# GENERALIZED LINEAR MIXED EFFECTS MODELS WITH APPLICATION TO FISHERY DATA

JEFFREY JOHN DOWDEN

# *Generalized Linear Mixed Effects Models with Application to Fishery Data*

by

©Jeffrey John Dowden

*A practicum submitted to the School of Graduate Studies in partial fulfillment of the requirement for the Degree of Master of Applied Statistics*

**Department of Mathematics and Statistics**
**Memorial University of Newfoundland**

October, 2007

St. John's Newfoundland Canada

# Abstract

The generalized linear model (GLIM) represents a versatile class of models suitable for several types of dependent variables. GLIMs are popular models and are often an appropriate choice for modelling fisheries data. However, fishery data and corresponding models tend to be complex, because of the complexity of the populations the data are sampled from. In this practicum we use generalized linear mixed effects models (GLMMs), which are GLIMs in which some parameters are random effects to model two different fisheries data sets. The first involves a time series of biological samples used to determine fish maturity, and the second involves paired-trawl catch data to determine if there is a difference in catch rates between two fishing vessels. In this research we find that GLMMs improve estimates of maturities in a selected fish stock and can be used to model differences in catch rates between fishing vessels effectively. This research also suggests that prediction and forecast accuracies are improved by using GLMMs. We also provide some simulation results and found that, overall, GLMMs appear to perform better than GLIMs in terms of bias, coverage errors, and power tests.

# Acknowledgements

It is difficult to overstate my gratitude to my supervisor, Dr. Noel Cadigan. His insight, dedication, and guidance have made it possible for me to complete this practicum. He has been been very generous with this idea and gave me a great opportunity to work on this interesting application of generalized linear mixed effects models. I would have been completely lost without him.

I would like to thank my co-supervisor, Dr. Gary Sneddon. His comments, suggestions, encouragment and ability to explain things clearly has helped me immensely throughout my academic programs.

I acknowledge the financial support provided by the School of Graduate Studies, Department of Mathematics and Statistics, Dr. Noel Cadigan, and Dr. Gary Sneddon in the form of Graduate Assistantship & Teaching Assistantships. Further, I wish to thank Fisheries and Oceans Canada for giving me the opportunity to gain relevant work experience throughout my program.

Last and most importantly, I am very much grateful to my parents, John and Marilyn Dowden, and my girlfriend Catherine Harty for there continuous love, support, and guidance throughout all my endeavors. To them I dedicate this practicum.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Motivation

The generalized linear model (GLIM; McCullagh and Nelder, 1989) represents a versatile class of models suitable for several types of dependent variables such as continuous, dichotomous, and count (see Nelder and Wedderburn, 1972). GLIMs are popular models that have been used in many research areas such as biological sciences, health sciencies, engineering and econometrics. *StatSci.org* (accessed May 16, 2007) presents a selected bibliography of technical work related to this subject.

The GLIM is often an appropriate choice for modelling fisheries data. For example, Jiao and Chen (2004) fitted a GLIM for a production model and sequential population analysis (SPA) to assess a stock of Atlantic cod. They illustrated the problems associated with normality assumptions and concluded that the GLIM should be used to identify appropriate error structures in modelling fish population dynamics. Another example of the application of GLIM's to fisheries data is Ye et. al. (2001).

In this report we use similar GLIM's, or extensions that are described shortly, to model two very different fisheries data sets. The first involves a time series of biological

samples used to determine fish maturity. This important information is used in fish stock assessments. The second data set involves paired-trawl catch data to determine if there is a difference in catch rates between two fishing vessels. This information is important when interpreting fishery survey results from different vessels, and fishery surveys are a fundamental component of most stock assessments. Although these data sets are different in nature, it turns out that similar statistical models can be used to estimate important parameters from these data.

However, fishery data and corresponding models tend to be complex, because of the complexity of the populations the data are sampled from. Realistic models usually have a much larger number of parameters than can be reliably estimated. Fortunately many of these parameters can be realistically viewed as random variables that can be described and also predicted by a much smaller number of variance parameters. This makes the complex fishery models more tractable to estimate. Hence, in our two fishery data sets, we find an advantage in using Generalized Linear Mixed Models (GLMM's), which are GLIM's in which some parameters are actually random effects. The main purpose of our report is to show how to use GLMM's to model two complex fishery data sets. We also provide some simulation results to assess the reliability of the GLMM estimates.

## 1.2 The Generalized Linear Mixed Effects Model

In this section we first describe the GLIM, followed by the GLMM. A GLIM consists of the following components:

First, the response variable vector $\mathbf{Y} = (Y_1, ...., Y_n)$ is denoted as an $n \times 1$ random vector whose distribution is from the exponential family (see Dobson, 2002). In this case the variance of the response depends on the mean ($\mu = E[\mathbf{Y}]$) through a variance function $\mathbf{V}$:

$$Var(\mathbf{Y}) = \Phi \mathbf{w}^{-\frac{1}{2}} \mathbf{V}(\mu) \mathbf{w}^{-\frac{1}{2}} \tag{1.1}$$

where $\Phi$ is a diagonal dispersion matrix which is either known or must be estimated, $\mathbf{w}$ is a diagonal matrix of known weights for each observation, and $\mathbf{V}(\mu)$ is a matrix of the variance function.

Secondly, a monotonic differentiable link function $g(\cdot)$ is specified which describes how the expected value $\mu$ of the response vector $\mathbf{Y}$ is related to a linear predictor $\eta$

$$g(\mu) = \eta. \tag{1.2}$$

The linear predictor incorporates information about the covariates into the model,

$$\eta = \mathbf{X}'\beta, \tag{1.3}$$

where $\mathbf{X}$ is an $n \times p$ matrix of covariates of $\text{rank}(\mathbf{X}) = p$ such that $\mathbf{X}'\mathbf{X}$ is non-singular and $\beta$ is a $p \times 1$ vector of unknown parameters which we also refer to as fixed effects.

Common GLIMs include linear regression, logistic regression and Poisson regression with the corresponding identity, logit, and log link functions respectively.

Fixed effects GLIMs are usually based on the assumption that all observations are independent of each other and are not appropriate for analysis of correlated data, in particular, clustered and/or longitudinal data (e.g. Zeger et. al. 1988).

A generalized linear mixed effects model (GLMM) is more appropriate for the analysis of correlated data. A GLMM is a natural extension to the GLIM whereby a random effect is added to the linear predictor to account for the correlation of the data. Many references on the method are available (e.g. Breslow and Clayton 1993; Lee and Nelder, 1996; Sutradhar, 2003). GLMMs are well suited for biological

and medical data, which normally display heterogeneous responses to treatments. GLMMs are used extensively for data that are not normally distributed. For example, Gilmour et al. (1985) analyzed binomial data using GLMMs, and Agresti et al. (2000) described a variety of social science applications of GLMMs when responses were categorical. Another advantage of the GLMM is the ability to combine data by introducing multilevel random effects (see Goldstein, 1995). Xiao et. al. (2004) cite numerous applications of GLMMs in fisheries sciences.

Suppose that $\mathbf{Y}$ is an $n \times 1$ random vector for the observed data and $\delta$ is an $r \times 1$ vector of random effects. The GLMM is based on the asumption that

$$\mu = E[\mathbf{Y}|\delta] = g^{-1}(\mathbf{X}\beta + \mathbf{Z}\delta) \tag{1.4}$$

where $g^{-1}(\cdot)$ is the inverse of the monotonic link function, $\mathbf{X}$ and $\beta$ are defined as in (1.3), and the matrix $\mathbf{Z}$ is an $n \times r$ matrix for the random effects. The random effects are usually assumed to be normally distributed with mean 0 and unknown variance-covariance matrix $\mathbf{G}$.

The GLMM contains a linear mixed model inside the inverse link function, this is referred to as the linear predictor,

$$\eta = \mathbf{X}\beta + \mathbf{Z}\delta. \tag{1.5}$$

The variance of the observations, conditioned on the random effects, is

$$Var[\mathbf{Y}|\delta] = \mathbf{A}_{\mu}^{1/2}\mathbf{R}\mathbf{A}_{\mu}^{1/2}. \tag{1.6}$$

Here $\mathbf{A}_{\mu}$ is a diagonal matrix containing evaluations at $\mu$ of a linear variance function for the GLMM and $\mathbf{R}$ is a variance-covariance matrix of unknowns (Wolfinger and O'Connell, 1993).

## 1.3 Estimation Methods for Generalized Linear Mixed Effects Models

The primary interest for GLMMs is in the estimation of fixed effects; however, Liang and Zeger (1986) and Zeger et. al. (1988) discuss the interpretation of their estimates in terms of subject-specific (SS) and population-averaged (PA) models. A SS approach focuses on the estimation of the fixed effects parameters $\beta$, the random effects $\delta$, and the variance of the random effects. The PA approach is primarily interested in the estimation of $\beta$ and the marginal variance of $\mathbf{Y}$ which is related to the variance of the random effects. The random effects themselves are treated as nuisance parameters. An example of SS modelling is the best linear unbiased prediction (BLUP; Robinson, 1991), and an example of PA approach (applied to count data) is given by Thall and Vail (1990).

Fitting a linear mixed model using a likelihood approach consists of specifying a distribution for the random effects and then estimating the unknown parameters using maximum likelihood (ML) or restricted maximum likelihood (REML). The REML approach produces unbiased estimates of variance parameters in some problems (e.g. Harville, 1977; McGilchrist, 1994). These methods are usually referred to as marginal approaches and typically involve numerical integrations over the random effects. The ML approach is also a PA approach because the random effects are not estimated.

Suppose that $\mathbf{Y_i}$ is a vector of observed data for each of $i$ subjects, $i = 1, ..., k$. $\mathbf{Y_i}$ is assumed to be independent across $i$, but within subject covariance is likely to exist because each of the elements of $\mathbf{Y_i}$ is measured on the same subject. Assume that a random effects vector $\delta_\mathbf{i}$ exists that is also independent across $i$. Assuming an appropriate model linking $\mathbf{Y_i}$ and $\delta_\mathbf{i}$ exists (i.e. a GLIM) and this model involves covariates $\mathbf{X_i}$ that are related to the mean of $\mathbf{Y_i}$ conditional on $\delta_\mathbf{i}$, the joint probablility density function is

$$p(\mathbf{Y_i}|\mathbf{X_i}, \beta, \mathcal{R}, \delta_i)q(\delta_i|\mathcal{G}) \tag{1.7}$$

where $p(\cdot)$ is the conditional probability density function of $\mathbf{Y_i}$, $q(\cdot)$ is the probability density function of $\delta_i$, $\mathbf{X_i}$ is a matrix of observed explanatory variables, $\beta$ is a vector of unknown parameters, $\mathcal{R}$ is a vector of unknown unique elements of $\mathbf{R}$ (the variance-covariance matrix of the observations), and $\mathcal{G}$ is a vector of unknown unique elements of $\mathbf{G}$ (the variance-covariance matrix of the random effects). Let $\vartheta = (\beta, \mathcal{R}, \mathcal{G})'$, likelihood inferences based about $\vartheta$ are based on the marginal likelihood function

$$M(\vartheta) = \prod_{i=1}^{k} \int \cdots \int p(\mathbf{Y_i}|\mathbf{X_i}, \beta, \mathcal{R}, \delta_i)q(\delta_i|\mathcal{G})d\delta_i. \tag{1.8}$$

In particular, the function

$$f(\vartheta) = -\log M(\vartheta) \tag{1.9}$$

is minimized over $\vartheta$ numerically in order to estimate $\vartheta$, and the inverse Hessian (second-order derivative) matrix provides an approximate variance-covariance matrix for the estimates of $\vartheta$. The function $f(\vartheta)$ is referred to as the negative log-likelihood function or the objective function for optimization.

There are limitations to the marginal method. Likelihood equations tend to be complex and difficult to derive. Rarely will closed form expressions exist for the marginal likelihood. In some instances the data may contain a large number of random effects which leads to a high dimensional integral for the marginal likelihood equation. Numerical integration techniques have been used such as guassian quadrature (e.g. Anderson and Aitkin, 1985; Davidian and Gallant, 1993) and Gibbs sampling (Zeger and Karim, 1991). High dimensional integrals can be very computationally intensive to solve numerically, and in some cases are not feasable (Stiratelli et. al, 1984).

Another approach for estimating $\beta$, $\delta$, $\mathbf{G}$ and $\mathbf{R}$ is the pseudo-likelihood (PL) and restricted pseudo-likelihood (REPL) procedure (Wolfinger and O'Connell, 1993). Implementation of the PL and REPL procedure first involves linearizing the data using a first order Taylor's series approximation expanding about initial estimates of the fixed regression parameters and random effects. Then normal linear model theory is used to estimate variance parameters. The variance parameters are then used to estimate fixed regression parameters and predict random effects which, in turn, are used to linearize the data again to estimate new variance parameters. This process is repeated until a specific tolerance level is obtained (i.e. convergence). This procedure is described in more detail later in this chapter.

Several other estimation techniques for GLMMs have appeared in the literature. Breslow and Clayton (1993) presented two estimation procedures referred to as penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL), although, these two methods correspond to the SS and PA models of Zeger et. al. (1988) respectively. As well, the implementation of PQL and MQL can be achieved using PL (Wolfinger and O'Connell, 1993).

Waclawiw and Liang (1992) predicted random effects of a GLMM by iteratively solving a set of Stein-type estimating equations. This SS approach is similar to the PL in its iterative nature, although they replace the mixed model and ML/REML equations with optimal estimating equations for fixed effects, random effects, and variance parameters.

Sutradhar and Rao (2001) considered an exact MQL approach, however, this PA approach was only developed for small values of the variance of the random effects. Sutradhar (2004) since improved on the exact MQL approach by proposing an exact quasi-likelihood or generalized quasi-likelihood (GQL) method whereby the covariance matrix needed to construct the estimating equation has been computed for small or large values of the variance of the random effects.

Lee and Nelder (1996) used a hierarchical likelihood (HL) approach to estimate

fixed parameters and random effects. In this SS approach, the random effects were treated as fixed effects and then were used to obtain estimates of the variance components. This approach was similar to the PQL method proposed by Breslow and Clayton (1993).

The PL/REPL approach appears to be an appropriate choice and is useful for modelling GLMMs since it provides a unified framework for both SS and PA inference and includes PQL and MQL as special cases. As well, PL/REPL algorithms can be implemented using mixed model software packages (see Section 1.3; *SAS/STAT*® PROC GLIMMIX).

We provide more details about the PL/REPL approaches below, summarized from Wolfinger and O'Connell (1993).

Let $\hat{\beta}$ and $\hat{\delta}$ be known estimates of $\beta$ and $\delta$ and recall that

$$E[\mathbf{Y}|\delta] = \mu = g^{-1}(\mathbf{X}\beta + \mathbf{Z}\delta) = g^{-1}(\eta), \tag{1.10}$$

which is a vector consisting of evaluations of $g^{-1}$ at each component of $\eta$. Now let

$$g^{-1}(\eta) = g^{-1}(\hat{\eta}) + \tilde{\Upsilon}(\mathbf{X}\beta - \mathbf{X}\hat{\beta}) + \tilde{\Upsilon}(\mathbf{Z}\delta - \mathbf{X}\hat{\delta}) \tag{1.11}$$

where

$$\tilde{\Upsilon} = \left(\frac{\partial g^{-1}(\eta)}{\partial \eta}\right)_{\hat{\beta},\hat{\delta}} \tag{1.12}$$

is a diagonal matrix with elements consisting of the first derivative of $g^{-1}$. Note that ( i ) is a first-order Taylor series approximation of $\mu$ expanding about $\hat{\beta}$ and $\hat{\delta}$. Rearranging the terms yields the expression

$$\tilde{\Upsilon}^{-1}(\mu - g^{-1}(\hat{\eta})) + \mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\delta} = \mathbf{X}\beta + \mathbf{Z}\delta. \tag{1.13}$$

The left-hand side is the expected value, conditional on $\hat{\delta}$ and $\hat{\beta}$, of

$$\tilde{\Upsilon}^{-1}(\mathbf{Y} - g^{-1}(\hat{\eta})) + \mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\delta} \equiv \mathbf{P} \tag{1.14}$$

and

$$Var[\mathbf{P}|\delta] = \tilde{\Upsilon}^{-1}\mathbf{A}_{\mu}{}^{1/2}\mathbf{R}\mathbf{A}_{\mu}{}^{1/2}\tilde{\Upsilon}^{-1}. \tag{1.15}$$

Thus we can consider the model

$$P = \mathbf{X}\beta + \mathbf{Z}\delta + \varepsilon \tag{1.16}$$

as a linear mixed model with pseudo-response (i.e. linear mixed pseudo-model) $\mathbf{P}$, fixed effects $\beta$, random effects $\delta$, and $Var[\varepsilon] = Var[\mathbf{P}|\delta]$.

Now define

$$\mathbf{V}(\theta) = \tilde{\Upsilon}^{-1}\mathbf{A}_{\mu}{}^{1/2}\mathbf{R}\mathbf{A}_{\mu}{}^{1/2}\tilde{\Upsilon}^{-1} + \mathbf{Z}\mathbf{G}\mathbf{Z}' \tag{1.17}$$

as the marginal variance in the linear mixed pseudo-model, where $\theta$ is a $q \times 1$ parameter vector containing all unknown parameters in $\mathbf{R}$ and $\mathbf{G}$ and $\mathbf{Z}'$ is the transposed matrix for the random effects. Based on this linearized model, an objective function can be defined, assuming the distribution of $\mathbf{P}$ is normal. The maximum log pseudo-likelihood for $\mathbf{P}$ is

$$\ell(\theta, \mathbf{p}) = -\frac{1}{2}\log|\mathbf{V}(\theta)| - \frac{1}{2}\mathbf{r}'\mathbf{V}(\theta)^{-1}\mathbf{r} - \frac{n}{2}\log(2\pi) \tag{1.18}$$

and the restricted maximum log pseudo-likelihood is

$$\begin{aligned} \ell_R(\theta, \mathbf{p}) &= -\frac{1}{2}\log|\mathbf{V}(\theta)| - \frac{1}{2}\mathbf{r}'\mathbf{V}(\theta)^{-1}\mathbf{r} \\ &\quad -\frac{1}{2}\log|\mathbf{X}'\mathbf{V}(\theta)^{-1}\mathbf{X}| - \frac{n-k}{2}\log(2\pi) \end{aligned} \tag{1.19}$$

where $\mathbf{r} = \mathbf{p} - \mathbf{X}(\mathbf{X}'\mathbf{V}(\theta)^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}(\theta)^{-1}\mathbf{p}$, $\mathbf{p}$ is a realization of the random vector $\mathbf{P}$, $n$ denotes the number of observations, and $k$ is the rank of $\mathbf{X}$. Numerical methods (i.e. Newton-Raphson, quasi-Newton) are generally required to maximize $\ell$ and $\ell_R$ over the parameters $\theta$. After obtaining estimates for $\theta$, estimates for $\beta$ and $\delta$ are computed as

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}(\hat{\theta})^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}(\hat{\theta})^{-1}\mathbf{p} \tag{1.20}$$

$$\hat{\delta} = \hat{\mathbf{G}}\mathbf{Z}'\mathbf{V}(\hat{\theta})^{-1}\hat{\mathbf{r}} \tag{1.21}$$

With $\beta$ and $\delta$ set to the estimates the linearization is re-computed, ( ) and ( ) are maximized to obtain updated estimates of $\mathbf{R}$ and $\mathbf{G}$. This is iterated until convergence. This involves two levels of iteration: one for the linearization, and one for the estimation of the variance parameters in the linearized model.

In some cases, the conditional distribution may contain a scale parameter ($\phi \neq 1$). The variance function becomes

$$\mathbf{V}(\theta^*) = \tilde{\Upsilon}^{-1}\mathbf{A}_\mu^{1/2}\mathbf{R}^*\mathbf{A}_\mu^{1/2}\tilde{\Upsilon}^{-1} + \mathbf{Z}\mathbf{G}^*\mathbf{Z}' \tag{1.22}$$

where $\theta^*$ is the covariance parameter vector with $q-1$ elements. The matrices $\mathbf{R}^*$ and $\mathbf{G}^*$ are re-parameterized versions of $\mathbf{R}$ and $\mathbf{G}$ in terms of $\phi$. The maximum log pseudo-likelihood for the linear mixed pseudo-model ( ) is

$$\ell(\theta^*, \mathbf{p}) = -\frac{1}{2}\log|\mathbf{V}(\theta^*)| - \frac{n}{2}\{\mathbf{r}'\mathbf{V}(\theta^*)^{-1}\mathbf{r}\} - \frac{n}{2}(1 + \log\{2\pi/n\}) \tag{1.23}$$

and the restricted maximum log pseudo-likelihood is

$$\begin{aligned}\ell_R(\theta^*, \mathbf{p}) &= -\frac{1}{2}\log|\mathbf{V}(\theta^*)| - \frac{n-k}{2}\{\mathbf{r}'\mathbf{V}(\theta^*)^{-1}\mathbf{r}\} \\ &\quad -\frac{1}{2}\log|\mathbf{X}'\mathbf{V}(\theta^*)^{-1}\mathbf{X}| - \frac{n-k}{2}(1+\log\{2\pi/(n-k)\}). \qquad (1.24)\end{aligned}$$

The solutions for $\hat{\beta}$, $\hat{\delta}$ and $\hat{\phi}$ are

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}(\hat{\theta}^*)^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}(\hat{\theta}^*)^{-1}\mathbf{p} \qquad (1.25)$$

$$\hat{\delta} = \hat{\mathbf{G}}\mathbf{Z}\mathbf{V}(\hat{\theta}^*)^{-1}\hat{\mathbf{r}} \qquad (1.26)$$

$$\hat{\phi} = \hat{\mathbf{r}}'\mathbf{V}(\hat{\theta}^*)^{-1}\hat{\mathbf{r}}/n^* \qquad (1.27)$$

where $n^*$ equals $n$ for PL and $n-k$ for REPL.

## 1.4 Statistical Software Packages

A key feature of the PL/REPL method is its ability to be implemented using standard statistical software packages. In this practicum, we use three software procedures developed by the SAS Institute for estimating parameters in GLIMs and GLMMs: PROC GENMOD (generalized linear models), PROC NLMIXED (non-linear mixed effects models), and PROC GLIMMIX (generalized linear mixed effects models). Each method will be compared in terms of the accuracy of parameter estimates and model (i.e. response) predictions.

### 1.4.1 The GENMOD Procedure

The GENMOD procedure fits a GLIM to the data by maximum likelihood estimation over the vector of unknown coefficients ($\beta$). In general, there is no closed form solution for the maximum likelihood estimates. GENMOD estimates the parameters of the model using an iterative fitting process. The dispersion parameter ($\phi$) is also estimated by either maximum likelihood, by the residual deviance, or by Pearson's chi-squared divided by the degrees of freedom. Covariances, standard errors, and $p$-values for the parameter estimates are computed based on the asymptotic normality of maximum likelihood estimators.

A number of link functions and probability distributions are available for the GENMOD procedure. The link functions include the identity, logit, probit, log, and complementary log-log. The distributions include normal, binomial, Poisson, gamma, inverse Gaussian, negative binomial, and multinomial.

The GENMOD procedure has the abilility to fit correlated response data by the generalized estimating equation (GEE) method (Liang and Zeger, 1986), although we do not utilize this feature of the software in our analyses.

### 1.4.2 The NLMIXED Procedure

PROC NLMIXED allows you to specify, conditional on the random effects, a distribution for the response variable that has either a standard form (normal, binomial, Poisson) or a general distribution defined by the user. PROC NLMIXED fits nonlinear mixed models by maximizing an approximation to the likelihood integrated over the random effects. Such marginal methods are commonly used with mixed models. Different integral approximations are available. These include Gaussian quadrature (Pinheiro and Bates, 1995) and first-order Taylor series approximation (Beil and Sheiner, 1988). Successful convergence of the optimization procedure results in parameter estimates along with their standard errors based on the Hessian matrix of

the likelihood function.

The NLMIXED procedure only implements maximum likelihood. This is because the analog to the restricted maximum likelihood method in PROC NLMIXED software would involve a high dimensional integral over all of the fixed-effects parameters, and this integral is typically not available in closed form.

### 1.4.3 The GLIMMIX Procedure

The GLIMMIX procedure fits GLMMS based on linearizations (see Section 1.3). A Taylor series expansion is used to approximate the GLMM as a linear mixed model. The advantage of the linearization is that only the variance parameters have to be estimated numerically because closed form expressions exist for the regression parameter estimates. The linearization method is doubly iterative. The approximate linear mixed model is fit which is itself an iterative process, then the new parameter estimates are used to update the linearization, which results in a new linear mixed model. The process stops when parameter estimates between sucessive linear mixed model fits change within a specified tolerance. The default estimation method in PROC GLIMMIX software for models containing random effects is restricted pseudo-likelihood (REPL). Maximum likelihood estimates of variance parameters tend to be biased for small sample sizes. The REPL may provide less biased estimation of random effect variance parameters.

An advantage of linearization based methods is that they can use a relatively simple form of the linearized model that typically can be fit based on only the mean and variance in the linearized form. Models for which the marginal distribution is difficult, or impossible, to compute can be fit with linearization. This approach is well suited for models with correlated errors and a large number of random effects.

A disadvantage of this approach is the absence of a true objective function and potentially biased estimates of covariance parameters, especially for binary data. In a

GLMM is not always possible to derive the exact log-likelihood of the data, therefore likelihood based tests and statistics are often hard to derive. PROC GLIMMIX produces Wald-type test statistics (e.g. Buse, 1982) and confidence intervals.

PROC GLIMMIX software provides marginal and conditional residuals. Conditional residuals are based on predictors of the random effects and estimates of the fixed effects regression parameters. The predictors of the random effects are the estimated best linear unbiased predictors (BLUPs) in the approximated linear model.

## 1.5 Scope of the Practicum

The following is an outline for the remainder of the practicum. In Chapter 2 we apply the GLMM to maturity data for a selected fish stock off the southern coast of Newfoundland. In Chapter 3 we apply the GLMM to fishery survey calibration data from two research vessels fishing in the Northern Gulf of St. Lawerence. In Chapter 4 a simulation study is presented on the properties of estimators based on fishery calibration data. This practicum contains various acroynms (i.e. GLIM, GLMM), a table of acronyms along with their corresponding descriptions is given in Appendix B.

# Chapter 2

# Application of GLMM: Fish Stock Maturities Data

## 2.1 Introduction

Generalized linear mixed models have become an increasingly important method for fisheries research in recent years (Xiao et al., 2004). These models are suitable for analyzing complex count data. Models for such data often have a large number of parameters to estimate, many of which can usefully be considered as random variables to improve estimation (or prediction). In this chapter we examine the application of mixed models to improve estimation and inference for fish stock maturation rates. This is an important problem in fisheries sciences. Maturation rates are fundamental to understanding the dynamics and productivity of fish stocks. Good estimates of maturation rates are required for successful management of commercially exploited fish stocks (see Olsen et. al., 2005).

The mature component of a fish stock is usually referred to as the spawning stock biomass (SSB). It can be defined as the product of biomass-at-age and proportion mature-at-age (maturities), summed over all ages in the stock. Good estimates of

maturities are required for good estimates of SSB. Poor estimates of maturities can have deleterious impacts on estimation of the potential yield of a fishery, population growth and the health of stocks (Welch and Foucher, 1988). Maturities can change over time because of many factors (eg. Beacham, 1983; Pitt, 1975; Templeman et al, 1978; Shelton and Armstrong, 1983; Morgan and Colbourne, 1999). Annual estimates are required; however, the most appropriate way to produce such estimates is by cohort (see Morgan, 2000). A cohort is a group of animals (or more generally individuals) with the same birth year.

Biological sampling programs for fish stocks provide counts of the number mature. Each year such sampling produces data on the number of fish examined, their age, and the number found to be mature. The probability of maturing is an increasing function of age within a cohort. A common model used to estimate maturities is logistic regression, where age is the covariate. This is a fixed effects generalized linear model (GLIM) with a logit link function (McCullagh and Nelder, 1989). This model is fitted to each cohort when maturity data for sufficient ages have been collected. However, there are problems with this approach. Data are updated annually for unfinished (eg. recent) cohorts and this can result in substantial changes from year to year in the estimated maturities for that cohort. For example, the maturity at age 5 in 2003 estimated using cohort data up to 2004 can be quite different than the 2003 estimate using data up to 2003.

Often the annual trends in cohort maturities are fairly smooth (see Needle et. al., 2003). The purpose of this chapter to investigate if a generalized linear mixed effects model (GLMM) can be used to improve estimates of maturities, particularly for unfinished cohorts, by utilizing the autocorrelation structure in cohort maturities. We examine an important commercial species in the Northwest Atlantic, the Atlantic cod (*Gadus morhua*). Also, fisheries managers often consider changes in SSB in stock projections for different future management scenarios and they require that maturities be forecasted in the next several years (or more) to compute SSB's. We also investigate if mixed models can improve forecasted maturities.

## 2.2 Materials and Methods

### 2.2.1 Data

North Atlantic cod data were collected during annual research vessel trawl surveys by Fisheries and Oceans Canada (DFO). The trawl surveys we consider involved towing a scientifically-standardized fishing trawl for a fixed distance and at a fixed speed. Tow sites (or sets) were selected at random using a depth-area based stratified sampling design. More details about the surveys are given in Chapter 3. We examined data for cod in Northwest Atlantic Fisheries Organization (NAFO) sub-division 3Ps collected during 1960-2005. NAFO fisheries management divisions are shown in Fig. [illegible]. The maturity data were also collected using a length stratified sampling scheme (Doubleday, 1981).

Some of the variables recorded for fish in each catch were year, cohort (year-age), sex, age, number per catch and proportion mature per catch. Due to sex-specific differences in maturation, females and males were treated separately (Barot et al., 2005; Swain and Poirer, 1997). Each fish was classified as mature or immature based on the criteria of Templeman et al. (1978). The 3Ps cod data contained 25810 observations in total, of which 12275 were males and 13535 were females. The number-per-set sampled for maturities for different years and ages ranged from 1 to 186. Zero catches were discarded since they provided no information on maturity. Observed proportions mature at age, $p_a$, were calculated by taking into account the length frequency of the population and the length stratified sampling (Morgan and Hoenig, 1997).

### 2.2.2 Fixed Effects Model

Define $p_c(a)$ to be the probability that a fish is mature at age $a$ in cohort $c$. If fish are sampled at random from the stock then the binomial model is appropriate to use

(see Cox and Snell, 1989). The binomial probability of observing $y_{ac}$ mature fish at age $a$ in cohort $c$ from a random sample of $n_{ac}$ fish is

$$Pr(Y_{ac} = x) = \binom{n_{ac}}{x} p_c(a)^x \{1 - p_c(a)\}^{n_{ac}-x}, \quad x = 0, ..., n_{ac}. \tag{2.1}$$

The mean and variance of the binomial distribution are $E(Y_{ac}) = n_{ac}p_c(a)$ and $Var(Y_{ac}) = n_{ac}p_c(a)\{(1 - p_c(a)\}$.

It is reasonable to assume that $p_c(a)$ is a smooth monotone increasing function of age within cohorts. A common model used in this situation is the fixed effects (FE) logistic regression model

$$p_c(a) = \frac{exp(\beta_{0c} + \beta_{1c} \times a)}{1 + exp(\beta_{0c} + \beta_{1c} \times a)}, \tag{2.2}$$

which is the canonical link function for the Binomial distribution (McCullagh and Nelder, 1989). This model is used in other areas of fisheries research such as fishing gear-selectivity studies (Millar, 1992). We used the maximum likelihood estimation method and $SAS/STAT^{®}$ PROC GENMOD (SAS Inst., 2005) to estimate the logistic regression parameters.

Define $A_{50}$ to be the age at 50% maturity, $p_c(A_{50}) = 0.5$. From ( ' ') this value is $A_{50} = \frac{-\beta_{0c}}{\beta_{1c}}$ and can be estimated as

$$\hat{A}_{50} = \frac{-\hat{\beta}_{0c}}{\hat{\beta}_{1c}}, \tag{2.3}$$

where $\hat{\beta}_{0c}$ and $\hat{\beta}_{1c}$ are the estimates of the logistic regression parameters. The maturity range ($MR$) is define to be the difference between the age at 75% maturity and the age at 25% maturity, which is $MR = A_{75} - A_{25} = \frac{\log(9)}{\beta_{1c}}$. It can be estimated as

$$\widehat{MR} = \hat{A}_{75} - \hat{A}_{25} = \frac{\log(9)}{\hat{\beta}_{1c}}. \tag{2.4}$$

In some instances over-dispersion may arise. Over-dispersion occurs when data have more variability than accounted for in our modelling assumptions. It can be caused by many factors such as population spatial heterogeneity. An approach to deal with over-dispersion is to use quasi-likelihood estimation (McCullagh and Nelder, 1989) with $Var(Y_{ac}) = \phi n_{ac} p_c(a)\{1 - p_c(a)\}$, where $\phi$ is an over-dispersion parameter. This is the approach used in PROC GENMOD to account for over-dispersion.

There were insufficient data for some cohorts to estimate model parameters. For example, the maturities for the 2001 3Ps cod cohort were only observed at ages 1 to 4 in years 2002 to 2005. This covers only the lower portion of the maturity ogive and does not cover a sufficient range of ages to estimate the regression parameters. Similarly, some of the earlier cohorts are only observed at older ages that also do not cover a large enough age range to estimate the model parameters. Therefore, we estimated parameters only for the 1954-2000 cohorts.

### 2.2.3 Mixed Effects Model

In the previous section we considered the parameters $\beta_{0c}$ and $\beta_{1c}$ as fixed effects; that is, fixed but unknown parameters to estimate. However, the maturities tend to change smoothly over time and $\beta_{0c}$s and $\beta_{1c}$s appear to be autocorrelated over cohorts (i.e. time). To account for this autocorrelation structure we use a mixed effects (ME) model. Recall that data are collected each year for unfinished cohorts and additional data collected each year can cause substantial changes in the parameter estimates from year to year. Accounting for the autocorrelation structure in $\beta_{0c}$ and $\beta_{1c}$ may help reduce this problem. This is our main rationale for investigating a ME model approach. We used $SAS/STAT^{®}$ PROC GLIMMIX to compute estimates for ME models with and without over-dispersion. We compared both estimates in terms of their variability (smoothness) and predictive capability (described below).

### Autocorrelated Cohort Effects

A model we investigate for the proportion mature-at-age is

$$p_c(a) = \frac{exp\{(\beta_0 + \delta_{0c}) + (\beta_1 + \delta_{1c}) \times a\}}{1 + exp\{(\beta_0 + \delta_{0c}) + (\beta_1 + \delta_{1c}) \times a\}}, \tag{2.5}$$

where $\delta_{0c}$ and $\delta_{1c}$ are the random cohort effects for the intercept and age coefficients, respectively, while $\beta_0$ and $\beta_1$ are fixed effects representing the average intercept and age effects for all cohorts. We assume that the $\delta$'s are random variables from a normal distribution with mean zero, but are autocorrelated over cohorts; $\delta_{0c} \sim N(0, \sigma^2_{\delta_0})$, $\delta_{1c} \sim N(0, \sigma^2_{\delta_1})$, $Corr(\delta_{0j}, \delta_{0k}) = \gamma_0^{|j-k|}$ and $Corr(\delta_{1l}, \delta_{1m}) = \gamma_1^{|l-m|}$. These are AR(1) correlations with $\gamma_0$ and $\gamma_1$ autocorrelations respectively. Diagnostics for this assumption are presented in the next section. A normal distribution is commonly used to model random effects.

The estimates of $\beta_0$ and $\beta_1$ are added to the BLUP's of $\delta_{0c}$ and $\delta_{1c}$ (see Chapter 1), respectively, to predict the logistic regression model's random slopes and intercepts for each cohort,

$$\hat{\beta}_{0c} = \hat{\beta}_0 + \hat{\delta}_{0c}, \tag{2.6}$$

$$\hat{\beta}_{1c} = \hat{\beta}_1 + \hat{\delta}_{1c}. \tag{2.7}$$

These predictions are used to predict $A_{50}$ and $MR$ using ( ) and ( ). The notation ˆ is used interchangeably between estimate and prediction. The difference will depend on whether the estimate of the effect is fixed (estimate) or random (prediction).

### Autocorrelated Cohort and Random Year Effects

There is a great deal of spatial and temporal variation in maturity at age, and size and maturation can be heterogeneous across the range of a population (Korsbrekke,

1999; Bromley, 2000; Armstrong et al., 2003; Gerritsen et al., 2003). If sampling does not cover the full range of a population's range or if there is annual variation in the distribution of samples across the population's range, then the calculated proportion mature may not be representative of the population. In addition, environmental conditions (food, temperature, etc) may be particularly good or bad in a given year, leading to more or less individuals making the decision to become adult. Sampling and environmental factors lead to year effects, apparent or real, when modelling the data.

A model that accommodates year effects for the probability that a fish from cohort $c$ is mature at age $a$ in year $y$, $p_{cy}(a)$, is

$$p_{c,y}(a) = \frac{exp\{(\beta_0 + \delta_{0c} + \eta_{0y}) + (\beta_1 + \delta_{1c}) \times a\}}{1 + exp\{(\beta_0 + \delta_{0c} + \eta_{0y}) + (\beta_1 + \delta_{1c}) \times a\}}. \tag{2.8}$$

In this model the random slope effects ($\delta_{1c}$s) for age are auto-correlated, similar to (.  ). However, the random intercept effects are composed of two separate effects, an autocorrelated cohort effect ($\delta_{0c}$) and a simple uncorrelated year effect $\eta_{0y}$. The $\eta_{0y}$s are i.i.d. $N(0, \sigma^2_\eta)$. Note that for this model we do not investigate additional over-dispersion. The year effects are assumed to account for any additional over-dispersion beyond the correlated random slope and intercept cohort effects.

### 2.2.4 Autocorrelation Diagnostics

The logistic regression intercepts and slopes ($\beta_{0c}$ and $\beta_{1c}$) appear to be autocorrelated over cohorts; however, there are many types of autocorrelation structures. It is important to identify the basic type of autocorrelation structures to improve estimation of $\beta_{0c}$ and $\beta_{1c}$. Two common models are the autoregressive process with order $p$, AR($p$), and the moving average process with order $q$, MA($q$). The AR($p$) process uses the idea that the present observation of some arbitrary series $z_t$ can be explained as a function of the $p$ past observations, $z_{t-1}, z_{t-2}, \ldots, z_{t-p}$, where $p$ determines the

number of steps (lags) in the past needed to forecast the current value and is defined as

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + ... + \phi_p z_{t-p} + e_t, \tag{2.9}$$

where $\phi_1, \phi_2, ..., \phi_p$ are constants and $e_t \sim N(0, \sigma_e^2)$. The MA($q$) process assumes that the observed value $z_t$ can be explained as a function of the past $q$ error terms, that is, the errors are combined linearly to form the observed data. In such cases, we write

$$z_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q}, \tag{2.10}$$

where $\theta_1, \theta_2, ..., \theta_q$ are parameters that determine the overall pattern of the process.

Identification of autocorrelation is necessarily inexact because many models that occur in practice depend on properties of the "real world" which cannot be determined purely by mathematics alone (Box and Jenkins, 1976). Graphical methods are employed to aid in identifying autocorrelation, particularly, plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF). The ACF and PACF is given, for example, in Shumway and Stoffer (2000).

Table 2.1: Behavior of the ACF and PACF plots for AR($p$) and MA($q$) models. These are common characteristics found in both the ACF and PACF plots and are used to indentifiy autocorrelation structure.

| | AR($p$) | MA($q$) |
|---|---|---|
| ACF | Tails off | Cuts off after lag $q$ |
| PACF | Cuts off after lag $p$ | Tails off |

Analysis of the ACF and PACF for FE estimates of $\beta_{0c}$ and $\beta_{1c}$ was conducted to determine a reasonable type of autocorrelation structure to use in the ME models. Some of the FE estimates were occasionally very large (negative or positive), although

poorly determined with large standard errors. More typical estimates would also fit the data well. This is not uncommon and often results from the correlation between slope and intercept parameter estimators; however, although such estimates do not result in anomalous estimates of maturities, they can have undue influence on the shapes of the ACF and PACF functions. A solution we propose for this problem is to use ME predictions of the slope and intercept random cohort effects assuming simple uncorrelated random errors to use in the ACF and PACF plots. These simple variance component (VC) estimates are reduced to the mean for all cohorts, with the amount of shrinkage depending on the estimated variance of the random effects. Anomalous estimates are much less likely compared to FE estimates, unless they explain a substantial amount of variance. We speculate that these VC predictions provide for more accurate diagnostics of the autocorrelation structure. However, validating this speculation is beyond the scope of our research.

### 2.2.5 Prediction and Forecast Accuracy

A cross-validation procedure was performed to measure prediction precision. A case (i.e. data for an age and cohort) was removed and then the number mature was predicted from the remaining data. This is classified as a leave-one-out cross-validation (Efron, 1983). The prediction accuracy was measured using Pearson's $\chi^2$ cross-validation statistic

$$X^2_{ac(-1)} = \sum_{a,c} \left( \frac{y_{ac} - \hat{\mu}_{ac(-1)}}{\sqrt{\hat{v}_{ac(-1)}}} \right)^2 . \tag{2.11}$$

This was compared for different models and also with the Pearson's $\chi^2$ goodness-of-fit statistic

$$X^2_{ac} = \sum_{a,c} \left( \frac{y_{ac} - \hat{\mu}_{ac}}{\sqrt{\hat{v}_{ac}}} \right)^2 . \tag{2.12}$$

We denote $E(y_{ac}) = \mu_{ac}$, $Var(Y_{ac}) = v_{ac}$, $\hat{\mu}_{ac} = n_{ac}\hat{p}_c(a)$, and $\hat{v}_{ac} = n_{ac}\hat{p}_c(a)1 - \hat{p}_c(a)$ as the estimate obtained using all of the data. The prediction of $\mu_{ac}$ and $v_{ac}$ obtained when that case was deleted is denoted as $\hat{\mu}_{ac(\ 1)}$ and $\hat{v}_{ac(\ 1)}$, respectively. These statistics were computed for all ages and cohorts.

