WILEY

Bayesian Modelling of Catch in a North-West Atlantic Fishery
Author(s): Carmen Fernández, Eduardo Ley and Mark F. J. Steel
Source: *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 51, No. 3 (2002), pp. 257-280
Published by: Wiley for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/3592652
Accessed: 12-12-2016 23:50 UTC

REFERENCES
Linked references are available on JSTOR for this article:
http://www.jstor.org/stable/3592652?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://about.jstor.org/terms

Wiley, Royal Statistical Society are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the Royal Statistical Society. Series C (Applied Statistics)*

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

Appl. Statist. (2002)
51, *Part* 3, *pp.* 257–280

Bayesian modelling of catch in a north-west Atlantic fishery

Carmen Fernández,

University of St Andrews, UK

Eduardo Ley

International Monetary Fund Institute, Washington DC, USA

and Mark F. J. Steel

University of Kent at Canterbury, UK

[Received March 2000. Final revision December 2001]

Summary. We model daily catches of fishing boats in the Grand Bank fishing grounds. We use data on catches per species for a number of vessels collected by the European Union in the context of the Northwest Atlantic Fisheries Organization. Many variables can be thought to influence the amount caught: a number of ship characteristics (such as the size of the ship, the fishing technique used and the mesh size of the nets) are obvious candidates, but one can also consider the season or the actual location of the catch. Our database leads to 28 possible regressors (arising from six continuous variables and four categorical variables, whose 22 levels are treated separately), resulting in a set of 177 million possible linear regression models for the log-catch. Zero observations are modelled separately through a probit model. Inference is based on Bayesian model averaging, using a Markov chain Monte Carlo approach. Particular attention is paid to the prediction of catches for single and aggregated ships.

Keywords: Bayesian model averaging; Categorical variables; Grand Bank fishery; Predictive inference; Probit model

1. Introduction

The mismanagement of the world's fisheries is one of the most important global environmental problems that we face today. Nine of the world's 17 major fisheries are in serious decline, and four others are classified as 'commercially depleted' by the Food and Agricultural Organization of the United Nations (Tibbets, 1994).

The Northwest Atlantic Fisheries Organization is one of several international organizations that tries to alleviate overexploitation through voluntary co-operation. It was established in 1978 to contribute to the optimal exploitation and rational use of fisheries resources in the Grand Bank outside Canada's exclusive economic zone (see `http://www.nafo.ca` for a map of the area covered by the treaty). Countries which are members of the Northwest Atlantic Fisheries Organization assign quotas among themselves and grant inspection rights to each other. Three inspection ships—two Canadian and one belonging to the European Union—

Address for correspondence: Mark F. J. Steel, Institute of Mathematics and Statistics, Cornwallis Building, University of Kent at Canterbury, Canterbury, CT2 7NF, UK.
E-mail: M.F.Steel@ukc.ac.uk

© 2002 Royal Statistical Society 0035–9254/02/51257

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

board vessels of member states and register the information in their log-books. In addition, ships from signatory countries report (through the so-called 'hails') their entry to and exit of the various zones of the fishing grounds. Finally, two daily flights over the Grand Bank and the Flemish Cap are made by inspection airplanes with the purpose of locating and identifying all ships fishing in the area. Boarding ships on high seas to verify catches is expensive and disrupts their operations. Furthermore, ships from non-signatory countries cannot be inspected. It then becomes important to construct models that allow for the prediction and monitoring of catches conditional on the information from aerial sightings and hails, ship characteristics and other variables (such as the month of the year). Thus, our aim is to model how all these variables influence the catch. This could provide useful information for regulatory measures and guidelines related to issues like mesh size and optimal size of the fleet. More importantly, it allows an estimate of the total amount caught by a group of ships operating in a certain area at a certain time of the year.

The data that we have consist of the daily catch (per ship) per species of fish. Since there are many days with zero catch for a given species of fish, our statistical model incorporates a positive probability of zero catch through a probit model. When a catch occurs, the logarithm of the quantity caught is modelled through a linear regression structure, where we formally treat the uncertainty concerning the choice of regressors through model averaging in a Bayesian setting using posterior model probabilities as weights. In view of the large number of potential models, we explore the posterior distribution by using Markov chain Monte Carlo sampling over the model space in the spirit of the 'MC^3' methodology of Madigan and York (1995). The Bayesian framework leads to exact small sample results, fully taking both parameter and model uncertainty into account. In the present application we have not used any strong prior information or a formal decision theory framework. Both of these can, however, easily be incorporated into a Bayesian analysis.

The aims of this paper are quite different from those of the large literature in stock assessment, where statistical methods are used to assess the size of fish stocks; see, for example, Hilborn and Walters (1992) for a general introduction and McAllister and Kirkwood (1998) for an overview of Bayesian stock assessment methods. A variety of statistical methods, such as Bayesian state space models (Millar and Meyer, 2000) and spatial methods (Newman, 1998), has recently been introduced into this literature. There is, in addition, substantial work on the estimation of year effects and abundance trends based on modelling the catch per hour fished; Quinn and Deriso (1999) have provided many examples. In contrast with these, and like Ferreira and Tusell (1996), our aim is to shed light on how the catch can be explained by certain observable characteristics—such as mesh size (Robichaud *et al.*, 1999)—and to provide operational forecasts of commercial landings of various species (Stergiou *et al.*, 1997). It is important to stress that the main aim of our analysis is not necessarily to develop a model that describes the dynamics of fisheries as closely as possible, but rather to provide a framework that can successfully be used for short-term predictions of quantities caught (of a certain species by a certain ship or group of ships) given an easily available information set. This will guide the modelling strategy and the choice of covariates that we shall consider.

Section 2 describes the data, whereas Section 3 introduces the statistical model. The zero observations are treated in Section 4, and the analysis of positive catches is discussed in Section 5. Section 6 focuses on prediction. The empirical results are presented in Section 7 and a final section concludes. Details of the computational implementation are presented in Appendix A. The data and program that was used to analyse them can be obtained from

`http://www.blackwellpublishers.co.uk/rss/`

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

2. The data

The original data were gathered by the inspection vessel of the European Union operating on the Grand Bank fishery. Inspectors board the fishing boats and record basic characteristics of the ship and the fishing equipment, as well as the quantities caught of different species and where and when this catch was effectuated. They use the ship's log-books to collect all the information accumulated since the last time that the ship was boarded. All data correspond to 1993 and the first half of 1994, leading to 6806 observations each corresponding to a particular ship at a given day. In all, there are 59 different ships.

The dependent variable is the live-weight of fish caught. Table 1 summarizes the regressors that we consider using. These include four categorical variables: the year when the catch is made (two levels), fishing technique (four levels), zone or division within the fishing grounds (four levels) and month of the year (12 levels). In addition, we have four continuous variables (Table 2), namely mesh size measured in millimetres, length of vessel measured in metres, gross registered tonnage (GRT) and engine power in kilowatts. See for example King (1995), chapter 2, for a description of fishing gear and methods.

Our data set also provides the nationality of the ship but we have decided not to consider this variable since one of the purposes of the analysis is to predict the catch of ships from non-signatory countries (for which we have no observations). However, we do have a year effect. This is because year class effects are important in fisheries and, from a biological point of view,

Table 1. Data statistics

Regressor	*% observations*
1, year 1993	75.36
2, year 1994	24.64
3, drift gill net	3.60
4, anchored gill net	1.44
5, otter trawl	79.64
6, otter trawl pair	15.32
7, zone 3L	34.64
8, zone 3M	25.69
9, zone 3N	35.05
10, zone 3O	4.62
11, January	4.89
12, February	10.74
13, March	15.05
14, April	12.06
15, May	13.99
16, June	9.48
17, July	7.02
18, August	7.71
19, September	7.98
20, October	7.04
21, November	3.48
22, December	0.56
23, gill net $\times$ $\log\{0.5 + \text{mesh size} - \min(\text{mesh size})\}$	
24, gill net $\times$ $\log\{0.5 + \text{mesh size} - \min(\text{mesh size})\}^2$	
25, trawl $\times$ $\log\{0.5 + \text{mesh size} - \min(\text{mesh size})\}$	
26, trawl $\times$ $\log\{0.5 + \text{engine power} - \min(\text{engine power})\}$	
27, log(length of vessel)	
28, log(GRT)	

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

Table 2. Values of the continuous variables

Variable	*Minimum*	*1st quartile*	*Median*	*3rd quartile*	*Maximum*
Mesh size (mm), gill netters	110	130	140	140	150
Mesh size (mm), trawlers	120	120	120	130	150
Engine power (kW), trawlers	588	845	1164	1470	2648
Length of vessel (m)	29.0	42.0	47.0	61.2	84.9
GRT	252.3	376.9	664.9	970.2	2382.0

it would not be sensible to assume equality of catches in, for example, May 1993 and May 1994. Inevitably, this complicates predictions for years for which no data are available.

The way that mesh size and engine power influence the catch is potentially very different for gill nets and otter trawls. Thus, we include these variables in terms of interactions with an indicator variable for the net type used: gill nets (adding drift and anchored) and trawls (both single and paired). In addition, there is prior reason to assume that the effect of mesh size might be non-linear for gill nets (for example, the catch would decrease if the mesh size were either too large or too small), so we include a quadratic interaction term for this fishing technique. An effect of engine power on the quantity caught is quite plausible for otter trawls (which are towed) but is very unlikely for gill nets (which are passive), so we do not include an interaction term for engine power and gill nets. To reduce the collinearity between these interaction terms and fishing techniques, the continuous variables power and mesh size are transformed as indicated in Table 1. This substantially increases the spread of the interaction variables and reduces the collinearity in the design matrix. The other continuous variables (length of the vessel and tonnage) are transformed to logarithms in the usual way.

