# Scenario-C:
The cod predation model.

SAMBA/09/2004

Mian Zhu
Tore Schweder
Gro Hagen

1st March 2004

Tittel/Title:
Scenario-C: The cod predation model.

Forfatter/Author:
Mian Zhu zhumian@hotmail.com
Tore Schweder tore.schweder@econ.uio.no
Gro Hagen Gro.Hagen@nr.no

Sammendrag/Abstract:

The Scenario C project aims at a dynamic minimal realistic model for capelin, herring, cod, harp seals and minke whales in the Barents Sea and for herring in the Norwegian Sea. This report is an attempt to estimate a consumption function for cod from estimated monthly abundances of capelin, herring and younger cod in regions of the Barents Sea, and the observed stomach contents of cods. The data are obtained from Institute of Marine Research.

A discrete choice model for the diet choice and consumption function of cod in the Barents Sea is implemented using the data of cod and three main prey species (herring, capelin, cod) from 1992-2000. It uses the notion of choice probability to describe the foraging behavior. In particular, it estimates how prey selection varies with respect to environmental changes. By using diet choice model, prey consumption can be easily computed and predicted from environmental variables.

In conclusion, the empirical results are quite consistent with the hypothesis that prey selection probability is highly related to the biomass amount of the specific prey. Particularly when the biomass is at low levels, even a small increase will lead to a large probability change, with the more available biomass, the smaller the growth. At high levels, the probability is nearly linear in biomasses. As a whole, the probability change is modeled to have a logistic pattern. In addition, it shows that cod prefers capelin, rather than young cod and herring.

1. Introduction

In this paper, we try to estimate a consumption function for cod from estimated monthly biomass abundances of capelin, herring and cod in regions of the Barents Sea, and observed stomach contents of cods. Biologists and marine researchers have developed several approaches for modeling cod predation. Schweder *et al* (1998) present models of the logistic type for choice probabilities for minke whale and cod predations, and fit the models to data of the type we will consider. We will continue the line of research sketched by Schweder in NAMMCO (2002, page 13) by introducing alternative models for choice probabilities, and fitting these models to cod data. A discrete choice model for the diet choice and consumption function for the cod stock in the Barents Sea is implemented using the stock data of cod and three main prey species (herring, capelin, young cod) from 1992-2000. It uses the notion of choice probability to describe the foraging behavior. In particular, it estimates how prey selection varies with respect to environmental change. By using the diet choice model, prey consumption can be easily computed and predicted from environmental variables.

We will regard the cod as a rational predator. The benefit to a cod of a meal of a certain kind and quantity is the supply of digestible calories and other beneficial food components. The cost is related to the process of digestion and search. It will maximize benefit and minimize cost. However, its utility of consuming a meal of herring is not directly measurable, nor is the cost of searching and capturing the meal. Instead of an explicit cost/benefit model, we will take the view that costs and benefits are combined in a net unobserved utility that is reflected in the choice probabilities for the four categories of meals (capelin, herring, young cod, and other food) conditional on the amount of available quantities of the three fish preys. No data is available regarding the category other food (other fish species, zooplankton) with respect to total quantity. I will therefore assume that it is available in an unknown but constant amount.

2. Data

2.1 Stomach contents analysis

The aim of our study is to model diet choice and consumption function for cod within a discrete choice model. We do not have an exact variable to represent the diet choice on the basis of stomach contents, so we need to transform the consumed amount by prey items, which is a continuous variable, into a discrete choice. What is found in the stomach is what was not processed further in the digestion process. The quantities are therefore not reflecting directly how much was eaten in the last meal or feeding event. Nevertheless, they are good (in fact, the only available) proxies to represent the prey selection results. By previewing the dataset we found the number of different prey items to range from one to three in most stomachs. The weight share of different prey

items varied greatly in the stomachs with mixed content. In order to simplify the statistical exercises, some main assumptions were made.

