Theses and Dissertations

Fall 2010

Partly parametric generalized additive model

Tianyang Zhang
University of Iowa

Recommended Citation

Zhang, Tianyang. "Partly parametric generalized additive model." PhD (Doctor of Philosophy) thesis, University of Iowa, 2010.
http://ir.uiowa.edu/etd/913.

PARTLY PARAMETRIC GENERALIZED ADDITIVE MODEL

by

Tianyang Zhang

<u>An Abstract</u>

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Statistics in the
Graduate College of The
University of Iowa

December 2010

Thesis Supervisor: Professor Kung-Sik Chan

ABSTRACT

In many scientific studies, the response variable bears a generalized nonlinear regression relationship with a certain covariate of interest, which may, however, be confounded by other covariates with unknown functional form. We propose a new class of models, the partly parametric generalized additive model (PPGAM) for doing generalized nonlinear regression with the confounding covariate effects adjusted nonparametrically. To avoid the curse of dimensionality, the PPGAM specifies that, conditional on the covariates, the response distribution belongs to the exponential family with the mean linked to an additive predictor comprising a nonlinear parametric function that is of main interest, plus additive, smooth functions of other covariates. The PPGAM extends both the generalized additive model (GAM) and the generalized nonlinear regression model. We propose to estimate a PPGAM by the method of penalized likelihood. We derive some asymptotic properties of the penalized likelihood estimator, including consistency and asymptotic normality of the parametric estimator of the nonlinear regression component. We propose a model selection criterion for the PPGAM, which resembles the BIC. We illustrate the new methodologies by simulations and real applications. We have developed an R package **PPGAM** that implements the methodologies expounded herein.

Abstract Approved: \underline{\hspace{5cm}}
\
Thesis Supervisor

\underline{\hspace{5cm}}
Title and Department

\underline{\hspace{5cm}}
Date

PARTLY PARAMETRIC GENERALIZED ADDITIVE MODEL

by

Tianyang Zhang

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Statistics in the
Graduate College of The
University of Iowa

December 2010

Thesis Supervisor: Professor Kung-Sik Chan

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Tianyang Zhang

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Statistics at the December 2010 graduation.

Thesis committee: _____
                  Kung-Sik Chan, Thesis Supervisor

                  _____
                  Richard L. Dykstra

                  _____
                  Joseph B. Lang

                  _____
                  Ying Zhang

                  _____
                  Dale L. Zimmerman

ACKNOWLEDGMENTS

Foremost, I would like to express my deepest gratitude to my advisor Professor Kung-Sik Chan for his guidance, his encouragement and his patience. His rigorous critiques, plenteous knowledge and excellent experience were pivotal for me to complete this research work. It is my honor to be his student. I believe that he is the best advisor I could have during my Ph.D. study.

I would like to thank the other four members in the advisory committee: Professor Richard L. Dykstra, Professor Dale L. Zimmerman, Professor Joseph B. Lang and Professor Ying Zhang for their insightful comments and questions. Their statistical expertise was great resource for me to accomplish my thesis.

Special thanks to Dr. Kevin Bailey of NOAA, Department of Commerce. His talents and knowledge in marine science contributed substantially to the applications of the new statistical methodologies proposed in my thesis.

Last but not the least, I would like to thank my parents, Yingzhen Ma and Ming Zhang, who gave birth to me, raised me with love and supported me spiritually. I would like to thank my wife Yuning Song for her encouragement and love, which are the most valuable wealth in my life.

ABSTRACT

In many scientific studies, the response variable bears a generalized nonlinear regression relationship with a certain covariate of interest, which may, however, be confounded by other covariates with unknown functional form. We propose a new class of models, the partly parametric generalized additive model (PPGAM) for doing generalized nonlinear regression with the confounding covariate effects adjusted nonparametrically. To avoid the curse of dimensionality, the PPGAM specifies that, conditional on the covariates, the response distribution belongs to the exponential family with the mean linked to an additive predictor comprising a nonlinear parametric function that is of main interest, plus additive, smooth functions of other covariates. The PPGAM extends both the generalized additive model (GAM) and the generalized nonlinear regression model. We propose to estimate a PPGAM by the method of penalized likelihood. We derive some asymptotic properties of the penalized likelihood estimator, including consistency and asymptotic normality of the parametric estimator of the nonlinear regression component. We propose a model selection criterion for the PPGAM, which resembles the BIC. We illustrate the new methodologies by simulations and real applications. We have developed an R package **PPGAM** that implements the methodologies expounded herein.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Parametric nonlinear regression (Bates and Watts, 1988) is widely useful for studying a complex relationship that can be described by a nonlinear regression function of known functional form up to some unknown parameters. However, in many applications, the functional form of the nonlinear relationship may only be partly specified, because the relationships between the response and some confounding covariates may have unknown functional form. For example, in a study of the dynamics of pollock larvae in the Gulf of Alaska (GOA) reported in Chapter 3, while the baseline survival distribution may be assumed to belong to some parametric family, e.g. Weibull or log-logistic, the baseline hatchdate distribution and the environmental effects on hatching and/or survival may be best modeled nonparametrically as there is little substantive knowledge about their functional forms. This motivates us to study the partly parametric generalized additive model (PPGAM).

The proposed PPGAM generalizes the highly popular generalized additive model (GAM) (Hastie and Tibshirani, 1986, 1990; Wood, 2006) by adding a parametric nonlinear component to the additive predictor on the link scale. Such a model structure has wide applications in scientific studies where some parametric nonlinear regression relationship is of main interest, but it is confounded by some confounding covariates whose relationship to the response is of unknown functional form and hence best estimated nonparametrically. In many cases, the preceding nonparametric effects may be adequately approximated by an additive model structure, i.e. as a sum of smooth functions of the confounding covariates, an assumption embodied in the PPGAM.

