# GLM standardised CPUE abundance indices for orange roughy off Namibia from 1994 now updated to include records up to 2005

*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*

*February 2006*

## Abstract

GLM analyses are used to standardise the CPUE data for Namibian orange roughy in a manner that deals with tows that record zero catch of 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. 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. The standardised CPUE values for 2004 (i.e. the July 2004 – June 2005 fishing year) are more often below than above those for the previous year.

## Introduction

At the Deep Water Fisheries Working Group meeting held in Swakopmund on March 2003 it was decided to adopt the delta-lognormal model, as first proposed in Brandão and Butterworth (2002), to standardise the commercial orange roughy CPUE data. This type of model addressed two problems encountered in the 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 2006). In this paper, the results of the updated standardised CPUE indices for orange roughy taking an extra year's data into account are presented, using a delta-lognormal model.

## The Model

The model applied to the CPUE time series of data for Namibian orange roughy is a delta-lognormal which takes into account the presence of tows with zero catch as described by Lo *et al.* (1992) and Stone and Porter (1999).

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 linear models 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:

$$CPUE_y = \sum_{agg} \left( CPUE_{agg,y}^{+ve} Prop_{agg,y}^{non-zero} \right) A_{agg} \qquad (1)$$

where:

$CPUE_{agg,y}^{+ve}$ is the standardised CPUE index for tows which have positive catches for a given sub-aggregation,

$Prop_{agg,y}^{non-zero}$ is the standardised measure of the proportion of tows that have positive catches for a given sub-aggregation, and

$A_{agg}$ is the geographical area for a given sub-aggregation (Table 1).

Standardised indices for the component related to the CPUE of positive catches were obtained by fitting a lognormal model that allows for possible differences in abundance trends in orange roughy in the various aggregations, and assume 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. Brandão and Butterworth (2003) found that only the vessel *Whitby* showed a significant difference in its first year of operation. New vessels that have operated in the fishery since this analysis have not shown a significant difference in their first year of operation and therefore only the *Whitby* vessel is differentiated with respect to its first year in the fishery and all subsequent years. The model to estimate the standardised index of positive catches is thus given by:

$$\ln(CPUE^{+ve}) = \mu + \alpha_{vessel} + \beta_{y} + \gamma_{month} + \lambda_{agg} + \eta_{y\times agg} + \varepsilon \quad (2)$$

where:

$\mu$ is the intercept,

*vessel* is a factor with 14 levels associated with each of the vessels that have operated in the fishery:

*Bell Ocean II*
*Conbaroya Quarto*
*Concasa*
*Dantago*
*Emanguluko*
*Harvest Nicola*
*Hurinis*
*Petersen*
*Sea Flower*
*Southern Aquarius*
*Ulzama*
*Whitby (first year)*
*Whitby (subsequent years)*
*Will Watch,*

*y* is a factor with 11 levels associated with the “fishing years” 1994–2004 (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*,

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

$\varepsilon$ is an error term assumed to be normally distributed.

In the case of the 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:

$$\ln\left(\frac{Prop^{non-zero}}{1-Prop^{non-zero}}\right) = \mu + \alpha_{vessel} + \beta_y + \gamma_{month} + \lambda_{agg} + \eta_{y\times agg} + \varsigma \qquad (3)$$

where

$\zeta$ is an error term assumed to be binomially distributed.

Standardised measures of the abundance of orange roughy in positive tows for a given (sub)-aggregation are estimated by calculating:

$$C\hat{P}UE^{+ve}_{agg,y} = \exp\left[\hat{\mu} + \hat{\beta}_y + \hat{\lambda}_{agg} + \hat{\eta}_{y\times agg}\right]\psi^{+ve}_y \qquad (4)$$

where in this application standardisation is with respect to the vessel *Southern Aquarius* and to the month of *August*, and where

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

$$\psi^{+ve}_y = g_m\left[\frac{m+1}{2m}\left(\hat{\xi}^2 - \hat{\xi}^2_{\hat{\theta}}\right)\right] \qquad (5)$$

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 $\hat{\mu} + \hat{\beta}_y + \hat{\lambda}_{agg} + \hat{\eta}_{y\times agg}$,

$\hat{\xi}^2_{\hat{\theta}}$ 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 are given by:

$$\hat{Prop}_{agg,y}^{non-zero} = \frac{\exp\left[\hat{\mu} + \hat{\beta}_y + \hat{\lambda}_{agg} + \hat{\eta}_{y \times agg}\right]}{1 + \exp\left[\hat{\mu} + \hat{\beta}_y + \hat{\lambda}_{agg} + \hat{\eta}_{y \times agg}\right]} . \tag{6}$$