1) Comparative importance

In mixed prey selection cases, we focus on the most important prey to avoid unclear and misleading selection. Prey items accounting for less than 10% of the whole weight of stomach content was therefore disregarded. Certainly, this is an arbitrary assumption, since the remaining of prey does not represent the actual consumption amount over a specific period of consumption. Digestion is a temporal process and a small amount in the stomach might result from big consumption of food eaten much earlier. Nevertheless, it is not unreasonable that the remains are proportional to the predation intensity. Moreover, we need a clear and simple index for the diet choice. Hence, the comparative importance of various prey items was considered.

2) Three independent feeding events:

The content of each stomach was assumed to consist of three independent feeding events. Each feeding event represents one single choice. Further, the feeding event was assumed entirely to consist of only one type of prey, and to be independent of previous choices made under similar conditions. The purpose of splitting the predation into three feeding events was to give all choices equal weights. The estimation results will be revised later by bootstrapping to reduce inferential errors caused by this simplification.

3) Cannibalism:

Cannibalism in the Northeastern Arctic cod stock plays an important role in cod predation. According to our research interest, we focus on the predation between old cod (older than 3 years) as predator and herring, capelin and young cod (younger than 3 years) as prey. This means we exclude the young cod from the predator data. According to a previous study (Bogstad, 2003), the cod length as predator ranges from 25cm to 125cm. In addition, the proportion of consumed young cod in the stomach of old cod under 50cm is very small. Therefore we have excluded cod whose length is under 50cm from the predator data.

Based on the above assumptions, the weight shares of each prey in a stomach were obtained. In order to decide the importance of prey choice under the three feeding event constraints, the following transformation rule based on the proportions was defined:

Transformation Rule[1]:

---

[1] There exists inconsistency between the individual event transformation rule and a total of three feeding events per stomach. We

Let $event_k$ be the number of feeding events related to prey k, and $share_k$ be the weight share of remains of prey k in one stomach, defined as follows

$$event_k = \begin{cases} 0 & \text{if } share_k \leq 0.1 \\ 1 & \text{if } 0.1 < share_k \leq 0.5 \\ 2 & \text{if } 0.5 < share_k \leq 0.9 \\ 3 & \text{if } 0.9 < share_k \end{cases}$$

subjected to the constraints $\sum_{k=1}^{4} event_k = 3$ and $\sum_{k=1}^{4} share_k = 1$, where k=herring, capelin, young cod, and other food.

According to this transformation rule, there exist two categories with respect to stomach content

1. single prey item, i.e. $event_k = 3$ for one prey k, where one stomach content will be regarded as three identical independent feeding events;

2. mixed prey items, i.e. $\sum_{k=1}^{4} event_k = 3$ for two or more prey items, according to the proportion of remains in stomach content to split them into one or two independent feeding events, and then each stomach will have three distinct feeding events.

Table 1 summarizes the stomach data.

Table 1 Summary of stomach data of cod

| Region | Number of Choices | | | | | |
	Herring	Capelin	Cod	Other	Mixture	Total
1	0	0	0	1090	0	1090
2	78	551	31	1534	70	2264
3	231	1642	779	11322	1636	15610
4	81	390	104	4948	233	5756
5	11	179	284	4562	572	5608
6	20	1501	384	8899	1394	12198
7	0	581	17	1067	602	2267
Total	421	4844	1599	33422	4507	44793

2.2 Prey abundance

Until now, all the discussions focus on the stomach contents data of cod. To apply a discrete choice model for cod predation, we need data on the attributes of the prey species. We consider prey abundances in the sampling region (see figure 1) and time, respectively, as the main attributes because the predation cost for a cod is assumed to be a function of these prey biomass abundances. First, we assume that the predation activities of cods only occur within one region, and that the characteristics of foraging environment differ between regions. Since the sampling of cods is monthly and distributed across the seven regions, we need to estimate the monthly prey abundances in each region. Therefore, we calculate regional distributions of monthly prey abundances based on stock abundance estimates and estimated migration patterns.