Table 1 indicates the empirical distribution of each of the categorical variables, and Table 2 indicates quantiles of the continuous variables (before transformation). Of course, Tables 1 and 2 provide only marginal information. Some complementary information is given in Fig. 1, where we present bivariate histograms (with lighter shades corresponding to higher relative frequencies) of some combinations of regressors for each of the years in the sample. Levels for the categorical variables are ordered as in Table 1 and continuous variables are categorized into five bins of equal width. From this we note a shift in 1993 from zone L in the period January–May to zone N for the remaining months of the year. The available months of 1994 show a somewhat more even spread over zones L, M and N. The month *versus* mesh size plots are presented for gill nets and trawls separately, which show that trawlers tend to use smaller meshes than gill netters. We also see a tendency towards the greater use of small mesh trawl nets (often 120 mm) in the months April–June 1993. Finally, the length and GRT of the ships are obviously positively correlated as can be seen from the last row of plots.

Table 3 lists the five most important species caught in the Grand Bank and has one category for all the other species ('rest'). Every time that we observe a ship, we observe its daily catches for all six species. A look at the data tells us that a ship's catch on any given day often does not include all species. In particular, we shall model the first five species listed in Table 3, for which the percentage of zeros in the data (6806 ship-days) ranges from 18.50% (halibut) to 88.33% (cod). Thus, this is an important aspect, which, if overlooked, would lead to a substantial overestimation of the catch. Hence, we shall model zero catches explicitly, by means of a probit

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

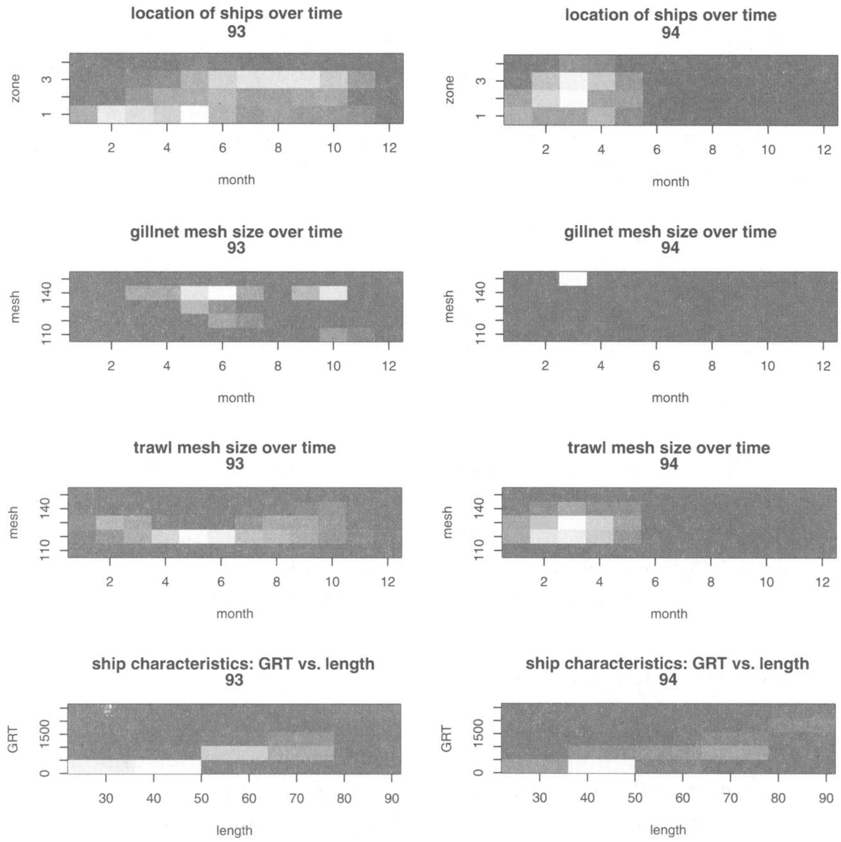

Fig. 1. Bivariate grey scale plots for regressors (lighter shades correspond to higher relative frequencies)

Table 3. Catch for various fish species

Species	*Description*	*Mean (kg)*	*Standard deviation (kg)*	*% zeros*	*% of catch*
1	Atlantic cod (*Gadus morhua*)	550.20	2517.48	88.33	9.80
2	Greenland halibut (*Reinhardtius hippoglossoides*)	3503.48	3610.55	18.50	62.40
3	Redfish (*Sebastes sp.*)	658.51	2794.49	85.73	11.73
4	Roundnose grenadier (*Coryphaenoides rupestris*)	213.46	502.18	43.20	3.80
5	Skate (*Raja sp.*)	503.39	1661.50	55.44	8.97
6	Rest	185.23	625.76	72.27	3.30

model. This feature of the data was not accounted for by Ferreira and Tusell (1996), who analysed the same data set but took only the positive observations into account. Table 3 also lists the fraction of the total live-weight that each species constitutes. We shall consider separate models for each of the species, to allow for the explanatory variables to affect the catch for each species differently.

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

3. The statistical model

In this section we outline our statistical model for the daily catch of a given species of fish per ship. The observations will be denoted by s_i, $i = 1, \ldots, n$ $(n = 6806)$, and we define $s = (s_1, \ldots, s_n)'$. Clearly, each of the n observations is non-negative, and a certain number of them, say Q, are strictly positive (those that correspond to a positive catch). For notational convenience, we shall order the observations so that the first Q observations are positive, whereas the remaining $n - Q$ observations are equal to 0.

As explained in the previous section, it is crucial to take account of the fact that there is a positive probability of zero catch. A natural approach is to use the probit model

$$\begin{aligned} s_i &= 0 \qquad \text{with probability } \Phi(x_i'\gamma), \\ s_i &> 0 \qquad \text{with probability } 1 - \Phi(x_i'\gamma), \end{aligned} \tag{3.1}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution, the vector $x_i \in \Re^{1+k}$ contains the element 1 as well as the explanatory variables presented in Table 1, and $\gamma \in \Re^{1+k}$ groups the parameters. Each categorical variable is handled through dummy variables taking the values 0 or 1, with one level excluded (thus, k is equal to 24 instead of 28). Throughout the paper, the design matrix $X \equiv (x_1, \ldots, x_n)'$ will be of full column rank.

If $s_i > 0$, we further assume a linear regression structure for $y_i \equiv \log(s_i)$. This is easier to handle than the probit model, so a more ambitious strategy is feasible. In particular, we shall allow for model uncertainty, where each of the potential models considered will assume that

$$y_i = \log(s_i) \text{ is distributed as normal}(\alpha + z_i'\beta, \sigma^2), \qquad i = 1, \ldots, Q, \tag{3.2}$$

and the vector z_i corresponds to a subset of the regressors in Table 1. For computational convenience, all the variables are now demeaned, so that each column of the resulting design matrix $Z \equiv (z_1, \ldots, z_Q)'$ sums to 0. The matrix Z is also of full column rank. In the generic model (3.2), $\alpha \in \Re$ is the intercept and $\sigma^2 > 0$ denotes the sampling variance, whereas the vector β groups the regression coefficients.

Note that models (3.1) and (3.2) have been defined entirely separately, using different parameters, and we shall also assume prior independence between the parameters in models (3.1) and (3.2). This is done partly for pragmatic reasons (as in this case we can conduct posterior inference independently, greatly simplifying the computations), but also because it is not obvious to us that the effects of a given variable on the probability of zero catch and on the actual amount caught (when the catch is positive) should be linked. We might possibly consider sign restrictions for the elements of γ in model (3.1) and β in the generic model (3.2). For example, for otter trawls, increasing mesh size could be expected to decrease the amount caught and to increase the probability of zero catch: this would imply a negative component in β and a positive component in γ, but it would not mean that their actual magnitudes are necessarily linked. Thus, such restrictions would not imply that both models should be analysed jointly. Here we have chosen not to impose prior constraints like these, and we shall instead let the data find the most appropriate parameter ranges. As we do not have strong prior information, our prior distribution (presented in the following two sections) will generally try to incorporate as little subjective input as possible.

We shall use the entire sample to make inference on γ and to predict the probability of zero *versus* positive catch; this analysis only uses the fact whether s_i is 0 or strictly positive. The actual value of the Q positive observations will be used to conduct inference on α, β and σ, and to predict the amount of catch given that it is positive. The probit model will be examined in Section 4, whereas Section 5 will be devoted to the model for positive catch. Lo *et al.* (1992) also

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

modelled zero observations separately from positive observations in the context of analysing relative fish abundance, using classical statistical procedures and a simple linear probability model for zero observations.

4. Analysis of zero observations

In this section, we focus on posterior inference on γ, the parameter in the probit model (3.1). We shall complement this sampling distribution with the prior

$$p(\gamma) = f_N^{1+k}\{\gamma|0, (h_0 X'X)^{-1}\}, \tag{4.1}$$

i.e. a $(1+k)$-variate normal distribution with zero mean and covariance matrix $(h_0 X'X)^{-1}$, where $h_0 > 0$. This corresponds to the g-prior introduced in Zellner (1986) and essentially says that the prior precision is a fraction h_0 of that of the sample. This prior is often used for relatively high dimensional parameters in the context of a lack of strong prior information, as it typically does not distort the information in the sample. We took 0 as the prior mean for γ, since, from model (3.1), $P(s_i = 0|\gamma = 0) = \frac{1}{2}$. For h_0 we adopt the value $h_0 = 1/n$, which roughly corresponds to the information in one observation and will easily be dominated in posterior and predictive inference. With the prior in equation (4.1) the predictive distribution is invariant with respect to the choice of the reference levels for the categorical variables, as is desirable.

Although the posterior distribution corresponding to a sample of n observations from model (3.1) and the prior in equation (4.1) cannot be computed analytically, we can use Gibbs sampling (with data augmentation) to approximate $p(\gamma|s)$ to any required precision (see Appendix A.1 for details, and Albert and Chib (1993) for a similar probit analysis).