We used a retrospective analysis to determine how accurately each model forecasted maturities. In the retrospective analysis, recent data were excluded from estimation, then maturities were predicted three years ahead and compared to the estimated logistic maturities obtained using all of the data. For example, if the retro year was 1997, maturities were predicted for 1998, 1999, and 2000; that is, data for 1998-2000 were not used to obtain the predictions. Each prediction was then compared to the corresponding estimated maturities using the data for 1998, 1999, and 2000. In the FE approach, predicted maturities were computed by averaging the three closest cohorts (Brattey et al., 2004). An averaging procedure was also used for recent unfinished cohorts that had insufficient data to estimate the maturity ogive. In the ME model approach, the correlation structure in the data was utilized to predict maturities. We examined retrospective performance for each year since 1995. A retrospective metric was used to quantify the retrospective error at each age:

$$\rho = \sum_{y=1995}^{Y} |\hat{p}_{a,y+3,y} - \hat{p}_{a,y+3,Y}|, \tag{2.13}$$

where $\hat{p}_{a,y+3,y}$ is the predicted proportion mature at age $a$ in year $y+3$ obtained using data up to retrospective year $y \leq Y$ and $Y$ is the last year in the full data set. This metric was computed for both the FE and ME models to measure their prediction accuracy.

### 2.2.6 Model Checking

It is important to check the validity of the linear binomial logistic model (see McCullagh and Nelder, 1989). Mis-specified assumptions, especially for the mean and variance, can result in poor estimates and predictions. GLMMs may be more sensitive to variance assumptions than GLIMs. Heagerty and Kurland (2001) demonstrated that large biases in regression parameter estimates can occur when random effects are misspecified. They recommended that careful attention be given to the random effects model assumptions when using GLMMs for regression inference with longitudinal data, such as our maturity data.

Residual plots are widely used as a way of checking systematic departures from assumptions in GLIMs. Lee and Nelder (1998) recommended using deviance residuals for checking model assumptions. To check the validity of the mean model, $E(y_{ac}) = n_{ac}p_c(a)$, we plot and examine standard deviance residuals versus age for each cohort to see if there are any patterns of variability not accounted for by our model. The binomial standard deviance residuals are defined as

$$r_{D_i} = \frac{sign(y_i - \mu_i)\sqrt{d_i}}{\sqrt{\phi(1 - h_i)}} \tag{2.14}$$

where $d_i$ is the contribution to the total deviance from observation $i$, $sign(y_i - \mu_i)$ is 1 if $y_i - \mu_i$ is postitve and -1 if $y_i - \mu_i$ is negative, and $h_i$ is the $ith$ diagonal element of the hat matrix $\mathbf{H}$ (see Dobson 2002; section 6.2.6).

We augment this by plotting observed and predicted proportions mature for each cohort to see if there are any large discrepancies. To check the validity of the variance model, absolute standard deviance residuals were plotted against fitted values. An incorrect variance function will often result in a trend in terms of the mean. Since PROC GLIMMIX uses a pseudo-likelihood approach it does not produce deviance residuals (which are based on maximum likelihood), therefore, for ME models we examined the mean and variance assumptions using $\chi^2$ residuals.

Over-dispersion is a naturally occurring phenomenon in binomial data (McCullagh and Nelder, 1989). Due to such factors as population spatial heterogeneity in maturities and the use of trawls to collect samples (i.e. cluster samples), it is important to test for over-dispersion. We used PROC GENMOD to fit a logistic model using separate slopes and intercepts for each cohort to estimate the over-dispersion parameter ($\phi$). We used PROC GLIMMIX to compute estimates for ME models with and without over-dispersion, and we compared both estimates in terms of their variability and predictive capability.

## 2.3 Results

### 2.3.1 Fixed Effects (FE) Model

The estimates $\hat{\beta}_{0c}$ and $\hat{\beta}_{1c}$ varied widely across cohorts (Fig. ' '; Table ' ) with little trend, but $\hat{A}_{50}$ was much less variable and declined over the 1954 to 2000 cohorts. $\widehat{MR}$ showed no long term trend but decreased for the 1963 cohort and increased for the 1980 cohort. The $\hat{p}_{ay}$ for $a = 4, ..., 8$ (top panel, Fig. ' ) increased over time, especially since 1990. Note that the flat lines at the beginning and end of the time series in the top panel of Fig. ' ' represent the average of estimates for adjacent cohorts used to hindcast and forecast maturities for cohorts with insufficient data for direct estimation.

The $\chi^2$ and $\chi^2_{-1}$ residuals (Fig. ' ) are based on the terms inside of the squares in equations ' ' and ' ' '. We focused on ages 4-8 because these were the ages that covered most of the dynamic range in the maturities. Outside this range, fish are usually either all immature (ages $\leq 3$) or all mature (ages $\geq 9$), and estimation is not controversial. The $\chi^2_{-1}$ residuals were very large for some ages and years; for example, age 6 in 1963. The total $\chi^2$ and $\chi^2_{-1}$ statistics (Table ' ') suggest that the predictive fit ($\chi^2_{-1}$) was much worse than the fit to all observations ($\chi^2$), although

this was mostly due to poor predictions for a small number of cases (Fig. ; bottom panel).

Retrospective analyses (Fig. ) showed a high degree of variability between forecasted and subsequently estimated maturities for ages 4-6. The variability was lower for older ages because the maturities at these ages have been close to one since 1995 and there is less scope for retrospective differences. We observed little to no retrospective error for $a \geq 10$ because essentially all fish in all cohorts were mature at these ages. Similarly, very few fish mature for $a \leq 4$ and consequently there is little scope for retrospective error at these ages. An extreme example of retrospective error was age 5 in 2005. In 2004, this maturity value was predicted (using the average value from the three previous years) to be 0.71. In 2005 this value was estimated to be 0.28. This substantial difference causes problems when estimating SSB (see Discussion).

Chi-square residuals (Fig. ) exhibited greater variability for some cohorts (e.g. the 1967 and 1973 cohorts). Also, some minor trends in residuals were apparent at younger ages for some cohorts (e.g. 1983). Large residuals occured when $\hat{p}_c(a)$ was close to one but a small number of immature fish were observed (Table ), However, $\hat{p}_c(a)$ were relatively close to the observed proportions for most cohorts (Fig. ).

In some years the deviance residuals mostly had the same sign (Fig. ), similar to the $\chi^2$ residuals (Fig. ). In six years 95% CI's for the mean residual did not cover zero which suggests some real year effects in the data. However, the evidence in not substantial because the CI's for these six years almost covered one, and no attempt was made to control the overall error rate of the multiple comparisons made in Fig. . A Bonferonni or similar adjustment would result in fewer significant year effects.

Absolute deviance residuals versus $n$ and $\hat{\mu} = n \times \hat{p}$ (Fig. ) showed an increasing trend. We expected that the average absolute residual would be approximately one, independent of $n$ or $\mu$, if the assumed mean and variance models were appropriate; however, 95% CI's for the trend in the residuals from a loess smoother did not

cover one for substantial ranges of $n$ or $\hat{\mu}$. Normally this would indicate the potential of model mis-specification; however, some preliminary simulations we conducted suggested this may not be the case. This point is considered further in the Discussion.

### 2.3.2 Autocorrelation Diagnostics

ACF and PACF plots for variance component (VC) model predictions of the random cohort effects for slopes and intercepts (Fig. ' i ) strongly suggested that the slopes and intercepts follow an AR(1) correlation structure. VC model estimates (Fig. ' ') were less variable than the corresponding FE estimates, although the results for $\hat{A}_{50}$ were very similar. ACF and PACF plots based on the FE model (Fig. ) did not seem reliable because they were heavily influenced by some anomalous estimates (see Fig. 2.12).

### 2.3.3 Mixed Effects (ME) Autoregressive (AR) Model

The $\hat{\beta}_{0c}$s (Fig. ') from ME AR models with or without overdispersion (OD) were much smoother than the FE estimates. The ME $\hat{\beta}_{0c}$s did not vary much between cohorts (Table ' .), especially for the AR OD model. For this model $\hat{\sigma}^2_{\delta_a}$ (Table ') was constrained at a lower bound we used for estimation. The $\hat{\beta}_{1c}$s increased over time, especially for the 1980 to 1990 cohorts. However, both ME model $\hat{A}_{50}$s were similar to the FE model estimates (Fig. ' '). The main differences between ME and FE models was in $MR$s, although the ME model estimates were similar with and without OD and showed a slight declining trend.

The ME model $\hat{p}_{ay}$s for $a = 4, ..., 8$ (Fig. ' ) varied more smoothly over time compared to the FE model estimates, and the AR OD model $\hat{p}_{ay}$s were smoother than those from the AR NOD model. However, the basic long-term trends in maturities from the ME AR and FE models were the same, showing an increase over time, especially since 1990.

The total $\chi^2$ fit statistic (Table . ,) was higher (i.e. worse fit) for the AR OD model compared to the AR NOD model, and the $\chi^2$ statistics for both ME models were higher than the FE model. In the ME models the variabilities of the cohort effects are constrained, and the amount of constraint depends on $\hat{\sigma}^2_{\delta_o}$ and $\hat{\sigma}^2_{\delta_1}$. The FE model parameters are not constrained, and this is why a better fit was obtained with this model. If $\sigma^2_{\delta_o}$ and $\sigma^2_{\delta_1}$ in the ME models were set at large values then these models would fit the data as well as the FE model. The AR NOD model fit better because $\hat{\sigma}^2_{\delta_o}$ and $\hat{\sigma}^2_{\delta_1}$ were larger than the AR OD model estimates (Table ). The AR OD model used the over-dispersion parameter $\phi$ to account for this extra lack-of-fit.

Fitting the data better does not mean a model predicts better. The reverse occurs for 3Ps cod (Table . '). The AR OD model had a smaller $\chi^2_1$ predictive fit statistic than the AR NOD model. The FE model had substantially poorer predictive performance than both ME models. This was also apparent in the $\chi^2_1$ residuals. For most ages and years the ME models (Figs. . ' and ',) predicted maturities better than the FE model (Fig. ',).

Similar to the FE model fit, there is evidence of year effects in Figs. ' ' , and ' , ,. In some years (i.e. follow diagonals in the figures) the residuals mostly had the same sign, especially for ages where maturities were not close to zero or one. This suggests some systematic variation exists in sampled maturities that cannot be accounted for by cohort effects. This is investigated more in the next section.

Retrospective results for the ME models (Fig. ' ) varied more smoothly over years compared to the FE results. Generally the forecast error, $\rho$, from the retrospective analysis for the AR OD model was smaller than the error for the AR NOD model, and both ME models had smaller forecast errors than the FE model. The relative performance of these models in terms of forecast error measured by $\rho$ was similar to the relative performance in terms of $\chi^2_{-1}$. However, substantial retrospective differences still occurred in some years.

The forecasts (and hindcasts) of maturities tend toward the overall average at

a rate that depends on the autocorrelation. If the autocorrelation is higher then the forecasted maturities change more smoothly. The AR OD forecasted maturities changed more slowly than the AR NOD results (Fig. ) because $\hat{\gamma}_o$ and $\hat{\gamma}_1$ for the AR OD model were greater than the estimates for the AR NOD model (Table ).

### 2.3.4 Mixed Effects Model with Random Year Effects (YE)

The ME model with autocorrelated logistic regression cohort effects ($\delta_{0c}$ and $\delta_{1c}$) and simple random year effects ($\eta_{0y}$) defined in equation , abbreviated as AR YE, fit the data better for ages 4-8 (Table ) and also had better predictive fit. This model had more potential parameters (i.e. $\eta_{0y}$s) than the FE model which explains why it could fit the data better. $\hat{\sigma}^2_\eta$ was much greater than $\hat{\sigma}^2_{\delta_o}$ and $\hat{\sigma}^2_{\delta_1}$ (Table ). The $\hat{A}_{50}$s from the AR YE model (Fig. ) were more different than the AR NOD, AR OD, or FE model estimates. These latter $\hat{A}_{50}$'s were usually very similar.

The AR YE model $\hat{A}_{50}$s and $\widehat{MIR}$s in Fig. did not directly include $\hat{\eta}_{0y}$s. The autoregressive year effects model parameters were treated like nuisance parameters that did not reflect real changes in population maturities, but rather sampling artifacts. The $\hat{A}_{50}$s and $\widehat{MIR}$s were based only on $\hat{\beta}_{0c}$ and $\hat{\beta}_{1c}$. We do not argue that $\eta_{0y}$s are nuisance parameters; however, if they are not then a single $A_{50}$ and MR cannot be identified for each cohort. We did not include them in Fig. to simplify the figure and make it directly comparable with Fig. .

The $\hat{\delta}_{0c}$s were all close to zero (Fig. ), which suggested that these effects were not significant. The small value and large standard error for $\hat{\sigma}^2_{\delta_0}$ in Table also indicated that $\hat{\delta}_{0c}$s were not significant. The $\hat{\delta}_{0c}$s from the AR OD model also did not seem significant. The $\hat{\gamma}_0$s were not significantly different from zero for both models (Table ). The $\hat{\delta}_{1c}$s were similar for the AR YE and AR OD models (Fig. ). Their $\hat{\sigma}^2_{\delta_1}$s (Table ) were not significantly different from zero; however, unlike $\hat{\gamma}_0$, $\hat{\gamma}_1$s for the AR OD and AR YE models were significant. The smooth trends in $\hat{\delta}_{1c}$ in Fig. also suggested the trend was real. AR YE model $\hat{\eta}_{0y}$s were significantly

different from zero in many years. The significant $\hat{\sigma}_\eta^2$ (Table ?.?) demonstrated the significant magnitude of YEs in the data.

The AR YE model $\hat{p}_{ay}$s (Fig. ?.?.) were very different depending on whether the YEs were treated as nuisance parameters (AR YE-) or not (AR YE+). In the latter case the estimates were more variable and similar to the AR NOD or FE model estimates, whereas in the first case the estimates were more similar to the AR OD estimates (Fig. ). The similarity between $\hat{p}_{ay}$s from the AR YE and FE models was also evident in the similarity of the residuals (Fig. ?.?) and cohort ogive comparisons (Fig. ?.?), although the AR YE model fitted the data better. Some of the improvement in fit was apparent in the $\chi^2$ at ages 4-8 (top panel, Fig. ?.?) compared to the FE results (Fig. ?.?). The improvement in predictive fit (bottom panels) was quite clear.

The improvement in fit is better illustrated in Fig. ?.?, which contains residuals for all observed maturities and not just ages 4-8. The AR YE model produced smaller residuals in 254 of 492 cases. More importantly, the AR YE model produced smaller residuals in 17 of 23 cases for which the FE model residuals were larger than two in absolute value. However, the total fit for the AR YE model, $\chi^2 = 841.48$, was worse than the FE model, $\chi^2 = 722.61$. This was due to a small number of observations. The per observation average $\chi^2$ was 1.71 and 1.47 for the AR YE and FE models, but the 1% trimmed means were 0.61 and 0.78. This indicates that for the large majority of the observations the AR YE model fit better than the FE model.

Confidence intervals for the average annual $\chi^2$ residuals all covered zero (Fig. ?.??). This was not the case for the FE model (Fig. ?.?). Note that there are five large residuals, $\lessapprox \pm 5$, in Fig. ?.?? but only four in Fig. ?.?? and in Table ?.?. This was because Fig. ?.?? was based on observations for all cohorts whereas Fig. ?.? and Table ?.? were base only on the 1954-2000 cohorts. Recall that other cohorts could not be reliably estimated with the FE model. Also, for seven other cases SAS PROC GENMOD software did not produce $\chi^2$ residuals for the FE model when the estimated

maturities were too close to 100% and not all fish were mature. This problem did not occur in the AR YE model. These cases were not included in Fig. but they were in Fig. . The additional cases in Fig. contributed one additional large residual which was truncated to display in this figure. The total AR YE model $\chi^2$ statistics reported in the previous paragraph did not include these seven cases or the cohorts prior to 1954 of after 2000.

The retrospective results varied more smoothly for the AR YE model when YEs were treated as nuisance parameters (AR YE-) than for the other ME models or the FE model (Fig. ). When YEs were included as predicted parameters (AR YE+) the retrospective results were more variable, but still usually better than the FE and AR NOD models.

## 2.4 Discussion

The overall results suggest that the GLMM is a more appropriate choice for modelling maturation rates in Atlantic cod stocks compared to the GLIM. The GLMM improved estimates of maturities, produced smaller residuals, and fit the data better for a large majority of observations (e.g. AR YE model). Furthermore, smaller retrospective metrics and cross-validation statistics suggests that prediction and forecast accuracies were also improved by the GLMM.

The usual approach of fitting maturity ogives seperately for each cohort using logistic regression can lead to large discrepancies in short term predictions of unfinished cohorts. For example, in 2004 the prediction of the estimate of the proportion mature at age 5 in 2005 was 0.70. In 2005 the revised estimate of this proportion mature was 0.30 (see Fig. , panel 2). This is a very large reduction. This age substantially contributes to the total biomass in 2004 (Brattey et. al., 2004). Hence, the change in estimated maturities caused by adding one more year of data is a major source of retrospective differences in SSB estimates between stock assessments in 2004 and

2005. Another source is retrospective difference in estimates of abundance produced by the stock assessment model. A description of this problem is beyond the scope of our work, but a recent reference dealing with this problem is Cadigan and Farrell (2004).

We also examined variability in terms of the binomial sample sizes. The rationale for this is that the sampling scheme used to collect the maturity data is complex and involves length-stratified cluster sampling, with post-sampling adjustments to account for the length distribution of the populations sampled. Explicitly dealing with this complex sampling scheme is beyond the scope of this paper; however, fish are not sampled completely at random and the effective sample size (see Kish, 1995) will be smaller than the total number sampled, at least for ages and cohorts that had large sample sizes. The smaller samples sizes will tend to represent clusters of size one, in which case the effective sample size is the same as the total number sampled. If the nature of the sampling scheme is important then we expect the binomial variance assumption to be incorrect for larger values of $n$, and result in deviance residuals with different variability than expected.

We use nonparametric regression methods, in particular local linear regression, to help identify trends in residuals. Alternatively, for variance diagnostics we also binned residuals and computed the average of the absolute standardized deviance residuals in each bin. Each bin had approximately 30 residuals (see Fig. ). However, some trends may be consistent with the binomial variance model because the distribution of the deviance residuals is not always approximately normal (Pierce and Schafer, 1986). We use a parametric bootstrap procedure (see Davidson and Hinkley, 1997) to examine if trends in residuals are inconsistent with the binomial variance model (Fig. ). We notice that an increasing trend does exist which is consistent with the binned residual analysis (Fig. )

We prefer PROC GLIMMIX for modeling more complicated random effects such as those with autocorrelation, compared to marginal approaches (i.e. PROC NLMIXED).

This is because that in a marginal approach, random effects are integrated out of the likelihood function, and this typically involves numerical integration. If complicated and/or numerous random effects exist, the numerical integrations may become complicated or computationally prohibitive.

Table 2.2: Summary statistics (over cohorts) of fixed effects model estimates and mixed autoregressive (AR) effects model predictions of intercepts, slopes, $\hat{A}_{50}$'s and $\hat{MR}$'s for 3Ps female cod.

| Model | Parameters | Summary | | | |
|---|---|---|---|---|---|
| | | Min | Max | Mean | Variance |
| Fixed effects (FE) | Intercept | -135.21 | -6.50 | -13.85 | 333.49 |
| | Slope | 1.03 | 22.57 | 2.33 | 9.33 |
| | $\hat{A}_{50}$ | 4.68 | 7.36 | 6.01 | 0.45 |
| | $\hat{MR}$ | 0.09 | 2.12 | 1.21 | 0.12 |
| Mixed AR effects without overdispersion (AR NOD) | Intercept | -11.30 | -9.87 | -10.56 | 0.13 |
| | Slope | 1.42 | 2.08 | 1.77 | 0.03 |
| | $\hat{A}_{50}$ | 4.76 | 7.42 | 6.02 | 0.45 |
| | $\hat{MR}$ | 1.06 | 1.54 | 1.23 | 0.02 |
| Mixed AR effects with overdispersion (AR OD) | Intercept | -10.44 | -10.38 | -10.40 | <0.01 |
| | Slope | 1.44 | 2.11 | 1.75 | 0.04 |
| | $\hat{A}_{50}$ | 4.93 | 7.23 | 6.03 | 0.42 |
| | $\hat{MR}$ | 1.04 | 1.53 | 1.27 | 0.02 |
| Mixed variance components effects (VC) | Intercept | -11.54 | -9.02 | -10.57 | 0.39 |
| | Slope | 1.41 | 2.09 | 1.77 | 0.03 |
| | $\hat{A}_{50}$ | 4.71 | 7.52 | 6.03 | 0.44 |
| | $\hat{MR}$ | 1.05 | 1.56 | 1.25 | 0.01 |
| Mixed AR effects with year effects (AR YE) | Intercept | -12.31 | -9.80 | -10.89 | 0.23 |
| | Slope | 1.44 | 2.18 | 1.77 | 0.04 |
| | $\hat{A}_{50}$ | 4.52 | 8.21 | 6.24 | 0.63 |
| | $\hat{MR}$ | 1.00 | 1.52 | 1.26 | 0.02 |

Table 2.3: Pearsons total $\chi^2$ statistics for 3Ps female cod, ages 4-8. Models described in Table 2.2.

| Model | Atlantic cod | |
|---|---|---|
| | $\chi^2$ | $\chi^2_{-1}$ |
| Fixed Effects | 401.97 | 61341.44 |
| Mixed effects AR model (no overdispersion) | 418.73 | 767.21 |
| Mixed effects AR model (overdispersion) | 465.18 | 662.40 |
| Mixed effects AR model with year effects | 274.09 | 642.96 |

Table 2.4: Mixed effects covariance parameter estimates (Est) with standard errors (S.E.), for 3Ps female cod maturities. Models described in Table [illegible].

| Parameter | AR NOD | | AR OD | | AR YE | |
|---|---|---|---|---|---|---|
| | Est. | S.E | Est. | S.E | Est. | S.E |
| $\hat{\sigma}^2_{\delta_0}$ | 0.2730 | 0.2009 | 0.0100 | - | 0.0265 | 0.0741 |
| $\hat{\sigma}^2_{\delta_1}$ | 0.0695 | 0.0745 | 0.0670 | 0.0746 | 0.0986 | 0.1288 |
| $\hat{\gamma}_0$ | 0.7936 | 0.1396 | 0.8187 | 0.7773 | 0.4541 | 1.068 |
| $\hat{\gamma}_1$ | 0.9659 | 0.03904 | 0.9702 | 0.0348 | 0.9846 | 0.0211 |
| $\hat{\phi}$ | - | - | 3.0216 | 0.1915 | - | - |
| $\hat{\sigma}^2_{\eta}$ | - | - | - | - | 0.3834 | 0.1188 |

Table 2.5: Large ($\lesseqgtr \pm 5$) $\chi^2$ residuals (Res) from the fixed effects model (FE) and the autocorrelated model with year effects (AR YE) for 3Ps cod. $\hat{p}$ is the estimated (or predicted) proportion mature.

| FE | | | | AR YE | | | |
|---|---|---|---|---|---|---|---|
| Cohort | Age | Res | $\hat{p}$ | Cohort | Age | Res | $\hat{p}$ |
| 1967 | 9 | -6.41 | >0.99 | 1966 | 9 | -4.19 | 0.99 |
| 1970 | 9 | -5.00 | 0.99 | 1967 | 9 | -6.40 | >0.99 |
| 1973 | 10 | -4.44 | 0.99 | 1970 | 9 | -4.56 | 0.99 |
| 1976 | 12 | -14.35 | >0.99 | 1973 | 10 | -13.88 | >0.99 |
| 1987 | 8 | -6.02 | >0.99 | 1976 | 12 | -16.45 | >0.99 |
| 1992 | 8 | -7.86 | >0.99 | 1996 | 8 | -6.41 | >0.99 |
| 1998 | 7 | -4.96 | >0.99 | | | | |

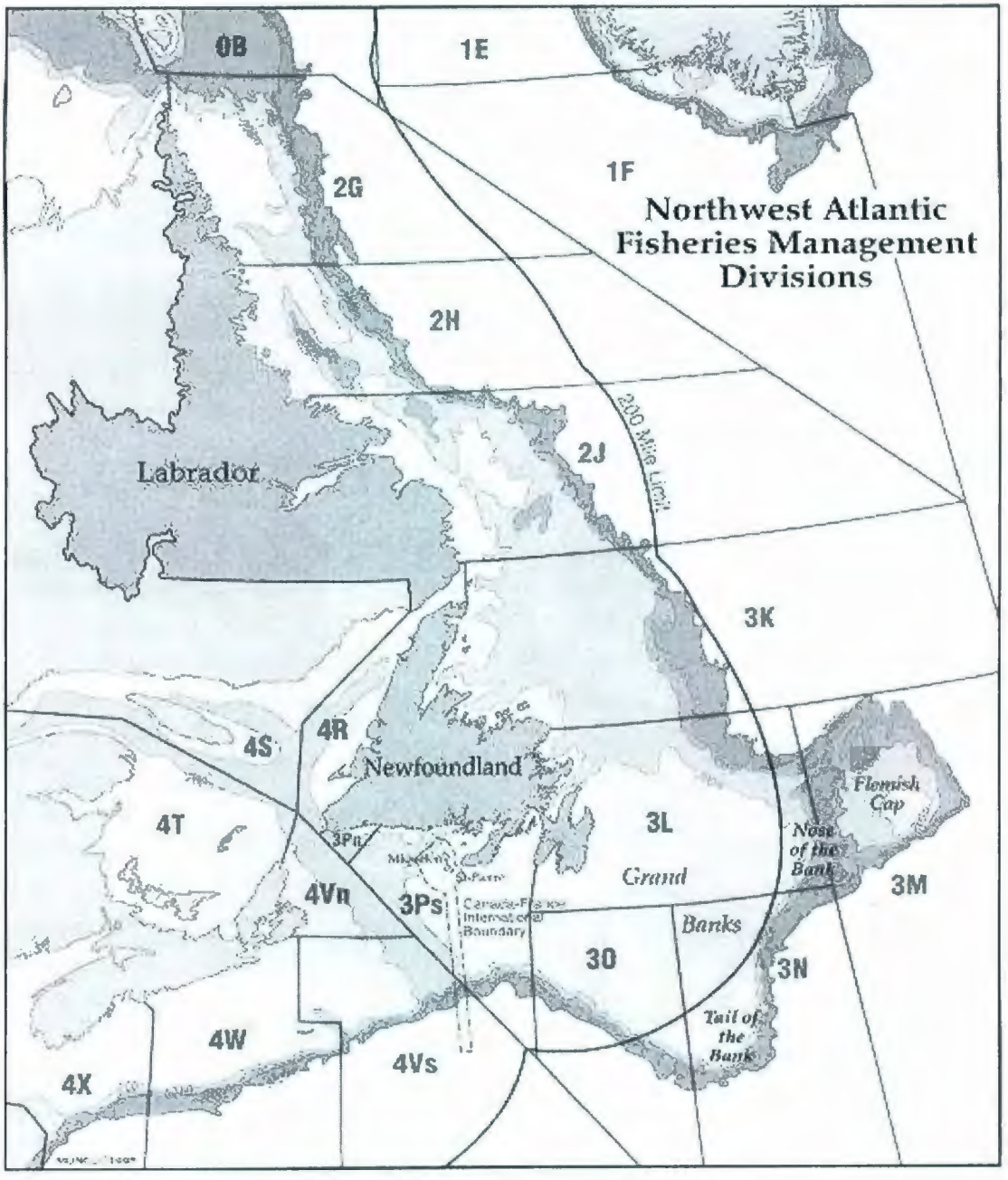


Figure 2.1: Northwest Atlantic Fisheries Organization (NAFO) fisheries management divisions.

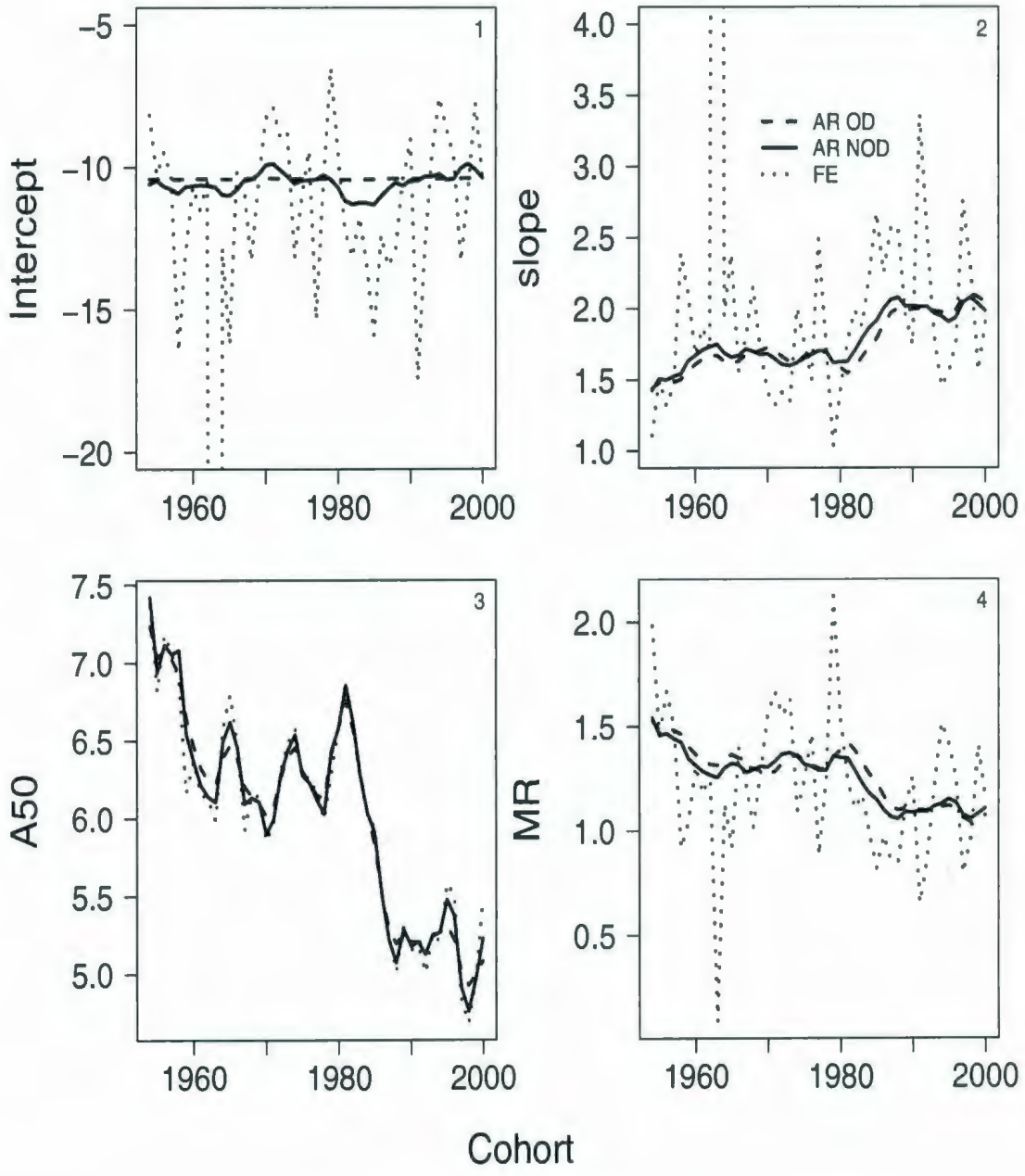


Figure 2.2: Estimates for 3Ps cod. Panel 1: intercepts. Panel 2: slopes. Panel 3: $\hat{A}_{50}$. Panel 4: $\hat{MR}$. AR NOD is the autoregressive (AR) mixed-effects model with no overdispersion (NOD), and AR OD has overdispersion. FE is the fixed-effects model.

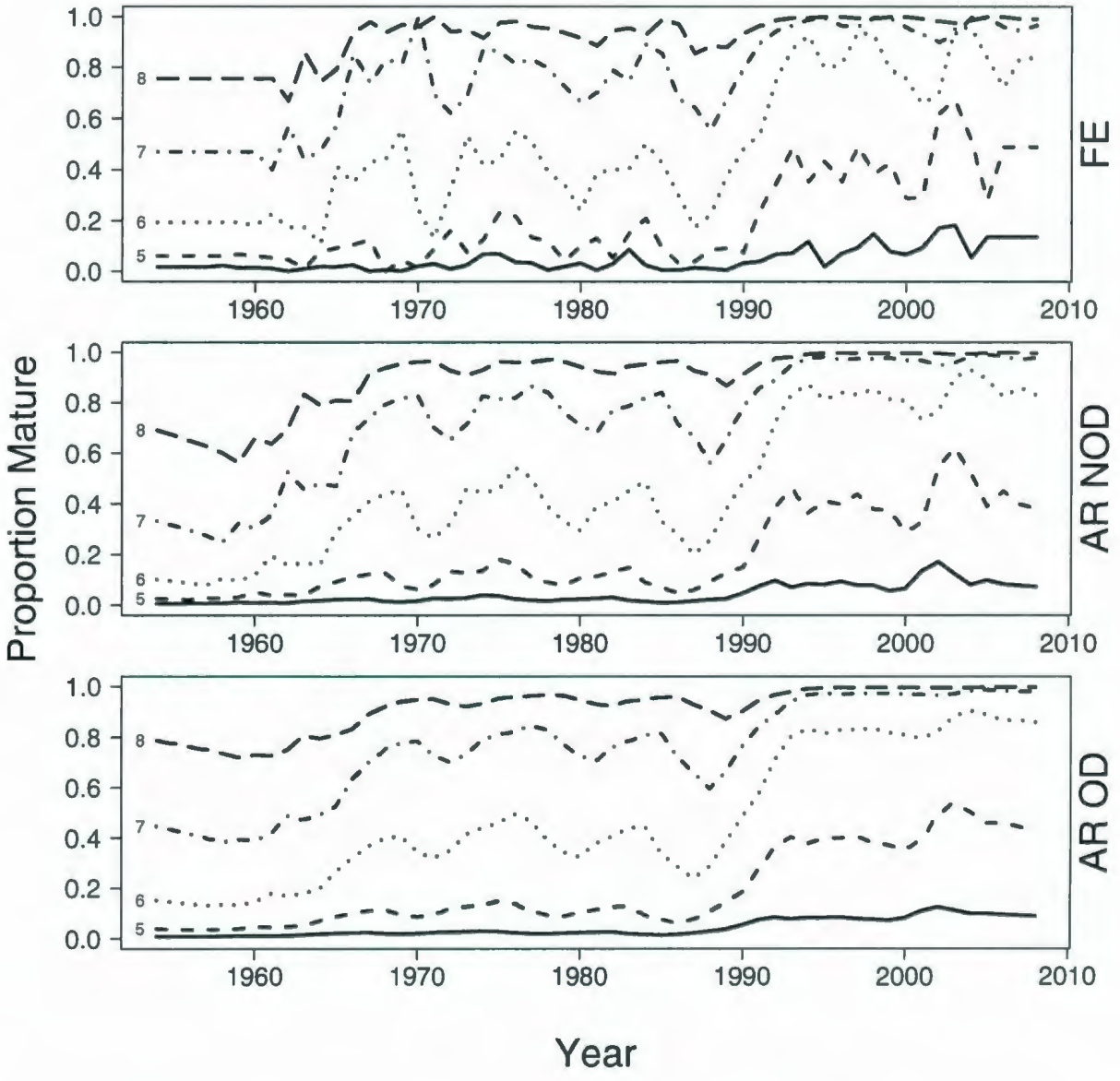


Figure 2.3: 3Ps cod proportions mature at ages 4-8 vs. year. Ages 5-8 are listed at the left-hand side. Top panel: Fixed-effects (FE) model. Middle panel: autoregressive (AR) mixed-effects model with no overdispersion (NOD). Bottom panel: AR model with overdispersion (OD).

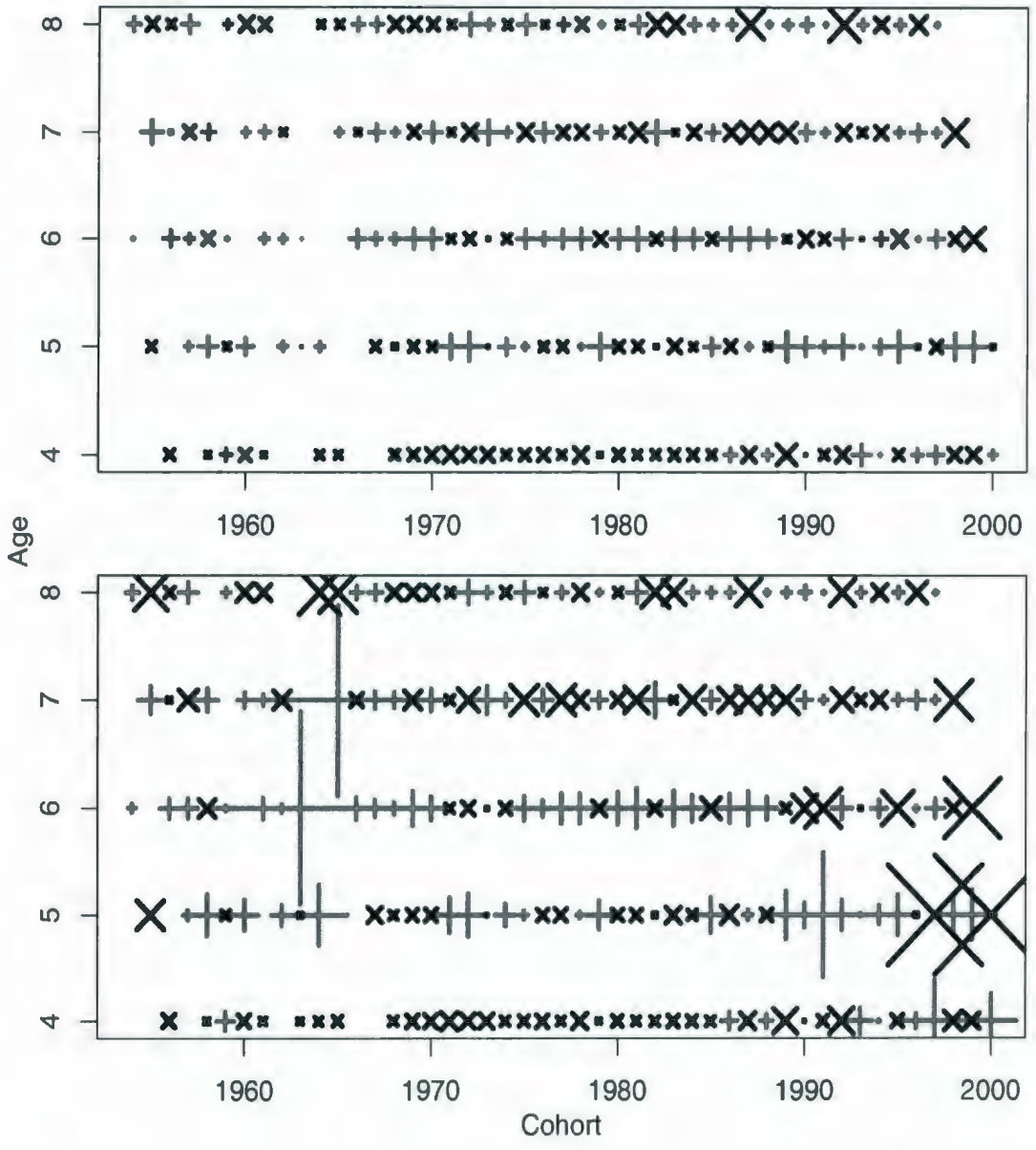


Figure 2.4: Residuals from the fixed effects (FE) model for 3Ps cod, + values are positive and × values are negative. Size is proportional to the absolute residual. Top panel: Chi-square ($\chi^2$) residuals. Bottom panel: Cross-validation chi-square ($\chi^2_{-1}$) residuals.

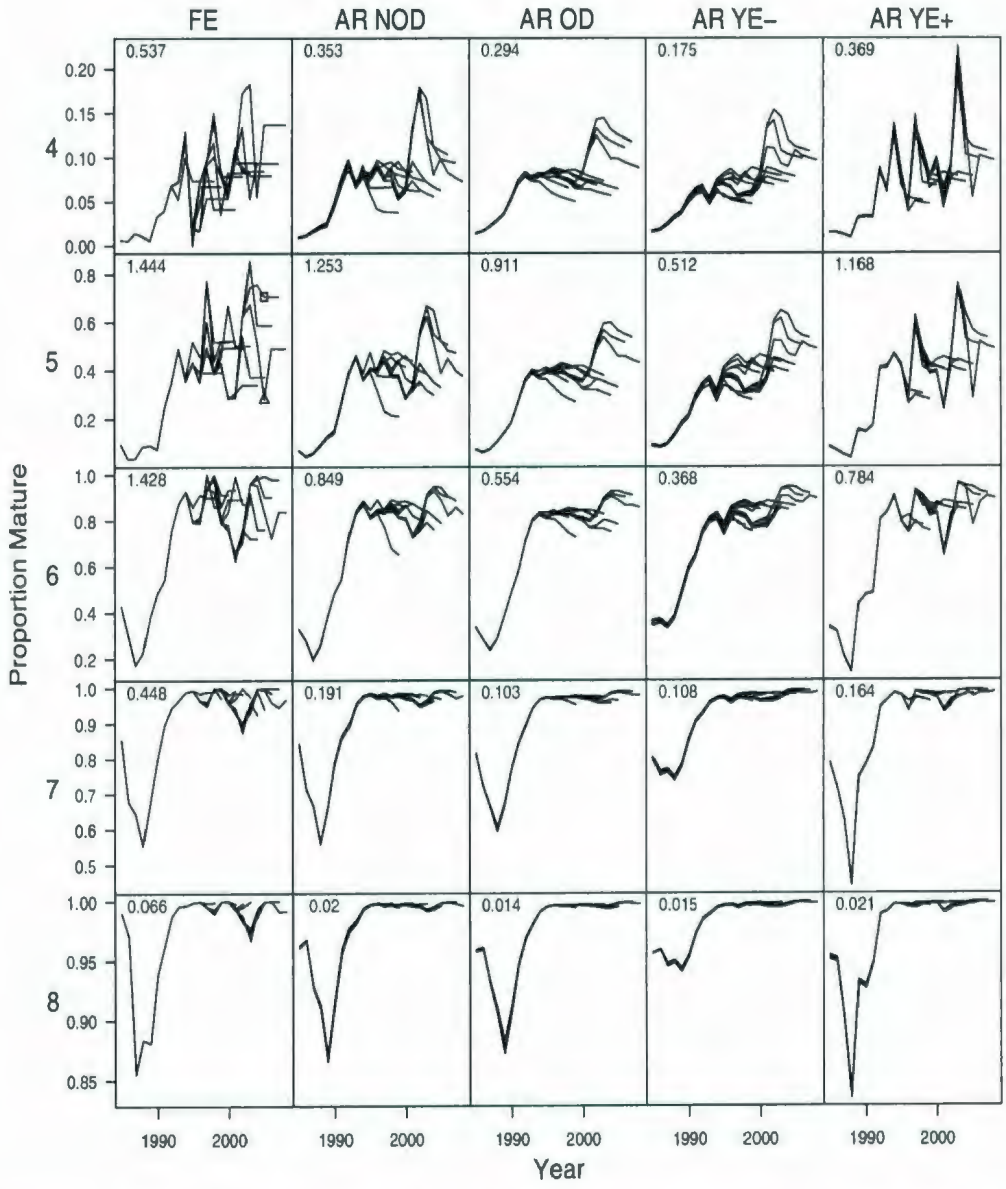


Figure 2.5: Retrospective analysis for 3Ps cod, ages 4-8 (listed in left margin). The retrospective $\rho$ metric is shown in the top left-hand corner of each panel. Column 1: Fixed-effects (FE) model. Column 2: autoregressive (AR) mixed-effects model with no overdispersion (NOD). Column 3: AR model with overdispersion (OD). Column 4. AR model with year-effects as nuisance parameters (YE-). Column 5. AR model with year-effects as predictive parameters (YE+). The maturity at age 5 predicted from the FE model using data up to 2004 is shown as a square, and the value estimated in 2005 is shown as a triangle.

