Not to be cited without permission of the author(s)

New Zealand Fisheries Assessment Research Document 96/1

Catch per unit effort (CPUE) analysis of the Campbell Island Rise southern blue whiting (*Micromesistius australis*) trawl fishery from 1986 to 1993

T.D. Chatterton

MAF Fisheries Greta Point
Present address
Ministry of Fisheries
PO Box 1020
Wellington

January 1996

Ministry of Fisheries, Wellington

This series documents the scientific basis for stock assessments and fisheries management advice in New Zealand. It addresses the issues of the day in the current legislative context and in the time frames required. The documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Catch per unit effort (CPUE) analysis of the Campbell Island Rise southern blue whiting (*Micromesistius australis*) trawl fishery from 1986 to 1993

T.D. Chatterton

N.Z. Fisheries Assessment Research Document 96/1 26 p.

1. EXECUTIVE SUMMARY

Standardised CPUE analysis of the Campbell Island Rise southern blue whiting (SBW) trawl fishery from 1986 to 1993 is described using three separate models. The models were a lognormal linear (LNL) model, and two generalised linear models - a gamma log-link (GLL) model and a delta lognormal (DLN) model. Catch-effort data came from the fishery statistics database and included all tows (n = 6971) within an area south of 50° 30' S and north of 53° 30' S, and east of 168° 30' E and west of 172° 00' E that targeted SBW or caught SBW or both. A total of 17 variables, either continuous or categorical, was used in all 3 models. The measure of CPUE was catch (kg) per n. mile.

A stepwise multiple regression procedure was used to fit the LNL model. Five variables were entered into the model, the most significant being year. The overall trend of the relative year effects from 1986 to 1992 was downwards, though there were small increases in 1989 and 1991. The index showed a marked increase from 1992 (0.07) to 1993 (0.47). Examination of plots of residuals and observed versus expected values indicated a poor fit of the model to the data.

Seven variables were selected into the GLL model using a stepwise procedure. The trajectory of the relative year effects from 1986 to 1992 was similar to that from the LNL model except the downward trend was monotonic. Again the index showed a sharp increase from 1992 (0.24) to 1993 (0.70). Plots of residuals and observed versus expected values indicated a much stronger fit of this model to the data, than of the LNL model. The GLL model was relatively insensitive to possible errors in high CPUE values.

The CPUE index given by the DLN model was calculated by combining results from two separate analyses; a GLL model of all successful (non-zero catch) tows and a binomial model (similar in structure to the GLL model) of the proportion of zero catches in each year. The combined index was very similar to the GLL index.

The Middle Depths Fishery Assessment Working Group recommended that the GLL index be used to tune the Virtual Population Analysis of the Campbell Island Rise stock because it gave a better fit to the data and handled zero tows in a more plausible way than the LNL model, and was relatively insensitive to potential errors in the dataset. The DLN model looks promising but requires further development.

2. INTRODUCTION

The southern blue whiting (SBW) trawl fishery is concentrated mainly on the August and September spawning aggregations on the Pukaki Rise and Bounty Platform, and near the Campbell Islands (Annala 1994). During the 1980s most SBW landings were taken from the Campbell Island spawning grounds. However, over the last 3 years large catches have also come from Bounty Platform and Pukaki Rise (Annala 1994).

The 1993 fishery assessment of the Campbell Island stock (Hanchet 1993) used catch-at-age data from 1986 to 1992 in a Virtual Population Analysis (VPA). The VPA was tuned using total effort standardised on a Soviet Atlantic class CPUE index. However, concerns were raised by the Middle Depths Fishery Assessment Working Group that the estimates of current biomass derived from the VPA were imprecise and negatively biased because of problems with the final point of the CPUE index (Hanchet 1993). The group recommended that a standardised CPUE analysis be carried out using a linear model that modelled the year effect after excluding other effects such as season, vessel size, and nation.

This paper describes the calculation of standardised CPUE indices for the period 1986 to 1993 using a lognormal linear (LNL) model and two generalised linear models; a gamma distribution, log-link function (GLL) model, and a delta-lognormal (DLN) model of the type described by Vignaux (1994).

3. THE DATA

3.1 Source of catch-effort data

The catch-effort data came from the fishery statistics database managed by the Information Technology Directorate. Records from all commercial tows made in August and September from 1986 to 1993 in the area south of latitude 50° 30' S and east of longitude 168° 30' E and from all tows outside this region that targeted SBW, or caught SBW, or did both during the same period were extracted. This selection ensured that all tows made in the Campbell Island fishery that did not catch SBW ("zero catches") and tows either catching SBW, or targeting SBW, or doing both within the Campbell Island area but with wrongly recorded or mis-typed position coordinates were included in the dataset. The final boundaries used for the Campbell Island fishery enclosed an area south of 50° 30' S and north of 53° 30' S, and east of 168° 30' E and west of 172° 00 E (Figure 1). All tows remaining outside this area after error checking (*see* Section 3.3), were deleted.

3.2 Contents

The following information was available for each tow.

Vessel identification number
Estimated catch (kg) of SBW

Date
Time at start of tow
Time at end of tow
Duration of tow (h)
Speed of tow (knots)
Distance of tow (n. miles)
Latitude (decimal degrees) at start of tow
Longitude (decimal degrees) at start of tow
Depth of groundrope (m) at start of tow
Depth of bottom (m) at start of tow
Headline height of net (m)
Target species

3.3 Error checks and data editing

The error checking procedures used on these data were similar to those described by Vignaux (1993). Briefly, if the cause of the anomaly was clear the record was altered so that it was consistent with adjacent records and other information on the tow (e.g., gear type, position of tow, bottom depth). Otherwise, the original paper record (i.e., the Trawl, Catch, Effort and Processing Return) was examined to determine the cause of the error, or if the tow had been witnessed by a fisheries observer then their records were compared. If the error was still not resolved, the tow was deleted. A summary of types and number of errors found and the proportion corrected is presented in Table 1.

Table 1: Number of errors associated with the following factors and the proportion able to be corrected

Factor	No. of errors	% corrected
Groundrope depth	177	26
Tow position	49	100
Date of tow	48	100
Speed of tow	31	100
Headline height	41	100
Bottom depth	21	100
Duplicate records	24	92
Start/end time	16	31
Catch (very big)	15	100
Total	422	66

After examination of the dataset, all 10 tows carried out in less than 300 m or over 610 m bottom depth within the area were deleted. These tows typically had a zero or very small SBW catch and were often targeting different species.

At the start of the 1991 fishing season MAF changed the definitions of the "Start time" and "End time" of a tow to be the times that the net went into the water and the time that it left the water, respectively. The original definitions were the times that the net reached fishing depths and that hauling began. Vignaux (1993) found, however, that only

30% of vessels in the hoki fleet used the new definition in the 1991 and 1992 fishing years. To avoid a downwards bias to these years of hoki CPUE, Vignaux eliminated all data from companies which used the new definition. The vessel (and company) composition of the SBW and hoki fishery in any one year is very similar (S. Hanchet, NIWA Greta Point, pers. comm.), so all SBW data from companies using the new definition in 1991 and 1992 were likewise excluded. Vignaux (1993) found no evidence of bias in the 1991 and 1992 year indices for hoki CPUE as a result of eliminating data from these companies. Hanchet (1993) also found that inclusion or exclusion of these tows made very little difference to his unstandardised CPUE indices. At the start of the 1993 season MAF changed the start and end time definitions back to their original meaning.

All tows that were not targeting SBW but were made at the same time of year in the same area at similar bottom depths (between 300 and 600 m) and with similar gear to the SBW fishery were examined carefully. These tows (67 in total) were targeting species such as hoki (23), ling (40), and orange roughy (3). Of these tows, 21 were deleted on the basis that the vessel either did not catch any SBW in these tows, or did not target SBW at any stage during the season, or the tows occurred at the beginning or end of the SBW season. The remaining tows, although not targeting SBW, often caught large quantities of SBW (up to 50 138 kg), or occurred during the middle of the SBW season, or both.

3.4 Zero catches

Initial examination of the "un-groomed" CPUE (catch per n. mile) data showed that for all years combined 8.7% had zero catches. Having a dataset with such a large proportion of zeros presents two problems when using parametric techniques such as the lognormal linear model. Firstly, the statistical distribution of the dataset may be distorted to such an extent that the assumptions of normality and equality of variances, which are implicit in the valid use of such models, may not be satisfied even after data transformation. Secondly, because there is a direct logarithmic transformation of CPUE a small constant must be added to avoid taking the logarithm of zero. Similar analyses have shown, however, that results can be sensitive to the value of this arbitrary constant (Porch & Scott 1992, Vignaux 1993). (In the gamma and delta lognormal models, zero data values are treated in different ways, as described below, compared to the lognormal linear model so the proportion of zero catches is not as problematic.) In an effort to reduce the number of zeros, and therefore minimise the above problems when running the lognormal linear model, several steps were taken.

(1) All tows which were 30 min or less in duration and in which no SBW were caught were deleted. These short unsuccessful tows were assumed to result from problems with the gear during shooting of the net. This eliminated 31 (5%) of the zero catches.

(2) Where possible, zero catches were compared with equivalent observer data. A total of 34 (5.5%) of the zero catches were recorded as non-zero catches by observers. For these tows the observer recorded catch was entered instead of the original estimated zero catch.

(3) Tows with a net depth of less than 100 m and zero SBW catch were

deleted. This eliminated 1 (0.2%) of the zero tows.

