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

New Zealand Fisheries Assessment Research Document 94/11

Catch per unit effort (CPUE) analysis of west coast South Island and Cook Strait spawning hoki fisheries, 1987-93

M. Vignaux
MAF Fisheries Greta Point
P.O. Box 297
Wellington

June 1994

MAF Fisheries, N.Z. Ministry of Agriculture and Fisheries

Catch per unit effort (CPUE) analysis of west coast South Island and Cook Strait spawning hoki fisheries, 1987–93
M. Vignaux
New Zealand Fisheries Assessment Research Document 94/11. 29 p

1 Executive summary

Standardised CPUE indices are calculated for the trawl fisheries based on spawning aggregations of hoki on the west coast of the South Island (WCSI) and in Cook Strait in winter. These indices are based on commercial catch and effort data and are standardised for effects such as changing distributions of vessel or net sizes using multiple regression.

A number of new methods are described for calculating standardised CPUE indices, and a number of different models are presented. Some of these new methods address the modelling problems associated with tows in which no hoki was caught. For the WCSI fishery, a linear model as presented in previous years, a linear model with interaction effects, and a generalised linear model (GLM) with a gamma error function are presented. For the Cook Strait fishery, a linear model, a linear model with interaction effects, and a combined linear and binomial model are presented.

The CPUE index from the WCSI linear model (the base-case index) shows a decline from 1.35 in 1987 to 0.5 in 1991 followed by an increase to an index of 0.67 in 1993. If the assumption is made that CPUE reflects abundance in this case, then this index should be corrected for the estimated proportion of fish that spawn inside the 25 nautical mile limit (and are hence inaccessible to the fishery) and used as a relative index of recruited mid-season biomass of the western stock of hoki.

The CPUE index from the Cook Strait linear model (the base-case index for Cook Strait) declines from 1 in 1989 to 0.53 in 1991, but recovers to 0.63 in 1992 and 0.57 in 1993. If the assumption is made that CPUE reflects abundance in this case, then this index should be used as a relative index of recruited mid-season biomass for the eastern stock of hoki.

2 Introduction

In 1992 a standardised CPUE index was calculated for the trawl fishery based on spawning aggregations of hoki on the west coast of the South Island (WCSI) in winter (Vignaux 1992). This CPUE index (for 1987 to 1991) was based on commercial catch and effort data and was standardised using multiple regression for effects such as changing distributions of vessel or net sizes. The index was recalculated (Vignaux 1993) to include data from the 1992 season, with a number of modifications and improvements to the method such as thorough error detection. This paper describes a further recalculation of the index to include data from the 1993 season.

This paper also describes the calculation of indices based on two alternative models. The first is a linear model in which first order interactions between predictor variables are considered, to allow a more complete description of the relationship between predictor variables and CPUE. The second is a generalised linear model with a gamma error distribution. Using a generalised linear model reduces the modelling problems associated with tows in which no hoki was caught.

Also, for the first time, CPUE indices for the Cook Strait winter spawning fishery are calculated. This dataset is modelled with three separate models. The first of these is a linear model, very similar to the WCSI linear model. The second is a linear model in which interactions between predictor variables are considered. The third is a combined binomial and linear model in which the binomial model describes the success rate of tows (the proportion of tows in which hoki was caught) and the linear model describes the catch rate in the successful tows. The combined model is another way of dealing with the modelling problems associated with tows in which no hoki was caught.

3 The data

Data for the 1993 west coast fishing season were extracted and checked for errors as described in Vignaux (1993). There were 6814 tow records extracted for 1993, taking the dataset to 47 333 records. A summary of the data available for each tow is in Table 1.

For the first time data for the Cook Strait fishery were also extracted, and similar error checking was applied to this data. This data set has 6985 records from 1989 to 1993. The number of records used in the Cook Strait fishery models is compared with the number used in the WCSI models in Table 2.

The Cook Strait fishery covers a much smaller area than the WCSI fishery (Figure 1), and is shallower (average bottom depth 380 m compared to 520 m). Thirty-one vessels fished in the Cook Strait fishery (compared to 193 in the WCSI fishery, 61 vessels in 1993 alone), all New Zealand vessels between 18 and 43 metres overall length. They used much

Table 1: Summary of variables in models. Cont indicates a continuous variable, Cat indicates a categorical variable with the given number of categories

Variable	Type	Description
Year	Cat 7	Year that tow occurred in
Season	Cont	Day relative to peak of season for that year
End time	Cont	Time in decimal hours at end of tow
Latitude	Cont	Position in decimal degrees at start of tow
Depth bottom	Cont	Depth in metres of bottom at start of tow
Depth net	Cont	Depth in metres of groundrope of net at start of tow
Net off bottom	Cont	Distance between groundrope and bottom at start of tow
Headline height	Cont	Headline height of net at start of tow
Vessel nation	Cat 6	Country of origin of the vessel (ex-USSR vessels are coded U)
Vessel length	Cont	Length overall of the vessel, in metres
Vessel tonnage	Cont	Gross tonnage of the vessel
Vessel year built	Cont	Year the vessel was built
Vessel power	Cont	Power of vessel engines in kilowatts
Vessel breadth	Cont	Breadth of the vessel, in metres
Vessel draft	Cont	Draft of the vessel, in metres
Vessel processing	Cat 4	Processing category of vessel
Vessel l*b*d	Cont	Vessel length * breadth * draft in cubic metres

Table 2: Number of records used in models of WCSI and Cook Strait CPUE. - means no data used

	Number of tows	
	WCSI	Cook Strait
1987	7110	-
1988	8227	-
1989	6525	400
1990	8056	1087
1991	5919	2105
1992	4682	1748
1993	6814	1645

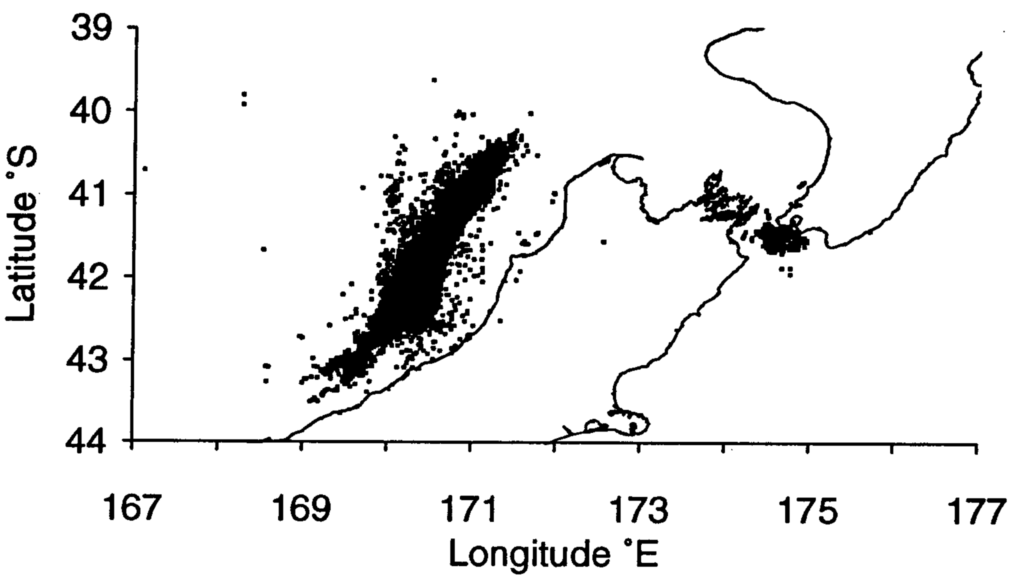

Figure 1: Locations of end of tow positions in WCSI and Cook Strait fisheries, 1989–93.

smaller nets than the vessels in the WCSI fishery (mean headline height 21 m compared to 48 m).

