Application of lognormal and delta-lognormal linear models to standardised CPUE abundance indices (1994 to 2002) for orange roughy off Namibia

Anabela Brandão and Doug S. Butterworth

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

March 2003

Abstract

GLM analyses are used to standardise the CPUE data for Namibian orange roughy. The possibility of there being a "learning" period of lower CPUE for a new vessel when it enters the fishery is taken into account. Alternative statistical approaches to deal with tows that record zero catch of orange roughy are considered. Further, to allow for areal expansion of the fishery at each aggregation, sub-aggregations are defined and CPUE trends estimated separately for each. Different methods for combining the results for the various sub-aggregations to provide a single index for an aggregation are considered. In broad terms standardised CPUE values for 2002 appear similar (with some slightly lower) to those for the previous year.

Introduction

In Brandão and Butterworth (2002) commercial CPUE data for orange roughy off Namibia were standardised by applying two Generalised Linear Models (GLMs) that addressed two problems encountered in such analyses for this fishery: i) a considerable number of tows with zero catches and ii) the areal distribution of effort shifting within and even beyond previously defined aggregations (especially notable for the *Johnies* aggregation). These standardised CPUE indices of abundance are then used as an input to a population model to assess the state of the stock (Brandão and Butterworth 2003). In this paper the GLMs of Brandão and Butterworth (2002) are updated using the re-entered data for commercial fishing with previously missing records included and with an extra year's data (for 2002).

The Models

Model to standardise the CPUE

The two models applied to the CPUE time series of data for Namibian orange roughy in Brandão and Butterworth (2002) consist of the base case model which is the GLM "full model" of Brandão and Butterworth (2001), which allows for the possibility that vessels might have different degrees of "effectiveness" in their first year in the orange roughy fishery compared to subsequent years, referred to as the "lognormal model". The second model, described by Lo *et al.* (1992) and Stone and Porter (1999), uses the delta-lognormal method to obtain standardised CPUE indices in the presence of tows with zero catch. This model is referred to as the "delta-lognormal model".

The lognormal model

The lognormal model allows for possible differences in abundance trends in orange roughy in the various aggregations, and assumes the possibility that vessels might operate differently in their first year in the fishery, but have the same degree of "effectiveness" in all subsequent years. When this model was fitted to the corrected re-entered data and with an extra year's information, only the vessel *Whitby* showed a significant difference in its first year of operation. Therefore only this vessel was differentiated from its first year in the fishery and all subsequent years. Also commercial CPUE data are now available for the vessels *Petersen*, *Sea Flower* and *Will Watch* which were missing from the previous analysis. This model is given by:

$$\ln(CPUE + \delta) = \mu + \alpha_{vessel} + \beta_{year} + \gamma_{month} + \lambda_{agg} + \eta_{year \times agg} + \varepsilon \qquad (1)$$

where:

μ is the intercept,

vessel is a factor with 11 levels associated with each of the vessels that have operated in the fishery:

Conbaroya Cuarto

Dantago

Emanguluko

Harvest Nicola

Hurinis

Petersen

Sea Flower

Southern Aquarius

Whitby (first year)

Whitby (subsequent years)

Will Watch,

year is a factor with 9 levels associated with the "fishing years" 1994–2002 (note: "1996", for example, refers to the period July 1996 to June 1997),

month is a factor with 12 levels (January– December),

agg is a factor with 12 levels associated with the four aggregations and their sub-aggregations:

Johnies:	*Johnies1*
	Johnies2
	Johnies3
	Johnies4
Frankies:	*21 Jump Street*
	Frankies Flats
	Frankies Outer
	Three Sisters
	Smifton
Rix:	*Rix Inner*
	Rix Outer
Hotspot,	

$year \times agg$ is the interaction between year and aggregation (this allows for the possibility of different trends for the different sub-aggregations),

δ is a small constant added to the orange roughy CPUE to allow for the occurrence of zero CPUE values, and

ε is an error term assumed to be normally distributed.

Standardised CPUE time series for a given (sub)-aggregation are obtained by calculating:

$$CPUE_{agg,y} = \exp\left[\mu + \beta_{year} + \lambda_{agg} + \eta_{year \times agg}\right] - \delta \qquad (2)$$

where in this application we are standardising on the vessel *Southern Aquarius* and on the month of *August*.

