Ecological Complexity 32 (2017) 121–130

Contents lists available at ScienceDirect

Ecological Complexity

journal homepage: www.elsevier.com/locate/ecocom

Original Research Article

Comparison of five methods for parameter estimation under Taylor's power law

Peijian Shi[a], David A. Ratkowsky[b], Ningtao Wang[c,*], Yang Li[d,*], Lei Zhao[e], Gadi V.P. Reddy[f], Bai-Lian Li[g]

[a] Co-Innovation Centre for Sustainable Forestry in Southern China, Bamboo Research Institute, Nanjing Forest University, Nanjing 210037, China
[b] Tasmanian Institute of Agriculture, University of Tasmania, Private Bag 98, Hobart, Tasmania 7001, Australia
[c] Department of Biostatistics, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA
[d] Department of Mathematics and Statistics, University of Minnesota Duluth, Duluth, MN 55804, USA
[e] Department of Ecology and Evolutionary Biology and Kansas Biological Survey, University of Kansas, Lawrence, KS 66047, USA
[f] Western Triangle Agricultural Research Centre, Montana State University, Conrad, MT 59425, USA
[g] Ecological Complexity and Modeling Laboratory, Department of Botany and Plant Sciences, University of California, Riverside, CA 92521-0124, USA

ARTICLE INFO

Article history:
Received 13 June 2017
Received in revised form 7 September 2017
Accepted 25 October 2017
Available online xxx

Keywords:
Measurement error
Nonlinear regression
Statistical modeling
Taylor's power law

ABSTRACT

Taylor's power law (TPL) can be applied to the mean-variance relationship for various quantities, e.g., population densities over space and time, biomass of plants in the growth processes, developmental rates and growth rates of arthropods at different temperatures, etc. When TPL holds in the grouped data (e.g., the biomass of plants at different investigation times with limited replicates at each investigation), we must consider the heterocedasticity when carrying out the parameter estimation for a given mathematical model. We propose two new target functions based on TPL and attempt to more accurately estimate the model parameter(s). By capturing the mean-variance relationship, the accuracy of parameter estimation and the power of hypothesis testing are improved compared to the traditional methods which assume homogeneous variance. Five approaches, including ordinary least squares (OLS), chi-squared (χ^2), weighted least squares (WLS), psi-squared (ψ^2) and maximum likelihood estimation (MLE), are compared under both linear and nonlinear scenarios regarding the prediction accuracy using both computer simulations and experimental data. We further simulate irregular measurements on the predictors to examine the robustness of parameter estimation. In computer simulations, psi-squared, WLS, and MLE outperform OLS and chi-squared with respect to parameter estimation. In the simulation study of irregular measurements, psi-squared and MLE outperform WLS when the sample mean is smaller than the population mean. On the other hand, WLS outperforms psi-squared and MLE when the sample mean is larger than the population mean. Psi-squared has an advantage over MLE when large irregular deviations are present. This study strongly suggests using psi-squared and WLS in place of OLS or chi-squared for both linear and nonlinear regressions, the choice of method depending on the observations.

1. Introduction

Taylor (1961) studied the mean (M) and variance (V) of spatial densities of insects and found an empirical formula $V = \alpha M^{\beta}$ with constant parameters α and β, which has been referred to in the scientific literature as Taylor's power law (TPL). It is essentially a typical power law, and similar relationships were discovered independently (e.g., Thompson, 1917; Newman, 2005; Imre and Novotný, 2016; Cohen et al., 2016), some of them even before Taylor's work (e.g., Smith, 1938). Anderson et al. (1982) further confirmed that TPL holds in describing the temporal dynamics of population densities. Great attention has been paid to estimate the exponent (β) in TPL, which usually ranges from 1 to 2 for various population densities (Anderson et al., 1982; Sawyer, 1989; Kilpatrick and Ives, 2003; Ballantyne, 2005; Ballantyne and Kerkhoff, 2007; Eisler et al., 2008; Cohen et al., 2016; Cohen and Xu, 2015; Giometto et al., 2015; Shi et al., 2016, 2017). Besides the population densities dynamics, TPL is also applicable to the mean-variance relationship of crop biomasses at different time points

* Corresponding authors.
E-mail addresses: ningtao.wang@uth.tmc.edu (N. Wang), yangli@e.unm.edu (Y. Li).

https://doi.org/10.1016/j.ecocom.2017.10.006
1476-945X/

during the growth process (Shi et al., 2017) as well as bamboo biomasses at different internodes (Cheng et al., 2017). In fact, TPL has found application in many abiotic areas (Eisler et al., 2008; Fronczak and Fronczak, 2010; Xiao et al., 2015) and is considered as a universal statistical law as a consequence of the random sampling of skewed distributions (Cohen and Xu, 2015).

When carrying out a nonlinear regression (e.g., fitting the three-parameter logistic model; see Ratkowsky, 1983, 1990), it is customary to assume that the variances of the response variable at different predictors' levels are homogeneous, so that the variance of the response variable is assumed to be constant regardless of whether the predictors are large or small (Bates and Watts, 1988). On the other hand, TPL states that the variance is a function of the mean, so that the larger the x value is, the larger will be the variance of y (Cohen and Xu, 2015). In experimental population ecology, we frequently use nonlinear regression to describe the complex relationships between the response variable and one or more independent variable(s). Since TPL holds widely in ecology (i.e., ecological data are often heteroscedastic), we need to examine whether the heteroscedasticity can largely affect parameter estimation in a nonlinear regression when the traditional method of ordinary least squares is used (Bates and Watts, 1988; Ratkowsky, 1983).