4 The west coast South Island fishery

4.1 Linear model

A stepwise regression procedure was followed (Appendix 1), as described last year, to produce relative year effects. As in last year's procedure, direct effect variables were added until the improvement in R^2 was less than 0.5%. The calculation of a relative index from the coefficients of this model is shown in Table 3. For 1987–92 this index shows little change from that presented last year, and for 1993 shows an increase (Figure 2). 22.24% of the variance could be explained with the seven variables selected into the model.

Of the seven variables included in the model last year, the first four (season, vessel nation, end time, and year) are identical to last year (Table 4). Headline height moves from being

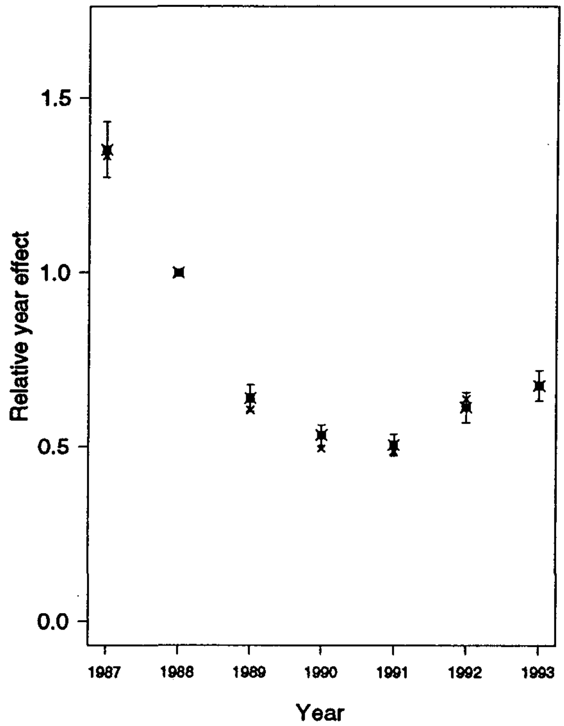

Figure 2: Comparison of base-case WCSI CPUE index (black squares) with results presented last year (crosses). Error bars are $\pm 2 s_{\hat{A}_j}$.

the fifth variable included in the model to the seventh position. Vessel length replaces vessel tonnage and vessel breadth. Latitude comes back into the model after being omitted last year.

The 1994 base-case index is compared to the index presented last year (1993 index), and the CPUE index corrected by the proportion of fish outside the 25 n. mile limit from acoustic surveys (Sullivan *et al* 1994) in Table 5. The corrected index is compared with the acoustic index in Figure 3.

The change in the indices when the arbitrary zero replacement catch is changed from 5 kg to 1 kg is slight as shown in Table 6.

4.2 Linear model with interaction effects

The data were refitted, at each stage considering the introduction not only of direct effect variables but also of first order interactions. Interactions involving the year effect itself were not considered. The variables included in the model are shown in Table 7. The most

Table 3: Relative year effects, $\hat{A}_j$, from linear model of log(catch per nautical mile) for 1987–93 WCSI (the base-case index). 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 (1988). $\hat{A}_j$ is the relative year effect and $s_{\hat{A}_j}$ is the standard error of the year effect

Year	R	s.e.	Cov.	$\hat{A}_j$	$s_{\hat{A}_j}$
1987	0.638	0.023	1.5e-05	1.35	0.04
1988	0.335	0.019	3.7e-04	1.00	0.00
1989	-0.114	0.020	-4.5e-05	0.64	0.02
1990	-0.296	0.019	-3.7e-05	0.53	0.02
1991	-0.350	0.021	-8.5e-05	0.50	0.02
1992	-0.154	0.024	-1.2e-04	0.61	0.02
1993	-0.059	0.022	-9.9e-05	0.67	0.02

Table 4: Comparison of variables used in the 1992, 1993, and 1994 base-case WCSI CPUE models. *See* Table 1 for definition of variables

Entry	1992 Model	1993 Model	1994 Model
1	Vessel breadth	Season	Season
2	End time	Vessel nation	Vessel nation
3	Season	End time	End time
4	Year	Year	Year
5	Vessel tonnage	Headline height	Vessel length
6	Vessel length	Vessel tonnage	Latitude
7	Latitude	Vessel breadth	Headline height

Table 5: Comparison of the 1994 base-case WCSI CPUE index with the index presented last year, and with the index corrected for the proportion of fish available to the fishery

Year	1993 Index	1994 base-case index	1994 corrected index
1987	1.34	1.35	1.74
1988	1.00	1.00	1.00
1989	0.60	0.64	1.06
1990	0.49	0.53	0.78
1991	0.48	0.50	0.63
1992	0.64	0.61	0.70
1993	-	0.67	0.94

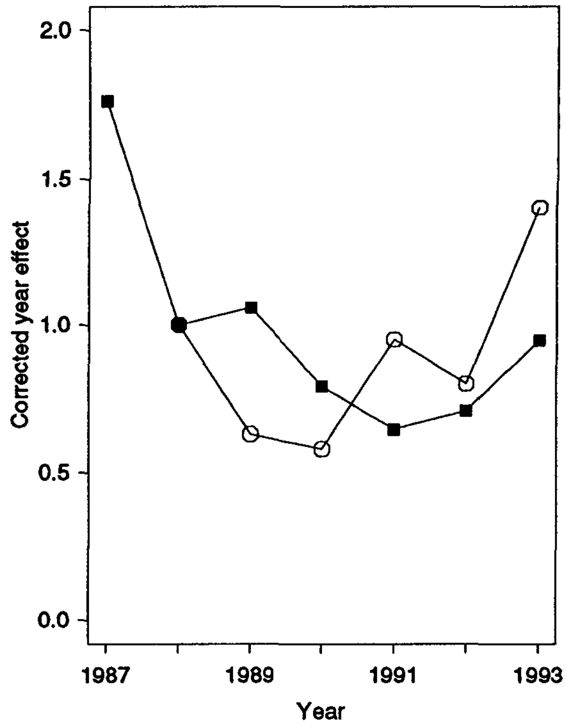

Figure 3: Comparison of base-case WCSI CPUE index corrected by proportion outside the 25 n. mile limit (black squares) with acoustic index (circles). Both indices are set to 1.0 in 1988. There is no acoustic index for 1987.

Table 6: Comparison of WCSI CPUE indices: base-case index calculated with a zero catch replacement of 5 kg, index calculated with a zero catch replacement of 1 kg, and an index with the interaction effect season-nation (and a 1 kg zero catch replacement)

	Index		
Year	Base-case	1 kg replacement	Interaction term
1987	1.35	1.38	1.46
1988	1.00	1.00	1.00
1989	0.64	0.63	0.57
1990	0.53	0.53	0.51
1991	0.50	0.51	0.54
1992	0.61	0.61	0.71
1993	0.67	0.67	0.68

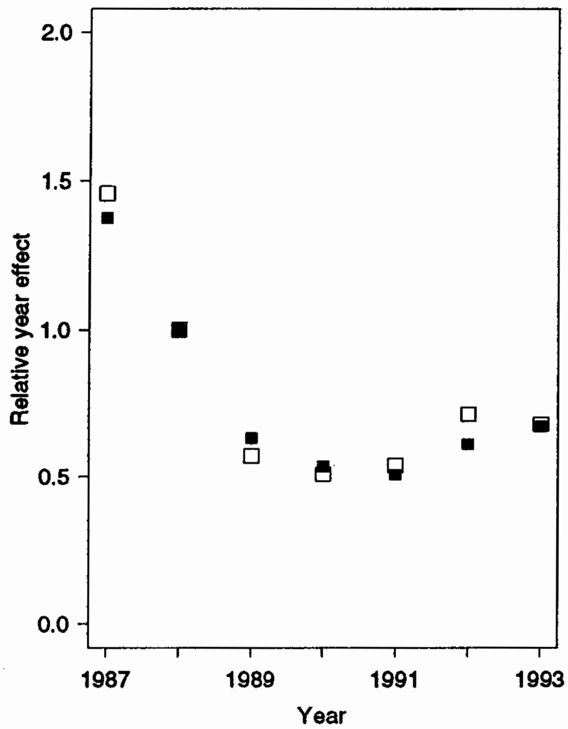

Figure 4: Comparison of base-case WCSI CPUE index with no interaction term (black squares) with a linear model with an interaction term (open squares).

significant change in this model is the inclusion of the interaction variable season-nation in the third iteration. This has a small effect on the year indices (note that the year indices are from the fit with the 1 kg zero replacement catch) (Table 6, Figures 4, 5).