Figure 2.1 Sampling regions
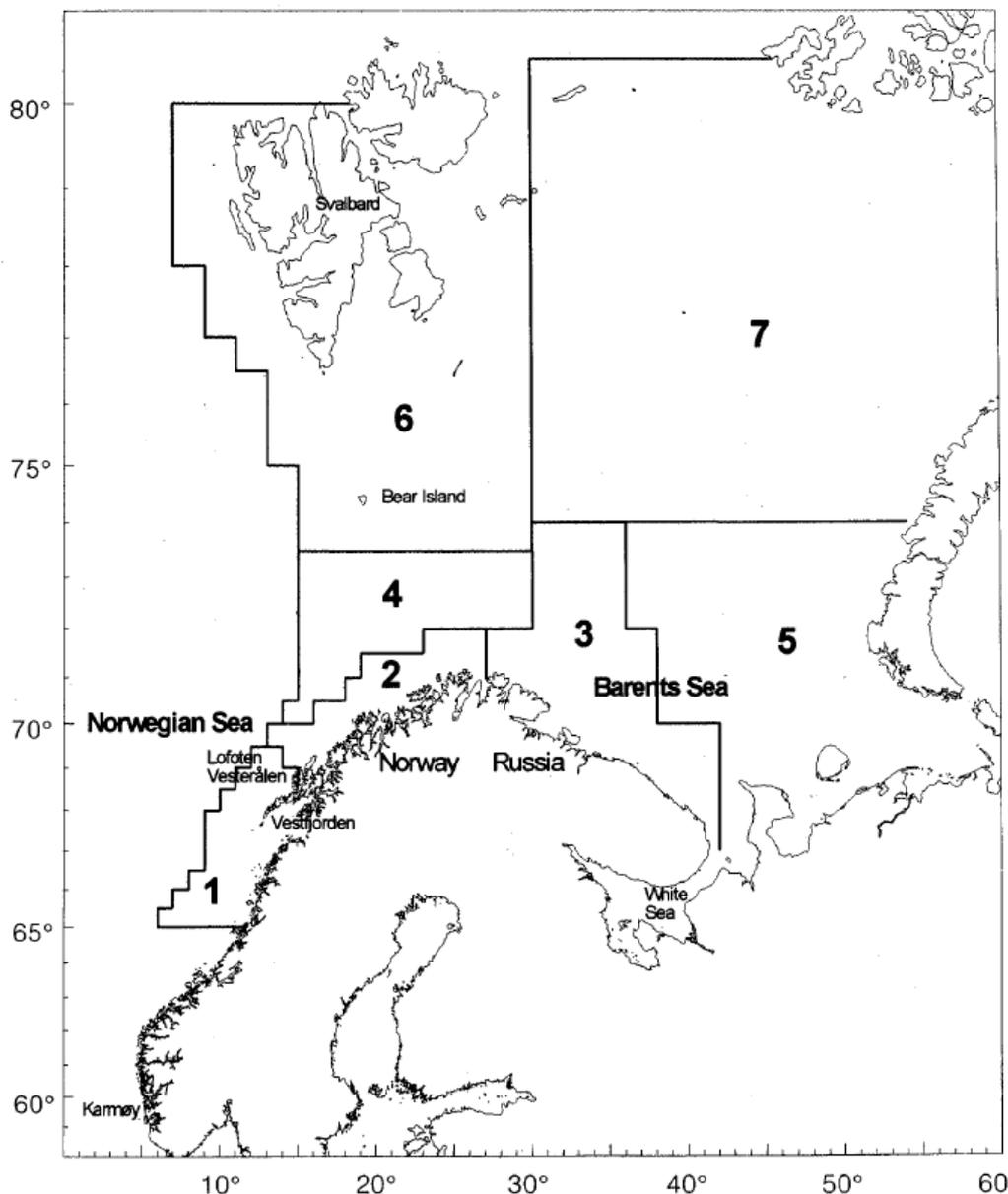
For all prey species, migration between areas is implemented using migration matrices that vary by month, and by age group, but are constant from year to year (Bogstad *et al*, 1997). In addition, we have 'number at age' stock data for all three species of fish, dated to January for the period 1992-2000, and we have used mortality tables to obtain monthly 'number at age' data.

