<u>Not to be cited without permission of the author</u>

New Zealand Fisheries Assessment Research Document 98/27

Feasibility of using CPUE as an index of stock abundance for hake

Terese H. Kendrick
NIWA
PO Box 14901
Kilbirnie, Wellington

November 1998

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.

Feasibility of using CPUE as an index of stock abundance for hake

Terese H. Kendrick

New Zealand Fisheries Assessment Research Document 98/27. 22 p.

1. EXECUTIVE SUMMARY

Hake is the most abundant bycatch species in the west coast South Island (WCSI) hoki fishery and is now being targeted off the west coast (HAK 7), on the Chatham Rise (HAK 4) and in the region of the Snares and Auckland Islands (HAK 1).

The catch per unit effort (CPUE) time-series for the mixed hoki/hake (WCSI) fishery has problems as an index of abundance because of a high proportion of reported zero hake catches which may or may not relate to the abundance of hake. The target fisheries in HAK 7, HAK 4, and HAK 1 are relatively small, and the timing of effort applied to them has been inconsistent, so these CPUE series may be uninformative.

Catch and effort data from each fishery, from 1989 to 1997, are standardised using three alternative multiple regression models, and the year effects are assessed for adequacy as possible indices of stock abundance. There were inadequate data to carry out CPUE analysis for the southern (HAK1) fishery. The models for the other fisheries were generally able to explain more than 0.3 of the variation in log(CPUE), but none of the indices suggested any consistent trends in CPUE that might be useful in assessing the state of the stocks.

The linear model is sensitive to the choice of arbitrary constant added to the catch and produced unreliable indices. The gamma model includes the zero catches without the addition of an arbitrary constant, but is not very responsive to the information content of them. The combined model models the success rate (whether or not the species of interest was encountered in the tow) and the catch rate (in successful tows) separately. The combined index presents the information content of both series in an intuitively sensible way.

The Stock Assessment Working Group chose to incorporate indices for the HAK7 mixed fishery, the HAK7, and the HAK4 target fisheries using the combined linear/logistic model in the modelling for the 1998 stock assessment.

2. INTRODUCTION

Hake is the most abundant bycatch species in the west coast South Island hoki fishery and is now being targeted off the west coast (HAK7), on the Chatham Rise (HAK 4), and in the region of the Snares and Auckland Islands (HAK1).

CPUE analysis for hake was first carried out on data from 1986 to 1990 by Colman *et al.* (1991). They considered the hoki bycatch fishery in HAK7 only, including only tows from south of 42° S. Because of evidence of under reporting of hake bycatch, only data from those vessels carrying scientific observers were used. They applied a multiple regression model to log-transformed CPUE. All zero catches were deleted from the dataset.

Colman & Vignaux (1992) refined and updated the CPUE analysis of the WCSI bycatch fishery for 1986–91. They incorporated all observer-collected data available rather than a subset and also included the zero catches. They applied a linear model to log-transformed data in which zero catches had been replaced with an arbitrary catch of 1 kg. The explanatory variables describing the tow were coarsely stratified and entered the model as categorical variables.

These earlier studies concluded that interpretation of hake CPUE indices derived from the hoki target fishery was difficult as it appeared to be highly sensitive to changes in fishing practices resulting from amendments to the legislation.

However, over the last few years there appear to have been no major changes in the fishery, so it was proposed that this study revisit CPUE for hake in the west coast hoki fishery and also analyse data from the various hake target fisheries to see if it is feasible to use CPUE analysis to derive useful abundance indices for hake stocks.

The approach taken in this paper was largely determined by the methods (repeated below) specified in the MFish contract for project MID9701. Deviation from these was the evaluation of three (rather than two) models currently in vogue for standardising CPUE.

- Where there are "adequate data" (i.e., at least 100 tows in each of at least 4 years) they will be applied to....
- ..two standardised multivariate CPUE models both of which attempt to minimise residual deviance: a lognormal linear model (Vignaux 1993) and a generalised linear model (Vignaux 1994).
- Parameters will be added to the models in a stepwise procedure until less than 0.5% improvement in the R^2 is seen.
- The final model for each fishstock will be considered a feasible method of monitoring stock abundance if it explains more than 30% of the residual deviance.

3. THE MODELS

Data from each of the fisheries examined was applied to three multiple regression models for standardising CPUE.

- A lognormal linear model described by Vignaux (1993) which has provided the base case for many similar CPUE standardisations in recent years.
- A poisson model with a gamma error function (gamma log-link model) described by Vignaux (1994).
- A combined model proposed by Vignaux (1994) which actually consists of two separate models, a logistic model, and a lognormal linear model, fitted consecutively. The resulting indices are combined mathematically.

The lognormal linear model is applied here to all commercial catch and effort data and consequently includes a much higher proportion of zero catches than Colman &Vignaux (1992) reported. An arbitrary constant was added to all catches (rather than just to zero catches), and for one stock (WCSI) the impact of the choice of that constant was assessed by recalculating the index for a range of values of constant. Stepwise selection of continuous variables (as linear terms until the improvement in R^2 was less then 0.5 %) has been retained and higher-order polynomials of those variables were included if their t-statistic was significant. The categorical variables year and nationality were then added to the model by creating dummy variables for each level (i.e., each year and each nationality). They were added (rather than selected) into the model so that use of the automatic stepwise selection tool in SAS v6.12 did not risk the exclusion of some levels of those factors.

The gamma log-link model deals with the zeros in a more mathematically robust way than the lognormal linear model. It allows the transformation of the response variable (in this case a log link) and the variance function (in this case proportional to the mean) to be specified separately. The advantage is that, because the fit is evaluated in untransformed space (Chambers & Hastie 1992), values of zero CPUE can be left in the model and no arbitrary replacement is required. For comparability, the same variables that were selected into the linear model were fitted using a log-link function and a gamma error distribution, to the raw CPUE of all tows.

The combined model was proposed by Vignaux (1994) for better combining catch rate information from successful tows, with the proportion of tows that were successful. Zero hake catches in tows targeted on hoki are thought more likely to represent a low density of hake than the choice of target species (Colman & Vignaux 1992) so, although a problem mathematically, they nevertheless may contain important information on this fishery.

The combined model splits the problem in two; first, how has the success rate of tows changed from year to year (where success is a non-zero tow) and secondly, how has the catch rate in the successful tows changed from year to year. It thus has two components fitted consecutively; a logistic regression on the success or otherwise (1 or 0) of each tow, and a lognormal linear regression on CPUE from just those tows that were successful (thus no zeros to contend with). Hence a change in abundance could be detected in either index and when the two indices are combined as described by Vignaux (1994) the effect is to modify the CPUE index (for successful tows only) by the ratio of successful to unsuccessful tows.