The delta-lognormal model

The delta distribution is often used in instances when there are a considerable number of zero observations, for which zero and non-zero data are consequently treated separately. Final estimates of abundance are obtained from the product of the proportion and the mean of non-zero observations. For the delta-lognormal model, two lognormal linear models (GLMs) are fitted to the commercial CPUE data, one to estimate the proportion of tows for which there is a positive catch, and the other to estimate the standardised CPUE for orange roughy for tows that have a positive catch.

Relative abundance indices of orange roughy are then given by:

$$CPUE_y = CPUE_y^{+ve} Prop_y^{non-zero} \tag{3}$$

where:

$CPUE_y^{+ve}$ is the standardised CPUE index for tows which have positive catches, and

$Prop_y^{non-zero}$ is the standardised measure of the proportion of tows that have positive catches.

Standardised indices for the component of positive catches were obtained by fitting the same lognormal model given in equation (1), i.e. the model to estimate the abundance of positive catches is given by:

$$\ln(CPUE^{+ve}) = \mu + \alpha_{vessel} + \beta_{year} + \gamma_{month} + \lambda_{agg} + \eta_{year \times agg} + \varepsilon \tag{4}$$

where the notation is as in equation (1).

In the case of orange roughy tow data the proportion of tows with a positive catch is either "0" or "1" for an individual tow, and therefore a model for the proportion positive assuming binomially distributed errors is considered, given by:

$$Prop^{non-zero} = \mu + \alpha_{vessel} + \beta_{year} + \gamma_{month} + \lambda_{agg} + \eta_{year \times agg} + \varsigma \tag{5}$$

where

ζ is an error term assumed to be binomially distributed.

Standardised measures of the abundance of orange roughy in positive tows are estimated by calculating:

$$C\hat{P}UE_y^{+ve} = \exp\left[\hat{\mu} + \hat{\beta}_{year} + \hat{\lambda}_{agg} + \hat{\eta}_{year \times agg}\right]\psi_y^{+ve} \tag{6}$$

where

ψ_v^{+ve} is a correction factor for bias (Lo *et al.* 1992), given by:

$$\psi_y^{+ve} = g_m\left[\frac{m+1}{2m}\left(\hat{\xi}^2 - \hat{\xi}_{\hat{\theta}}^2\right)\right] \tag{7}$$

where

$\hat{\xi}^2$ is the residual variance,

m is the degrees of freedom for the estimate of residual variance,

$\hat{\theta}$ is given by $\hat{\mu} + \hat{\beta}_{year} + \hat{\lambda}_{agg} + \hat{\eta}_{year \times agg}$,

$\hat{\xi}_{\hat{\theta}}^2$ is the variance of $\hat{\theta}$, and

$$g_m(t) = \sum_{p=0}^{\infty}\left[\frac{m^p(m+2p)}{m(m+2)\cdots(m+2p)}\left(\frac{m}{m+1}\right)^p \frac{t^p}{p!}\right]$$

where t is the argument of the function.

Standardised measures of the proportion of positive catches of orange roughy is given by:

$$P\hat{r}op_y^{non-zero} = \frac{\exp\left[\hat{\mu} + \hat{\beta}_{year} + \hat{\lambda}_{agg} + \hat{\eta}_{year \times agg}\right]}{1 + \exp\left[\hat{\mu} + \hat{\beta}_{year} + \hat{\lambda}_{agg} + \hat{\eta}_{year \times agg}\right]}. \tag{8}$$

Model Implementation

To take into account movement of orange roughy within a known aggregation, the analyses in Brandão and Butterworth (2002) took into consideration not only tows that lie within the inner strata of an aggregation, but also tows that take place in the outer strata of the aggregation. The levels of the factor for aggregations in the GLMs are then given as the various sub-aggregations. The definition of aggregations and their sub-aggregations of Brandão and Butterworth (2002) are used in this paper.

Commercial tow information inside the known aggregations of orange roughy in Namibia for the fishing years (July – June) 1994 to 2002, as provided by E. Johnsen has been used. The year 2002 is incomplete as this fishing year ends only in June 2003. Data until the end of September were available. A total of 16 289 data points was available for the analyses. Bottom distances were calculated from the GPS positions for each tow. For tows that did not have haul positions, but did have bottom time information, bottom distances were calculated by the following regression relationship:

Bottom distance [km] = *bottom time* [h] * 5.6082+0.1259.