The constant added to all CPUE values was 0.05 kg, giving (CPUE + 0.05). The final distribution for the logarithmic transformed data is shown in Figure 2.

3.5 Summary of tow data

Table 2 gives the number of vessels providing tow data, number of tows, and number of zero catches in each year in the final "groomed" dataset. The start position of all 6971 tows is shown in Figure 1.

Table 2: Summary of "groomed"Campbell Island SBW CPUE data

Year	'86	'87	'88	'89	'90	'91	'92	'93	Total
No. of vessels	15	17	15	23	33	28	33	20	
No. of tows	893	663	843	1008	1005	1057	1091	411	6971
Zero catches	41	35	60	47	78	39	204	44	548
(%)	(4.6)	(5.3)	(7.1)	(4.7)	(7.8)	(3.7)	(18.7)	(10.7)	(7.9)

3.6 Vessel data

The vessel data came from the vessel registration database and the Lloyds Register (Lloyds 1990). The following information was obtained for each vessel.

Nationality of vessel
Overall length (m)
Gross tonnage
Year built
Breadth of vessel (m)
Draught of vessel (m)
Company vessel was registered to in each of the years from 1986 to 1993

Processing capability and engine power were not used as variables because of the unreliability of this data. For example, in 1992 and 1993 some surimi vessels were not converting fish into surimi, but head and gutting instead. Also information from industry and observers indicated that some vessels had the ability to "gear-up" with extra power in excess of that registered, if required, through the use of auxiliary generators. The exact power capacity of some boats is, as yet, unknown. Until reliable data can be collected on these two variables they will be excluded from the CPUE analysis.

4. THE VARIABLES

The variables used in all three models were either continuous or categorical (Table 3). Categorical variables included year, tow positions, end time, start time, headline height, and vessel nation, and were included in the models with the constraint that the sum of the regression coefficients should be zero. Continuous variables were included up to a cubic power.

Table 3: Summary of variables used in CPUE models. cont = continuous variable, cat = categorical variable with a given number of categories

Variable	Type	Description
Year	cat 8	Year in which tow occurred
Tow position	cat 4	Position at start of tow (*see* text)
End time	cat 6	Time of end of tow in 4 h periods (eg, 0001–0400 h)
Start time	cat 6	Time of start of tow in 4 h periods (eg, 0001–0400h)
Depth net	cont	Depth (m) of groundrope of net at start of tow
Depth bottom	cont	Depth (m) of bottom at start of tow
Season	cont	Day relative to date at which 10% of fish spawning for that year
Headline height	cat 4	Height (m) of headline from groundrope (1) 0–15 m, (2) 16–40, (3) 41–75, (4) 76+
Target species	cat 3	Species targeted by tow
Vessel nation	cat 6	Country of origin of the vessel
Vessel length	cont	Length (m) overall of the vessel
Vessel tonnage	cont	Gross tonnage of the vessel
Vessel year built	cont	Year the vessel was built
Vessel breadth	cont	Breadth (m) of the vessel
Vessel draught	cont	Draught (m) of the vessel
Vessel l*b*d	cont	Vessel length * breadth * draught in cubic metres
Vessel company	cat 21	Company the vessel was fishing for in that year

The year of the tow is included as a categorical value which allows the regression coefficients of each year to vary independently. Relative year effects can be calculated from these regression coefficients (*after* Doonan 1991):

$$\hat{A}_i = \exp(\hat{Y}_i - \hat{Y}_{1986})$$

where $\hat{Y}_i$ is the year coefficient for year i, $\hat{Y}_{1986}$ is the year coefficient for 1986, and $\hat{A}_i$ is the estimate of the year effect in year i relative to the year effect in 1986. The variance of this estimate is;

$$S^2_{\hat{A}i} = \hat{A}_i^{\,2}\exp(\sigma^2)[\exp(\sigma^2) - 1]$$

where;

$$\sigma^2 = Var(\hat{Y}_i) + Var(\hat{Y}_{1986}) - 2Cov(\hat{Y}_i, \hat{Y}_{1986})$$

The relative year effects represent the changes in CPUE over time, all other effects having been taken into account, and therefore represent an index of abundance.

Tow position is a categorical variable in which the start position of each tow has been assigned to one of four area strata (Figure 3). The strata are based on the typical fishing pattern of the fleet and the main spawning sites within the Campbell Island Rise fishery: 1 is a northern area fished at the beginning of the season; 2 is the northern spawning ground; 3 is an area often fished at the end of the spawning season; and 4 is the southern spawning ground. Plots of CPUE against each stratum showed that there were apparent differences in catch per unit effort between areas (Figure 4a).

The end time and start time of each tow were separated into 4 h periods (i.e., 0001 to 0400 h, 0401 to 0800 h, and so on), thus creating categorical variables, after preliminary analysis had shown a possible diel pattern in CPUE. A plot of log(CPUE + 0.05) against end time is shown as an example (Figure 4b).

Season is a continuous variable based on the day relative to the date at which 10% of the fish were actively spawning for the first time that year. The season midpoint (i.e., 10% of fish spawning) has a value of 0, the days before the midpoint = -1, -2, -3, ..., and the days after the midpoint = +1, +2, +3, etc. Gonad staging data from fisheries observers were used to determine the season midpoint (Hanchet 1993). CPUE reaches a peak about 2 days before the middle of the season (Figure 4c), then declines quite sharply before showing a steady increase from about day +9. The late season increase may coincide with the time of second spawning which typically follows the first spawning by about 2 weeks, but is a much smaller event (Hanchet 1993).

Headline height was separated into four groups (Table 3) based on preliminary analysis of CPUE data and type of net used for the tow. CPUE appeared to increase with increasing headline height before reaching a plateau at between 40 and 75 m (Figure 4d). Catch effort decreased as net openings exceeded 75 m. The 0 – 15 m category was included as the likely range of headline heights for bottom trawl gear. Net openings over 15 m probably came from midwater trawl gear.

5. THE MODELS

The measure of CPUE used in all models was catch (kg) per n. mile. Other studies (Doonan 1991, Hanchet 1991, Vignaux 1992), have shown that catch per nautical mile gives a more consistent measure of effort than catch per tow or catch per day which can be confounded by factors such as change in length of tow time between years and processing time limitations.

5.1 Lognormal linear (LNL) model

The LNL model was similar to the multiple regression model described by Doonan (1991) and Vignaux (1992). The model was of the form:

$$\log(CPUE_t + 0.05) = M + Y_{it} + A_{jt} + B_{kt} + D_{lt} + \ldots + \varepsilon_t$$

or, equivalently

$$(CPUE_t + 0.05) = \exp(M) * \exp(Y_{it}) * \exp(A_{jt}) * \exp(B_{kt}) * \exp(D_{lt}) * \ldots * \exp(\varepsilon_t)$$

where $(CPUE_t + 0.05)$ is the catch per unit effort for a particular tow t, M is an overall mean for $\log(CPUE_t + 0.05)$, Y_{it} is the effect on $\log(CPUE_t + 0.05)$ of tow t being in the ith year, A_{jt} is the effect of variable A having value j, B_{kt} is the effect of variable B having value k, D_{lt} is the effect of variable D having value l, and so on. ε_t is the error in $\log(CPUE_t + 0.05)$, with zero mean and independently normally distributed.

A stepwise procedure was used in constructing the model. At the first iteration all main effects were regressed individually against the CPUE estimate and the one explaining the most variation, that is giving the highest R^2, was selected as the first variable in the model. In the second iteration, all main effects plus the first selected variable were again fitted separately and the one giving the most improvement in R^2 was added to the model. This procedure continued until the inclusion of an additional variable in the model led to a less than 1% improvement in R^2.

Five variables were entered into the model (Table 4), with the most significant being year. The overall trend of the relative year effects from 1986 to 1992 was downwards, although there were small increases in 1989 and 1991 (Table 5, Figures 5 and 6). The index showed a marked increase from 1992 (0.07) to 1993 (0.47). The index appears to be influenced by the proportion of zero tows in each year: when the percentage of zero catches increases the index falls, and vice versa.

The pattern of the residuals calculated from the final model was non-random (Figure 7a) and a plot of observed CPUE values against fitted values showed a poor correlation, especially for zero data values (Figure 7b). These diagnostic features indicate a weak fit of the linear model to the log transformed data.