In the logistic component of the combined model, variables were selected in a stepwise fashion as before. The criterion for entry into the model was a significance level of 0.05. The year effect must be calculated by applying the inverse of the logit function to the coefficients of the model [that is, 1/(1+exp(coeff.))]. The pattern of these year effects is similar to that of the proportion of tows for which a zero catch of hake was reported. In the lognormal linear component of the combined model, variables were selected in the same way as described for the linear model of all tows. A constant was not added to the catch in this case.

When considering the results, the linear index calculated on successful tows only represents the extreme view that zero tows are irrelevant; that they represent either unreported catch or active avoidance of hake, and that incorporating them would bias the CPUE index downwards. The combined index attempts to incorporate the information content of the zero catches as if they represented zero abundance of hake.

The combined index was calculated from that of the linear component (of catch rate of successful tows), and that of the logistic component, using the proportion of zero tows in the first year $P0_1$

Combined index = $(1/(1\text{-}P0_1 * (1\text{- logistic index}))) * \text{linear index}$

where the logistic index is the exponential of the negative relative year effect coefficient from the logistic model, and the linear index is the exponential of the relative year effect coefficient from the linear model of non-zero catches. For those years with a higher than mean proportion of zero catches, the effect is to modify the CPUE index downwards and in years with a higher success rate, modify it upwards.

4. THE DATA

Catch and effort data from Trawl, Catch, Effort and Processing Returns (TCEPR), entered in the Ministry of Fisheries database as at 16 December 1997, were extracted and provided by MFish for all tows that targeted hake, and additionally, for the area defined as HAK7, for all tows where hoki was the target species. In each case, tows with zero catch of hake were included. The data for 1997 are probably incomplete and hence may not be representative owing to the seasonal nature of the fishery. Variables are identified within the test by italics.

All variables were examined by frequency histogram to establish reasonable bounds and then outliers and gross errors examined and documented (T. Kendrick, unpublished data). The many and obvious errors in *bottom depth* (usually by a factor of 10) and in *speed* (usually 4 misread as a 7) were corrected as described by Vignaux (1993). Records with negative or zero *tow duration* were deleted. There were many records with missing *total catch*, but the hake catch was often substantial and reasonable, so they were retained. Fishing gear descriptors contained many errors (i.e., values outside of expected bounds or not correlated when they should be, such as *net width* and *net height*). Often these values might readily be corrected on the basis of surrounding records, but in this study they were simply not offered to the models.

Selecting subsets of the available data as a means of standardisation was avoided as much as possible, the data extracts were constrained within certain geographical limits, *month* of the year and *target species*, but these were broad, and within them, tows from both bottom and midwater trawls were included with *fishing method* defined as a categorical variable and the model allowed to account for any overall difference in catch rate. Similarly, for HAK7 where the hoki fishery was revisited, tows targeted on both hoki and hake were included, *target species* was used as a categorical factor, and the model allowed to account for the difference.

4.1 WCSI

This dataset (representing the WCSI combined hoki/hake fishery) includes those tows targeted on either hake or hoki, between longitudes 167 and 172° and north of 46 ° S (to exclude tows in Cook Strait and at Puysegur), from depths greater than 275 m, from June to December inclusive, for the years 1989 to 1997, (Tables 1 and 2).

4.2 HAK7

This is a subset of the WCSI dataset; it includes only those tows targeted on hake (the HAK7 target fishery).

Table 1: The number of tows available for the WCSI and HAK7 analyses. The dataset for the combined fishery (WCSI) includes all tows, the dataset for the target fishery (HAK7) includes only those where hake was the target species

Target species	**Fishing method**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
Hake	Bottom Trawl	5	34	0	130	85	36	6	44	22
	Mid water	2	42	4	51	379	192	405	128	12
Target (HAK7)		7	76	4	181	464	228	411	172	34
Hoki	Bottom Trawl	508	1121	1336	813	1585	1394	1273	1578	1153
	Mid water	5879	6918	6874	5324	5219	7177	6752	5320	4669
Combined (WCSI)		6394	8115	8214	6318	7268	8799	8436	7070	5856

Table 2: The proportion of tows for WCSI and HAK7 for which a zero hake catch was reported. Data are missing (-) where there were no tows reported

Target species	**Fishing method**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
Hake	Bottom trawl	0.000	0.206	-	0.023	0.106	0.083	0.167	0.091	0.091
	Mid water	0.000	0.071	0.000	0.078	0.058	0.031	0.052	0.047	0.000
Target (HAK7)		0.000	0.132	0.000	0.039	0.067	0.039	0.054	0.058	0.059
Hoki	BottomTrawl	0.344	0.370	0.499	0.590	0.437	0.383	0.408	0.245	0.212
	Mid water	0.418	0.450	0.619	0.775	0.834	0.858	0.677	0.504	0.522
Combined (WCSI)		0.412	0.435	0.598	0.730	0.698	0.761	0.605	0.436	0.455

4.3 HAK4

This dataset includes only those tows targeted on hake in MFish Statistical Area 404 during September to January inclusive for 1989 to 1997. January was included in the previous year for the purpose of this analysis (Table 3).

Table 3: Number of tows available for the HAK4 analysis

Fishing method	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
Bottom trawl	107	42	181	378	206	324	305	235	11
Mid water trawl	0	0	13	104	148	159	31	5	0
Target fishery	107	42	194	482	354	483	336	241	11

Table 4: Proportion of tows for HAK4 for which a zero hake catch was reported. Data are missing (-) where there were no tows reported

Fishing method	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
Bottom trawl	0.028	0.071	0.022	0.000	0.005	0.012	0.000	0.030	0.000
Mid water trawl	-	-	0.077	0.115	0.000	0.025	0.161	0.000	-
Target fishery	0.028	0.071	0.026	0.025	0.003	0.017	0.015	0.029	0.000

4.4 HAK1

This dataset includes all tows targeted on hake in MFish Statistical Areas 28, 504, and 602 between June and December inclusive, for 1989 to 1997 (Table 5).

There were not sufficient data to proceed with any meaningful CPUE analysis.

Table 5: Number of tows available for the HAK1 analysis

Fishing method	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
Bottom trawl	46	26	4	31	11	40	71	54	1
Mid water trawl	0	0	0	0	16	0	0	0	0
Target fishery	46	26	4	31	27	40	71	54	1

5. THE VARIABLES

The response variable, CPUE, is taken to be catch (kg) per nautical mile. While this can mean something quite different depending on the size of the net opening, the quality of the net description data was poor, and *fishing method* was allowed to account in the model for some of the observed difference. *Length of tow* was calculated from *speed* of the vessel and *duration* of the tow. Variables are given in Table 6.

The logarithm of CPUE (calculated on catch plus a small constant) was used in the linear model, but CPUE, calculated on the actual catch with no constant added, was used in the gamma model and in the linear part of the combined model. A binary variable ***success*** (whether a catch of hake was recorded or not) was the response variable for the logistic part of the combined model.