The interaction terms are plotted in Figure 6 as a function of season for each of the nations. This plot suggests that while New Zealand, Japanese, and Norwegian vessels tend to have a better than average start of season, vessels from China and Poland have a worse than average start. This may be related to the relationship between the fishing strategies of the different vessels and the density or distribution of fish at this time of the season. For example, a strategy which involved a lot of exploratory fishing might do better when the fish were distributed widely in small dense patches than when fish were aggregated into a single large patch. On the other hand, these interaction effects may be artifacts of the severely unbalanced nature of the model, and may be misleading. Further investigation of fishing strategy and vessel and fish distributions may allow greater understanding.

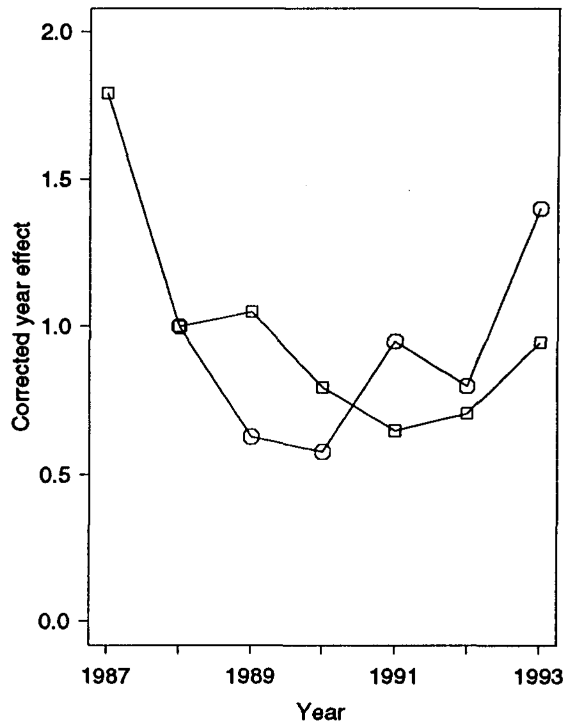

Figure 5: Comparison of WCSI CPUE index with interaction term corrected for the proportion outside the line (squares) compared with acoustic index (circles). Both indices are set to 1.0 in 1988. There is no acoustic index for 1987.

Table 7: Variables used in the linear model with interaction effects of WCSI CPUE. *See* Table 1 for definition of variables

Entry	Variable
1	Season
2	Vessel nation
3	Season-Nation interaction
4	End time
5	Year
6	Vessel length
7	Headline height

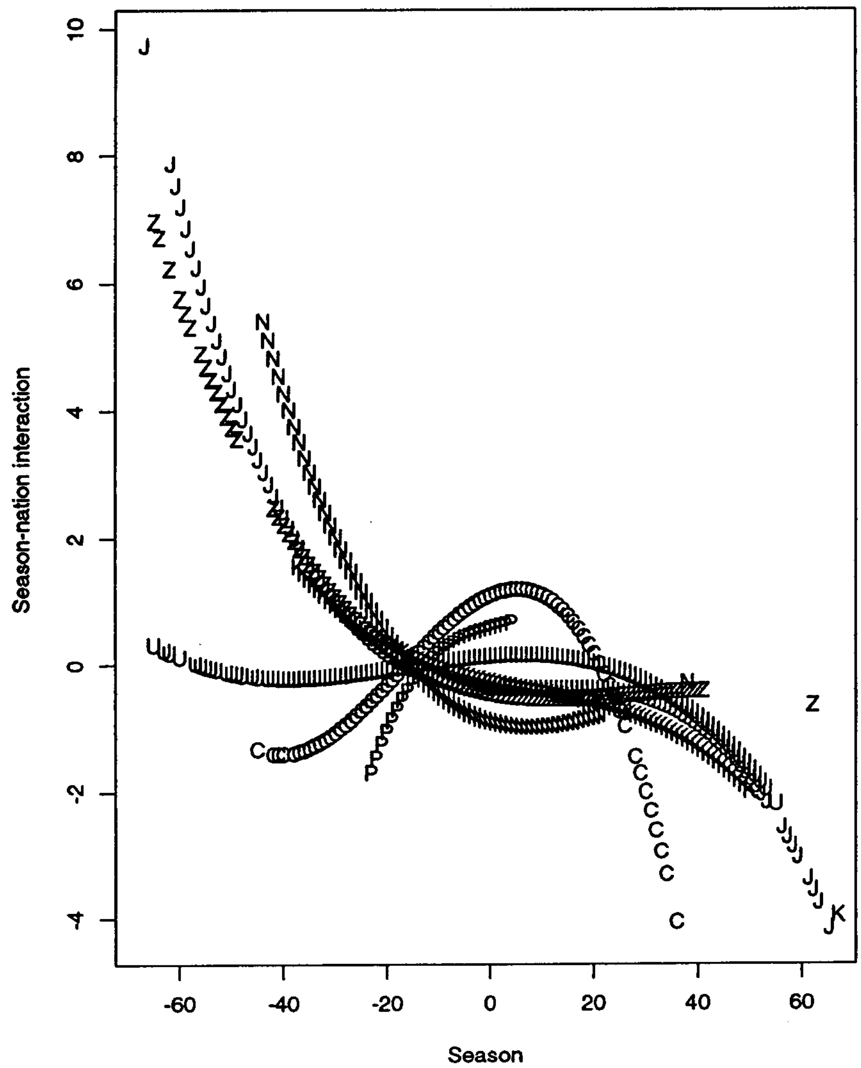

Figure 6: The season-nation interaction term in the WCSI linear model of CPUE plotted as a function of season for each of the nations (C=China, J=Japan, K=Korea, N=Norway, P=Poland, U=ex-USSR, Z=New Zealand). Season is the number of days before or after the day (in a particular year) with peak catch rates. season-nation interaction is the relative effect of the season-nation interaction term on catch rates for each of the seven nations.

4.3 Generalised linear model

In order to use a linear model to describe a variable such as CPUE which has a distinctly non-normal distribution, it was necessary to transform the response variable (CPUE) with a logarithmic transformation. This transformation gave the response variable a nearly normal distribution, and, because of the mathematical properties of the log function, also made the effects of the additive variables in the model multiplicative on real catch rates. It also made the variance more homogeneous (i.e., so that the variance was the same at all values of CPUE). However, we could not apply the logarithmic transformation when the catch was zero, so it was necessary to introduce an arbitrary constant.

Generalised Linear Models (GLM) are based on a generalisation of the linear model in which, instead of transforming the response variable, we explicitly state howd the model will be non linear. This allows us to specify both a transformation of the response variable (in this case a log link) and a variance function (in this case proportional to the square of the mean) separately. Instead of working with the transformed response variable (i.e. log(CPUE)) the model works with the untransformed variable (CPUE) and tries to minimise deviations from this (Chambers & Hastie 1992).

The main advantage for this work is that because the response variable is compared with the fitted values in untransformed space (i.e. without taking logarithms), values of zero CPUE can be left in the model and no arbitrary replacement is required. Because the fit is evaluated in untransformed space, the indices scale as the arithmetic mean of the CPUE values rather than as the geometric means, so the index has slightly different values from the linear model index.

