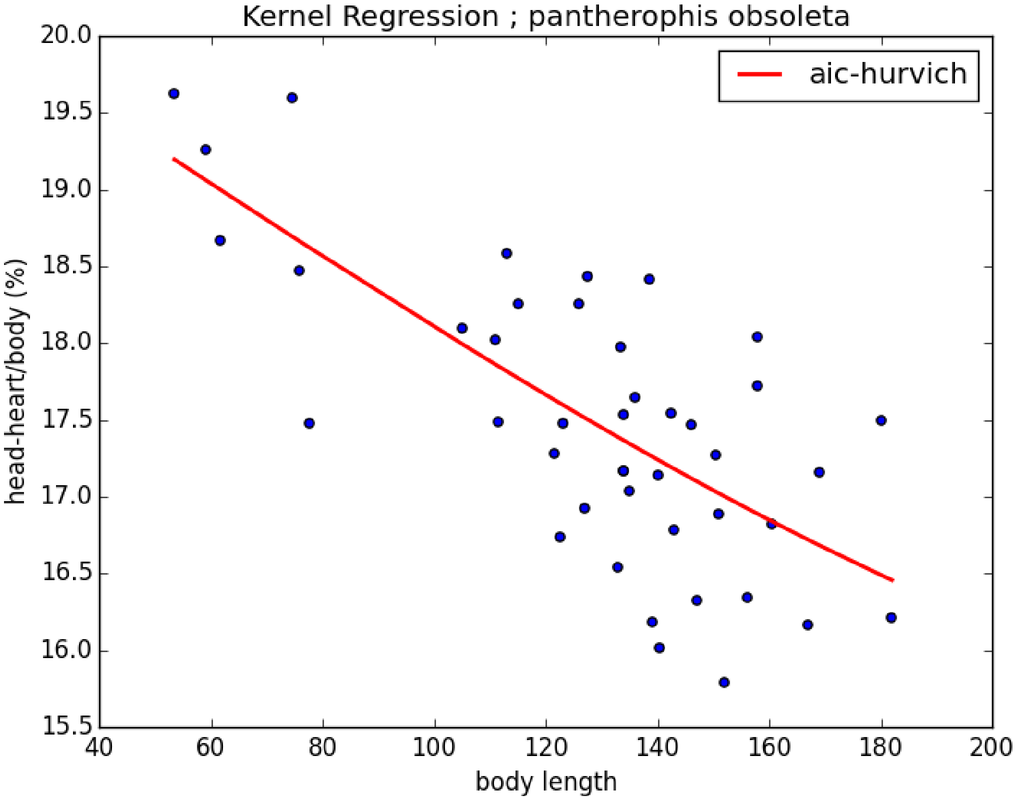
8.5 *Hydrophis platurus*

We plot the ratio regression (3.1).