Dummy variables were created for each level of the categorical explanatory variables *year*, *nationality*, *fishing method*, and *target species*, but where there was a choice, continuous variables were offered for inclusion in the models rather than categorical, e.g., *latitude* and *longitude*, rather than *statistical area*. *Depth of the gear* off the bottom was calculated and included as a continuous variable. It was often selected even though highly correlated with *fishing method.*

Table 6: Explanatory variables considered for inclusion in the models. Variables in bold were categorical, the others were all considered up to cubic order polynomials

Variable	Description
year	Calendar year except for HAK4 (September to January)
season	Day of year
(month)	Month of year (as alternative to season)
target species	Target species for that tow
latitude	Latitude at start of tow
longitude	Longitude at start of tow
bottom depth	Depth of bottom
fishing method	Bottom or mid-water gear used
height of headline	Difference between gear depth and bottom depth
nationality	Nationality of vessel
vessel length	Overall length of vessel
vessel draught	Draft of vessel
vessel breadth	Breadth of vessel
vessel tonnage	Gross tonnage of vessel
kilowatts	Power of vessel engine in kilowatts
year built	Year the vessel was built

6. RESULTS

6.1 WCSI Mixed Fishery (targeted on hoki or hake)

A feature of this dataset is the large proportion of zero hake catches reported (Table 2).

The Linear model

The order in which variables were selected, the model R^2, and the order polynomial at which they were selected, are given in Appendix 1. The four most important explanatory variables were *bottom depth*, *latitude*, *height of the headline* off the bottom, and *fishing method.*

The linear model of catch rate of all tows explained 0.34 of the variation in log(CPUE). The relative year effects and standard errors are shown in Figure 1 and given in Table 8. The index shows a decline from 1989 to 1992, (this contrasts with observer-only CPUE reported by Vignaux (1992) then, after 4 years at less than 10% of the 1989 rate, an apparently dramatic increase in 1996 and 1997 (Figure 1). The increase in year effects in 1996 and 1997, however, is also associated with a similar increase in standard errors.

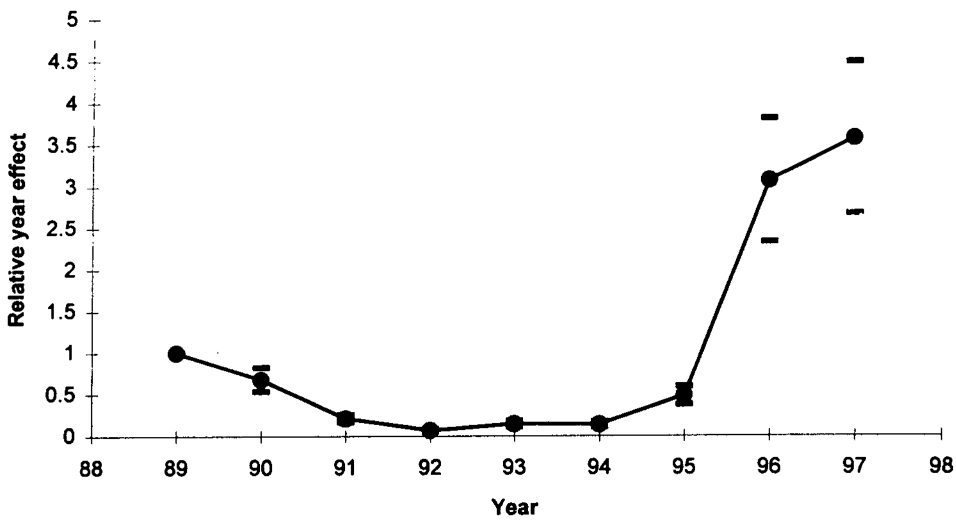

Figure 1: The linear model CPUE indices for the WCSI mixed hoki/hake fishery. The error bars are ± 2 standard errors.

The index is sensitive to the choice of constant added to the catch (Figure 2). Model selection is also sensitive to the value of the constant. When different values for the constant (c) (0.0001, 1.0, and 5.0 kg) were used, the indices produced were quite different in terms of magnitude, though some of the difference between series may be due to the different variables selected in each model.

The value of the constant is arbitrary, it distorts the regression, and the smaller values carry a lot of weight in a log transform. In this case, however, comparison of R^2 suggested a better fit when using a very small constant (0.0001 kg) [R^2 values ranged from 0.34 (C = 0.0001 kg) to 0.321 (C = 5 kg)].

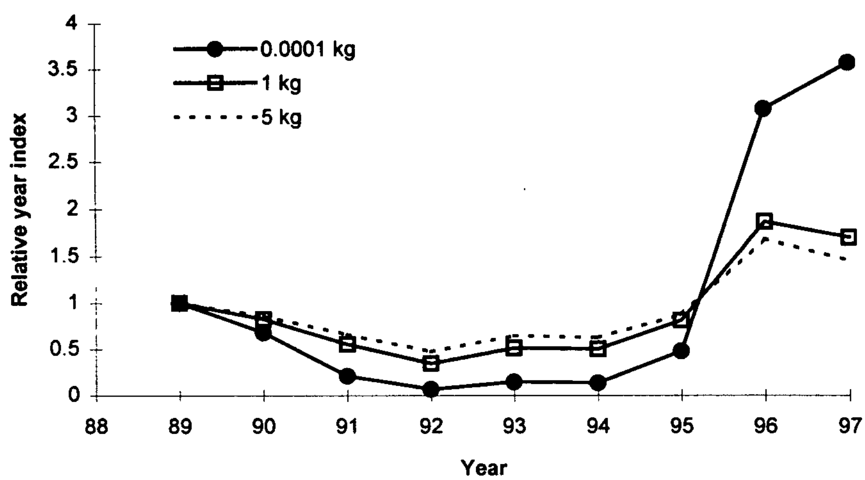

Figure 2: Sensitivity of the WCSI mixed fishery linear CPUE indices to the value of C (the constant added to catch). The model selection procedure was applied separately in each case.

The gamma log-link model

The same variables that were selected for the linear model (*see* Appendix 1) were fitted here to raw CPUE of all tows. The relative year effects and standard errors are shown in Figure 3 and given in Table 8.

The index shows a lot of variation with no clear trend (*see* Figure 3). It differs from the linear index by the absence of the large increase from 1995 to 1996.