Table 4: Lognormal linear (LNL) model. Choice of variables, in order of importance, in regression against log(CPUE + 0.05) for data from 1986 to 1993. cat = categorical variable

	R^2 at iteration					
Variable	1	2	3	4	5	6
Year	7.86					
Tow position	1.97	9.21				
End time	1.25	9.15	10.49			
Season	2.56	9.05	10.23	11.55		
Vessel nation	1.72	8.72	10.09	11.40	12.55	
Bottom depth	1.53	8.86	9.95	11.22	12.29	13.28
Target species	0.30	8.29	9.67	10.97	12.03	13.08
Vessel breadth	0.73	8.50	9.87	11.19	12.25	12.99
Net depth	0.67	8.31	9.55	10.94	11.88	12.87
Start time	0.95	8.88	10.28	10.66	11.72	12.73
Year vessel built	1.20	8.29	9.62	10.91	11.94	12.74
Vessel l*b*d	1.00	8.53	9.91	11.23	12.30	12.97
Vessel draught	0.94	8.46	9.85	11.16	12.25	12.95
Vessel length	0.20	8.41	9.74	11.09	12.15	12.87
Headline height	0.17	7.98	9.37	10.69	11.69	12.70
Vessel tonnage	0.24	8.35	9.66	11.00	12.08	12.88
Vessel company	1.30	7.88	9.24	10.53	11.56	12.56
Polynomial	cat	cat	cat	3	cat	
Improvement (%)		1.35	1.28	1.06	1.00	0.73

Table 5: Relative year effects $\hat{A}_i$ from the lognormal linear (LNL) model of CPUE for the period 1986 to 1993. R are the regression coefficients of the year variable, s.e. are the standard errors of these coefficients in the regression, and Cov. is the covariance between the coefficients and the coefficient of the base year (1986). $\hat{A}_i$ is the relative year effect and $s_{\hat{A}i}$ is the standard error of the year effect

Year	R	s.e	Cov.	$\hat{A}_i$	$s_{\hat{A}i}$
1986	0.868	0.094	0.009	1.000	0.000
1987	0.354	0.106	-0.001	0.598	0.090
1988	0.196	0.096	-0.001	0.511	0.071
1989	0.537	0.088	-0.001	0.718	0.097
1990	-0.361	0.096	-0.002	0.293	0.043
1991	0.114	0.119	-0.002	0.470	0.080
1992	-1.824	0.093	-0.001	0.068	0.009
1993	0.115	0.134	0.001	0.471	0.075

5.2 Gamma, with log-link function (GLL) model

The GLL model procedure has been described in detail by Vignaux (1994). Briefly, it attempts to fit a linear model to the CPUE data in its original form. This is in contrast to the usual parametric procedure of fitting a model to data that has been transformed (e.g., log(CPUE + 0.05)) to meet assumptions of normality and equality of variances, as in the LNL model. Initially, the raw data are transformed using a logarithmic function and a linear model is calculated. A set of linear predicted (expected) values is then generated and back-transformed using a link function. The difference between the expected and observed data values is then modelled using a distribution similar to the known nature of the data, in this instance a gamma error distribution (variance proportional to the square of the mean), and a residual deviance calculated. The model attempts to minimise the residual deviance by changing the parameters of the linear model. Because expected and observed values are compared in untransformed space, zero data values can be left in the model and no arbitrary replacement is required.

A stepwise procedure similar to that used for the LNL model was used to calculate the full GLL model. The procedure continued until the addition of an extra variable led to a less than 1% improvement in the deviance explained compared to the null deviance, which is analogous to the sum of squares of the errors in a linear model.

Seven variables were selected into the GLL model (Table 6). Year was again the most significant, followed by vessel nation, tow position, end time, net depth, vessel tonnage and season. Net depth and vessel tonnage were the two extra variables to enter this model compared to the LNL model. The trajectory of the relative year effects from 1986 to 1992 was similar to that shown by the LNL model except the downward trend was monotonic (Table 7, Figures 8, 9, and 10). Again the index showed a sharp increase from 1992 to 1993.

Plots of residuals (Figure 7c) and observed versus fitted values (Figure 7d) indicated that the GLL model fitted the CPUE data better than the LNL model. The residuals showed a more random distribution, and there was a better correlation between the expected and observed values.

Because the GLL model is fitted in untransformed space the indices scale as the arithmetic mean of the CPUE values rather than the geometric means, as with the LNL model. As a result, the GLL model can be more sensitive to errors in the data in the high CPUE tail of the distribution (Vignaux 1994). To test for this, a sensitivity analysis was done in which all CPUE values over 10 000 (about 1% of data) were removed from the dataset and both the LNL and GLL models recalculated. Comparison of the indices from the full dataset and from the dataset less the high extremes showed greater differences for the GLL model than the LNL model, but the overall pattern remained the same (Figure 11). The relatively small change in the indices suggests that, for this dataset, the GLL model may not be very sensitive to any remaining errors in high catch effort values.

Table 6: Gamma, log link function (GLL) model. Choice of variables, in order of importance, in model of CPUE data from 1986 to 1993. cat = categorical variable

	Residual deviance at iteration						
Variable	1	2	3	4	5	6	7
Year	8431						
Vessel nation	8533	8144					
Tow position	8686	8240	7965				
End time	8717	8256	7968	7776			
Net depth	8708	8292	8007	7806	7547		
Vessel tonnage	8770	8178	7995	7831	7614	7388	
Season	8640	8297	7998	7802	7602	7402	7250
Headline height	8793	8295	8010	7825	7621	7424	7298
Start time	8726	8255	7973	7777	7731	7485	7323
Year vessel built	8845	8376	8122	7954	7762	7530	7366
Vessel breadth	8687	8169	8000	7826	7619	7390	7372
Vessel length	8736	8179	8023	7859	7655	7414	7380
Bottom depth	8831	8434	8143	7946	7764	7540	7380
Vessel l*b*d	8674	8183	8026	7853	7655	7437	7382
Vessel draught	8689	8228	8068	7892	7698	7486	7366
Vessel company	8726	8422	8135	7955	7758	7533	7387
Target species	8943	8497	8197	8020	7825	7601	7443
Polynomial	cat	cat	cat	cat	3	3	3
Improvement		287	179	189	229	160	137

Table 7: Relative year effects $\hat{A}_i$ from the gamma, log-link (GLL) model of CPUE for the period 1986 to 1993. R are the regression coefficients of the year variable, s.e. are the standard errors of these coefficients in the regression, and Cov. is the covariance between the coefficients and the coefficient of the base year (1986). $\hat{A}_i$ is the relative year effect and $s_{\hat{A}i}$ is the standard error of the year effect

Year	R	s.e	Cov.	$\hat{A}_i$	$s_{\hat{A}i}$
1986	0.652	0.051	0.002570	1.000	0.000
1987	0.186	0.057	-0.000421	0.627	0.049
1988	0.106	0.051	-0.000006	0.579	0.042
1989	-0.018	0.058	-0.000588	0.512	0.041
1990	-0.121	0.051	-0.000358	0.462	0.036
1991	-0.326	0.063	-0.000623	0.376	0.033
1992	-0.781	0.053	-0.000526	0.238	0.019
1993	0.301	0.076	0.000518	0.704	0.061

5.3 Delta lognormal (DLN) model

A second generalised linear model, the DLN model (Vignaux 1994), was also used to determine annual changes in CPUE. This model has the same advantage as the GLL model in that it avoids introducing an arbitrary constant to cope with zero catches. The model is made up of two parts. A GLL model of the catch rate of successful (non-zero) tows and a binomial model of the proportion of zero catches. The GLL model determines how the catch rate of non-zero tows has changed from year to year, while the binomial model determines how the success rate (i.e., presence/absence of SBW in tows) has changed from year to year. A combined index is calculated in which the year effects from the GLL model are combined with the year effects from the binomial model.

5.3.1 GLL model of catch rate of successful tows

The catch rate of successful tows, with all zero tows excluded, was modelled using a gamma distribution and a log link function. Seven variables were included in the model (Table 8). These variables were identical to those included in the GLL model for all tows (*see* Table 6), and they entered the model in the same order. The CPUE indices calculated for both models were very similar (Table 9).

5.3.2 Binomial model of proportion of zero catches

A binomial model was used to model the proportion of zero catches. The response variable was 1 if the catch was zero, and 0 otherwise. Four variables were included in the model (Table 10) and the index calculated showed the same pattern as the proportion of zero tows in each year, relative to 1986 (*see* Table 9).

5.3.3 Combining the GLL and binomial models

The method for combining the two indices is described in full in appendix 3 of

Vignaux (1994). The combined index differs little from the one given by the GLL model (*see* Table 9, Figure 10). The index decreased montonically from 1986 to a low of 0.23 in 1992 and then increased sharply to 0.69 in 1993.

Table 8: Delta lognormal (DLN) model. Choice of variables, in order of importance, in gamma, log link function model run on non-zero tows only. CPUE data from 1986 to 1993

	Residual deviance at iteration							
Variable	1	2	3	4	5	6	7	8
Year	9464							
Vessel nation	9580	9174						
Tow position	9734	9269	8992					
End time	9767	9286	8996	8806				
Net depth	9759	9323	9035	8832	8571			
Vessel tonnage	9821	9210	9023	8856	8642	8411		
Season	9690	9328	9025	8825	8622	8423	8271	
Headline height	9844	9327	9039	8853	8644	8448	8320	8193
Target species	9879	9413	9110	8930	8733	8509	8351	8209
Start time	9777	9287	9001	8802	8755	8509	8346	8209
Year vessel built	9894	9407	9150	8979	8785	8553	8388	8243
Vessel breadth	9737	9199	9028	8851	8646	8413	8395	8258
Vessel length	9787	9211	9051	8884	8677	8447	8402	8259
Bottom depth	9875	9461	9166	8968	8784	8559	8398	8261
Vessel l*b*d	9723	9213	9054	8879	8678	8460	8405	8266
Vessel draught	9741	9258	9097	8917	8721	8510	8408	8268
Vessel company	9776	9454	9164	8981	8781	8556	8410	8271
Percent improvement		2.9	1.9	1.9	2.3	1.6	1.5	0.7

Table 9: The proportion of zero tows, the proportion relative to the 1986 value, the indices from the GLL model, the no zeroes GLL model, the binomial model and the DLN model. *s.e.* = standard error