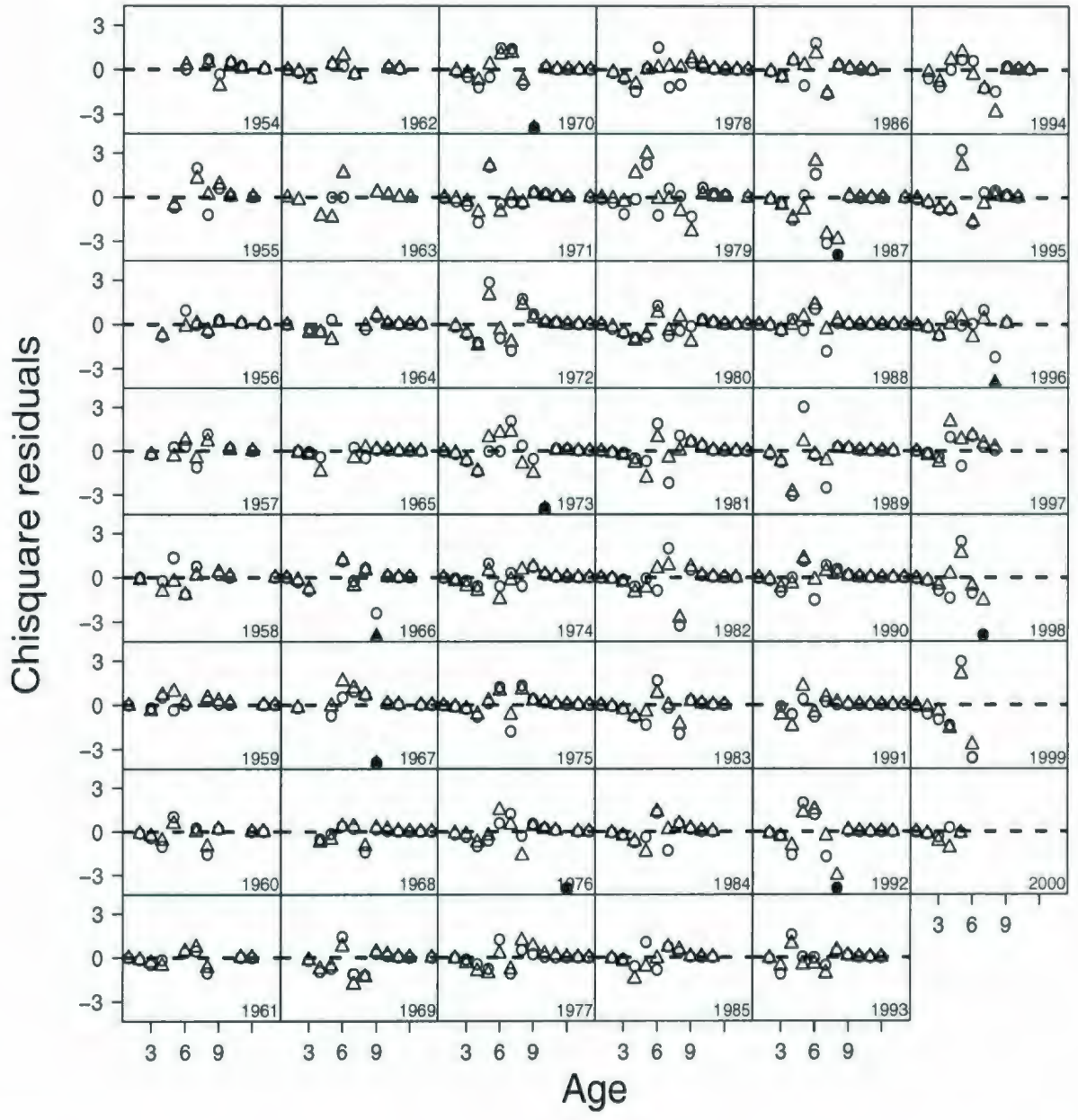


Figure 2.6: Chi-square ($\chi^2$) residuals by age and cohort for 3Ps cod. Fixed effects (FE) residuals are plotted are circles (○), and residuals from the autocorrelated model with year effects (AR YE) are plotted as triangles (△). Solid symbols are truncated ($\leq\geq \pm 4$), and are presented in Table 2.5.

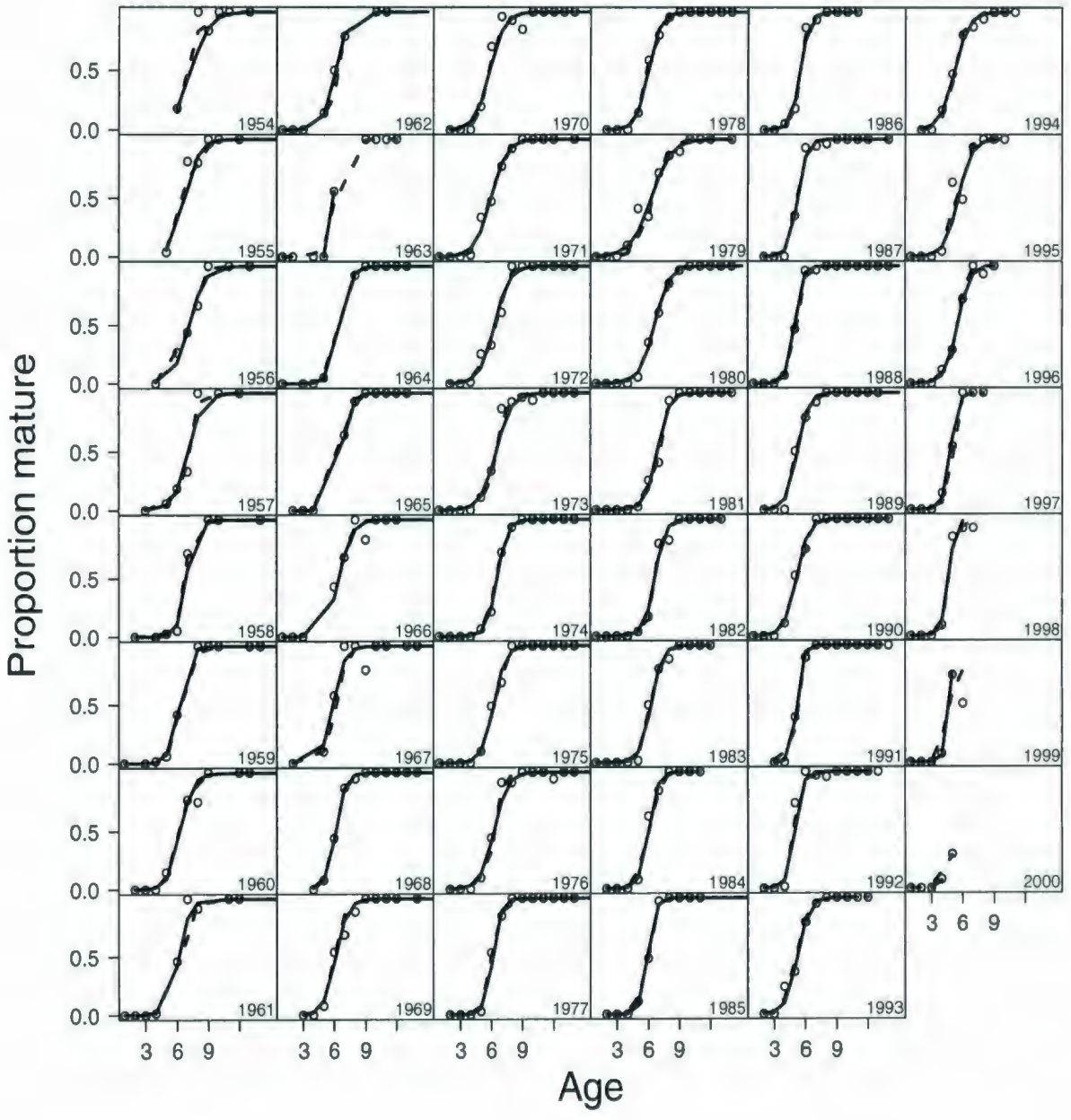


Figure 2.7: 3Ps cod proportions mature at age estimated from the fixed effects model (FE; solid lines) and the autocorrelated model with year effects (AR YE; dashed lines). Observations are plotted as circles (○).

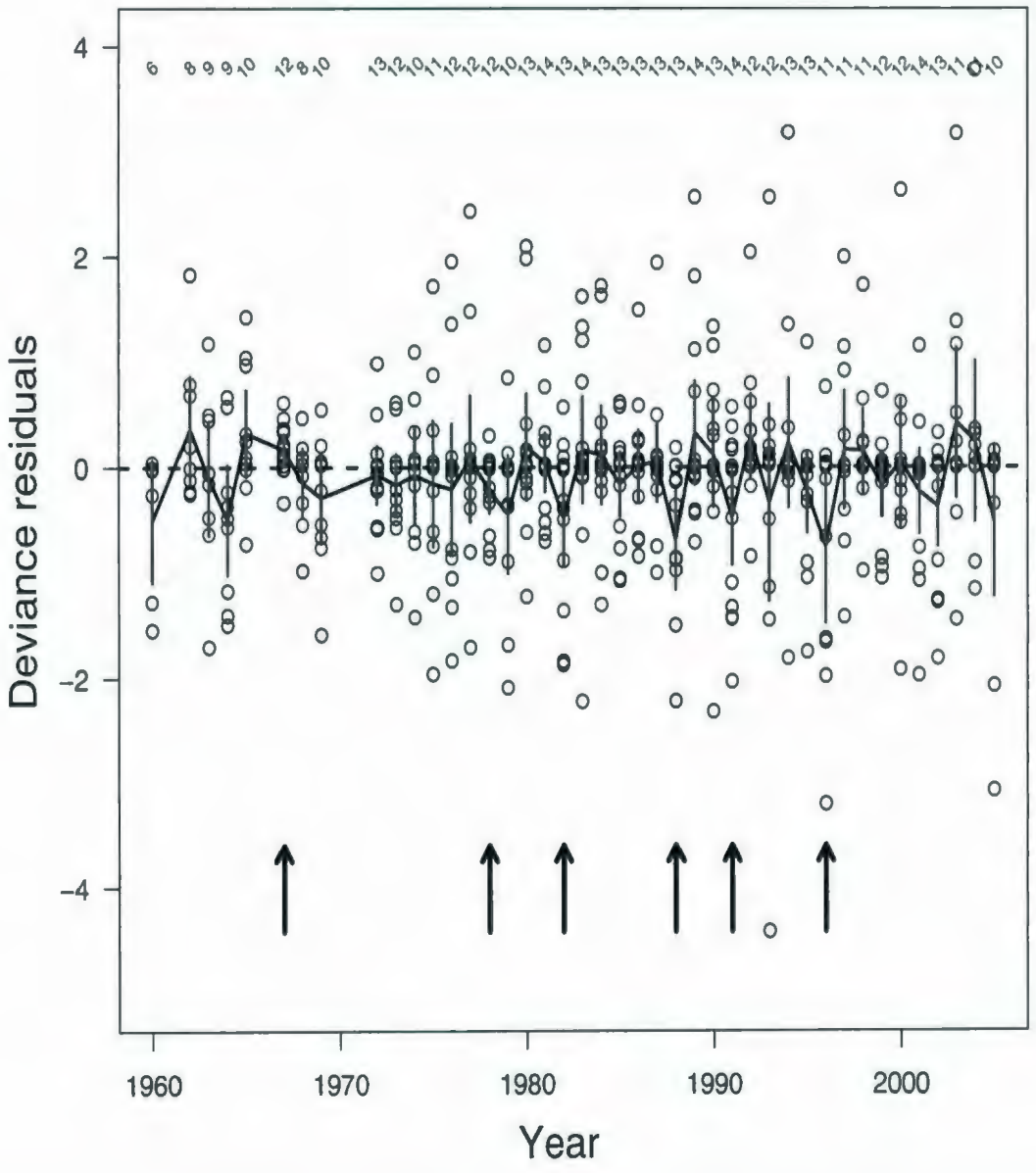


Figure 2.8: Average annual deviance residuals (solid lines) from the fixed effects (FE) model for 3Ps cod. Vertical lines demark 95% confidence intervals (CI's). Residuals are plotted as circles (○). Arrows denote years for which the CI's do not cover zero. Numbers of residuals are listed at the top.

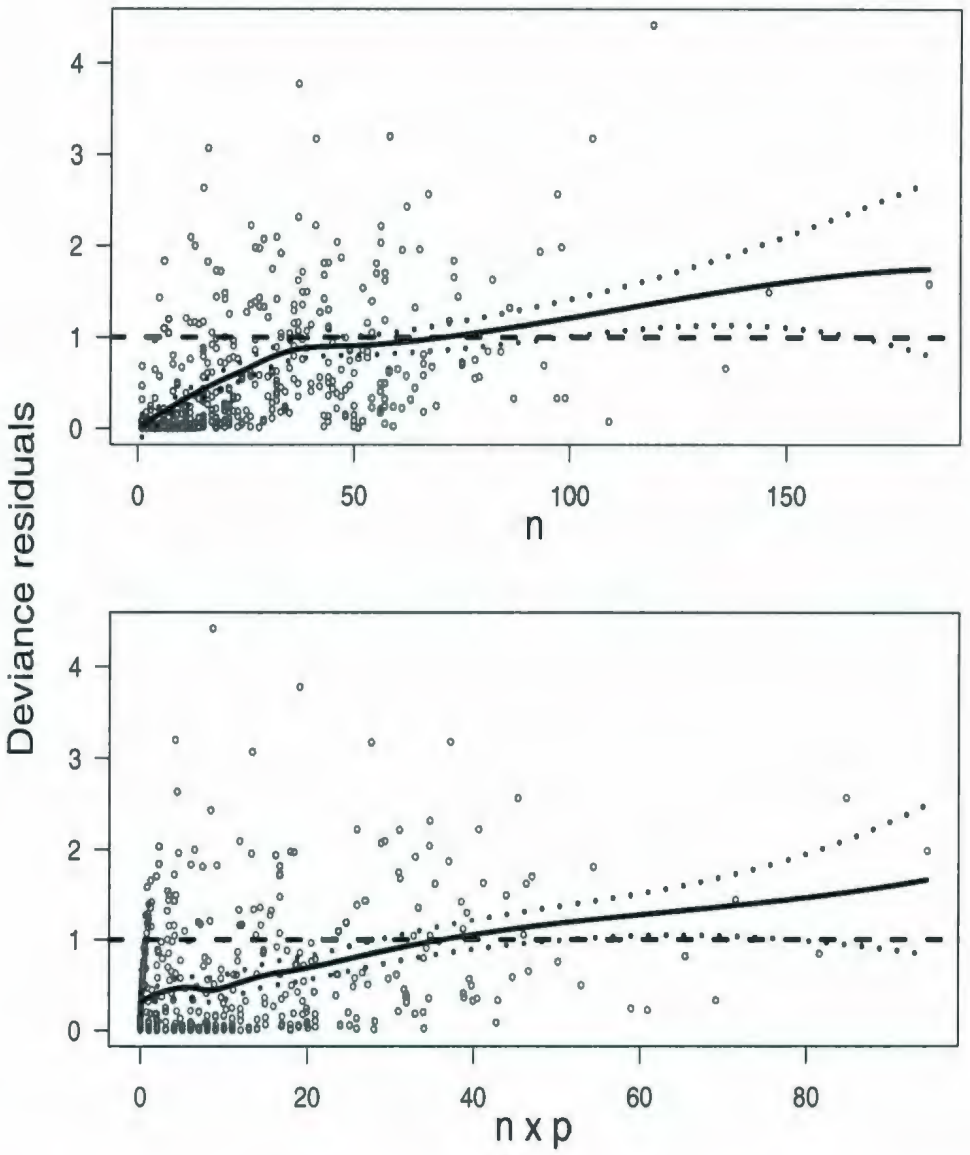


Figure 2.9: Absolute values of deviance residuals (o's) for 3Ps cod from the fixed effects (FE) model vs. $n$ (panel 1) and $\hat{\mu} = n \times \hat{p}$ (panel 2). The solid line is the fit from a loess smoother, and the dotted lines represent 95% confidence limits for the smoother. The dashed line is a reference line at 1 and represents the approximate expected value of the absolute residuals.

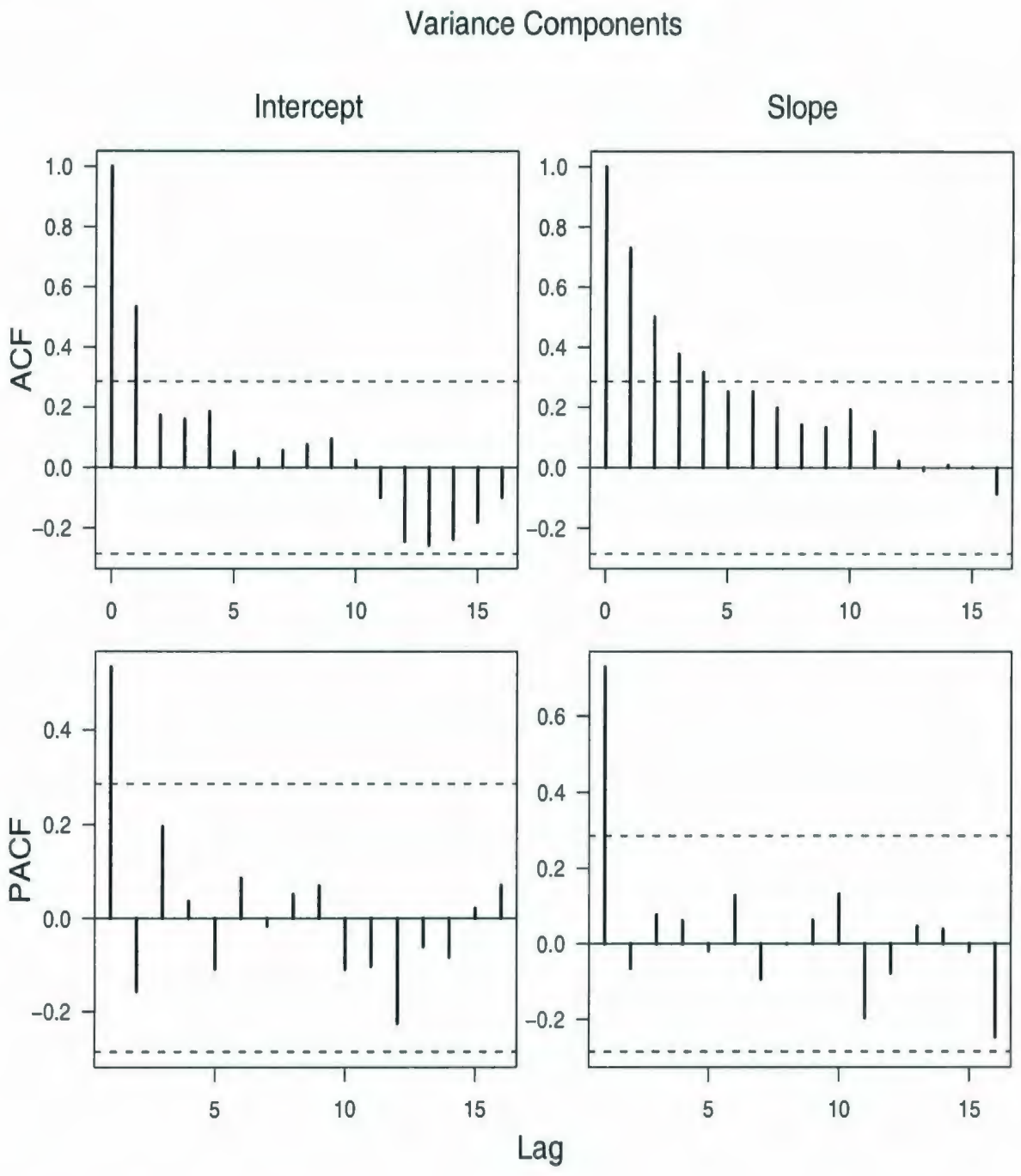


Figure 2.10: 3Ps cod autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of the intercepts and slopes from the variance components (VC) model.

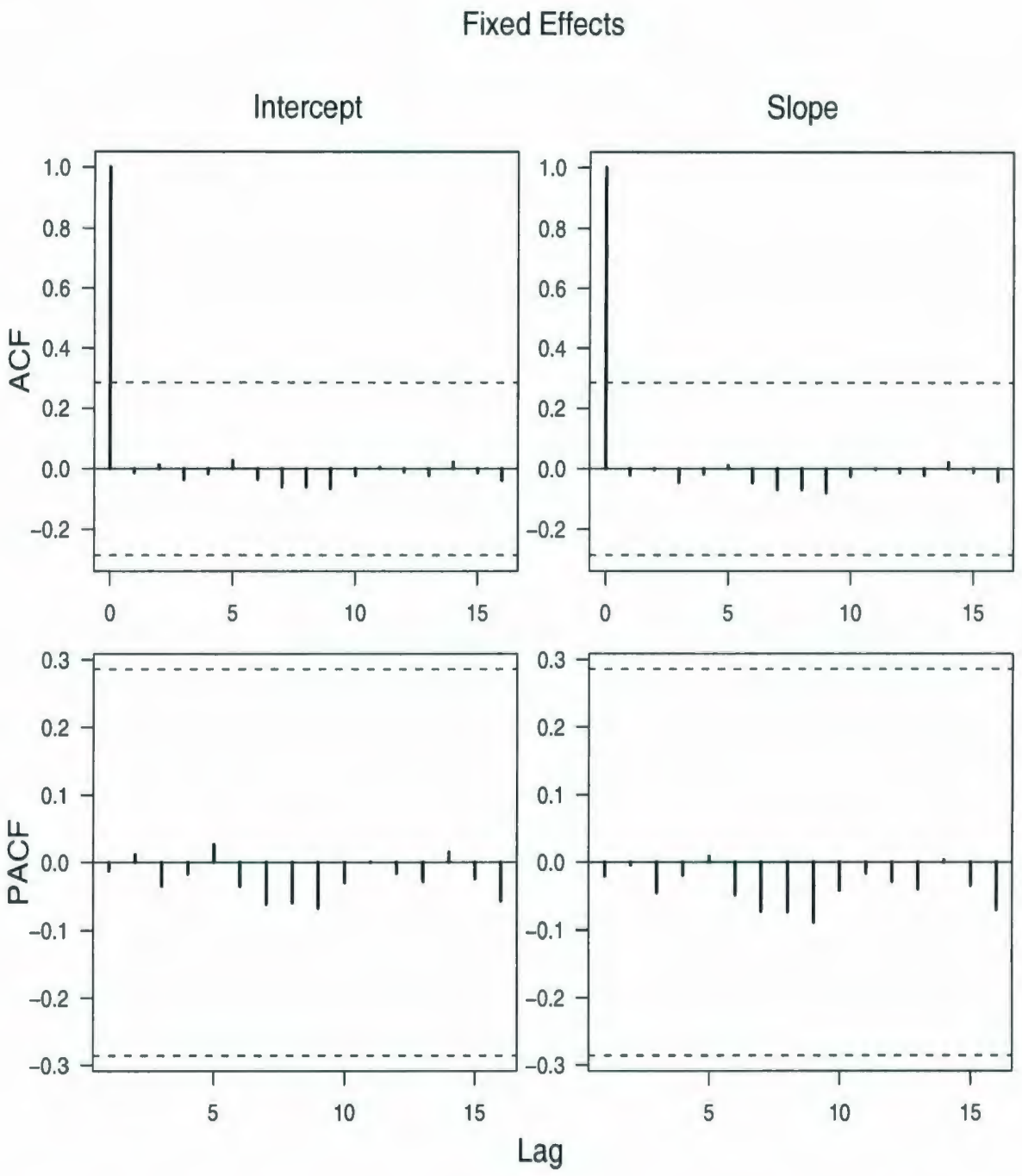


Figure 2.11: 3Ps cod autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of the intercepts and slopes from the fixed-effects (FE) model.

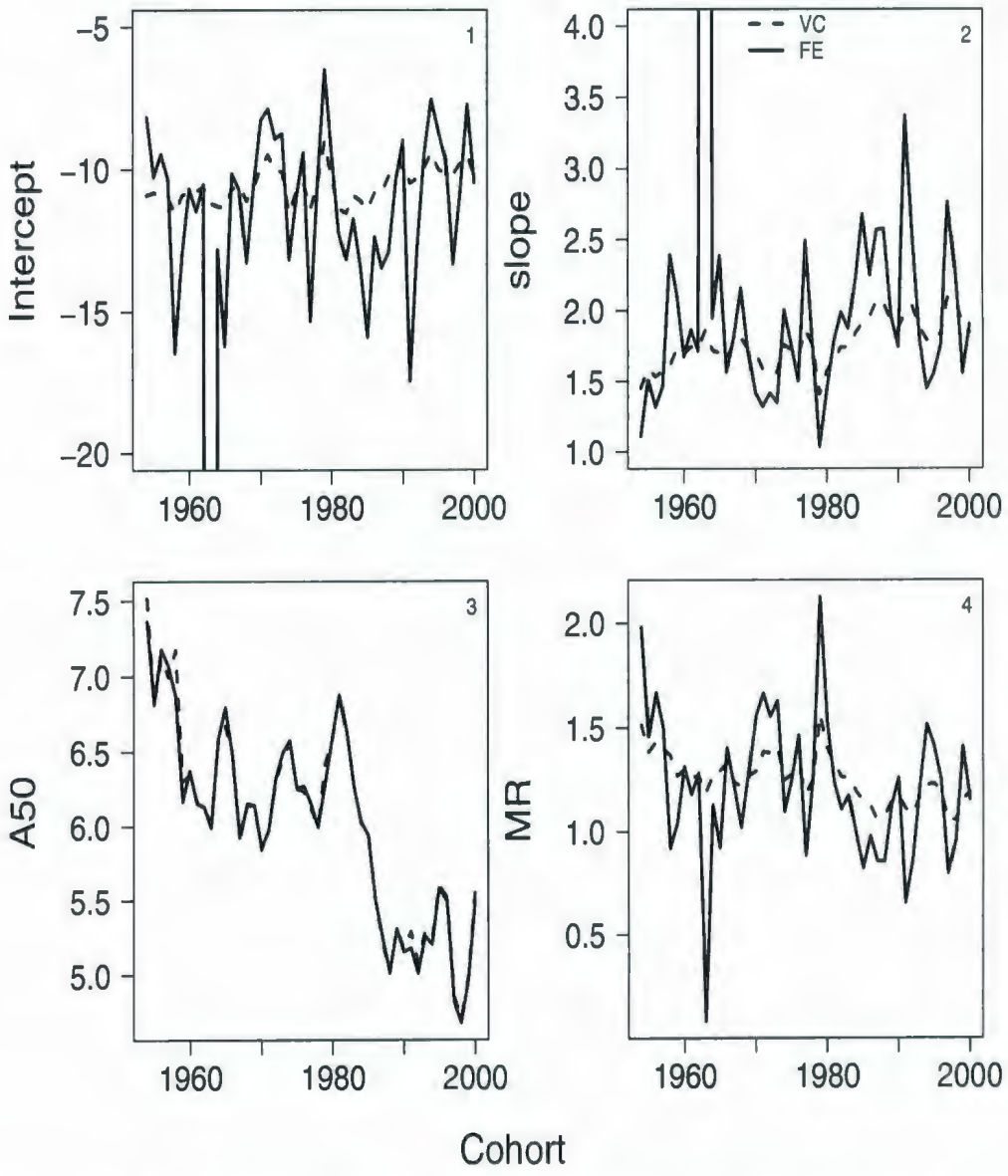


Figure 2.12: Estimates for 3Ps cod. Panel 1: intercepts. Panel 2: slopes. Panel 3: Age at 50% maturity, $\hat{A}_{50}$. Panel 4: Maturity range, $\hat{MR}$. VC is the mixed-effects variance components model and FE is the fixed-effects model.

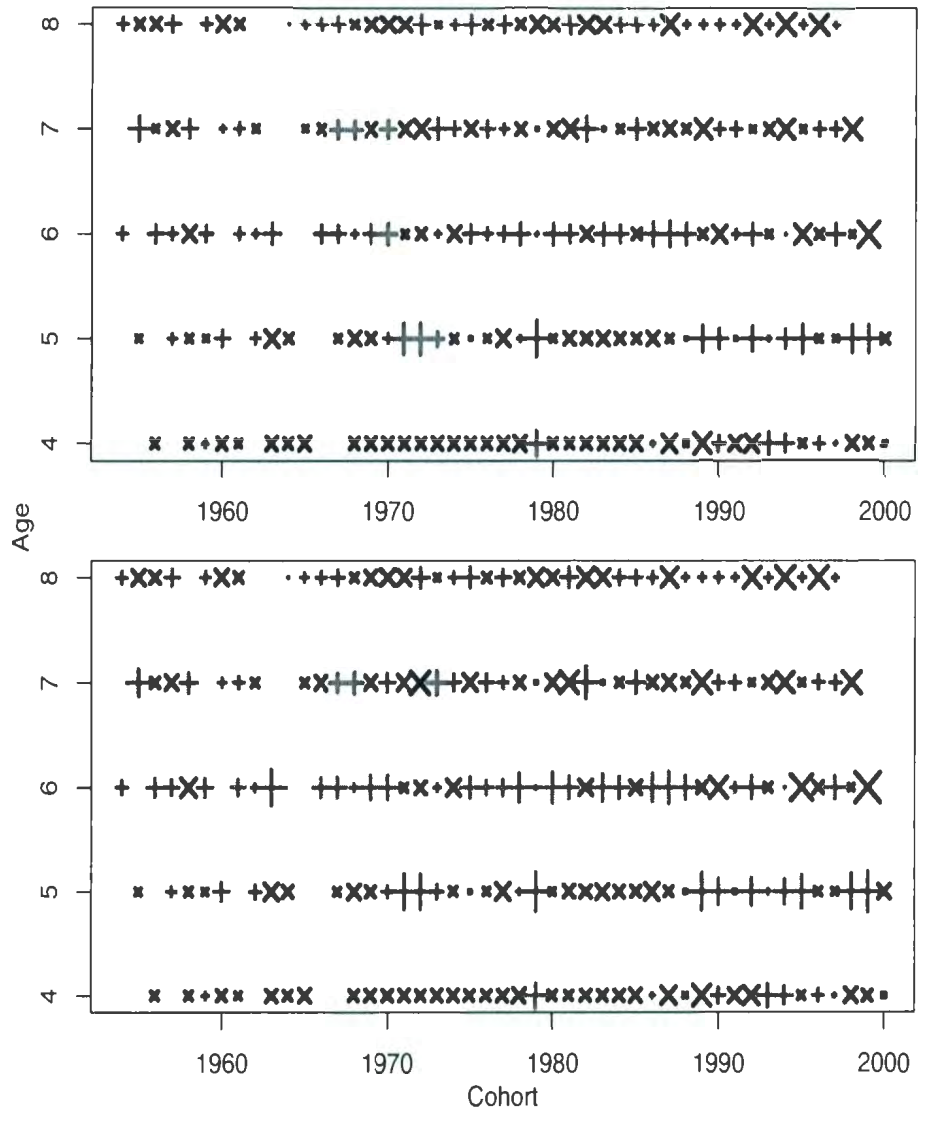


Figure 2.13: Residuals from the autoregressive mixed-effects model with no overdispersion (AR NOD) for 3Ps cod, + values are positive and × values are negative. Size is proportional to the absolute residual. Top panel: Chi-square ($\chi^2$) residuals. Bottom panel: Cross-validation chi-square ($\chi^2_{-1}$) residuals.

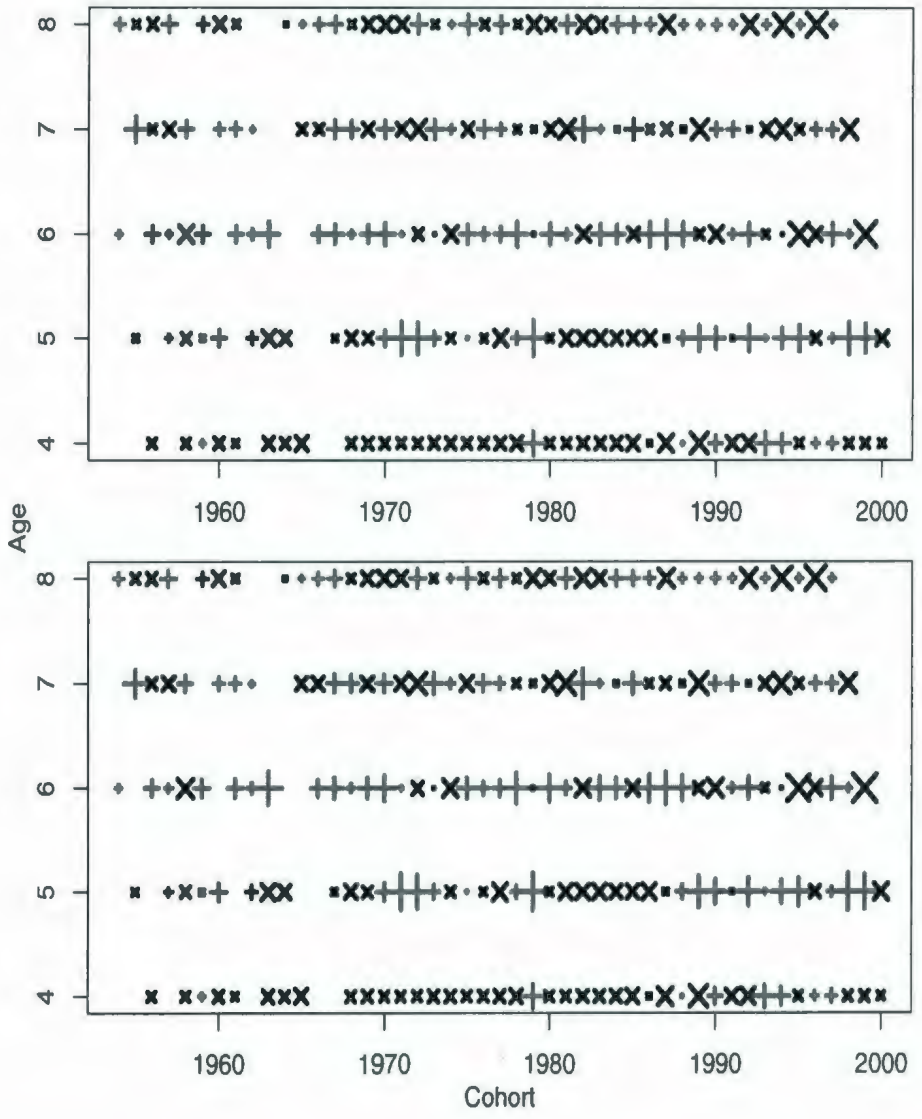


Figure 2.14: Residuals from the autoregressive mixed-effects model with overdispersion (AR OD) for 3Ps cod, + values are positive and × values are negative. Size is proportional to the absolute residual. Top panel: Chi-square ($\chi^2$) residuals. Bottom panel: Cross-validation chi-square ($\chi^2_{-1}$) residuals.

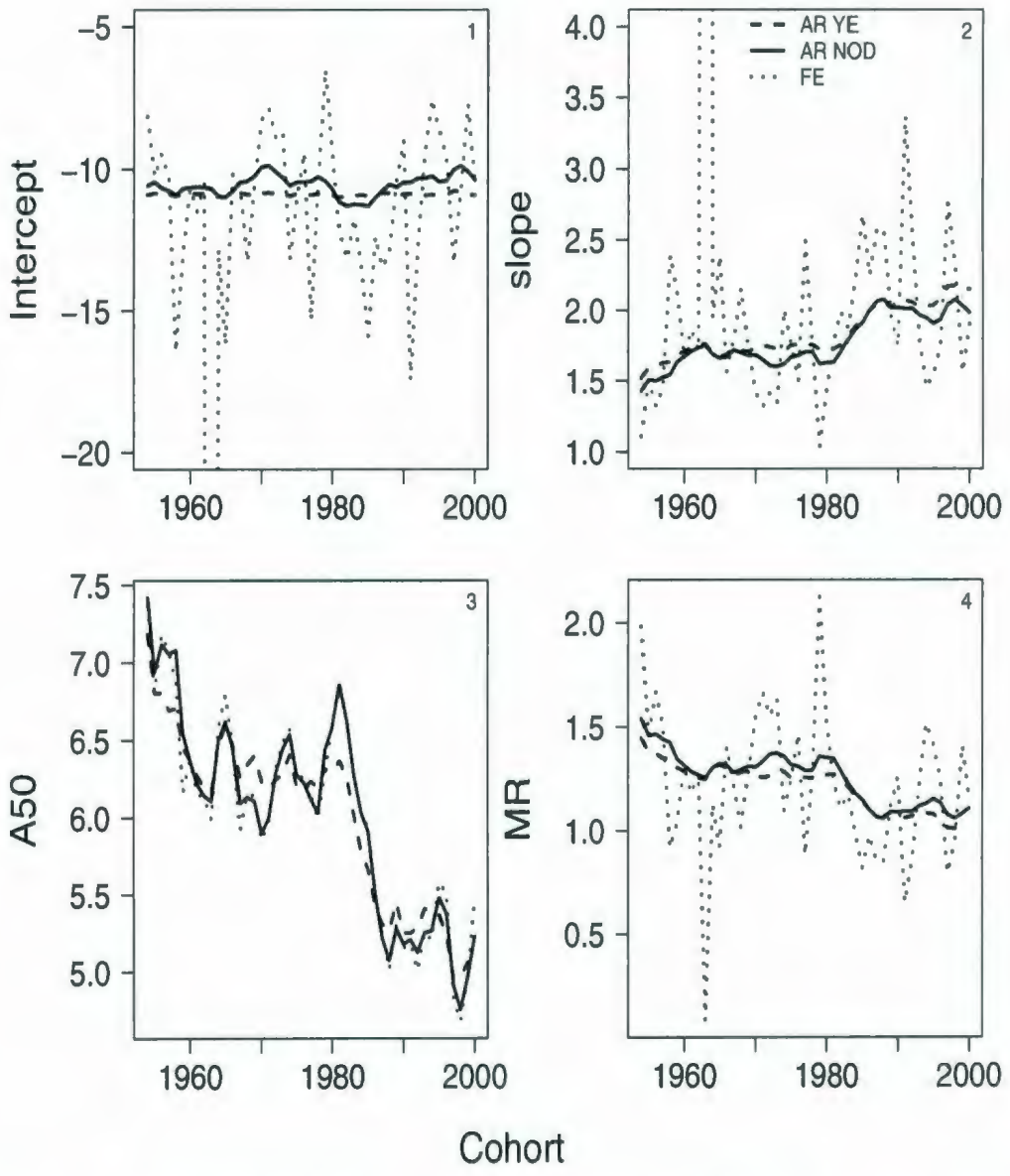


Figure 2.15: Estimates for 3Ps cod. Panel 1: intercepts. Panel 2: slopes. Panel 3: $\hat{A}_{50}$. Panel 4: $\hat{MR}$. AR NOD is the autoregressive (AR) mixed effects model with no overdispersion (NOD), and AR YE has year effects. FE is the fixed effects model.

