A Semiparametric Method to Test for Correlated Evolution in a Phylogenetic Context

Luke J. Harmon[1] **and Liam J. Revell**[2]

[1]**Department of Biological Sciences, University of Idaho, Moscow, ID, USA**
[2]**Department of Biology, University of Massachusetts Boston, Boston, MA, USA**

Corresponding author:
Luke J. Harmon[1]

Email address: `lukeh@uidaho.edu`

ABSTRACT

This article describes a new semi-parametric method to test for an evolutionary correlation between continuously valued species traits.

1 INTRODUCTION

Statistical phylogenetic comparative methods are an invaluable tool in modern evolutionary research (Felsenstein 1985; Revell and Harmon 2022). This set of methodologies allows us to test a broad range of evolutionary hypotheses for species-level data, while taking into account non-independence due to their shared histories (reviewed in Harmon 2019; Revell & Harmon 2022). Typically, phylogenetic comparative methods assume a particular underlying model of trait evolution, such as Brownian motion (Felsenstein 1973) for continuously valued traits. Although there are some circumstances under which phenotypic traits might genuinely evolve via a Brownian process (e.g. Lynch 1990), violations of the Brownian model are likely be common in real data sets (Estes and Arnold 2007; Hunt 2006; Uyeda et al. 2011; Freckleton and Harvey 2006). Nonetheless, the Brownian motion model has persisted due to its simplicity and apparent robustness to deviations from model assumptions under some circumstances (Garland et al. 1993; Stone 2011).

There are a few methods available to identify violations of Brownian motion models and then to account for these deviations statistically. First, one can apply various diagnostic tests for independent contrasts that are sensitive to a variety of violations of the Brownian model (Garland 1992; Freckleton and Harvey 2006; Blomberg et al. 2003). For, if the true model is Brownian motion, then standardized independent contrasts of Felsenstein (1985) should be independently and identically distributed (i.i.d). If, on the other hand, standardized contrasts are correlated with their expected variance (the amount of evolutionary branch length that subtends the contrast), this suggests that there is more or less change along long branches of the tree (depending on the sign of the correlation) than expected under Brownian motion. One can then attempt various transformations of the branch lengths of the tree so that the contrasts no longer violate our test (Garland et al., 1992). Another similar diagnostic is the node-height test, in which we check for a relationship between the standardized contrasts and the height of the node over which they are calculated (Freckleton and Harvey 2006). Once again, this correlation should be absent if model assumptions hold. In both instances, one can consider these tests to be measurements of model adequacy, where failure of the test suggests that the underlying assumption of Brownian motion may be invalid (Pennell et al. 2015).

Under circumstances in which the Brownian motion assumption is violated, one can always consider alternative models. In phylogenetic comparative methods,however, our alternative models are typically limited to a few simple alternatives. One can attempt, for example, different evolutionary models that are nonetheless Gaussian, such as the Ornstein-Uhlenbeck or a speciational trait evolution models. Likewise, one can transform the branches of the phylogenetic tree or the elements of the variance-covariance matrix

45   derived from the tree. Nonetheless, there is no general solution and on occasion there may be no way to
46   coax certain data sets to fit the particular assumptions of our analyses.

47   In statistics, nonparametric tests can sometimes provide robustness in cases where models are not
48   known or provide poor fits to data (e.g. Wasserman, 2005). As comparative methods are employed to
49   tackle a wider and wider range of problems, from genomics to cancer biology to evolutionary studies at the
50   broadest scale, there is a growing need for methods that are robust to model violations. Among existing
51   comparative methods, there are nonparametric methods available for the analysis of discrete characters
52   (e.g. Maddison, 1990) and morphometric shape data (Adams 2014). To date, however, the only available
53   nonparametric comparative method designed specifically for continuously distributed trait data is Ackerly's
54   (2000) contrast sign test (CST). This test compares the signs (positive or negative) of phylogenetically
55   independent contrasts for two characters. If the two characters are evolving independently, then the
56   number of pairs of contrasts with the same sign should follow a binomial distribution with probability
57   of success p=0.5. This test provides a robust alternative to parametric comparative methods, but is not
58   commonly used. The main drawback of Ackerly's approach is that it has low power, partially because it
59   uses only the signs, and not the magnitudes, of the contrasts.