The PPGAM may also be considered as an extension of the generalized nonlinear model (GNLM) (Bates and Watts, 1988), and can accommodate both continuous and discrete responses, as the response distribution is assumed to conditionally belong to the exponential family. By the theory of reproducing kernel Hilbert space, each additive function of the nonparametric part of the PPGAM is spanned by the (thin-plate) splines (Duchon, 1977; Wood, 2006), so we propose the method of penalized (quasi-)likelihood for estimating a PPGAM. Under some regularity conditions and making heavy use of the empirical processes theory, we prove some large-sample properties of the penalized likelihood estimators of the parameters indexing the nonlinear component (to be referred as the nonlinear parametric estimator below), including consistency and asymptotic normality; as well we derive the consistency of the additive function estimators.

The PPGAM can be further generalized by relaxing the additivity assumption, which results in the partly parametric generalized nonlinear regression (PPGNR) model. The PPGNR is a class of semiparametric models. Semiparametric models have been extensively studied in the literature. Partial linear model (Wahba, 1990) is a simple semiparametric model with the regression function being the sum of a linear parametric component and unknown nonparametric functions. Mammen and van de Geer (1997) derived some asymptotic properties of the penalized quasi-likelihood estimator for a general partial linear model with sub-exponential error terms. Wang and Ke (2009) proposed a general semiparametric nonlinear regression model with a similar mean structure as PPGNR but assuming Gaussian errors. Huang and Chen (2008) considered a regression model with two additive components in the regression function, namely parametric nonlinear function and nonparametric smooth function. Hence, their model shares a similar conditional mean structure as the PPGAM. However, the main difference between their model and the PPGAM proposed here

is that their approach makes no distributional assumption for the response and they simply assume independent and identically distributed (iid) errors with zero mean and finite variance, whereas the proposed PPGAM explicitly specifies that the conditional response distribution belongs to the exponential family.

The rest of the thesis is organized as follows. In Chapter 2, we provide an in-depth study of the PPGAM. We give the formulation of the PPGAM in Section 2.1. The penalized likelihood estimation approach and a proposed model selection criterion based on the Laplace approximation of the marginal likelihood are introduced in Section 2.2. We use empirical processes theory to derive the asymptotic properties of the penalized (quasi-)likelihood estimator. Some asymptotic properties of the functional estimator and the parametric estimator are reported in Section 2.3. In Section 2.4, we report several simulation studies on the empirical performance of the method of the penalized likelihood estimation and the empirical effectiveness of the proposed model selection criterion. We illustrate the PPGAM with a real dataset from the atomic bomb study in Section 2.5. The proofs of all theorems are collected in Section 2.6. The approximation of the asymptotic variance and the Laplace approximation of the marginal likelihood are justified in Sections 2.7 and 2.8, respectively. Chapter 3 contains a relatively complete case study on the dynamics of pollock larvae. In the application, the PPGAM successfully models the hatching process of pollock in the Gulf of Alaska (GOA) and the survival process of pollock larvae, and untangles the human and/or natural intervention effects. In Chapter 4, we elaborate two forecasting models for pollock recruitment, based on the threshold generalized additive models. Both Chapters 3 and 4 are based on joint works with Dr. Kevin Bailey of NOAA, Department of Commerce. The methodologies developed in Chapter 2 are implemented in an R package **PPGAM**, which is detailed in Chapter 5. We conclude in Chapter 6.

CHAPTER 2
PARTLY PARAMETRIC GENERALIZED ADDITIVE MODEL

The partly parametric generalized additive model (PPGAM) assumes that the conditional response distribution belongs to exponential families whose mean is linked to a predictor function that is a sum of a parametric nonlinear component and a finite number of smooth functions of the covariates. To avoid overfitting of the nonparametric functions, we propose to estimate a PPGAM by the method of penalized likelihood; the penalized log likelihood equals the log likelihood plus a roughness penalty for each additive function in the model (Hastie and Tibshirani, 1990; Marx and Eilers, 1998). Our theoretic investigation shows that the penalized likelihood estimator of the parameter in the nonlinear component (simply referred to as the parametric estimator below) is consistent and follows an asymptotically normal distribution under some mild regularity conditions. In practice, the covariance matrix of the asymptotic normal distribution for the parametric estimator can be approximated by the inverse of the observed Fisher information, which can be justified in some simple case and the approximation is further studied via simulation study. Moreover, the penalty in the penalized estimation method admits a Bayesian interpretation as some prior belief on the smoothness of the functions. Based on the Bayesian approach, we develop a model selection criterion via Laplace approximation of the marginal likelihood. We illustrate the PPGAM with an analysis of the cancer risk of the survivors from the atomic bomb in Hiroshima and Nagasaki, Japan; this application concerns a non-Gaussian response, and it demonstrates the

versatility of the PPGAM for handling non-normal data.

2.1   Model Structure

The PPGAM is similar to the GAM except that the conditional mean has a parametric nonlinear component. Specifically, consider independent and identical distributed random vectors $(Y_1, T_1), \cdots, (Y_n, T_n)$ which share the same distribution of (Y, T). Given the covariate T, the conditional distribution of the response Y belongs to the exponential family with its probability density (mass) function being given by $f_\vartheta(y) = exp[\{y\vartheta - b(\vartheta)\}/a(\phi) + c(y, \phi)]$ where ϑ is the canonical parameter, $a(\phi) = \phi/\omega$, with ω being a known constant that denotes the weight of the datacase and ϕ being the known dispersion parameter, and $c(y, \phi)$ is the normalizing constant. The canonical parameter ϑ relates to the conditional mean $\mu(t) = E(Y|T = t)$ by the formula that the first derivative $\dot{b}(\vartheta) = \mu(t)$. For the GAM, the conditional mean is linked to some additive function of the covariate. Here, we consider an extension of the GAM with the covariate T comprising of a number of random vectors, namely X and $Z = (Z_v, \nu = 1, \cdots, w)$ and the conditional mean is linked to a sum of additive functions of the Z_v's as well as an additive parametric term involving X alone:

