GLM and GLMM standardisation of the commercial abalone CPUE for Zone F

Anabela Brandão and Doug S. Butterworth

Marine Resource Assessment & Management Group (MARAM)
Department of Mathematics and Applied Mathematics
University of Cape Town
Rondebosch, 7701, South Africa

August 2012

Abstract

This paper presents a GLM standardisation of the abalone CPUE data for Zone F. A GLMM approach to standardising abalone CPUE data is also investigated.

Introduction

Similar catch-per-unit-effort (CPUE) General Linear Model (GLM) and General Linear Mixed Model (GLMM) standardisation procedures described in Brandão and Butterworth (2012) for Zones A to D have been applied to the commercial abalone data for Zone F. The principle objective of the GLM (and GLMM) analyses are to obtain series of relative abundance indices that have been standardised by incorporating important covariates in the explanation of abalone CPUE variation.

The data

Commercial catch data (as kg whole mass), and effort data (as total duration of dives in minutes for each day dived) are available for the period 1981 to 2011. A year in this paper refers to a Model-year, where a Model-year *y* runs from October of year *y*-1 to September of year *y*. The covariates included in the GLM analysis include the date (in terms of Model-year and season (3-monthly periods)), and the divers. Records with a dive time less than 10 minutes were excluded as well as outliers based upon observations with large residuals (> 6 standard deviations) in an initial GLM fit. Records for divers that had less than three dives in the whole database were excluded. A total of 1 364 data points remained for the analysis. Table 1 gives the number of records used in the final GLM analysis per Model-year.

General Linear Model (GLM) to standardise the CPUE

The following GLM model, which allows for possible annual temporal differences in abalone distribution, is used to standardise the commercial abalone CPUE data:

$$\ln(CPUE) = \mu + \alpha_{year} + \beta_{season} + \varphi_{diver} + \varepsilon \qquad (1)$$

where:

CPUE	is the catch-per-unit-effort defined as catch (kg) divided by dive time (minutes),
μ	is the intercept,
year	is a factor with 27 levels associated with the Model-years 1981–2011 (excluding the years 1984, 1985, 1999 and 2009. In 2009 the fishery was closed),
season	is a factor with 4 levels associated with the season effect (1 = Jan-Mar; 2 = Apr-Jun; 3 = Jul-Sep; 4 = Oct-Dec),
diver	is a factor with 143 levels associated with the diver code, which includes both the entitlement holders coded in the database as well as "divers". Some recent divers not yet allocated a code were given a temporary code of 555 for the purposes of this analysis[1], and
ε	is the error term assumed to be normally distributed.

Interactions between *year* and *season* factors have not been considered in the GLM because of the scarcity of records on a per year and per season basis.

The standardised CPUE series is given by the $\exp\left(\alpha_{year}\right)$ term as there are no interactions with *year* in the model.

General Linear Mixed Model (GLMM) to standardise the CPUE

The GLM used to standardise commercial CPUE indices assumes that all factors in the model are fixed effects with the variance of the response values being that of the error term ε. In a GLM analysis we model only the mean (i.e. the fixed effects) of the data. A GLMM has the ability to model not only the mean of the data but also its variance. In fact, a GLMM also allows for the presence of random variables (called random effects) which describe additional variability in the data apart from that reflected by the error term of equation (1). One of the fixed effects factor in the GLM is "divers" with 143 different levels (in the present analysis) associated with different divers

---

[1] For the years 2007 to 2011 over which this code was used, such records comprise 0.004% of the total.

with some of the divers in the fishery having very few dives. An alternative approach proposed in this paper is to treat “divers” as a random effect in the GLMM.

The GLMM applied to the abalone commercial CPUE data is of the form:

$$\ln\left(CPUE\right)) = \mathbf{X}\alpha + \mathbf{Z}\beta + \varepsilon\,, \tag{3}$$

where :

α	is the unknown vector of fixed effects parameters (this vector includes all the parameters for the effects in the GLM model of equation (1) above, except for diver,
$\mathbf{X}$	is the design matrix for the fixed effects,
β	is the unknown vector of random effects parameters (here diver),
$\mathbf{Z}$	is the design matrix for the random effects,
ε	is an error term assumed to be normally distributed and independent of the random effects.