60   Here we describe a new semiparametric comparative method, the phylogenetic rank correlation (PRC)
61   test.

62   This test, which is based on the correlation of ranks of independent contrasts on a phylogenetic tree,
63   is robust under several models of trait evolution, and has power comparable to standard independent
64   contrasts and much greater than the CST. The PRC test thus may provide a robust alternative to parametric
65   methods such as PGLS and independent contrasts.

2 DESCRIPTION OF THE METHOD

67   The PRC test has six steps (Fig. 1):

68   1. Calculate standardized independent contrasts (Felsenstein, 1985) for two continuously distributed
69   traits.

70   2. Reflect the contrasts, multiplying both the X and Y value of each contrast by -1. The data will now
71   include two 'copies' of each contrast, one with the signs of each contrast reversed.

72   3. Create ranks independently for each variable. For tied scores, assign the mean rank to all tied
73   contrasts.

74   4. Calculate the test statistic. Here, we use Spearman's rank correlation coefficient, but one could also
75   use Kendall's τ.

76   5. Generate the null distribution of the test statistic via randomization of the original contrasts. To do
77   this, permute the independent contrasts for one variable. Then, double these permuted contrasts,
78   create new ranks, and calculate the test statistic. Repeat this permutation procedure a large number
79   of times (say, 9,999) to generate a null distribution.

80   6. Compare the test statistic with the null distribution to obtain a P-value.

81   Several of these steps require additional comments or explanations.

82   For step one, standardized contrasts are calculated. The calculation of these contrasts, as originally
83   described by Felsenstein (1985), assumes a Brownian motion model of evolution; however, since the
84   standardized contrasts are converted into ranks, violating this assumption has little effect on the analysis
85   (see below). In practice, by converting contrasts into ranks, we assume only that contrasts subtending
86   longer branches of the tree tend to be larger in absolute magnitude. This assumption is compatible with a
87   wide range of evolutionary models as long as variance among species tends to increase through time (see
88   Estes and Arnold 2007).

89   Reflecting the contrasts (step 2) is necessary because the direction of any particular contrast is arbitrary.
90   Reflecting contrasts across the axes ensures that the analysis is unchanged by rotating the nodes in the
91   tree. This procedure is the statistical equivalent of forcing the regression through the origin, which is
92   required when testing for correlations of independent contrasts. We do not have to worry about inflating
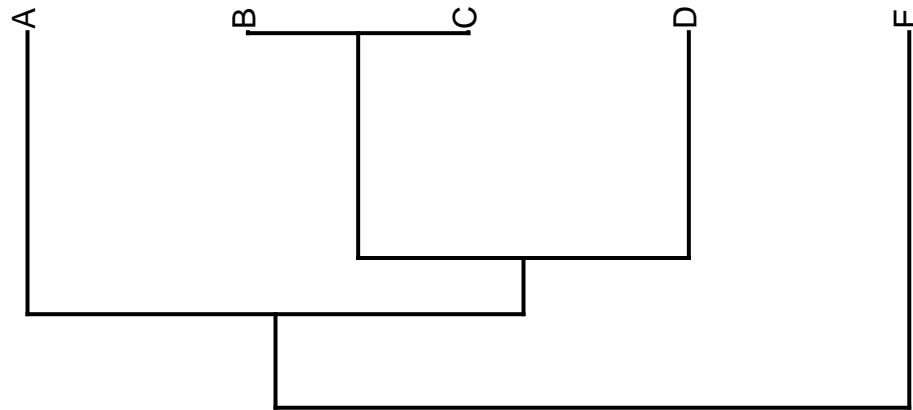
Figure 1. Example figure.

degrees of freedom by doubling the number of data points because statistical significance is determined by permutations.

Finally, in step 4, any nonparametric test statistic could be used for this test (see, e.g. Wasserman, 2005). We used Spearman's rank correlation, which is simply the Pearson product-moment correlation of the ranks of the two variables. Kendall's τ is another viable option.

We provide a worked example of the calculations involved in this test in Figure 2. Software to carry out this test, written in the R statistical computing language (R Development Core Team, 2024), is provided in the Appendix, and will be made publicly available as part of the GEIGER package (Harmon et al., 2007; Pennell et al., 2014).