$$\mu(x, z) = F(h_\theta(x) + \sum_{v=1}^{w} s_v(z_v)), \tag{2.1}$$

where F is a known smooth monotonic inverse link function, $h_\theta(x)$ is the parametric (possibly) nonlinear component of known functional form up to some unknown parameter θ, and $s(z_v)$'s are the nonparametric components with each $s(\cdot)$ being a smooth function to be estimated nonparametrically. Let $g_\theta(x, z) = h_\theta(x) + \sum_{v=1}^{w} s_v(z_v)$, which is referred to as the predictor (function) on the link scale. Often, F is taken as the cannonical inverse link function $\dot{b}(\cdot)$, e.g. the exponential function for the Poisson family. In the absence of the additive nonparametric components,

the above model is just the generalized nonlinear regression model. In some applications, the main parametric nonlinear term and its corresponding parameter are of main interest, and the additive nonparametric effects are of secondary importance.

2.2 Estimation and Model Selection

2.2.1 Estimation Algorithm

For estimating the PPGAM, we adopt the penalized likelihood estimation approach by imposing a penalty on the roughness of the unknown functions. Here, the roughness of a smooth function, say s, is defined as follows:

$$J^2(s) = \int_0^1 (s^{(k)}(z))^2 dz,$$

where $k \geq 1$ is a fixed integer and $s^{(k)}$ is the k-th derivative of function s. Unless stated otherwise, k is taken to be 2. The unknown smooth functions s_ν and the parameter θ of the model defined by (2.1) can then be estimated by the argument maximizing the normalized penalized (log-)likelihood function

$$\bar{l}_p = \frac{1}{n} \sum_{i=1}^n Q(Y_i, \mu(T_i)) - \sum_{\nu=1}^w \lambda_\nu J^2(s_\nu), \tag{2.2}$$

where Q is the (log-)likelihood function:

$$Q(y; \mu) = \int_y^\mu \frac{(y-u)}{V(u)} du, \tag{2.3}$$

with V being the conditional variance of the response Y, $\mu(T_i) = F(g_\theta(X_i, Z_i))$, $g_\theta(X_i, Z_i) = h_\theta(X_i) + \sum_{v=1}^w s_v(Z_{iv})$, and λ_ν are positive smoothing parameters to be determined from the data.

First consider the case of known smoothing parameters. Optimization of (2.2) can be achieved iteratively by alternatively updating the function estimates of s_ν and θ as follows. With θ fixed, the model is simply a GAM so the estimation of the smooth functions can be done via the penalized iteratively re-weighted least squares (P-IRLS) (Wood, 2006, p.138). Specifically, it can be readily shown that the function estimates that maximizes (2.2) are natural (thin-plate) smoothing splines; c.f.

(Wood, 2006, p.154). By the theory of reproducing kernel Hilbert space, the function estimates can be represented as linear combinations of the spline basis functions. Consequently, $s_v(z_{iv})$ can be written as $B_{i,\nu}^* \beta_\nu^*$, where $B_{i,\nu}^*$ is the ith row of the design matrix B_ν^* of the spline basis, and β_ν^* is the corresponding parameter vector. Furthermore, the penalties in (2.2), $\lambda_\nu J^2(s_\nu)$, can be expressed as $\lambda_\nu \beta_\nu^{*T} S_\nu \beta_\nu^* / 2$, where S_ν are known matrices (Gu, 2002; Wood, 2006). The estimation of β^*'s can then be carried out by the P-IRLS algorithm. Similarly, for fixed smooth fiunctions, the model becomes a generalized nonlinear regression model and θ can be estimated by maximum likelihood. The iteration can be stopped when the algorithm converges according to some stopping criteria.

In practice, the smoothing parameters are unknown and have to be estimated. We choose the smoothing parameters by minimizing the Generalized Cross Validation score (GCV); see Craven and Wahba (1979) and Golub et al. (1979). This can be conveniently done in the step that updates the function estimates of s_ν via the GAM fitting.

2.2.2   Observed Information

Below in Section 2.3, we derive the asymptotic normality for the penalized likelihood estimator $\hat{\theta}_n$, under some regularity conditions. However, the covariance matrix of the asymptotic normal distribution is rather complex and requires the evaluation of some intractable conditional expectations. Here, we provide an alternative approach to assessing the uncertainty of the estimator that is based on the observed information calculated from the penalized likelihood (2.4). For simplicity, henceforth in this section, we assume that $\nu = 1$ so that Z has one component, hence the subscript ν is dropped from the notation; it is straightforward to extend the formulas to the case of multiple components in Z. Refering to the form of the density (mass) function of the exponential family, define $V^*(\mu) = \ddot{b}(\vartheta)/\omega$ so that

$V = V^*(\mu)\phi$ in the expression of the function Q. The penalized (log-)likelihood can be written as:

$$l_p = \sum_{i=1}^{n} \int_{Y_i}^{\mu(T_i)} \frac{y-u}{\phi V^*(u)} du - n\lambda \beta^{*T} S \beta^*/2. \tag{2.4}$$