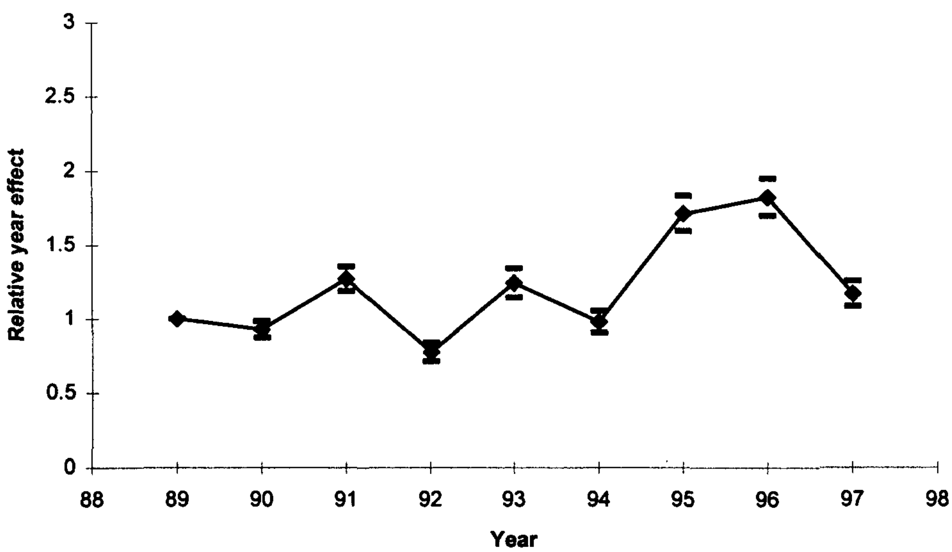

Figure 3: The gamma log-link CPUE indices for the WCSI mixed hoki/hake fishery. The error bars are ± 2 standard errors

The combined model

The logistic component of the combined model was fitted to the success or otherwise (0 or 1) of all tows. The order in which variables were selected, the score chi-square, and the order polynomial at which they were selected are given in Appendix 2. The five most important explanatory variables were *bottom depth*, *longitude*, *fishing method*, *height of the headline* off the bottom, and *latitude*, which agrees closely with those selected by the linear model of CPUE of all tows. The relative year effects are given in Table 7: they resemble the pattern of the proportion of tows for which a zero catch of hake was reported.

The linear component of the combined model was fitted to the log(CPUE) of only those tows that were successful. The order in which variables were selected, the model R^2, and the order polynomial at which they were selected are given in Appendix 3. The most important explanatory variable was target species, which makes sense, and the next five were *latitude*, *kilowatts*, *bottom depth*, *height of headline* off bottom, and *fishing method*. With the exception of the power of the vessel (kilowatts) this agrees closely with those selected by the other models. The relative year effects and standard errors are shown in Figure 4 and given in Table 7. The model explained 0.22 of the variance in log(CPUE)[1]. This model yielded an index similar to that from the gamma log-link model in spite of it being fitted to a much smaller dataset (*see* Table 1).

[1] Using the same model as selected for the linear model including zero tows (Section 6.1.1) yielded an R^2 of 0.224.

The two indices give similar signals about the stock, with recent years of highest (relative to 1989) CPUE coinciding with lowest (relative to 1989) proportions of zero catch. The combined index (Figure 4, Table 7) resembles that from the linear model of all tows (*see* Figure 1).

Table 7: Relative year effects from the logistic model of the proportion zero tows, the linear model of catch rate of successful tows , and the combined index from 1989-97 in the WCSI mixed hoki/hake fishery. Base year is 1989

Model	1989	1990	1991	1992	1993	1994	1995	1996	1997
Logistic	1.000	1.051	1.273	1.503	1.406	1.418	1.044	0.793	0.719
Linear	1.000	0.933	1.157	0.736	1.076	1.007	1.501	1.845	1.215
Combined	1.000	0.894	0.884	0.401	0.689	0.633	1.446	2.150	1.484

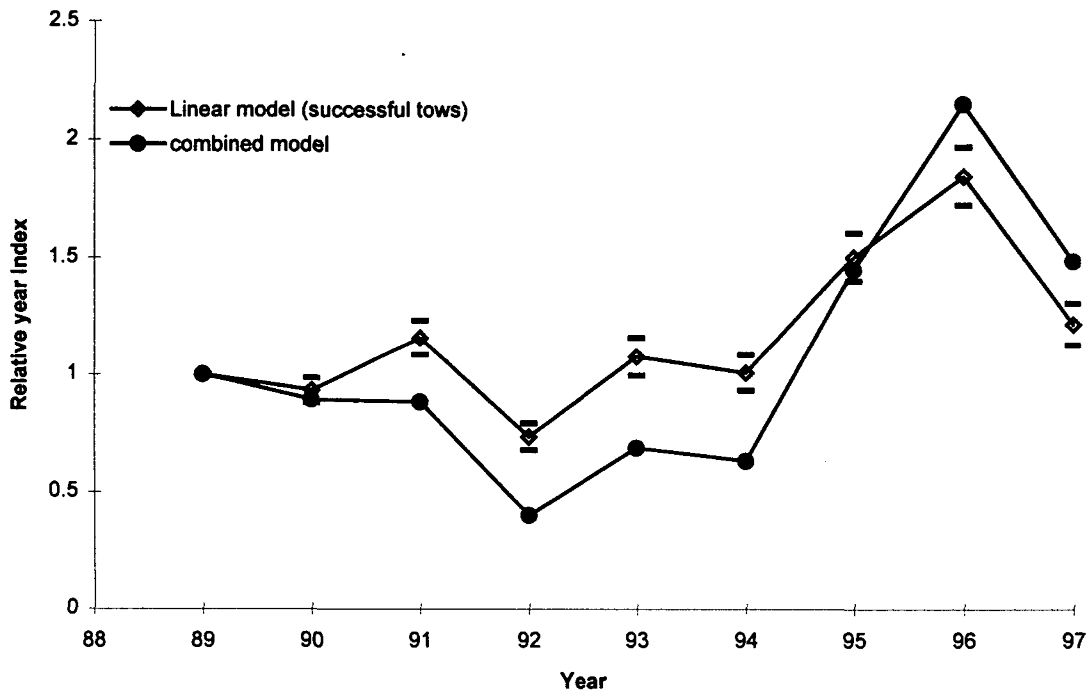

Figure 4. Relative year effects from the linear model of catch rate on successful days and from the combined model, for the mixed hoki/hake WCSI fishery. The error bars are ± 2 standard errors of the linear index.

Table 8: Relative year effects and standard errors, where calculable, from the linear model, the gamma log-link model, and the combined model, for 1989–97 in the WCSI mixed hoki/hake fishery

Year	Linear index	Std. error	Gamma index	Std. error	Combined index
1989	1.000		1.000		1.000
1990	0.679	0.072	0.927	0.027	0.894
1991	0.212	0.023	1.269	0.041	0.884
1992	0.069	0.008	0.773	0.032	0.401
1993	0.145	0.017	1.242	0.049	0.689
1994	0.138	0.015	0.979	0.038	0.633
1995	0.481	0.055	1.711	0.060	1.446
1996	3.064	0.367	1.821	0.062	2.150
1997	3.564	0.449	1.168	0.043	1.484

6.2 WCSI (HAK7) Target Fishery

All the data available, from 1989 to 1997, were included in the models for this target fishery though an insufficient number of data points means that the first 3 years are unreliable. The base year for these analyses is 1993 to avoid using zero for $P0_1$ in the combined model.