## 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 thus correspond to the various sub-aggregations. The definitions of aggregations and their sub-aggregations given by 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 2004, as provided by NatMIRC, has been used. Only eleven tows were available for the 2005 fishing year as data were available only until the end of July 2005 (i.e. for only one month of fishing). As a restriction is applied to the data records used in the GLM analyses that there must be 20 or more records within a fishing year in each sub-aggregation, insufficient data were available for the analyses to include the 2005 fishing year.  A total of 17 971 tows was available for the analyses. Of these, 15 119 recorded a non-zero catch. Bottom distances were calculated from the GPS positions for each tow. For tows that did not have haul positions (the majority of tows in the last few years), 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

developed in earlier analyses (Brandão and Butterworth 2003).

## GLM Results and Discussion

The lognormal model applied to tows with a positive catch (equation (2)) accounts for 46.5% of the total variation of orange roughy positive CPUE. Table 2 shows the parameter estimates obtained for the factor vessel for the CPUE of positive catches and for the proportion of positive tows. Tables 3 to 6 show the index of abundance provided by 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. Two of 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 referred to as the "proportional" method, assumes that although no observations were made, there was orange roughy present. It then further assumes that the amount present is in the same proportion relative to the previous year to that observed in the other constituent sub-aggregation of that aggregation for that year. If, however, there are no data for any of the sub-aggregations for the year under consideration, this method assumes a proportional change from the previous year (*y*-1) that continues the trend between years *y*-2 and *y*-1. The overall standardised index for each aggregation is obtained by summing the standardised CPUE for each sub-aggregation multiplied by its associated geographical area (equation(1)).

Figures 1 to 4 show the index of abundance provided by 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 delta-lognormal model to the CPUE data for the two methods of combining the individual indices of the sub-aggregations. All aggregations show differences between the two methods of combining individual indices (Tables 3 to 5 and Figs. 1 to 3). These differences are most marked in the first few years of the series (mostly for pre–1997). For comparison purposes, nominal CPUE series are also shown in Figs. 1 to 4. Again differences in the series are most marked in the first few years of the series.

The standardised CPUE values for 2004 are lower than those for the previous year except for *Hotspot* and *Johnies* under the "proportional" method of dealing with empty cells (no data are available for *Rix* in 2004 as this aggregation was closed for commercial fishing on 1st August 2004).

## Acknowledgements

Data were provided by NatMIRC for this study. Funding from the Namibian Deepwater Fishing Industry is gratefully acknowledged.

## References

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/Doc 1.

Brandão, A. and Butterworth, D.S. 2003. Application of lognormal and delta-lognormal linear models to standardised CPUE abundance indices (1994 to 2002) for orange roughy off Namibia. Namibian Ministry of Fisheries and Marine Resources document: DWFWG/WkShop/Mar03/Doc 2.

Brandão, A. and Butterworth, D.S. 2006. Namibian orange roughy age-structured production model assessments and projections for 2006. Namibian Ministry of Fisheries and Marine Resources document: DWFWG/WkShop/Feb06/Doc 3.

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.** 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 |

* Too few tows fall within the *Hotspot Outer* sub-aggregation for specific account to be taken of this sub-aggregation, and therefore these tows are omitted from the GLM analyses.

**Table 2.** Parameter estimates for the vessel factor when the lognormal model is applied to tows with a positive catch (equation (2)) and the model for the proportion positive (equation (3)) are fitted.

| **Vessel** | **Vessel factor = $e^{\alpha_{vessel}}$ (positive catches)** | **Vessel factor = $e^{\alpha_{vessel}}$ (proportion positive)** |
|---|---|---|
| *Bell Ocean II* | 0.382 | 0.173 |
| *Conbaroya Cuarto* | 0.304 | 1.271 |
| *Concasa* | 0.188 | 1.361 |
| *Dantago* | 0.314 | 0.777 |
| *Emanguluko* | 0.450 | 1.136 |
| *Harvest Nicola* | 0.214 | 0.497 |
| *Hurinis* | 0.321 | 0.630 |
| *Petersen* | 0.438 | 4.276 |
| *Sea Flower* | 0.509 | 2330* |
| ***Southern Aquarius*** | **1.000** | **1.000** |
| *Ulzama* | 1.249 | 0.505 |
| *Whitby* (first year) | 0.503 | 212* |
| *Whitby* (subsequent years) | 1.035 | 1.021 |
| *Will Watch* | 1.007 | 2036* |

* Note: These large values are not unrealistic, but rather are a consequence of the logit transformation used in equation (3) [which restricts the final factor applied to lie between 0 and 1] and the fact that these three vessels had no records of zero tows.

