Fisheries Research 147 (2013) 137–144

Contents lists available at SciVerse ScienceDirect

Fisheries Research

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

Growth estimates of cardinalfish (*Epigonus crassicaudus*) based on scale mixtures of skew-normal distributions

Javier E. Contreras-Reyes[a,*], Reinaldo B. Arellano-Valle[b]

[a] División de Investigación Pesquera, Instituto de Fomento Pesquero, Av. Almirante Blanco Encalada 839, 2361827 Valparaíso, Chile
[b] Departamento de Estadística, Pontificia Universidad Católica de Chile, 7820436 Santiago, Chile

ARTICLE INFO

Article history:
Received 28 December 2012
Received in revised form 3 May 2013
Accepted 4 May 2013

Keywords:
Von Bertalanffy model
Age–length
Skew-*t*
Cardinalfish
Heteroskedasticity
Local influence

ABSTRACT

Our article presents a robust and flexible statistical model of the age–length relationship of cardinalfish (*Epigonus crassicaudus*). Specifically, we consider a non-linear regression model in which the error distribution allows for heteroskedasticity and belongs to the skew-normal (SMSN) distributions family of scale mixtures, thus eliminating the need to transform the dependent variable using techniques such as the Box–Cox transformation. The SMSN is a tractable and flexible class of asymmetric, heavy-tailed distributions that is useful for robust inference when the normality assumption for the error distribution is questionable. Two well-known important members of this class are the proper skew-normal and skew-*t* distributions. In this work, the skew-*t* model is emphasised. However, the proposed methodology can be adapted for each of the SMSN models with some basic changes. The present work is motivated by a previous analysis of cardinalfish where the oldest specimen was 15 years of age. In this study, we use the proposed methodology on a data set based on an otolith sample where the determined longevity is higher than 54 years.

1. Introduction

Interest in describing the growth of biological species is increasing. This interest has motivated the use of biological models to describe growth in terms of an age–length relationship. The von Bertalannfy (VB) growth curve is a notable model (von Bertalanffy, 1938; Kimura, 1980). This model explains the length of a species in terms of its age by means of a non-linear function depending on three parameters: L_∞, K, and t_0. L_∞ represents the asymptotic length of the species under study; K is the growth rate, also known as the Brody growth rate coefficient (Brody, 1945); and t_0 is the theoretical age at length zero. Specifically, if $y(x)$ represents the observed length at age x, then a deterministic expression of the VB growth curve is given by

$$y(x) = L_\infty(1 - e^{-K(x-t_0)}). \quad (1)$$

To fit Eq. (1) using an empirical dataset, (y_i, x_i), $i = 1, \ldots, n$, where y_i (length) and x_i (age) are the response and explanatory variables for the ith sample subject, respectively, the VB growth curve can be described as a non-linear regression

$$y_i = \eta_i + \varepsilon_i, \quad (2)$$

where $\eta_i = \eta(\beta\,; x_i) = L_\infty(1 - e^{-K(x_i-t_0)})$, $\beta = (L_\infty, K, t_0)^T$ is the vector of unknown parameters and the ε_i are independent, random errors.

Kimura (1980) studied Eq. (2) under the assumptions of independence and normality for the random errors models, $\varepsilon_i \sim N(0, \sigma^2)$, and proposed the maximum likelihood method to fit the model (see also Allen, 1966). Other models include normal errors with a constant coefficient of variation (CV) (Candy et al., 2007) and multiplicative log-normal errors (Millar, 2002). The robustness of the log-normal distributional assumption was tested by Wang and Ellis (1998) (gamma and truncated normal errors were additionally considered in that study) to estimate parameters in the VB model when individual variability in growth is ignored in a simplified context of uniform recruitment at constant length and mortality among individuals. They found that, in the presence of individual variability, the values of K and L_∞ varied between individuals, and the existing methods provided positively biased parameter estimates. More recently, Cubillos et al. (2009) studied the VB model using the Cope–Punt methodology (Cope and Punt, 2007), which considers a random error in assigning the age that is determined by two different readers. Although this last model considers the independence and normality assumptions for the error terms, it assumes that the assigned age is determined by an exponential or gamma distribution, guaranteeing a real age composition.