5. Analysis for positive observations

5.1. Model specification

We shall incorporate model uncertainty in the sense that we allow for any subset of the variables in Table 1 to appear as regressors in the generic model (3.2). This means that instead of a single model we have a set $\mathcal{M} = \{M_j: j = 1, \ldots, J\}$, where each model corresponds to a particular choice of regressors.

To understand our model space $\mathcal{M}$ fully, we need to explain carefully how we deal with categorical variables in this situation. We shall treat different levels of a categorical variable separately, so that a model in $\mathcal{M}$ can include or exclude any level with the only restriction that not all levels of a categorical variable can be included in the same model. This gives us extra flexibility with respect to the simpler approach that treats categorical variables as single entities which can only be fully excluded (which means that all levels have exactly the same effect) or included (which implies that all levels have different effects). With our treatment, we also allow for intermediate situations where several levels of a variable have the same effect (and are, therefore, excluded from the model) whereas other levels (those included in the model) have different effects. This is an issue of empirical relevance as the results in Section 7 will illustrate. Our approach implies that we cannot fix a reference level, as we want to treat all levels in a symmetric fashion. As an example, consider the categorical variable month of the year, which has 12 levels. If we were to designate, say, December as a reference level we would be able to capture a situation where, for example, January has the same effect as December (by also excluding January), but not a situation where January has the same effect as, say, February, yet not the same as December. By allowing a free reference level, we can accommodate any combination of levels having the same effect.

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

With K continuous variables and R categorical variables with $L_1, L_2, \ldots, L_R$ levels, this strategy implies a model space $\mathcal{M}$ with

$$J = 2^K \prod_{r=1}^{R} (2^{L_r} - 1)$$

elements, which for our application leaves us with 176904000 possible models. We stress that, whereas all the $\tilde{k} = 28$ variables in Table 1 will appear in some of the models, the maximum number of regressors that any single model can contain is $k = 28 - 4 = 24$ (since there are four categorical variables). Whenever a model contains all except one levels of a categorical variable, we say that the model is 'full' in that categorical variable. Note that models that are full in one or several categorical variables appear with different representations in $\mathcal{M}$, each corresponding to a particular choice of reference level. This feature will be taken into account when setting a prior distribution for the models.

5.2. Priors under different models

We now turn to the issue of eliciting priors for the parameters in expression (3.2) given a particular model M_j. For these parameters we specify a prior distribution that incorporates minimal prior information while leading to analytical tractability. On the intercept α and the scale parameter σ, which are present in all the models, we assume the usual non-informative distributions, respectively defined through

$$\begin{aligned} p(\alpha) &\propto 1, \\ p(\sigma) &\propto \sigma^{-1}. \end{aligned} \tag{5.1}$$

For the vector $\beta_{(j)}$, which groups the relevant regression coefficients under model M_j, we assume the g-type prior

$$p(\beta_{(j)}|\sigma, M_j) = f_N^{k_j}\{\beta_{(j)} \,|\, 0, \sigma^2(g_0 Z_j' Z_j)^{-1}\}, \tag{5.2}$$

where k_j is the number of explanatory variables included in M_j and Z_j denotes the corresponding design matrix. This prior specification requires minimal judgmental input from the user, since only the scalar g_0 is left to be chosen. We shall take $g_0 = 1/\max(Q, \tilde{k}^2)$, where Q is the number of positive observations and $\tilde{k}$ is the number of available regressors in Table 1. This choice is inspired by Fernández *et al.* (2001a), who found that the use of such a strategy for g_0 leads to very satisfactory identification of the correct model in simulation exercises, whereas the out-of-sample predictive behaviour is also quite good. Besides their empirical simulation justification, they also derived some theoretical properties of this prior. Finally, model M_j assumes that its excluded explanatory variables do not matter, i.e. that their associated regression coefficients are equal to 0. Now that we have specified the prior distribution, we can immediately conduct Bayesian inference under model M_j, by combining this distribution with the corresponding sampling model from expression (3.2). Since this prior distribution resembles a natural conjugate prior, computing the posterior and predictive distributions is quite simple, as will be explained later in the paper.

5.3. Model averaging

So far we have considered a single model M_j from the space of all possible models $\mathcal{M}$. From a Bayesian perspective, model uncertainty can be treated in a coherent fashion by further specifying a prior distribution $P(M_j)$ on the models. Here we shall consider a uniform distribution on the space of genuinely different models. By this we mean that we take into account that $\mathcal{M}$

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

contains multiple copies of models which are full in some categorical variable, downweighting their prior probabilities accordingly. If desired, other prior distributions could be considered with only minor modifications to our framework.

The posterior distribution of a quantity is now given by a mixture of the posterior distributions under each of the models, with mixing probabilities corresponding to the posterior model probabilities. Thus, Bayesian inference provides a coherent framework for treating model uncertainty, leading to an inferential procedure which averages over the inferences resulting from each of the individual models. Madigan and Raftery (1994), Raftery *et al.* (1997) and Fernández *et al.* (2001b) found in a series of empirical applications that, in the presence of model uncertainty, Bayesian model averaging leads to the best predictive performance, as measured by a logarithmic scoring rule. In a decision theory context, mixing over models can be shown to be optimal under predictive squared error loss, provided that the set of models considered is exhaustive (Min and Zellner, 1993). We follow this approach and consider model averaging rather than selecting one single model.

Applying the Bayes theorem, the posterior probability of model M_j is given by

$$P(M_j|y) \propto l_y(M_j)\ P(M_j), \tag{5.3}$$

where $P(M_j)$ is the prior probability and $l_y(M_j)$ the marginal likelihood of model M_j. The latter is obtained from expression (3.2), integrating out the parameters with their prior distribution described in Section 5.2. It is easy to show that $l_y(M_j)$ is finite if and only if the sample $y = (y_1, \ldots, y_Q)'$ contains at least two different observations. This condition will be both necessary and sufficient for posterior and predictive inference throughout the paper.

Although we can derive an explicit expression for $l_y(M_j)$ (see expression (A.1) in Appendix A.2), direct computation of the posterior probability in expression (5.3) is very difficult owing to the large number of models in $\mathcal{M}$ (approximately 177 million in our application). Therefore, we shall approximate the posterior distribution of the models via simulation, using a Markov chain Monte Carlo sampler on the model space $\mathcal{M}$. Appendix A.2 provides more details on the particular sampler that we have adopted, which is of the Metropolis–Hastings type. In case we have no categorical variables, the sampler essentially simplifies to the MC^3 method of Madigan and York (1995) that was also used in Raftery *et al.* (1997).

5.4. Inference on regression coefficients

We now consider inference on a linear combination

$$b'\beta \equiv \sum_{l=1}^{\tilde{k}} b_l \beta_l$$

of the elements of the $\tilde{k}$-dimensional regression vector β, where $\tilde{k} = 28$, corresponding to all the variables in Table 1. To do this, we need to apply the model averaging ideas explained in the previous subsection. Under model M_j, $b'\beta$ takes the value 0 if none of the regressors corresponding to a non-zero element of b is included in M_j and has a Student t-distribution otherwise. The exact form of the posterior distribution of $b'\beta$ is

(a) with probability $p \equiv \Sigma_{j:B_j b=\mathbf{0}} P(M_j|y)$,

$$b'\beta = 0, \tag{5.4}$$

and

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

(b) with probability $1-p$, $b'\beta$ has density

$$\frac{1}{1-p} \sum_{j:B_j b \neq \mathbf{0}} f_S \left\{ b'\beta \middle| Q-1, \frac{b'B_j'(Z_j'Z_j)^{-1}Z_j'y}{g_0+1}, \frac{Q-1}{G_j} \frac{g_0+1}{b'B_j'(Z_j'Z_j)^{-1}B_j b} \right\} P(M_j|y), \tag{5.5}$$

where B_j is the relevant selection matrix under model M_j in the sense that $\beta_{(j)} = B_j\beta$, with $\beta_{(j)}$ corresponding to the regressors included in M_j, $\mathbf{0}$ is a vector of 0s of the appropriate dimension and $f_S(x|\nu, m, a)$ denotes the probability density function (PDF) of a Student t-distribution with ν degrees of freedom, location m (the mean if $\nu > 1$) and precision a (with variance $\nu/\{(\nu-2)a\}$ provided that $\nu > 2$). Finally, G_j is defined in expression (A.2) in Appendix A.2. From expressions (5.4) and (5.5) it is clear that, once we have run the Markov chain on $\mathcal{M}$ to compute $P(M_j|y)$, we can obtain the distribution of $b'\beta$ analytically.

6. Prediction

We now focus on forecasting the value of a new observable, say s_f, given a vector of explanatory variables and the observed sample s. Our forecast for s_f will be based on the out-of-sample predictive distribution, which is obtained from expression (3.1)–(3.2) after integrating out all the parameters and all possible models using their respective posterior distributions. From expression (3.1)–(3.2) it is immediate that the predictive distribution for s_f will be a mixture of a point mass at zero and a continuous distribution. In particular, we have the following.

(a)

$$s_f = 0 \tag{6.1}$$

with probability

$$\omega_f \equiv \int \Phi(x_f'\gamma)\, p(\gamma|s)\, \mathrm{d}\gamma, \tag{6.2}$$

where $x_f \in \Re^{1+k}$ contains the element 1 and the explanatory variables for s_f. The integral in expression (6.2) can be calculated by averaging $\Phi(x_f'\gamma)$ over the draws of γ generated through the Gibbs sampler in Appendix A.1.

(b) With probability

$$1-\omega_f, \tag{6.3}$$

$s_f > 0$ and it has PDF

$$p(s_f|y) = \frac{1}{s_f} \sum_{j=1}^{J} f_S \left[\log(s_f) \middle| Q-1, \bar{y} + z_{f(j)}' \frac{(Z_j'Z_j)^{-1}}{g_0+1} Z_j'y, \right.$$
$$\left. \frac{Q-1}{G_j} \left\{ \frac{Q+1}{Q} + z_{f(j)}' \frac{(Z_j'Z_j)^{-1}}{g_0+1} z_{f(j)} \right\}^{-1} \right] P(M_j|y), \tag{6.4}$$

where $z_{f(j)}$ is the k_j-dimensional vector that contains the explanatory variables (demeaned as indicated after expression (3.2)) that are relevant under model M_j.