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

| **Year** | **Standardised indices** | |
|---|---|---|
| | **"Zero" method** | **"Proportional" method** |
| **1994** | 5.756 | 7.627 |
| **1995** | 0.868 | 1.151 |
| **1996** | 1.218 | 1.614 |
| **1997** | 1.608 | 0.294 |
| **1998** | 0.587 | 0.107 |
| **1999** | 0.261 | 0.048 |
| **2000** | 0.224 | 0.041 |
| **2001** | 0.126 | 0.023 |
| **2002** | 0.158 | 0.029 |
| **2003** | 0.134 | 0.024 |
| **2004** | 0.059 | 0.042 |

**Table 4.** Standardised CPUE series (each normalised to their mean over the years considered) for the *Frankies* aggregation obtained by fitting the delta-lognormal model, assuming binomial errors for the proportion positive, to the observed CPUE data for Namibian orange roughy. Two methods ("zero" and "proportional") for dealing with years in which no observations were made in the sub-aggregations are considered. The *Frankies* aggregation was closed in 1999 and has been partially reopened since 2002 and fully reopened since 2005 (calendar years). Therefore the indices for the fishing years that span those calendar years are based on very few data.

| Year | Standardised indices | |
|---|---|---|
| | "Zero" method | "Proportional" method |
| **1995** | 1.362 | 7.461 |
| **1996** | 4.305 | 1.447 |
| **1997** | 1.361 | 0.457 |
| **1998** | 0.649 | 0.218 |
| **1999** | 0.292 | 0.104 |
| **2000** | — | 0.050* |
| **2001** | 0.432 | 0.172 |
| **2002** | 0.151 | 0.071 |
| **2003** | 0.428 | 0.015 |
| **2004** | 0.020 | 0.003 |

* See text for brief explanation of convention under this method in cases of no data for any sub-aggregation.

**Table 5.** Standardised CPUE series (each normalised to their mean over the years considered) for the *Rix* aggregation obtained by fitting the delta-lognormal model, assuming binomial errors for the proportion positive, to the observed CPUE data for Namibian orange roughy. Two methods ("zero" and "proportional") for dealing with years in which no observations were made in the sub-aggregations are considered.

| **Year** | **Standardised indices** | |
|---|---|---|
| | **"Zero" method** | **"Proportional" method** |
| **1995** | 0.956 | 3.116 |
| **1996** | 0.717 | 2.337 |
| **1997** | 4.709 | 2.547 |
| **1998** | 2.042 | 1.105 |
| **1999** | 0.404 | 0.218 |
| **2000** | 0.419 | 0.226 |
| **2001** | 0.300 | 0.162 |
| **2002** | 0.300 | 0.162 |
| **2003** | 0.154 | 0.083 |
| **2004** | — | 0.044* |

* See text for brief explanation of convention under this method in cases of no data for any sub-aggregation.

**Table 6.** Standardised CPUE series (each normalised to their mean over the years considered) for the *Hotspot* aggregation obtained by fitting 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” and “proportional” methods become identical.

| **Year** | **Standardised indices** |
|---|---|
| **1994** | 5.782 |
| **1995** | 2.463 |
| **1996** | 0.871 |
| **1997** | 0.314 |
| **1998** | 0.493 |
| **1999** | 0.261 |
| **2000** | 0.102 |
| **2001** | 0.167 |
| **2002** | 0.352 |
| **2003** | 0.093 |
| **2004** | 0.101 |