Taking the first derivatives of l_p, we get

$$\frac{\partial l_p}{\partial \beta_j^*} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\phi V^*(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j^*} - n\lambda [S\beta^*]_j \qquad j = 1, \ldots, k^* \tag{2.5}$$

$$\frac{\partial l_p}{\partial \theta_q} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\phi V^*(\mu_i)} \frac{\partial \mu_i}{\partial \theta_q} \qquad q = 1, \ldots, r, \tag{2.6}$$

where $\mu_i = F(g_i)$, $g_i = B_i^* \beta^* + h_\theta(x_i)$, and $[\cdot]_j$ denotes the j-th component of the enclosed vector. Define $\delta_i = \frac{y - \mu_i}{\phi V^*(\mu_i)\dot{g}(\mu_i)}$ and $\dot{\delta}_i = \frac{\partial \delta_i}{\partial \mu_i}$, where $\dot{g}(\mu_i) = \frac{\partial g_i}{\partial \mu_i}$. The second derivatives of l_p are given as follows:

$$\frac{\partial^2 l_p}{\partial \beta^* \partial \beta^{*T}} = B^{*T} G_\delta B^* - n\lambda S$$

$$\frac{\partial^2 l_p}{\partial \beta^* \partial \theta^T} = B^{*T} G_\delta \dot{H}_\theta$$

$$\frac{\partial^2 l_p}{\partial \theta \partial \theta^T} = \dot{H}_\theta^T G_\delta \dot{H}_\theta + \ddot{H}_{\delta\theta}, \tag{2.7}$$

where G_δ is an $n \times n$ diagonal matrix whose diagonal elements are $\frac{\dot{\delta}_i}{\dot{g}(\mu_i)}$, $\dot{H}_\theta$ is an $n \times r$ matrix whose (i, q)th element equals $\dot{H}_{\theta iq} = \frac{\partial h(x_i)}{\partial \theta_q}$, and $\ddot{H}_{\delta\theta}$ represents an $r \times r$ matrix with its (q, q')th element equal to $\ddot{H}_{\delta\theta qq'} = \sum_{i=1}^{n} \delta_i \frac{\partial^2 h(x_i)}{\partial \theta_q \partial \theta_{q'}}$. Then the observed Fisher information equals

$$\mathcal{I} = \begin{bmatrix} -B^{*T} G_\delta B^* + n\lambda S & -B^{*T} G_\delta \dot{H}_\theta \\ -\dot{H}_\theta^T G_\delta B^* & -\dot{H}_\theta^T G_\delta \dot{H}_\theta - \ddot{H}_{\delta\theta} \end{bmatrix}, \tag{2.8}$$

evaluated at $\beta^* = \hat{\beta}_n^*, \theta = \hat{\theta}_n$.

The covariance matrix of the penalized likelihood estimator can be approximated by the inverse of the observed Fisher information. This approach yields a covariance matrix for $\hat{\theta}_n$ that is asymptotically equivalent to that given by the limiting normal approximation rigorously derived below, at least for a simple case with conditionally normal responses.

2.2.3 Model Selection

In practical analysis, there are often several competing models to choose from. Hence, it is desirable to develop some model selection criteria for choosing a model appropriate for the data on hand. We adopt a Bayesian approach for using marginal likelihood for choosing between several competing partly parametric GAMs. Indeed, the roughness penalty can be regarded as some sort of prior, in which case the maximum penalized likelihood estimator is equivalent to using the posterior mode for point estimation. The marginal likelihood of a model, say M_j, equals

$$P(D|M_j) = \int P(D|\zeta, M_j)P(\zeta|M_j)d\zeta \qquad j = 1, \ldots, J, \tag{2.9}$$

where D denotes the data, ζ the model parameter, $P(D|\zeta, M_j)$ the likelihood of ζ under M_j, and $P(\zeta|M_j)$ is the prior probability of ζ under M_j. When the prior probabilities $P(M_j)$ of all potential models are the same, the marginal likelihoods are proportional to the posterior model probabilities $P(M_j|D)$, which quantify the evidence for the models lent by the data.

For the PPGAM, the model parameter ζ^T equals (β^{*T}, θ^T). β^{*T} can be further decomposed into (β^T, β_+^T), where β consists of the coefficients of the basis functions of zero roughness, i.e. 1 and z, for $k = 2$, and β_+ are those corresponding to the basis functions with positive penalty, so that $\beta^{*T}S\beta^* = \beta_+^T S_+ \beta_+$. Let the dimensions of β^*, β and β_+ be k^*, k and k_+ respectively. The prior probability of ζ under M_j, $P(\zeta|M_j)$ can be written as $P(\beta_+|\beta, \theta, M_j)P(\beta, \theta|M_j)$. The quadratic form of the roughness penalty motivates us to adopt the following prior:

$$P(\zeta|M_j) = \frac{|n\lambda S_+|^{1/2}}{(2\pi)^{k_+/2}} \exp(-\frac{1}{2}n\lambda\beta_+^T S_+ \beta_+)P(\beta, \theta|M_j). \tag{2.10}$$

We shall assume that the smoothing parameters are known. (In practice, the smoothing parameters will be replaced by the estimates based on GCV.) Substituting the likelihood function and the prior (2.10) into (2.9) yields

$$P(D|M_j) = \frac{|n\lambda S_+|^{1/2}}{(2\pi)^{k_+/2}} \int \exp\{l_p(\zeta)\}P(\beta, \theta|M_j)d\zeta. \tag{2.11}$$

The integral in (2.11) has a closed-form solution only for some simple cases, so we apply Laplace approximation of the integral to obtain

$$
\begin{aligned}
\log P(D|M_j) \approx & l_p(\hat{\zeta}_n) + \log(P(\hat{\beta}_n, \hat{\theta}_n|M_j)) + \frac{k^* + r - k_+}{2} \log(2\pi) \\
& - \frac{1}{2} \log\{\det |H|\} + \frac{1}{2} \log |n\lambda S_+|.
\end{aligned}
\tag{2.12}
$$

where H is the negative Hessian matrix of the log penalized likelihood l_p evaluated at the penalized likelihood estimator. Since there is no specific prior information on the parameters β and θ, the improper constant (flat) priors may be used; hence, the term $\log(P(\hat{\beta}_n, \hat{\theta}_n|M_j))$ in the above approximation (2.12) can be dropped. Alternatively, the preceding claim can be verified subject to some mild regularity conditions on the prior of $P(\beta, \theta|M_j)$; c.f. Schwarz (1978) and Liu and Chan (2009). The criterion derived above is similar to the BIC, in that it "does not depend on the prior distribution" (Schwarz, 1978).