GLM Results and Discussion

The lognormal linear model of equation (1) was fitted to the commercial CPUE data. In this instance, a value of δ taken to be 10% of the average of the orange roughy CPUE data (=0.016) was used. Examination of the results, especially the interaction terms between vessel and year, revealed that generally the only large effects observed occur in the first year in which the vessel *Whitby* took part in the fishery. Given these results it was decided to include extra levels of the *Whitby* vessel factor to account for the first year showing a different pattern from other years, and thus omitting a year-vessel interaction from the GLM.

The lognormal model (equation (1)) accounts for 40% of the total variation of orange roughy CPUE. Table 1 shows the parameter estimates obtained for the factor vessel. The lognormal model applied to tows with a positive catch (equation (4)) accounts for 43% of the total variation of orange roughy positive CPUE. A total of 13 600 tows have a positive catch. Tables 2 to 5 show the index of abundance provided by the lognormal model, and the delta-lognormal model assuming binomial errors for the proportion positive for each aggregation. Observations are not available for all years in all of the sub-aggregations. The three methods of combining the standardised CPUE indices from each individual sub-aggregation to obtain a standardised CPUE index for each aggregation of Brandão and Butterworth (2002) were used to deal with such empty cells. The first method, referred to as the “zero” method, assumes that empty cells mean that there was no orange roughy in those areas for those years. The second method (“same”) assumes that although no observations were made, there was orange roughy present. It is further assumed that the same amount was present as at the first time an observation was made, or the same as last observed is present for subsequent years. The third method referred to as the "proportional" method, makes the same assumption as the previous method, except that now the amount is taken to be in the same proportion relative to the previous year to that observed in another sub-aggregation for that year. The overall standardised index for each aggregation is obtained by summing the stanardised CPUE for each sub-aggregation multiplied by its associated geographical area (Table 6)

Figures 1 to 7 show the index of abundance provided by the lognormal model and the delta-lognormal model assuming binomial errors for the proportion positive for each aggregation. For each aggregation (except *Hotspot* for which there are no empty cells) a comparison is provided of the indices of abundance of orange roughy obtained by fitting the lognormal model to the CPUE data for the three methods of combining the individual indices of the sub-aggregations. A comparison is also shown for the two models fitted to the CPUE data using the "zero" method of combining individual indices from sub-aggregations. All aggregations show differences between the three methods of combining individual indices (Tables 2 to 4 and Figs. 1, 3 and 5). Differences are most marked in the first few years of the series (mostly for pre–1997). In all aggregations, the indices obtained from fitting a lognormal model hardly differ from those obtained from fitting a delta-lognormal model assuming binomial errors for the proportion positive.

In broad terms standardised CPUE values for 2002 appear similar (with some slightly lower) to those for the previous year.

Acknowledgements

A number of people have willingly provided data for this study. Assistance from Espen Johnsen, in particular, of NatMIRC is gratefully acknowledged, as is funding from the Namibian Deepwater Fishing Industry and the Namibian Ministry of Fisheries and Marine resources.

References

Brandão, A. and Butterworth, D.S. 2001. The application of GLM analyses to the CPUE data for Namibian orange roughy. Namibian Ministry of Fisheries and Marine Resources document: DWFWG/WkShop/Mar01/doc1.

Brandão, A. and Butterworth, D.S. 2002. Standardised CPUE abundance indices for orange roughy off Namibia based on lognormal and delta-lognormal models. Namibian Ministry of Fisheries and Marine Resources document: DWFWG/WkShop/Feb02/doc1.

Brandão, A. and Butterworth, D.S. 2003. Stock assessment of Namibian orange roughy populations using an age-structured production model and all available indices of abundance from 1994 to 2002, and making allowance for annually variable aggregation of the stocks. Namibian Ministry of Fisheries and Marine Resources document: DWFWG/WkShop/Mar03/doc.

Lo, N.C., Jacobson, L.D., and Squire, J.L. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aqaut. Sci. 49: 2515–2526.

Stone, H.H. and Porter, J.M. (1999) Standardized CPUE indices for Canadian bluefin tuna fisheries based on commercial catch rates. International Commission for the Conservation of Atlantic Tunas, collective Volume of Scientific Papers: XLIX (2): 206–221.

Table 1. Parameter estimates for the vessel factor when the lognormal model (equation (1)) is fitted. The value of $\delta = 0.1$ is chosen.