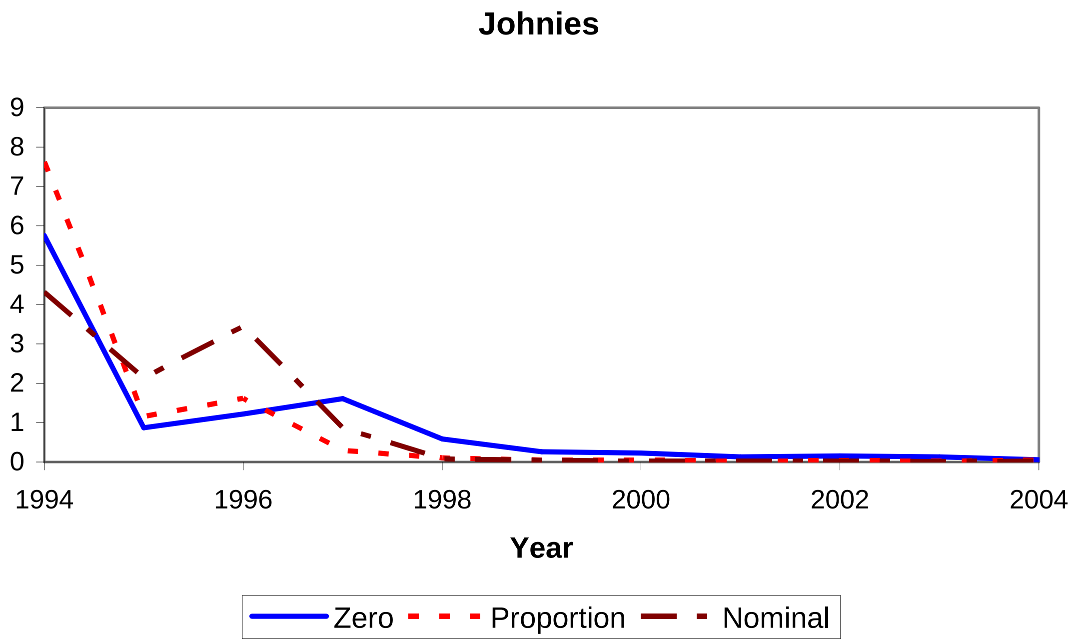


**Figure 1.** Index of abundance for the *Johnies* aggregation (normalised to its mean over the ten year period) for Namibian orange roughy obtained from fitting the delta-lognormal model assuming binomial errors for the proportion positive. Results are shown for the two methods of dealing with empty cells when combining the indices from sub-aggregations. For comparison, the nominal CPUE series is also shown.

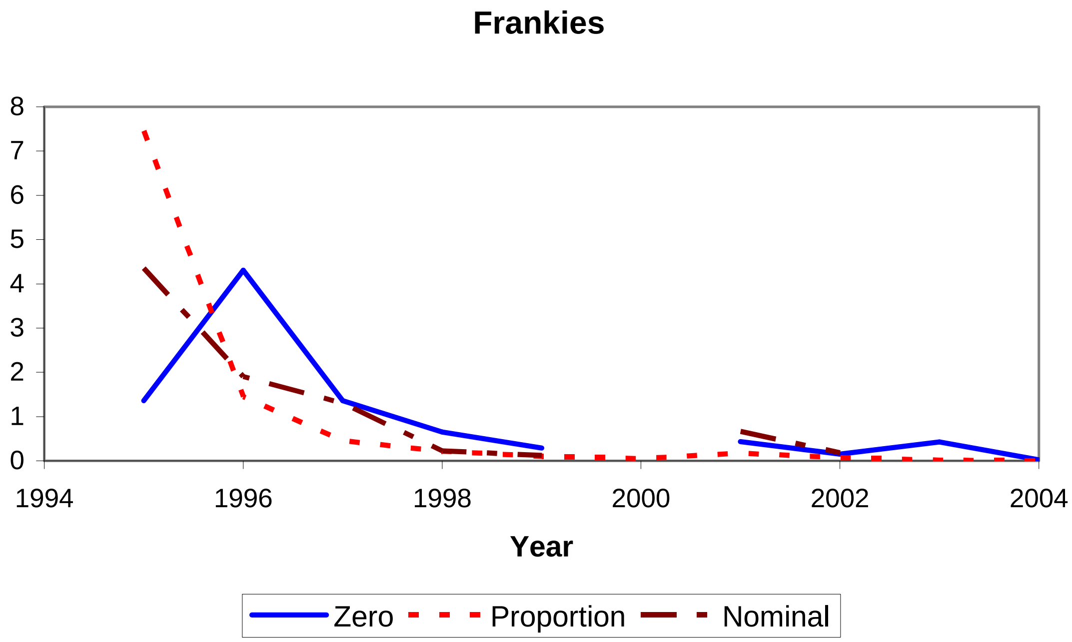


**Figure 2.** Index of abundance for the *Frankies* aggregation (normalised to its mean over the ten year period) for Namibian orange roughy obtained from fitting the delta-lognormal model assuming binomial errors for the proportion positive. Results are shown for the two methods of dealing with empty cells when combining the indices from sub-aggregations. For comparison, the nominal CPUE series obtained is also shown.