The CPUE was therefore modelled with a GLM, using a log link function and a gamma error distribution (producing a variance proportional to the square of the mean). Figure 7 confirms that in the data the variance is proportional to the square of the mean. Variables were included in the model in a stepwise fashion as with the linear model. The process stopped when there was less than 0.5% improvement in the deviance explained compared to the null deviance (analogous to the sum of squares of the errors or SSE in a linear model).

Six variables were selected into the GLM model. These were Vessel processing, Season, Year, Vessel tonnage, Latitude, and End time. Four of these are also in the linear model, and Vessel tonnage was in the linear model last year. Only the Vessel processing category (no processing, head and gut, filleting, or surimi) is a new variable, and this is closely related to Vessel nation. A comparison of the variables selected in the base-case linear and the GLM model is in Table 8.

The year indices and its standard error from the GLM is shown in Table 9 and Figure 8. The indices are broadly similar to the results of the linear model, but show less decline between 1988 and 1991. The GLM index shows a decrease rather than an increase between 1992 and 1993, but when corrected by the proportion outside the 25 n. mile limit, the

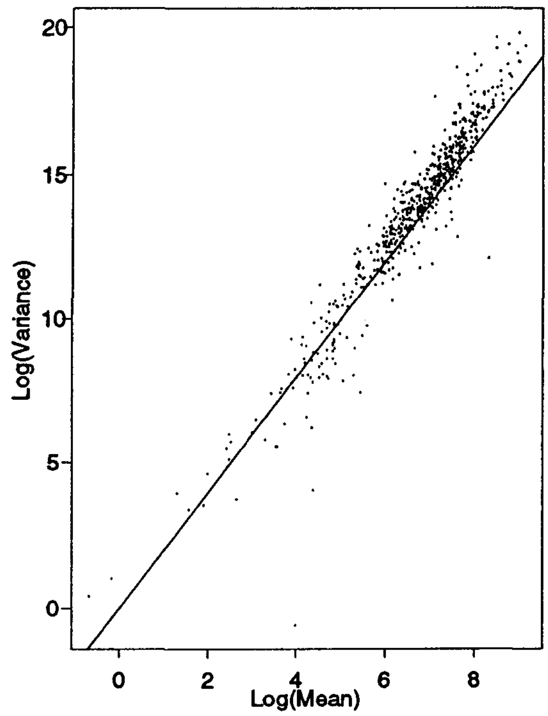

Figure 7: Relationship between the variance of the catch rates of tows on a day versus the mean of the catch rates on each day (682 points, one for each unique day in the dataset). The line that passes through the origin and has a gradient of 2 is shown for comparison.

Table 8: Comparison of variables used in the base-case WCSI linear model and the GLM. *See* Table 1 for definition of variables

Entry	Linear model	GLM
1	Season	Vessel processing
2	Vessel nation	Season
3	End time	Year
4	Year	Vessel tonnage
5	Vessel length	Latitude
6	Latitude	End time
7	Headline height	-

Table 9: Relative year effects $\hat{A}_j$ from generalised linear model (gamma distribution and log link) of catch per n. mile for 1987–93 in WCSI. 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 (1988). $\hat{A}_j$ is the relative year effect and $s_{\hat{A}_j}$ is the standard error of the year effect

Year	R	s.e.	Cov.	$\hat{A}_j$	$s_{\hat{A}_j}$
1987	0.521	0.024	-1.753e-06	1.40	0.04
1988	0.187	0.020	4.161e-04	1.00	0.00
1989	-0.070	0.022	-4.513e-05	0.77	0.02
1990	-0.358	0.020	-3.039e-05	0.58	0.02
1991	-0.352	0.023	-9.181e-05	0.58	0.02
1992	0.074	0.027	-1.447e-04	0.89	0.03
1993	-0.002	0.023	-1.023e-04	0.83	0.03

Table 10: The WCSI GLM CPUE index corrected for the proportion of fish available to the fishery

Year	GLM index	Corrected index
1987	1.40	1.83
1988	1.00	1.00
1989	0.77	1.28
1990	0.58	0.86
1991	0.58	0.74
1992	0.89	1.03
1993	0.83	1.17

GLM index increases (Table 10, Figure 9).

However, because the fit is evaluated in untransformed space, so that the indices scale as the arithmetic mean of the CPUE values rather than as the geometric means, the GLM may be more sensitive to errors in the data in the high CPUE tail of the distribution. Table 11 shows the linear and GLM indices for the full dataset and for the dataset with all records with CPUE above 20 t. per n. mile (0.65% of the data, but 1.9% in 1987 and only 0.14% in 1991) trimmed off. The GLM model is clearly much more sensitive to the outliers.

Although the data were thoroughly error checked, it is very likely that errors remain. As the GLM model is very sensitive to the outlying data, it could easily be biased by a few errors in these outlying values. There are, however, other advantages to using the GLM model, and it may be that future work, or future quality improvements in the data, will make this method practical for this dataset.

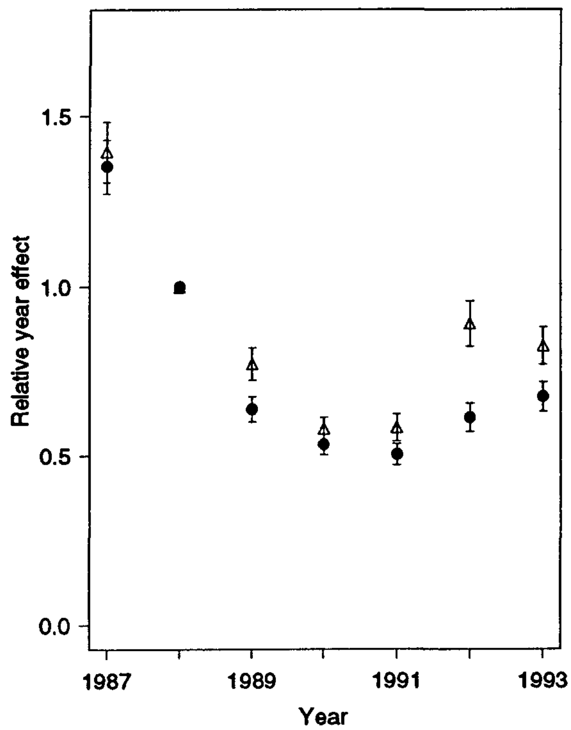

Figure 8: Comparison of the index from the base-case WCSI linear model of CPUE (black squares) and the GLM (triangles).

Table 11: Comparison of the sensitivity of the linear and the GLM models of WCSI CPUE to trimming all CPUE values above 20 t. per n. mile

	Linear		GLM	
Year	Full	Trimmed	Full	Trimmed
1987	1.35	1.30	1.40	0.88
1988	1.00	1.00	1.00	1.00
1989	0.64	0.65	0.77	0.74
1990	0.53	0.54	0.58	0.99
1991	0.50	0.51	0.58	0.85
1992	0.61	0.61	0.89	0.85
1993	0.67	0.68	0.83	1.13