Vessel	**Vessel factor =** $e^{\alpha_{vessel}}$
Conbaroya Cuarto	0.611
Dantago	0.682
Emanguluko	0.743
Harvest Nicola	0.623
Hurinis	0.676
Petersen	0.643
Sea Flower	0.963
Southern Aquarius	**1.000**
Whitby (first year)	1.183
Whitby (subsequent years)	0.658
Will Watch	1.552

Table 2. Standardised CPUE series (each normalised to their mean over the years considered) for the *Johnies* aggregation obtained by fitting the "lognormal model" and the delta-lognormal model assuming binomial errors for the proportion positive to the observed CPUE data for Namibian orange roughy. Three methods ("zero", "same" and "proportional" of dealing with years in which no observations were made in the sub-aggregations are considered.

Year	"Zero" method		"Same" method		"Proportional" method	
	Lognormal model	Delta-lognormal model (binomial errors)	Lognormal model	Delta-lognormal model (binomial errors)	Lognormal model	Delta-lognormal model (binomial errors)
1994	2.209	2.878	2.485	2.921	5.020	5.407
1995	0.506	0.663	1.376	1.476	1.150	1.245
1996	0.643	0.734	1.465	1.522	1.461	1.378
1997	1.798	1.796	1.171	1.171	0.436	0.369
1998	0.998	0.876	0.650	0.572	0.242	0.180
1999	0.775	0.584	0.505	0.381	0.188	0.120
2000	0.818	0.665	0.533	0.434	0.199	0.137
2001	0.659	0.441	0.429	0.288	0.160	0.091
2002	0.594	0.362	0.387	0.236	0.144	0.074

Table 3. Standardised CPUE series (each normalised to their mean over the years considered) for the *Frankies* aggregation obtained by fitting the "lognormal model", the delta-lognormal model assuming binomial errors for the proportion positive, and the delta-lognormal model assuming lognormal errors for the proportion positive to the observed CPUE data for Namibian orange roughy. Three methods ("zero", "same" and "proportional" of dealing with years in which no observations were made in the sub-aggregations are considered.

Year	**"Zero" method**		**"Same" method**		**"Proportional" method**	
	Lognormal model	**Delta-lognormal model (binomial errors)**	**Lognormal model**	**Delta-lognormal model (binomial errors)**	**Lognormal model**	**Delta-lognormal model (binomial errors)**
1995	0.453	0.578	2.487	2.329	4.544	4.028
1996	2.597	2.292	2.202	1.949	1.174	0.921
1997	1.190	1.117	1.009	0.950	0.538	0.449
1998	1.026	1.097	0.870	0.933	0.464	0.441
1999	0.392	0.333	0.385	0.327	0.190	0.142
2000			0.363	0.433	0.079	0.047
2001	0.342	0.583	0.342	0.540	0.182	0.259
2002			0.342	0.540	0.828	1.714

Table 4. Standardised CPUE series (each normalised to their mean over the years considered) for the *Rix* aggregation obtained by fitting the "lognormal model" and the delta-lognormal model assuming binomial errors for the proportion positive to the observed CPUE data for Namibian orange roughy. Three methods ("zero", "same" and "proportional" of dealing with years in which no observations were made in the sub-aggregations are considered.

Year	**"Zero" method**		**"Same" method**		**"Proportional" method**	
	Lognormal model	**Delta-lognormal model (binomial errors)**	**Lognormal model**	**Delta-lognormal model (binomial errors)**	**Lognormal model**	**Delta-lognormal model (binomial errors)**
1995	0.511	0.705	1.788	1.822	2.040	2.464
1996	0.400	0.355	1.717	1.591	1.595	1.241
1997	2.729	2.493	1.730	1.648	1.680	1.543
1998	1.675	1.813	1.062	1.198	1.031	1.122
1999	0.602	0.639	0.382	0.422	0.371	0.396
2000	0.903	0.943	0.572	0.623	0.556	0.583
2001	0.578	0.520	0.366	0.343	0.356	0.322
2002	0.603	0.533	0.382	0.352	0.371	0.330

Table 5. Standardised CPUE series (each normalised to their mean over the years considered) for the *Hotspot* aggregation obtained by fitting the "lognormal model" and the delta-lognormal model assuming binomial errors for the proportion positive to the observed CPUE data for Namibian orange roughy. There are no zero- cells for *Hotspot*, so the "zero", "same" and "proportional" options need not be considered.