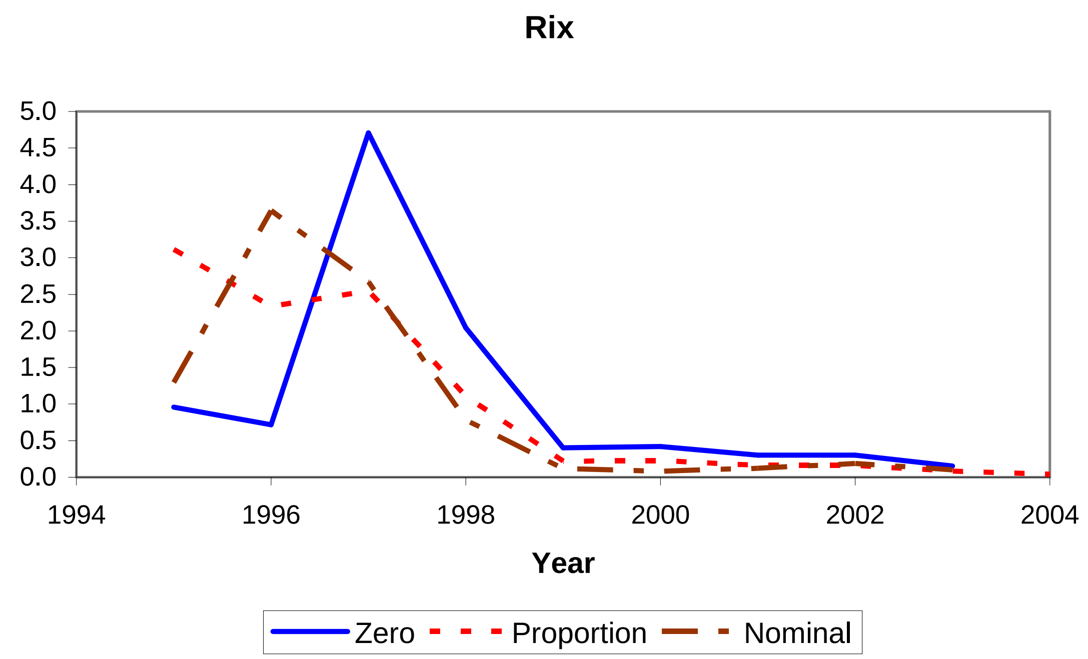


**Figure 3.** Index of abundance for the *Rix* aggregation (normalised to its mean over the ten year period) for Namibian orange roughy obtained from fitting the delta-lognormal model assuming binomial errors for the proportion positive. Results are shown for the two methods of dealing with empty cells when combining the indices from sub-aggregations. For comparison, the nominal CPUE series is also shown.

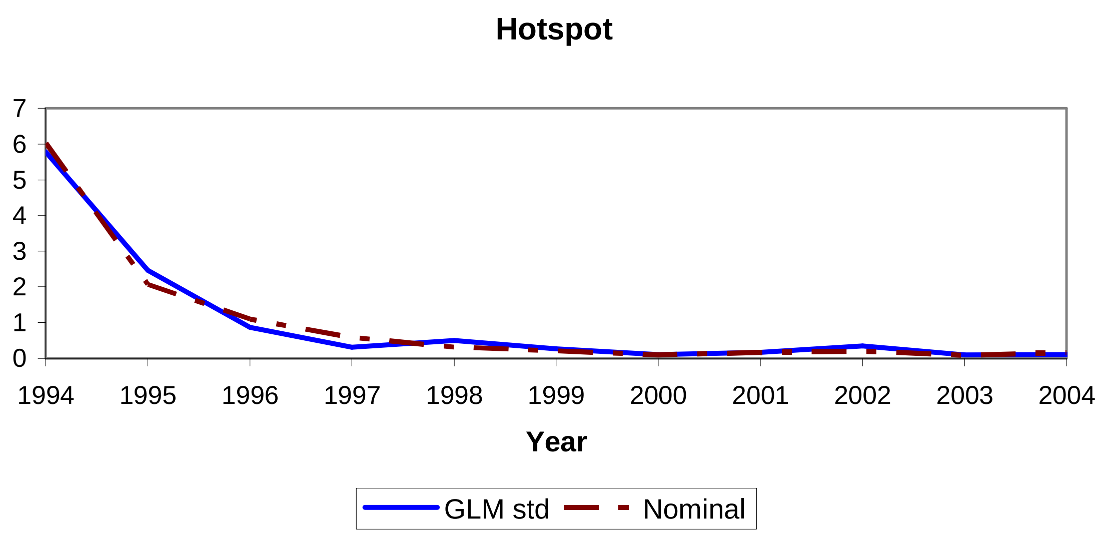


**Figure 4.** Index of abundance for the *Hotspot* aggregation (normalised to its mean over the eleven year period) for Namibian orange roughy obtained from fitting the delta-lognormal model assuming binomial errors for the proportion positive. For comparison, the nominal CPUE series is also shown.