To obtain regional distribution data we multiply the monthly 'number at age' data with the migration pattern matrices,

$$N_t^{as} = D_t^{as} M_t^{as} \quad , \qquad (3.1)$$

where *s* is species index of prey (herring, capelin and cod), *t* is time (month number), *a* is age. N_t^{as} is a 7-element row vector containing the number of prey *s* by region at month *t* and age *a* . D_t^{as} is a row vector of 'number at age' stock data of prey *s* at month *t* and age *a*, and M_t^{as} is the distribution vector obtained from the migration matrices and indicate the proportion of the prey *s* by region at month *t* , and age *a*.

From these monthly "number at age at region" stock abundances for each prey, we multiply the mean weight by age to find biomass distribution B_t^{as} .

$$B_t^{as} = N_t^{as} W_t^{as} , \qquad (3.2)$$

where W_t^{as} is the mean weight of prey *s*, at time *t* , for age *a*.

We do not assume any age-preference pattern in the prey selection process of old cod to herring and capelin. Thus, in order to reduce the dimensionality problem and make computations more convenient, we sum up the age-specific biomass distributions to B_t^{s}, from January to December in 1992-2000, for herring and capelin.

Cannibalism plays an important role in cod predation. From the stomach data, we find that cod length as prey is almost always less than 50cm. According to the length/age ratio the mean length of 3 years old cods is approximately 50 cm, and so we use the biomass density of young cod (younger than 3 years) as available prey abundance instead of the whole cod abundance.

Actually cod predation depends more on availability (density) than on the biomass value itself. The areas of the 7 regions range from 14000 to 90000 square nautical miles. Thus we use the biomass density level to represent the prey abundance instead of absolute biomass. The prey density level unit is kg/square nautical mile.

The "number at age" stock data and mortality data are mainly taken from the ICES (International Council for the Exploration of the Sea) reports, but also partly from the Institute of Marine Research in Bergen(provided by Bjarte Bogstad). The migration pattern and the seven

regions are those used in the MULTSPEC project (Bogstad *et al*, 1997). Since the biomass density levels are not evenly distributed in our sampling region and period, we show three figures to illustrate the distribution of biomass density levels of the three prey species more clearly (figures 2.2, 2.3 and 2.4). In these figures the x-axis represent the biomass density (kg per square nautical mile) and the y-axis represent the proportion of total cell numbers (cell in this context means the combination of region and time).

**Figure 2.2 distribution of herring biomass density**

distribution of herring biomass density

N = 134379   Bandwidth = 364.9
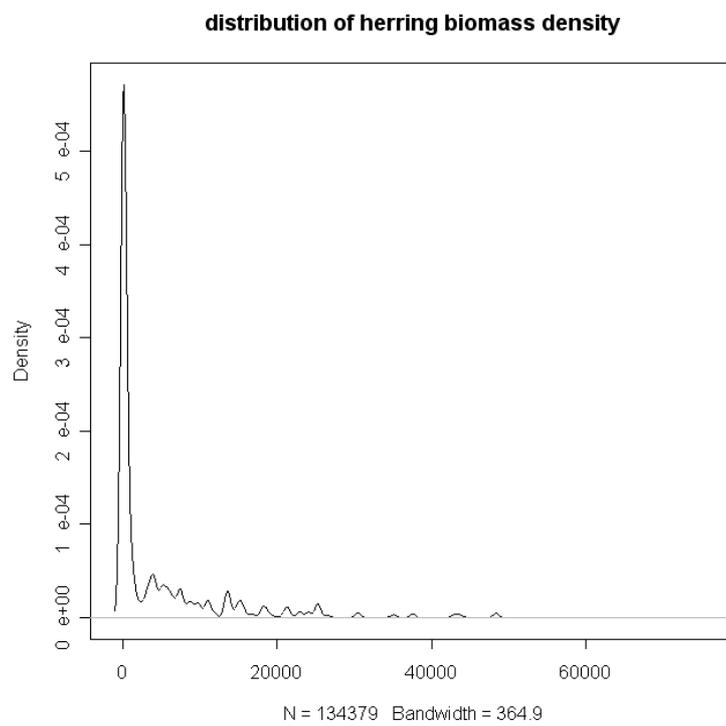
Figure 2.3 distribution of capelin biomass density