2.3   Asymptotic Properties

In this section, we derive some large-sample properties of the penalized likelihood estimator. These properties will be derived under some general regularity conditions that do not require the conditional distribution of the response variable to belong to the exponential family. Hence the objective function will be referred to below as the penalized quasi-(log-)likelihood function. For simplicity, we consider the case of univariate X and Z with a bounded joint support which, with no loss of generality, equals $[0, 1]^2$. Thus, the conditional mean function on the link scale equals

$$
g_\theta(x, z) = h_\theta(x) + s(z).
\tag{2.13}
$$

It is assumed that

(A1) the function s is k-times differentiable and its k-th derivative $s^{(k)}$ is square

integrable, i.e. it belongs to the Sobolev class $\{s : J(s) < \infty\}$, where

$$J^2(s) = \int_0^1 (s^{(k)}(z))^2 dz, \tag{2.14}$$

and $k \geq 1$ is a fixed positive integer.

In practice, k is generally taken to be 2, corresponding to the use of natural cubic splines. By the Sobolev-embedding theorem, s can be written as:

$$s(z) = s_1(z) + s_2(z), \tag{2.15}$$

where

$$s_1(z) = \sum_{j=1}^{k} \beta_j z^{j-1} = \beta^T \phi(z) = (\beta_1, \ldots, \beta_k)(\phi_1(z), \ldots, \phi_k(z))^T, \tag{2.16}$$

with $\phi_j(z) = z^{j-1}, j = 1, \ldots, k$ and and $|s_2(z)| \leq J(s_2) = J(s) < \infty$. Thus, $h_\theta(x) + s_1(z)$ is the parametric part whereas $s_2(z)$ is the purely nonparametric part of the conditional mean function on the link scale.

The quasi-(log-)likelihood for a single datum equals $Q(y; \mu) = \int_y^\mu \frac{(y-u)}{V(u)} du$, where V is the conditional variance of the response Y. Hence, the penalized quasi-likelihood estimator equals (with the smoothing parameter now denoted as λ_n indicating its dependence on the sample size)

$$\hat{g}_n = \arg\max_{g \in G}[\bar{Q}_n(F(g)) - \lambda_n J^2(s)] \tag{2.17}$$

$$= \arg\max_{g \in G}[\frac{1}{n}\sum_{i=1}^{n} Q(Y_i, \mu(T_i)) - \lambda_n J^2(s)].$$

The quasi-likelihood estimator is expressed as $\hat{g}_n(x, z) = h_{\hat{\theta}_n}(x) + \hat{s}_n(z)$, and the estimated conditional mean is denoted by $\hat{\mu}_n = F(\hat{g}_n)$. The following two functions will be needed below: $f(g) = \dot{F}(g)$ and $l(g) = f(g)/V(F(g))$. For any function $a(Y, T)$: $\Re \times [0, 1]^2 \to \Re$, its empirical L_2 norm is defined as $\|a\|_n = \sqrt{\sum_{i=1}^{n} a^2(Y_i, T_i)/n}$.

Besides condition (A1), the following conditions are required for the consistency of the penalized quasi-likelihood estimator:

(A2) The smoothing parameter converges to 0 with the following rate: $\lambda_n = o_p(n^{-1/2})$, $1/\lambda_n = O_p(n^{2k/(2k+1)})$.

(A3) The error terms $W = Y - \mu_0(T)$ is conditionally sub-exponential almost surely,

i.e. there exists a positive constant C_0 such that

$$E(\exp(|W|/C_0)|T) \le C_0 \qquad \text{a.s..}$$

(A4) The parametric component satisfies a Lipschitz condition: there exists a bounded measurable function w such that $|h_{\theta_1}(x) - h_{\theta_2}(x)| \le w(x)\|\theta_1 - \theta_2\|, \forall\, \theta_1, \theta_2 \in \Theta$, the parameter space (a subset of some Euclidean space) and $\forall\, x$.

(A5) The parametric part of the conditional mean function satisfies the following inequality: $\exists\, K > 0$ such that

$$\left\| h_\theta(X) + \sum_{j=1}^{k} \beta_j Z^{j-1} \right\|_n = \| h_\theta(X) + \beta^T \phi(Z) \|_n \ge K(\|\theta\| + \|\beta\|),$$

(2.18)

with probability $\to 1$, as $n \to \infty$.

(A6) The variance function is bounded below from 0: there exists a constant $C_1 > 0$ such that $V(u) \ge 1/C_1$ for all u in the range of the inverse link function F.

(A7) The function l is bounded away from 0 and infinity, i.e. there exists a constant $C_2 > 0$ such that $1/C_2 \le |l(\xi)| \le C_2$ for all $\xi \in \Re$.

We remark that, except for (A5), these conditions are generally similar to the assumptions employed in the literature of semi-parametric regression, e.g. Mammen and van de Geer (1997). (A5) is akin to the condition of positive-definite design matrix in the study of partly linear regression models. Now, we can state the consistency result.

Theorem 2.3.1 *Under conditions (A1)-(A7), it holds that*

$$\|\hat{g}_n - g_0\|_n = O_p(\lambda_n^{1/2}),$$

(2.19)

$$J(\hat{g}_n) = O_p(1),$$

(2.20)

where g_0 is the true conditional mean function on the link scale.

With the convergence rate assumption on λ_n in (A2), it follows from the preceding theorem that $\hat{g}_n$ is a consistent estimator of g_0 with convergence rate $O_p(n^{-1/4})$. For

the parametric estimator $\hat{\theta}_n$ the convergence rate can be improved to $O_p(n^{-1/2})$ and, in fact, asymptotically normal as shown in the theorem below.