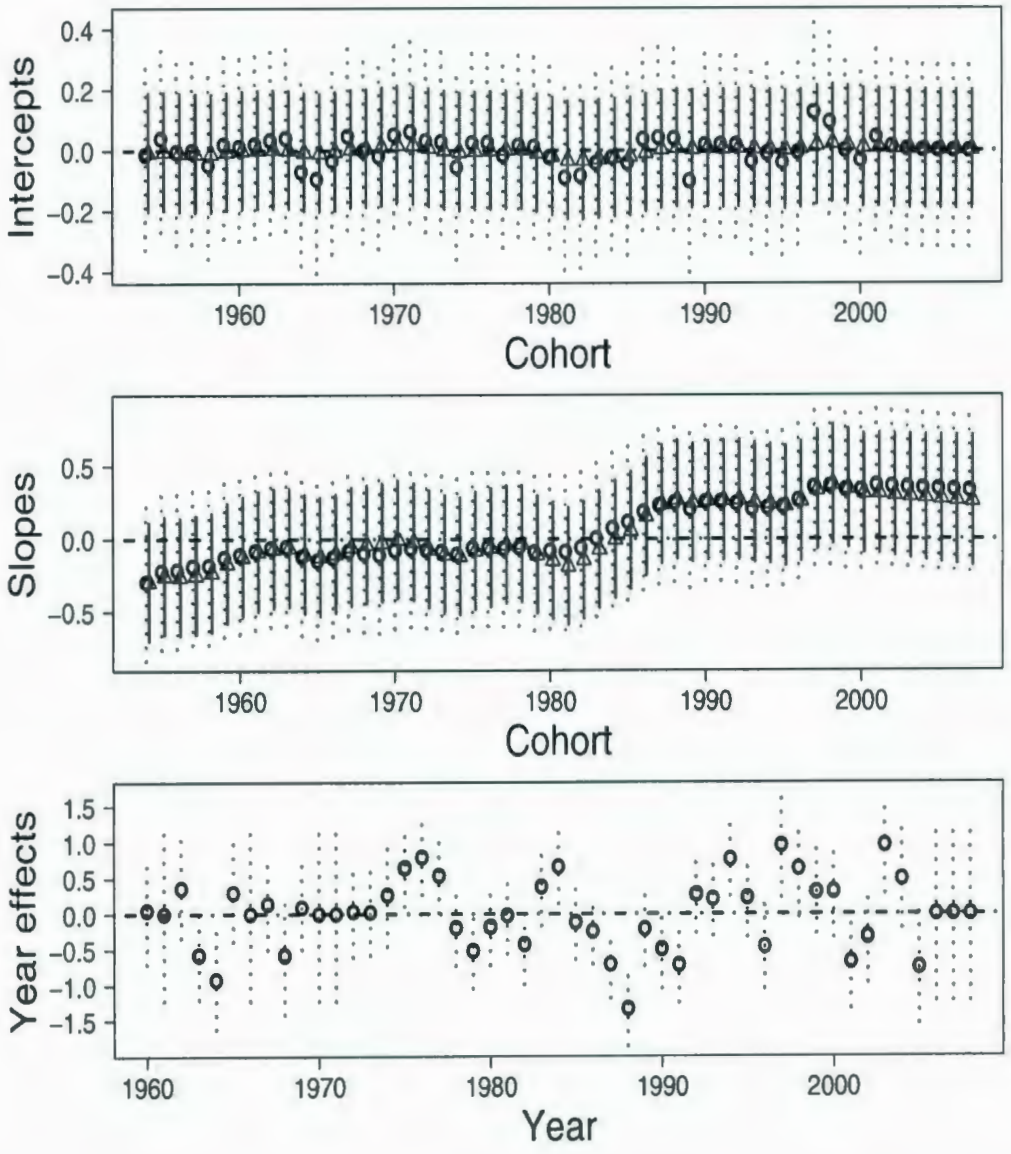


Figure 2.16: Predictions of random effects for 3Ps cod, with 95% confidence intervals (vertical lines). Triangles and solid lines are for the autoregressive mixed effects model with overdispersion (AR OD). Circles and dashed lines are for the AR model with year effects (AR YE).

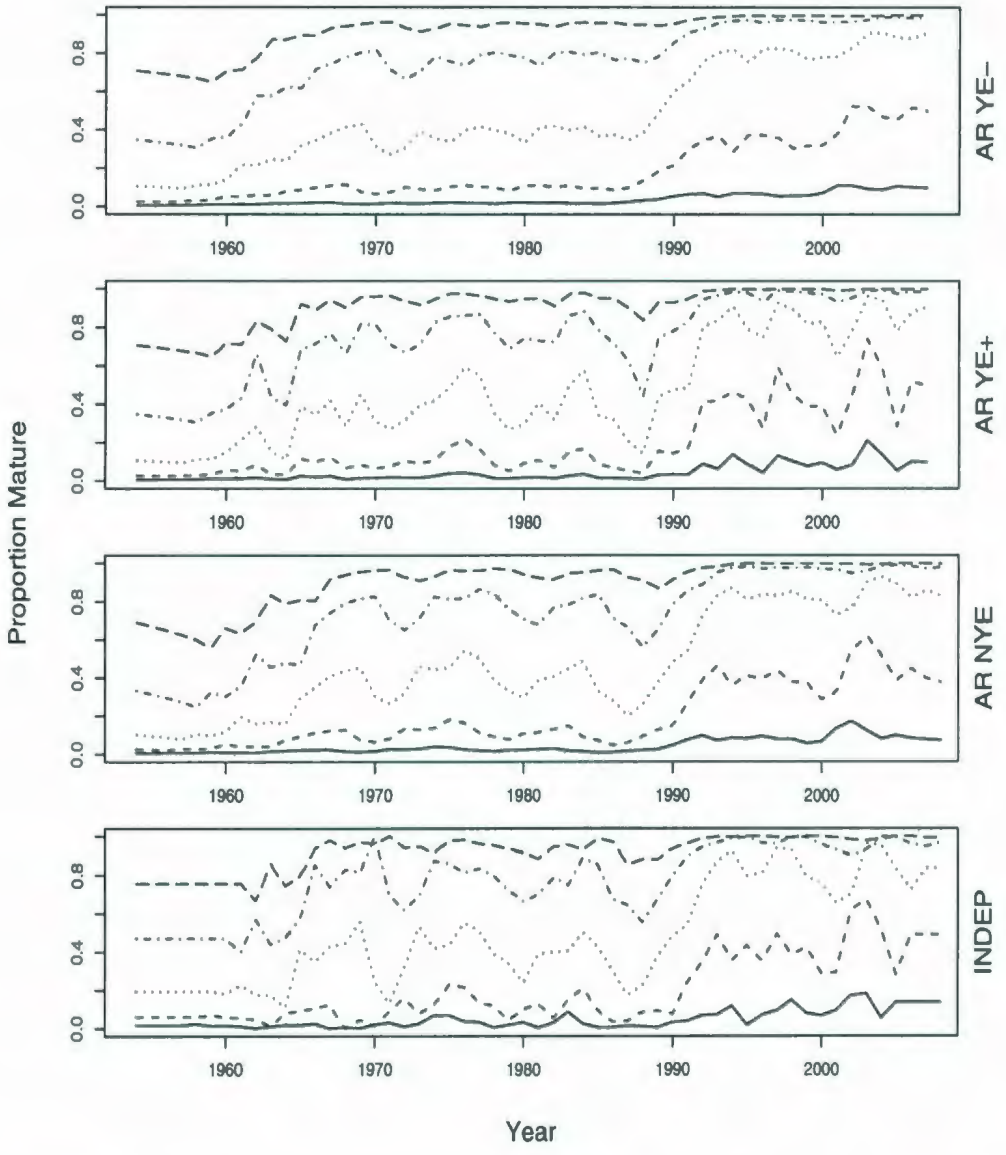


Figure 2.17: 3Ps cod proportions mature at ages 4-8 vs. year. Ages 5-8 are listed at the left-hand side. Top panel: Autoregressive (AR) mixed effects model with year effects as nuisance parameters (YE-). Second panel: AR with year effects as predictive parameters (YE+). Third panel: autoregressive (AR) mixed-effects model with no overdispersion (NOD). Bottom panel: Independent fixed effects (FE) model.

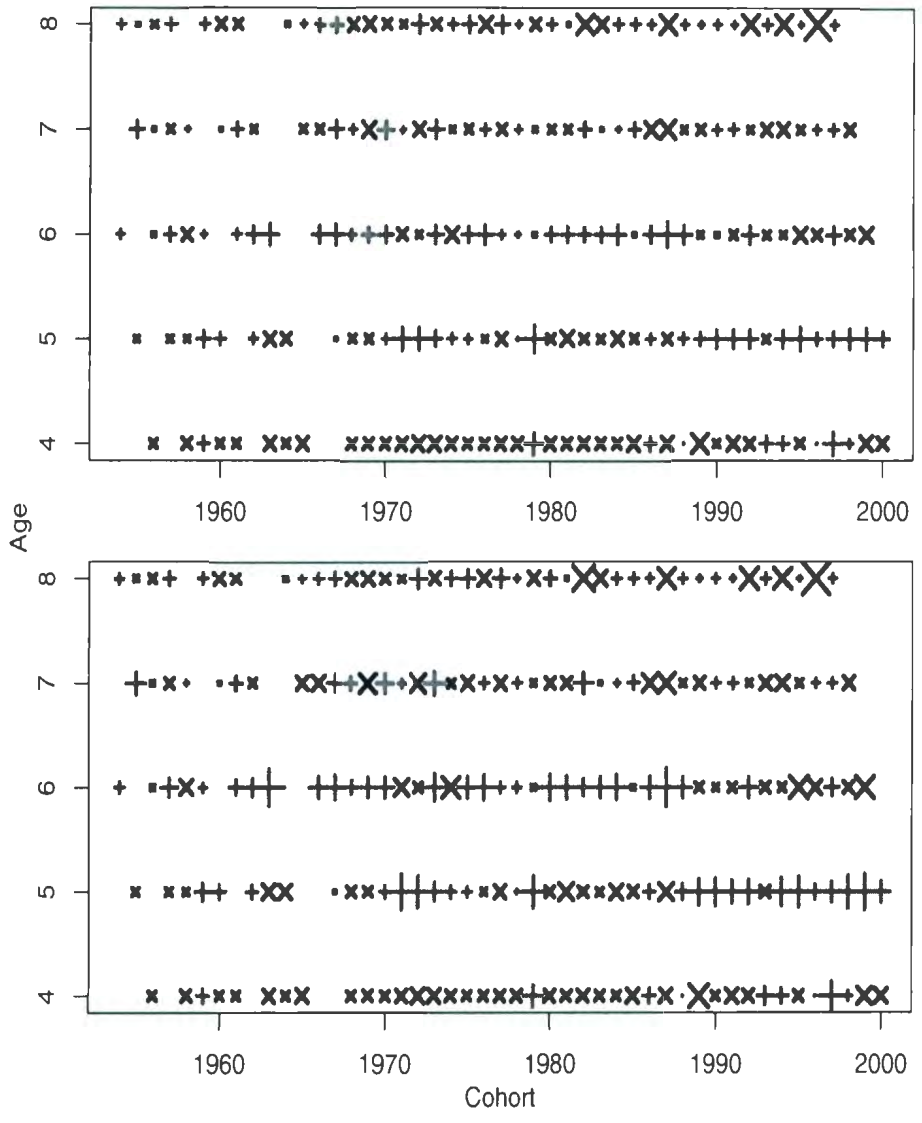


Figure 2.18: Residuals from the autoregressive mixed-effects model with year effects (AR YE) for 3Ps cod, + values are positive and × values are negative. Size is proportional to the absolute residual. Top panel: Chi-square ($\chi^2$) residuals. Bottom panel: Cross-validation chi-square ($\chi^2{}_1$) residuals.

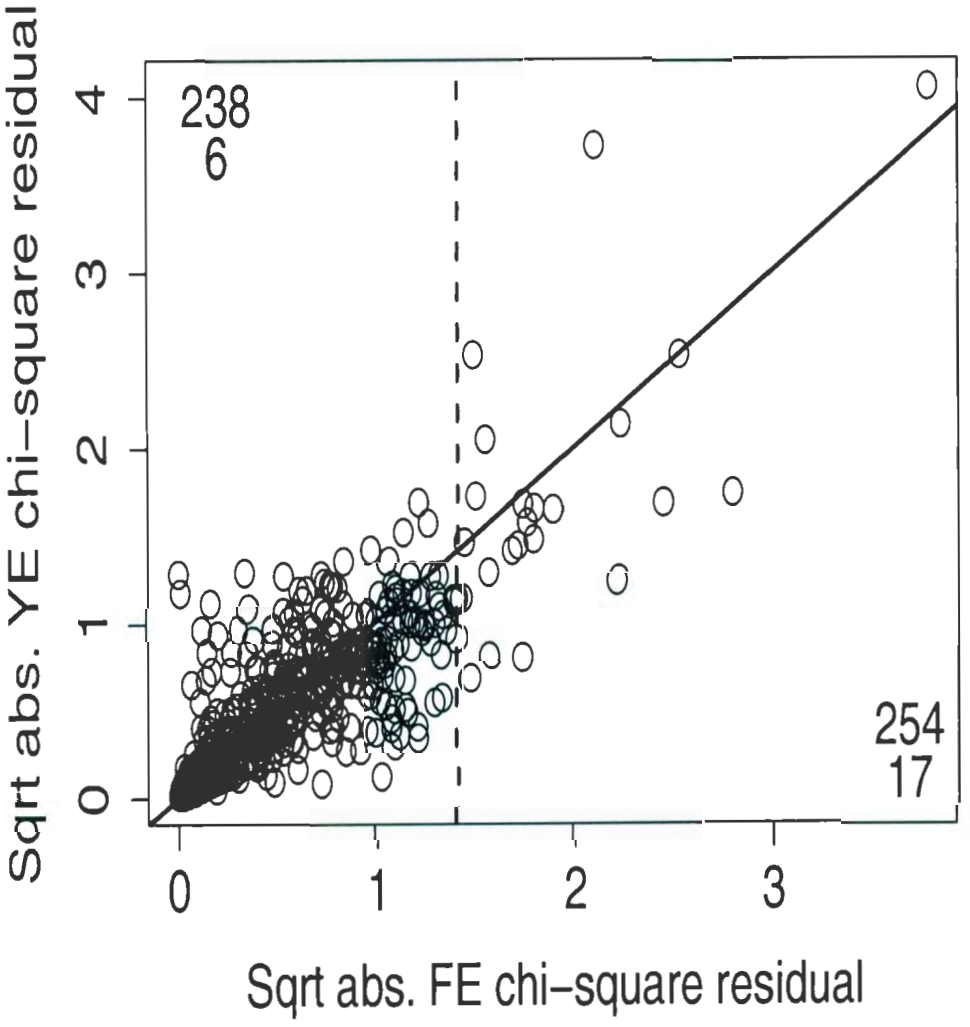


Figure 2.19: Square root of absolute values of the $\chi^2$ residuals from the autoregressive mixed-effects model with year effects (AR YE) and the fixed effects (FE) model for 3Ps cod. The 1:1 line is shown (solid) and the dotted line delineates FE residuals greater than $\sqrt{2}$. The number of points above and below the 1:1 line are shown, and beneath these values are the corresponding number of points whose FE residuals are greater than $\sqrt{2}$.

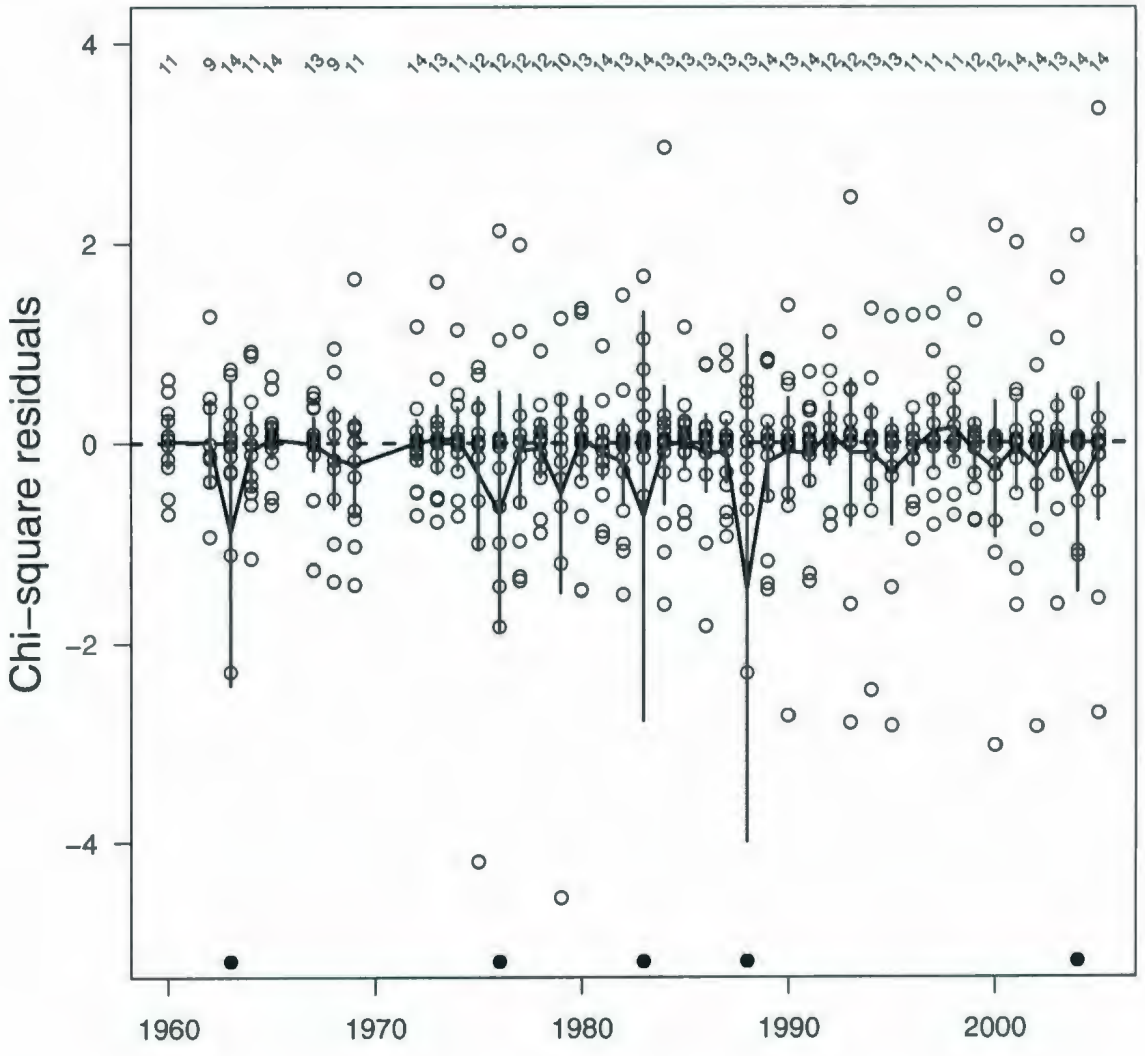


Figure 2.20: Absolute values of the $\chi^2$ residuals from the autoregressive mixed-effects model with year effects (AR YE) and the fixed effects (FE) model for 3Ps cod. The 1:1 line is shown (solid) and the dotted line delineates FE residuals greater than 2. The number of points above and 2below the 1:1 line are shown, and beneath these values are the corresponding number of points whose FE residuals are greater than 2.

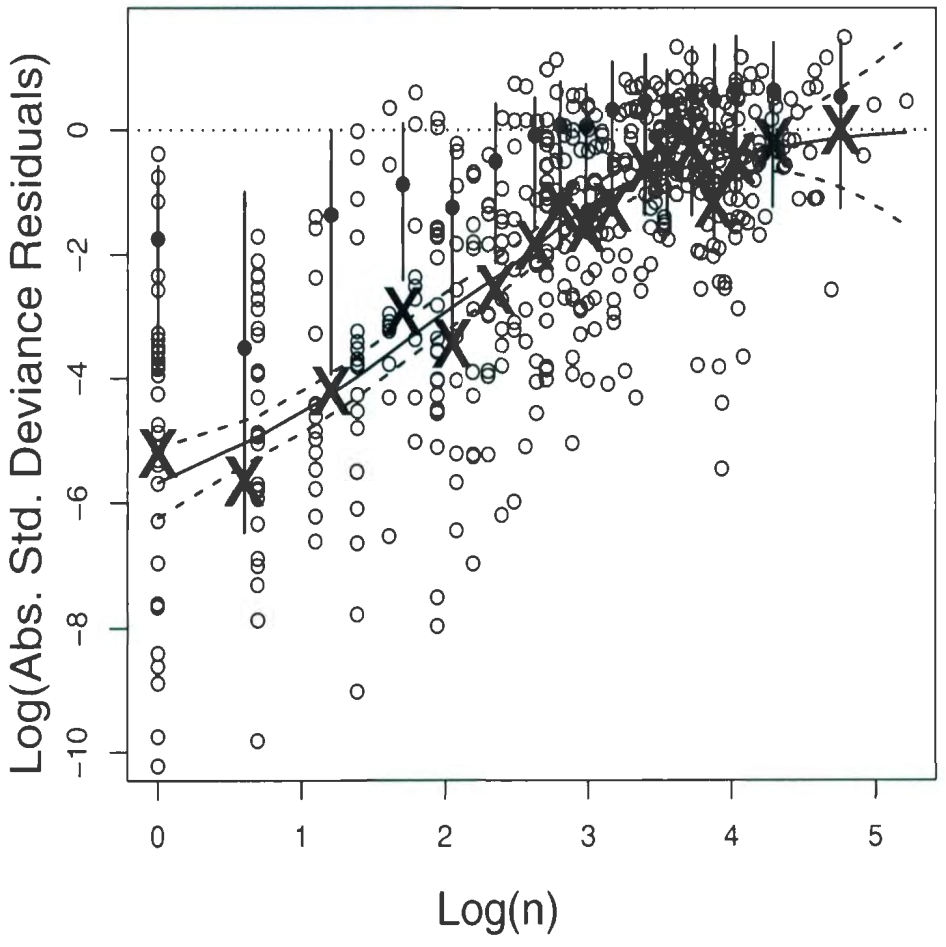


Figure 2.21: Log of absolute standardized deviance residuals.(○'s) for 3Ps cod from the fixed effects (FE) model vs. $log(n)$. The ×'s are average absolute *log* deviance residual in each bin. The •'s are the average *log* absolute deviance residuals in each bin using the parametric bootstrap procdeure. Vertical lines represent the 95% CI limits for each averaged *log* absolute deviance residual in each bin using the parametric bootstrap. The solid line is the fit from a loess smoother, and the dashed lines represents the 95% confidence limits for the smoother. The dotted line is a reference line at $log(1) = 0$, and represents the approximate expected value of the absolute standardized residuals.

# Chapter 3

# Application of GLMM: Fishery Survey Calibration Data

## 3.1 Introduction

Surveys are an important part of fisheries science. They are often multi-species bottom-trawl surveys and they are used extensively in stock assessments. The information collected is used to provide forecasts of stock status and for many other purposes such as determining species at risk (e.g. Smedbol et al., 2002). The sampling unit in a bottom trawl survey is defined as the area over the bottom covered by a trawl towed at a fixed speed for a fixed distance. A trawl sample is commonly referred to as a set, or a tow. A variety of information from each tow is collected for many species. In this chapter we focus on the total number caught for a species. A full review of bottom trawl surveys is given in Gunderson (1993).

Multi-species surveys are generally standardized in their survey protocols from year to year. These standardized protocols include the type of net used for the survey, the mesh size, the distance travelled per tow and the speed of the vessel. Typically a survey trawl does not catch all the fish in the sampled (i.e. swept) area. Some fish avoid the net and some fish escape through the mesh. The fraction of fish caught is commonly referred to as the trawl catchability. If the same protocols are used each

year then the trawl catchability should remain constant and the information collected (i.e. catch numbers, catch weight) should primarily reflect stock size. Annual changes in survey catches should reflect annual changes in stock size.

Sometimes it is necessary to introduce a new vessel, gear type, or some other aspect of the fishing plan which may affect catchability and impact the continuity of the survey time series (Lewy et al., 2004). In this case, it is important to compare the catchabilities between the two vessels, or more generally the two survey protocols, and derive estimates that will correct potential changes in catchability. In this chapter we examine if a change in survey vessels and trawl gear has an impact on catch rates from bottom trawl surveys.

The two research vessels used for the comparative fishing study are the *Canadian Coast Guard Alfred Needler* (AN) and the *Canadian Coast Guard Teleost* (TEL). Both vessels have many different characteristics. The AN is a 50m trawler and the TEL is a 63m trawler equipped with a more powerful engine. The AN has been used to conduct multi-species trawls in the northern Gulf of St. Lawrence (NAFO Divisions 4R and 4S, and SubDivision 3Pn; Figure 3.1) since 1990 and had always been equipped with a Uri 81'/114' shrimp trawl (Bourdages et al., 2007). However, since 2004 the TEL has been used for surveys in this region. The TEL is equipped with a Campelen 1800 shrimp trawl and has been used to conduct multi-species trawl surveys off the Atlantic coast since the mid-1990s (McCallum and Walsh, 2002). Differences in vessel characteristics are listed in Table 3.1. We examine paired-trawl experiments to estimate the relative difference between AN and TEL catch rates. Cadigan et al. (2006) conducted similar research whereby paired trawl catch rates between two "sister" vessels, the Alfred Needler and Wilfred Templemen, were examined to estimate differences in catchabilities (see Discussion in Section 5).

In paired-trawl experiments two vessels are fished as close together as possible to minimize spatial heterogeneity between stock densities fished by each vessel; therefore any differences in catches should reflect differences in catchabilities. In the past,

a common approach for parameter estimation was to log transform the catches and use normal linear model analysis. This approach does not properly account for the stochastic nature of the data since it involves arbitrary choices for dealing with zero catches. A better approach is to treat the catches from both vessels as Poisson or over-dispersed Poisson random variables, which are appropriate distributions for count data. This approach was used by Benoît and Swain (2003), but the analysis is complex since many fish density parameters for each tow have to be estimated (see Section 3.3). Pelletier (1998) assumed that catches followed a Negative Binomial distribution. Pelletier (1998) suggested that the number of parameters to be estimated could be reduced if fish densities were assumed to be constant between paired tows, or that the densities were random with the same mean between tows. However, these assumptions will not be appropriate in a large-scale comparative fishing survey since it is not possible to ensure that the conditions hold true. Cadigan et. al. (2006) also assumed catches to be Poisson random variables and used an associated conditional distribution whereby the total catch-at-length from both vessels was treated as fixed. This approach eliminates from the analysis the large number of fish density parameters. The conditional approach has been used in other areas of fisheries science such as fishing gear (net) size-selectivity studies (e.g. Millar, 1992).

A problem with the Poisson approach is that spatial variation and complex sampling can lead to Poisson over-dispersion and inaccurate estimates of standard errors. Benoit and Swain (2003) and Lewy et al. (2004) used an over-dispersion parameter to account for extra variation in the data. Cadigan et. al. (2006) addressed complex sampling and local spatial variability in stock densities fished between each trawler using a generalized linear mixed model (GLMM) which provides a flexible approach for parameter estimation with clustered data. In this chapter, we use the conditional distribution approach and GLMM to analyze catch rates for paired trawl calibration studies. A comprehensive discussion of GLMM's can be found in, for example, Demidenko (2004).

## 3.2 Methods

The objective of the comparative fishing exercise was to determine if differences in catches exist between two vessels when the AN used a Uri 81'/114' survey trawl and the TEL used a Campelen 1800 survey trawl (Parson et. al, MS 1997). Data from paired tows were collected to quantify potential differences. Ranges in catch sizes, fish sizes per catch and swept area were collected. The location of the comparative fishing was NAFO Divisons 4RS and Subdivision 3Pn (see Fig. ). Tow stations were selected randomly as a part of the survey protocol (Fig. ).

### 3.2.1 Paired-trawl fishing protocols

For each fishing tow made by the AN, a parallel tow in the same direction was conducted by the TEL. Both vessels operated at a distance of about 0.5 nautical miles apart. In order to avoid any bias based on the positioning of one vessel compared to the other, TEL would change sides (port, starboard) with the AN at each fishing tow (Bourdages et al., 2007). If the vessels could not fish one next to the other due to the narrowness of the stratum, they would then fish simultaneously one behind the other (about 0.5 nautical miles behind). If this situation occurred more than once, the vessels alternated between tows. During the entire comparative fishing activities, the operations of one vessel had to remain visible from the other vessel's wheelhouse. In order to make sure that fishing operations were conducted simultaneously, fishing operations on board the AN began at the same time as on the TEL.

Fishing operations on board the AN were conducted in the same manner as during previous surveys, including the duration of tows (24 minutes), the towing speed (3 knots) and the ratio of trawl warp length on fishing depth (from a predefined chart used by the AN). The duration of a tow was calculated between the time the trawler winches stopped because the trawl was down and the time they started again to bring the trawl back in. The data from the Scanmar probes (depth, vertical opening of the

trawl and distance between doors) were also recorded for each tow.

Fishing operations on board the TEL were conducted according to standard procedures established by the Newfoundland Region scientists for surveys using the Campelen trawl. The duration of a fishing tow was 15 minutes, calculated from the time the Scanmar probe signalled that the trawl had hit bottom. Tow speed was 3 knots. The ratio of trawl warp length on fishing depth was based on the chart usually used for similar depths. As with the AN, data from the Scanmar probes (depth, vertical opening of the trawl, door-spread and wing-spread) were recorded for each tow.

## 3.3 Statistical Models

The following section uses statistical models and terminology similar to Cadigan et al. (2006). Let $N_{ilk}$ be the number of fish at length $l$ caught at tow station $i$ by vessel $k$. We refer to each vessel as either the test or control vessel. We assume the control vessel ($k = c$) is the AN and the test vessel ($k = t$) is the TEL, although the results could be easily adjusted if the vessels were reversed. Let $\lambda_{ilk}$ denote the density of length $l$ fish encountered by vessel $k$ at tow station $i$.

Let $q_{lk}$ denote the probability that a fish is captured, which is assumed to be the same at each site $i$ but may differ for each vessel, $k$, and length, $l$. We expect $q_{lk}$ to vary smoothly in terms of $l$. The relative efficiency of the AN compared to the TEL is defined as

$$\rho_l = \frac{q_{lc}}{q_{lt}}. \tag{3.1}$$

### 3.3.1 Fixed effects model

With this model we assume that each vessel encounters the same local fish densities; i.e. $\lambda_{ilt} = \lambda_{ilc} = \lambda_{il}$ for all lengths $l$. If fish are captured independently from a large population then the catch by the test vessel is a Poisson random variable with mean

$$E(N_{ilt}) = q_{lt}\lambda_{il} = \mu_{ilt} \tag{3.2}$$

Likewise, the catch by the control vessel is also a Poisson random variable with mean

$$E(N_{ilc}) = q_{lc}\lambda_{il} = \rho_l\mu_{ilt}. \tag{3.3}$$

If fish enter the trawl as a Poisson process and are caught independently with probability $q$ then this is a Poisson thinning process (e.g. Grimmett and Stirzaker, 1992), and the catch is also a Poisson random variable. For the Poisson distribution, $Var(N) = E(N)$.

Each $\rho_l$ can be estimated using a Poisson generalized linear model (GLIM; e.g. McCullagh and Nelder, 1989) or $\rho_l$ can be modelled as a function of $l$ and the functional parameters estimated using a GLIM. This approach was used by Benoit and Swain (2003), in which they adjusted for extra Poisson variability, $Var(N) = \phi E(N)$. The Poisson GLIM approach is complicated since many tow stations and length classes are sampled, which means there will be many $\mu_{ilt}$ nuisance parameters to estimate.

If trawl catches are Poisson distributed then a better approach for inferences about $\rho_l$'s is to use the conditional distribution of $N_{ilk}$ given $N_{il}$ (Cox and Snell, 1989; Reid, 1995), where $N_{il}$ is the total number caught at length for both vessels. Let $n_{il}$ be the observed value of $N_{il}$. The conditional distribution of $N_{ilt}$ given $N_{il} = n_{il}$ is

$$Pr(N_{ilt} = x | N_{il} = n_{il}) = \binom{n_{il}}{x} p_l^x (1 - p_l)^{n_{il}-x}. \tag{3.4}$$

This is a binomial distribution where $p_l = \rho_l/(1+\rho_l)$ is the probability that a captured fish is caught by the AN (control vessel). The only unknown parameters in this distribution are the $\rho_l$'s.

The mean and variance of the binomial distribution are $E(N_{ilt}) = n_{il}p_l$ and $\mathrm{Var}(N_{ilt}) = n_{il}p_l(1-p_l)$, respectively. In some instances over-dispersion may arise. Over-dispersion occurs when data have more variability than accounted for in our modelling assumptions. It can be caused by factors such as population spatial heterogeneity and/or complex sampling. In paired-fishing experiments it is impossible to insure that exactly the same densities (i.e. $\lambda$'s) are fished by each vessel, and we think this is a major source of over-dispersion in our data. An approach to deal with over-dispersion is to use quasi-likelihood estimation (McCullagh and Nelder, 1989) with $Var(N_{ilt}) = \phi n_{il}p_l(1-p_l)$, where $\phi$ is an over-dispersion parameter. Note that the over-dispersed Poisson approach may give different inferences (e.g. standard errors) compared to the over-dispersed binomial approach. The binomial approach seems perferable for reasons outlined in Cox and Snell (1989) and Reid (1995).

Assume that $\rho_l$ is a smooth non-negative function of length. A suitable parametric model for $\rho_l$ is $\rho(l) = exp(\beta_0 + \beta_1 l)$. This leads to the logistic regression model

$$p(l) = \frac{exp(\beta_0 + \beta_1 l)}{1 + exp(\beta_0 + \beta_1 l)}, \tag{3.5}$$

which is the canonical link function for the binomial distribution (McCullagh and Nelder, 1989). This model is used in other areas of fisheries research such as fishing gear size-selectivity studies (Millar, 1992).

The logistic regression model is referred to as the FE2 model since there are two fixed effects parameters to be estimated. We also examine a model in which $\beta_1$ is fixed at zero, which is referred to as the FE1 model. In this case we can also pool data over lengths because $\rho_l$ is constant for all $l$. We refer to this as the FEP1 model. Pooling

is commonly used to avoid complications that arise when within-set catches are correlated. We use the maximum likelihood estimation method and SAS/STAT®PROC GENMOD software to estimate the logistic regression parameters.

Further adjustments are required due to variations in tow distances by each vessel at each tow site. Eqs. (3.2) and (3.3) can be multiplied by $A_{ik}$, the swept area of vessel $k$ at tow station $i$ and the $\lambda$ parameters can be thought of as a density per unit of swept area. It is then easy to show that

$$\log\left(\frac{p_{il}}{1-p_{il}}\right) = \beta_0 + \beta_1 l + \log\left(\frac{A_{ic}}{A_{it}}\right). \tag{3.6}$$

The last term in the logistic regression model is treated as an offset (McCullagh and Nelder, 1989). The adjustments are presented for the FE2 model, but the results can be used the same way with mixed effects models presented in the next section.

### 3.3.2 Mixed effects model

In this approach we do not assume that $\lambda_{ilc} = \lambda_{ilt}$. Let $\delta_{il} = \log(\lambda_{ilc}/\lambda_{ilt})$ and let $z_{il}$ be the offset term as defined in (3.6). The model for the AN proportion of catch is

$$\log\left(\frac{p_{il}}{1-p_{il}}\right) = \beta_0 + \beta_1 l + z_{il} + \delta_{il}. \tag{3.7}$$

If the length distributions of fish encountered by both vessels are exactly the same then $\delta_{il} = 0$. In practice this does not normally happen since the length distributions can be different, and the differences may vary systematically with length. This is illustrated in Fig. 3.2. In this hypothetical example the TEL caught larger fish than the AN and $\delta_{il}$ decreased with length. We suggest that in general $\delta_{il}$ will vary smoothly with length. However, we expect that for each length $l$ the variations in $\delta_{il}$ will be independent between sites (i.e. uncorrelated across tow sites).

A mixed effects model is used to account for this error structure. A mixed model contains both fixed parameters and random effects. We assume that the $\delta$'s are random variables from a normal distribution with mean zero, but are autocorrelated over lengths; that is, $\delta_{il} \sim N(0, \sigma_\delta^2)$ and $Corr(\delta_{im}, \delta_{in}) = \gamma^{|m-n|}$. This is an AR(1) correlation structure with $\gamma$ as the autocorrelation parameter. The $\delta_{il}$ are assumed to be uncorrelated between sites; that is, $Corr(\delta_{im}, \delta_{jn}) = 0$ for tow sites $i \neq j$ and for all lengths $m, n$. We use SAS/STAT®PROC GLIMMIX software for estimation.

Similar to the fixed effects models, we denote the mixed effects models with only a vessel effect (e.g. $\beta_1 = 0$) as ME1 and the mixed effects model with both vessel and length effect as ME2. If we pool the data, we denote the method as MEP1.

## 3.4 Results

During the 2004 and 2005 multi-species survey in Divisions 4RS and Subdivision 3Pn, a total of 154 successful paired tows (sets) were completed involving the AN and TEL. The sets were located in the offshore portions of 4RS and 3Pn (Fig. [illegible]). The distance between paired tows was 0.5 nautical miles. Average tow depths ranged from 42m in set 74 to 396m in set 223.

Four species of fish (Table [illegible]) were selected to assess the relative efficiency of the AN. More fish were measured on the TEL than on the AN. Species of crab and shrimp were also measured but are not considered here. The largest difference in catch totals between vessels occured for American plaice. Catch totals for Witch flounder were less for both vessels as compared to the other species.

### 3.4.1 Fixed effects model (FE1)

In this model the length parameter in ( [illegible] ) is fixed at zero and the intercept ($\beta_0$) is treated as an unknown fixed quantity to estimate. In the first analysis catches were pooled over length classes within each set. Estimates were significantly different from

zero (Table ) for each stock. Each estimate was negative indicating that the AN had an overall lower relative efficiency than the TEL. Consistent with Table , all estimates were significantly different than zero for the un-pooled analysis (Tables and ). The standard errors were smaller in the un-pooled analysis. If observations are correlated within sets, as we expect, then these standard errors will be too small. In this situation the pooled analysis may give more reliable results than the un-pooled analysis, but a mixed-model should do even better. Note that $\hat{\beta}_0$ was identical for the pooled and un-pooled catches since there was no differential subsampling among lengths (see Cadigan et. al., 2006).

Pooled catches from each vessel are shown in Fig. . The estimated relative efficiency from the FEP1 model is shown (dashed line) as a line through the origin with slope $\hat{\rho} = exp(\hat{\beta}_0)$. Catch totals were divided by the area swept in each tow (scaled catch). The scaled catches (for all sets) by the TEL were greater than the AN for all species. The $\rho$ for the four species were substantially different from one (slope of the dotted line).

### 3.4.2 Mixed effects model (ME1)

In the first analysis catches were pooled within sets, which we refer to as the MEP1 model. The random effects $\delta_i = \sum_l \delta_{il}$ were assumed to be independent and identically distributed (iid) N(0,$\sigma^2$), $i = 1, 2, ...$ The $\beta_0$ estimates were negative (Table ) for each species indicating again that the AN had a lower relative efficiency than the TEL and the effects were significant for all species. Note that $\hat{\beta}_0$s (Table ; Fig. ) tended to be similar to the FEP1 results but standard errors were more different (Table ). The predicted random effects (Fig. ) were smaller for Greenland halibut and Witch flounder than for Atlantic cod or American plaice. This corresponds to the smaller estimates of $\sigma^2$ for Greenland halibut and Witch flounder (Table ).

We also investigated an ME1 (not pooled over lengths) model that assumed the $\delta_{il}$s were constant across all lengths for each set and equal to $\delta_i$ which were iid N(0,$\sigma^2$)

for $i = 1, 2, ...$ This ME1 model was comparable to the MEP1 model. The un-pooled results (Table ) were identical to the pooled results for estimates, standard errors, and confidence intervals.

### 3.4.3 Outliers I

Outliers are common in comparative fishing experiments and are difficult to deal with in practice, especially when one has to examine hundreds of data sets for different species and areas.

The FEP1 and MEP1 models were re-estimated after potential outliers were removed (Fig. ; black solid circles). The purpose of this is to simply explore the sensitivity of the various methods to outliers. Two sets (i.e. trawl pairs) were removed for each species. MEP1 estimates of $\beta_0$ appear to be more stable for Atlantic cod and American plaice (Fig. ), they did not change as much when outliers were removed compared to FEP1 results.

The vessel effect was significant for both the FEP1 and MEP1 models when two potential outliers were removed (Tables and ), similar to the results based on all of the data.

### 3.4.4 Fixed effects model (FE2)

In this model, both parameters in Eq. ( ) are treated as as unknown fixed quantities to estimate, based on the un-pooled data. Parameter estimates, standard errors, and confidence intervals are shown in Table . Significant differences in the relative efficiency were found in all four species. Fig. presents the estimated $\rho_l$'s and the proportion of the total catch-at-length taken by the AN. The estimated $\rho_l$'s were less than one, except for larger lengths of American plaice, and increased with length for three of the four stocks. This suggests that the AN had a lower catchability than the TEL but the differences in catchabilities decreased for larger sized fish. The observed

proportions of catch for the AN were close to the model predictions except for small and large lengths. The sample sizes are not reflected in the proportions and would tend to be much smaller for small and large lengths.

The length distributions of total catches and residuals are presented in Figs. - . The differences in length frequencies are large. The largest difference occured for American plaice (Fig. ). AN total catches per swept area are much more comparable to the TEL total catches per swept area when adjusted by the relative efficiency ($\rho_l$), for all species. Some potential outliers are apparent (i.e. American plaice and Greenland halibut), although the residuals-at-length do not deviate substantially from zero.

### 3.4.5 Mixed effects model (ME2)

In the ME2 model the (Eq. ) random effects were modelled as autocorrelated random variables. There was a significant length effect for American plaice and Atlantic cod (Table ). The estimated $\rho_l$'s were similar to the FE2 model for American plaice, Atlantic cod, and Witch flounder but differed for Greenland halibut (see Fig. ).

The autocorrelation estimates (Table ) for $\delta$ were greater than 0.9 and the error variance $\sigma^2 >> 0$ which indicates that there was a substantial length dependency between the within-set proportion of total catch by the AN, beyond what was accounted for by the fixed length effect $\beta_1$. The length distributions of total catches and residuals are presented in Figs. - . AN total catches per swept area are more comparable to the TEL total catches per swept area when adjusted by the relative efficiency ($\rho_l$), for all species. The adjusted AN catches deviate slightly more from the TEL catches compared to the FE2 model. The predicted random effects deviated substantially from zero which indicates the ratio of catch densities for some lengths and sets were quite different from the overall average $\rho_l$. No large set outliers

were apparent from the conditional chi-squared residuals, except for perhaps American plaice (Fig. ). Outliers in the FE2 analyses appear to have been accounted for by the random effects. Standardized residuals versus length are similar to those from the fixed effects model.

### 3.4.6 Outliers II

The FE2 and ME2 models were re-estimated after two potential outliers were removed from the data (Fig. ). These were the same sets removed for the FEP1 and MEP1 models (Fig ; solid black circles). Estimates of $\beta_0$ and $\beta_1$ apear to be equally sensitive for both models.

## 3.5 Discussion

The results overall suggest that there was a significant difference in catch rates between the AN and TEL. The AN and TEL total length frequencies were substantially different. The sign of the vessel effects for all species examined was negative which provides evidence that the catchability of the AN was lower than the TEL.