distribution of capelin biomass density

N = 134379   Bandwidth = 403.8
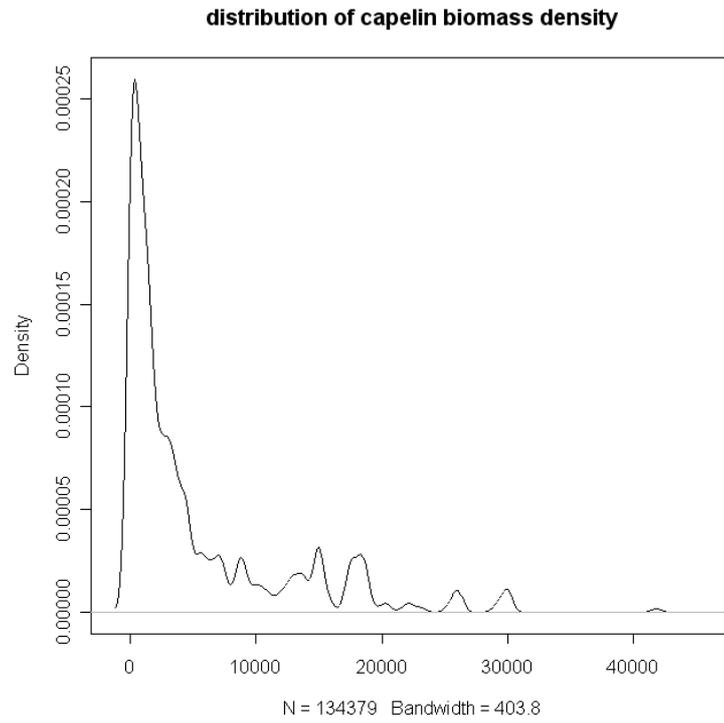
Figure 2.4 distribution of young cod biomass density

distribution of young cod biomass density

N = 134379   Bandwidth = 50.47
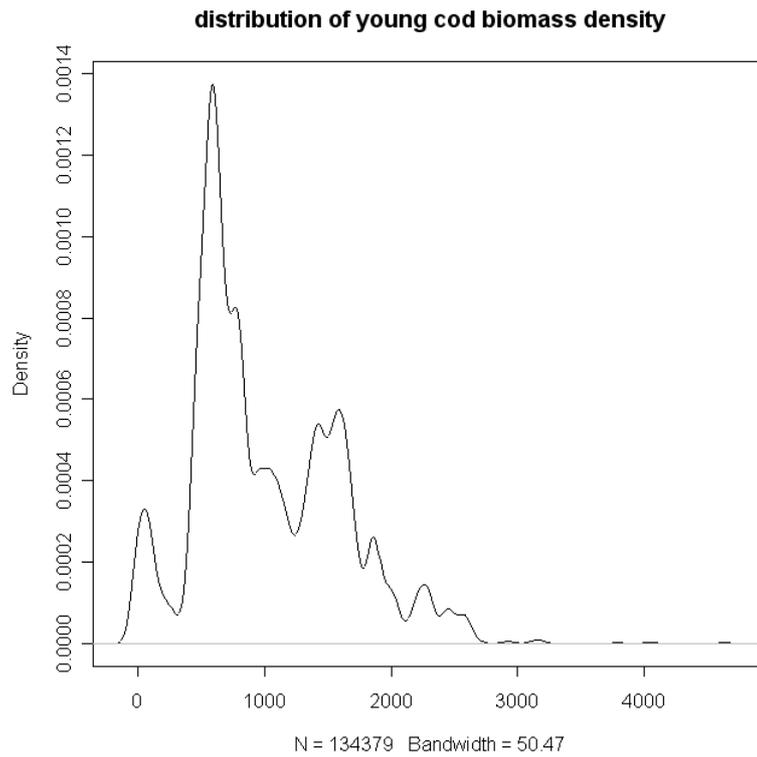
3. Predation model