Year	Lognormal model	Delta-lognormal model (binomial errors)
1994	3.806	4.148
1995	1.983	2.177
1996	0.832	0.675
1997	0.488	0.411
1998	0.561	0.469
1999	0.440	0.383
2000	0.303	0.277
2001	0.329	0.285
2002	0.258	0.174

Table 6. Geographical area for each sub-aggregation of orange roughy off Namibia.

Aggregation	Sub-aggregation	Area (km^2)
Johnies	***Johnies1***	82.8
	Johnies2	457.2
	Johnies3	198.2
	Johnies4	587.1
Frankies	***21 Jump Street***	39.2
	Frankies Flats	17.8
	Frankies Outer	1 255.0
	Three Sisters	39.6
	Smifton	15.8
Rix	***Rix Inner***	99.4
	Rix Outer	685.6
Hotspot	***Hotspot Inner***	97.3
	Hotspot Outer*	89.0

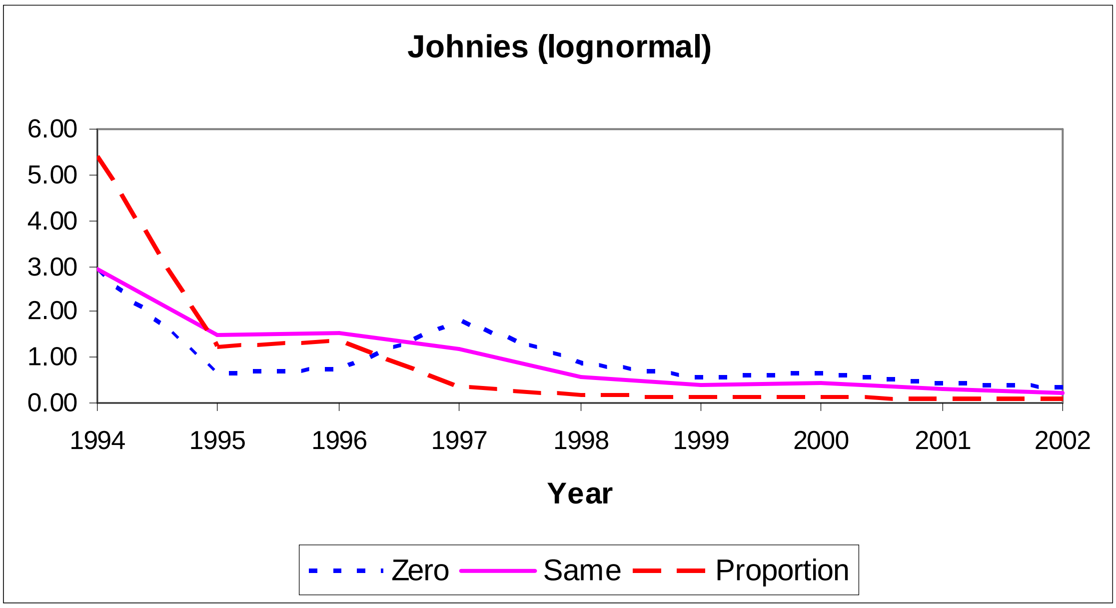

Figure 1. Index of abundance fro the *Johnies* aggregation (normalised to its mean over the nine year period) for Namibian orange roughy obtained from fitting the lognormal model. Results are shown for the three methods of dealing with empty cells when combining the indices from sub-aggregations.