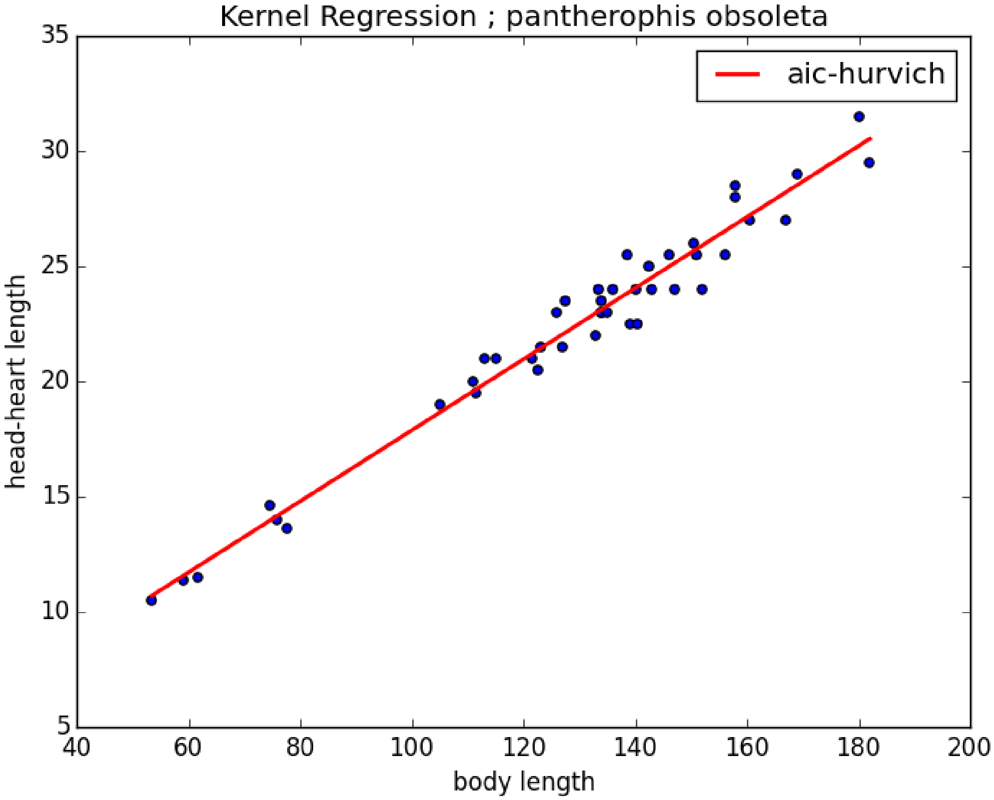


8.6 *Pantherophis obsoleta*

We plot the ratio regression (3.1).



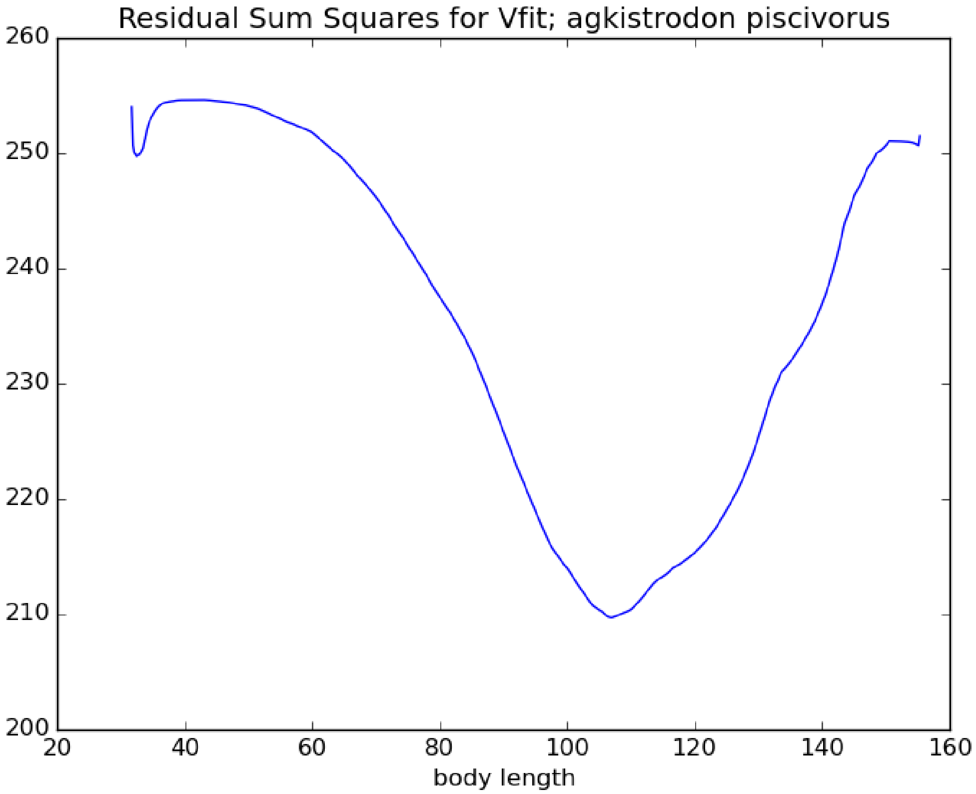
We plot the regression in the original variables.