We assume that only one type of prey is consumed in each feeding event. Further we assume that the energy requirement of the predator is obtained from physiological considerations, e.g. energetics. We want to model the probability $P(Y_i = k \mid x_1,.....x_K)$ of the food being of type k out of K prey types, given the available amounts $x_1,.....x_K$ in the period and region. The prey capelin, young cod and herring are available in highly variable quantities. Since the cost of consuming a unit of herring depends highly on the amount of herring present in the region, the choice probability must depend continuously on the available amount of herring, and in fact continuously on $x_1,.....x_K$. Thus we need models in a continuous way. As the amounts of the particular prey items shrink towards zero, the consumption of this item should also shrink towards zero. Because of this, the choice probability is modeled as a function of the logarithm of the prey abundance. We shall use the following terminology. Let $L(\theta) = (L_1(\theta_1),...,L_K(\theta_K))$ be the multinomial logit model with predictors θ_k which are linear as functions of the components of the covariate vector $\mathbf{x}_k$ and with parameter vector $\boldsymbol{\beta}_k$. The covariate vector $\mathbf{x}_k$ is based on the biomass of three prey species ($B_{herring}, B_{capelin}, B_{cod}$):

$$L_k(\theta) = P(Y_i = k) = \frac{e^{\theta_k}}{\displaystyle\sum_{j=1}^{K} e^{\theta_j}} \quad , \qquad (3.1)$$

$$\theta_k = \boldsymbol{\beta}_k' \mathbf{x}_k \qquad , \qquad (3.2)$$

Though studies have indicated that the choice probabilities are highly related to the biomass density levels, we don't have enough knowledge about the specific model form. There are three candidate models in our study; which differ in the linear predictor of expression (3.2). We will compare three candidates and choose the appropriate one according to criteria of model fit and convenience for further study. The candidates are as follows:

$$\theta_k = \beta_{k0} + \beta_{k1}B_{herring} + \beta_{k2}B_{capelin} + \beta_{k3}B_{cod} + \beta_{k4}\log(B_k) \quad (3.3)$$

$$\theta_k = \beta_{k0} + \beta_{k1}B_k + \beta_{k4}\log(B_k) \qquad (3.4)$$

$$\theta_k = \alpha_1 d_{her} + \alpha_2 d_{cap} + \alpha_3 d_{cod} + \beta_1 B_k + \beta_4 \log(B_k) \quad (3.5)$$

for the three prey species capelin, cod and herring. For the "other food" category we let the linear predictor be zero in all models.

The differences between these three models are found in the components of the covariate vector and parameter constraints. The linear predictor (3.3) depends on the abundance of all three

prey items instead of only the choice prey types that is the case in (3.4). (3.5) simplifies the complexity of model (3.4) further with an assumption that the parameters of the linear predictor are the same for all prey types. The constant term will, however, depend on which prey types that are available. In the model (3.3), the probability of choice k is dependent on other available prey items in two ways: one is directly from the linear predictor θ_k that includes the attributes of other prey types; the other is indirectly from the denominator in (3.1), $\sum_{j=1}^{K} e^{\theta_j}$, that includes attributes of other prey items as well. In (3.4) and (3.5), only the indirect dependence is effective.