	Proportion of zeroes		GLL model	GLL model	Binomial	DLN
Year	Raw	Relative	All tows	No zeroes	model	model (*s.e.*)
1986	0.046	1.00	1.00	1.00	1.00	1.00 (0.05)
1987	0.053	1.15	0.63	0.65	1.60	0.63 (0.04)
1988	0.071	1.54	0.58	0.61	1.79	0.58 (0.04)
1989	0.047	1.02	0.51	0.52	1.29	0.52 (0.04)
1990	0.078	1.70	0.46	0.51	2.28	0.48 (0.04)
1991	0.037	0.80	0.38	0.38	1.05	0.37 (0.03)
1992	0.187	4.07	0.24	0.29	7.12	0.23 (0.02)
1993	0.107	2.33	0.70	0.75	3.04	0.69 (0.07)

Table 10: Delta lognormal (DLN) model. Choice of variables, in order of importance, in binomial model. CPUE data from 1986 to 1993

	Residual deviance at iteration				
Variable	1	2	3	4	5
Year	3633				
Bottom depth	3776	3583			
Vessel nation	3791	3598	3539		
Season	3804	3602	3553	3502	
End time	3813	3608	3557	3512	3477
Vessel draught	3797	3601	3543	3522	3484
Tow position	3798	3598	3569	3526	3487
Vessel l*b*d	3790	3599	3540	3524	3488
Vessel length	3822	3618	3564	3524	3489
Vessel breadth	3805	3603	3545	3525	3489
Year vessel built	3803	3616	3564	3527	3491
Start time	3829	3622	3571	3528	3492
Target species	3830	3622	3572	3528	3492
Vessel tonnage	3827	3625	3572	3532	3496
Headline height	3826	3618	3568	3535	3498
Vessel client	3810	3632	3581	3535	3500
Net depth	3828	3622	3575	3535	3500
Percent improvement		1.3	1.2	1.0	0.7

6. DISCUSSION

This analysis of CPUE in the Campbell Island Rise SBW fishery gives a number of improvements compared with that done by Hanchet (1993). Firstly, all data from all vessel types and nations were used; secondly, zero tows were included; and lastly, the CPUE analysis was standardised which allowed the relative year effects to be modelled after removing the effect of other significant variables.

After evaluating the three different models, the Middle Depths Fishery Assessment Working Group recommended that the GLL model index be used to tune the VPA model because it

- gives a much better fit to the CPUE data than does the LNL model;
- deals with zero tows in a more plausible way;
- was relatively insensitive to potential errors in the dataset; and
- has become widely used in fisheries assessment and its methodology is well understood (Porch & Scott 1992). The DLN model is a more recent generalised linear model that has not yet been fully developed. The Working Group did suggest, however, that the sensitivity of the VPA analysis to the DLN CPUE index be tested.

There are some improvements that could be considered for next year's CPUE analysis. The effect of interactions on the various models could be examined, though interpretation of such effects can be difficult (Vignaux 1994). Inclusion of a power variable for vessel may also be useful, if accurate data can be obtained. Also, further developments of the DLN model may eventually make this a more appropriate model to use. Given the increasing proportion of zero tows occurring in the fishery, such a model is intuitively more appealing because it determines the change in annual CPUE by separate analysis of the successful tows and the proportion of unsuccessful tows.

7. ACKNOWLEDGMENTS

I thank Marianne Vignaux for her generous support, advice, and mathematical assistance with this analysis; Stuart Hanchet for his assistance and advice; and the Middle Depths Fishery Assessment Working Group for useful discussions.

8. REFERENCES

Annala, J.H. (Comp.) 1994: Southern blue whiting (SBW). *In* Report from the Fishery Assessment Plenary, May 1994: stock assessments and yield estimates, pp. 217–225. (Unpublished report held in MAF Fisheries Greta Point library, Wellington).

Doonan, I. 1991: Orange roughy fishery assessment, CPUE analysis - linear regression, NE Chatham Rise 1991. New Zealand Fisheries Assessment Research Document 91/9. 31 p.

Hanchet, S.M. 1991: Southern blue whiting fishery assessment for the 1991–92 fishing year. New Zealand Fisheries Assessment Research Document 91/7. 48 p.

Hanchet, S.M. 1993: Southern blue whiting fishery assessment for the 1993–94 fishing year. New Zealand Fisheries Assessment Research Document 93/17. 56 p.

Lloyds, 1990: Lloyds Register of Shipping. Lloyds, London 3 Vols.

Porch, C.E. and Scott, G.P. (1992): A numerical evaluation of GLM methods for indices of abundance from West Atlantic bluefin tuna catch per trip data when a high proportion of trips are unsuccessful. ICCAT Working Document, SCRS 92/93-75

Vignaux, M. 1992: Catch per unit of effort (CPUE) analysis of the hoki fishery. New Zealand Fisheries Assessment Research Document 92/14. 23 p.

Vignaux, M. 1993: Catch per unit of effort (CPUE) analysis of the hoki fishery, 1987-92. New Zealand Fisheries Assessment Research Document 93/14. 23 p.

Vignaux, M. 1994: Catch per unit of effort (CPUE) analysis of west coast South Island and Cook Strait spawning hoki fisheries, 1987–93. New Zealand Fisheries Assessment Research Document 94/11. 23 p.

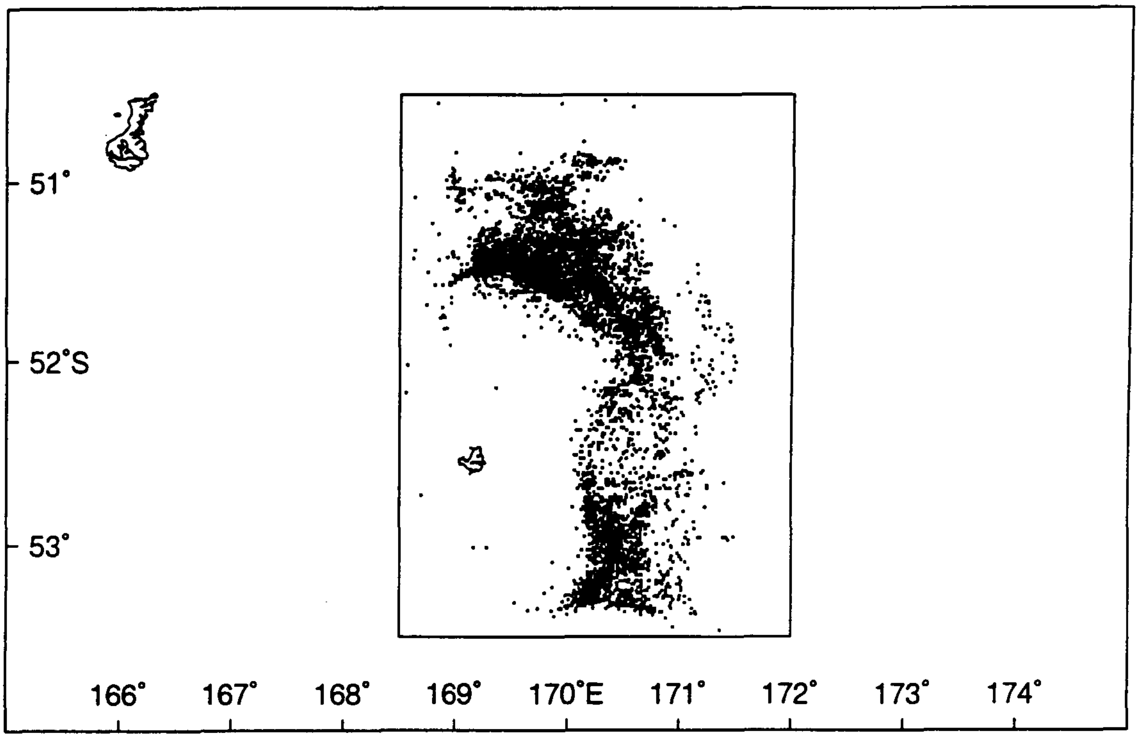

Figure 1: Map of tows used in CPUE analysis, 1986 - 1993.