In a practical context, we may be interested in predicting the aggregate catch of a group of ships during a certain spell of time. This means that we focus on the predictive distribution of

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

$$s_{\text{sum}} \equiv \sum_{i=1}^{I} s_{f_i}$$

rather than considering one single observable s_f as was the case above. The predictive distribution of s_{sum} is computed by averaging its sampling distribution over parameters and models by using the relevant posterior distributions. It is clear from model (3.1) that in the sampling s_{sum} is 0 with probability

$$\omega(\gamma) \equiv \prod_{i=1}^{I} \Phi(x'_{f_i}\gamma)$$

(where $x_{f_i} \in \Re^{1+k}$ corresponds to the explanatory variables for s_{f_i}) and has some PDF with probability $1 - \omega(\gamma)$. This means that we forecast

(a) $s_{\text{sum}} = 0$, with probability $\omega_{\text{sum}} \equiv \int \omega(\gamma)\, p(\gamma|s)\, \mathrm{d}\gamma$, which, as before, we compute by averaging $\omega(\gamma)$ over the Gibbs draws of γ and
(b) with probability $1 - \omega_{\text{sum}}$, $s_{\text{sum}} > 0$ and has a predictive distribution given through some PDF on $(0, \infty)$. Although an explicit expression for this PDF is complicated to derive, we can approximate this distribution via simulation, drawing a set of values from model (3.1)–(3.2) where the parameters are, in turn, drawn from the posterior distribution (taking model averaging into account).

7. Discussion of results

7.1. Computational issues and model probabilities

Most of the discussion in this subsection will focus on the Markov chain on model space, since it is the most computationally demanding aspect of our model. In the interest of the practical importance of this methodology, and to enhance its appeal to applied researchers, we have made particular efforts to create an efficient set of programs that can deal with problems of empirical relevance. The programs are coded in Fortran 77 and make efficient use of central processor unit time, e.g. through storing results for already visited models in stacks (saving recalculations when a model is revisited by the chain). As a consequence, for example, the entire single-ship analysis presented in what follows takes between 1 and 3 h (depending on the species) on a 200 MHz PowerPC-based desktop computer.

Throughout, we shall split the available observations into a subsample used for posterior inference (the 'estimation subsample') and the remaining observations, which will be used for comparison with the predictive distribution (the 'prediction subsample'). Observations are randomly assigned to the estimation subsample with probability 0.75 and the resulting number of observations in this subsample is $n = 5087$ with Q in Table 4 indicating the number of positive observations in this subsample. The total number of regressors is $\tilde{k} = 28$ (all those in Table 1) for halibut, redfish and grenadier. For cod, $\tilde{k} = 26$ because there are no catches in November or December. For skate, $\tilde{k} = 27$ as we leave out the quadratic interaction term between mesh size and gill nets to avoid problems of collinearity. Thus, we obtain $Q > \tilde{k}^2$ for halibut, grenadier and skate, which leads to choosing $g_0 = 1/Q$ in prior (5.2), whereas for cod and redfish we choose $g_0 = 1/\tilde{k}^2$.

The Markov chain that was used for computing posterior model probabilities is described in Appendix A.2. Table 4 lists the number of drawings retained and the initial number of discarded draws (the 'burn-in'), as well as the total number of models visited. We consider several strategies for assessing the convergence of this chain. Since the marginal likelihood for model M_j, $l_y(M_j)$, can be calculated explicitly, we shall apply formula (5.3) to compute posterior probabilities on

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

Table 4. Monte Carlo performance and posterior probabilities

	Results for the following species:				
	Cod	*Halibut*	*Redfish*	*Grenadier*	*Skate*
Number of observations Q	583	4161	727	2891	2256
Number of retained drawings	1000000	500000	1000000	500000	2000000
Number of discarded drawings	500000	100000	500000	100000	1000000
Number of models visited	32739	1906	18264	2840	5202
Number of non-equivalent models visited	24229	485	15940	1766	3266
Window estimate and empirical frequency correlation coefficient	0.9890	0.9782	0.9919	0.9909	0.9659
Weighted average q	0.1602	0.0554	0.2834	0.0776	0.0392
Posterior probability covered by chain	0.8811	0.9438	0.9530	0.9695	0.9988
Posterior probability of best model	0.0335	0.0510	0.0791	0.1019	0.0883
Number of models required for 90% posterior probability	4022	144	1494	156	235
Posterior probability of stepwise model	7.4×10^{-13}	7.3×10^{-5}	4.2×10^{-5}	0.0640	0.0010

the basis of the models visited by the Markov chain (instead of using the empirical frequencies of visiting each model). This idea, called 'window estimation' by Clyde *et al.* (1996), implies that the computed posterior odds (ratios of posterior probabilities) between any two models visited are the actual posterior odds. From Table 4 we see that the correlation coefficient between the posterior probabilities of all models visited computed on the basis of empirical frequencies and window estimation is always above 0.96. This provides an indication of convergence of the chain.

A second diagnostic of convergence is based on the fact that models that are full in one or more categorical variables have exactly equivalent counterparts in the model space (that only differ in the chosen levels of the categorical variables for which they are full). Asymptotically, such equivalent models are visited equally often, which suggests looking at

$$q \equiv \left\{ \max_i(\text{freq}_i) - \min_i(\text{freq}_i) \right\} \Big/ \sum_i \text{freq}_i,$$

where freq_i is the number of times that the chain visits representation i of the same model. Clearly, $q \in [0, 1]$ with $q = 0$ the best result and $q = 1$ the worst, indicating that only one of the equivalent representations was visited. Table 4 reports a weighted average of the q-values, with weights proportional to the posterior probabilities of each model representation. The reported q-values are all reasonably small. Table 4 also lists the number of truly different models that were visited.

A third measure of convergence is provided by an estimate of the total posterior model probability covered by the chain following George and McCulloch (1997). This estimate is based on comparing visit frequencies and the aggregate marginal likelihood for a predetermined subset of models. Table 4 presents this estimate for the various species, which is never below 88% (and typically well above 90%).

All diagnostics indicate that convergence is never a problem, which was corroborated by the fact that other independent runs started from randomly chosen models led to virtually identical results.

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

The chains visit a relatively small number of models: except for cod always less (and usually much less) than one model in every 9600. Throughout, the acceptance probability of proposals in the Markov chain Monte Carlo algorithm is between 6% and 18%. The best model (the model with the highest posterior mass) contains between seven (cod) and 18 (halibut) regressors and often receives quite a substantial posterior probability, but never so large that model averaging becomes unnecessary. The number of highest probability models that is needed to cover 90% of the total visited posterior mass (also presented in Table 4) gives a further indication of the substantial spread of the posterior mass in model space.

Marginal posterior inclusion probabilities of the various regressors ($1-p$ with p obtained from expression (5.4)) are given in Table 5. Clearly, the models visited for halibut are always full in the variables year and zone (which means that each of the two years has probability $1-\frac{1}{2}=0.5$ of inclusion and each of the four zones has probability $1-\frac{1}{4}=0.75$ of inclusion). Note the large differences in the posterior probabilities of inclusion across the various species, which supports our decision to model each species separately.

Convergence of the Gibbs sampler for the probit model was assessed by monitoring the posterior moments of γ in different runs of various lengths. Retaining 20000 draws after a burn-in of 5000 was found to be more than sufficient.

Table 5. Marginal posterior inclusion probabilities of regressors†

Regressor	*Probabilities for the following species:*				
	Cod	*Halibut*	*Redfish*	*Grenadier*	*Skate*
Year 1993	0.49	0.50	0.03	0.49	0.41
Year 1994	0.49	0.50	0.03	0.49	0.41
Drift gill net	0.12	0.01	0.15	0.96	0.06
Anchored gill net	0.14	0.04	0.16	0.05	0.06
Otter trawl	0.93	0.22	0.23	0.54	0.03
Otter trawl pair	0.21	1.00	0.56	0.47	0.03
Zone 3L	0.12	0.75	0.02	0.12	0.99
Zone 3M	0.85	0.75	0.99	0.82	0.16
Zone 3N	0.20	0.75	0.02	0.97	0.85
Zone 3O	0.12	0.75	0.99	0.04	0.02
January	0.03	0.99	0.05	0.05	0.89
February	0.04	1.00	0.87	1.00	0.23
March	0.05	0.88	0.90	1.00	0.23
April	0.16	0.89	0.99	1.00	0.23
May	0.75	0.88	1.00	1.00	0.49
June	0.10	0.37	0.10	1.00	0.23
July	0.04	0.15	0.36	0.02	0.79
August	0.03	0.16	0.30	0.02	0.80
September	0.08	0.97	0.91	0.02	0.80
October	0.10	0.99	0.15	0.05	0.81
November	—	0.99	0.10	0.02	0.88
December	—	0.04	0.17	0.02	0.02
Gill net × f(mesh size)	0.19	1.00	0.08	0.06	0.98
Gill net × $f(\text{mesh size})^2$	0.13	0.94	0.10	0.09	—
Trawl × f(mesh size)	0.58	0.03	0.97	1.00	1.00
Trawl × f(engine power)	0.13	1.00	0.92	0.23	1.00
log(length of vessel)	1.00	1.00	0.11	0.02	1.00
log(GRT)	1.00	1.00	0.15	1.00	0.60

†$f(\cdot)$ denotes the transformation indicated in Table 1.

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

7.2. Posterior results

Here we present some posterior results for the regression coefficients in β and γ, limiting ourselves to some illustrative findings.