The results from the pooled (over lengths) fixed effects GLIM suggested that differences between vessels were significant. However, this model was based on the erroneous assumption that differences in stock densities fished by each vessel were the same. We also analyzed the pooled data using a mixed effects model with an iid random normal intercept effect for each set. The parameter estimates from this model were similar to the estimates from the fixed effects model, but the standard errors were larger from the mixed effects model, even though in the GLIM model the binomial variability was adjusted for over-dispersion. Benoit and Swain (2003) suggested that standard errors for comparative fishing data were too small. They used a randomization approach to get more reasonable standard errors; however, the mixed-model we use appears to be an easier approach. We will study the efficacy of

the fixed and mixed models for estimating relative efficiency from comparative fishing data in the Chapter 4.

The magnitude of the differences in catchabilities appear to overwhelm any differences due to methodology, although, this was not the case in Cadigan et al. (2006). These authors found that the fixed effects model incorrectly indicated significant differences in catchabilities when these differences were small or nonexistent. The mixed effects did not, although it is possible that the mixed effects model may simply have very low power. The analyses we presented here demonstrate that when differences in catchabilities are large then the mixed effects approach detects these differences as significant.

We also showed with the mixed effects approach that exactly the same estimates and inferences can be obtained with pooled and un-pooled data provided the random effects assumptions used with the un-pooled data were consistent with the pooled data.

We suggested that differences in stock densities will not be completely random but will vary smoothly as a function of length. When this random structure is accounted for using a length autocorrelated random component in the logistic model for the un-pooled, the relative efficacy was found to be significantly different from one for the four stocks we examined. We suggest these mixed effects results are more reliable because the basis for statistical inference is more reliable. However, the efficacy of the mixed model approach for estimating relative efficiency and determining statistical significance requires further evaluation. Simulations would be useful for this purpose (see Chapter 4).

We presented evidence that mixed effects models can accomodate paired trawl outliers as well as fixed effects models. Cadigan et. al. (2006) found evidence that the ME approach was more robust to some large outliers as compared to the FE approach. This implies that the mixed effects model may be more advantageous for analyzing data when outliers are present.

Pooling catches within each set may be appropriate for fixed effects models when catches are correlated between sets. This correlation can be accomodated for the mixed effects model and pooling is not necessary with this approach as seen in Tables 4.1 and 4.2. This is advantageous since pooling can be problematic when catches are subsampled differently (see Section 2.2.2 of Cadigan et al., 2006). There is also some loss in the precision of estimators based on pooled data, although this depends on the amount of within-set correlation.

Cadigan et al. (2006) estimated relative efficiency between two "sister" vessels, the Alfred Needler and Wilfred Templeman. These vessels are sister vessels since they were similar in terms of structure (i.e. hull geometry and horsepower) and used identical fishing gear (Campelen 1800 survey trawl). The results showed overall that the mixed effects model performed better than the fixed effects model for similar reasons noted in this chapter.

Table 3.1: Differences in vessel characteristics of the Canadian Coast Guard vessels Alfred Needler (AN) and Teleost (TEL).

| Parameter | AN | TEL |
|---|---|---|
| Length (m) | 50.3 | 63.0 |
| Breadth (m) | 11.0 | 14.2 |
| Draft (m) | 4.9 | 7.2 |
| Displacement (t) | 958 | 2215 |
| Power (hp) | 2600 | 4000 |
| Winch System | fixed brake | autotrawl |
| Warp Diameter (mm) | 28.6 | 25.4 |
| Warp Weight (kg/m) | unassigned | 3.0 |

Table 3.2: Catch summaries. $N_c$ and $N_t$ are the AN and TEL measured catches. $n$ is the total number of obsevations (lengths and tows) where $N_t + N_c > 0$.

| Species | $n$ | $N_c$ | $N_t$ | $N_t + N_c$ | $N_t/N_c$ |
|---|---|---|---|---|---|
| American plaice | 1768 | 2918.6 | 10441.58 | 13360.18 | 3.58 |
| Atlantic cod | 1749 | 911.86 | 3580.44 | 4492.3 | 3.92 |
| Greenland halibut | 2523 | 10987.12 | 14079.31 | 25066.43 | 1.28 |
| Witch flounder | 1272 | 471.00 | 1984.00 | 2455.00 | 4.21 |

Table 3.3: FEP1 model results. SE - standard error. L,U - profile likelihood confidence intervals. pv - $\chi^2$ p-value. Significant estimates in bold.

| Species | $\hat{\phi}$ | $\hat{\beta}_0$ | SE | 95% L | 95 % U | $\chi^2$ | pv |
|---|---|---|---|---|---|---|---|
| American plaice | 3.73 | **−1.51** | 0.0781 | -1.6654 | -1.3590 | 373.46 | $< 0.0001$ |
| Atlantic cod | 2.73 | **−1.59** | 0.1014 | -1.7913 | -1.3935 | 245.46 | $< 0.0001$ |
| Greenland halibut | 3.62 | **−0.498** | 0.0461 | -0.5887 | -0.4081 | 116.99 | $< 0.0001$ |
| Witch flounder | 1.08 | **−1.69** | 0.0554 | -1.7954 | -1.5783 | 926.71 | $< 0.0001$ |

Table 3.4: FE1 model results. SE - standard error. L,U - profile likelihood confidence intervals. pv - $\chi^2$ p-value. Significant estimates in bold.

| Species | $\hat{\phi}$ | $\hat{\beta}_0$ | SE | 95% L | 95 % U | $\chi^2$ | pv |
|---|---|---|---|---|---|---|---|
| American plaice | 1.58 | **−1.51** | 0.0331 | -1.5753 | -1.4455 | 2801.01 | $< 0.0001$ |
| Atlantic cod | 1.16 | **−1.59** | 0.0430 | -1.6735 | -1.5049 | 1364.29 | $< 0.0001$ |
| Greenland halibut | 1.45 | **−0.498** | 0.0185 | -0.5350 | -0.4625 | 726.88 | $< 0.0001$ |
| Witch flounder | 1.59 | **−1.69** | 0.0818 | -1.8484 | -1.5274 | 424.42 | $< 0.0001$ |

Table 3.5: MEP1 model results. SE - standard error. L,U - profile likelihood confidence intervals. pv is the $t$-statistic p-value. Significant estimates in bold.

| Species | $\hat{\sigma}^2$ | $\hat{\beta}_0$ | SE | 95% L | 95 % U | $t$ | pv |
|---|---|---|---|---|---|---|---|
| American plaice | 1.08 | **−1.355** | 0.1078 | -1.5682 | -1.1417 | -12.56 | $< 0.0001$ |
| Atlantic cod | 1.02 | **−1.686** | 0.1320 | -1.9480 | -1.4232 | -12.77 | $< 0.0001$ |
| Greenland halibut | 0.2721 | **−0.639** | 0.0588 | -0.7554 | -0.5224 | -10.87 | $< 0.0001$ |
| Witch flounder | 0.457 | **−1.682** | 0.0896 | -1.8597 | -1.5051 | -18.78 | $< 0.0001$ |

Table 3.6: ME1 model results. SE - standard error. L,U - profile likelihood confidence intervals. pv is the $t$-statistic p-value. Significant estimates in bold.

| Species | $\hat{\sigma}^2$ | $\hat{\beta}_0$ | SE | 95% L | 95 % U | $t$ | pv |
|---|---|---|---|---|---|---|---|
| American plaice | 1.08 | **−1.355** | 0.1078 | -1.5682 | -1.1417 | -12.56 | $< 0.0001$ |
| Atlantic cod | 1.02 | **−1.686** | 0.1320 | -1.9480 | -1.4232 | -12.77 | $< 0.0001$ |
| Greenland halibut | 0.2721 | **−0.639** | 0.0588 | -0.7554 | -0.5224 | -10.87 | $< 0.0001$ |
| Witch flounder | 0.457 | **−1.682** | 0.0896 | -1.8597 | -1.5051 | -18.78 | $< 0.0001$ |

Table 3.7: FEP1 model results when two potential outliers (i.e trawl pairs) were removed. SE - standard error. L,U - profile likelihood confidence intervals. pv - $\chi^2$ p-value. Significant estimates in bold.

| Species | $\hat{\phi}$ | $\hat{\beta}_0$ | SE | 95% L | 95 % U | $\chi^2$ | pv |
|---|---|---|---|---|---|---|---|
| American plaice | 3.13 | **−1.60** | 0.0702 | -1.7405 | -1.4651 | 519.58 | $< 0.0001$ |
| Atlantic cod | 2.53 | **−1.73** | 0.1014 | -1.9441 | -1.5358 | 278.17 | $< 0.0001$ |
| Greenland halibut | 3.40 | **−0.469** | 0.0440 | -0.5557 | -0.3833 | 113.95 | $< 0.0001$ |
| Witch flounder | 1.55 | **−1.66** | 0.0795 | -1.8152 | -1.5034 | 434.20 | $< 0.0001$ |

Table 3.8: MEP1 model results when two potential outliers (i.e trawl pairs) were removed. SE - standard error. L,U - profile likelihood confidence intervals. pv is the $t$-statistic p-value. Significant estimates in bold.

| Species | $\hat{\sigma}_2$ | $\hat{\beta}_0$ | SE | 95% L | 95 % U | $t$ | pv |
|---|---|---|---|---|---|---|---|
| American plaice | 0.987 | **−1.346** | 0.1052 | -1.5541 | -1.1381 | -12.80 | $< 0.0001$ |
| Atlantic cod | 1.03 | **−1.72** | 0.1345 | -1.9885 | -1.4536 | -12.80 | $< 0.0001$ |
| Greenland halibut | 0.256 | **−0.634** | 0.0578 | -0.7478 | -0.5195 | -10.97 | $< 0.0001$ |
| Witch flounder | 1.55 | **−1.66** | 0.0795 | -1.8152 | -1.5034 | -20.88 | $< 0.0001$ |

Table 3.9: FE2 model results. SE - standard error. L,U - profile likelihood confidence intervals. pv - $\chi^2$ p-value. Significant estimates in bold.

| Species | $\hat{\phi}$ | | Estimate | SE | 95% L | 95 % U | $t$ | pv |
|---|---|---|---|---|---|---|---|---|
| American plaice | 1.53 | $\hat{\beta}_0$ | **−2.7552** | 0.1307 | -3.0128 | -2.5001 | 444.05 | $< 0.0001$ |
| | | $\hat{\beta}_1$ | **0.0524** | 0.0052 | 0.0421 | 0.0626 | 100.60 | $< 0.0001$ |
| Atlantic cod | 1.13 | $\hat{\beta}_0$ | **−2.7558** | 0.1523 | -3.0576 | -2.4605 | 327.55 | $< 0.0001$ |
| | | $\hat{\beta}_1$ | **0.0267** | 0.0032 | 0.0204 | 0.0331 | 67.71 | $< 0.0001$ |
| Greenland halibut | 1.45 | $\hat{\beta}_0$ | **−0.2348** | 0.0536 | -0.3399 | -0.1297 | 19.17 | $< 0.0001$ |
| | | $\hat{\beta}_1$ | **−0.0089** | 0.0017 | -0.0122 | -0.0056 | 27.36 | $< 0.0001$ |
| Witch flounder | 1.07 | $\hat{\beta}_0$ | **−2.1178** | 0.1669 | -2.4498 | -1.7951 | 160.96 | $< 0.0001$ |
| | | $\hat{\beta}_1$ | **0.0171** | 0.0061 | 0.0051 | 0.0292 | 7.79 | 0.0053 |

Table 3.10: ME2 model results. SE - standard error. L,U - profile likelihood confidence intervals. pv is the $t$-statistic p-value. Significant estimates in bold.

| Species | $\hat{\sigma}^2/\hat{\gamma}$ | | Estimate | SE | 95% L | 95 % U | $t$ | pv |
|---|---|---|---|---|---|---|---|---|
| American plaice | 1.296 | $\hat{\beta}_0$ | **−2.5031** | 0.2369 | -2.9714 | -2.0348 | -10.57 | $< 0.0001$ |
| | 0.959 | $\hat{\beta}_1$ | **0.0448** | 0.0080 | 0.0292 | 0.0605 | 5.62 | $< 0.0001$ |
| Atlantic cod | 0.986 | $\hat{\beta}_0$ | **−2.6597** | 0.2249 | -3.1069 | -2.2125 | -11.82 | $< 0.0001$ |
| | 0.995 | $\hat{\beta}_1$ | **0.0224** | 0.0042 | 0.0104 | 0.0307 | 5.28 | $< 0.0001$ |
| Greenland halibut | 0.374 | $\hat{\beta}_0$ | **−0.5635** | 0.1273 | -0.8159 | -0.3110 | -4.42 | $< 0.0001$ |
| | 0.929 | $\hat{\beta}_1$ | −0.0014 | 0.0035 | -0.0083 | 0.0054 | -0.420 | 0.676 |
| Witch flounder | 0.697 | $\hat{\beta}_0$ | **−2.0847** | 0.2216 | -2.5233 | -1.6461 | -9.41 | $< 0.0001$ |
| | 0.933 | $\hat{\beta}_1$ | 0.01497 | 0.0079 | -0.0005 | 0.03048 | -0.82 | 0.059 |

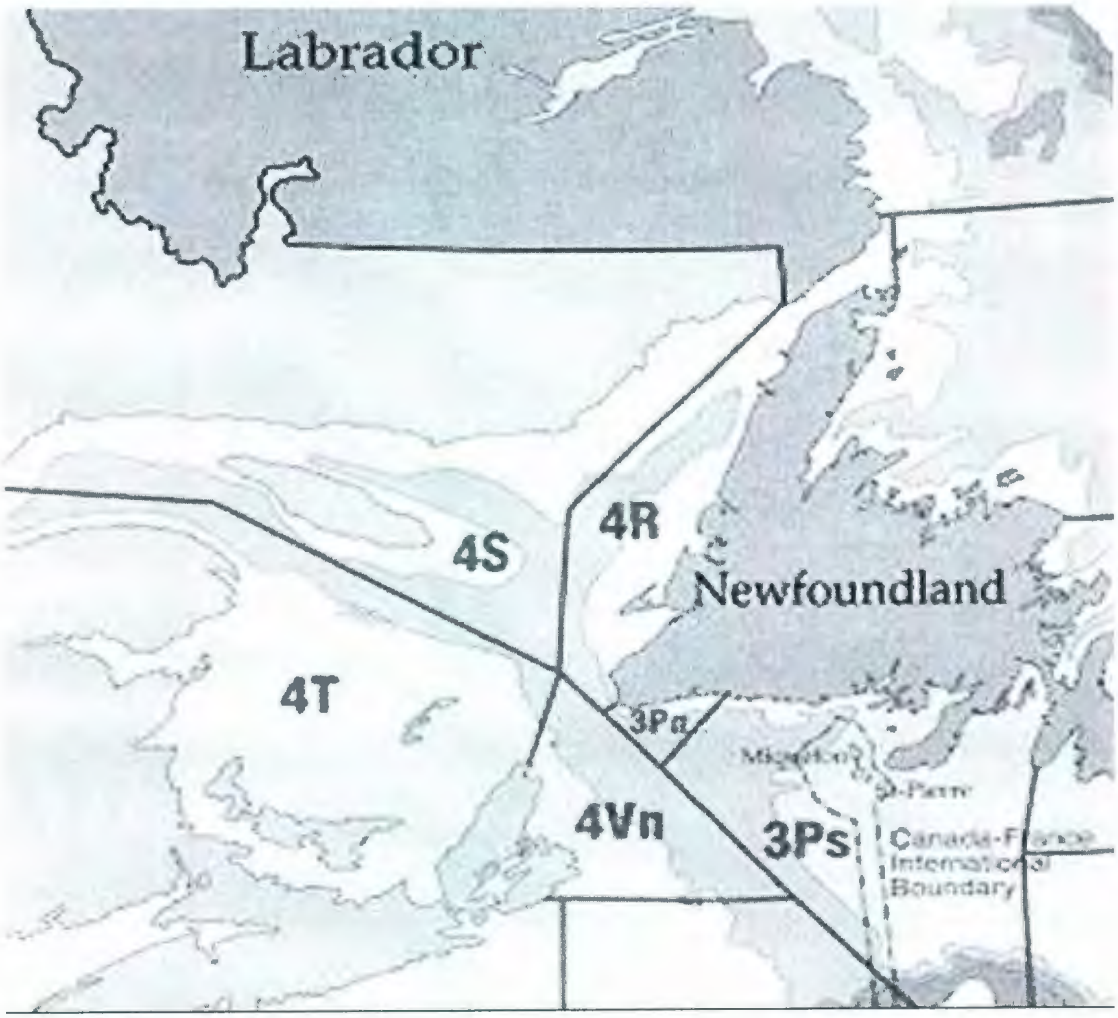


Figure 3.1: Northwest Atlantic Fisheries Organization (NAFO) northern Gulf fisheries management Divisions 4RS and Subdivision 3Pn.

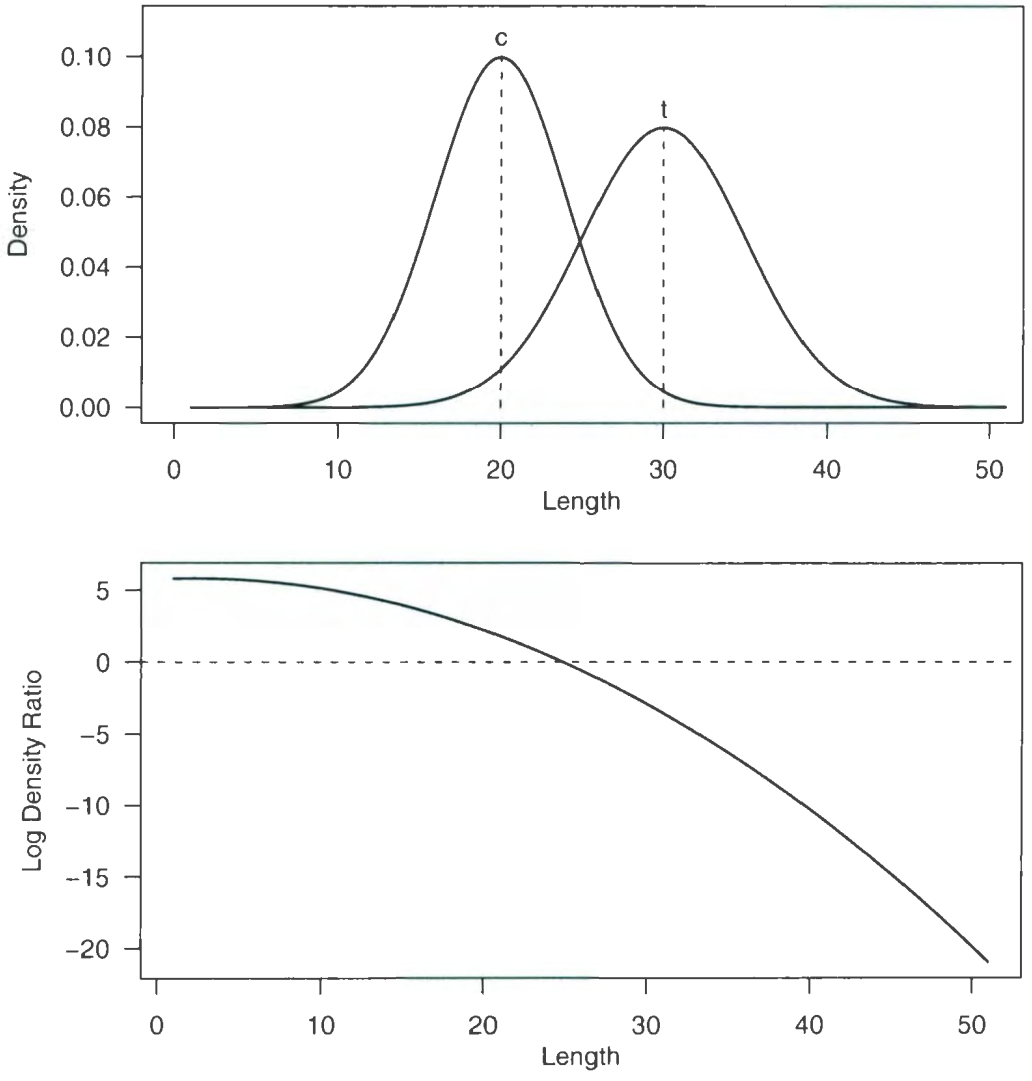


Figure 3.2: Top panel: Hypothetical length distributions sampled by each trawl, $\lambda_{lc}$: the fish density encountered by the Alfred Needler and $\lambda_{lt}$: the fish density encountered by the Teleost. Bottom panel: log density ratio, $\delta_l = \log(\lambda_{lc}/\lambda_{lt})$.

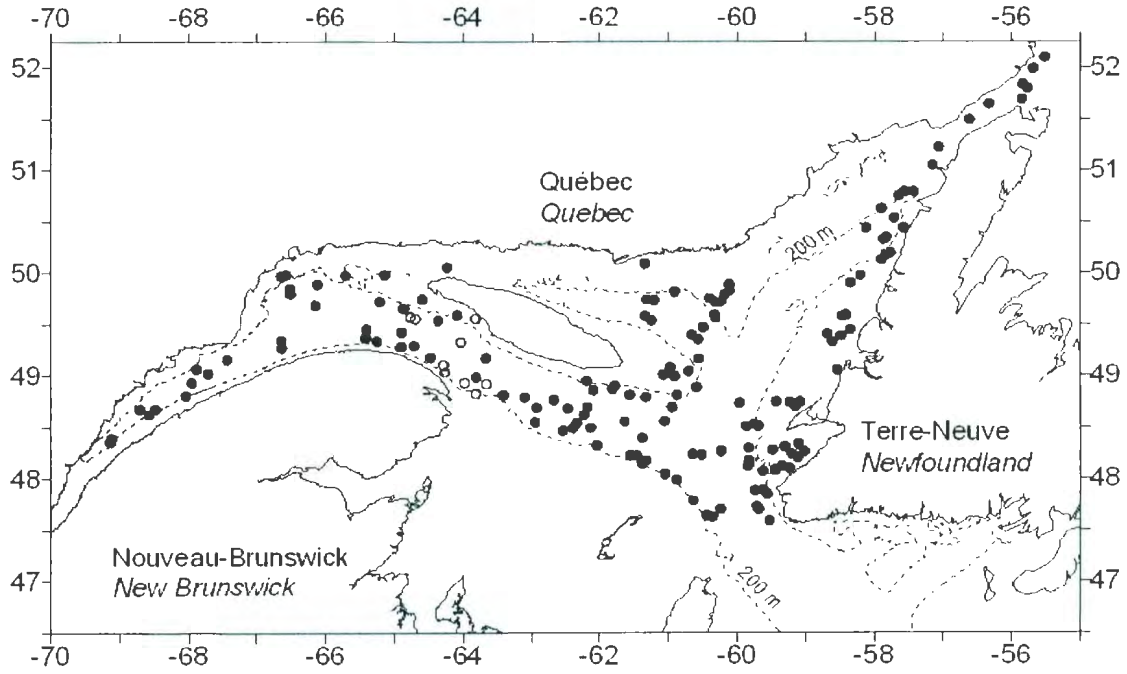


Figure 3.3: Location of the comparative fishing for NAFO Divisons 4RS and Subdivision 3Pn in 2004 (○) and 2005 (●).

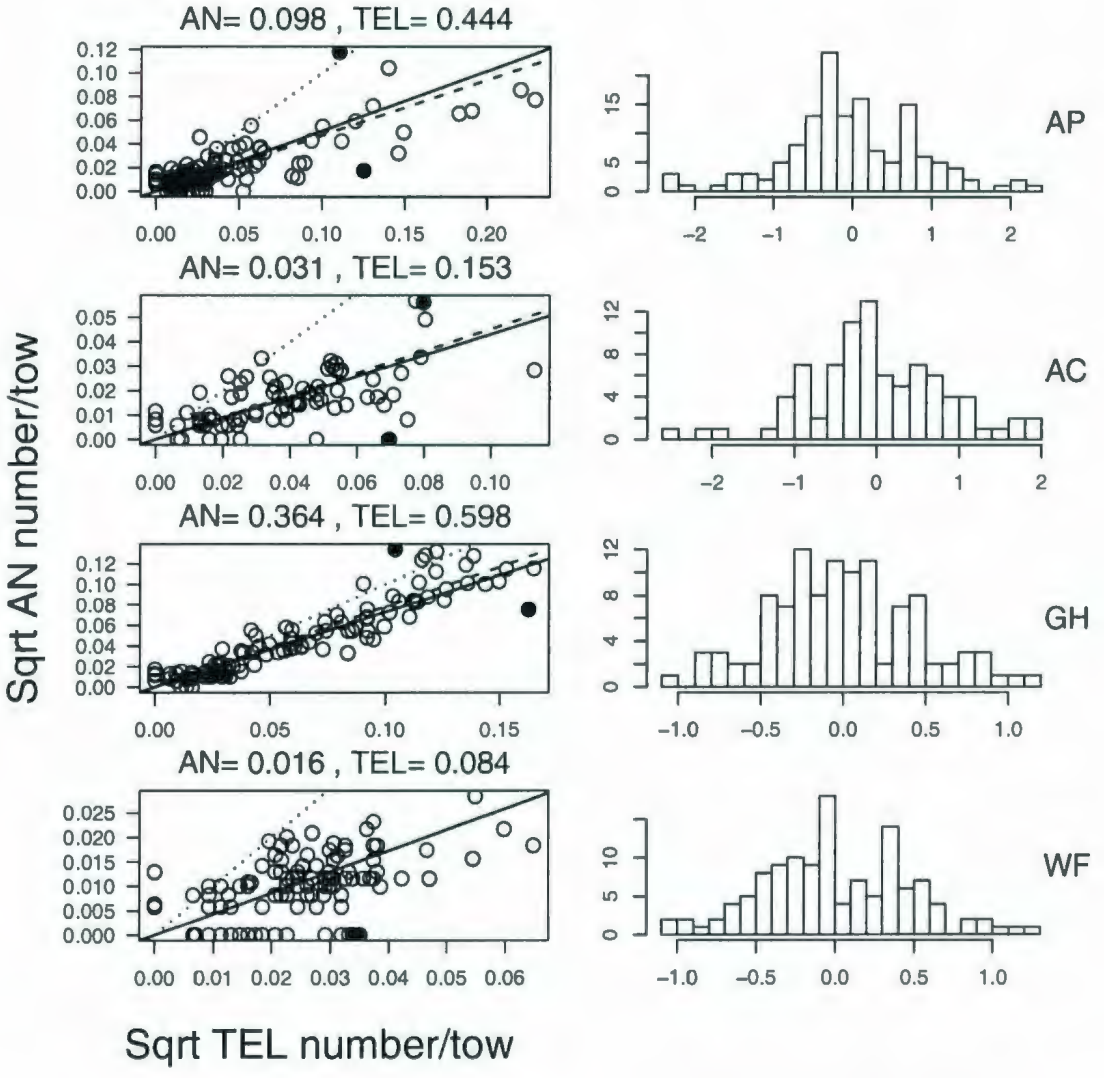


Figure 3.4: Left column: Total scaled catches from each haul, AN vs. TEL. The total per swept area for all sets are listed at the top. The dotted line has a slope of one. The dashed line has a slope equal to the relative efficiency ($\rho$) for the FEP1 model and the solid line denotes the mean relative efficiency for the MEP1 model. Solid black circles represent potential outliers. Right column: The predicted random effects histograms. The rows indicate each species with codes given in the right margin: AP - American plaice; AC - Atlantic cod; GH - Greenland halibut; WF - Witch flounder.

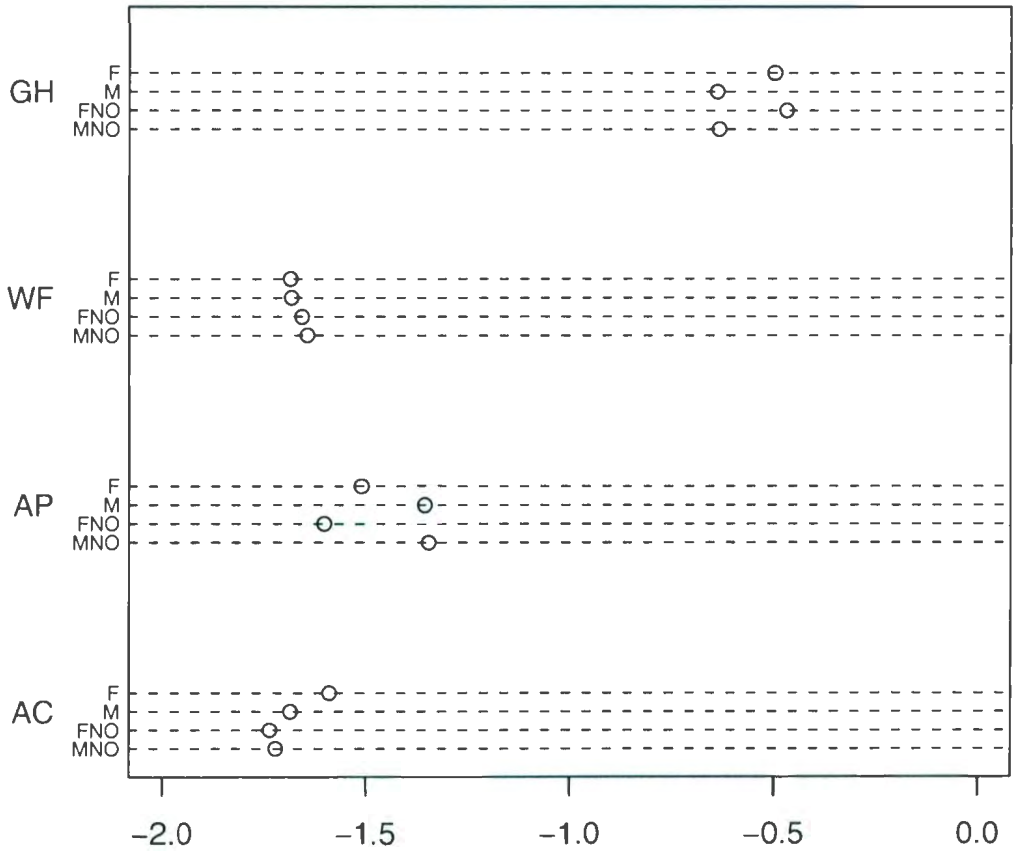


Figure 3.5: Estimates of $\hat{\beta}_0$ from the FEP1 and MEP1 models, and models with two potential outliers removed (FNO and MNO). Species codes: AP - American plaice; AC - Atlantic cod; GH - Greenland halibut; WF - Witch flounder.

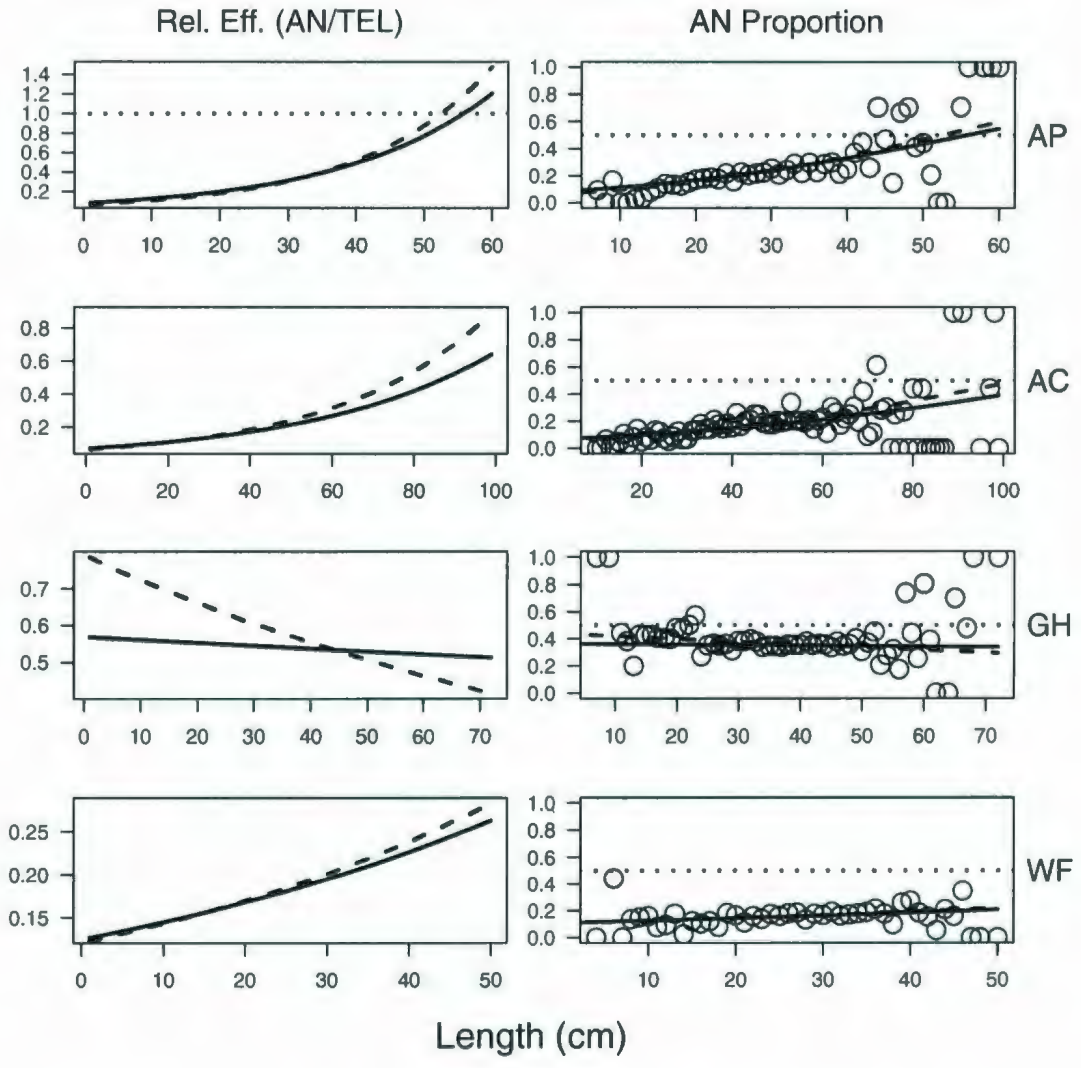


Figure 3.6: Left column: Estimates of relative efficiency ($\hat{\rho}_l$). Solid line represents the ME2 model estimate, dashed line represents FE2 model estimate. Right column: Observed (o's) and estimated (lines) proportions of AN scaled catches. Rows are for each species, with codes indicated in the right margin: AP - American plaice; AC - Atlantic cod; GH - Greenland halibut; WF - Witch flounder.

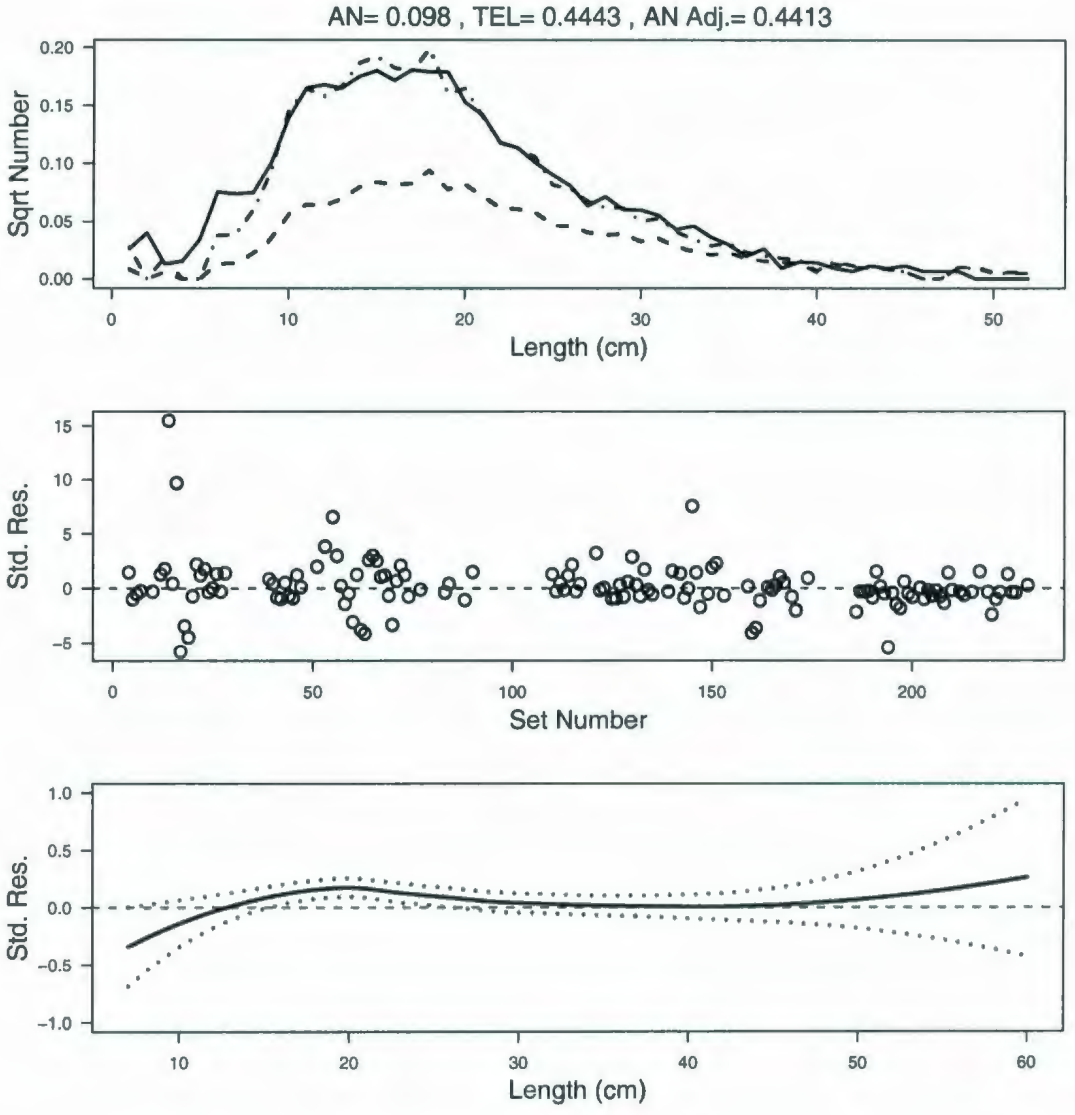


Figure 3.7: FE2 model results for American plaice. Total scaled catches per swept area for both vessels and AN catches per swept area adjusted by relative efficiency ($\rho_l$) are given at the top. Top panel: Total length frequencies for TEL (dashed-dotted line), AN (dashed line) and AN adjusted by relative efficiency (solid line), over all sets. Middle panel: Standardized (by standard deviation) total chi-square residuals for each set. Bottom panel: A local linear smoother versus length (solid line) of the standardized chi-square residuals. The dashed lines are 95% confidence intervals for the average residuals.

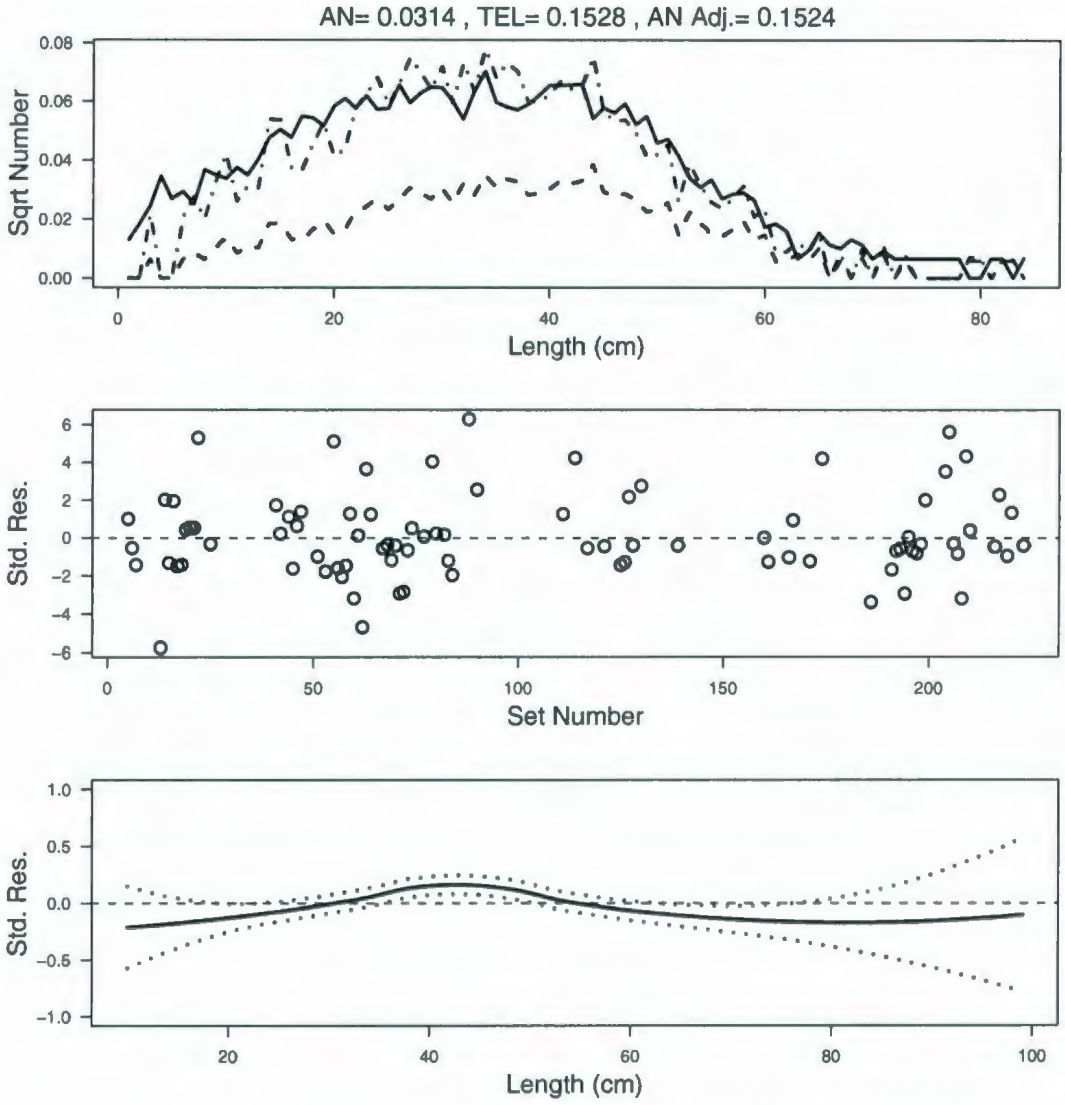


Figure 3.8: FE2 model results for Atlantic cod. Total scaled catches per swept area for both vessels and AN catches per swept area adjusted by relative efficiency ($\rho_l$) are given at the top. Top panel: Total length frequencies for TEL (dashed-dotted line), AN (dashed line) and AN adjusted by relative efficiency (solid line), over all sets. Middle panel: Standardized (by standard deviation) total chi-square residuals for each set. Bottom panel: A local linear smoother versus length (solid line) of the standardized chi-square residuals. The dashed lines are 95% confidence intervals for the average residuals.