We use the method of maximum likelihood to estimate the parameters specifying the various versions of the logistic choice probability model. The iteration method is Newton's method. According to the expression of gradients and Hessians in each model and iteration formula, we use S-plus 6.1 to get the maximum likelihood estimates. The iterations stop when the tolerance of convergence is less than 10^{-7}.

We have assumed three feeding events in every stomach observation. As a matter of fact, this is a rather arbitrary assumption. To avoid the confusion and misleading conclusions derived from this assumption, it is important to bootstrap each analysis to obtain reasonable standard errors and also appropriately down scaled likelihood to compare the three candidate model forms.

Suppose we have estimated parameters $\hat{\beta}$ on the basis of covariate vector $\mathbf{x}$ which is the original sample of stomach observations. We assume that every stomach content includes three feeding events as before, and let $(d,\mathbf{x})$ be the object describing the feeding events and prey availabilities for a randomly chosen stomach. We draw a bootstrap sample, $(d,\mathbf{x})^*$, by drawing 40000 cases of $(d,\mathbf{x})$ with replacement. The maximum likelihood estimate is then obtained from the bootstrapped data $\{(d,\mathbf{x})^*\}$. This procedure is repeated 500 times. According to three feeding events assumption, we split each stomach observation into three feeding events, and each feeding event is assumed independent in the likelihood. One stomach can be split into three, six, or more feeding events on account of replacement sampling.

After 500 times of bootstrap sampling $(d,\mathbf{x})$ of 40000 cases with replacement, we obtain a group of bootstrap estimators $\hat{\beta}_{ML}^*(b)_L$ $b=1,\ldots,500$, $L=40000$. Then we calculate the bootstrap estimate of standard errors for parameters $\hat{\beta}$.

$$SE(\hat{\beta}) = \sqrt{\sum_{b=1}^{500} \frac{(\hat{\beta}_{ML}^*(b) - \hat{\beta}_{ML}^*(\cdot))^2}{499}} \qquad (3.6)$$

10

where $\hat{\boldsymbol{\beta}}_{ML}^{*}(\cdot) = \sum_{b=1}^{500} \hat{\boldsymbol{\beta}}_{ML}^{*}(b)/500$.

I also check whether my estimator is unbiased by considering $\hat{\boldsymbol{\beta}}_{ML}^{*}(b) - \hat{\boldsymbol{\beta}}_{ML}^{*}(\cdot)$ (3.7)

To find the downscaling constant of the likelihood, we assume that

$$2\log L(\hat{\boldsymbol{\beta}}_{ML}) \sim 2\log L(\boldsymbol{\beta}_{true}) + \tau\chi_{df}^{2} \qquad (3.8)$$

where df (degrees of freedom of χ^2 distribution) is the number of free parameters and τ is a scaling parameter.

From the bootstrap runs I calculate $\hat{\boldsymbol{\beta}}_{ML}^{*}$ for each replicate, and obtain the estimation equation

$$\text{var}\left[2\log L(\hat{\boldsymbol{\beta}}_{ML}^{*})\right] = \tau^2 * \text{var}(\chi_{df}^2) = \tau^2 * 2df \qquad (3.9)$$

then

$$\hat{\tau} = sqrt(\frac{2\,\hat{\text{var}}\left[\log L(\hat{\boldsymbol{\beta}}_{ML}^{*})\right]}{df}) \qquad (3.10)$$

then all the likelihoods can be rescaled by

$$\log L(\hat{\boldsymbol{\beta}}_{corrected}) = \frac{1}{\hat{\tau}}\log L(\hat{\boldsymbol{\beta}}_{original}) \qquad (3.11)$$

then, on the basis of rescaled likelihood, we use AIC to select the best model in the given model set which has the smallest AIC index.