The linear model

The linear model of catch rate of all tows explained 0.19 of the variation in log(CPUE). The order in which variables were selected, the model R^2, and the order polynomial at which they were selected are given in Appendix 4. The four most important explanatory variables were *bottom depth*, *vessel draught*, *height of the headline* off the bottom, and *year built*. The relative year effects and standard errors are shown in Figure 5 and given in Table 10.

The index shows a decline from 1992 to 1996. There is inadequate data for 1997 and for the years 1989 through 1991, and the index for those years may be unreliable.

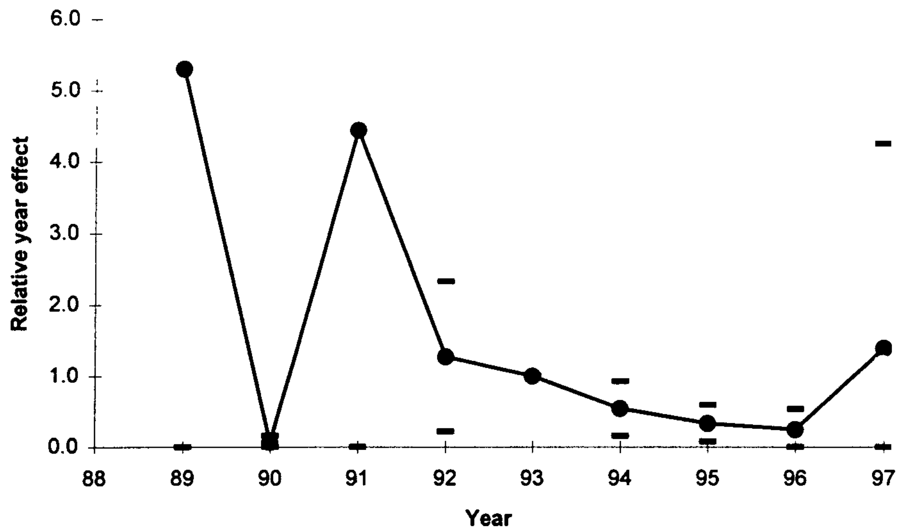

Figure 5: The linear model CPUE indices for the HAK7 target fishery. The constant added to catch is 0.0001 kg. The error bars are ± 2 standard errors

The gamma log-link model

The gamma log-link model of catch rate was fitted to the raw CPUE of all tows. The same variables that were selected for the linear model (*see* Appendix 3) were fitted here to raw CPUE of all tows. The relative year effects and standard errors are shown in Figure 6 and given in Table 10.

The index shows a lot of variation with no clear trend (Figure 6).

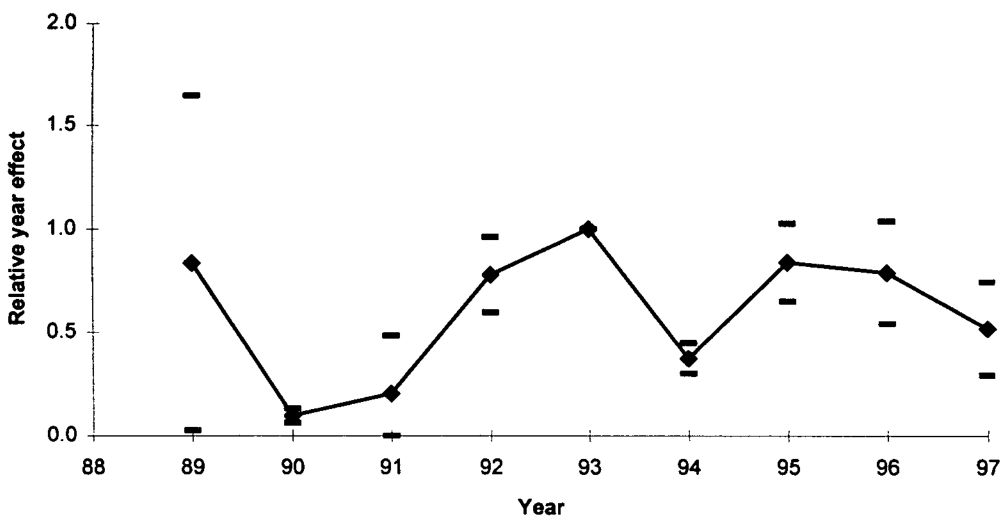

Figure 6: The gamma log-link CPUE indices for the HAK7 target fishery. The error bars are ± 2 standard errors

The combined model

The logistic component of the combined model was fitted to the success or otherwise (0 or 1) of all tows. The only variables selected were *bottom depth* (cubic), *vessel draught* (cubic), and *height of the headline* off the bottom. These were also the three most important variables selected into the linear model of CPUE of all tows.

The linear component of the combined model was fitted to log(CPUE) of only those tows which were successful, i.e., reported a catch of hake. The order in which variables were selected, the score chi-square, and the order polynomial at which they were selected are given in Appendix 5. The three most important explanatory variables were *bottom depth, vessel draught,* and *year built* which agrees closely with the other models: *latitude and longitude* however, were more important than *height of the headline* off the bottom. The relative year effects and standard errors are shown in Figure 7 and given in Table 9. The model explained 0.33 of the variation in log(CPUE).

It is interesting that the logistic and linear components give contradictory signals, i.e., years of higher CPUE and low proportion of zero catches do not coincide. However, the proportion of zero catches is low (table 2), as would be expected in a target fishery, and the index obtained from modelling the successful tows is not modified greatly by combination with the logistic index. The combined model index confirms the pattern seen in the gamma model index and though it appears quite different from that from the linear model of all tows, it nevertheless confirms the observation of a lot of variation in CPUE with no clear trend.