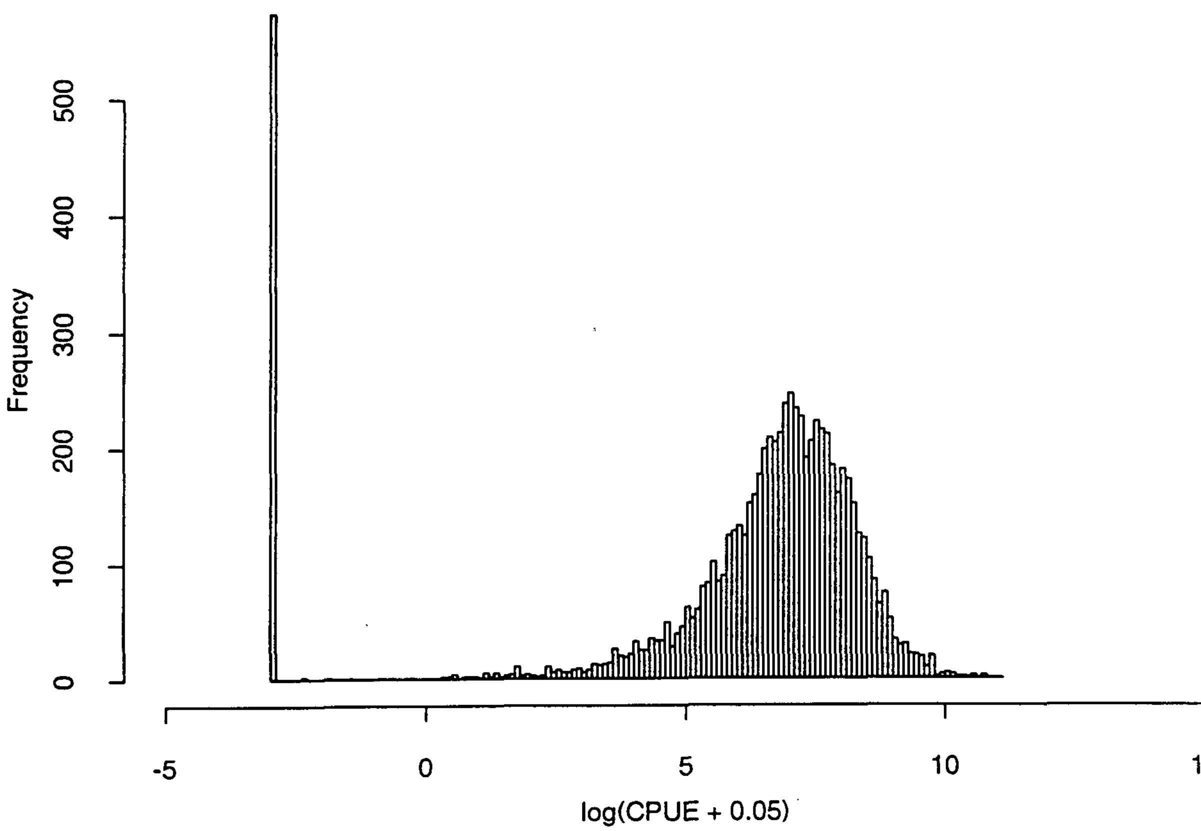

Figure 2: Histogram of log(CPUE) with the arbitary zero replacement catch set to 0.05 kg.

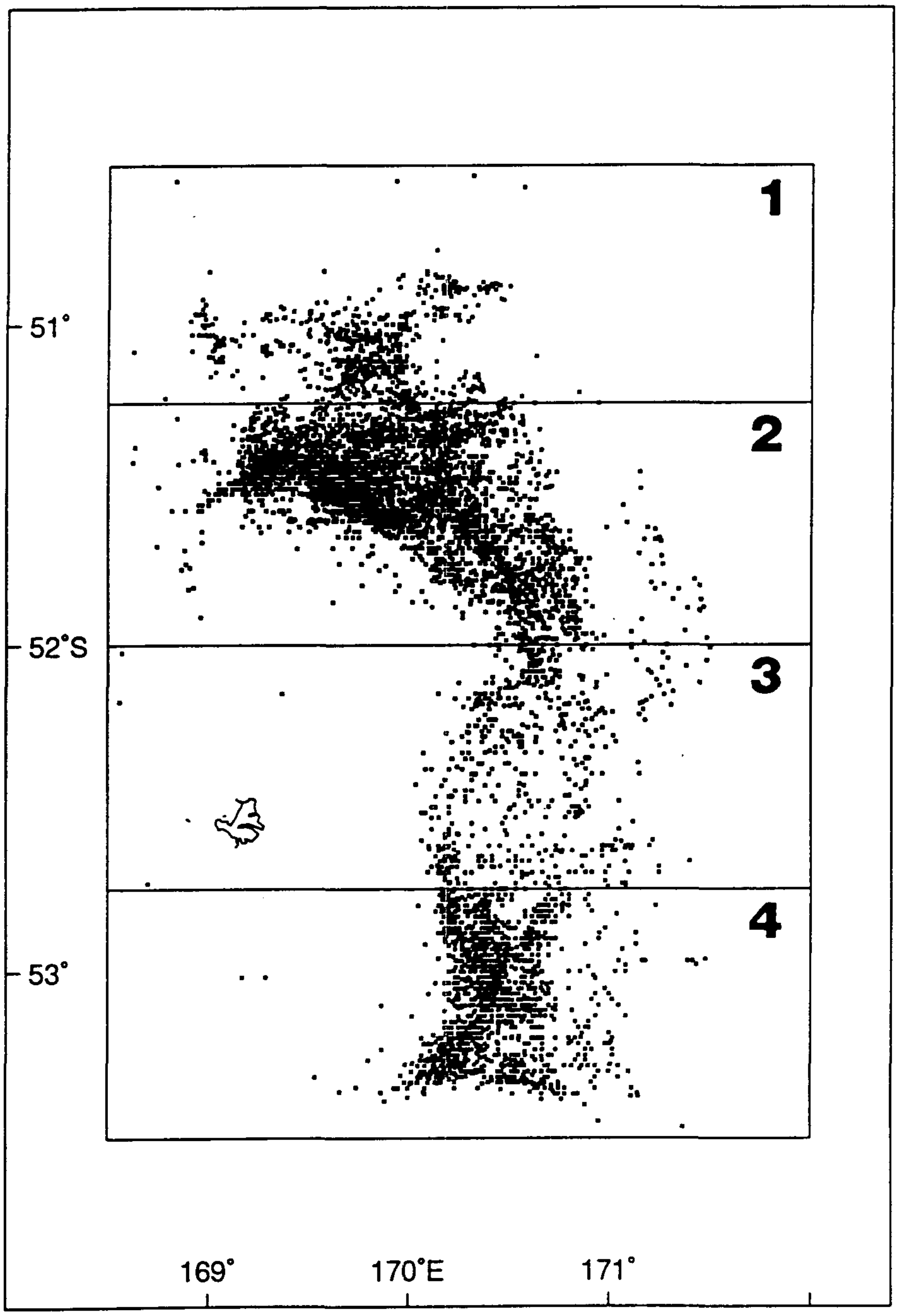

Figure 3: Start positions of tows and area strata used in the CPUE analysis, 1986 - 93.

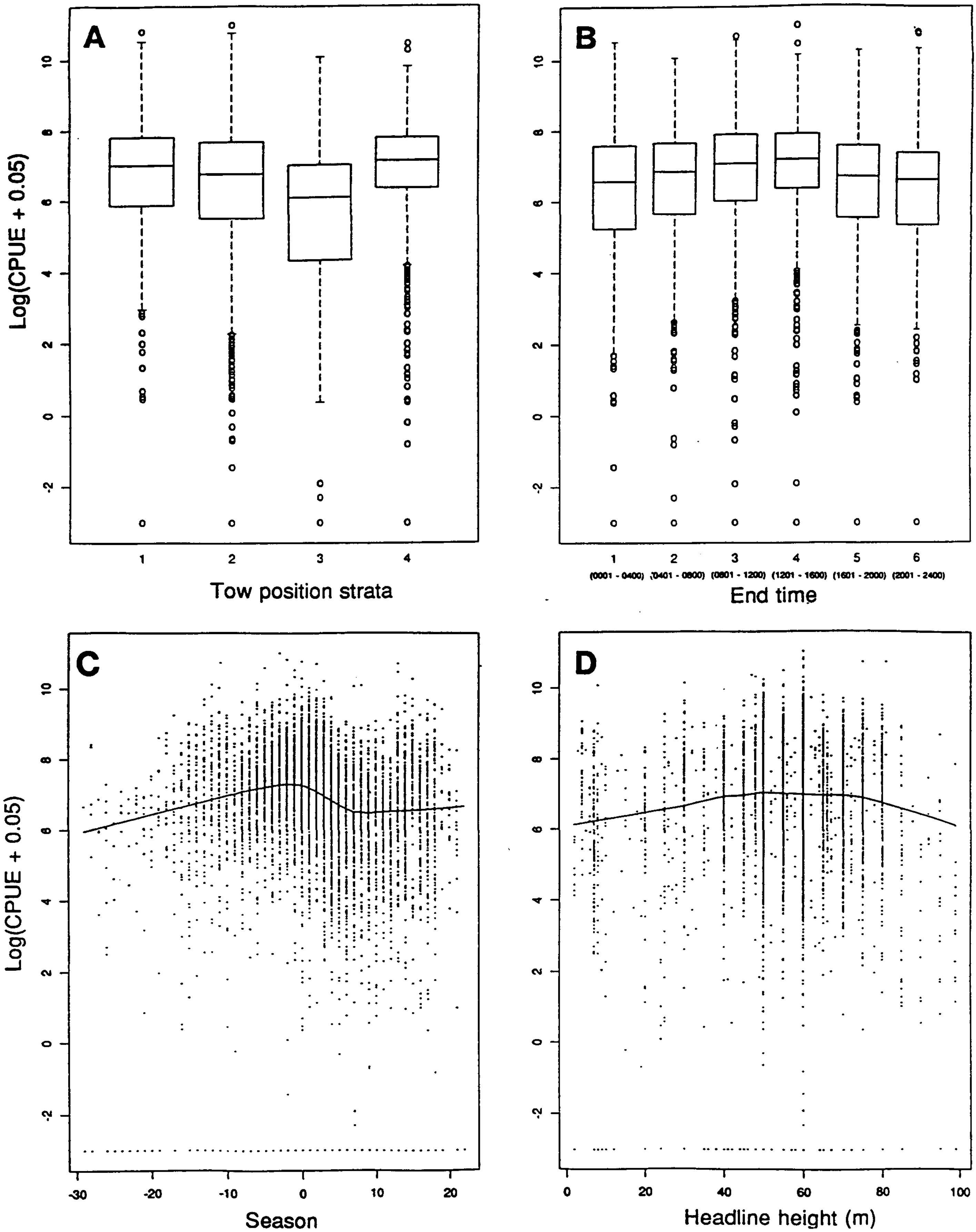

Figure 4. Plots of log(CPUE + 0.05) against; (A) tow position strata, (B) end time of tow, (C) season, and (D) headline height (m). Plots (C) and (D) have a lowess curve fitted to the data. *See* text for details of each variable. Box plots show median, quartiles, and standard range of data (1.5*(inter-quartile range).