This approach assumes that both the random effects and the error term have zero mean, i.e. E(β) = E(ε) = 0, so that $\mathrm{E}\left(\ln\left(CPUE\right)\right)\right) = X\alpha$. The variance-covariance matrix for the residual errors (ε) is denoted by **R** and that for the random effects (β) by **G**. The analyses undertaken here assume that the residual errors as well as the random effects are homoscedastic and are uncorrelated, so that both **R** and **G** are diagonal matrices given by:

$$\mathbf{R} = \sigma_{\varepsilon}^{2}\mathbf{I}$$
$$\mathbf{G} = \sigma_{\beta}^{2}\mathbf{I}$$

where **I** denotes an identity matrix. Thus, in the mixed model, the variance-covariance matrix (**V**) for the response variable is given by:

$$\mathrm{Cov}\left(\ln\left(CPUE\right)\right) = \mathbf{V} = \mathbf{ZGZ}^{\mathrm{T}} + \mathbf{R}\,,$$

where $\mathbf{Z}^{\mathrm{T}}$ denotes the transpose of the matrix **Z**.

The estimation of the variance components (**R** and **G**), the fixed effects (α) and the random effects (β) parameters in GLMM requires two steps. First the variance components are estimated by the method of residual maximum likelihood (REML), which produces unbiased estimates for the variance components as it takes into account the degrees of freedom used in estimating the fixed effects. Once estimates of **R** and **G** have been obtained, estimates for the fixed effects parameters (α) can be obtained as well as predictors for the random effects parameters (β).

GLMM standardised CPUE series is again obtained by the $\exp\left(\alpha_{year}\right)$ term.

Results and Discussion

GLM

The examination of the residuals of an initial fit showed evidence of slight heteroscedasticity (Figure 1). To account for this heteroscedasticity, the iterative, inverse-weighting procedure applied by Plagányi and Edwards (2007) to Zones A to D has been applied in which reduced weight is given to the data points with the largest variance in the model.

The GLM model accounts for 60.9% of the total variation of abalone CPUE. Table 2 lists the nominal and standardised CPUE indices provided by the model and Figure 2 shows graphical comparisons of the same. Broadly speaking, the standardisation makes relatively little difference to the nominal trends, except that the GLM-standardised series show a much sharper drop in the CPUE over the first three years of the series.

GLMM

Table 2 lists the GLMM-standardised CPUE indices. Figure 2 shows these graphically and compares them to those obtained by a GLM; there isn’t too much difference between the results for the two different methods of standardisation, with the GLMM standardisation in general following the nominal series more closely than the GLM-standardised one.

Reference

Brandão, A. and Butterworth, D.S. 2012. GLM and GLMM standardization of the commercial abalone CPUE for Zones A-D. FISHERIES/2012/AUG/SWG-AB/04.

Plagányi, É.E. and Butterworth, D.S. 2010. A spatial- and age-structured assessment model to estimate the impact of illegal fishing and ecosystem change on the South African abalone *Haliotis midae* resource. African Journal of Marine Science, 32(2):207-236.

Table 1. The number of data entries per Model year available for the final GLM analysis to standardise the commercial abalone CPUE series for zone F are shown. The abalone fishery was closed in February 2008 and reopened in 2010. Model-years are defined as the period October of the preceding year to September of the year indicated.

Model year	Number of records
1981	24
1982	16
1983	31
1984	
1985	
1986	65
1987	48
1988	48
1989	63
1990	61
1991	42
1992	63
1993	23
1994	32
1995	13
1996	15
1997	29
1998	30
1999	
2000	33
2001	35
2002	53
2003	57
2004	24
2005	24
2006	22
2007	85
2008	243
2009	
2010	68
2011	117

Table 2. Nominal, GLM and GLMM-standardised commercial CPUE series for abalone for Model-years (October of the preceding year to September of the year indicated) 1981 to 2011 for Zone F. The nominal and both sets of standardised values have been divided by the mean value of the respective series.

Model year	Nominal CPUE	GLM standardised CPUE	GLMM standardised CPUE
1980	1.468	1.456	1.191
1981	1.468	0.545	1.114
1982	1.195	0.531	1.084
1983			
1984			
1985	1.295	1.144	1.145
1986	1.197	1.050	1.197
1987	1.476	1.431	1.614
1988	0.713	0.645	0.791
1989	1.412	1.395	1.522
1990	1.097	1.119	1.088
1991	0.763	0.706	0.807
1992	0.939	0.893	1.007
1993	1.809	1.660	2.026
1994	1.257	1.693	1.531
1995	1.381	1.596	1.556
1996	1.275	1.080	1.246
1997	0.894	1.092	1.008
1998			
1999	0.534	0.526	0.428
2000	0.667	0.729	0.597
2001	0.793	0.832	0.664
2002	0.667	0.750	0.597
2003	1.014	1.276	1.023
2004	1.007	1.300	1.089
2005	1.078	1.328	1.105
2006	0.362	0.568	0.383
2007	0.470	0.535	0.400
2008			
2009	0.385		0.404
2010	0.383	0.551	0.385
2011	1.468	1.456	1.191