In addition to studying the age–length relationship, Cubillos et al. (2009) considered samples of otolith of cardinalfish obtained

* Corresponding author. Tel.: +56 032 2151 455; fax: +56 032 2151 645.
E-mail addresses: jecontrr@uc.cl (J.E. Contreras-Reyes), reivalle@mat.puc.cl (R.B. Arellano-Valle).

0165-7836/$ – see front matter. Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2013.05.002

from 1998 to 2007 in the Chilean south central coastal zone (Latitude 33–42°S), from which a random selection of 96 otoliths was obtained within the range of 20–37 cm from cardinalfish under 15 years of age (Gálvez et al., 2000). However, results from a new method for reading otoliths (Ojeda et al., 2010) show that cardinalfish can live for over 54 years. This species lives in waters from 100 to 550 m in depth but generally stays between depths of 250 and 300 m. According to commercial log books, lengths mostly vary between 17 and 47 cm, with no significant difference between the sexes (Wiff et al., 2005).

In this paper, we study the VB growth model (1) using a flexible class of non-normal distributions for the random error ε_i. Specifically, as in Basso et al. (2010), we consider the class of scale mixture of the skew-normal (SMSN) distributions (Branco and Dey, 2001) for random errors. The SMSN is an attractive class of asymmetric, heavy-tailed distributions that is useful for robust inference when the normality assumption for the error distribution is unrealistic. The flexibility of these distributions allows us to fit observations with a high presence of skewness and heavy tails. These distributions are additionally useful to model distributions with extreme values that generate residual heterogeneity in classic models (Kimura, 1990). Arellano-Valle et al. (2013) conclude that including skewness and heavy tails in the model to fit untransformed data produces different decisions than those obtained by applying the normal model to transformed data using the Box–Cox technique, thus ensuring the effective maximisation of the likelihood function. Our study additionally incorporates Cook's (1986) local influence analysis and the conformal normal curvature of Poon and Poon (1999). The datasets considered in this study contain observations of cardinalfish up to 61 years of age, a substantially wider range of ages than in the Gálvez et al. (2000) study.

2. Methodology

In this section, we study the VB non-linear regression model using a similar approach to the one used by Basso et al. (2010), Lachos et al. (2010, 2011) and Labra et al. (2012). Specifically, we consider a non-linear regression model (2) with the assumption that the random errors ε_i are independent, heteroskedastic and distributed according to the SMSN class of distributions. In other words, we suppose that

$$\varepsilon_i = u_i^{-1/2} e_i + \mu_i, \tag{3}$$

where e_i and u_i are independent random quantities and the μ_i are location parameters. More precisely, the e_i are independent skew-normal random errors, $e_i \sim SN(0,\ \sigma_i^2,\ \lambda_i)$, with density function

$$h(e_i;\sigma_i,\lambda_i) = \frac{2}{\sigma_i}\phi\left(\frac{e_i}{\sigma_i}\right)\Phi\left(\lambda_i\frac{e_i}{\sigma_i}\right), \quad -\infty < e_i < \infty,$$

where $\sigma_i > 0$ and $-\infty < \lambda_i < \infty$ are scale and shape/skewness parameters, respectively, and $\phi(z)$ and $\Phi(z)$ are the density and distribution function of the standardised normal distribution, respectively. In (3), the u_i are positive (scale) random factors perturbing the skew-normality and are assumed to be independent and identically distributed (iid) with distribution function $G(u\,;\nu)$ defined on $(0, \infty)$, depending on the unknown parameter ν (possibly vectorial).

The mean and variance of the skew-normal random errors $e_i \sim SN(0, \sigma_i^2, \lambda_i)$ are given by $E(e_i) = \sqrt{2/\pi}\,\delta_i\sigma_i$ and $\mathrm{Var}(e_i) = \{1 - (2/\pi)\,\delta_i^2\}\sigma_i^2$, where $\delta_i = \lambda_i/\sqrt{1+\lambda_i^2}$. Thus, if we assume in (3) that the moments $\kappa_k = E(u_i^{-k/2})$, $k = 1, 2$, are finite, then the mean and variance of the SMSN random errors ε_i exist and are given by $E(\varepsilon_i) = \mu_i + \sqrt{2/\pi}\,\kappa_1\sigma_i\delta_i$ and $\mathrm{Var}(\varepsilon_i) = \kappa_2\sigma_i^2\{1 - (2/\pi)\,\delta_i^2\}$. To have errors with a zero mean, we impose the condition $\mu_i = -\sqrt{2/\pi}\,\kappa_1\sigma_i\delta_i$. Under this condition, we then have for the response variable y_i