3 STATISTICAL PROPERTIES OF THE METHOD

We compared the statistical performance of the PRC method with both non-phylogenetic (standard regression) and phylogenetic comparative methods (independent contrasts, IC, and the contrast sign test, CST; see Fig. 2 legend for simulation details). Like independent contrasts and the contrast sign test, the PRC test has appropriate type-I error under a Brownian motion model (Fig. 2a). However, both alternative tests outperform both standard correlation tests and independent contrasts when models of evolution are non-Brownian (Fig. 2b). Independent contrasts fail when there is little phylogenetic structure in the data, as under an Ornstein-Uhlenbeck model with a high α parameter (Fig. 2b). Similarly, standard correlation fails when data has a strong phylogenetic component, as under Brownian motion (Fig. 2b). Both alternative methods have type I error near 5/% under a range of α values. Type I error rates for the rank test are only slightly elevated (7/%) compared to its nominal level.

Likewise, our PRC method has only slightly less power than standard independent contrasts, and considerably more power than Ackerly's (2000) sign test, over both a range of tree sizes and character correlation values (Fig. 2c-d). The PRC test also outperforms other comparative methods when there is error in phylogenetic branch length estimation. To illustrate this, we simulated data on pure-birth phylogenetic trees (as in Figure 1) but then ran analyses after setting all branch lengths in these trees equal. Under these conditions, both PRC and the contrast sign test have appropriate type I error (CST 2.0/%, PRC 5.7/%), while type I error is elevated for independent contrasts (11.4/%). When the expected correlation between characters is 0.5, PRC again has power similar to independent contrasts (IC 74.6/%, PRC 62.9/%). This test also mimics the situation where one knows tree topology, but branch lengths are either unknown or estimated with high error.

We have characterized this method as "semiparametric" because it is not truly distribution-free. One key component of the method involves calculating standardized independent contrasts, which assumes that the characters are evolving under a Brownian motion model. However, by converting these contrasts to ranks, this method greatly lessens the effects of violations of this assumption for testing correlations between characters. In fact, to preserve contrast rank order, one must assume that changes on each branch

Table 1. Example table. See main text for more details.

	log(L)	estimated α	d.f.	AIC	weight
directional one-rate Mk	–	–	–	–	–
ordered two-rate Mk	–	–	–	–	–
ordered ten-rate Mk	–	–	–	–	–

are independent, and that there is more change on longer branches in the tree than along short branches. Violations of this assumption will lead to increased type I error of this test. One example of violations that would have negative impacts on all tests described here (standard regression, independent contrasts, CST, and PST) is rate heterogeneity across the tree (O'Meara et al., 2006). For example, if evolution is faster along certain branches in the tree, nearby contrasts values and signs will be correlated, and all methods relying on their independence will exhibit inflated type I error.

Table 1 shows a summary of our results.

4 CONCLUSIONS

Since the model of evolution underlying continuous characters is frequently unknown, and may depart from Brownian motion,we recommend using this method to test correlations between characters. If standard contrasts are well-behaved, then standard techniques are still useful for estimating slopes and other parameter values. However, this semiparametric test provides a method for generating robust P-values across a range of models. The technique is nearly as powerful as independent contrasts but much more robust to deviations from the underlying model of evolution.

5 DATA AVAILABILITY

This article was written in Rmarkdown (Xie et al. 2018, 2020; Allaire et al. 2023), and developed with the help of both *bookdown* (Xie 2016, 2023) and the posit Rstudio IDE (RStudio Team 2020). All data and markdown code necessary to exactly rebuild the submitted version of this article (including its analyses and figures) are available at `https://github.com/liamrevell/Revell-and-Harmon.NonParametricPCM`.

6 ACKNOWLEDGMENTS

LITERATURE CITED

Allaire, J., Y. Xie, J. McPherson, J. Luraschi, K. Ushey, A. Atkins, H. Wickham, J. Cheng, W. Chang, and R. Iannone. 2023. Rmarkdown: Dynamic documents for R.

RStudio Team. 2020. RStudio: Integrated development environment for R. RStudio, PBC., Boston, MA.

Xie, Y. 2023. Bookdown: Authoring books and technical documents with r markdown.

Xie, Y. 2016. Bookdown: Authoring books and technical documents with R markdown. Chapman; Hall/CRC, Boca Raton, Florida.

Xie, Y., J. J. Allaire, and G. Grolemund. 2018. R markdown: The definitive guide. Chapman; Hall/CRC, Boca Raton, Florida.

Xie, Y., C. Dervieux, and E. Riederer. 2020. R markdown cookbook. Chapman; Hall/CRC, Boca Raton, Florida.