AIC (Akaike information criteria):

$$AIC = -2\log L + 2K \qquad (3.12)$$

where K is the number of parameters to be estimated.

4. Result

4.1 Model selection

Until now, we have results for three different models. According to the information criteria, we can compare the AIC of these models to decide which one is the best model in the model set. After bootstrapping, we get rescaled parameters for each model, and after likelihood correction we find that model (3.5) has the smallest AIC. Therefore, the most appropriate model in this study is model (3.5).

Table 4.1 Model selection

	Model(3)	Model(4)	Model(5)
$\log L(\hat{\boldsymbol{\beta}}_{original})$	-86026	-86129	-86281
K	15	9	5
$\hat{\tau}$	164	159	167
$\log L(\hat{\boldsymbol{\beta}}_{corrected})$	524.55	541.69	516.65
AIC	1064.1	1092.4	1038.1

4.2 Estimation Result

Estimation results for model (3.5) are summarized in Table 4.2. Here, our sample consists of 134379 feeding events. Note that the likelihood ratio index used to test the fit of the model is 0.5369, which indicates that the fit of model (3.5) is quite good.

Table 4.2 Estimation result for conditional logit model (Cod)

Variable	Coefficient estimate	Bootstrap standard errors	Bootstrap bias	t statistic	p value
d_{her}	-5.124763	0.136833	0.083241	39.23	0.000
d_{cap}	-2.935873	0.116738	0.064539	26.36	0.000
d_{cod}	-3.511843	0.116310	0.073538	29.16	0.000
B	0.000008	0.000003	-0.000002	3.18	0.000
$\log(B)$	0.328281	0.015402	0.024157	21.33	0.000

Summary statistics:

Number of feeding events:	134379	Rescale parameter: $\hat{\tau} = 167$
Log likelihood at $\beta = 0$		186289/167=1115.50
Log likelihood at convergence		86281/167=516.65
Log likelihood index		1-(516.65/1115.50)=0.5369

From above result, the highly significant coefficient for logarithmic biomass estimate of 0.3282, confirms our basic hypothesis that probability is positively related to the biomass of prey, particularly at low prey biomass level.

5. Discussion

In conclusion, the empirical results are quite consistent with the hypothesis that prey selection probability is highly related to the biomass amount of the specific prey. Particularly when the biomass is at low levels, even a small increase will lead to a large probability change. At high levels, the probability is nearly linear in biomasses. As a whole, the probability change has a logistic pattern. Moreover, this model provides another test for prey preference, since we regard the "other food" as the base category. The three choice specific constants represent special preferences of herring, capelin and younger cod to other food. The estimates show that all of them are significantly different from zero, which means that even when other food can provide the same biomasses as those of herring, capelin and cod, the predator would choose herring, capelin and cod with high probability on account of this prey preference. However, there is a surprising effect in the prey-specific constants, which shows that cod prefers capelin to young cod, then to herring.

References:

Bogstad, B., Hauge, K., & Ulltang, Ø. (1997) : "MULTSPEC – A Multi-species Model for Fish and Marine Mammals in The Barent Sea". Journal of Northwest Athlantic Fishery Science. Vol.22:317-341.

Bogstad, B. (2003): A model for size preferences in cannibalism in Northeast Arctic cod. Manuscripts

NAMMCO (2002) : "Modelling Marine Mammal – Fisheries Interactions in The North Atlantic". Workshop Report. NAMMCO.

Schweder, T., Hagen, G., & Hatlebakk, E. (1998) : "On the Effect on Cod and Herring Fisheries of Returning the Revised Management Procedure for Minke Whaling in the Greater Barents Sea". Fisheries Research. Vol.37 : 77-95.

Zhu, M. (2003): "Minke Whale: A rational consumer in the sea". Master Thesis in Department of Economics, University of Oslo