$$E(y_i) = \eta_i \quad \text{and} \quad \mathrm{Var}(y_i) = \kappa_2\sigma_i^2\left\{1 - \left(\frac{2}{\pi}\right)\delta_i^2\right\}, \tag{4}$$

where $\eta_i = \eta(\beta\,; x_i)$ is the VB curve defined as in (2).

We can additionally observe from (3) that, given the scale mixture factors u_i, the random errors ε_i have skew-normal distribution $SN(\mu_i, u_i^{-1}\sigma_i^2, \lambda_i)$ and are independent. Hence, we have from (2) that, conditionally on u_i, the response variables y_i have a distribution $y_i|u_i \sim SN(\eta_i + \mu_i,\ u_i^{-1}\sigma_i,\ \lambda_i)$. The marginal density of y_i is therefore

$$f(y_i;\beta,\sigma_i^2,\lambda_i,\nu) = \frac{2}{\sigma_i}\int_0^\infty \sqrt{u_i}\,\phi(\sqrt{u_i}\,z_i)\Phi(\sqrt{u_i}\,\lambda_i z_i)dG(u_i;\nu), \tag{5}$$

where $z_i = (y_i - \eta_i - \mu_i)/\sigma_i$.

The SMSN class of densities in (5) provides several asymmetric, heavy-tailed models that are useful for robust inference in the presence of influential observations or outliers. All these distributions contain the skew-normal as a special case. In addition, for $\lambda_i = 0$, the SMSN class reduces to the symmetric class of scale mixtures of normal distributions considered in Lange and Sinsheimer (1993).

We additionally assume that $\lambda_i = \lambda$, so that $\delta_i = \delta$, and $\sigma_i^2 = \sigma^2 m(\rho; x_i)$, where $m(\rho\,; x_i)$ is a nonnegative variance function such that $m(0\,; x_i) = 1$. Consequently, in (5) the model parameters are given by $\beta = (L_\infty, K, t_0)^\top$, σ^2, λ, ρ and ν. Additionally, the specifications presented in Table 1 below are considered for the variance function $m(\rho\,; x_i)$. Note that in Table 1, the Kimura, exponential and power functions allow for $\rho = 0$ and $\rho \neq 0$, which are the most obvious examples of homoskedastic and heteroskedastic variability, respectively. Furthermore, for $\rho < 0$, all of these variance models produce a trend toward decreasing variance, whereas in the case of $\rho > 0$, a trend of increasing variance is produced (Kimura, 1990). Finally, the CV constant is a particular case of the Kimura model when $\rho = 1$.

Table 1
Some variance functions.

Model	$m(\rho\,; x_i)$
Homoskedastic	1
CV constant (Candy et al., 2007)	η_i^2
Kimura (Kimura, 1990)	$\eta_i^{2\rho}$
Exponential (Labra et al., 2012)	$e^{\rho x_i}$
Power (Labra et al., 2012)	$x_i^{2\rho}$

2.1. The skew-t model

In this section, we focus our attention principally on the skew-t case with ν ($\nu > 0$) degrees of freedom (Branco and Dey, 2001; Azzalini and Capitanio, 2003; Genton, 2004; Arellano-Valle et al., 2013). This model assumes, as in (3), that the mixing random factors u_i are iid Gamma($\nu/2, \nu/2$) with density $g(u_i\,;\nu) = dG(u_i\,;\nu)/du_i$ given by

$$g(u_i;\nu) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}u_i^{\nu/2-1}e^{-\nu u_i/2}, \quad u_i > 0.$$

In this case, we have

$$\kappa_1 = \left(\frac{\nu}{2}\right)^{1/2}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}, \quad \nu > 1, \quad \kappa_2 = \frac{\nu}{\nu-2}, \quad \nu > 2.$$