We recall that all the available regressors in Table 1 are used for the probit model, where we exclude a reference level (arbitrarily chosen as year 1994, otter trawl by pair, zone 3O and December) for each categorical variable. Since the elements in γ are not directly interpretable, we present posterior results for transformations with a clear interpretation. For the categorical regressors, we compute the difference in the probability of zero catch between a category and its reference case—e.g. year 1993 *versus* year 1994—when all other explanatory variables are evaluated at typical values. Thus, for categorical variables we compute $\Phi(\bar{x}_{\mathrm{c}}'\gamma) - \Phi(\bar{x}_{\mathrm{r}}'\gamma)$, where $\bar{x}_{\mathrm{c}}$ and $\bar{x}_{\mathrm{r}}$ are vectors of 'typical' values, identical except for the relevant categorical variable. For these typical values we take the modal level for categorical variables and median values for continuous variables. We shall consider two sets of values throughout: one corresponding to a typical gill netter (taking modes and medians over the gill net observations, and taking anchored gill net as the reference level for fishing technique) and one corresponding to a typical trawler. For the continuous variables, we consider the derivative of the probability of zero catch with respect to the logarithm of the continuous variable. This gives us the (local) effect on the probability of zero catch of a proportionate change in the underlying continuous variable. As with the categorical variables, this effect will be evaluated at typical values for all regressors.

Since all these measures (called 'effects' in what follows) are functions of γ, we can compute their full posterior distributions. Table 6 presents the posterior mean and standard deviation of the effects of all relevant variables for both typical ships considered. We present results for halibut and redfish only, which are the most important species in terms of live-weight caught. In addition, halibut is the species with the lowest proportion of zero catch (18.5%), whereas redfish has one of the highest proportions of zeros (85.7%).

From Table 6 we see that the regressors can have a large effect on the probability of zero catch, and that the effect is rather specific to the species considered. In view of the decline of the Grand Bank fisheries at the time that the data were collected, we could have expected the year to have a large effect. However, only for grenadier (not presented in Table 6) have we found a substantially lower probability of positive catch in 1994. For the other species the difference is small. We now briefly discuss some results for halibut and merely note that the findings for redfish are often very different, as can be seen directly from Table 6. The probability of catching halibut with a gill net is higher with a drift gill net than an anchored one (which serves as the reference case for computing the effects for gill netters), and a single otter trawl does better than a trawl by pair. As far as the location of catch is concerned, the probability of catching halibut is lowest in the reference zone 3O and highest in zone 3L. The time of the year also has a substantial effect: December is the worst month of the year, whereas March and April seem best. Increasing the mesh size of a gill net in a neighbourhood of the median value (140 mm) has a positive effect on the probability of catch: locally increasing the mesh size by 1% increases the probability of catching halibut by 0.002–0.028. The local effect of changes to mesh size for a typical trawler, however, is much smaller. This illustrates the importance of treating gill nets and otter trawls separately. The engine power of ships with trawl gear does not seem to play a substantial role either, although more power is consistently associated with a higher probability of catch. Finally, longer vessels tend to have a lower probability of zero catch, but this is partly offset by the opposite effect of GRT.

Let us now focus on results for the continuous part, modelled as in Section 5. The coefficient β_l corresponding to a categorical variable has the following interpretation: $\exp(\beta_l)$ is the ratio between the median catch with the corresponding dummy variable equal to 1 and the median

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

Table 6. Posterior moments of some effects in the probit†

Regressor	*Moments for the following species and ships:*			
	Halibut		*Redfish*	
	Typical gill net	*Typical trawl*	*Typical gill net*	*Typical trawl*
Year 1993	−0.00 (0.04)	−0.00 (0.00)	0.02 (0.01)	0.01 (0.01)
Drift gill net	−0.16 (0.09)	—	−0.58 (0.08)	—
Otter trawl	—	−0.39 (0.03)	—	−0.01 (0.01)
Zone 3L	−0.61 (0.05)	−0.49 (0.06)	0.47 (0.06)	0.28 (0.05)
Zone 3M	−0.22 (0.04)	−0.21 (0.05)	0.03 (0.02)	0.07 (0.04)
Zone 3N	−0.48 (0.04)	−0.41 (0.05)	0.52 (0.06)	0.29 (0.05)
January	−0.31 (0.13)	−0.08 (0.05)	0.25 (0.10)	0.10 (0.05)
February	−0.40 (0.12)	−0.09 (0.05)	0.04 (0.02)	0.08 (0.05)
March	−0.52 (0.12)	−0.10 (0.05)	0.03 (0.02)	0.08 (0.05)
April	−0.55 (0.12)	−0.10 (0.05)	0.03 (0.01)	0.08 (0.05)
May	−0.33 (0.12)	−0.08 (0.05)	0.03 (0.02)	0.08 (0.05)
June	−0.21 (0.12)	−0.07 (0.05)	0.09 (0.03)	0.09 (0.05)
July	−0.25 (0.12)	−0.08 (0.05)	0.01 (0.01)	0.05 (0.06)
August	−0.31 (0.12)	−0.08 (0.05)	0.00 (0.01)	0.02 (0.06)
September	−0.35 (0.12)	−0.09 (0.05)	0.00 (0.01)	0.03 (0.06)
October	−0.34 (0.12)	−0.09 (0.05)	0.00 (0.01)	0.04 (0.05)
November	−0.38 (0.12)	−0.09 (0.05)	0.01 (0.01)	0.06 (0.06)
Mesh size	−1.49 (0.67)	−0.05 (0.04)	−0.88 (0.28)	−0.08 (0.04)
Engine power	—	−0.03 (0.01)	—	0.01 (0.00)
Length of vessel	−1.19 (0.15)	−0.10 (0.03)	0.15 (0.08)	0.04 (0.02)
GRT	0.81 (0.08)	0.07 (0.02)	−0.19 (0.06)	−0.05 (0.01)

†Entries are posterior means with standard deviations in parentheses.

catch in case this dummy variable is 0. If a continuous regressor is the logarithm of a variable (length and GRT), then the corresponding regression coefficient β_l is unequivocally interpreted as an elasticity (i.e. it approximately reflects the relative percentage change in median catch as a consequence of a 1% relative change in the original untransformed continuous regressor). For the interactions with trawls, to which the more complicated transformation indicated in Table 1 was applied, the elasticity of the median catch with respect to that regressor is given by β_l times a positive factor (which depends on where we evaluate the elasticity). For the gill net mesh interaction, the elasticity is a linear combination of both the intervening components of β.

The $\tilde{k}$-dimensional ($\tilde{k} = 28$ for most species) regression vector β has a rather complicated posterior distribution, which is a mixture of point masses at zero and continuous parts. It is therefore quite challenging to present this distribution in an easily interpretable format. In what follows, we shall illustrate some aspects of this posterior distribution for halibut. Again, the results vary considerably across species.

Figs 2–4 present, for some selected linear combinations of the components of β, $b'\beta$, the posterior PDF (5.5) for halibut. In addition, the gauge on top (black shading) indicates the posterior probability that $b'\beta \neq 0$. The vertical lines presented in some of these graphs relate to the classical estimate and 90% confidence interval obtained from a stepwise regression technique, as explained and discussed later in Section 7.4.

Fig. 2 focuses on the elements of β corresponding to year and zone. From Table 5 we note that all models visited are 'full' in these two categorical variables (i.e. $L_r - 1$ out of the L_r

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

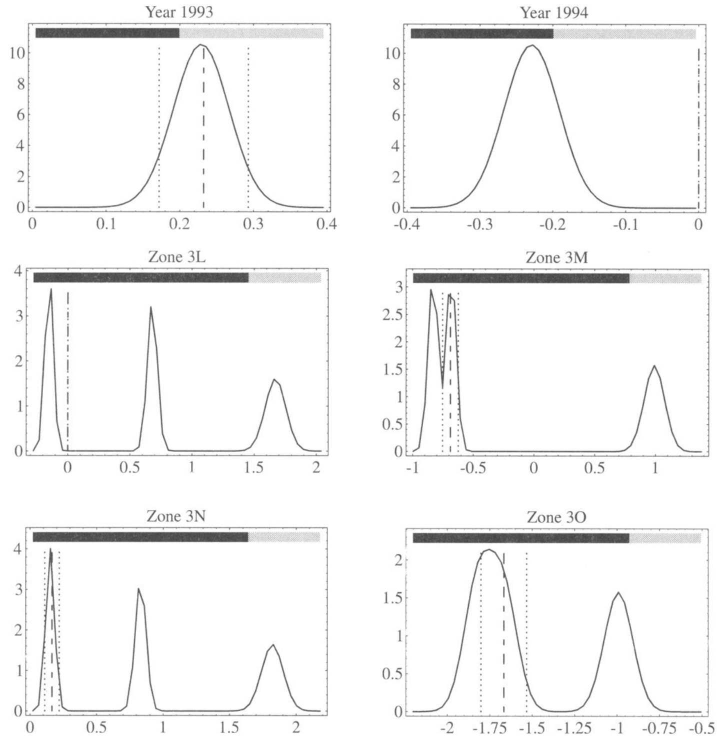

Fig. 2. Halibut: year and zone

possible levels are always included in the model). This induces $L_r - 1$ modes in the marginal posterior PDF for the regression coefficients, where every mode corresponds to a different level being excluded (and, thus, treated as the reference level). For year we have $L_r = 2$ possible levels, leading to unimodal distributions which indicate that 1993 is clearly a better year than 1994. For zone we have $L_r = 4$ levels and we observe the expected $L_r - 1 = 3$ modes. From the relative locations of the modes, it is easy to derive that, for example, the three modes for zone 3L correspond to taking zone 3N, 3M and 3O (from left to right) as reference levels. There is a clear ranking in that zone 3O is the worst, followed by 3M, 3L and 3N, in that order. The difference between zones 3L and 3N is not very large (about 0.15 between the modes, or a factor of 1.16 between median catches), which accounts for the apparent bimodality of the PDF corresponding to zone 3O. Zone 3O is the zone with by far the least observations, leading to Student t-distributions with a large spread in expression (5.5), which means that the modes corresponding to reference cases 3N and 3L can no longer be separately identified in Fig. 2 for zone 3O. In a case such as this, where models are full in a categorical variable, it does not matter which level is taken as a reference level (since all levels are always identified as being different), and we could equivalently fix the reference level and present conditional results instead of the marginal results that are given here. For example, if we give results for zone conditioned on the reference level zone 3O, only the extreme right modes appear for the other zones. However, when more than one level at a time is excluded from models visited (as is usually the case), we need the extra flexibility that is provided by our framework where reference levels are not fixed in advance.