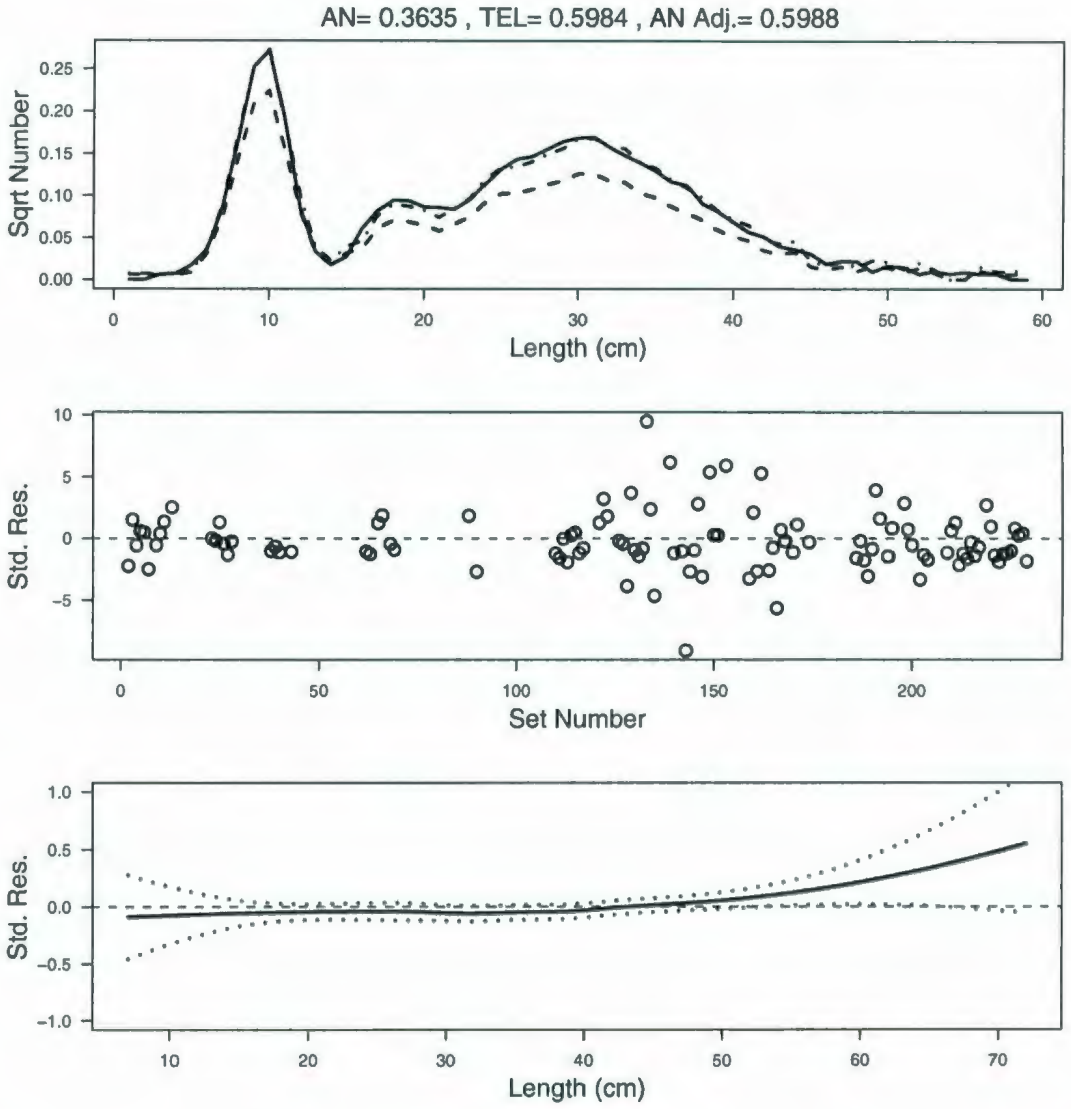


Figure 3.9: FE2 model results for Greenland halibut. Total scaled catches per swept area for both vessels and AN catches per swept area adjusted by relative efficiency ($\rho_l$) are given at the top. Top panel: Total length frequencies for TEL (dashed-dotted line), AN (dashed line) and AN adjusted by relative efficiency (solid line), over all sets. Middle panel: Standardized (by standard deviation) total chi-square residuals for each set. Bottom panel: A local linear smoother versus length (solid line) of the standardized chi-square residuals. The dashed lines are 95% confidence intervals for the average residuals.

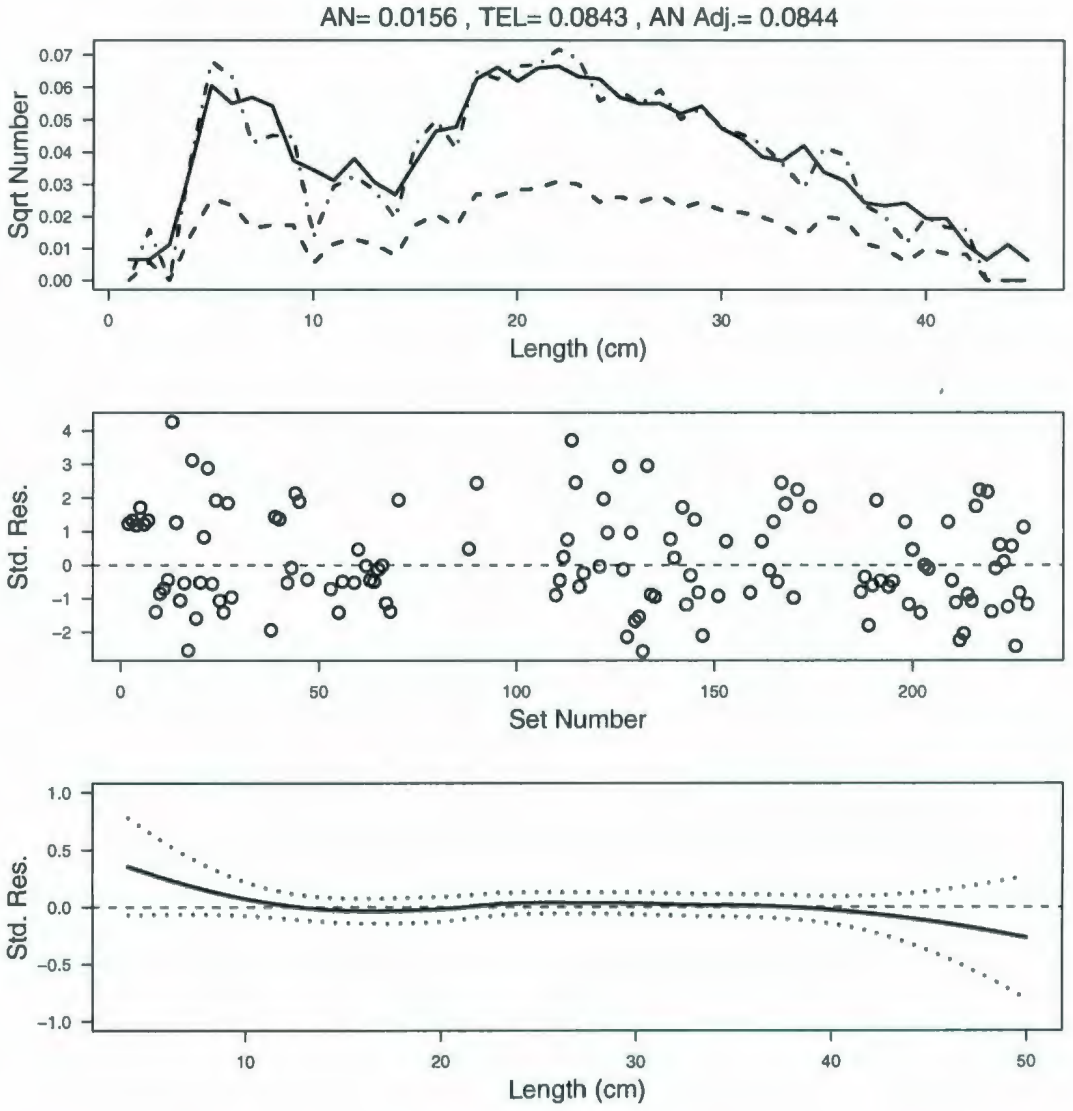


Figure 3.10: FE2 model results for Witch flounder. Total scaled catches per swept area for both vessels and AN catches per swept area adjusted by relative efficiency ($\rho_l$) are given at the top. Top panel: Total length frequencies for TEL (dashed-dotted line), AN (dashed line) and AN adjusted by relative efficiency (solid line), over all sets. Middle panel: Standardized (by standard deviation) total chi-square residuals for each set. Bottom panel: A local linear smoother versus length (solid line) of the standardized chi-square residuals. The dashed lines are 95% confidence intervals for the average residuals.

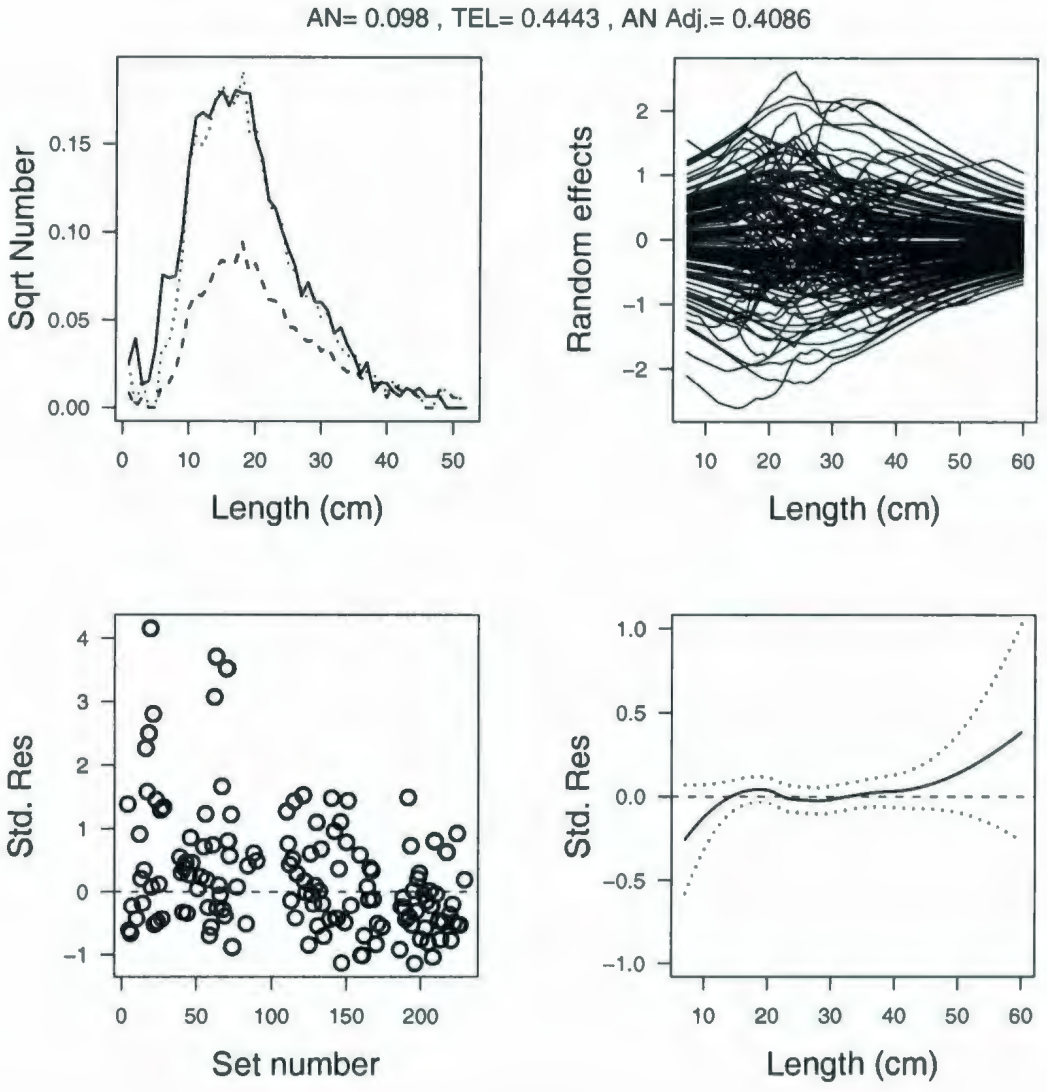


Figure 3.11: ME2 model results for American plaice. Total scaled catches per swept area for both vessels and AN catches per swept area adjusted by relative efficiency ($\rho_l$) are given at the top. Top left panel: Total length frequencies for TEL (dashed-dotted line), AN (dashed line) and AN adjusted by relative efficiency (solid line), over all sets. Top right panel: Predicted random effects, $\hat{\delta}_{il}$, vs length for each set. Bottom left panel: Standardized (by standard deviation) total chi-square residuals for each set. Bottom right panel: A local linear smoother versus length (solid line) of the standardized chi-square residuals. The dashed lines are 95% confidence intervals for the average residuals.

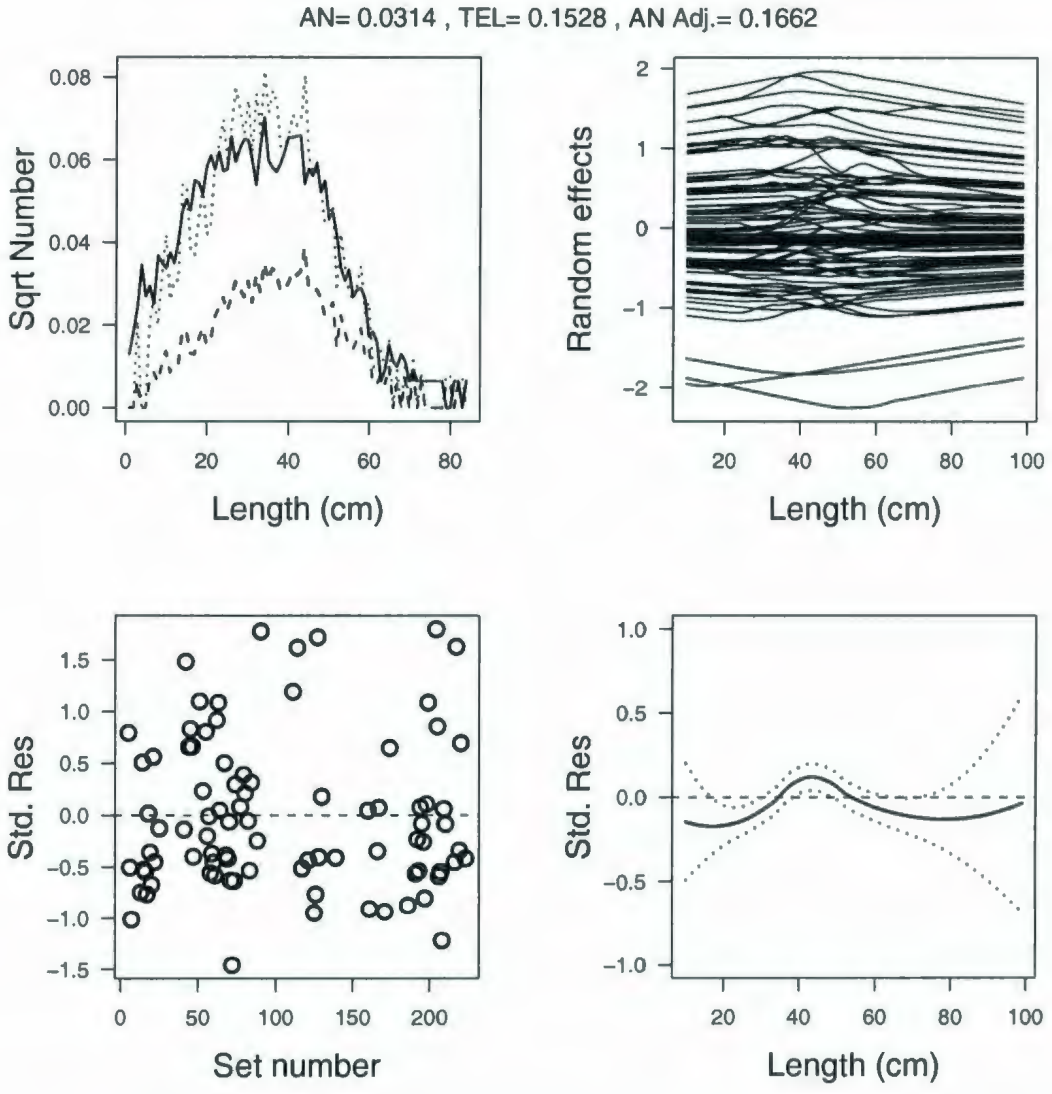


Figure 3.12: ME2 model results for Atlantic cod. Total scaled catches per swept area for both vessels and AN catches per swept area adjusted by relative efficiency ($\rho_l$) are given at the top. Top left panel: Total length frequencies for TEL (dashed-dotted line), AN (dashed line) and AN adjusted by relative efficiency (solid line), over all sets. Top right panel: Predicted random effects, $\hat{\delta}_{il}$, vs length for each set. Bottom left panel: Standardized (by standard deviation) total chi-square residuals for each set. Bottom right panel: A local linear smoother versus length (solid line) of the standardized chi-square residuals. The dashed lines are 95% confidence intervals for the average residuals.

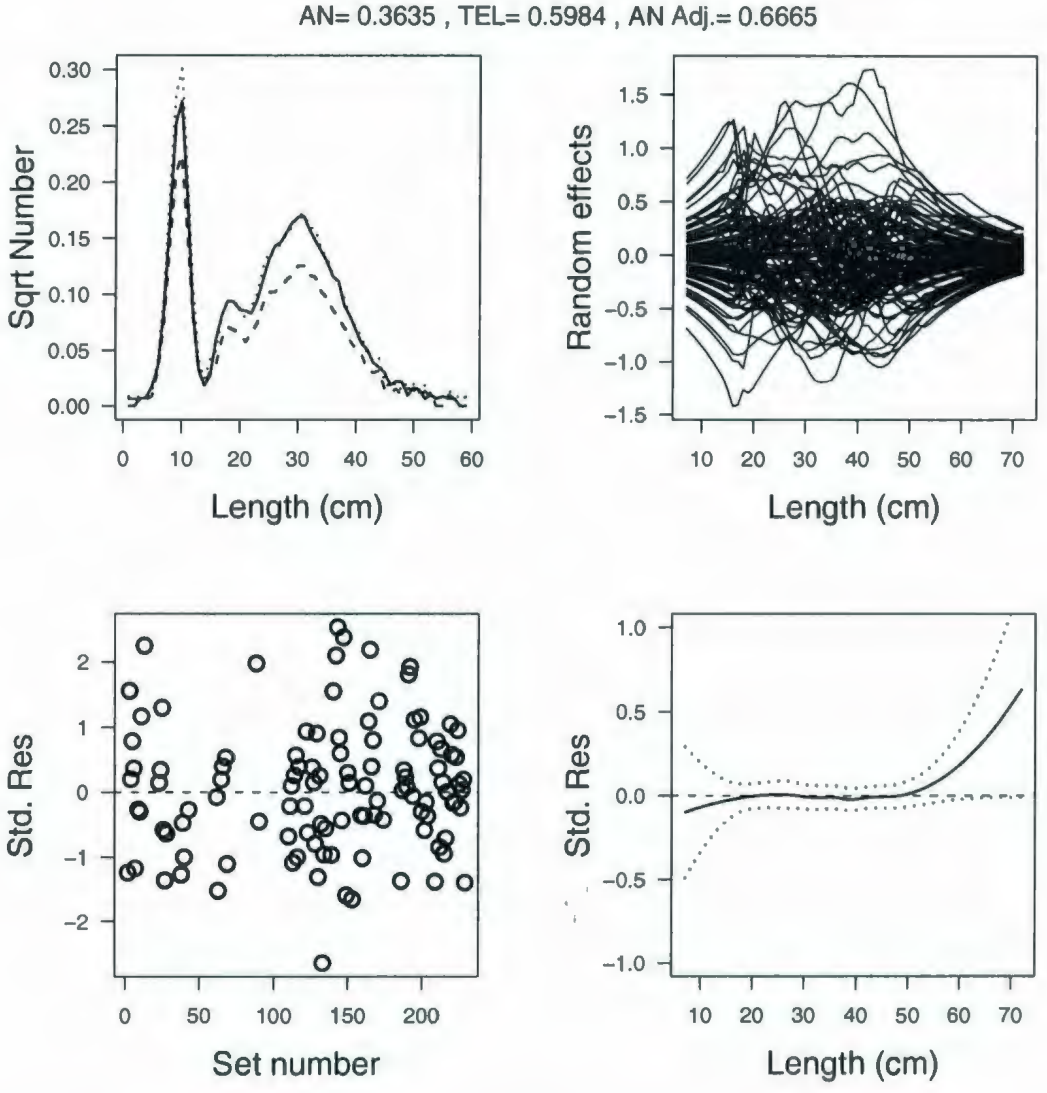


Figure 3.13: ME2 model results for Greenland halibut. Total scaled catches per swept area for both vessels and AN catches per swept area adjusted by relative efficiency ($\rho_l$) are given at the top. Top left panel: Total length frequencies for TEL (dashed-dotted line), AN (dashed line) and AN adjusted by relative efficiency (solid line), over all sets. Top right panel: Predicted random effects, $\hat{\delta}_{il}$, vs length for each set. Bottom left panel: Standardized (by standard deviation) total chi-square residuals for each set. Bottom right panel: A local linear smoother versus length (solid line) of the standardized chi-square residuals. The dashed lines are 95% confidence intervals for the average residuals.

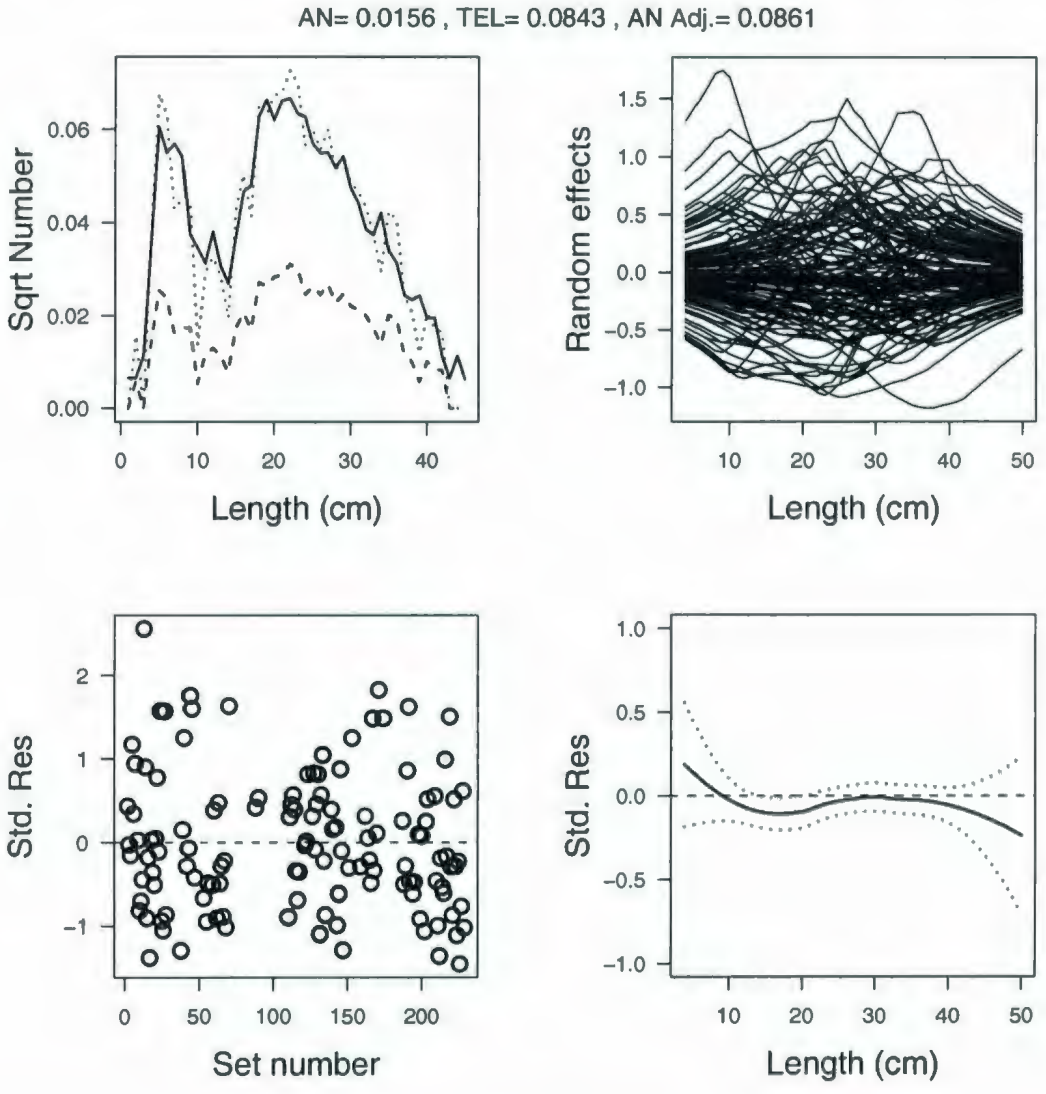


Figure 3.14: ME2 model results for Witch flounder. Total scaled catches per swept area for both vessels and AN catches per swept area adjusted by relative efficiency ($\rho_l$) are given at the top. Top left panel: Total length frequencies for TEL (dashed-dotted line), AN (dashed line) and AN adjusted by relative efficiency (solid line), over all sets. Top right panel: Predicted random effects, $\hat{\delta}_{il}$, vs length for each set. Bottom left panel: Standardized (by standard deviation) total chi-square residuals for each set. Bottom right panel: A local linear smoother versus length (solid line) of the standardized chi-square residuals. The dashed lines are 95% confidence intervals for the average residuals.

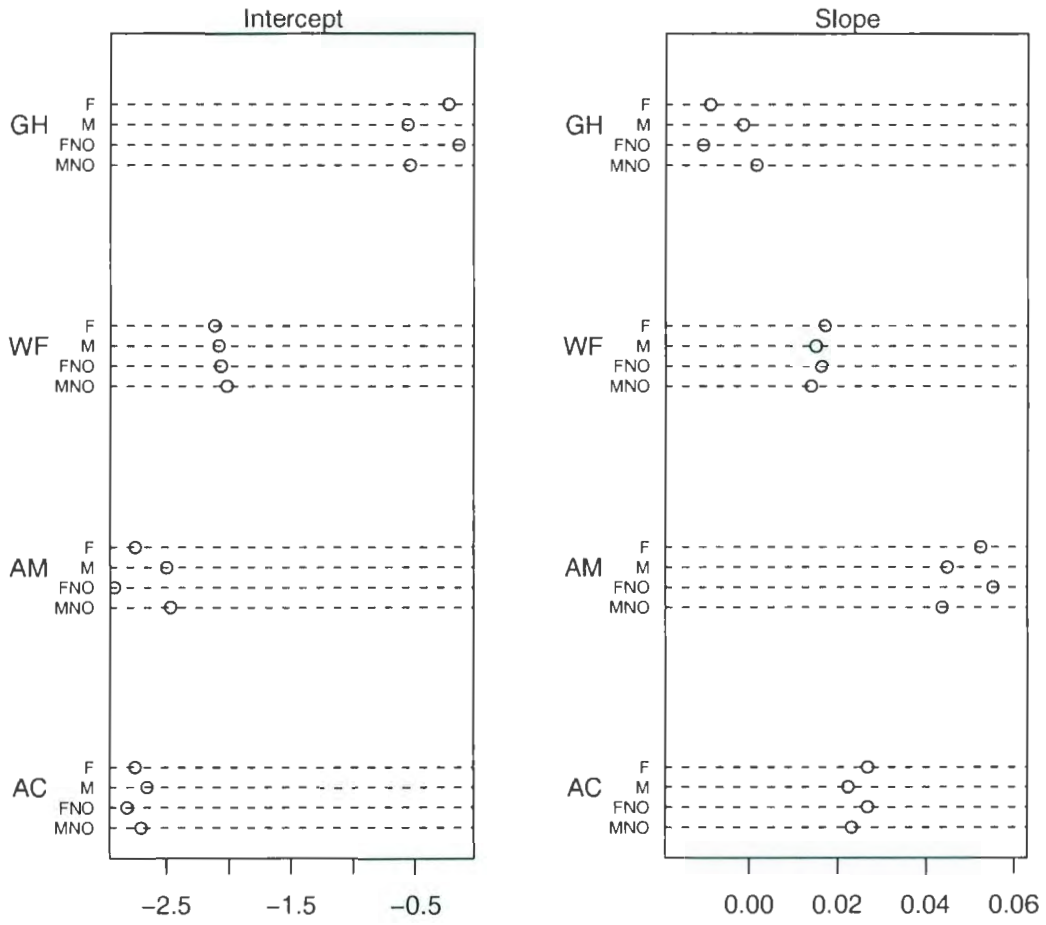


Figure 3.15: Estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ from the FE2 (F) and ME2 (M) models with two potential outliers removed (FNO and MNO). Species codes: American plaice - AM; Atlantic cod - AC; Greenland halibut - GH; WF - Witch flounder.

# Chapter 4

# Simulation Study

## 4.1 Introduction

Simulation studies were conducted to evaluate the reliability of parameter estimates and associated confidence intervals for the log relative efficiency parameter ($\beta_0$) in the fixed effects and mixed effects models. Our simulations were based on pooled models (FEP1 and MEP1) from Chapter 3 on comparative fishing.

Estimation of the MEP1 model using the conditional pseudo-likelihood approach (Wolfinger and O'Connell, 1993; PROC GLIMMIX), is denoted as MEP1c . In many cases marginal estimation methods are commonly used with mixed effects models (Zeger and Liang, 1986). These models estimate parameters by maximizing the true likelihood, integrated over the random effects, which is different from the conditional approach. We implement the marginal approach using PROC NLMIXED. Estimation of the MEP1 model using the marginal approach is denoted as MEP1m. In this chapter we compare the sensitivity and robustness of its parameter estimates to the FEP1 and MEP1c estimates.

## 4.2 Simulation Set-up

We considered three fish stocks, Atlantic cod, deepwater redfish and Greenland halibut on which to base our simulations; that is, the sample size (number of sets and total catch per set) was the same as from real comparative fishing data between the two research vessels, Alfred Needler (AN) and Wilfred Templeman (WT). This data was collected from the 2005 DFO Spring and Fall bottom-trawl surveys in NAFO Divisions 3LN and Subdivison 3Ps (Fig. 2.1, Chapter 2). We used the total number of catches per tow, $N$, and randomly generated catches for the control vessel (i.e. WT), $N_c$. The control vessel catch is conditionally a binomially distributed random variable. The probability that a captured fish is taken by the control vessel is

$$p_c = \frac{e^{\beta_0+\delta}}{1+e^{\beta_0+\delta}}. \tag{4.1}$$

In the simulation, $\beta_0$, the true fixed effect log relative efficiency parameter, ranged from 0 to 2 in steps of 0.25 and $\delta$, the random effect, was normally distributed with a constant variance, $\sigma^2$. We chose $\sigma^2 = 0.1, 0.5.$, and 0.9 which represented typical values of the random effect variances (see Cadigan et al., 2006). We also chose $\sigma^2 = 0$ to check how well the mixed effects model performed when random effects did not exist (i.e. fixed effects model). The simulation had 36 factors per species in total, nine levels for $\beta_0$ times four levels for $\sigma^2$.

For each simulation, we generated $K = 2000$ sets of data for each species, estimated the parameters under the various models, and computed the bias for $\beta_0$ for each factor combination ($\beta_0$s and $\sigma^2$s) and for each model. The bias is computed as

$$Bias(\hat{\beta}_0) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\beta}_{0k} - \beta_0), \tag{4.2}$$

where $\beta_0$ is the true parameter value and $\hat{\beta}_{0k}$ is the parameter estimate of $\beta_0$ from

the $k^{th}$ simulation iteration for each model. We examined the coverage errors for 95% nominal confidence intervals provided by PROC GENMOD and PROC GLIMMIX.

We used the profile likelihood confidence intervals (e.g. Venzon and Moolgavkar, 1988) for the GLIM parameters provided by PROC GENMOD. Suppose $\beta = (\beta_1, ..., \beta_p)$ is a $p \times 1$ vector of fixed parameters. A profile-likelihood confidence interval for $\beta_i$ is as follows. Let

$$\ell^*(\beta_i) = \max_{\tilde{\beta}} \ell(\beta) \tag{4.3}$$

where $\tilde{\beta}$ is the vector $\beta$ with the $i^{th}$ element fixed at $\beta_i$ and $\ell$ is the log-likelihood function. If $\ell(\hat{\beta})$ is the log-likelihood evaluated at the maximum likelihood estimate $\hat{\beta}$, then $2(\ell(\hat{\beta}) - \ell^*(\beta_i))$ has a limiting $\chi^2$ distribution with one degree of freedom if $\beta_i$ is the true parameter. A $(1-\alpha)\times 100\%$ confidence interval for $\beta_i$ is

$$\{\beta_i : \ell^*(\beta_i) \geq \ell_0 = \ell(\hat{\beta}) - 0.5\chi^2_{1-\alpha,1}\}. \tag{4.4}$$

PROC GENMOD finds the endpoints of the confidence intervals numerically. This is achieved by starting at the maximum likelihood estimate of $\beta$ and approximating the log-likelihood with a quadratic surface, for which an exact solution is possible. The process is iterated until convergence to an endpoint is attained. The process is repeated for the other endpoint.

PROC GLIMMIX produces Wald-type confidence intervals for the parameter estimates of a GLMM. The $(1-\alpha)\times 100\%$ Wald-type confidence interval for parameter $\beta_i$ is defined as

$$\hat{\beta} \pm z_{(1-\alpha/2)} SE(\hat{\beta}_i) \tag{4.5}$$

where $z_{(1-\alpha/2)}$ is the $100(1-\alpha/2)$th percentile from a standard normal distribution, $\hat{\beta}$,

is a fixed parameter estimate, and $SE(\hat{\beta}_t)$ is the estimate of its standard error. PROC GLIMMIX produces parameter estimates for GLMMs using pseudo-likelihood or restricted pseudo-likelihood estimation (details given in Chapter 1). It does not provide profile likelihood confidence intervals because it does not use the exact likelihood for most mixed models (recall that the likelihood can be computationally prohibitive for many mixed models).

Since we know the true $\beta_0$ values in each simulation, we would expect about 95% of the confidence intervals (CI) we construct to contain the true parameter value.

We also investigated the power of the test

$$\begin{aligned} H_0 &: \beta_0 = 0 \\ H_a &: \beta_0 \neq 0. \end{aligned} \tag{4.6}$$

We did this as follows: for each simulated dataset, we constructed a 95% CI for $\beta_0$. If this CI did not include 0, we rejected $H_0$. We counted how many times we rejected $H_0$ among our 2000 datasets within a particular factor combination (i.e. $\beta_0$ and $\sigma^2$) and computed the power of the test (i.e. power curve)

$$P = \frac{\Omega}{2000}, \tag{4.7}$$

where $\Omega$ is the number of times $H_0$ is rejected. We did this for each of the 36 factors in the simulation study.

We also investigated the robustness of our model with respect to the assumption of the distribution of the random effects ($\delta$). It is reasonable to assume that densities sampled from both the test and control vessels are independent and each follows a gamma distribution with mean $\mu$ and variance $\phi\mu^2$:

$$\lambda_i \sim Gamma\left(\frac{1}{\phi}, \phi\mu\right), \quad i = 1, 2. \tag{4.8}$$

It then follows that $w_i = \log[(\phi\mu)^{-1}\lambda_i]$ is log-gamma (e.g. Lawless, 1980) with density

$$f(w_i) = \frac{1}{\Gamma(\frac{1}{\phi})} exp\left(\frac{w_i}{\phi} - e^{w_i}\right), \quad i = 1, 2. \tag{4.9}$$

An appropriate probabilistic model for $\delta$ is the difference of two log-gamma densities ($\delta = w_1 - w_2$) which has a mean and variance that depends only on $\phi$. In Appendix A we derive the distribution of $w_i$ and present the mean and variance of $\delta$. Table 4.1 presents the values of $\phi$ that give us means and variances corresponding to the normal case. Histograms and probability plots of $\delta$ display distributions that are only slightly skewed (see Fig. 4.1).

We repeated the simulation study using the difference of two log-gamma distributed random variables for $\delta$ with choices of $\phi$ given in Table 4.1. We then estimated $\beta_0$ and $\sigma^2$ under the (incorrect) assumption that $\delta$ was normally distributed.

## 4.3 Results

Tables 4.2 and 4.3 present summary statistics (pooled over lengths) of total catch per set for the control and for the totals from the control and test vessels, respectively. The mean catch per tow was largest for deepwater redfish compared to Atlantic cod and Greenland halibut for both the control and total vessels. The between tow variance of catches for each species were quite large for both vessels.

### 4.3.1 Analysis with Normal Distributed Random Effects

In this analysis catches were pooled over lengths within sets. The bias of $\hat{\beta}_0$ is given in Fig. 4.2, plotted against the nine fixed $\beta_0$ values used to generate the simulation

data. The bias was extremely close to zero for all models and species when random effects were non-existent ($\sigma^2 = 0$). When random effects existed, the bias for the FEP1 model (solid line) was negative for most values of $\beta_0$ but was zero when $\beta_0 = 0$. This indicated that the FEP1 model underestimated most values of $\beta_0$. The bias for each of the three species was worse as each level of $\sigma^2$ and $\beta_0$ increased. The bias for the MEP1c (dotted line) was negative for most values of $\beta_0$ and got worse as the range of $\beta_0$ and $\sigma^2$ values increased for each species. The MEP1c model appeared to underestimate $\beta_0$, although the bias was smaller compared to the FEP1 model. The bias for the MEP1m (dash-dotted line) was very small in magnitude and in some cases equal to zero (i.e. Atlantic cod when $\sigma^2 = 0.5$). The MEP1m model had little to no trend over all values $\beta_0$ and $\sigma^2$. This showed that the MEP1m model estimated $\beta_0$ quite accurately.

The 95% coverage errors of confidence intervals for $\beta_0$ are presented in Figs. -. The solid line represents the lower 95% simulated coverage error, the dotted line represents the upper 95% simulated coverage error, and the dash-dotted line represents the total (upper + lower) simulated coverage errors. Coverage errors were reasonably close to the expected levels for all models and species when random effects were non-existant (see Fig. ), although the confidence intervals obtained from PROC GLIMMIX performed somewhat worse. Total coverage errors produced by the FEP1 model were large ($>> 0.05$) for each species and increased for larger values of $\sigma^2$ (Figs. - ). The total coverage errors for the MEP1c model were closer to the expected critical value of 0.05, especially for deepwater redfish. Atlantic cod and Greenland halibut produced somewhat larger coverage errors for larger values of $\beta_0$. Coverage errors for the MEP1m model were close to the expected levels for most species and values of $\sigma^2$ with the only exceptions being Atlantic cod (Fig. ; row 1, column 3) and Greenland halibut (Fig. ; row 3, column 3) when $\sigma^2 = 0.1$. In these cases, the coverage errors for species increased as values of $\beta_0$ increased.

Fig. displays 95% confidence interval (CI) widths of $\beta_0$ for each model. The CI widths were approximately equal for all models and species when random effects

did not exist. FEP1 CI widths (solid line) were smaller over the range of $\beta_0$ values as compared to the MEP1c (dotted line) and MEP1m (dash-dotted line) for all species when random effects existed. CI widths for the MEP1c and MEP1m models were very similar for all species and $\sigma^2$. The largest differences in CI widths occured for Atlantic cod and Greenland halibut when $\sigma^2 = 0.9$ (Fig. ; row 1 column 4, and row 3 column 4, respectively).

Power curves are presented in Fig. . The power of the test (Eq. ) was equivalent for each model when random effects did not exist (Fig. ; column 1). For the FEP1 model (solid line), the power approached 100% when the true value of $\beta_0 \geq 0.5$. Power curves for all three models were similar when $\sigma^2 = 0.1$. The power of the test tended to be weaker for FEP1 as $\sigma^2$ increased. This is false power however, because at the $\beta = 0$ case, the FEP1 model concluded $\beta_0 \neq 0$ up to 40% of the time depending on the species and $\sigma^2$ (i.e. Atlantic cod when $\sigma^2 = 0.9$). This should only happen 5% of the time, and this indicates that it is not fair to compare the type II error rates because the type I error rate of the FEP1 test is quite different than the ME tests.

### 4.3.2 Analysis When Random Effects Follow a Difference of Two Log-Gamma Random Variables

Results for bias (Fig. ), 95% coverage error (Figs. - ), 95% CI widths (Fig. ) and power curves (Fig. ) are presented when the random effects are the difference of two log-gamma random variables (see Section 1.2). The results for all models were very similar to the previous case when the random effects were normally distributed.

## 4.4 Discussion

Results from the simulation study suggested that, overall, mixed effects models appeared to perform better than fixed effects models in terms of bias, coverage errors, and power of the test for rejecting $\beta_0 = 0$, both when random effects were normally distributed and when the random effects were a difference of two log-gamma random variables. When random effects were present, and their variance was not small, the fixed effects GLIM model performed poorly, with substantial bias in estimates of log relative efficiency and poor confidence intervals for this parameter. Mixed effects models performed equally as well when random effects did not exist.