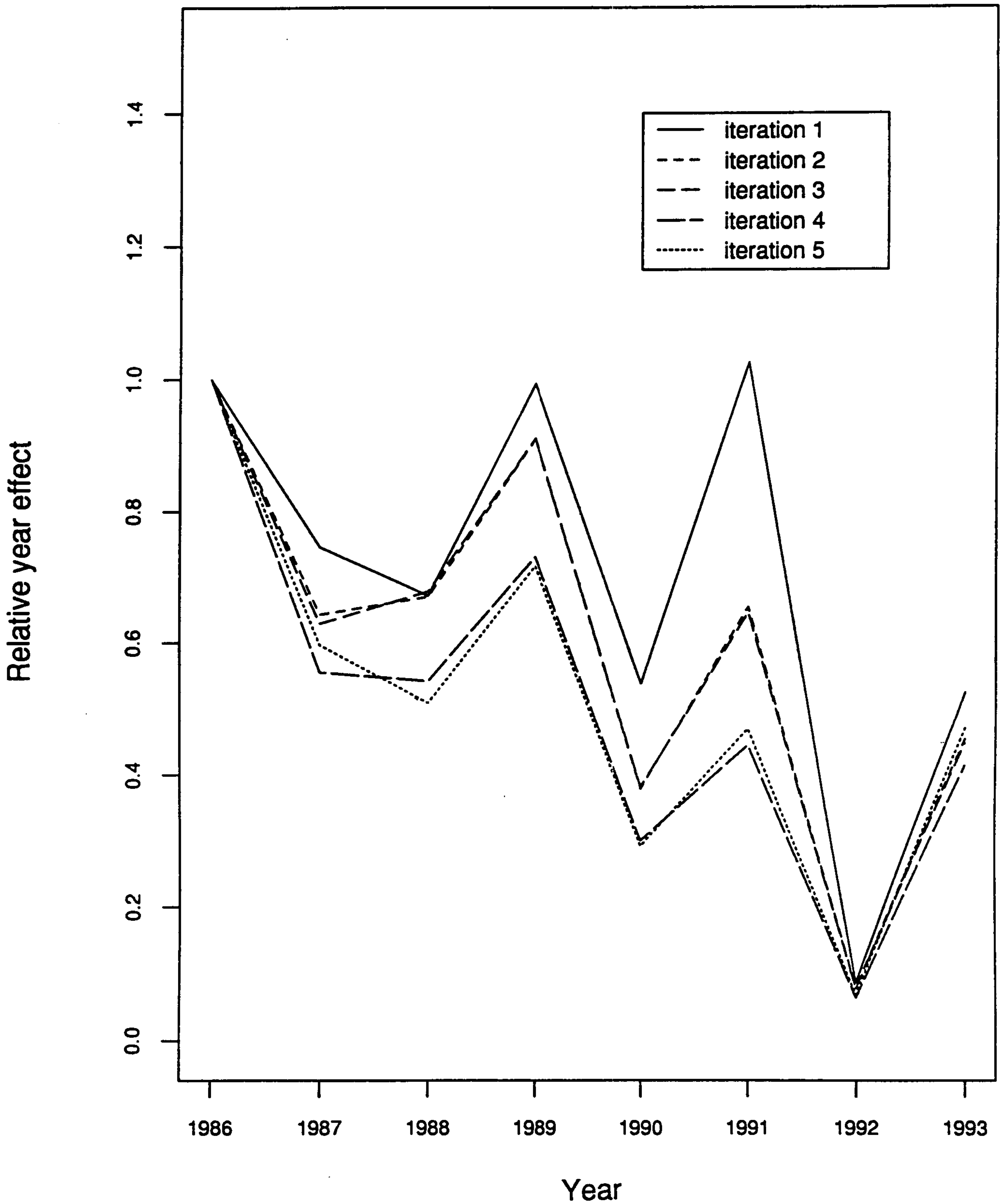

Figure 5. Relative year effects estimated from the lognormal linear (LNL) model iterations 1 to 5, each iteration including an extra variable.

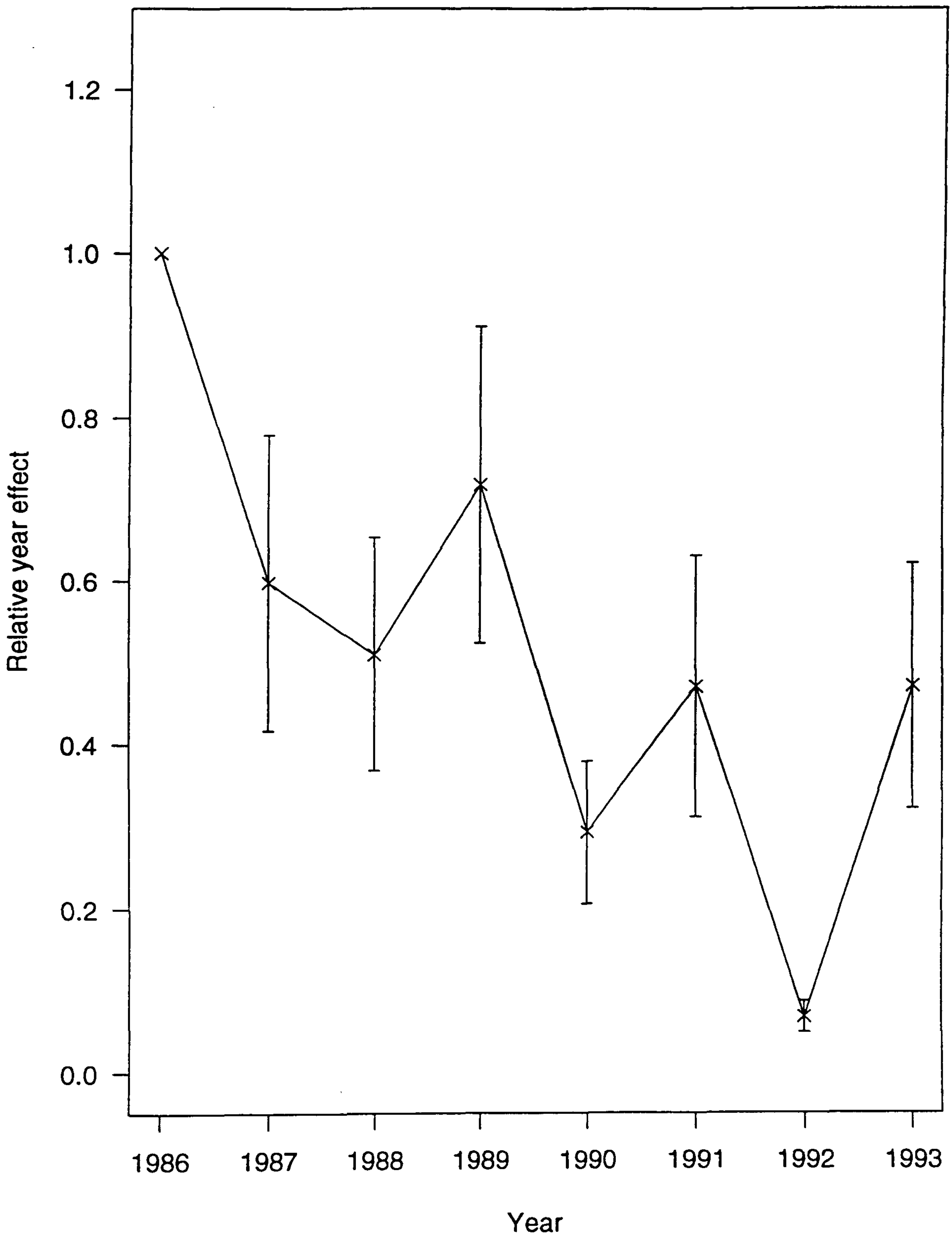

Figure 6. Lognormal linear (LNL) model relative year effects $\pm 2s_{\hat{\lambda}_j}$.