In the present study, we propose a new method based on TPL for parameter estimation and compare it with four traditional methods to check which models are better than the remaining candidates in reducing the prediction errors of model parameters. Three nonlinear models with known parameters are used to simulate the data for comparing the results of parameter estimations using five methods.

2. Materials and methods

2.1. Five methods for parameter estimation

Assume q groups of observations with each group representing a different level of x. For example, considering the dry weight of plants at q investigation times during the growth season, n_i (i = 1, 2, 3, . . . , q) plants are weighed at each time for a total of $n = \sum_i^q n_i$ plants. A linear or nonlinear growth function $f(x;\ \boldsymbol{\theta})$ can be formulated, where $\theta = (\theta_1, \theta_2, \theta_3, \ldots, \theta_p)$ represents p parameters of the growth function (e.g., in a simple linear regression $y = a + bx$, $p = 2$). Let $\mu_i = f(x_i;\ \theta)$, where x_i represents the predictor observations in the i-th group. Here we consider five methods to estimate $\boldsymbol{\theta}$.

(i) Ordinary least squares (OLS)

$$\hat{\boldsymbol{\theta}} = \underset{\theta}{\text{argmin}} \sum_{i=1}^{q} \sum_{j=1}^{n_i} y_{ij} - \mu_i)^2 \tag{1}$$

where y_{ij} represents the dry weight of the j-th plant (j = 1, 2, 3, . . . , n_i) in the i-th group.

(ii) Chi-squared (χ^2)

$$\boldsymbol{\theta} = \underset{\theta}{\text{argmin}}\ \chi^2 = \underset{\theta}{\text{argmin}} \sum_{i=1}^{q} \sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_i)^2}{|\mu_i|} \tag{2}$$

(iii) Psi-squared (ψ^2)

$$\boldsymbol{\theta} = \underset{\theta}{\text{argmin}}\ \psi^2$$

$$= \underset{\theta}{\text{argmin}} \sum_{i=1}^{q} \left[\left(\frac{\mu_i}{\sqrt[q]{\prod_{k=1}^{q} \mu_k})^{\beta}} \frac{\frac{\sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2}{n_i - \frac{n_i}{n} p}}{} \right] \tag{3}$$

where β is the exponent of TPL that can be estimated in practice by the linear regression between ln(mean) and ln(variance) of y_{ij}. See Appendix A1 in Supplementary data for the derivation of Eq. (3).

(iv) Weighted least squares (WLS)

$$\boldsymbol{\theta} = \underset{\theta}{\text{argmin}} \sum_{i=1}^{q} [w_i \cdot \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2] \tag{4}$$

where $w_i = \frac{1}{\sum_{j=1}^{n_i} \frac{(y_{ij} - \bar{y}_i)^2}{n_i - 1}}$, and $\bar{y}_i$ represents the mean of the i-th group of response variables. In fact, chi-squared can be regarded as a special case of WLS. Here we use the most commonly adopted WLS with weights equal to the reciprocal of the group-wise variances.

(v) Maximum likelihood estimation (MLE) with TPL

$$\hat{\boldsymbol{\theta}} = \underset{\theta}{\text{argmax}}\ \ln(L)$$

$$= \underset{\theta}{\text{argmax}} \left\{ - \left[\frac{\frac{n}{2} ln(2\pi\alpha) + \frac{1}{2}\beta \sum_{i=1}^{q} n_i ln(\mu_i) + \sum_{i=1}^{q} \sum_{j=1}^{n_i} y_{ij} - \mu_i)^2}{2\alpha {\mu_i}^{\beta}} \right] \right\} \tag{5}$$

where L represents the likelihood. See Appendix A1 in Supplementary data for the derivation of Eq. (5).

2.2. Examples of parameter estimation

To evaluate the accuracies of the parameter estimates obtained using the five methods, three mathematical models, i.e., $f(x;\ \boldsymbol{\theta})$, were implemented (Fig. S1-S3). These models were used because the TPL in the grouped data sets had been found to hold. However, it does not mean that the above five methods in Section 2.1 for parameter estimation only apply to the following three models. If TPL holds in other grouped data sets that use other ecological models, the five methods can also be employed to implement parameter estimation. The three models used here have been reported in the relevant references (see below for details).

(i) The temperature effect on the developmental rates of insects in the mid-temperature range has been formulated by using the simple linear model

$$r_d = a + bT \tag{6}$$

where r_d is the developmental rate (response variable) and T is the temperature (independent variable) (Uvarov, 1931; Campbell et al., 1974); a and b are parameters to be estimated.

(ii) The temperature effect on the growth rates of poikilotherms in the whole thermal range has been formulated by using right-skewed bell-shaped curves (Sharpe and DeMichele, 1977; Ratkowsky et al., 2005; Deutsch et al., 2008; Ratkowsky and Reddy, 2017). The Lobry–Rosso–Flandrois (LRF) model,

$$r_g = \left\{ \frac{r_{\text{opt}}(T - T_{max})T - T_{min})^2}{(T_{\text{opt}} - T_{min})\left[(T_{\text{opt}} - T_{min})(T - T_{\text{opt}}) - (T_{\text{opt}} - T_{max})(T_{\text{opt}} + T_{min} - 2T)\right] \text{if } T \in (T_{min}, T_{max}) 0 \text{if } T \notin (T_{min}, T_{max})} \right., \tag{7}$$