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

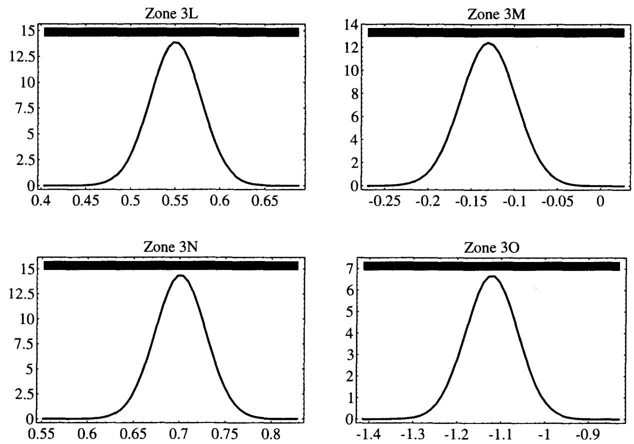

Fig. 3. Halibut: zone with centring

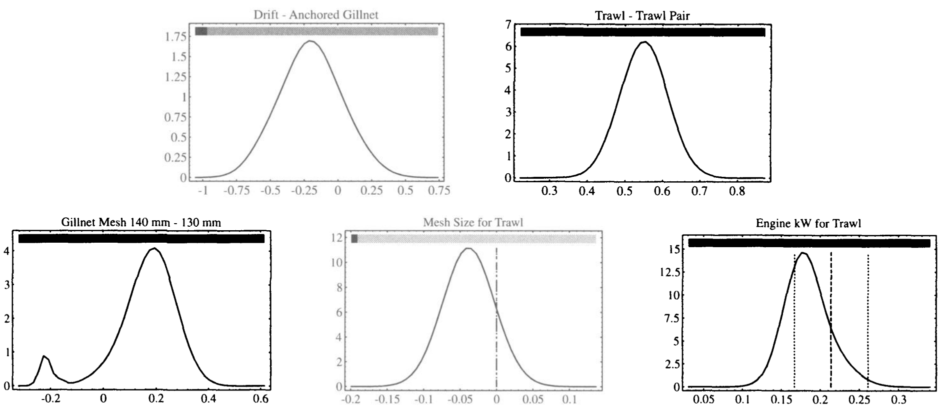

Fig. 4. Halibut: fishing techniques and their interactions

In general, we should aim to present results for quantities that have the same meaning regardless of the choice of reference levels. An interesting way to present regression coefficients of categorical variables is in the form of centred coefficients, i.e.

$$\delta_l \equiv \beta_l - \sum_{i=1}^{L_r} \beta_i / L_r \qquad (l = 1, \ldots, L_r),$$

for a categorical variable with L_r levels and original coefficients $(\beta_1, \ldots, \beta_{L_r})$. Clearly, $\Sigma_{l=1}^{L_r} \delta_l = 0$ and δ_l indicates the difference between level l and the average, so its meaning is not dependent on any particular choice of reference level. Fig. 3 presents the marginal posterior distributions of the centred coefficients associated with each zone. The ranking of zones mentioned above is now immediately obvious from Fig. 3.

The effects of the fishing techniques and their interactions with mesh size and engine power are examined in Fig. 4. From Table 5, we see that the categorical variable corresponding to fishing technique (with $L_r = 4$ levels) is not fully represented in every model. Some levels (the gill net techniques) are almost never included and otter trawl by pair is always included. Thus,

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

otter trawl by pair is never treated as a reference level (indicating that it is quite different from the other levels) whereas often more than one of the other levels are excluded (and thus treated as equal). The fact that models now exclude either one, two or three of these levels at the same time creates more possibilities for modes in the marginal distributions of the associated regression coefficients, and interpretation becomes much more difficult. However, now we would lose flexibility if we fixed a reference level (for example, if we had chosen otter trawl by pair as the reference level, we could not have accommodated the situation described above, where trawl by pair is different from all the others and some of the other levels are equal).

Evaluating the relative merits of the fishing techniques is complicated by the presence of interactions with mesh size and engine power. Therefore, Fig. 4 presents the posterior distribution of the differences between the regression coefficients associated with drift and anchored gill net (which are equally affected by the interactions) and also between those for otter trawl and otter trawl by pair. These are interpretable quantities (logarithms of median catch ratios) and reveal little difference between both gill nets, whereas single trawls tend to do better than trawls by pair. To obtain a rough idea of the overall effects of the different fishing methods, we can consider the configuration of the best model (the model with highest posterior probability), which includes trawl by pair as the only technique and all interactions except for trawl with mesh size. On the basis of the posterior mode of the included regression coefficients for this model, and evaluating the effect at median values for the continuous regressors, we obtain the following ranking from better to worse: otter trawls, trawls by pair (median catch about 58% of otter trawls) and both gill nets (median catch about 10% of otter trawls). These numbers are roughly consistent with the observed values (which are, of course, affected by other factors as well). Fig. 4 also graphs the difference in log(median catch) for two gill net mesh sizes, suggesting a higher median catch for 140 mm mesh (median and third quartile from Table 2) than for 130 mm (first quartile). Finally, for trawls, mesh size is almost never included in the model, whereas engine power has a positive effect on the median positive catch of halibut.

For brevity, we have not shown the posterior density functions of the regression coefficients of the months or the size variables. The main messages here are that the months January until May have a positive effect, whereas July until November lead to a lower median catch of halibut. Finally, length has a positive effect and GRT a negative effect. From Fig. 1, we know that both variables are strongly positively correlated and, on balance, the effect of size on the median catch of halibut will be quite small.

7.3. Predictive results

On the basis of the posterior results partially described above, we shall now predict observations in the subsample that was not used for posterior inference.

First, let us examine how well we predict the probability of zero catch. For every observation in the prediction subsample, we compute $\omega_f = P(s_f = 0|s)$ as in expression (6.2). An interesting check on the adequacy of our probit model is then to compare these predictive probabilities with the actual occurrences of zero or positive catches. Table 7 presents the means and standard deviations of ω_f computed over the zero and the positive observations in the prediction sample. Clearly, ω_f tends to take much higher values for those observations that turn out to be 0, indicating that the probit model does far better than simply assuming that the probability of a zero catch is constant across observations.

Let us now use the predictive results for the continuous part described by density (6.4) to assess the predictive adequacy of the modelling of positive observations. For all the positive observations in the prediction subsample we record in which percentile of the continuous part

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

Table 7. Predictive zero catch probability†

	Probabilities for the following species:				
	Cod	*Halibut*	*Redfish*	*Grenadier*	*Skate*
Zero observations	0.94 (0.14)	0.59 (0.30)	0.92 (0.13)	0.63 (0.30)	0.66 (0.20)
Positive observations	0.45 (0.21)	0.09 (0.14)	0.45 (0.28)	0.29 (0.16)	0.45 (0.19)

†Entries are means with standard deviations in parentheses.

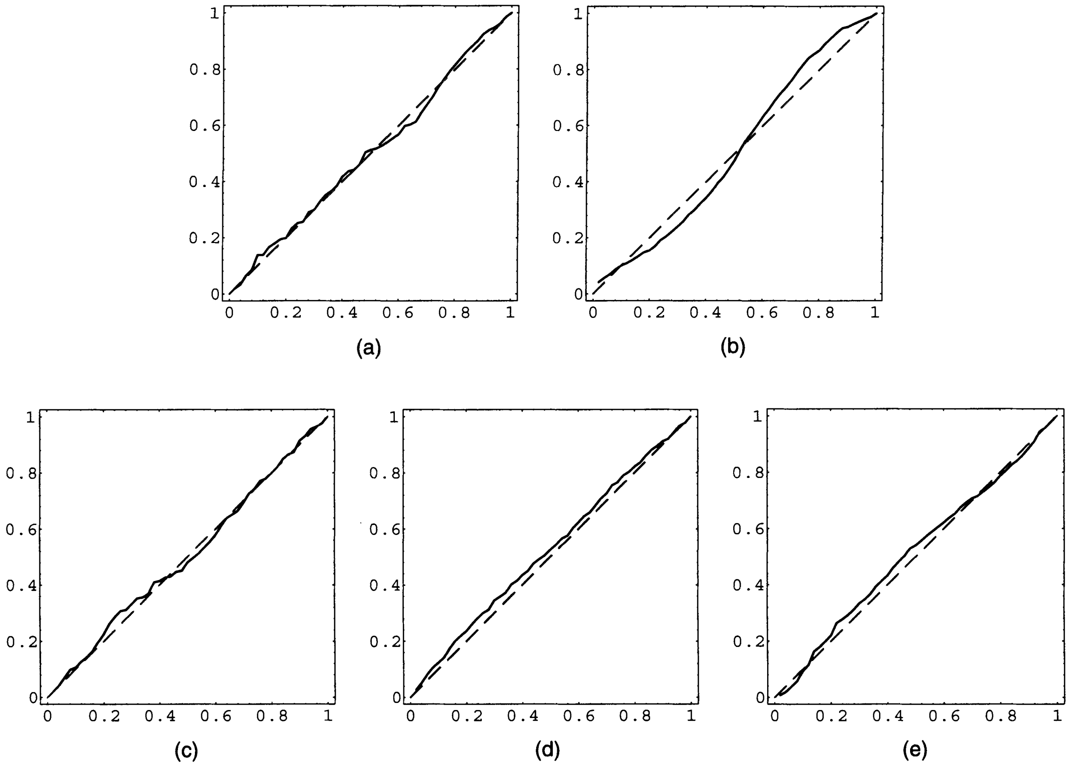

Fig. 5. Q–Q-plots for predictions of individual observations: (a) cod; (b) halibut; (c) redfish; (d) grenadier; (e) skate

of the predictive distributions (using the corresponding values of the regressors) the actual observations fall. Contrasting predictive with empirical quantiles leads to a Q–Q-plot that indicates how well the model (estimated on the basis of the estimation subsample) fits the data in the prediction subsample. As the assignment of observations to either subsample is random, we would expect such plots to be a good measure of model accuracy. Fig. 5 presents these Q–Q-plots for all five species, indicating that the model fit is always quite good.