Before stating the theorem on the asymptotic distribution of the parametric estimator $\hat{\theta}_n$, we introduce some notations needed below. Let $\dot{h}_\theta(x)$ be the gradient vector $\dot{h}_\theta = \frac{\partial h_\theta}{\partial \theta}$. Define

$$e_1(z) = \frac{E_0(\dot{h}_0(X)f_0(T)l_0(T)|Z = z)}{E_0(f_0(T)l_0(T)|Z = z)},$$

where the subscript 0 signifies that all expressions are evaluated at the true model. Let

$$e_2(x, z) = \dot{h}_0(x) - e_1(z).$$

Theorem 2.3.2 *Suppose the assumptions for Theorem 2.3.1 are satisfied. Additionally, assume*

(A8) the functions l and f are Lipschitz continuous in some neighborhood of the true parameter value θ_0.

(A9) For all $\theta \in \Theta_1$ and z, $E([\dot{h}_\theta(X) - E(\dot{h}_\theta(X)|Z = z)][\dot{h}_\theta(X) - E(\dot{h}_\theta(X)|Z = z)]^T)$ is positive definite.

(A10) The probability density function of Z is bounded away from 0 on its support.

(A11) $J(e_1) < \infty$.

(A12) $E(f_0 l_0 e_2 e_2^T)$ is positive definite.

(A13) h_θ is continuously differentiable up to order two in some neighborhoods of θ_0.

Then,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = [E(f_0 l_0 e_2 e_2^T)]^{-1}(1/\sqrt{n}\sum_{i=1}^{n} W_i l_0(T_i)e_2(T_i)) + o_p(1)$$

$$\rightsquigarrow N(0, \Sigma), \tag{2.21}$$

where $\Sigma = [E(f_0 l_0 e_2 e_2^T)]^{-1}$.

2.4 Simulation Study

In this section, we report some simulation studies that address the following issues. First, it is desirable to use simulation for evaluating the empirical performance of the parametric estimator of a PPGAM. Second, we use simulation to study how well the inverse of the Fisher information matrix approximates the covariance matrix of the penalized likelihood estimator. Third, we use simulations to assess the effectiveness of the proposed model selection criterion. Fourth, we apply a simulation study to assess the loss in estimation using spline method.

Since the mean function of a PPGAM is the sum of a parametric nonlinear component and some nonparametric functions, we simulated from a number of models using different combinations of parametric and nonparametric components. For simplicity, the predictor on the link scale equals the sum of a parametric nonlinear function and a smooth function that has either a 1-dimensional or 2-dimensional argument. The parametric nonlinear part has a simple form as αx^{ξ}, with the parameters α and ξ varying over different values for achieving different curvatures. For the nonparametric component it is one of the following two functions (c.f. Wood (2006), p. 197):

$$s_1(z_1) = z_1^{11}(10(1-z_1)^6) + 10(10z_1)^3(1-z_1)^{10} - 1.4, \tag{2.22}$$

$$s_2(z_2, z_3) = 0.4 \times 0.6\pi(1.2e^{-(z_2-0.2)^2/0.6^2-(z_3-0.3)^2} + 0.8e^{-(z_2-0.7)^2/0.4^2-(z_3-0.8)^2/0.6^2}), \tag{2.23}$$

where $z_1, z_2, z_3 \in [0,1]$. The smooth functions s_1 and s_2 are shown in Fig. 2.1 and 2.2, respectively. The response is conditionally Gaussian with variance 0.01. All simulation results are based on 1000 replications.

Table 2.1 summarizes the results for the method of penalized likelihood introduced in Section 2.2. Data were generated from two PPGAMs with the same parametric nonlinear component ($\alpha x^{\xi} = 0.2x^{0.75}$), but different smooth functions ($s_1(z_1)$ or $s_2(z_2, z_3)$). For each model, we tried three sample sizes, namely $n =$

Figure 2.1: One dimensional smooth curve
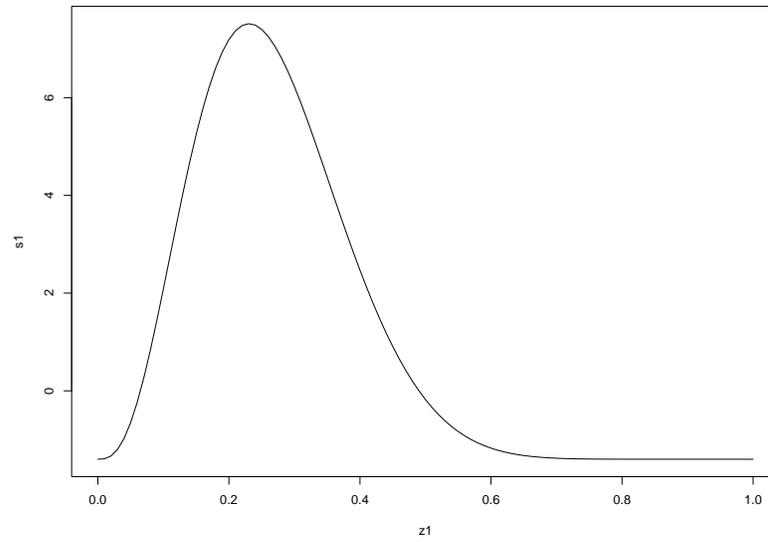
Figure 2.2: Two dimensional smooth plate
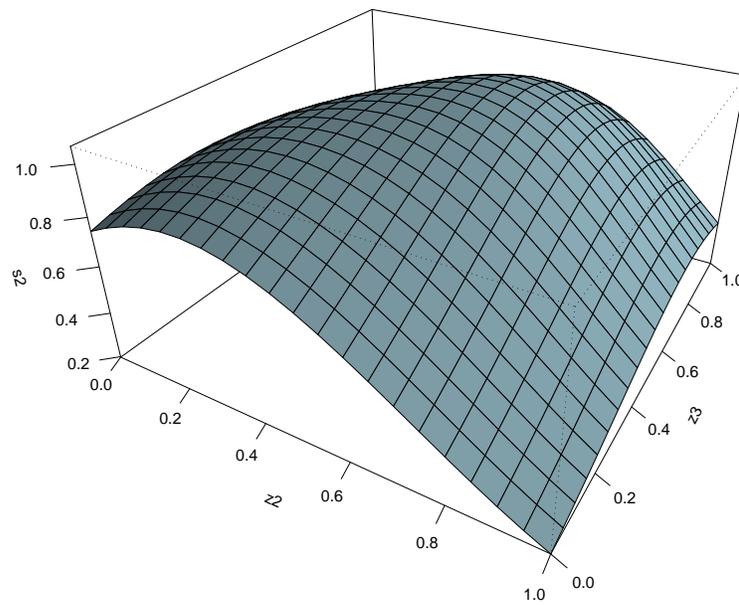
100, 750 and 2000 respectively. The biases and the standard errors (se's), obtained by inverting the observed Fisher information, decrease with increasing sample size, and so do the mean square errors. The biases are generally small except for the case when the smooth function has a 2-dimensional argument and small sample size (100). The standard errors and the sample standard deviations of the parametric estimators (sd's) are close to each other, especially for large samples.