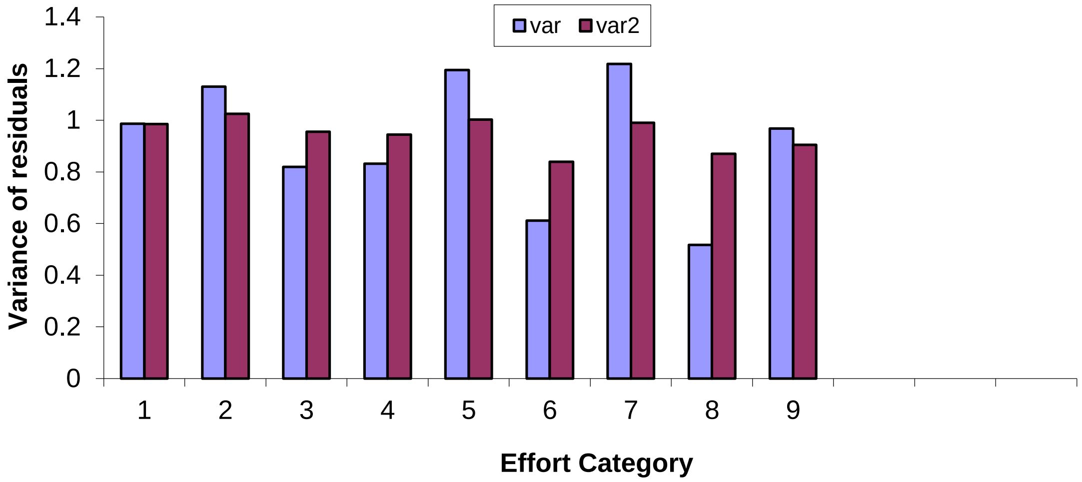

Figure 1. Plot of the average variance of the residuals versus effort category (50 minutes interval) for an initial GLM fit of the data (var, showing evidence of heteroscedasticity) and for the final weighted GLM (var2, showing homoscedastic residuals).

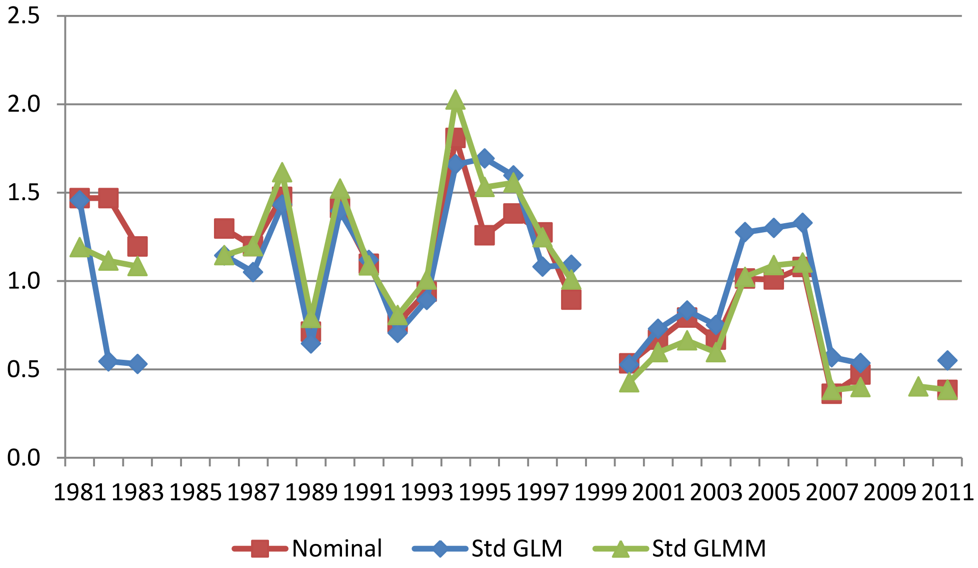

Figure 2. GLM and GLMM-standardised CPUE trends (normalised to their means over the 27 year period) for Zone F. For comparison, the nominal series (also normalised to its mean over the 27 year period) is also shown.