For illustration, we now show some predictive distributions for particular observations in the prediction subsample. Fig. 6 graphs the predictive PDFs of the non-zero catch of halibut for observations $i = 196$ (270 kg) and $i = 1189$ (3600 kg)—i.e. density (6.4). From the probit

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

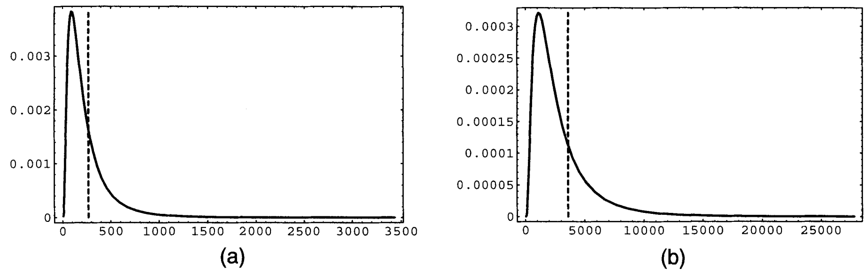

Fig. 6. Halibut: predictive densities and actual observed values for (a) ship-day 196 and (b) ship-day 1189

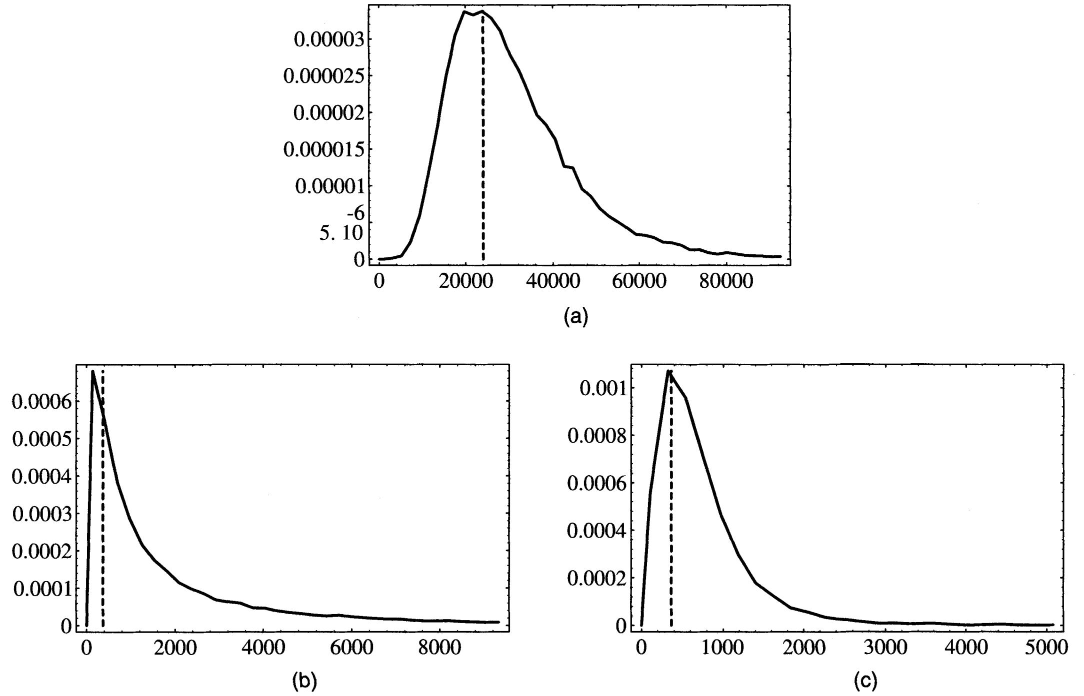

Fig. 7. Predictive densities and actual observed values for clusters: (a) halibut, cluster 277; (b) redfish, cluster 140; (c) grenadier, cluster 236

analysis, the probability of zero catch for observation 196 is 0.43, whereas observation 1189 has a probability of only 0.03 of being equal to 0. The rather different predictive distributions illustrate the importance of the fishing gear. The main difference between the observations is that 196 corresponds to a vessel using drift gill nets whereas 1189 is with an otter trawl. In both cases, the actual catch (indicated by a broken vertical line in Fig. 6) is quite compatible with the predictive distributions.

For policy purposes, it might be interesting to predict, not the catch of one single ship, but the aggregate catch of a number of ships, that are known to be in a certain area of the Grand Bank at a certain time of the year. If we group the data into clusters of 5 ship-days, we can analyse how such predictions, based on the estimation subsample, compare with the actual retained clusters. Clusters of ships that are in the same zone on the same day are likely to be

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

of most interest for practically relevant predictions. To mimic such clusters, we have sorted the prediction subsample by year, day and zone (in that order) and selected clusters of five consecutive observations from that ordering. The predictive distribution, computed as described at the end of Section 6, leads to Q–Q-plots (not shown) that indicate adequate predictions for clusters. Fig. 7 presents some individual cluster predictive densities for the non-zero catch of halibut, redfish and grenadier. The probability of zero aggregate catch varies dramatically across these clusters: from less than $10^{-6}\%$ for halibut to 0.91% for grenadier and 58.8% for redfish. Again, the quantities caught are well matched by the corresponding predictive distributions. These predictive densities immediately lead to probability statements, e.g. about a fleet of certain characteristics exceeding a certain catch, which could straightforwardly be used in a decision theory context.

7.4. Classical methods

In a classical statistical framework, posterior model probabilities are not readily available and, usually, a particular model is selected instead of averaging over models. Given the substantial spread of the posterior mass over the models in $\mathcal{M}$ (see Table 4), that does not seem an adequate strategy for the analysis of these data.

Nevertheless, if we wish to use classical methods for the selection of variables, a popular technique is stepwise regression. Table 4 records the posterior probabilities of the models selected by using forward selection and backward elimination as in Lo *et al.* (1992). Here we base the choice of reference levels for the categorical variables on the best model: we can choose any reference level for those categorical variables in which the best model is full and for the other categorical variables we choose from the levels that are excluded in the best model the level with the lowest posterior inclusion probability (see Table 5).

The models chosen by this stepwise regression technique have between 6 (cod) and 21 (halibut) variables. For grenadier and skate, this method identifies the important variables reasonably well: no variables with posterior probabilities over 0.8 are left out and only one regressor with a posterior inclusion probability under 0.2 is selected (for skate). Accordingly, the posterior probability of the stepwise model is relatively high for these species (see Table 4). For halibut and redfish the performance of stepwise regression is much less in line with the posterior inclusion probabilities. For cod there is an even larger conflict between the stepwise model and the posterior inclusion probabilities, and, as a consequence, the stepwise model picks up virtually no posterior mass.

The classical 90% confidence intervals corresponding to the models selected by stepwise regression are indicated in Fig. 2 and the last two plots of Fig. 4 by dotted vertical lines. The estimated value is indicated by a broken line and a single dash–dot line at zero represents exclusion of the corresponding regressor. Even though some confidence intervals roughly contain 90% of the posterior mass, they can be quite different from the corresponding Bayesian credible intervals.

8. Concluding remarks

In this paper we have outlined the modelling of daily live-weight catches of different species of fish in the Grand Bank fishery. An important feature of the data is the fact that on most days not all species are caught by a certain ship. Thus, modelling of these implicit zero observations is crucial. This was done through a probit model. For the positive observations, we have used a log-normal regression model, where we allow for any combination of regressors from a set of different explanatory variables. We deal with model uncertainty through Bayesian model averaging. Many of the regressors are categorical variables, and we pay particular attention to

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

the treatment of categorical variables in a model uncertainty context. In particular, we allow for any combination of levels to be included in the models, as long as no categorical variable appears with all possible levels (thus, no reference level is fixed in advance). To deal with the resulting 177 million possible models, we apply a Markov chain Monte Carlo algorithm, based on the Metropolis–Hastings sampler, to generate a Markov chain of drawings in this large model space. Throughout, we use a carefully chosen prior distribution which also takes into account that models that are full in categorical variables have equivalent counterparts (corresponding to different choices for the excluded level of these categorical variables), and we examine posterior and predictive inference. The former can be instrumental in policy decisions regarding the effect of certain ship characteristics or regulations concerning, for example, mesh size or fishing techniques. The latter is required if we wish to predict catches per species from easily obtained information regarding the presence of vessels with known characteristics in a certain area at a certain time, rather than having to board these vessels (which is much more costly and altogether impossible for ships from countries outside the Northwest Atlantic Fisheries Organization). The methods would also be useful for estimating the total catch by area when misreporting and illicit landings are common. Bayesian model averaging naturally takes into account all uncertainty concerning parameter values as well as model uncertainty. Thus, realistic predictions can be made for one or more ship-days, duly taking into account the ships' characteristics, location and month as well as parameter and model uncertainty. Using efficient code, new data can easily be processed and posterior and predictive inference can be conducted without excessive computational requirements. We find that the model proposed fits our data relatively well, and that results differ crucially between species.