Table 2.1: Parameter estimation results

true parameter	$s_1(z_1)$				$s_2(z_2, z_3)$			
	mean	bias$\times 10^{-3}$	sd	se	mean	bias$\times 10^{-3}$	sd	se
$n = 100$								
$\alpha = 0.2$	0.217	17.0	0.092	0.088	0.212	12.0	0.082	0.088
$\xi = 0.75$	0.756	6.4	0.152	0.151	0.812	62.0	0.328	0.344
$n = 750$								
$\alpha = 0.2$	0.201	1.1	0.028	0.029	0.201	0.6	0.022	0.023
$\xi = 0.75$	0.752	1.7	0.051	0.054	0.758	7.7	0.097	0.103
$n = 2000$								
$\alpha = 0.2$	0.199	-0.8	0.018	0.017	0.200	-0.03	0.014	0.014
$\xi = 0.75$	0.753	3.2	0.033	0.033	0.755	5.0	0.063	0.062

According to Theorem 2.3.2, the parametric estimator of a PPGAM is asymptotic normal. For each simulated dataset, we construct 95% confidence intervals for the parameters α and ξ, based on normal assumption and computing the covariance by inverting the observed Fisher information. Table 2.2 reports the empirical coverage rates of the 95% confidence intervals so constructed and they are generally close to the nominal confidence level of 0.95, with better agreement as the sample size increases. Therefore, inverting observed Fisher information can provide effective

approximation for the variance of the parametric estimator of a PPGAM.

Table 2.2: Confidence interval coverage

Model Structure	Nonparametric	$s_1(z_1)$	$s_1(z_1)$	$s_1(z_1)$	$s_2(z_2, z_3)$
	Parametric	$0.3x^{0.5}$	$0.2x^{0.75}$	$0.1x^{0.9}$	$0.2x^{0.75}$
$n = 100$					
95 % CI coverage	α	0.905	0.926	0.913	0.941
	ξ	0.936	0.955	0.936	0.953
$n = 250$					
95 % CI coverage	α	0.920	0.952	0.923	0.950
	ξ	0.952	0.949	0.931	0.947
$n = 750$					
95 % CI coverage	α	0.937	0.958	0.941	0.960
	ξ	0.948	0.959	0.950	0.966
$n = 2000$					
95 % CI coverage	α	0.943	0.945	0.937	0.945
	ξ	0.957	0.953	0.948	0.947

Table 2.3 reports some simulation results on the empirical performance of the proposed model selection criterion that compares a true model with an alternative model. From the simulation results based on 500 samples, the proposed selection criterion invariably selected the true model except for one case where the parametric nonlinear component is very close to the alternative model with a linear component. For that exceptional case, the true and the alternative models have the same smooth function s_1, while the parametric component of the true model is $0.1x^{0.9}$ and the counterpart of its competitor is a linear function. Considering that the curve $0.1x^{0.9}$ is very close to a line, it is not unexpected that the alternative model outperforms

the true one as it has one less parameter. When the sample size increases to 2000, the correct selection rate between these two models increases to an acceptable level of 0.527. Further increasing of the sample size to 5000 leads to the correct selection rate increasing to a desirable level of 0.816. Overall, the proposed model selection criterion performs well, especially with large samples.

Table 2.3: Relative frequency of selecting the true model

Alternative Models		True	Models	Alternative Models		True Models
		$s_1(z_1)$	$s_1(z_1)$			$s_2(z_2, z_3)$
		$0.2x^{0.75}$	$0.1x^{0.9}$			$0.2x^{0.75}$
nonparametric	parametric			nonparametric	parametric	
$n = 500$						
$s(z_1)$	τx	0.964	0.315	$s(z_2, z_3)$	τx	0.657
$s(z_1)$	$\exp(\alpha x^{\xi})$	0.861	0.815	$s(z_2, z_3)$	$\exp(\alpha x^{\xi})$	0.966
$s(z_1) + s(z_4)$	αx^{ξ}	0.989		$s(z_2, z_3)$	$\exp(\alpha x)$	0.903
$s(z_1) + s(z_4)$	τx	0.999		$s(z_2) + s(z_3)$	αx^{ξ}	1.000
$s(z_1) + s(z_4)$	$\exp(\alpha x^{\xi})$	0.993		$s(z_2)$	αx^{ξ}	1.000
$n = 2000$						
$s(z_1)$	τx		0.527			
$n = 5000$						
$s(z_1)$	τx		0.816			

So far, we consider the case that the main effect of interest admits a parameterized nonlinear form, whereas the other confounding covariate effects are modeled additively and nonparametrically. An interesting question arises regarding the extent of loss in estimation efficiency due to modeling other covariates nonparametrically. If the functional form of the confounding covariate effects is known and made

use in the estimation, this should enhance the estimation efficiency of the nonlinear parametric effects. On the other hand, if the functional form of the confounding covariate effects is mis-specified, then this may induce loss of estimation efficiency.