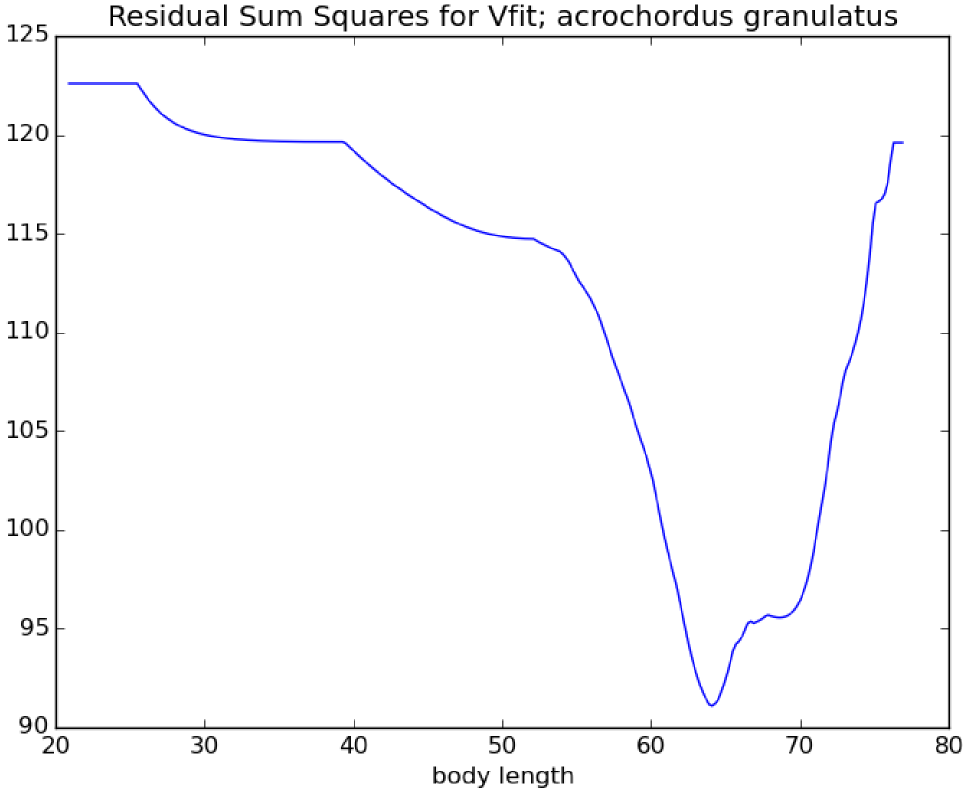


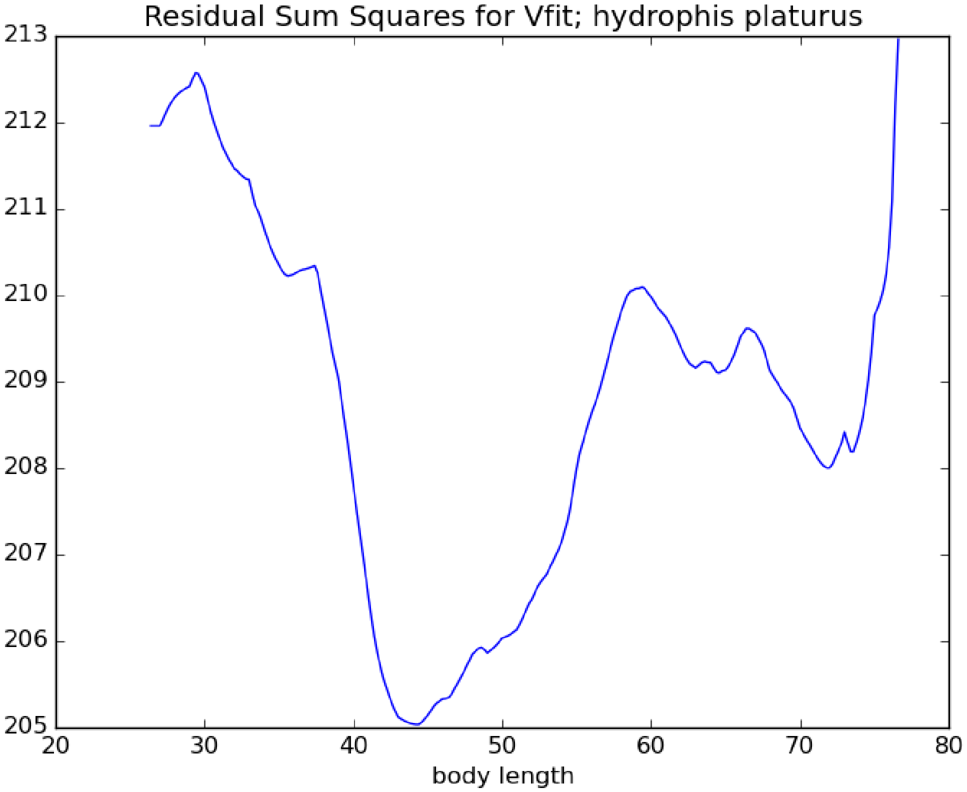
Appendices

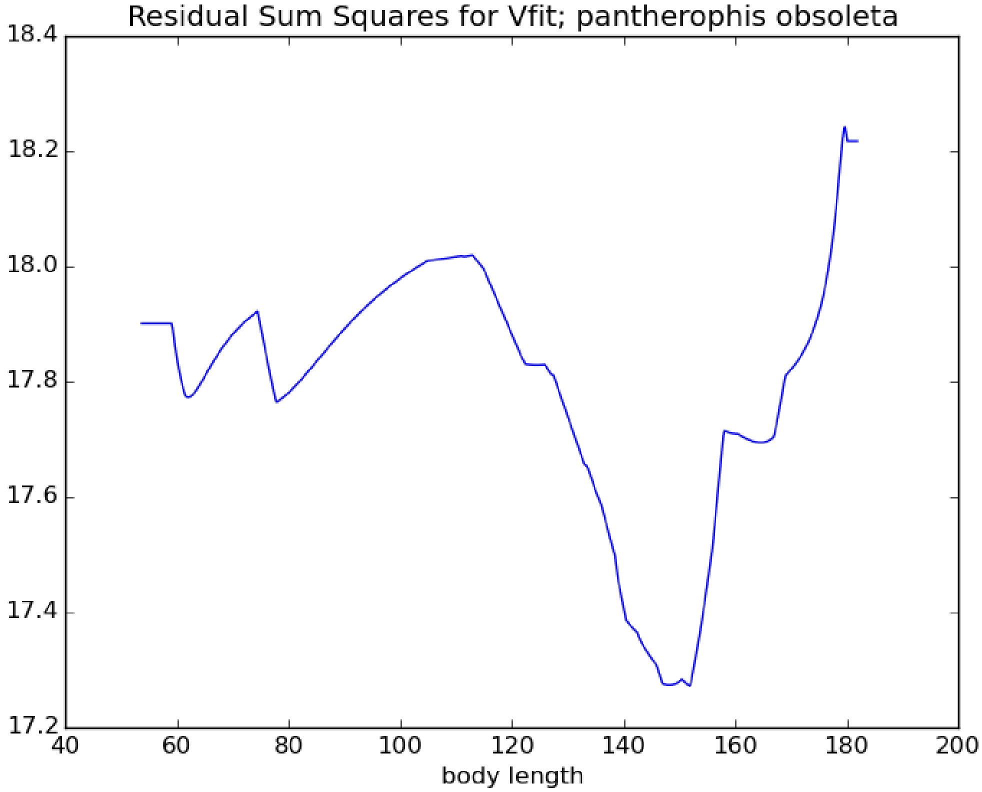
A V-fit Residual Sum-Squares Plots

We plot the function (4.7) for the data sets.


Residual Sum Squares for Vfit; agkistrodon piscivorus
260
250
240
230
220
210
200
20
40
60
80
100
120
140
160
body length


Residual Sum Squares for Vfit; acrochordus granulatus
125
120
115
110
105
100
95
90
20
30
40
50
60
70
80
body length


Residual Sum Squares for Vfit; hydrophis platurus
213
212
211
210
209
208
207
206
205
20
30
40
50
60
70
80
body length



B Software Details

All computation in this report was performed on a Linux Ubuntu 64-bit operating system, using the following software.

Software	**Version**
Python	2.7.11
statsmodels (python)	0.6.1
scipy (python)	0.15.1
R	2.14.1
np (R)	0.60-2

Regressions and ANCOVA analysis presented in the main paper were performed on Windows using Statview SAS 5.0.1.

References

[1] A.C. Davison and D.V. Hinkley. *Bootstrap Methods and their Application.* Cambridge University Press, 1997.

[2] Paul Feder. The log likelihood ratio in segmented regression. *The Annals of Statistics*, 3(1):84–97, 1975.

[3] Paul Feder. On asymptotic distribution theory in segmented regression problems–identified case. *The Annals of Statistics*, 3(1):49–83, 1975.