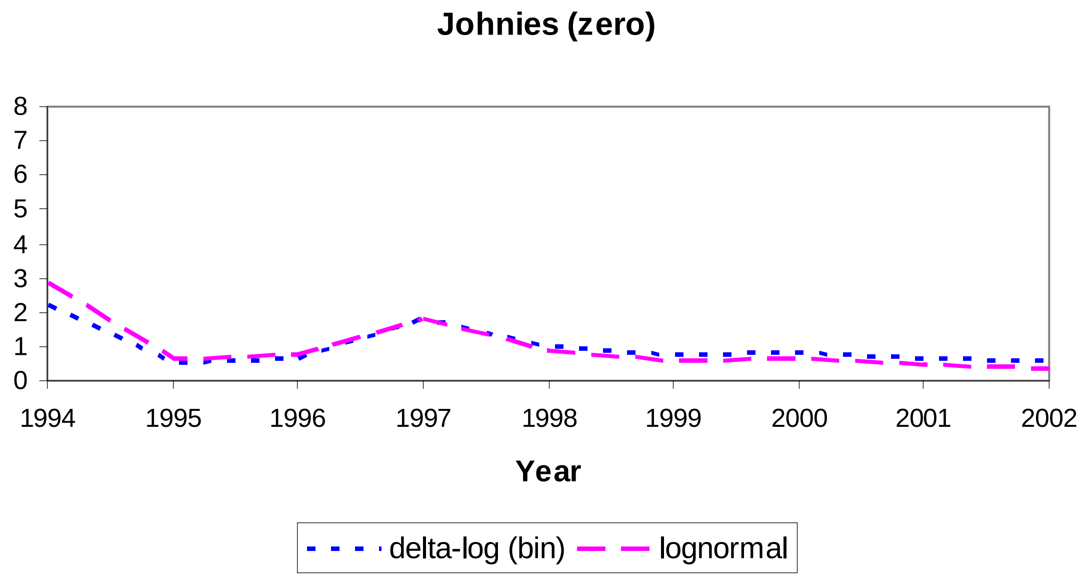

Figure 2. Index of abundance for the *Johnies* aggregation (normalised to its mean over the nine year period) for Namibian orange roughy obtained from fitting the lognormal model and the delta-lognormal model assuming binomial errors for the proportion positive. Results are shown for the "zero" method of dealing with empty cells when combining the indices from sub-aggregations.

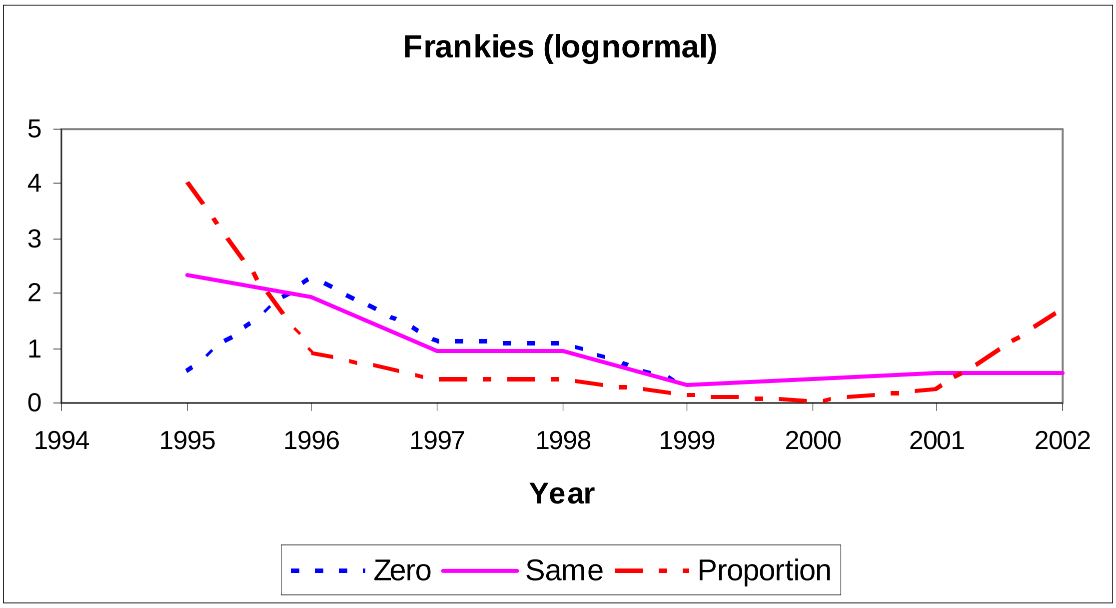

Figure 3. Index of abundance for the *Frankies* aggregation (normalised to its mean over the nine year period) for Namibian orange roughy obtained from fitting the lognormal model. Results are shown for the three methods of dealing with empty cells when combining the indices from sub-aggregations.