There are several ways in which the model used here could be extended. A possible elaboration would be to include random ship effects—i.e. ship-specific intercepts—in either the discrete or the continuous part of the model. They could pick up certain quality aspects of the vessels that are not captured in the regressors. A potential interpretation of such individual effects would be as the skill of the captain of the vessel, which was equated with technical efficiency in a stochastic frontier model by Kirkley *et al.* (1998). Barring rather restrictive forms for the distribution of the random effects, this would result in substantial complications: for example, our computations for the continuous part rely on the fact that the marginal likelihood for each model can be computed analytically. We have also avoided including dynamic effects in the model; such effects might provide a 'closer fit' but are not in line with the aim of providing easily computed operational predictions on the basis of available information (which typically does not include a recent history of quantities caught by a cluster of ships considered). In addition, their inclusion would be at the cost of adding to the theoretical and computational complexity of the model. Also, it might be a useful exercise to examine the effects of allowing for heteroscedasticity in the error term of the generic model (3.2) by making σ depend on, for example, the size of the ship. Of course, both the theory and the practical implementation would become more cumbersome as a consequence (unless such a dependence would be fixed, rather than estimated from the data). Finally, again at a considerable cost in terms of added complexity, one might propose a multivariate model for all species with correlated error terms.

Acknowledgements

We thank María del Carmen Gallastegui, Fernando Tusell and two referees for very useful comments. We are grateful to Fernando Tusell and Alain Laurec of the European Commission Fisheries Directorate-General for kindly making the data available to us. During part of this research Carmen Fernández was at the Department of Mathematics, University of Bristol,

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

Eduardo Ley was a Fellow at the Energy and Natural Resources Division of Resources for the Future, Washington DC, and was subsequently at Fundación de Estudios de Economía Aplicada, Madrid, while Mark Steel was at the Department of Economics, University of Edinburgh. Carmen Fernández and Mark Steel were also affiliated to the Center for Economic Research and the Department of Econometrics, Tilburg University, the Netherlands, during much of this work, where Carmen Fernández was supported by training and mobility of researchers grant ERBFMBICT 961021 awarded by the European Commission.

Appendix A: Samplers

A.1. Gibbs sampler for probit model

We introduce independent latent variables m_i $(i = 1, \ldots, n)$, with m_i distributed as normal$(x_i'\gamma, 1)$. From model (3.1), it is immediate that $s_i = 0$ is equivalent to $m_i > 0$, whereas $s_i > 0$ is equivalent to $m_i < 0$. The posterior distribution is, therefore,

$$p(\gamma|s) = p(\gamma|m_i < 0 \text{ for } i = 1, \ldots, Q;\, m_i > 0 \text{ for } i = Q+1, \ldots, n).$$

A Gibbs sampler, augmenting with $m \equiv (m_1, \ldots, m_n)'$, consists of drawing from

$$p(\gamma|m, s) = p(\gamma|m) = f_N^{1+k}[\gamma|\, \{(1+h_0)X'X\}^{-1}\, X'm,\ \{(1+h_0)X'X\}^{-1}]$$

and

$$p(m|\gamma, s) \propto \left\{ \prod_{i=1}^{Q} f_N^1(m_i|x_i'\gamma, 1) I_{(m_i<0)} \right\} \prod_{i=Q+1}^{n} f_N^1(m_i|x_i'\gamma, 1) I_{(m_i>0)}.$$

A.2. Markov chain Monte Carlo sampler on model space

Suppose that the chain is currently at M_s, which has k_s continuous regressors and n_r levels for categorical variable r (where $0 \leqslant k_s \leqslant K$, $r = 1, \ldots, R$ and $n_r \in \{0, 1, \ldots, L_r - 1\}$). Suppose that there are f_s full categorical variables, $c_1, \ldots, c_{f_s}$—i.e. $n_{c_1} = L_{c_1} - 1, \ldots, n_{c_{f_s}} = L_{c_{f_s}} - 1$. The number of regressors in M_s is then $N_s = k_s + n_1 + \ldots + n_R$, whereas the maximum amount of regressors in any model is $N_{\text{tot}} = K + L_1 + \ldots + L_R - R$. The Metropolis–Hastings algorithm proceeds along the following steps.

Step 1: propose a new model M_{can} in several stages. First propose N_{can}:

$$N_{\text{can}} = \begin{cases} N_s + 1 & \text{with probability } (N_{\text{tot}} - N_s)/N_{\text{tot}}, \\ N_s - 1 & \text{with probability } N_s/N_{\text{tot}}. \end{cases}$$

Now propose M_{can} conditionally on the drawn value of N_{can}.

(a) If $N_{\text{can}} = N_s+1$, sample M_{can} by uniformly adding one regressor to M_s, excluding levels of categorical variables in which M_s is already full. We can choose from $N_{\text{tot}} - N_s + R - f_s$ variables, so the probability of adding each is $1/(N_{\text{tot}} - N_s + R - f_s)$. Define $T_{\text{can},\,s} = (N_{\text{tot}} - N_s + R - f_s)/(N_{\text{tot}} - N_s)$. Proceed to step 2.

(b) If $N_{\text{can}} = N_s - 1$, uniformly drop one regressor from M_s to form M_{can}; each choice has probability $1/N_s$. Define $T_{\text{can},s} = (N_{\text{tot}} - N_{\text{can}})/(N_{\text{tot}} - N_{\text{can}} + R - f_{\text{can}})$. Proceed to step 2.

Step 2: compute

(a) $B_{\text{can},s} = l_y(M_{\text{can}})/l_y(M_s)$, where

$$l_y(M_j) \propto \left(\frac{g_0}{g_0+1}\right)^{k_j/2} G_j^{-(Q-1)/2}, \tag{A.1}$$

with

$$G_j = \frac{1}{g_0+1} y' M_{W_j} y + \frac{g_0}{g_0+1} (y - \bar{y}\iota_Q)'(y - \bar{y}\iota_Q), \tag{A.2}$$

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms

where ι_Q is the Q-dimensional vector of 1s, $\bar{y} = \iota_Q' y/Q$, $W_j = (\iota_Q : Z_j)$ and $M_{W_j} = I_Q - W_j (W_j' W_j)^{-1} W_j'$, and

(b) $L_{\text{can},s} = (\Pi_{i=1}^{f_s} L_{c_i}) / \Pi_{i=1}^{f_{\text{can}}} L_{c_i}$, with f_{can} denoting the number of full categorical variables in M_{can}.

Step 3: with probability $q = \min(1, B_{\text{can},s} L_{\text{can},s} T_{\text{can},s})$ the chain moves to M_{can}, whereas with probability $1 - q$ it stays at M_s.

Step 4: record the new state of the chain (be it M_{can} or M_s) after uniformly redrawing the reference level for each of the full categorical variables.

References

Albert, J. H. and Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Ass.*, **88**, 669–679.

Clyde, M., Desimone, H. and Parmigiani, G. (1996) Prediction via orthogonalized model mixing. *J. Am. Statist. Ass.*, **91**, 1197–1208.

Fernández, C., Ley, E. and Steel, M. F. J. (2001a) Benchmark priors for Bayesian model averaging. *J. Econometr.*, **100**, 381–427.

——(2001b) Model uncertainty in cross-country growth regressions. *J. Appl. Econometr.*, **16**, 563–576.

Ferreira, E. and Tusell, F. (1996) Un modelo aditivo semiparamétrico para estimación de capturas: el caso de las pesquerías de Terranova. *Invest. Econ.*, **20**, 143–157.

George, E. I. and McCulloch, R. E. (1997) Approaches for Bayesian variable selection. *Statist. Sin.*, **7**, 339–373.

Hilborn, R. and Walters, C. J. (1992) *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. New York: Chapman and Hall.

King, M. (1995) *Fisheries Biology, Assessment and Management.* Oxford: Fishing News Books.

Kirkley, J., Squires, D. and Strand, I. E. (1998) Characterizing managerial skill and technical efficiency in a fishery. *J. Prod. Anal.*, **9**, 145–160.

Lo, N. C., Jacobson, L. D. and Squire, J. L. (1992) Indices of relative abundance from fish spotter data based on delta-lognormal models. *Can. J. Fish. Aquat. Sci.*, **49**, 2515–2526.

Madigan, D. and Raftery, A. E. (1994) Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Am. Statist. Ass.*, **89**, 1535–1546.

Madigan, D. and York, J. (1995) Bayesian graphical models for discrete data. *Int. Statist. Rev.*, **63**, 215–232.

McAllister, M. K. and Kirkwood, G. P. (1998) Bayesian stock assessment: a review and example application using the logistic model. *ICES J. Mar. Sci.*, **55**, 1031–1060.

Millar, R. B. and Meyer, R. (2000) Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. *Can. J. Fish. Aquat. Sci.*, **57**, 43–50.

Min, C. and Zellner, A. (1993) Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *J. Econometr.*, **56**, 89–118.

Newman, K. B. (1998) State-space modeling of animal movement and mortality with application to salmon. *Biometrics*, **54**, 1290–1314.

Quinn, T. J. and Deriso, R. B. (1999) *Quantitative Fish Dynamics.* Oxford: Oxford University Press.

Raftery, A. E., Madigan, D. and Hoeting, J. A. (1997) Bayesian model averaging for linear regression models. *J. Am. Statist. Ass.*, **92**, 179–191.

Robichaud, D., Hunte, W. and Oxenford, H. A. (1999) Effects of increased mesh size on catch and fishing power of coral reef fish traps. *Fish. Res.*, **39**, 275–294.

Stergiou, K. I., Christou, E. D. and Petrakis, G. (1997) Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. *Fish. Res.*, **29**, 55–95.

Tibbets, J. (1994) Ocean commotion. *Environ. Hlth Perspect.*, **104**, 380–385.

Zellner, A. (1986) On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques—Essays in Honor of Bruno de Finetti* (eds P. K. Goel and A. Zellner), pp. 233–243. Amsterdam: North-Holland.

This content downloaded from 129.219.247.33 on Mon, 12 Dec 2016 23:50:26 UTC
All use subject to http://about.jstor.org/terms