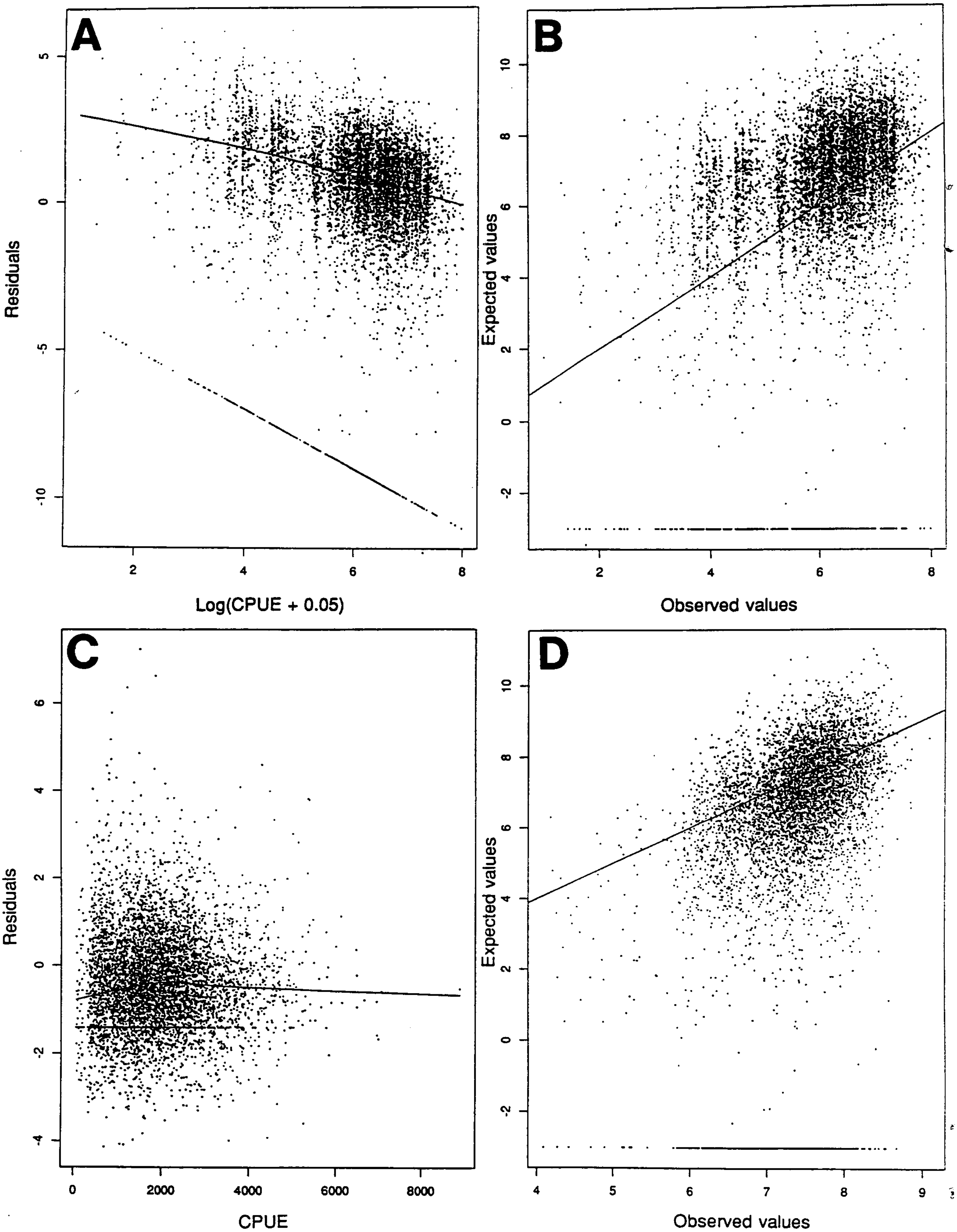

Figure 7. Plots of residuals and observed against expected values. 7a; plot of residuals from lognormal linear (LNL) model, with LOWESS curve fitted, 7b; plot of observed versus expected values from LNL model, with $Y = Y_i$ line fitted, 7c; plot of residuals from gamma, log-link function (GLL) model, with LOWESS curve fitted, 7d; plot of observed versus expected values from GLL model, with $Y = Y_i$ line fitted.

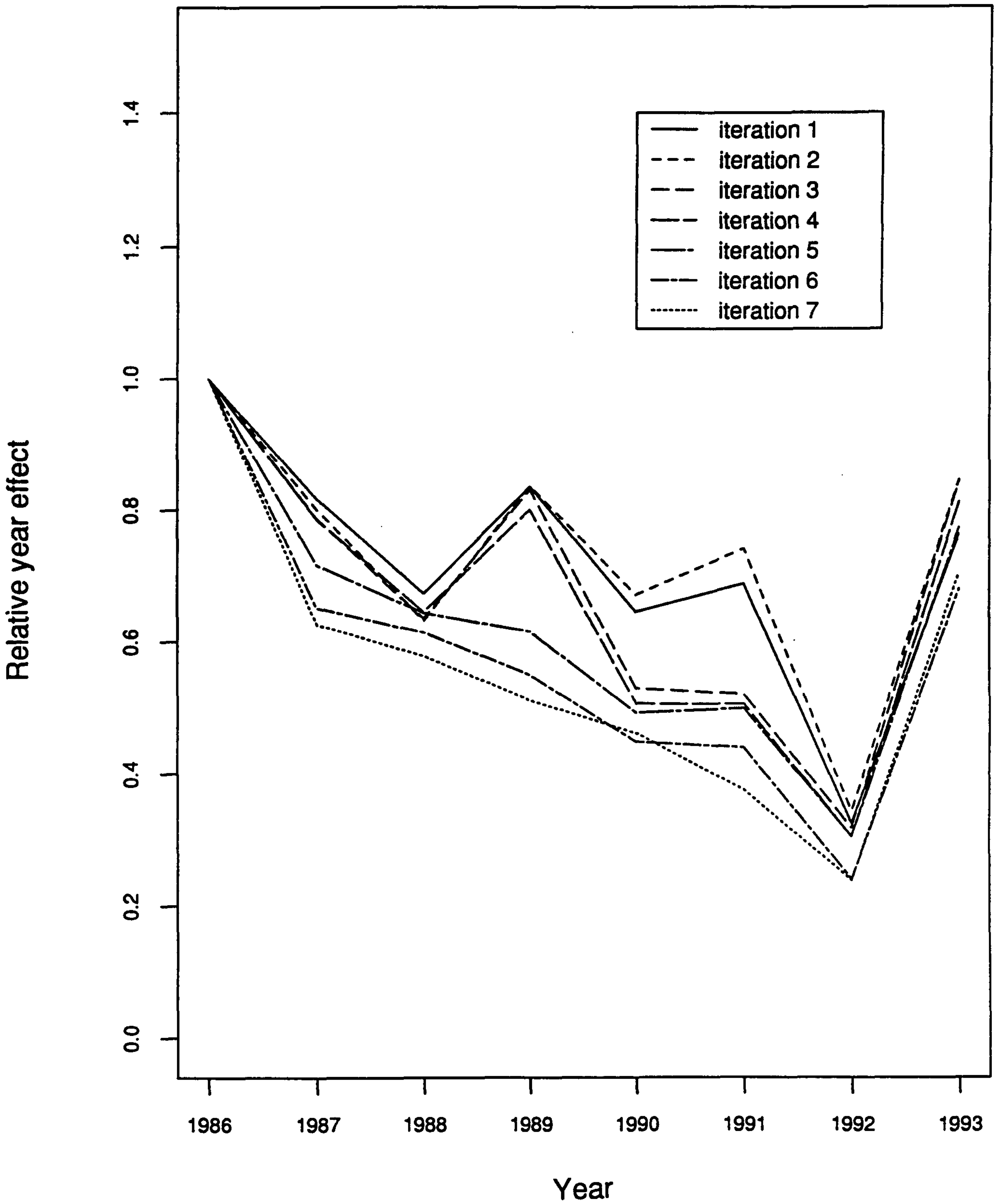

Figure 8. Relative year effects estimated from the gamma, log-link (GLL) model iterations 1 to 7, each iteration including an extra variable.