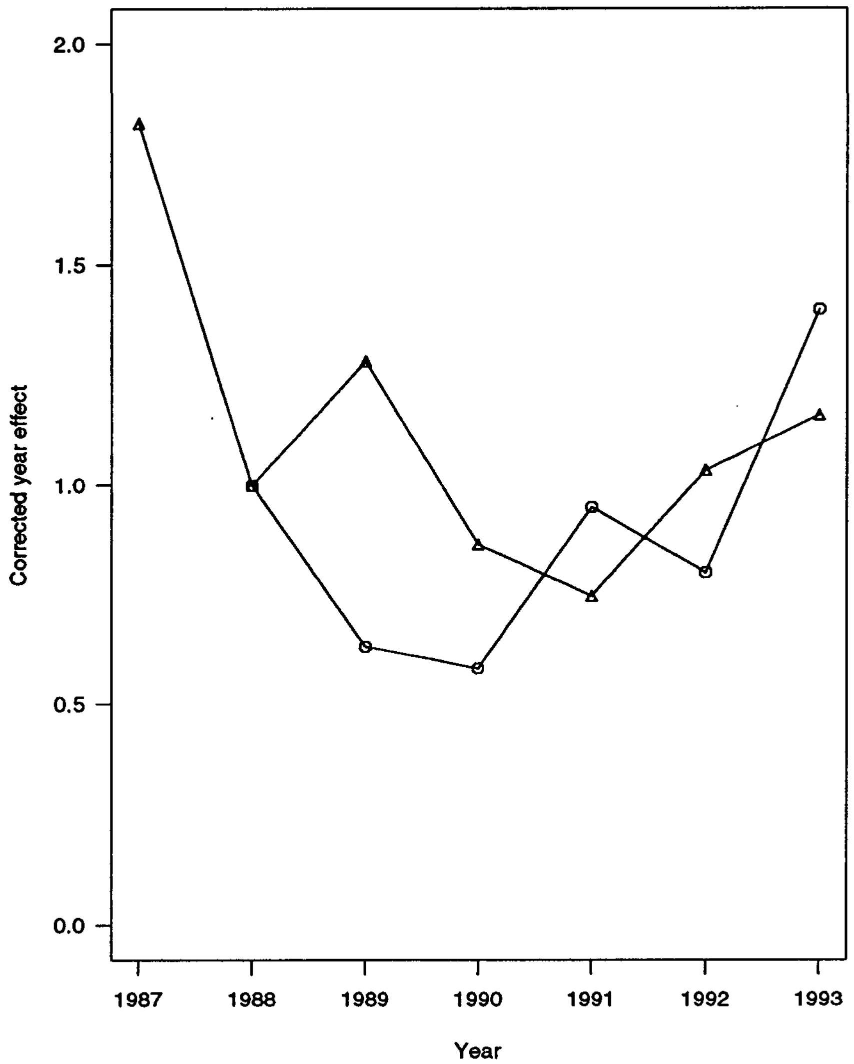

Figure 9: Comparison of the WCSI GLM CPUE index corrected by proportion outside the 25 n. mile limit (triangles) and the acoustic index (circles). Both indices are set to 1.0 in 1988. There is no acoustic index for 1987.

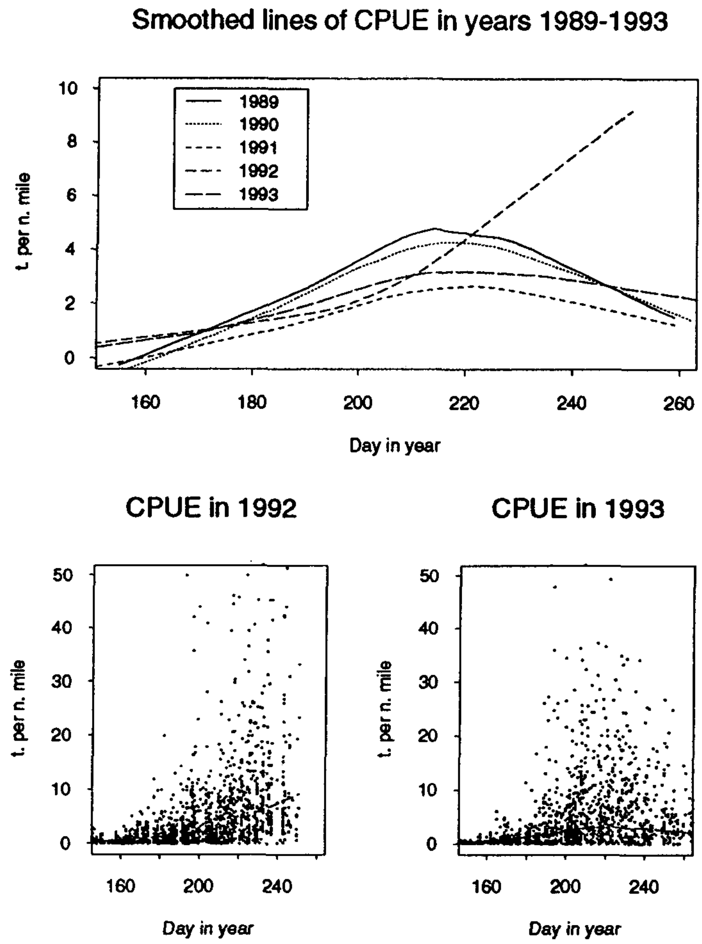

Figure 10: Trends in raw catch rates in the Cook Strait fishery during each of the fishing seasons, 1989–1993.

5 The Cook Strait fishery

5.1 Linear model

A linear model similar to that described above was fitted to the Cook Strait data (Appendix 2). Season was not modelled relative to the peak of the season as in 1992 there was no peak, i.e., the catch rates apparently increased up to the end of fishing (see Figure 10).

The base-case index calculated from the linear model with zero catches replaced with 5 kg is given in Table 12 and shown in Figure 11. Table 13 shows the variables included in the model. For comparison a second model was fit with zero catches replaced with 1 kg. This makes little difference to the indices (Table 14).

Table 12: Relative year effects $\hat{A}_j$ from linear model of log(catch per nautical mile) for 1989–93 (the base-case index) with zero catches replaced with 5 kg in Cook Strait. 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 (1989). $\hat{A}_j$ is the relative year effect and $s_{\hat{A}_j}$ is the standard error of the year effect

Year	R	s.e.	Cov.	$\hat{A}_j$	$s_{\hat{A}_j}$
1989	0.351	0.072	0.0052	1.00	0.00
1990	0.253	0.049	-0.0013	0.91	0.09
1991	-0.290	0.039	-0.0011	0.53	0.05
1992	-0.111	0.042	-0.0014	0.63	0.06
1993	-0.204	0.045	-0.0003	0.57	0.05

Table 13: Summary of variables used in the base-case Cook Strait linear model. *See* Table 1 for definition of variables

Entry	Variable
1	Season
2	Vessel draft
3	Year
4	Headline height

Table 14: Comparison of base-case Cook Strait CPUE index with an index calculated with a zero catch replacement of 1 kg, and with an index with the interaction term season-vessel draft

			Index
Year	Base-case	1 kg replacement	Interaction term
1989	1.00	1.00	1.00
1990	0.91	0.93	0.93
1991	0.53	0.54	0.53
1992	0.63	0.61	0.63
1993	0.57	0.55	0.54