Table 9: Relative year effects from the logistic model of the proportion zero tows, the linear model of catch rate of successful tows, and the combined index from 1989-97 in the HAK7 target Hake fishery. Base year was 1993

Model	1989	1990	1991	1992	1993	1994	1995	1996	1997
Logistic	0.000	1.149	0.000	0.842	1.000	0.786	1.146	1.300	0.728
Linear	0.825	0.081	0.198	0.690	1.000	0.411	0.854	0.689	0.502
Combined	0.884	0.079	0.212	0.703	1.000	0.420	0.835	0.652	0.517

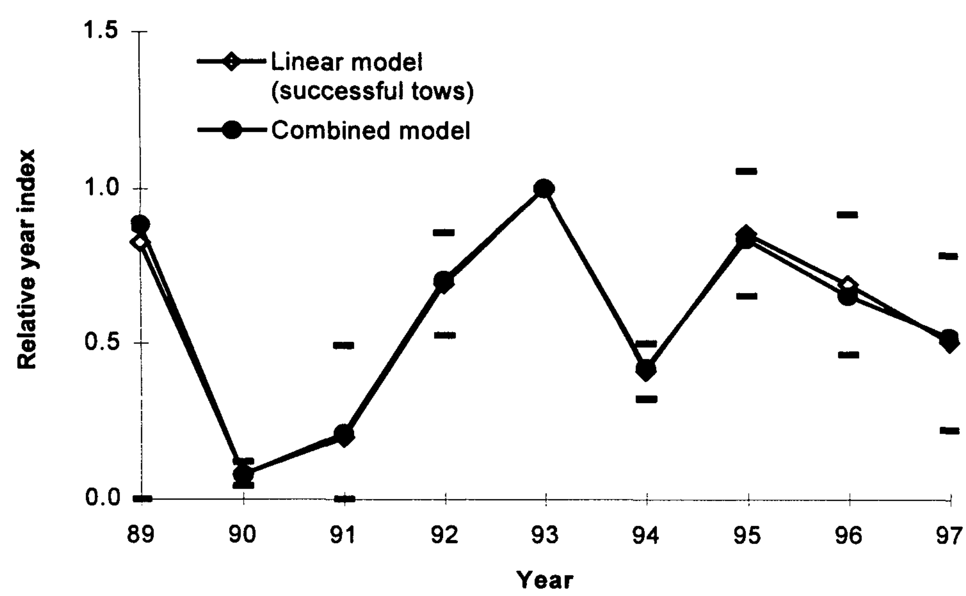

Figure 7. Relative year effects from the linear model of successful tows, and the combined model, for the HAK7 target fishery. The error bars are ± 2 standard errors of the linear model index

Table 10: Relative year effects and standard errors where calculable, from the linear model, the gamma log-link model, and the combined model for the 1989-97 HAK7 target fishery

year	Linear index	Std. error	Gamma index	Std. error	Combined Index
1989	5.3005	40.534	0.8350	0.405	0.8838
1990	0.0681	0.046	0.0967	0.017	0.0793
1991	4.4266	168.101	0.2028	0.140	0.2121
1992	1.2728	0.526	0.7784	0.092	0.7031
1993	1.0000	0.000	1.0000	0.000	1.0000
1994	0.5400	0.191	0.3737	0.037	0.4204
1995	0.3333	0.130	0.8372	0.094	0.8345
1996	0.2447	0.144	0.7881	0.124	0.6520
1997	1.3865	1.421	0.5181	0.113	0.5170

6.3 THE CHATHAM RISE (HAK 4) TARGET FISHERY

The linear model

The linear model of catch rate of all tows explained 0.14 of the variation in log(CPUE). The order in which variables were selected, the model R^2, and the order polynomial at which they were selected are given in Appendix 6. The three most important explanatory variables were *draught* of the vessel, *fishing method*, and *latitude*. The relative year effects and standard errors are shown in Figure 8 and given in Table 12.

The index estimates are imprecise and the downward trend is not significant (Figure 8).

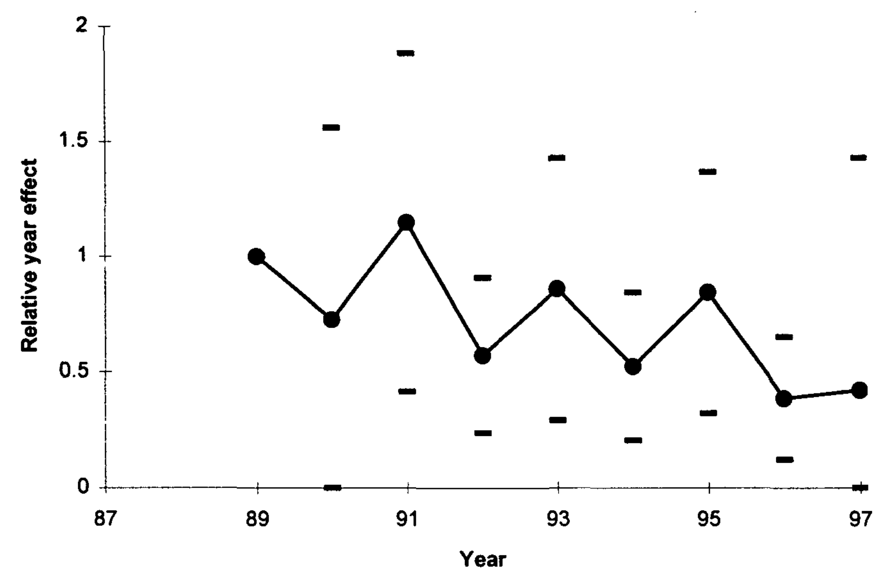

Figure 8: The linear CPUE indices for the HAK 4 target fishery. The error bars are ± 2 standard errors

The gamma log-link model

The gamma log-link model of catch rate was fitted to the raw CPUE of all tows. The same variables that were selected for the linear model (*see* Appendix 6) were fitted here to raw CPUE of all tows. The relative year effects and standard errors are shown in Figure 9 and given in Table 12.

The index shows a downward trend from 1989 to 1997 (Figure 9), although most of the estimates are imprecise.

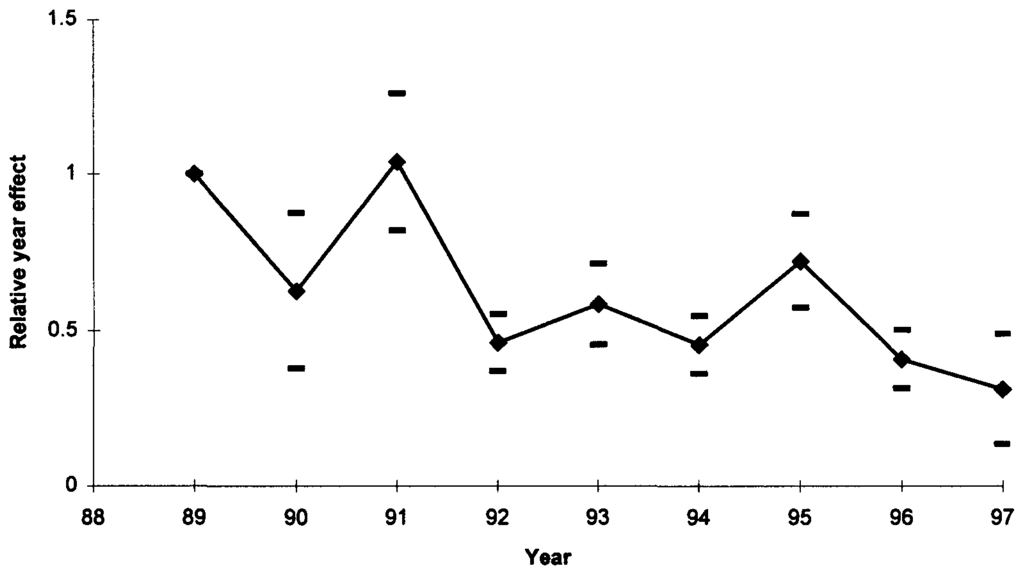

Figure 9: The gamma log-link CPUE indices for the HAK 4 target fishery. The error bars are ± 2 standard errors.