Maximum likelihood estimation based on the marginal likelihood for mixed models (e.g. PROC NLMIXED) yielded results that were less biased than the conditional estimation methods of PROC GLIMMIX. However, for some models we are interested in (see previous chapters) marginal methods are not feasible and it would be desirable to reduce the bias in conditional estimators. Kuk (1995) and Lin and Breslow (1996), among others, provided asymptotic bias corrections for regression parameters and variance component estimates in GLMMs. Other possible bias reduction techniques could include re-sampling techniques such as the jackknife and bootstrap (see Efron, 1982; Wu, 1986, Efron and Tibshirani, 1993), or Taylor series expansion (see Box, 1971; Cook et al., 1986; Cordeiro and McCullagh, 1991). Each technique is useful for reducing bias in parameter estimates but there are limitations. An investigation of bias reduction techniques is worthwhile, but beyond the scope of this practicum.

The fixed, conditional, and marginal approaches for un-pooled simulations were also examined. In this set-up, lengths for each species were not pooled over each set. Random effects were the same for each length within a set, but differed between sets. Random effects were both normally distributed and a difference of two log-gamma random variables. We examined purely fixed effects models (FE2), mixed effects models

with random intercepts (ME2RI), and mixed effects models with random autocorrelated length effects within each set (ME2AR). Conditional pseudo-likelihood methods were used for both the random intercept model and the autocorrelated random length effects model, denoted as ME2RIc and ME2ARc, respectively. The marginal approach was applied only to the random intercept model (ME2RIm) since PROC NLMIXED does not have an option to specify autocorrelation structures. This type of random effect is computationally too difficult for numerical integration and the marginal approach. Estimates for both the fixed intercept ($\beta_0$) and the fixed length ($\beta_1$) parameters were computed, but bias of the parameter estimates and coverage errors of confidence intervals were only considered for $\beta_0$. The results were similar to those of the pooled analysis except for ME2ARc, although, further analysis is needed to clarify this result.

Simulations were conducted to test the efficacy of the mixed model approach for estimating relative efficiency, although, simulations on the variances of the estimates would also be useful to fully understand the performance of these estimates. One method for reporting the performance of variance estimates is to compute the mean square errors (MSE; see Gunst and Mason, 1977). However, this was beyond the scope of this practicum.

Simulation studies to examine the robustness and sensitivity of the various methods to outliers in the data would also be useful. One such technique is to simulate datasets with and without outliers and compare parameter estimates from both simulations in terms of senstivity (magnitude of deviation) due to outliers. However, this was beyond the scope of this practicum.

Table 4.1: Values of the scale parameter ($\phi$) and the corresponding values for the mean (E[$\delta$]) and variance (Var[$\delta$]) of $\delta$.

| $\phi$ | E($\delta$) | Var($\delta$) |
|---|---|---|
| 0.049 | 0 | 0.1 |
| 0.223 | 0 | 0.5 |
| 0.372 | 0 | 0.9 |

Table 4.2: Catch summaries for the control vessel (WT) pooled over lengths. $n_c$ is the total number of tows per species, $\bar{n}_c$ is the mean catch per tow and $Var(n_c)$ is the between-tow variance of catch for each species. $N_c$ is the total number of catches over all tows.

| Species | Scientific Name | $n_c$ | $\bar{n}_c$ | $Var(n_c)$ | $N_c$ |
|---|---|---|---|---|---|
| Atlantic cod | *Gadus morhua* | 91 | 20.86 | 3125.45 | 1899 |
| Deepwater Redfish | *Sebastes mentella* | 63 | 86.81 | 9631.35 | 5496 |
| Greenland halibut | *Reinhardtius hippoglossoides* | 56 | 10.82 | 180.22 | 606 |

Table 4.3: Catch summaries for the both control and test vessels (WT + AN) pooled over lengths. $n_{c+t}$ is the total number of tows per species, $\bar{n}_{c+t}$ is the mean catch per tow and $Var(n_{c+t})$ is the between-tow variance of catch for each species. $N_{c+t}$ is the total number of catches over all tows.

| Species | Scientific Name | $n_{c+t}$ | $\bar{n}_{c+t}$ | $Var(n_{c+t})$ | $N_{c+t}$ |
|---|---|---|---|---|---|
| Atlantic cod | *Gadus morhua* | 91 | 43.14 | 8276.36 | 3926 |
| Deepwater Redfish | *Sebastes mentella* | 63 | 191.57 | 39934.93 | 12069 |
| Greenland halibut | *Reinhardtius hippoglossoides* | 56 | 24.27 | 844.23 | 1359 |

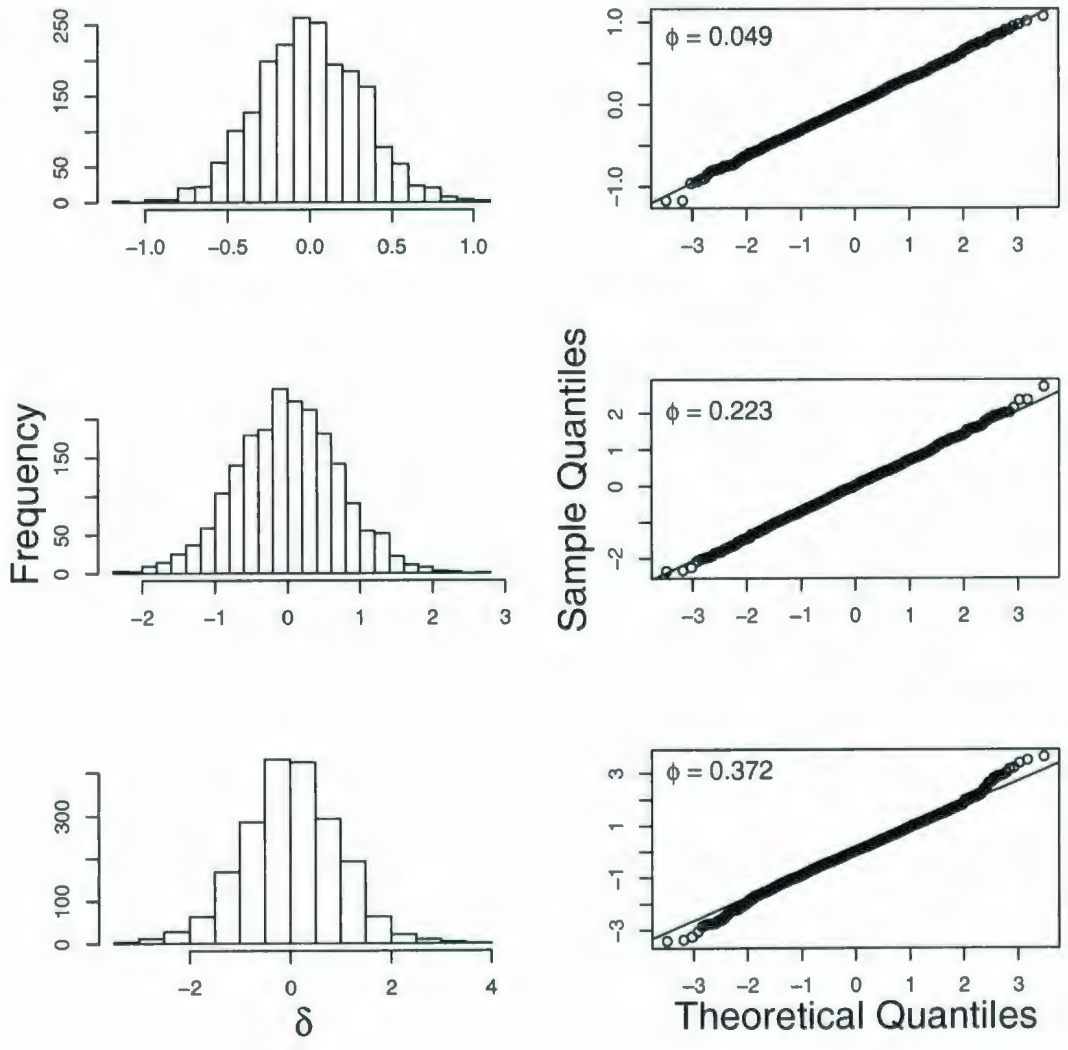


Figure 4.1: Right: histograms of the random effects, $\delta$, where $\delta = \log[(\mu\phi)^{-1}\lambda_1] - \log[(\mu\phi)^{-1}\lambda_2]$. Both $\lambda_1$ and $\lambda_2$ are gamma random variables with mean $\mu$ and variance $\phi\mu^2$. Left: probability plots for the random effects. Rows indicate random effects for corresponding $\phi$ values which are given in the upper left corner of the probability plot.

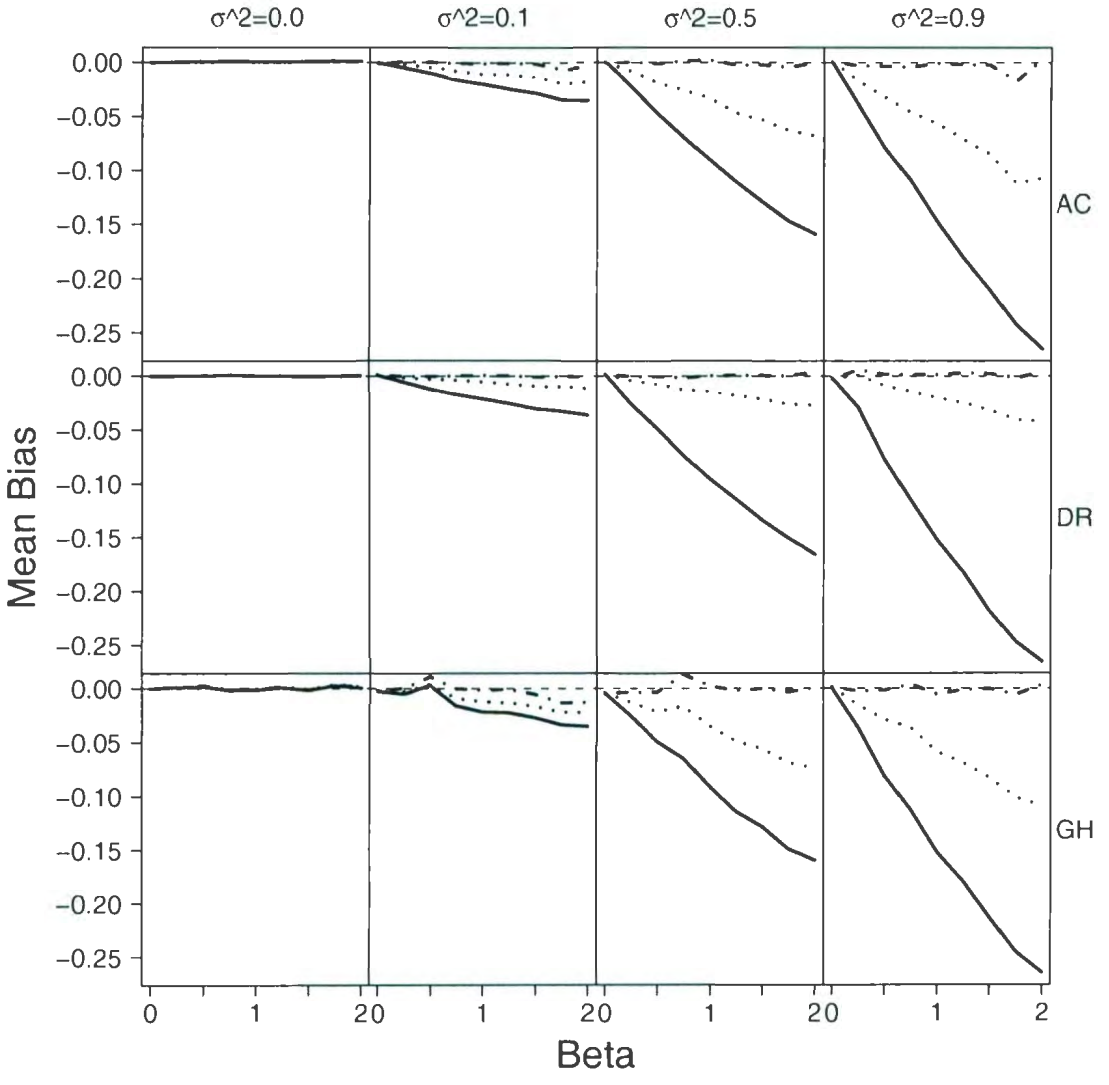


Figure 4.2: Bias of $\hat{\beta}_0$ for FEP1 (solid line), MEP1c (dotted line), and MEP1m (dash-dotted line) models. Random effects are normally distributed with 0 mean and variances $\sigma^2 = 0.0, 0.1, 0.5, 0.9$, respectively. The dashed line represents the horizontal line at 0. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

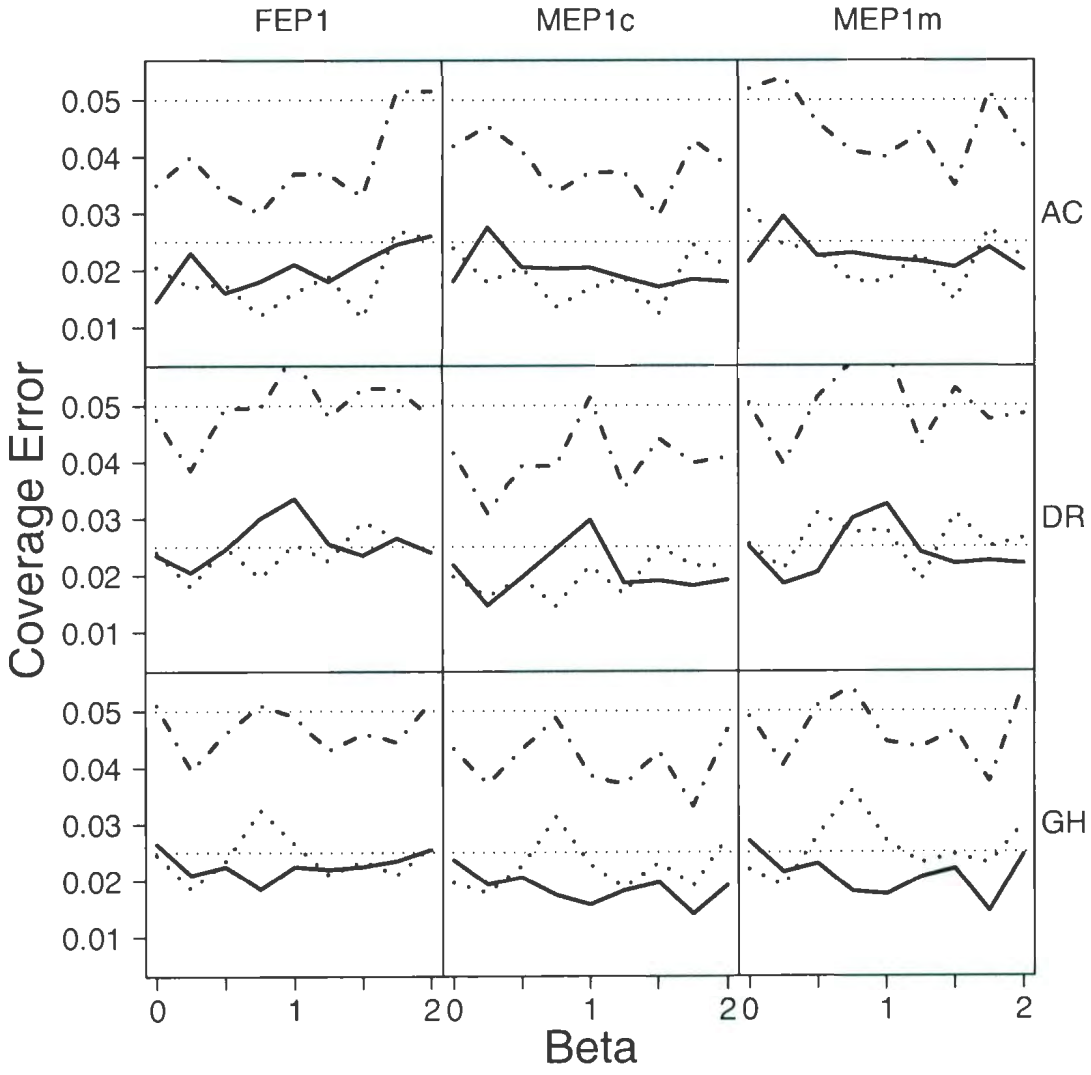


Figure 4.3: 95% coverage errors of the confidence intervals from the parameter estimates for FEP1, MEP1c, and MEP1m models when random effects do not exist. The solid line represents lower coverage errors, the dotted line represents upper coverage errors, the dash-dotted line represents total coverage errors (lower + upper), and the horizontal dotted lines represent critical values $\alpha = 0.05$ and $\frac{\alpha}{2} = 0.025$. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

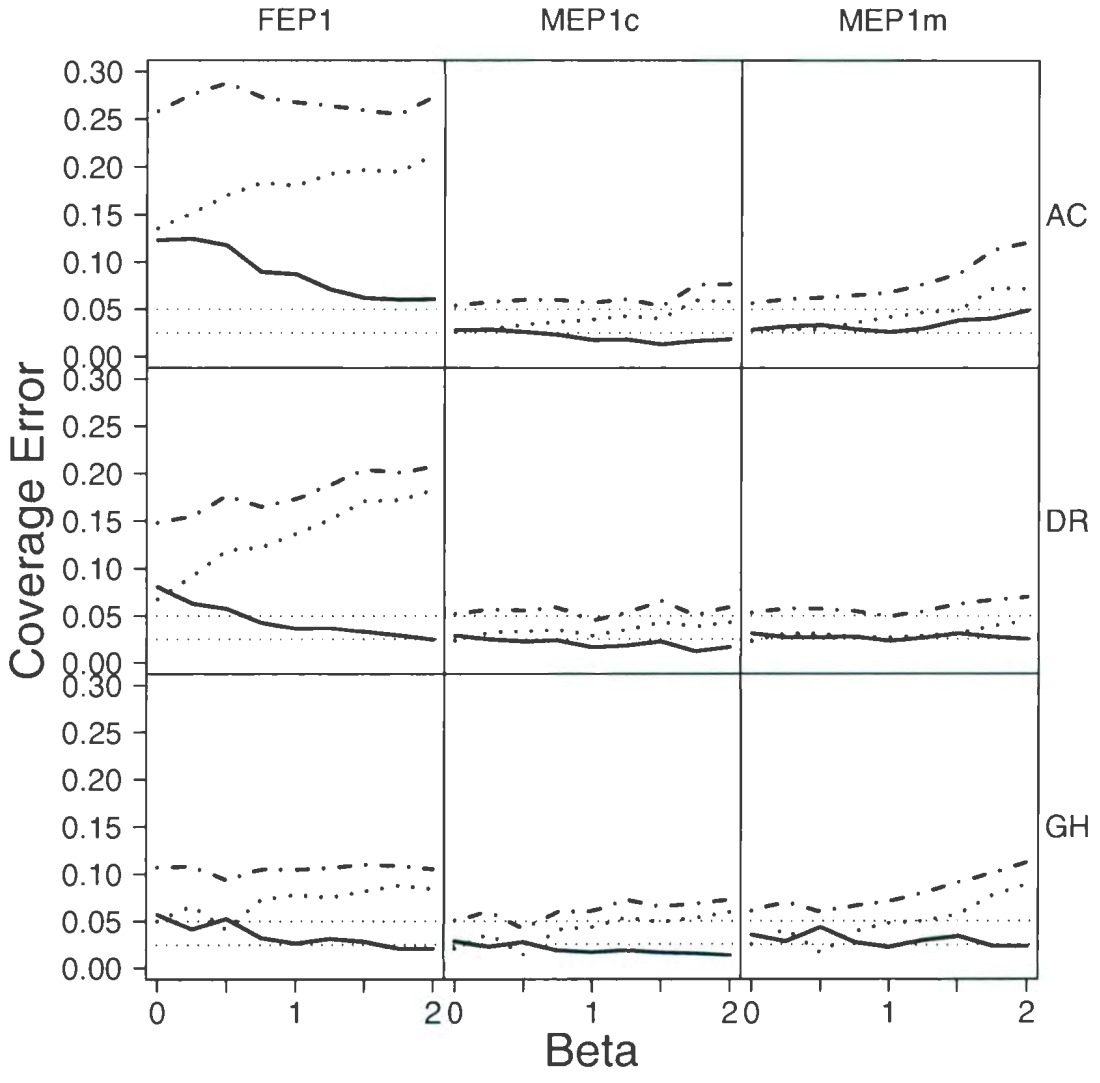


Figure 4.4: 95% coverage errors of the confidence intervals from the parameter estimates for FEP1, MEP1c, and MEP1m models. Random effects are normally distributed with 0 mean and variance $\sigma^2 = 0.1$. The solid line represents lower coverage errors, the dotted line represents upper coverage errors, the dash-dotted line represents total coverage errors (lower + upper), and the horizontal dotted lines represent critical values $\alpha = 0.05$ and $\frac{\alpha}{2} = 0.025$. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

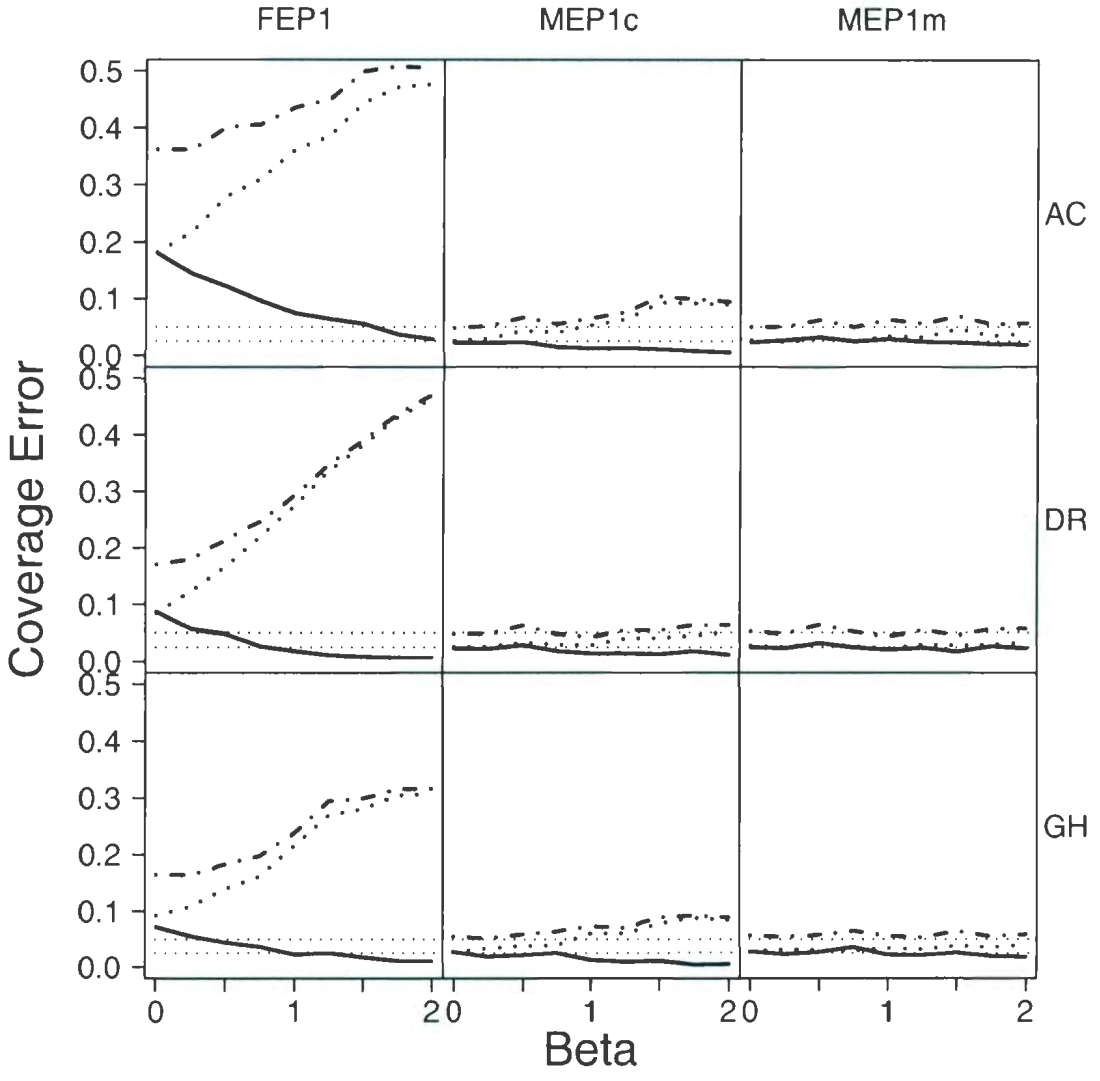


Figure 4.5: 95% coverage errors of the confidence intervals from the parameter estimates for FEP1, MEP1c, and MEP1m models. Random effects are normally distributed with 0 mean and variance $\sigma^2 = 0.5$. The solid line represents lower coverage errors, the dotted line represents upper coverage errors, the dash-dotted line represents total coverage errors (lower + upper), and the horizontal dotted lines represent critical values $\alpha = 0.05$ and $\frac{\alpha}{2} = 0.025$. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

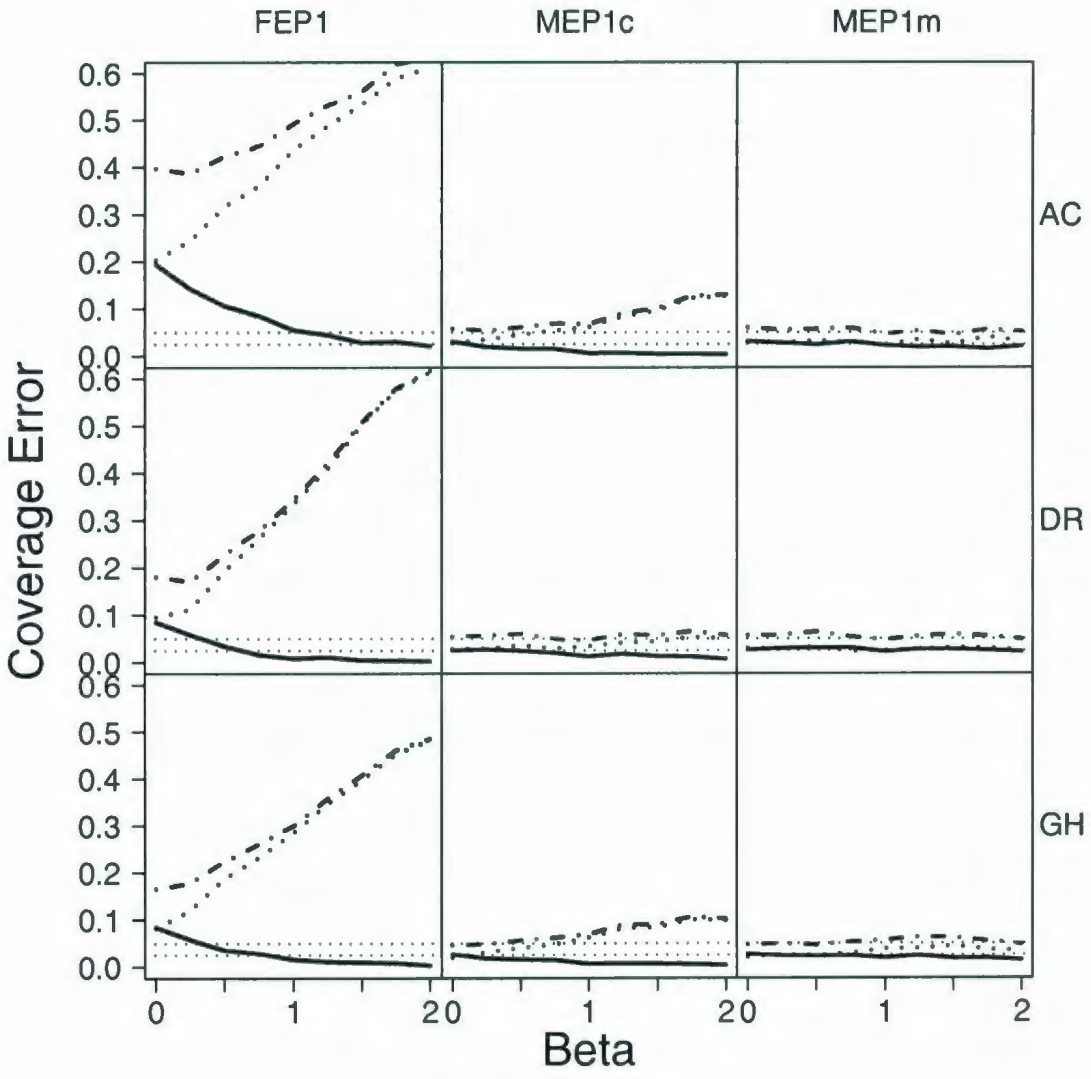


Figure 4.6: 95% coverage errors of the confidence intervals from the parameter estimates for FEP1, MEP1c, and MEP1m models. Random effects are normally distributed with 0 mean and variance $\sigma^2 = 0.9$. The solid line represents lower coverage errors, the dotted line represents upper coverage errors, the dash-dotted line represents total coverage errors (lower + upper), and the horizontal dotted lines represent critical values $\alpha = 0.05$ and $\frac{\alpha}{2} = 0.025$. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

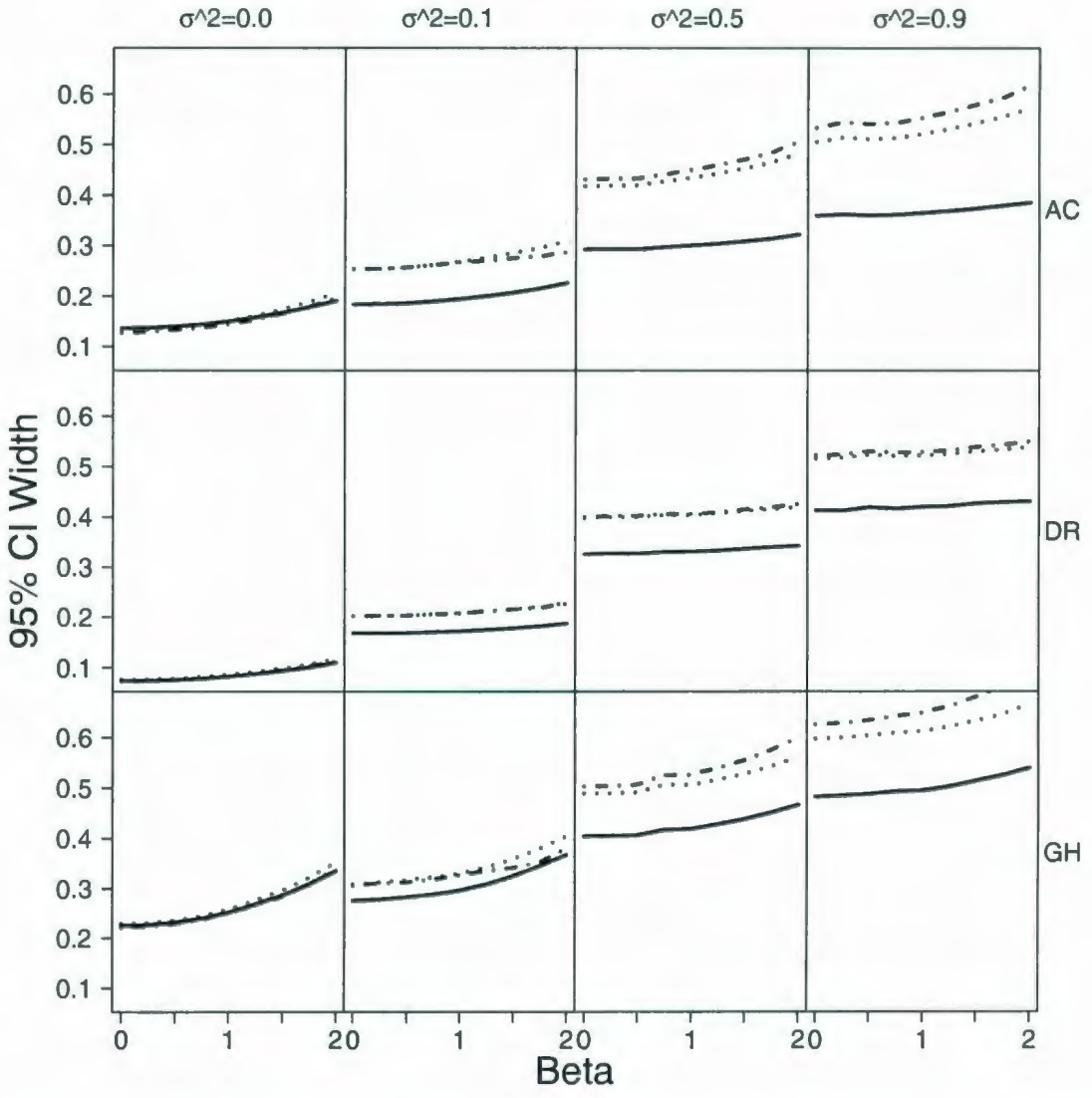


Figure 4.7: The 95% confidence widths of the parameter estimates for FEP1 (solid line), MEP1c (dotted line), and MEP1m models (dash-dotted line). Random effects are normally distributed with 0 mean and variances $\sigma^2 = 0.0, 0.1, 0.5, 0.9$, respectively. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

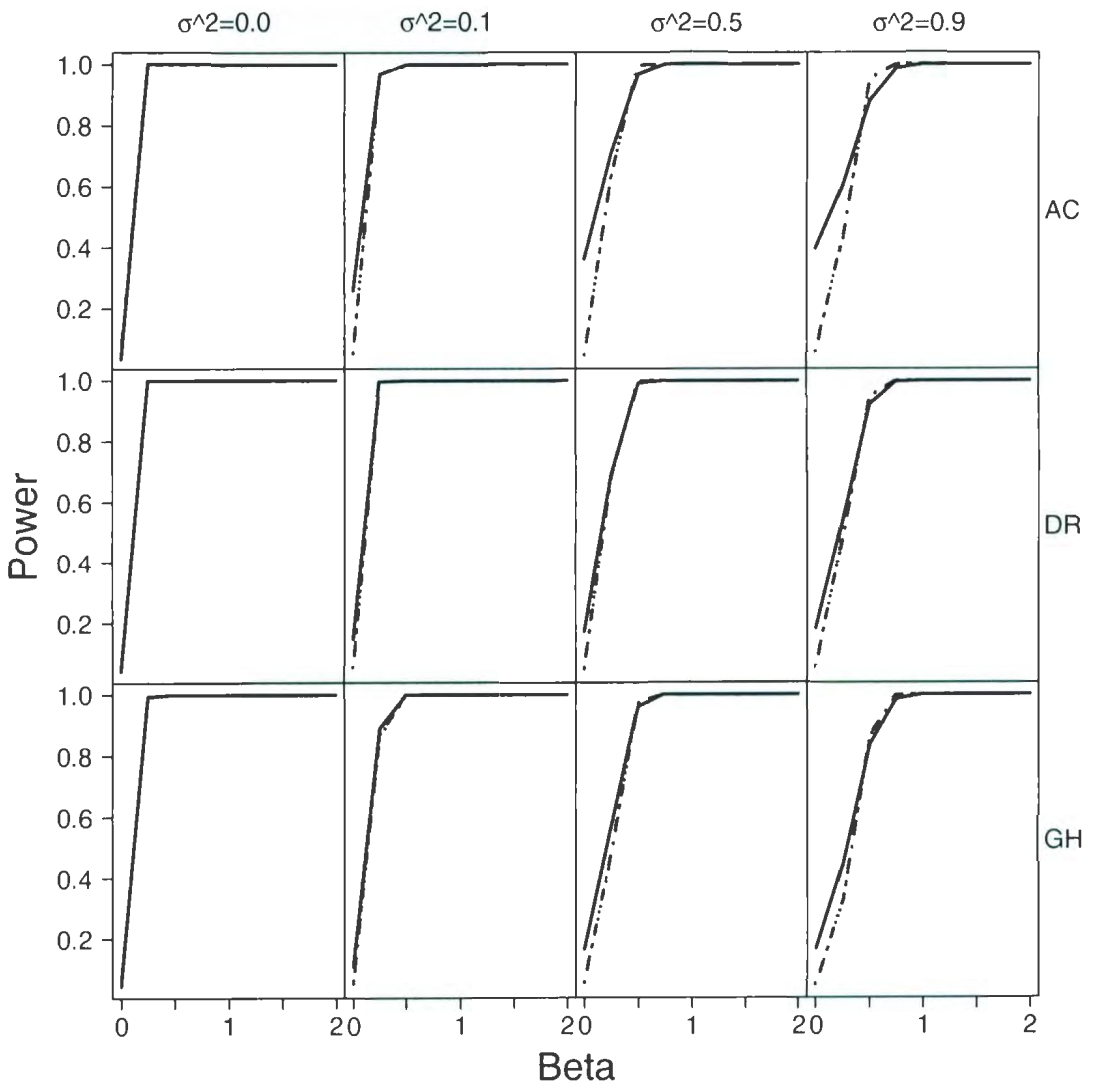


Figure 4.8: Power curves for FEP1 (solid line), MEP1c (dashed line), and MEP1m (dash-dotted) with normally distributed random effects. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

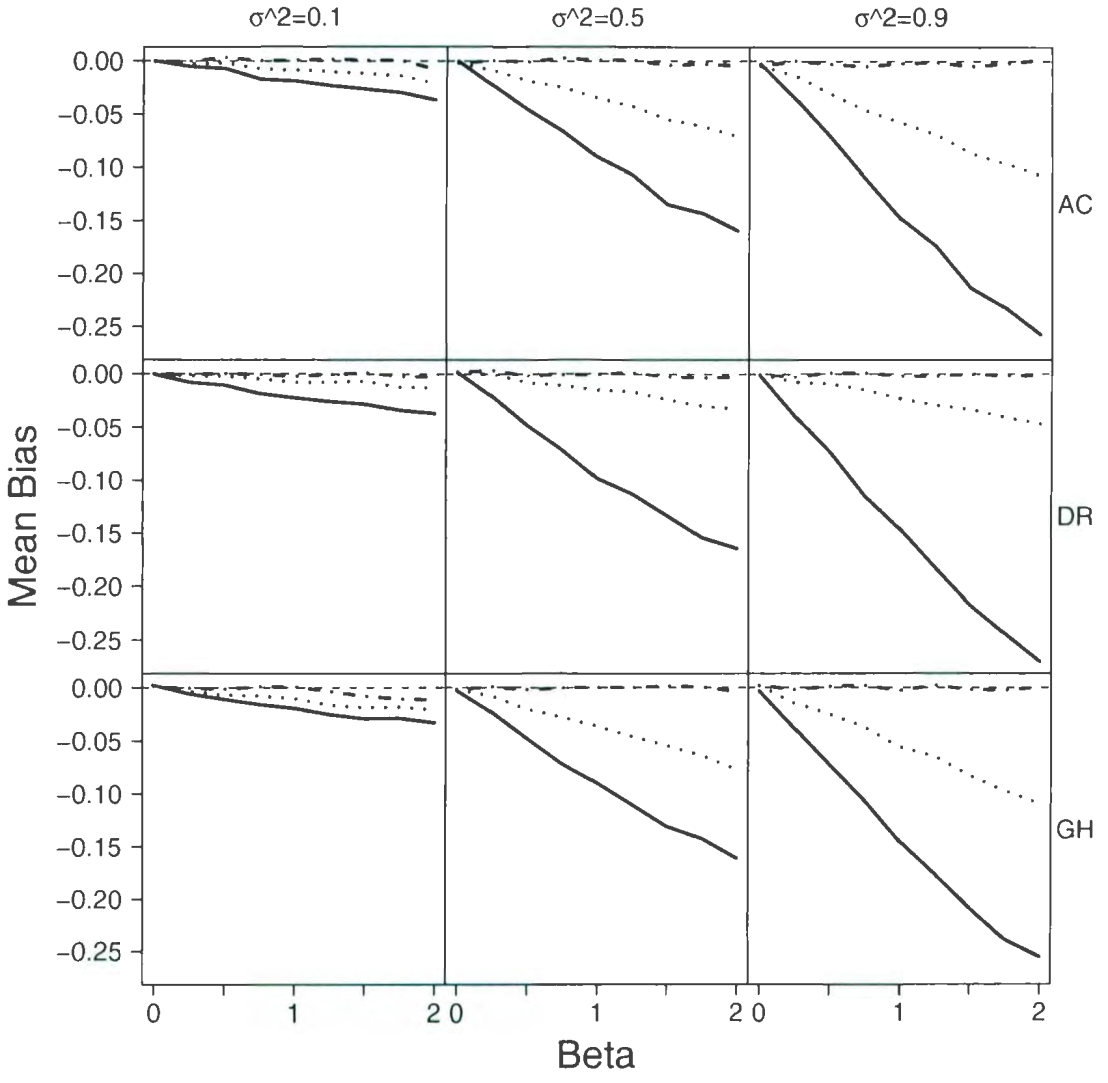


Figure 4.9: Bias of $\hat{\beta}_0$ for FEP1 (solid line), MEP1c (dotted line), and MEP1m (dash-dotted line) models. Random effects are a difference of two log-gamma distributed random variables with 0 mean and variances $\sigma^2 = 0.1, 0.5, 0.9$, respectively. The dashed line represents the horizontal line at 0. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