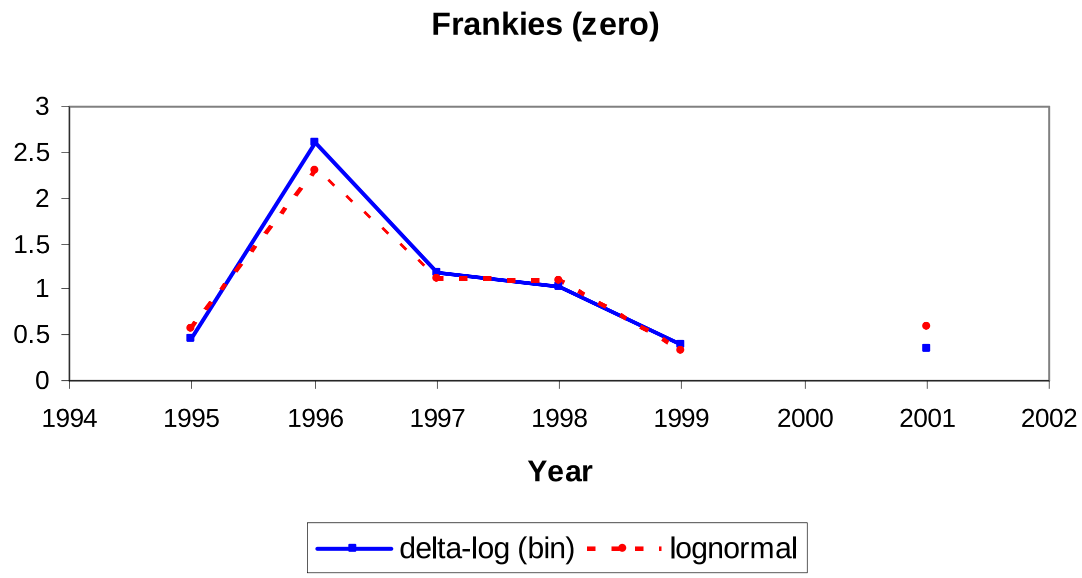

Figure 4. Index of abundance for the *Frankies* aggregation (normalised to its mean over the nine year period) for Namibian orange roughy obtained from fitting the lognormal model and the delta-lognormal model assuming binomial errors for the proportion positive. Results are shown for the "zero" method of dealing with empty cells when combining the indices from sub-aggregations.

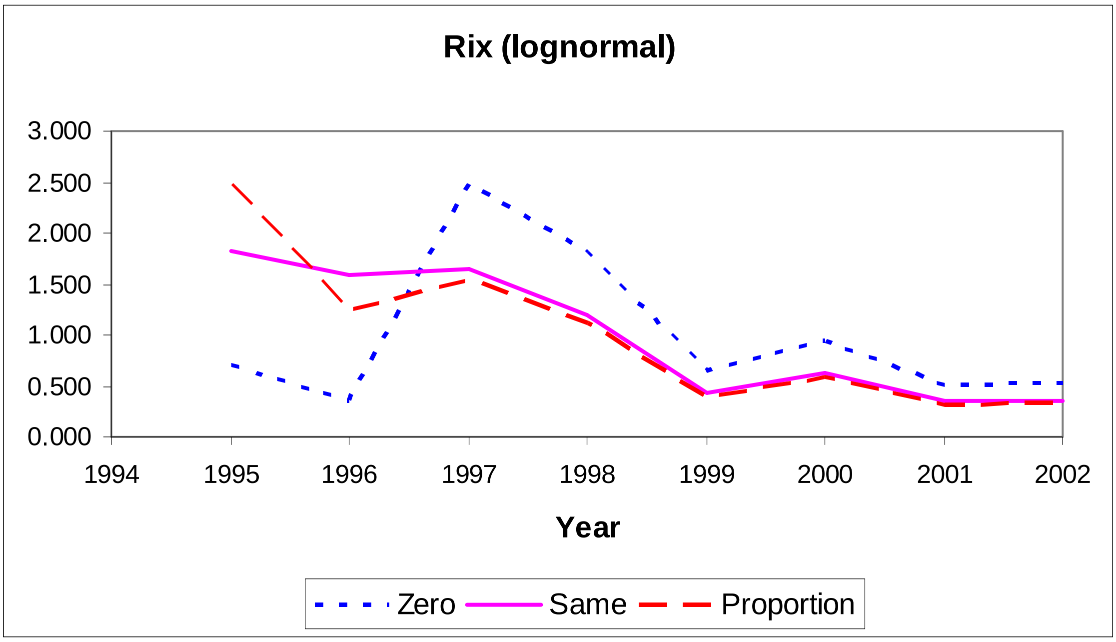

Figure 5. Index of abundance for the *Rix* aggregation (normalised to its mean over the nine year period) for Namibian orange roughy obtained from fitting the lognormal model. Results are shown for the three methods of dealing with empty cells when combining the indices from sub-aggregations.