The combined model

The logistic component of the combined model was fitted to the success or otherwise (0 or 1) of all tows. The only variables selected were *fishing method* and *latitude*. These were included in the first three selected into the linear model of CPUE of all tows. The relative year effects are given in Table 11. The pattern of the index is similar to that of the proportion of tows for which a zero catch of hake was reported.

The linear component of the combined model was fitted to log(CPUE) of only those tows which were successful, i.e., reported a catch of hake. The model explained 0.32 of the variation in log(CPUE). The order in which variables were selected, the model R^2, and the order polynomial at which they were selected are given in Appendix 7. The three most important explanatory variables were *latitude* of the vessel, *draught,* and *fishing method,* which agrees closely with those selected by the other models. The relative year effects and standard errors are shown in Figure 10 and given in Table 11.

The two constituent series both give the same signal about the stock, with the (standardised) proportion of zero catches increasing as CPUE decreases. The combined index (Figure 10, Table 11) shows a downward trend, but precision of estimates is poor.

Table 11: Relative year effects from the logistic model of the proportion of zero tows, the linear model of catch rate of successful tows, and the combined index from 1989-97 in the HAK4 target fishery. Base year is 1989

Model	1989	1990	1991	1992	1993	1994	1995	1996	1997
Logistic	1.000	1.190	0.223	0.155	0.024	0.244	0.181	0.705	0.000
Linear	1.000	0.692	0.975	0.424	0.514	0.407	0.574	0.336	0.316
Combined	1.000	0.683	1.000	0.436	0.529	0.417	0.589	0.340	0.325

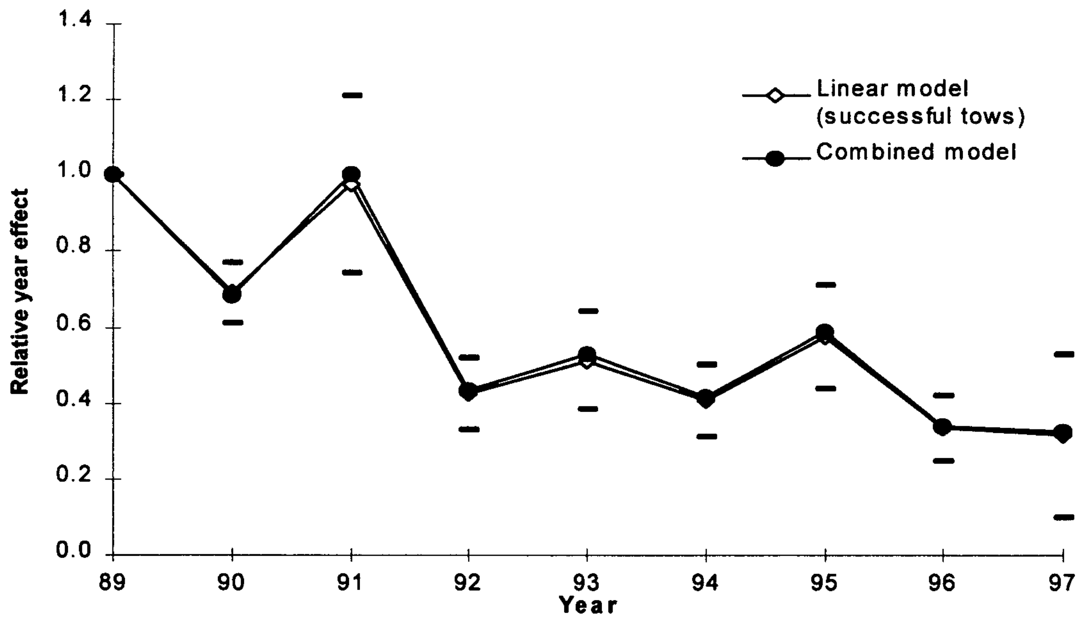

Figure 10: Relative year effects from the linear model of catch rate for successful tows, and the combined model, for the HAK4 target hake fishery. The error bars are ± 2 standard error's of the linear index

The three models all yield very similar indices for this stock (Table 12).

Table 12: Relative year effects and standard errors where calculable, from the linear model, the gamma log-link model, and the combined model, for 1989-97 in the HAK4 target fishery.

Year	Linear index	Std. error	Gamma index	Std. error	Combined index
1989	1.000		1.000		1.000
1990	0.727	0.416	0.627	0.124	0.683
1991	1.149	0.367	1.038	0.110	0.999
1992	0.572	0.168	0.461	0.045	0.436
1993	0.859	0.284	0.584	0.065	0.529
1994	0.524	0.159	0.454	0.046	0.417
1995	0.844	0.261	0.721	0.074	0.589
1996	0.386	0.132	0.408	0.047	0.340
1997	0.421	0.504	0.311	0.088	0.325

6.4 THE SOUTHERN (HAK1) TARGET FISHERY

There are not enough data available from this target fishery to warrant doing any CPUE analyses.

7. ADDENDUM

The Stock Assessment Working Group were interested to see a seasonal effect, particularly for the newer target fisheries. Season, defined as day of the year, was not selected into any of the models for the target fisheries, suggesting no regular pattern of catch rate with day and perhaps too fine a temporal scale. Season was replaced by the categorical variable month, which entered as dummy variables. The combined model index was recalculated for each fishery using otherwise the same models as before. The portion of the time series that is greyed is that for which there are adequate data, and which the Stock Assessment Working Group chose to incorporate into the modelling for the 1987–88 hake stock assessment.