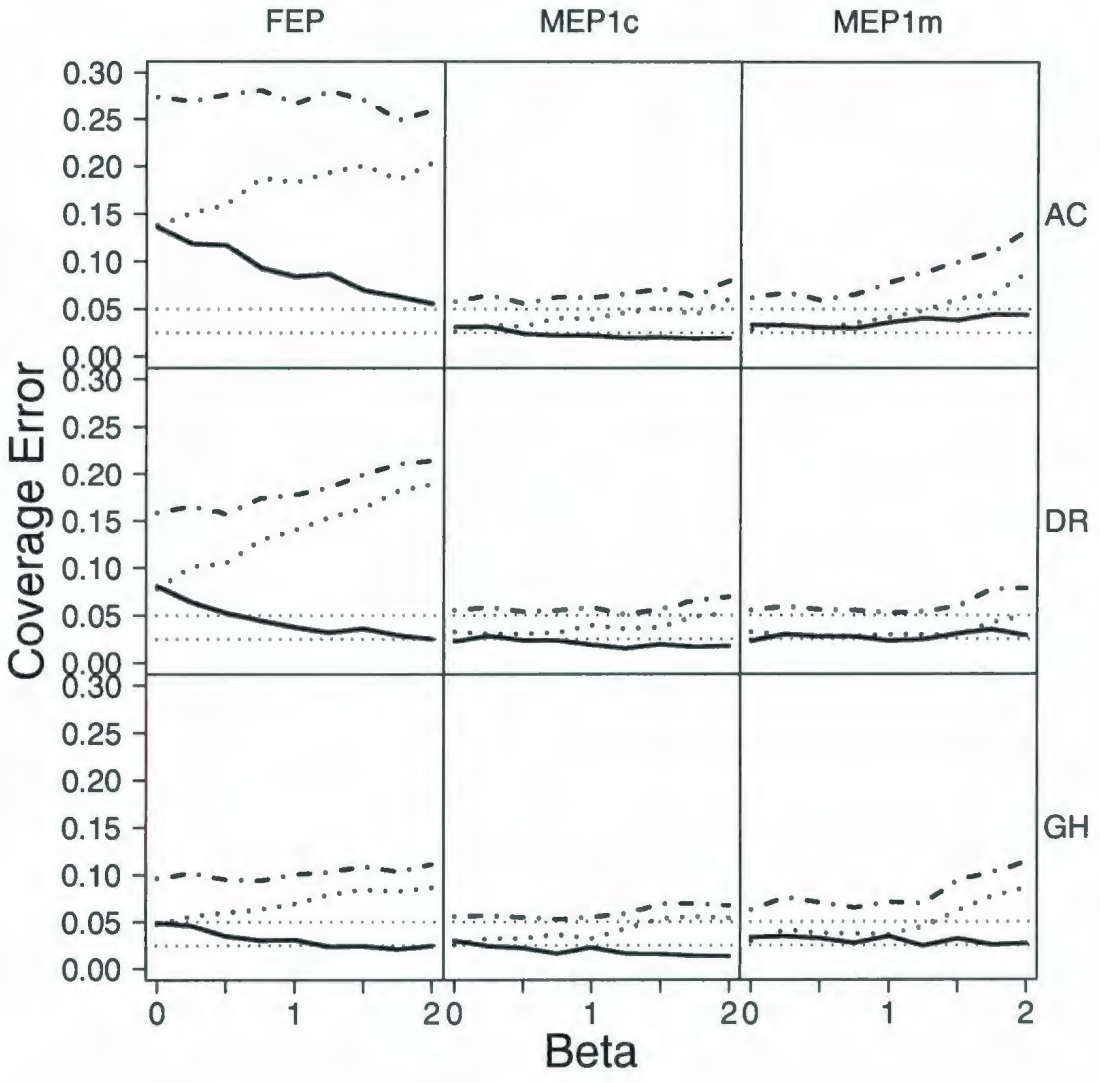


Figure 4.10: 95% coverage errors of the confidence intervals from the parameter estimates for FEP1, MEP1c, and MEP1m models. Random effects are a difference of two log-gamma distributed random variables with mean $E(\delta) = 0$ and variance $Var(\delta) = 0.1$. The solid line represents lower coverage errors, the dotted line represents upper coverage errors, the dash-dotted line represents total coverage errors (lower + upper), and the horizontal dotted lines represent critical values $\alpha = 0.05$ and $\frac{\alpha}{2} = 0.025$. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

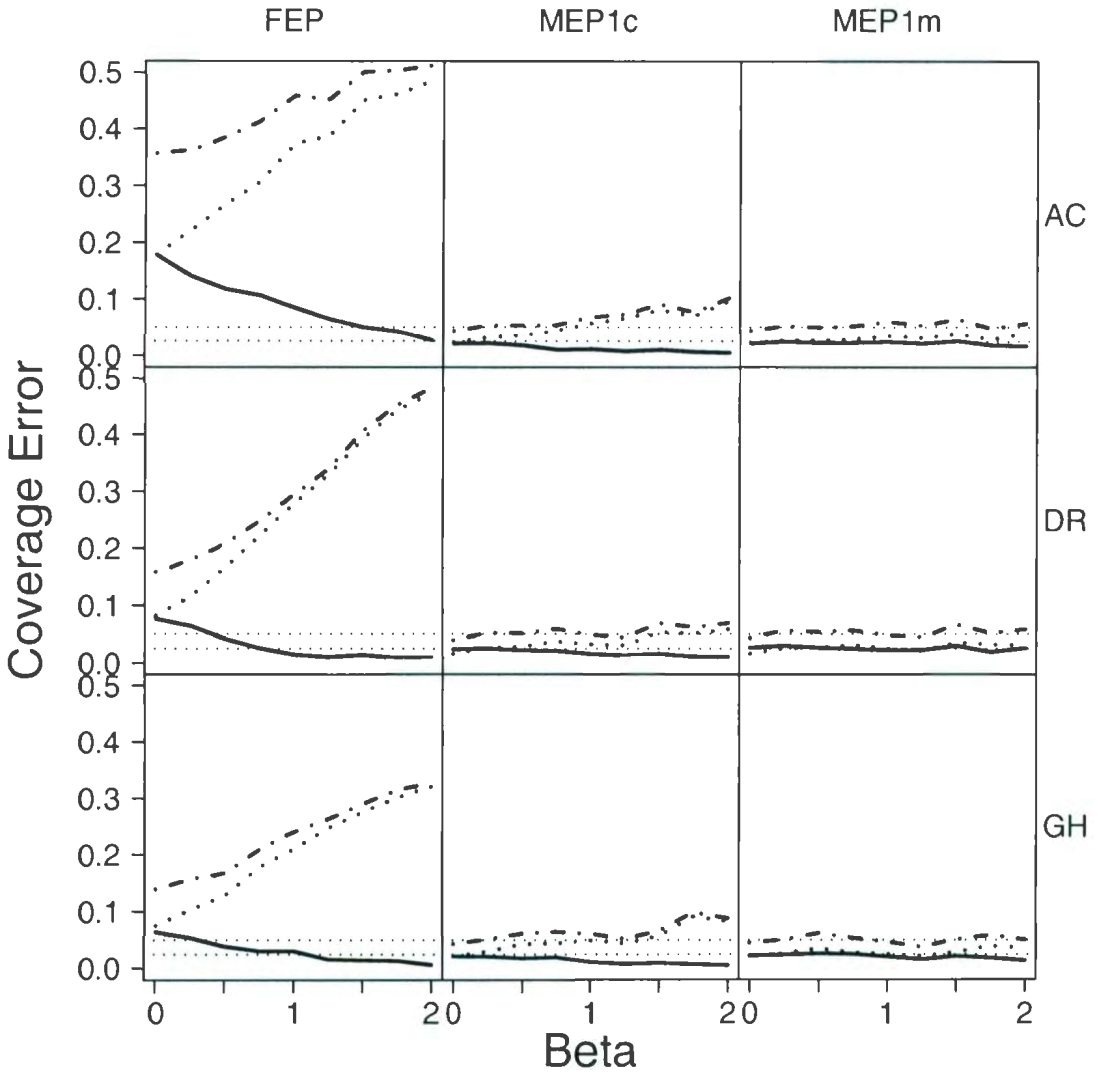


Figure 4.11: 95% coverage errors of the confidence intervals from the parameter estimates for FEP1, MEP1c, and MEP1m models. Random effects are a difference of two log-gamma distributed random variables with mean $E(\delta) = 0$ and variance $Var(\delta) = 0.5$. The solid line represents lower coverage errors, the dotted line represents upper coverage errors, the dash-dotted line represents total coverage errors (lower + upper), and the horizontal dotted lines represent critical values $\alpha = 0.05$ and $\frac{\alpha}{2} = 0.025$. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

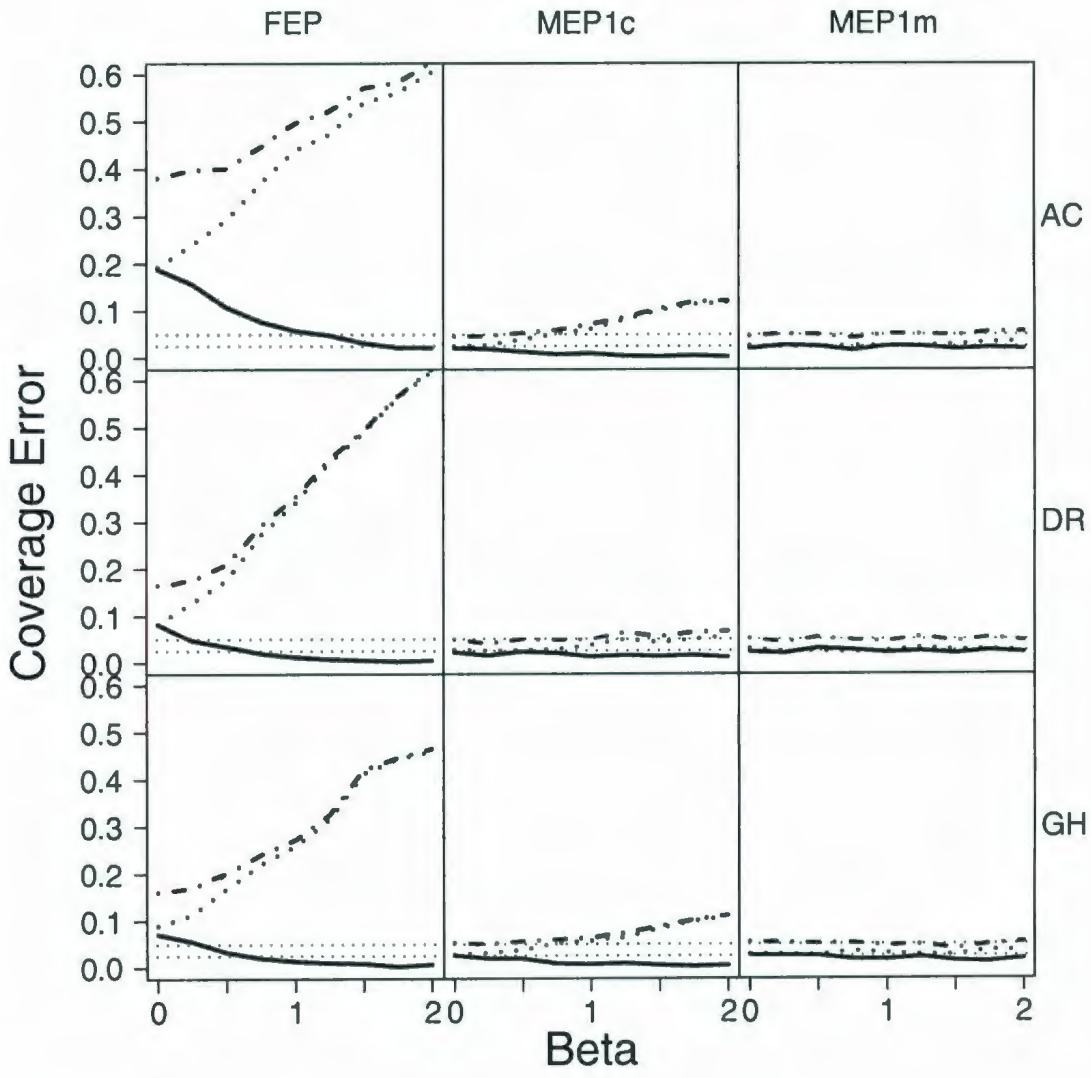


Figure 4.12: 95% coverage errors of the confidence intervals from the parameter estimates for FEP1, MEP1c, and MEP1m models. Random effects are a difference of two log-gamma distributed random variables with mean $E(\delta) = 0$ and variance $Var(\delta) = 0.9$. The solid line represents lower coverage errors, the dotted line represents upper coverage errors, the dash-dotted line represents total coverage errors (lower + upper), and the horizontal dotted lines represent critical values $\alpha = 0.05$ and $\frac{\alpha}{2} = 0.025$. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

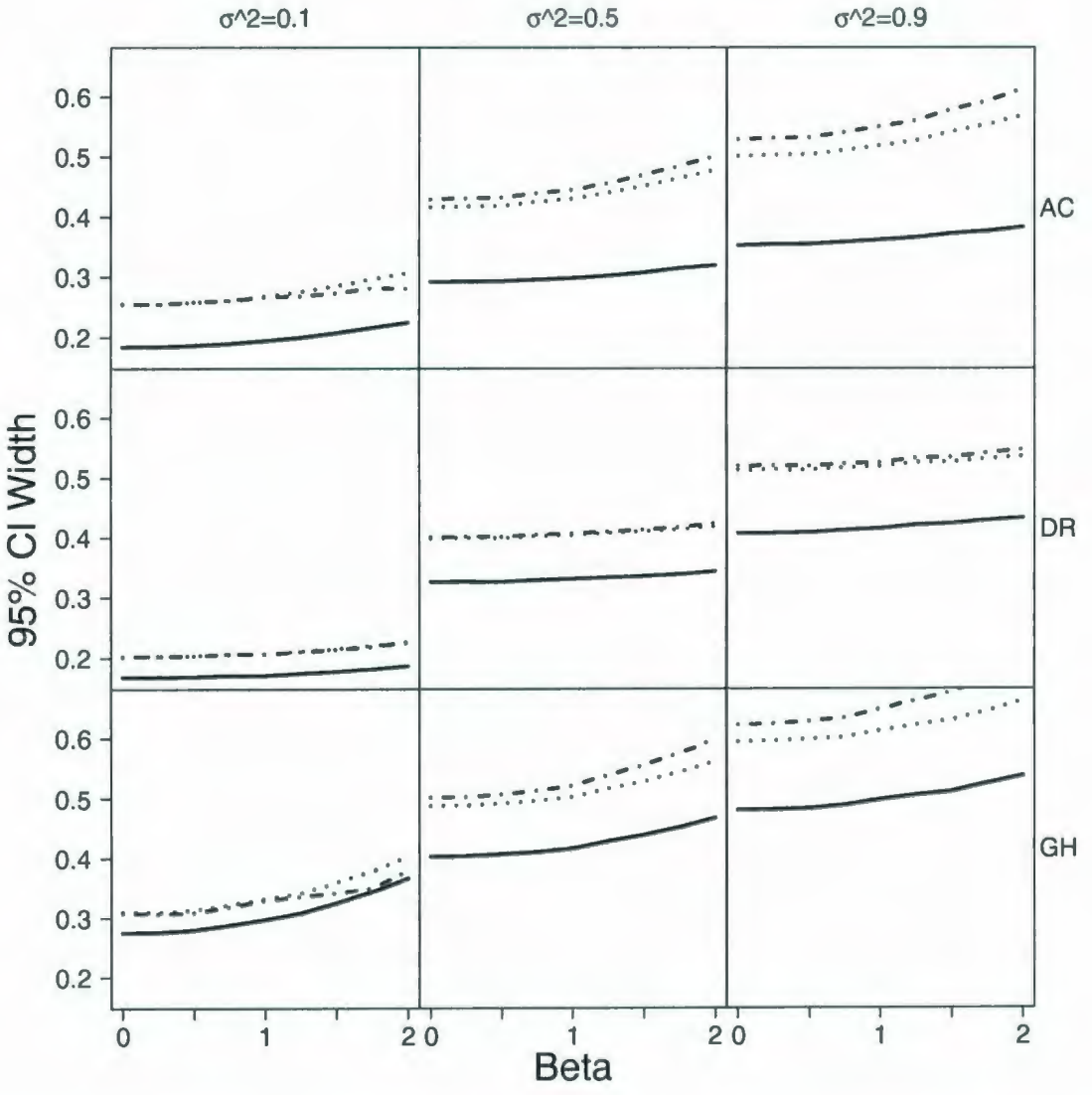


Figure 4.13: 95% confidence width of the parameter estimates for FEP1 (solid line), MEP1c (dotted line), and MEP1m models (dash-dotted line). Random effects are a difference of two log-gamma distributed random variables with 0 mean and variances $\sigma^2 = 0.1, 0.5, 0.9$, respectively. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

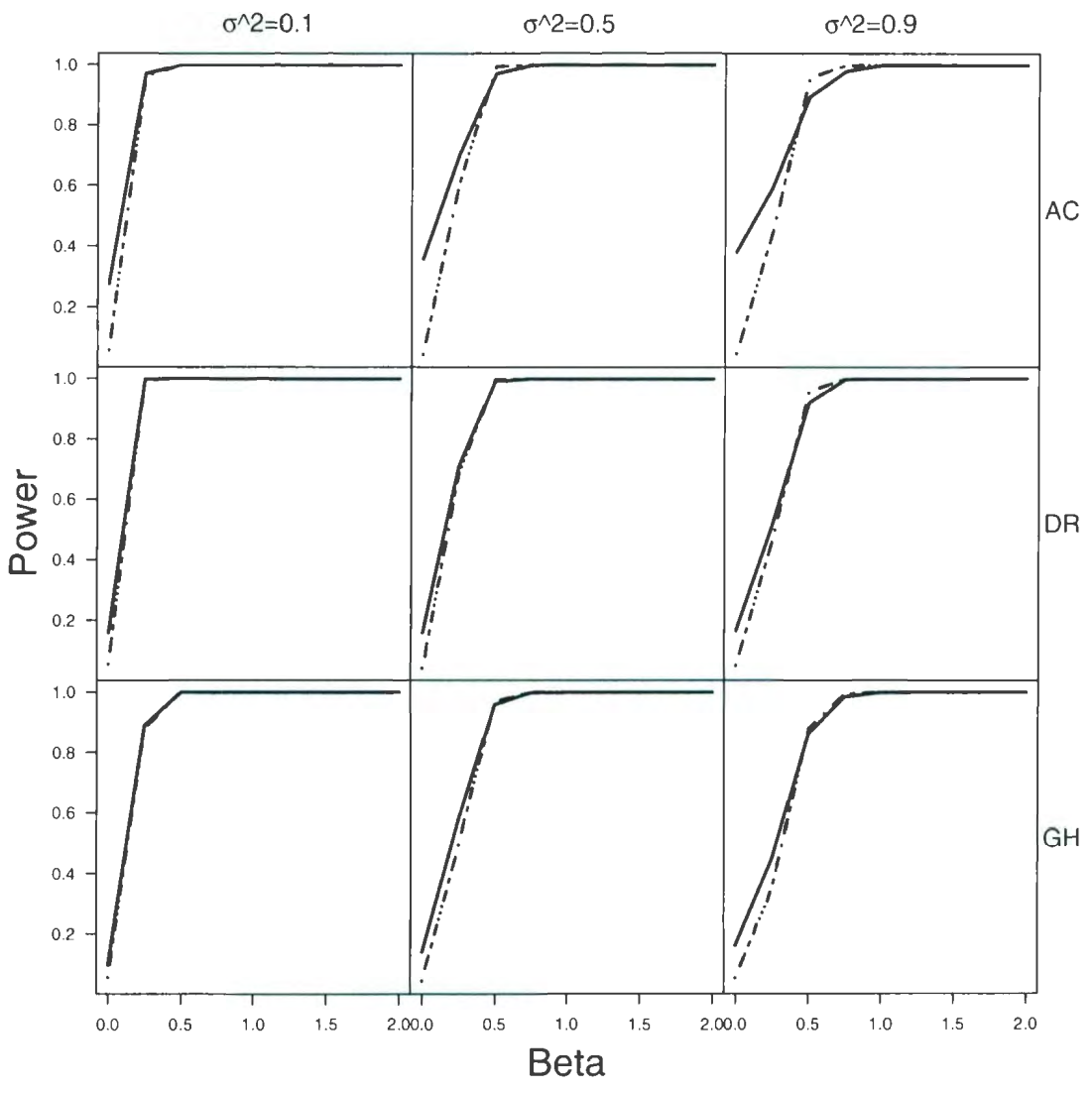


Figure 4.14: Power curves for FEP1 (solid line), MEP1c (dashed line), and MEP1m (dash-dotted). Random effects are a difference of two log-gamma distributed random variables. Rows are for species, with codes indicated at the right hand-side: AC - Atlantic cod; DR - deepwater redfish; GH - Greenland halibut.

# Bibliography


[1] Agresti, A., Booth, J.C., Hobart, J.P., and Caffo, B. 2000. Random-effects modeling of categorical response data. *Socio. Meth.*, **30**: 27–80.

[2] Anderson, D.A., and Aitkin, M. 1985. Variance component models with binary response. *JRSS B*, **47**: 203–210.

[3] Armstrong, M.J., Gerritsen, H.D., Allen, M., McCurdy, W.J., and Peel, J.A.D. 2004. Variability in maturity and growth in a heavily exploited stock: cod (Gadus morhua L.) in the Irish Sea. *ICES J. Mar. Sci.*, **61**: 98–112.

[4] Barot, S., Heino, M., Morgan, M.J., and Dieckmann, U. 2005. Maturation of Newfoundland American plaice (Hippoglossoides platessoides): long term trends in maturation reaction norms despite low fishing mortality. *ICES J. Mar. Sci.*, **62**: 56–64.

[5] Beacham, T.D. 1983. Variability in median size and age at sexual maturity of Atlantic cod, Gadus morhua, on the Scotian Shelf in the Northwest Atlantic Ocean. *Fish Bull. U.S.*, **32**: 303–321.

[6] Beal, S.L., and Sheiner, L.B. 1988. Heteroskedastic nonlinear regression. *Technometrics*, **30**: 327–338.

[7] Benoît, H.G., and Swain, D.P. 2003. Accounting for length and depth-dependent diel variation in catchability of fish and invertebrates in an annual bottom-trawl survey. *ICES J. Mar. Sci.*, **60**: 1298–1317.

[8] Bourdages, H., Savard, L., Archambault, D., and Valois, S. 2007. Results from the August 2004 and 2005 comparative fishing experiments in the northern Gulf of St. Lawerence between CCGS Alfred Needler and CCGS Teleost. *Can. Tech. Rep. Fish. Aqua. Sci.*, 26xx: ix+ 57pp.

[9] Box, M.J. 1971. Bias in non-linear regression. *JRSS B.*, **32**: 171–201.

[10] Box, G.E.P., and Jenkins, G.M. 1976. *Time Series Analysis: Forecasting and Control.* 2nd ed. Holden-Day, San Francisco.

[11] Brattey, J., Cadigan, N.G., Healy, B.P., Lilly, G.R., Murphy, E.F., Stansbury, D.E., and Mahe, J.-C. 2004. An assesment of the cod (Gadus Morhua) stock in NAFO Subdiv. 3Ps in October 2004. *Can. Stock Asses. Sec. Res. Doc.*, 2004/083.

[12] Breslow, N.E., and Clayton, D.G. 1993. Approximate inference in generalized linear mixed models. *JASA*, **88**: 9 25.

[13] Bromley, P.J. 2000. Growth, sexual maturation and spawning in central North Sea plaice (Pleuronectes platessa L.), and the generation of maturity ogives from commercial catch data. *J. Sea Res.*, **44**: 27 43.

[14] Buse, A. 1982. The likelihood ratio, Wald, and Lagrange multiple tests: An expository note. *Am. Stat.*, **36**: 153 157.

[15] Cadigan, N.G., and Farrell, P.J. 2004. Local influence diagnostics for the retrospective problem in sequential population analysis. *ICES J. Mar. Sci*, **62**: 256 265.

[16] Cadigan, N.G., Walsh, S.J., and Brodie, W. 2006. Relative efficiency of the Wilfred Templemen and Alfred Needler research vessels using a Campelen 1800 shrimp trawl in NAFO Subdivision 3Ps and Divisions 3LN. *DFO Can. Sci. Advis. Sec. Res. Doc.* 2006/085.

[17] Cook, R.D., Tsai, C.-L., and Wei, B.C. 1986. Bias in nonlinear regression. *Biometrika*, **73**: 615 623.

[18] Cordeiro, G.M., and McCullagh, P. 1991. Bias correction in generalized linear models. *JRSS B.*, **53**: 629 643.

[19] Cotter, A.J.R. 2001. Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves. *ICES J. Mar. Sci.*, **19**: 41 49.

[20] Cox, D.R., and Snell, E.J. 1989. *Analysis of Binary Data.* 2nd ed. Chapman and Hall, London.

[21] Davidian, M., and Gallant, R.A. 1993. The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, **80**: 475 488.

[22] Davidson, A.C., and Hinkley, D.V. 1997. *Bootstrap Methods and their Applications.* Cambridge University Press, New York.

[23] Demidenko, E. 2004. *Mixed Models Theory and Applications.* John Wiley & Sons, Hoboken, New Jersey.

[24] Dobson, A.J. 2002. *An Introduction to Generalized Linear Models.* 2nd Ed. Chapman & Hall/CRC Texts in Statistical Science, Boca Raton, Florida.

[25] Doubleday, W.G. 1981. Manual on groundfish surveys in the Northwest Atlantic. *NAFO. Sci. Coun. Studies*, **2**: 7 55.

[26] Efron, B. 1982. The jackknife, the bootstrap and other re-sampling plans. *SIAM.* Philadelphia, pp. 92.

[27] Efron, B. 1983. Estimating the error rate of prediction rule: improvement on cross-validation. *JASA*, **78**: 316–331.

[28] Efron, B, and Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*, New York: Chapman-Hall.

[29] *Generalized Linear Models.* Retrieved May 16, 2007, from http://www.statsci.org/glm/bibliog.htm.

[30] Gerritsen, H.D., Armstrong, M.J., Allen, M., McCurdy, W.J., and Peel, J.A.D. 2003. Variability in maturity and growth in a heavily exploited stock: whiting (Merlangius merlangus L.) in the Irish Sea. *J. Sea Res.*, **49**: 69 82.

[31] Gilmour, A.R., Anderson, R.D., and Rae, A.L. 1985. The analysis of binomial data by generalized linear mixed models. *Biometrika*, **72**: 593 599.

[32] Goldstein, H. *Multilevel Statistical Models.* 2nd Ed., Halstead Press, New York.

[33] Grimmit, G.R., and Stirzaker, D.R. 1994. *Probability and Random Processes.* 2nd Ed., Oxford University Press Inc., New York. 541 pp.

[34] Gunderson, D.R. 1993. *Surveys of Fisheries Resources.* John Wiley & Sons, New York.

[35] Gunst, R.F., and Mason, R.L. 1977. Biased estimation in regression: An evaluation using mean square error. *JASA*, **72**: 616 628.

[36] Harville, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *JASA*, **72**: 320–328.

[37] Heagerty, P.J., and Kurland, B.F. 2001. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, **88**: 973 985.

[38] Jiao, Y., and Chen, Y. 2004. An application of generalized linear models in production model and sequential population analysis. *Fish. Res.*, **70**: 367 376.

[39] Kish, L. 1995. *Survey Sampling.* John Wiley & Sons, New York.

[40] Korsbrekke, K. 1999. Variations in maturity of haddock in the Barents Sea in relation to year-class strength, age, size, sex and area. *J. Northw. Atl. Fish. Sci.*, **25**: 37 45.

[41] Kuk, A.Y.C. 1995. Asymptotically unbiased estimation in generalized linear models with random effects. *JRSS B*, **57**: 395 407.

[42] Lawless, J.F. 1980. Inference in the generalized gamma and log gamma distribution. *Technometrics*, **22**: 409 419.

[43] Lawless, J.F. 1982. *Statistical Models for Lifetime Data.* 2nd Ed., Wiley-Interscience, Hoboken, N.J.

[44] Lee, Y., and Nelder, J.A. 1996. Hierarchical generalized linear models. *JRSS B*, **58**: 619 678.

[45] Lee, Y., and Nelder, J.A. 1998. Generalized linear models for the analysis of quality-improved experiments. *Can. J. Stat.*, **26(1)**: 95 105.

[46] Lewy, P., Nielson, J.R., and Hovgard, H. 2004. Survey gear calibration independent of spatial distribution. *Can. J. Fish. Aquat. Sci.* **61**: 636 647.

[47] Liang, K-.Y., and Zeger, S.L. 1986. Longitudinal data anlysis using generalized linear models. *Biometrika*, **73**: 13 22.

[48] Lin, X., and Breslow, N.E. 1996. Bias correction in generalized linear mixed models with multiple components of dispersion. *JASA*, **91**: 1007 1018.

[49] Lindstrom, M.J., and Bates, D.M. 1990. Nonlinear mixed effects models for repeated measures data. *Biometrika*, **46**: 673 687.

[50] McCallum, B., and Walsh, S.J. 2002. An update on the performance of the Campelen 1800 during the bottom trawl survey in NAFO subareas 2 and 3 in 2001. *NAFO SCR Doc.*, 02/32.

[51] McCullagh, P., and Nelder, J.A. 1989. *Generalized Linear Models.* 2nd ed. Chapman and Hall, London.

[52] McGilichrist, C.A. 1994. Estimation in generalized linear models. *JRSS B*, **56**: 61 69.

[53] Morgan, M.J. 2000. Estimating spawning stock biomass in 2J3KL cod using a cohort maturation model and variable sex ratio. *Can. Stock Assess. Sec. Res. Doc.* 2000/110.

[54] Morgan, M.J., and Colbourne, E.B. 1999. Variation in maturity at age and size in three populations of American plaice. *ICES J. Mar. Sci.*, **56**: 673 688.

[55] Morgan, M.J., and Hoenig, J.M. 1997. Estimating age at maturity from length stratified sampling. *J. Northw. Fish. Sci.*, **21**: 51 63.

[56] Millar, R.B. 1992. Estimating the size-selectivity of fishing gear by conditioning on the total catch. *JASA*, **420**: 962 968.

[57] Needle, C.L., O'Brien, C.M., Darby, C.D., and Smith, M.T. 2003. Incorporating time series structure and environmental information in medium-term stock projections. *Sci. Mar.*, **67** (Suppl. 1): 201 209.

[58] Nelder, J.A., and Wedderburn, R.W.M. 1972. Generalized linear models. *JRSS A*, **135**: 370 384.

[59] Olsen, E.M., Lilly, G.R., Heino, M., Morgan, M.J., Brattey, J., and Dieckmann, U. 2005. Assessing changes in age and size at maturity in collapsing populations of Atlantic cod (Gadus Morhua). *Can. J. Fish. Aqua. Sci.*, **62**: 811 823(13).

[60] Parsons, D.G, Kulka, D.W., and Veitch, P.J. MS 1997. Distribution, biomass, abundance and demography of shrimp (Pandalus borealis) in Flemish Cap (NAFO Div. 3M) based on data obtained during a Canadian research trawl survey, SeptemberOctober, 1996. *NAFO SCR Doc.*, No. 81, Serial No. N2927, 14 p.

[61] Pelletier, D. 1998. Intercalibration of research survey vessels in fisheries: a review and application. *Can. J. Fish. Aquat. Sci.* **55**: 2672 2690.

[62] Pierce, D.A., and Schaffer, D.W. 1986. Residuals in generalized linear models. *JASA*, **88**: 973-985.

[63] Pinheiro, J.C., and Bates, D.M. 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J. Comp. Graph. Stat.*, **4**: 12 35.

[64] Pitt, T.K. 1975. Changes in abundance and certain biological characteristics of Grand Bank American plaice, Hippoglossoides platessoides. *J. Fish. Res. Board Canada*, **32**: 1383 1398.

[65] Reid, N. 1995. The rdes of conditioning in inference. *Stat. Sci.* **10**: 138 199.

[66] Robinson, G.K. 1991. That BLUP is a good thing: the estimation of random effects (with discussion). *Stat. Sci.*, **6**: 15 51.

[67] SAS Institute Inc. 2005. *SAS Technical Report, SAS/STAT Software, the Genmod Procedure, Release 9.1.* SAS Institute Inc., Cary, NC.

[68] SAS Insitute Inc. 2005. *SAS Technical Report, SAS/STAT Software, the Glimmix Procedure, Release 9.1.* SAS Institute Inc., Cary, NC.

[69] SAS Institute Inc. 2005. *SAS Technical Report, SAS/STAT Software, the NLMixed Procedure, Release 9.1.* SAS Institute Inc., Cary, NC.

[70] Schall, R. 1991. Estimation in generalized linear models with random effects. *Biometrika*, **78**: 719 727.

[71] Shelton, P.A., and Armstrong, M.J. 1983. Variations in parent stock and recruitment of pilchard and anchovy populations in the southern Benguela system. In:, Proceedings of the Expert Consultation to Examine Change in Abundance and Species Composition of Neritic Fish Resources, San Jose, Costa Rica, 18-29 April 1983. G.D. Sharp and J. Csirke (eds.) *FAO Fish. Rep.*, **291(3)**: 1113 1132.

[72] Shumway, R.H., and Stoffer, D.S. 2000. *Time Series Analysis and Its Applications.* Springer texts in statistics, New York.

[73] Smedbol, R.K., Shelton, P.A., Swain, D.P., Frechet, A., and Chouinard, G.A. 2002. Review of population structure, distribution and abundance of cod (*Gadus Morhua*) in Atlantic Canada in a species-at-risk context. *DFO Can. Sci. Advis. Sec. Res. Doc.* 2002/082.

[74] Stiratelli, R., Laird, N., and Ware, J. 1984. Random effects models for serial observations with binary responses. *Biometrics*, **40**: 961 971.

[75] Sutradhar, B.C. 2003. An overview on regression models for discrete longitudinal responses. *Stat. Sci.*, **3**: 377 393.

[76] Sutradhar, B.C. 2004. On exact quasilikelihood inference in generalized linear mixed models. *Sankhya: Ind. J. Stat.*, **66**: 263 291.

[77] Sutradhar, B.C., and Rao, R.P. 2001. On marginal quasi-likelihood inference in generalized linear mixed models. *J. Multi. Anal.*, **76**: 1 34.

[78] Swain, D.P., and Poirer, G.A. 1997. Distributions of Atlantic Cod and American Plaice during the September 1996 survey of the southern Gulf of St. Lawrence and their relation to historical data. *DFO Can. Sci. Advis. Sec. Res. Doc.*, 1997/066.

[79] Templeman, W., Hodder, V.M., and Wells, R. 1978. Sexual maturity and spawning in haddock, Melanogrammus aeglefinus, of the southern Grand Bank. *ICNAF Res. Bull.*, **13**: 53 65.

[80] Thall, P.F., and Vail, S.C. 1990. Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**: 657 671.

[81] Venzon, D.J., and Moolgavkar, S.H. 1988. A method for computing profile-likelihood based confidence intervals. *App. Stat.*, **37**: 87 94.

[82] Welch, D.W., and Foucher, R.P. 1988. A maximum likelihood methodology for estimating length-at-maturity with application to Pacific cod (Gadus macrocephalus) population dynnamics. *Can. J. Fish. Aqua. Sci.*, **45(2)**: 333 343.

[83] Wolfinger, R., and O.Connell, M. 1993. Generalized linear mixed models: A psuedo-likelihood approach. *J. Stat. Comp. Sim.*, **6**: 233 243.

[84] Wolfinger, R., Tobias, R., and Sall, J. 1994. Computing guassian likelihoods and their derivatives for general linear mixed models. *SIAM J. Sci. Comp.*, **15(6)**, 1294 1310.

[85] Wu, C.F.J. 1986. Jackknife, bootstrap, and other resampling methods in regression analysis. *Ann. Stat.*, **14**: 1261 1295.

[86] Xiao, Y., Punt, A.E., Millar, R.B., and Quinn II, T.J. 2004. Models in fisheries research: GLMs, GAMS and GLMMs. *Fish. Res.* **70**: 137-139.

[87] Ye, Y., Al-Husani, M, and Al-Baz, A. 2001. Use of generalized linear models to analyze catch rates having zero values: the Kuwait driftnet fishery. *Fish. Res.*, **53**: 151 168.

[88] Zeger, S.L., and Karim, M.R. 1991 Generalized linear models with random effects; a Gibbs sampling approach. *JASA*, **86**: 171 178.

[89] Zeger, S.L., and Liang, K.-Y. 1986. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* **42**: 121 130.

[90] Zeger, S.L., Liang, K-.Y., and Albert, P.A. 1988. Models for longitudinal data: A generalized estimating equations approach. *Biometrika*, **58**: 1049 1060.

# Appendix A

# Derivation of a Log-Gamma Distribution

It is sufficient to assume that fish densities fished by two separate vessels follow a gamma distribution such that

$$\lambda_1, \lambda_2 \sim Gamma\left(\frac{1}{\phi}, \phi\mu\right) \tag{A.1}$$

where $\lambda_1, \lambda_2$ are iid random variables, $\phi$ is an overdispersion parameter, and $\mu$ and $\phi\mu^2$ are the respective means and variances of both populations. The probability density functions of $\lambda_1$ and $\lambda_2$ are

$$f(\lambda_i) = \frac{1}{\Gamma(\frac{1}{\phi})(\phi\mu)^{\frac{1}{\phi}}}\lambda_i^{(\frac{1}{\phi}-1)}e^{-\lambda_i/\phi\mu}, \quad 0 \leq \lambda_i < \infty, \ i = 1, 2. \tag{A.2}$$

Now we examine the difference of the log of the two densities, denoted as

$$\begin{aligned} \delta &= \log\left(\frac{(\phi\mu)^{-1}\lambda_1}{(\phi\mu)^{-1}\lambda_2}\right) \\ &= \log[(\phi\mu)^{-1}\lambda_1] - \log[(\phi\mu)^{-1}\lambda_2] \\ &= w_1 - w_2 \end{aligned} \tag{A.3}$$

where $w_1$ and $w_2$ are iid random variables. Examining $w_i$ $(i = 1, 2)$, we can derive their corresponding distribution functions. The Jacobians of the transformed variables are

$$J_i = \frac{\partial \lambda_i}{\partial w_i} = (\phi\mu)e^{w_i}. \tag{A.4}$$

Using a transformation of variables, the corresponding distributions for $w_1$ and $w_2$ are computed as

$$\begin{aligned} g(w_i) &= f\left(\phi\mu e^{w_i}\right) \times |J_i| \\ &= \frac{1}{\Gamma(\frac{1}{\phi})(\phi\mu)^{\frac{1}{\phi}}}(\phi\mu e^{w_i})^{\frac{1}{\phi}-1} exp\left(-\phi\mu\frac{1}{\phi\mu}e^{w_i}\right) \times (\phi\mu)e^{w_i} \\ &= \frac{1}{\Gamma(\frac{1}{\phi})} exp\left(\frac{w_i}{\phi} - e^{w_i}\right), \quad -\infty < w_i < \infty, \ (i = 1, 2) \end{aligned} \tag{A.5}$$

which are log-gamma density functions. The mean and variance of $w_i$ are defined as

$$E(w_i) = \psi\left(\frac{1}{\phi}\right) \tag{A.6}$$

$$Var(w_i) = \psi'\left(\frac{1}{\phi}\right) \tag{A.7}$$

where $\psi(\frac{1}{\phi})$ and $\psi'(\frac{1}{\phi})$ are the digamma and trigamma functions, respectively (see Lawless 1982; Appendix B).

It then follows that the expected value of $\delta = w_1 - w_2$ (the difference of two log-gamma random variables) is

$$E(\delta) = E(w_1 - w_2)$$

$$\begin{aligned} &= E(w_1) - E(w_2) \\ &= \psi\left(\frac{1}{\phi}\right) - \psi\left(\frac{1}{\phi}\right) \\ &= 0 \end{aligned} \tag{A.8}$$

and the variance is

$$\begin{aligned} Var(\delta) &= Var(w_1 - w_2) \\ &= Var(w_1) + Var(w_2) - 2Cov(w_1, w_2) \\ &= \psi'\left(\frac{1}{\phi}\right) + \psi'\left(\frac{1}{\phi}\right) \\ &= 2\psi'\left(\frac{1}{\phi}\right) \end{aligned} \tag{A.9}$$

where $Cov(w_1, w_2) = 0$ since $w_1$ and $w_2$ are iid random variables.

# Appendix B

# Table of Acronyms

| Acronym | Description |
|---|---|
| $A_{50}$ | Age at 50% maturity |
| AC | Atlantic cod |
| ACF | Autocorrelation function |
| AM | American Plaice |
| AN | Canadian Coast Guard research vessel Alfred Needler |
| AR NOD | Autoregressive mixed effects model without over dispersion parameter |
| AR OD | Autoregressive mixed effects model with over disperison parameter |
| AR YE | Autoregressive mixed effects model with year effects |
| AR YE- | Autoregressive mixed effects model with year effects as nuisance parameters |
| AR YE+ | Autoregressive mixed effects model with year effects as predictive parameters |
| AR(p) | Autoregressive process with order p |
| BLUP | Best linear unbiased predictor |
| CI | Confidence interval |
| DFO | Department of Fisheries and Oceans |
| DR | Deepwater redfish |
| FE | Fixed effects model |
| FE1 | Fixed effects model with vessel effect parameter |
| FE2 | Fixed effects model with vessel and length effect parameters |
| FEP1 | Fixed effects model with vessel effect parameter (pooled over lengths) |
| GEE | Generalized estimating equation |
| GH | Greenland halibut |
| GLIM | Generalized linear model |
| GLMM | Generalized linear mixed effects model |
| HL | Hierarchical likelihood |
| MA(q) | Moving average process with order q |
| ME1 | Mixed effects model with vessel effect parameter |
| ME2 | Mixed effects model with vessel and length effect parameters |
| ME2ARc | Conditional mixed effects model with autocorrelated random lengths within sets |

| Acronym | Description |
|---|---|
| ME2RIc | Conditional mixed effects model with random intercepts |
| ME2RIm | Conditional mixed effects model with random intercepts |
| MEP1 | Mixed effects model with vessel effect parameter (pooled over lengths) |
| MEP1c | Conditional mixed effect model pooled over lengths |
| MEP1m | Marginal mixed effects model pooled over lengths |
| ML | Maximum likelihood |
| MQL | Marginal quasi-likelihood |
| MR | Maturity range |
| NAFO | Northwest Atlantic Fisheries Organization |
| PA | Population-averaged |
| PACF | Partial autocorrelation function |
| PL | Pseudo-likelihood |
| PQL | Penalized quasi-likelihood |
| REML | Restricted maximum likelihood |
| REPL | Restricted pseudo-likelihood |
| SE | Standard error |
| SS | Subject-specific |
| SSB | Spawning stock biomass |
| TEL | Canadian Coast Guard research vessel Teleost |
| VC | Variance components |
| WF | Witch flounder |
| WT | Canadian Coast Guard research vessel Wilfred Templeman |