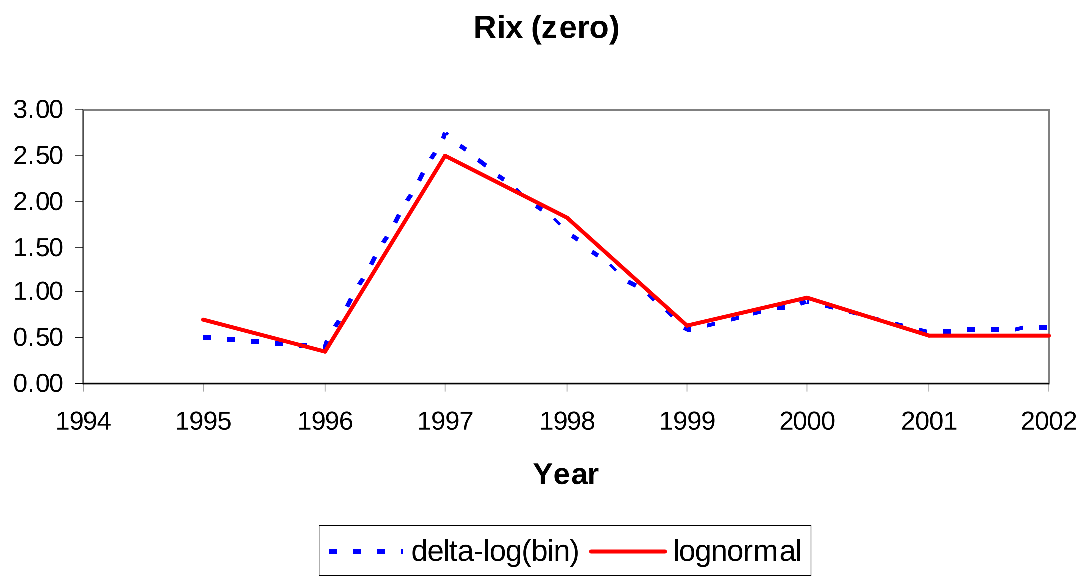

Figure 6. Index of abundance for the *Rix* aggregation (normalised to its mean over the nine year period) for Namibian orange roughy obtained from fitting the lognormal model and the delta-lognormal model assuming binomial errors for the proportion positive. Results are shown for the "zero" method of dealing with empty cells when combining the indices from sub-aggregations.

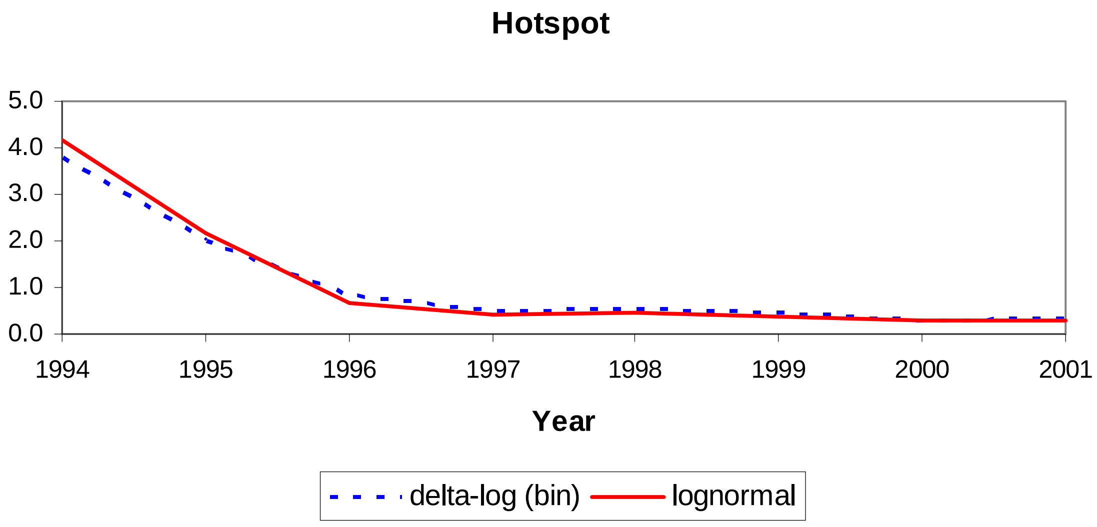

Figure 7. Index of abundance for the *Hotspot* aggregation (normalised to its mean over the nine year period) for Namibian orange roughy obtained from fitting the lognormal model.