

# Dependent variable selection in phylogenetic generalized least squares regression analysis under Pagel's lambda model

Zheng-Lin Chen<sup>1</sup>, Hong-Ji Guo<sup>1</sup>, and Deng-Ke Niu<sup>1</sup>

<sup>1</sup>Beijing Normal University

June 16, 2023

## Abstract

Phylogenetic generalized least squares (PGLS) regression is widely used to detect evolutionary correlations. In contrast to the equal treatment of analyzed traits in conventional correlation methods such as Pearson and Spearman's rank tests, we must designate one trait as the independent variable and the other as the dependent variable. However, in our PGLS regression analyses (using Pagel's  $\lambda$  model) of both empirical and simulated datasets, switching independent and dependent variables yielded many conflicting results. A serious problem with PGLS regression that has not been noticed before is that selecting an inappropriate trait as the dependent variable will often result in an error. To assess correlations in simulated data, we established a gold standard by analyzing changes in traits along phylogenetic branches. Next, we tested seven potential criteria for dependent variable selection: log-likelihood, Akaike information criterion,  $R^2$ ,  $p$ -value, Pagel's  $\lambda$ , Blomberg et al.'s  $K$ , and the estimated  $\lambda$  in Pagel's  $\lambda$  model. We determined that the last three criteria performed equally well in selecting the dependent variable and were superior to the other four. For practicality, we suggest using the trait with a higher  $\lambda$  or  $K$  value as the dependent variable in future PGLS regressions. In analyzing the evolutionary relationship between two traits, we should designate the trait with a stronger phylogenetic signal as the dependent variable even if it could logically assume the cause in the relationship.

## Hosted file

0616 response variable.docx available at <https://authorea.com/users/629715/articles/649826-dependent-variable-selection-in-phylogenetic-generalized-least-squares-regression-analysis-under-pagel-s-lambda-model>