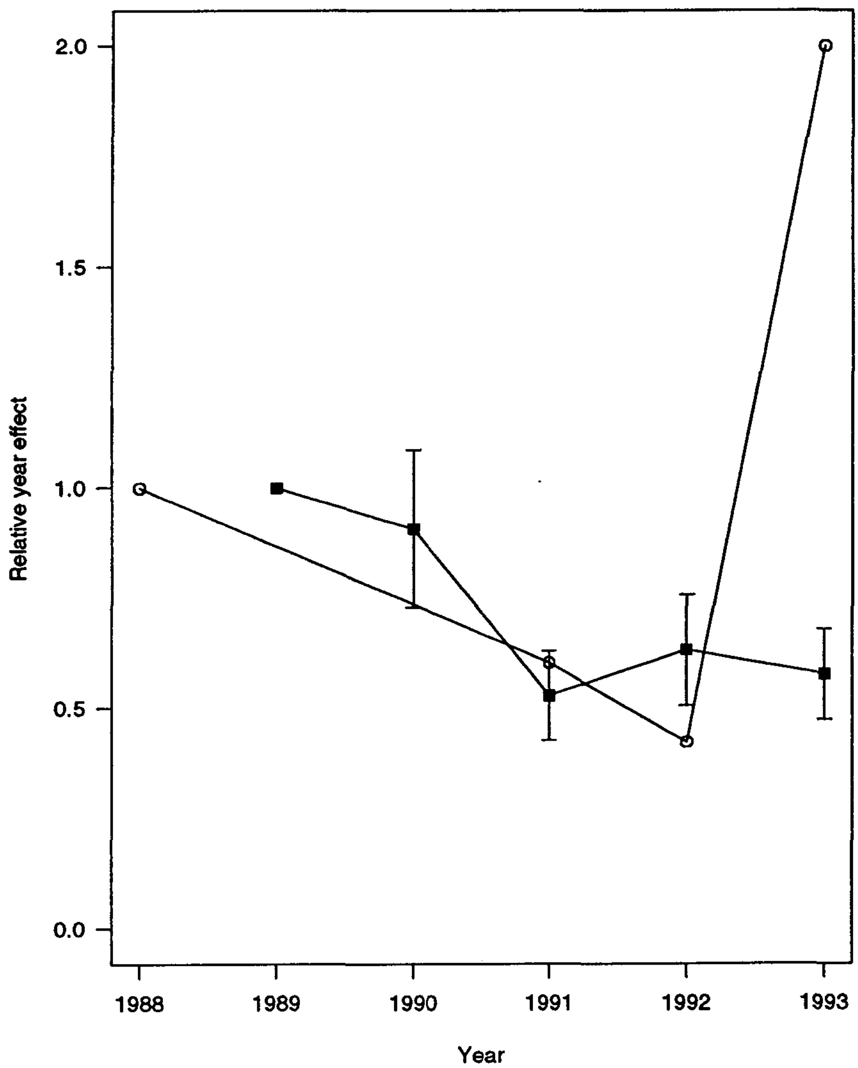

Figure 11: Comparison of base-case Cook Strait CPUE index (black squares) with acoustic index (circles). Error bars are $\pm 2s_{\hat{A}_j}$.

The base-case index is compared with the acoustics index in Figure 11. Both indices show a similar decline from the start of the fishery to 1991, but the acoustics index shows a large increase in 1993 which is not shown in the CPUE index. This increase may be due to the recruitment of a large year class to the fishery. The indices show the greatest decline between 1988 and 1991. This is not easily explained by the catches which were low until the 1991 season, and high in 1991, 1992, and 1993 (Sullivan *et al* 1994).

Although vessels are allowed to fish in all areas of Cook Strait, the fishery is mainly centered on Cook Strait Canyon, which is only one of six strata used in the acoustics index. The CPUE index should perhaps be adjusted by the proportion of fish in this stratum each year, in a similar way that the WCSI CPUE index is corrected for the proportion of fish outside the line. However, the present coverage of the strata by the acoustic surveys is not sufficiently complete to allow this to be done.

5.2 Linear model with interaction effects

The Cook Strait data were refitted, at each stage considering the introduction not only of direct effect variables but also of first order interactions. Interactions involving the year effect itself were not considered. Table 15 shows the variables included in the model.

An interaction between the continuous variable season and the continuous variable draft was introduced into the model as a cubic polynomial in the fourth iteration. The introduction of this variable changes the basic season and vessel draft effects (Figure 12). The season variable changes from an asymmetric peak to an increasing curve, the draft variable changes from a curve with a peak to another increasing curve. The season:draft interaction is a decreasing curve. This could be interpreted as that the end of the season has better catch rates, and larger vessels have greater catch rates, but at the end of the season, the catch rates of large vessels are not as good as might have been expected. The effect on the year indices of the introduction of this interaction term is small (Table 14).

Table 15: Variables used in the Cook Strait CPUE linear model with interaction effects. *See* Table 1 for definition of variables

Entry	Variable
1	Season
2	Vessel draft
3	Year
4	Season:vessel draft interaction
5	Headline height

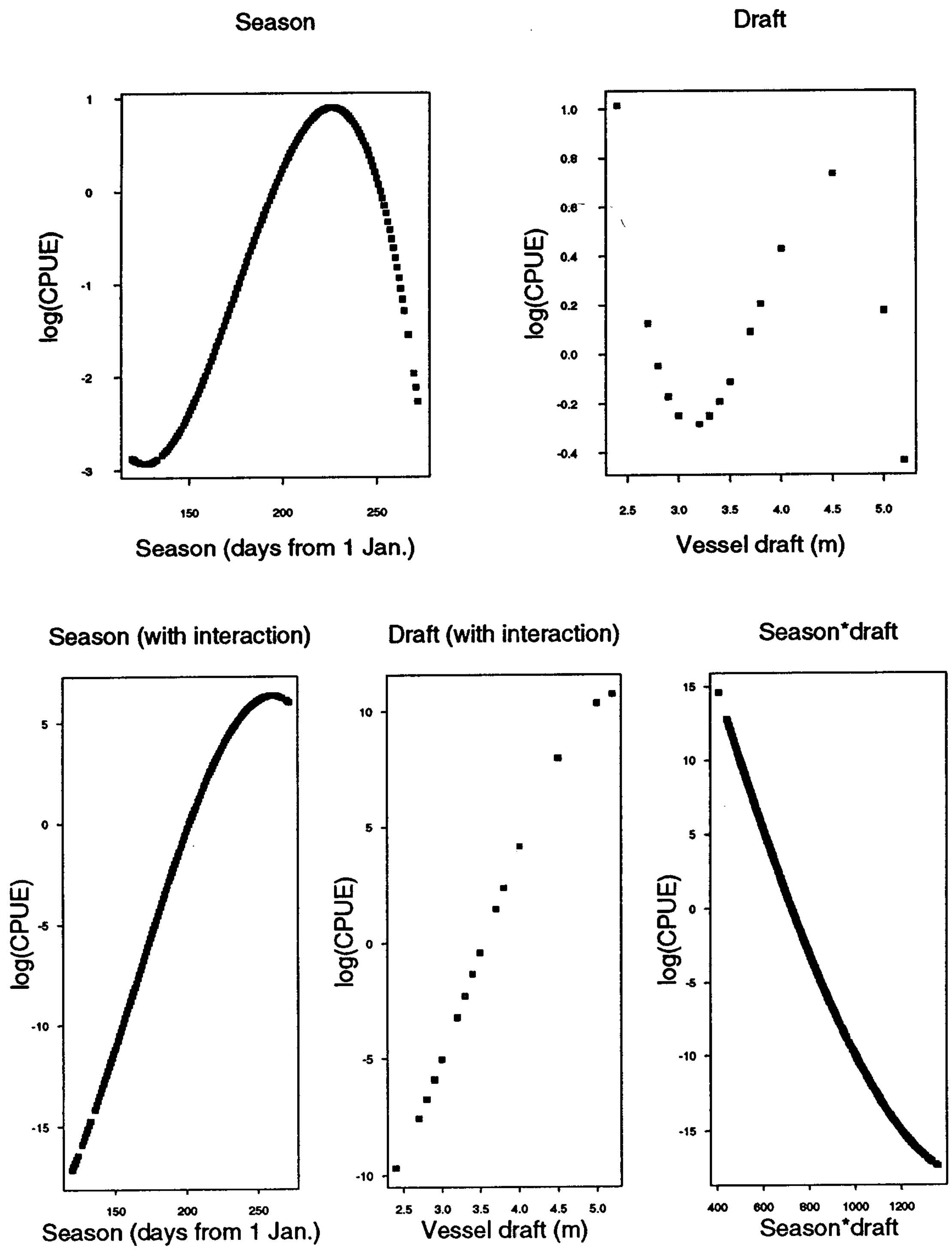

Figure 12: Changes in the seasonal and vessel draft variables when an interaction between season and vessel draft is introduced to the linear model of Cook Strait CPUE.