[4] Z. D. Feng and C. E. McCulloch. Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society*, 58(3):609–617, 1996.

[5] Glenn Firebaugh and Jack P. Gibbs. User's guide to ratio variables. *American Sociological Review*, 50(5):713–722, 1985.

[6] C Hsiao, Q Li, and JS Racine. A consistent model specification test with mixed categorical and continuous data. *Journal of Econometrics*, 140:802–826, 2007.

[7] Derek J. Hudson. Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association*, 61(316):1097–1129, 1966.

[8] Hyune-Jy Kim, Michael P. Fay, Eric J. Feuer, and Douglas N. Midthune. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics In Medicine*, 19:335–351, 2000.

[9] Helmut Küchenhoff. An exact algorithm for estimating breakpoints in segmented generalized linear models. *Computational Statistics*, 12:235–247, 1997.

[10] G. J. McLachlan and S. Rathnayake. On the number of components in a gaussian mixture model. *Wiley Interdisciplinary Reviews (WIREs): Data Mining and Knowledge Discovery*, 4:341–355, 2014.

[11] Cosma Shalizi. *Advanced Data Analysis from an Elementary Point of View.* Cambridge University Press, 2015.

[12] Binbing Yu, Michael J. Barrett, Hyune-Ju Kim, and Eric J. Feuer. Estimating joinpoints in continuous time scale for multiple change-point models. *Computational Statistics and Data Analysis*, 51:2420–2427, 2007.