7.1 HAK7 Mixed fishery

R^2 decreased from 0.226 to 0.225 The revised indices are;

1989	1990	1991	1992	1993	1994	1995	1996	1997
1.00	0.90	0.89	0.40	0.70	0.64	1.43	2.18	1.50

7.2 HAK7 Target fishery

R^2 increased from 0.329 to 0.333. The month effects were not significant. The revised indices are;

1989	1990	1991	1992	1993	1994	1995	1996	1997
0.88	0.06	0.13	0.72	1.00	0.38	0.68	0.60	0.44

7.3 HAK4 Target fishery

R^2 increased from 0.324 to 0.341. The month effects were not significant. The revised indices are;

1989	1990	1991	1992	1993	1994	1995	1996	1997
1.00	0.58	0.95	0.42	0.55	0.39	0.61	0.30	0.27

10.0 DISCUSSION

Analysis of CPUE data has been completed in order to obtain CPUE indices of abundance for the targeted and bycatch west coast hake fisheries of the South Island and the targeted hake fishery on the Chatham Rise. Three alternative models are presented. The indices are generally imprecise and do not reveal significant trends during the period 1989–97. Indices for the HAK7 and HAK4 target fisheries are truncated due to insufficient data in the earlier years and in the most recent year. There were insufficient data to carry out any CPUE analysis for the HAK1 target fishery.

The lognormal linear model of catch rate of all tows is sensitive to the value of the arbitrary constant added to catch. The choice of constant is arbitrary, and in a fishery such as the WCSI mixed fishery, the high proportion of zero tows and the log transform means that the constant has a lot of influence. The linear model thus yields an unreliable index.

The gamma model is superior mathematically in its approach to dealing with zeros, but the pattern of the index from this model resembles the pattern from the linear model of catch rate when all zero tows have been removed. In the WCSI fishery, where the zero tows may be informative, the gamma model may be inappropriate.

The combined model considers both the success rate of tows and the catch rate of successful tows, and combines both sources of information in a standardised way.

There was generally very close agreement between models for each fishery about the relative importance of explanatory variables. The linear model of all tows yielded indices very different from the other two models, and this may be due to the arbitrary constant that was added to catch. In the HAK4 fishery, all three models yielded very similar CPUE indices.

As expected, the fisheries for which more observed data are available gave greater precision of estimates, though this might not reflect better information about stock abundance. The pattern of the combined model index for the WCSI mixed fishery reflects the significant increase in CPUE and the corresponding decrease in zero tows seen for hake from 1994 to 1996. This is in spite of anecdotal information that fishing company policy has been to minimise hake catch. It is not clear, therefore, that either series is indexing abundance of hake.

The combined model index for the HAK7 target fishery shows a lot of variation and no trend. The indices of catch rate and of success rate send contradictory signals, though the proportion of zero tows is low and contributes little to the index. There has not been a consistent seasonal pattern of effort in this fishery across years, and it is not clear that the fishery representatively samples the fishstock.

The combined model index for the HAK4 target fishery shows a fairly consistent decline from 1991. The proportion of zero tows increases as CPUE decreases in a manner consistent with CPUE indexing abundance.

The Stock Assessment Working Group chose to incorporate three series into the stock assessment modelling. They are the HAK7 combined (1989–97), HAK7 target (1992–96) and HAK4 target (1992–96) indices using the combined model and a month effect.

8. ACKNOWLEDGMENTS

This work was financially supported by the Ministry of Fisheries, project number MID9701. Sincere thanks are due to A. David McDonald (CSIRO) for help and advice and to Paul Breen and Alistair Dunn who suggested many improvements to earlier versions of the manuscript.

9. REFERENCES

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

Colman, J.A., Stocker, M., & Pikitch, E. 1991: Assessment of hake (*Merluccius australis*) stocks for the 1991–92 fishing year. *New Zealand Fisheries Assessment Research Document 91/14.* 29 p.

Colman, J.A. & Vignaux, M. 1992: Assessment of hake (*Merluccius australis*) for the 1992–93 fishing year. *New Zealand Fisheries Assessment Research Document 92/17.* 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.* 29 p.

Appendix 1: Summary of the stepwise procedure for fitting the linear model to the mixed hoki/hake WCSI fishery catch rate

Variable entered	Order of polynomial	Model R^2
bottom depth	3	0.1780
start latitude	2	0.2261
net off bottom	3	0.2535
fishing method	categorical	0.2662
target species	categorical	0.2740
draught	3	0.2806
breadth	3	0.2957
kilowatts	3	0.2957
season	2	0.2984
length	3	0.2999
year & nationality included	dummy variables	**0.3418**

Appendix 2: Summary of the stepwise procedure for fitting the logistic model to the success rate of tows in the mixed hoki/hake WCSI fishery

Variable entered	Order of polynomial	chi-square
bottom depth	2	11360.5
start longitude	1	3127.4
fishing method	categorical	1970.2
net off bottom	3	1125.3
start latitude	3	986.1
draught	3	616.0
breadth	3	917.3
kilowatts	3	682.6
target species	categorical	319.1
season	3	220.4
length	3	196.0
year built	1	98.5
tonnage	3	69.5
year & nationality included	dummy variables	

Appendix 3: Summary of the stepwise procedure for fitting the linear model to the catch rate of successful tows in the mixed hoki/hake WCSI fishery

Variable entered	polynomial	Model R^2
target species	categorical	0.0748
start latitude	2	0.1142
kilowatts	1	0.1375
bottom depth	1	0.1554
net off bottom	2	0.1601
fishing method	categorical	0.1668
season	3	0.1685
year built	1	0.1696
length	1	0.1706
tonnage	3	0.1724
year & nationality included	dummy variables	**0.2261**

Appendix 4: Summary of the stepwise procedure for fitting the linear model to the catch rate in the HAK7 target fishery

Variable entered	Order of polynomial	Model R^2
bottom depth	3	0.0754
draught	1	0.0948
net off bottom	3	0.1106
year built	1	0.1186
fishing method	categorical.	0.1217
start latitude	1	0.1307
start longitude	1	0.1322
year & nationality included	dummy variables	**0.1947**

Appendix 5: Summary of the stepwise procedure for fitting the linear model to the catch rate of successful tows in the target HAK7 fishery

Variable entered	Order of polynomial	Model R^2
bottom depth	3	0.0685
draught	1	0.1154
year built	1	0.1553
start latitude	1	0.1624
start longitude	1	0.1724
net off bottom	2	0.1790
fishing method	categorical	0.1866
length	3	0.1918
tonnage	1	0.1952
kilowatts	3	0.1969
year & nationality included	dummy variables	**0.3299**

Appendix 6: Summary of the stepwise procedure for fitting the linear model to catch rate in the HAK4 target fishery

Variable entered	Order of polynomial	Model R^2
draught	1	0.0079
fishing method	categorical	0.0192
start latitude	3	0.0224
length	3	0.0238
tonnage	2	0.0264
year & nationality included	dummy variables	**0.1403**

Appendix 7: Summary of the stepwise procedure for fitting the linear model to the catch rate of successful tows in the HAK4 target fishery

Variable entered linear	Order of polynomial	Model R^2
start latitude	1	0.0744
draught	2	0.1059
fishing method	categorical	0.1103
bottom depth	2	0.1140
tonnage	3	0.1153
length	3	0.1285
year & nationality included	dummy variables	**0.3237**