5.3 Combined linear and binomial model

In Section 4.3, a method was described to avoid introducing an arbitrary constant to cope with zero catches. In this section an alternative method is described. This method splits the problem into two - firstly, how has the success rate of tows changed from year to year (where a successful tow is a non-zero tow) and secondly, how has the catch rate in successful tows changed from year to year. Hence a change in catch rates could be detected in either index. The method is loosely based on the delta-lognormal method described by Lo *et al*(1992) who were able to use a log-normal model for both parts of the model because the data was aggregated rather than tow by tow.

5.3.1 Linear model of catch rate of successful tows

The catch rate of successful tows was modelled using an ordinary linear model as described above, but with all zero tows excluded. Table 16 shows the variables used in the linear model with no zero catches. This index is very similar to the base-case index, until 1992, when it is much higher (Table 17).

Table 16: Variables used in the Cook Strait linear model of CPUE from a dataset with all zero catches removed. *See* Table 1 for definition of variables

Entry	Variable
1	Season
2	Vessel draft
3	Year
4	Vessel length

Table 17: The proportion of zero tows, the proportion relative to the 1989 value, the indices from the base-case and the no zeroes linear models, and the indices from the binomial models with just the year variable and with all variables

	Proportion zero		Linear model		Binomial model	
Year	Raw	Relative	Base-case	No zeroes	Year only	All variables
1989	0.032	1.00	1.00	1.00	1.00	1.00
1990	0.025	0.78	0.91	0.91	0.76	0.72
1991	0.017	0.53	0.53	0.54	0.52	0.59
1992	0.032	1.00	0.63	0.82	0.99	1.35
1993	0.040	1.25	0.57	0.81	1.24	2.26

5.3.2 Binomial model of proportion of zero catches

The success rate (or equivalently, proportion of zero catches) was modelled using a GLM as described in Section 4.3 above, but instead of a gamma family model, a binomial model was used. The response variable was 1 if the catch was zero, and 0 otherwise. When only the year effect is included in the model, the coefficients are similar to the proportion of zeroes in each year (Table 17).

A full binomial model was fitted using all variables, in the same stepwise manner as used in the linear models. The process stopped when there was less than 0.5% improvement in the deviance explained compared to the null deviance (analogous to the sum of squares of the errors or SSE in a linear model). Four variables were included in the model- latitude, year, vessel draft, and headline height (Table 18).

Table 18: Variables used in the full binomial model of Cook Strait CPUE. *See* Table 1 for definition of variables

Entry	Variable
1	Latitude
2	Year
3	Vessel draft
4	Headline height

5.3.3 Combining the linear and binomial models

A combined index, in which the year effects from the linear model are combined with the year effects from the binomial model, is shown in Table 19 and Figure 13. The method of combining the two indices is described in Appendix 3. This index is lower in the last 2 years than the linear model (no zeroes) alone, but not as low as the base-case model.

The standard error of the relative year effects for the combined index was estimated using a bootstrap procedure, in which 6985 records were sampled with replacement from the set of 6985 records in the dataset and the indices recalculated (using the same linear and binomial models). The resampling was repeated 500 times.

Table 19: Indices from base-case linear model, linear model with no zeroes, binomial model, combined model and the standard errors for the index from the combined model for Cook Strait CPUE. The calculation of the index for the combined model (as described in Appendix 3) uses a value of π_0 (the proportion of zero catches in the first year) of 0.0325. The column s.e. is an estimate of the standard error of each of the year effects for the index from the combined model, calculated using a bootstrap procedure

Year	Base-case	Linear (no zeroes)	Binomial	Combined	s.e.
1989	1.00	1.00	1.00	1.00	0.00
1990	0.91	0.91	0.72	0.92	0.13
1991	0.53	0.54	0.59	0.55	0.07
1992	0.63	0.82	1.35	0.81	0.10
1993	0.57	0.81	2.26	0.78	0.09

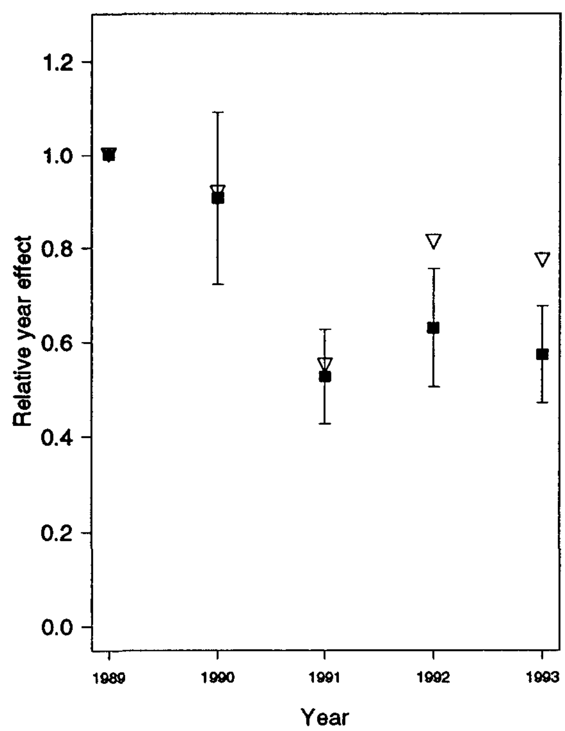

Figure 13: Comparison of combined index (triangles) with base-case index for Cook Strait CPUE (black squares).

6 Discussion

When the base-case index for the WCSI fishery is calculated, as calculated last year, the indices for 1987–92 are not greatly changed by the addition of 1993 season data. The index for 1993 once again shows an increase on the previous year.

The model for the WCSI which includes interaction effects is likely to be an improvement in general, but because it is difficult to interpret the interaction effects, it is possible that the effects are artifacts of the unbalanced nature of the model, and may be misleading. In any case, including them makes little difference to the indices. Further work may determine whether such interaction effects should be included in future.

The GLM for the WCSI is also a superior model in principle, but in this analysis it seems to be too sensitive to potential errors in the data and should not be used. However, there may be other data sets (possibly with a less widely ranging tail in the distribution of CPUE) where it is appropriate. It may be that future work, or future quality improvements in the data, will make this method practical for WCSI hoki as well.

The analysis of Cook Strait data is entirely new. The Cook Strait fishery is a different fishery to the WCSI fishery, with smaller vessels fishing a smaller area. Even if CPUE provides an index of abundance on the west coast, it may not do so in Cook Strait. However, I know of no way to evaluate objectively the applicability of CPUE as an index of abundance in a particular real fishery.

The combined model, though relatively undeveloped at this stage, suggests a promising method for dealing in a sensible way with the inevitable problem of zero catches. Development of this method will continue for use in future assessments.

7 Acknowledgments

I thank Elizabeth Bradford, Ian Doonan, and members of the Hoki Stock Assessment Working Group for useful discussions.

8 References

Chambers, J.M. & Hastie, T.J. 1992: Statistical models in S. Wadsworth & Brooks/Cole, Pacific Grove, California. 608 p.

McCullagh, P. & Nelder, J.A. 1989: Generalised linear models. Chapman & Hall, London. 511 p.

Sullivan, K.J., Cordue, P.L., & Ballara, S.L. 1994: A review of the 1992–93 hoki fishery and assessment of hoki stocks for 1995. Draft N.Z. Fisheries Assessment Research Document.

Vignaux, M. 1992: Catch per unit of effort (CPUE) analysis of the hoki fishery. N.Z. Fisheries Assessment Research Document 92/14. 31 p.