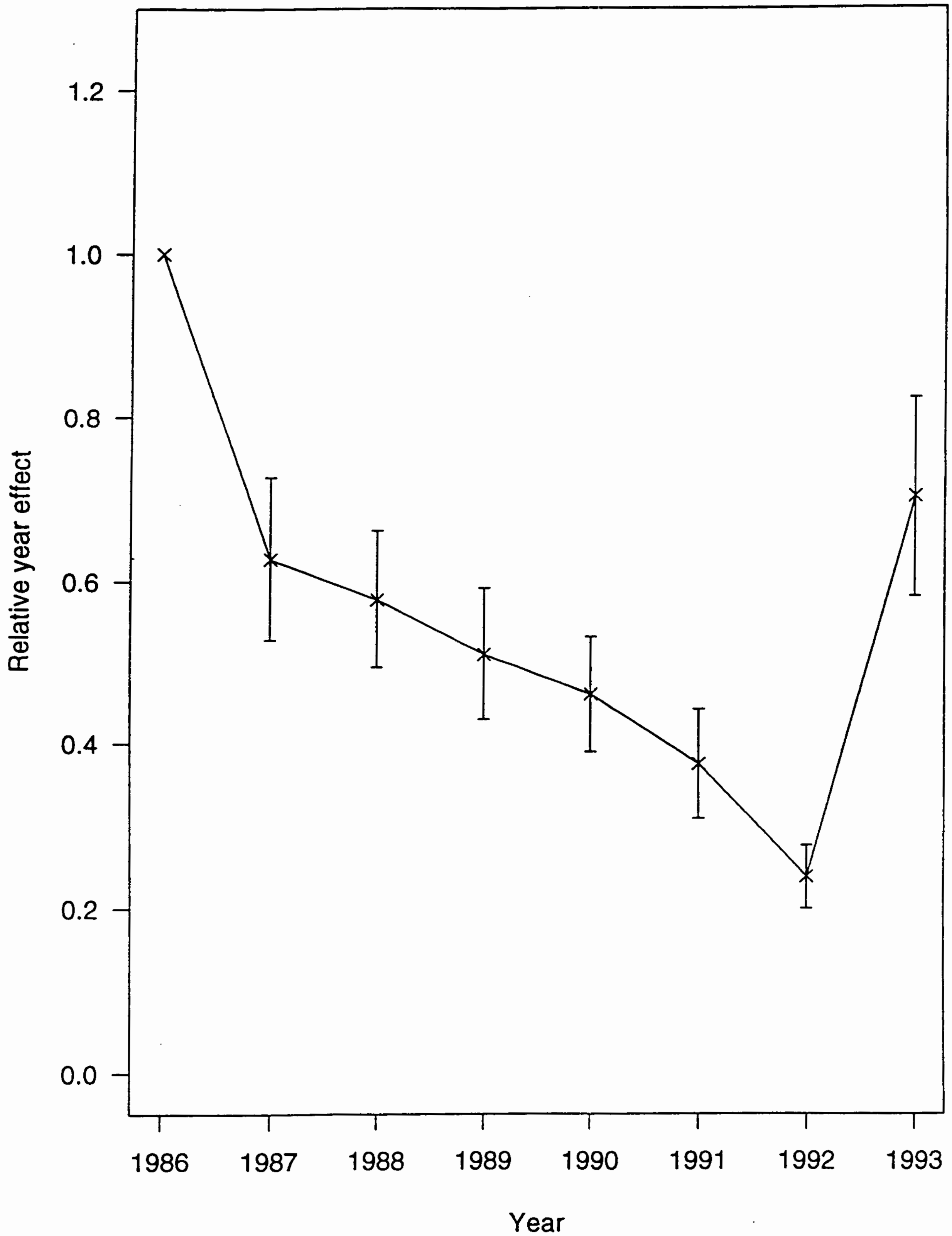

Figure 9. Gamma, log-link (GLL) model relative year effects $\pm\ 2s_{\hat{A}j}$.

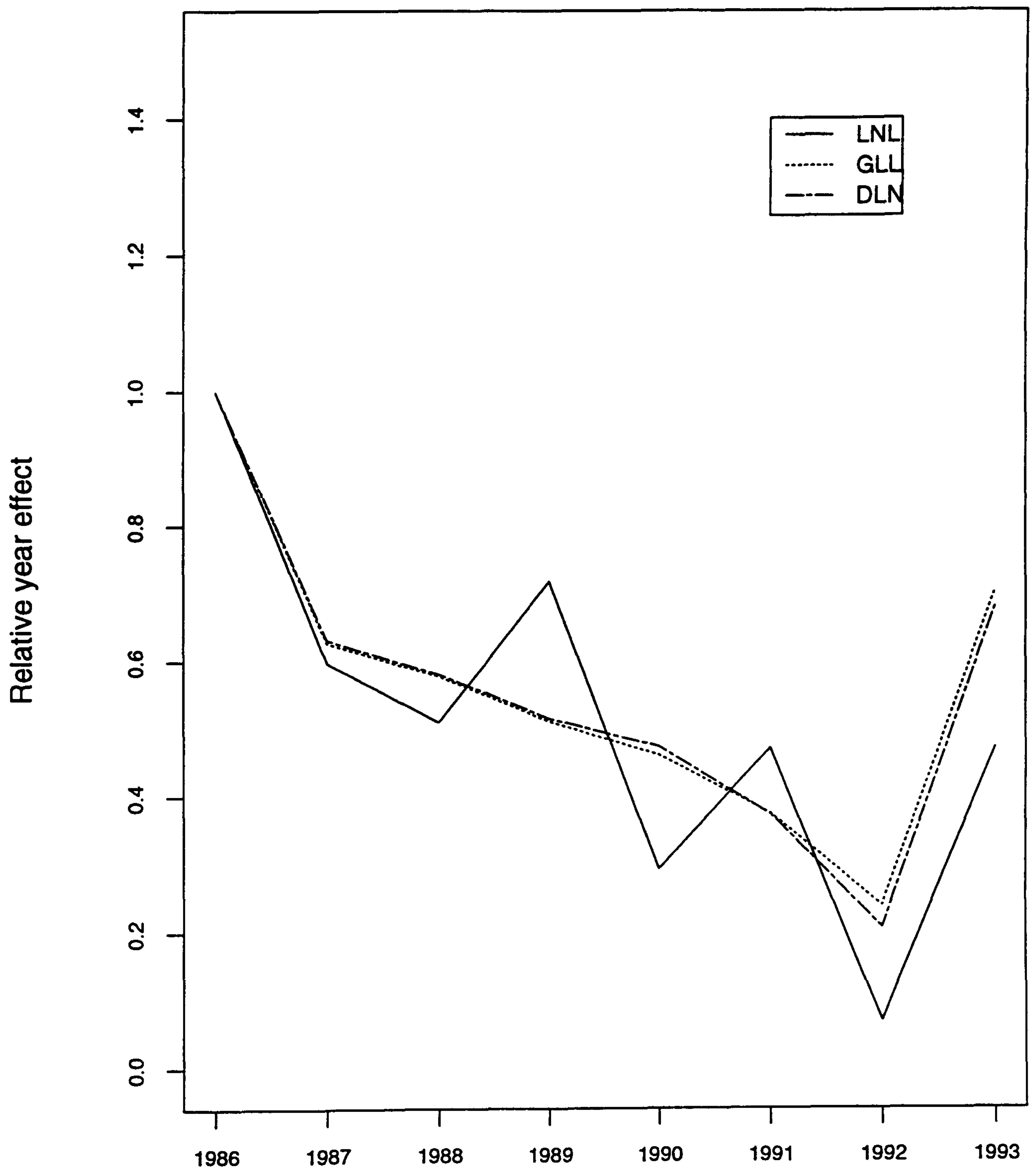

Figure 10. Comparison of relative year effects estimated from the lognormal linear (LNL) model, the gamma log-link (GLL) model, and the delta log-normal (DLN) model.

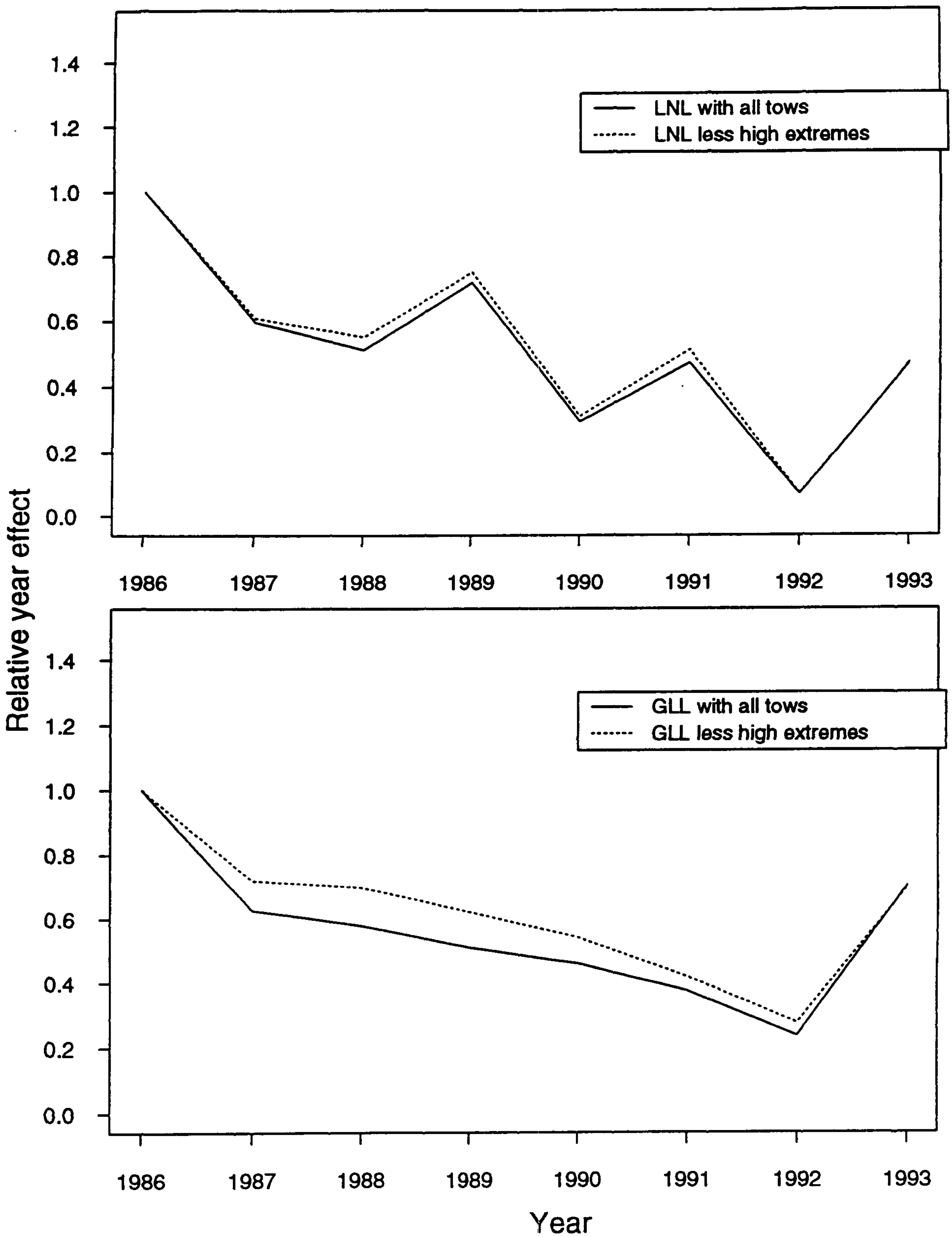

Figure 11. Sensitivity of the LNL and GLL models to removing high extremes.