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

Lo, N.C-H, Jacobson, L.D., & Squire, J.L. 1992: Indices of relative abundance from fish spotter data based on delta-lognormal models *Canadian Journal of Fisheries and Aquatic Sciences* 49: 2515-2526

9 Appendix 1: Details of WCSI base-case model

Choice of variables in linear model (base-case) of log(catch per nautical mile) for WCSI fishery 1987–93 in order of importance. R^2 is the multiple regression coefficient. In the first iteration, log(CPUE) was regressed against each of the variables in turn to find the variable which explained the most variation in log(CPUE) i.e. had the highest multiple regression coefficient R^2. This variable was included in the model. In iteration 2, log(CPUE) was regressed against the new model plus each of the other variables in turn to find the next most significant variable. The process continued for seven iterations.

	R^2 at iteration							
Variable	1	2	3	4	5	6	7	8
Season	10.23							
Vessel nation	4.52	16.30						
End time	1.69	11.83	18.25					
Year	2.22	12.09	18.14	20.14				
Vessel length	3.70	13.93	17.56	19.65	21.58			
Latitude	7.90	10.77	17.23	18.44	20.78	22.33		
Headline height	2.34	11.89	17.05	19.11	21.50	22.26	22.97	
Vessel year built	0.22	10.66	16.62	18.59	20.73	22.06	22.86	23.34
Vessel tonnage	2.99	13.18	17.56	19.59	21.48	22.01	22.77	23.19
Depth net	2.42	10.55	16.55	18.76	20.37	21.82	22.53	23.16
Vessel lbd	4.78	15.47	17.51	19.57	21.44	21.81	22.56	23.15
Vessel draft	3.98	15.08	17.24	19.27	21.16	21.79	22.55	23.14
Vessel processing	4.83	16.02	17.37	19.42	21.50	21.96	22.68	23.13
Depth bottom	1.14	11.04	16.84	18.56	20.29	21.72	22.48	23.12
Vessel power	3.77	14.40	17.48	19.53	21.45	21.75	22.50	23.09
Vessel breadth	4.22	14.57	17.49	19.58	21.54	21.66	22.44	23.07
Net off bottom	2.36	12.39	16.68	18.76	20.32	21.69	22.44	23.01
Improvement	10.23	6.07	1.95	1.89	1.44	0.75	0.64	0.37
Degrees of freedom	47329	47323	47320	47313	47310	47308	47305	

10 Appendix 2: Details of Cook Strait base-case model

Choice of variables in linear model (base-case) of log(catch per nautical mile) for Cook Strait fishery 1987–93 in order of importance. R^2 is the multiple regression coefficient. In the first iteration, log(CPUE) was regressed against each of the variables in turn to find the variable which explained the most variation in log(CPUE) i.e. had the highest multiple regression coefficient R^2. This variable was included in the model. In iteration 2, log(CPUE) was regressed against this model plus each of the other variables in turn to find the next most significant variable. This process continued in a series of five iterations.

	R^2 at iteration				
Variable	1	2	3	4	5
Season	20.73				
Vessel draft	1.26	22.57			
Year	0.88	21.55	23.53		
Headline height	9.70	21.66	23.14	24.21	
Vessel length	1.86	21.31	23.17	23.96	24.68
Latitude	1.43	21.22	23.09	24.03	24.64
Depth bottom	2.46	21.33	22.84	23.82	24.59
Vessel lbd	4.60	21.33	23.23	24.02	24.59
Vessel power	0.88	21.57	23.12	23.97	24.57
End time	0.05	20.98	22.87	23.85	24.53
Vessel tonnage	0.30	21.31	23.00	23.88	24.49
Net off bottom	1.00	21.05	22.81	23.76	24.47
Vessel breadth	0.50	21.35	23.02	23.88	24.47
Vessel year built	0.46	21.03	23.03	23.87	24.43
Depth net	4.30	21.01	22.67	23.60	24.31
Improvement	20.73	1.84	0.96	0.68	0.47
Degrees of freedom	6981	6978	6973	6971	

11 Appendix 3: Combining the linear and binomial indices

When we use a model with a log link function, such as the linear model (no zeroes) the year effects are calculated by applying the inverse of the link function (the exponential or anti-log function) to the regression coefficients from the fit of the model. Relative year effects are calculated by dividing each of the year effects by the year effect in the index year.

For example, consider a model of the form

$$y = \log(CPUE) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$$

where α_0 is the coefficient corresponding to the intercept of the regression, α_1 is the coefficient related to the explanatory variable x_1, and α_2 is the coefficient related to the explanatory variable x_2. If x_2 was a categorical variable, such as year, then α_2 would take on the values α_{20} if the year was the reference year, year 0, and α_{2i} if the year was some other year, year i.

So the index for year i relative to year 0 is just

$$\begin{aligned}\hat{A}_i &= \frac{\exp(y_i)}{\exp(y_0)} \\ &= \frac{\exp(\alpha_0 + \alpha_1 x_1 + \alpha_{2i})}{\exp(\alpha_0 + \alpha_1 x_1 + \alpha_{20})} \\ &= \exp(\alpha_{2i} - \alpha_{20})\end{aligned}$$

However, the binomial model was fitted using a GLM with a binomial distribution. A binomial model has a variance function proportional to $\pi(1-\pi)$ where π is the probability of a zero catch and a logit link function

$$\eta = \log(\frac{\pi}{1-\pi})$$

where η is the transformation of π. So in this case we cannot just exponentiate the regression coefficients to obtain the relative year effects.

The function that links the regression coefficients to the year effects is the inverse of the logit link function.

$$\pi = \frac{e^\eta}{1+e^\eta}$$

where η is the transformation of the variable representing the probability of a zero catch π.

For example, using a model

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

where β_0 is the coefficient corresponding to the intercept of the regression, β_1 is the coefficient related to the explanatory variable x_1, and β_2 is the coefficient related to the explanatory variable x_2. If x_2 was a categorical variable, such as year, then β_2 would take on the values β_{20} if the year was the reference year, year 0, and β_{2i} if the year was some other year, year i.

We have

$$\begin{aligned}
\pi_0 &= \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_{20})}{(1 + \exp(\beta_0 + \beta_1 x_1 + \beta_{20}))} \\
\pi_i &= \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_{2i})}{(1 + \exp(\beta_0 + \beta_1 x_1 + \beta_{2i}))} \\
1 - \pi_0 &= \frac{1}{(1 + \exp(\beta_0 + \beta_1 x_1 + \beta_{20}))} \\
1 - \pi_i &= \frac{1}{(1 + \exp(\beta_0 + \beta_1 x_1 + \beta_{2i}))}
\end{aligned}$$

Let

$$a_i = \frac{\pi_i}{1 - \pi_i} / \frac{\pi_0}{1 - \pi_0} = \exp(\beta_{2i} - \beta_{20})$$

Then

$$\begin{aligned}
\pi_i(1 - \pi_0) &= \pi_0(1 - \pi_i)a_i \\
\pi_i &= a_i\pi_0 + \pi_i\pi_0(1 - a_i) \\
\pi_i(1 - \pi_0(1 - a_i)) &= a_i\pi_0 \\
\pi_i &= \frac{a_i\pi_0}{1 - \pi_0(1 - a_i)} \\
\frac{1 - \pi_i}{1 - \pi_0} &= \frac{\pi_i}{a_i\pi_0} = \frac{1}{1 - \pi_0(1 - a_i)}
\end{aligned}$$

The combined index in year i relative to year 0 is therefore the ratio of

$$C_i = (1 - \pi_i)\exp(y_i)$$

to

$$C_0 = (1 - \pi_0)\exp(y_0)$$

where C_i is the product of the success rate in year i and the catch rate in the successful tows. So

$$\begin{aligned}
\frac{C_i}{C_0} &= \frac{(1 - \pi_i)\exp(y_i)}{(1 - \pi_0)\exp(y_0)} \\
&= \frac{1}{1 - \pi_0(1 - a_i)}\exp(\alpha_{2i} - \alpha_{20})
\end{aligned}$$