For studying these issues, we simulated a data with the conditionally normal response whose mean equals $h(x) + s(z)$ where,

$$h(x) = \alpha x^{\xi} = 0.2x^{0.75}$$

$$s(z) = \beta \sin(z) = 0.8 \sin(z),$$

and suppose z is the confounding variate, with its effects estimated via three approaches: (1) assuming the true functional form is known up to the parameter β, (2) assuming a mis-specified functional form obtained from a 13-th order Taylor approximation, up to the unknown multiplicative coefficient β, i.e. $s(z) \approx \beta P13(z) = \beta(z - z^3/3! + z^5/5! - z^7/7! + z^9/9! - z^{11}/11! + z^{13}/13!)$, and (3) modeling it nonparametrically as a natural cubic spline $\hat{s}(z)$. Also, $h(x)$ is assumed to be known up to the parameters α and ξ. Contrasting these three estimation approaches may shed light on the impact of the functional form of the confounding covariate effects on the estimation efficiency of the parametric estimator. From Table 2.4, we find that when employing nonparametric approach in fitting $s(z)$, the biases and mean square errors (MSEs) of the parameter estimation of $h(x)$ are larger than those when the true functional form of $s(z)$ is known. However, for large samples ($n = 750$ and $n = 2000$), the nonparametric fitting approach for $s(z)$ results in more accurate estimation (i.e. with smaller biases and MSEs) of the parameter of $h(x)$ than the fitting approach based on the polynomial approximation, $P13(z)$. Therefore, while there is some loss in the estimation efficiency for the parametric estimator when using the nonparametric approach to model other confounding covariate effects, the loss may be much smaller than that due to mis-specification of the confounding covariate effects.

Table 2.4: Parameter estimation with different fitting approaches for $s(z)$

parameter	true value	$\beta sin(z)$ bias$\times10^{-3}$	mse$\times10^{-4}$	$\beta P13(z)$ bias$\times10^{-3}$	mse$\times10^{-4}$	$\hat{s}(z)$ bias$\times10^{-3}$	mse$\times10^{-4}$
$n=100$							
α	0.2	-0.17	1.90	-24.7	9.06	2.85	22.5
ξ	0.75	1.60	12.6	49.9	50.5	6.52	75.9
$n=750$							
α	0.2	-0.10	0.25	-24.3	6.29	-0.81	2.60
ξ	0.75	0.41	1.67	46.5	25.0	2.84	9.14
$n=2000$							
α	0.2	0.08	0.10	-23.8	5.80	-0.95	0.98
ξ	0.75	-0.20	0.63	44.8	21.3	2.11	3.43

2.5 Leukemia Cancer Risk Analysis with Data from the Atomic Bomb Study

We illustrate the PPGAM with an analysis on the effects of radiation on the propensity of leukemia. Our analysis is based on some data collected by the Atomic Bomb Casualty Commission (ABCC); see `http://www.rerf.jp/index_e.html`. The ABCC was charged by President Turman to conduct investigations of the late effects of radiation among the atomic bomb survivors in Hiroshima and Nagasaki. We formulate a PPGAM that extends a power dose-response curve that describes the relationship between the degree of exposure to radiation from the atomic bomb to the time of onset of leukemia. The analysis adjusts for possible sex and city effects, as well as age effect. (Sex was found to be not significant, and hence dropped from the final fitted model.) Moreover, the age effect is modeled nonparametrically. In order to remove confounding factors, we restrict the analysis to the atomic bomb

survivors who died from diseases and hence excluded the survivors who died from accidents, etc. Furthermore, the analysis is based on the life span data of the atomic bomb survivors in Hiroshima and Nagasaki from 1950 to 1982. The model specifies that the probability of leukemia is linked by the logistic function to the predictor as follows:

$$\text{logit}(p_{c,k,d}) = \beta_0 + \beta_1 1_{c=1} + s(k) + \alpha d^\gamma, \tag{2.24}$$

where c represents the city, with $c = 0$ for Hiroshima and $c = 1$ for Nagasaki; k denotes the survivor's age on the date of the atomic bomb explosion; and d is the dose of radiation that the subject received, more specifically the DS86 Kerma dose. Here, $p_{c,k,d}$ is the (conditional) probability for a survivor to die from leukemia when his or her death was caused by some diseases, and it is assumed that the number of the survivors who died from leukemia follows a distribution, binomial$(n = m_{c,k,d}, p = p_{c,k,d})$, where $m_{c,k,d}$ equals the number of survivors died from some diseases during the study period. The nonparametric function $s(k)$ models the age effect on the risk of dying from leukemia. We adopt the power dosage function αd^γ for modeling the radiation effect on the onset of leukemia (Wijesinha and Piantadosi, 1995). The coefficient α is constrained to be non-negative for interpretation. To avoid constrained optimization, we re-parameterize the power dosage function as αd^γ to $\exp[\beta + \gamma \log(d)]$, so that (2.24) becomes:

$$\text{logit}(p_{c,k,d}) = \beta_0 + \beta_1 1_{c=1} + s(k) + \exp[\beta + \gamma \log(d)], \tag{2.25}$$

Fig. 2.3a displays the estimated age effects, which shows the general trend that younger people exposed to the atomic bomb have a higher probability to die from leukemia during the study period. The dosage function shown in Fig. 2.3b is based on the estimated parameters, $\hat{\beta} = -2.24(0.54)$ and $\hat{\gamma} = 0.55(0.09)$, and the increasing trend in the curve indicates that higher DS86 Kerma dose is associated with higher risk of dying from leukemia.

We consider a simpler case that the dosage is linearly correlated to the logit of $p_{c,k,d}$, i.e. $\gamma = 1$ in the model defined by Eqn. (2.24). Using the model selection criterion introduced in Section 2.2.3, the more general model defined by Eqn. (2.25) has a higher log marginal likelihood (-206.84) than the restricted model (-212.71), which indicates significant curvature in the dose-response relationship.

Figure 2.3: Estimated age and radiation-dose effects on leukemia risk
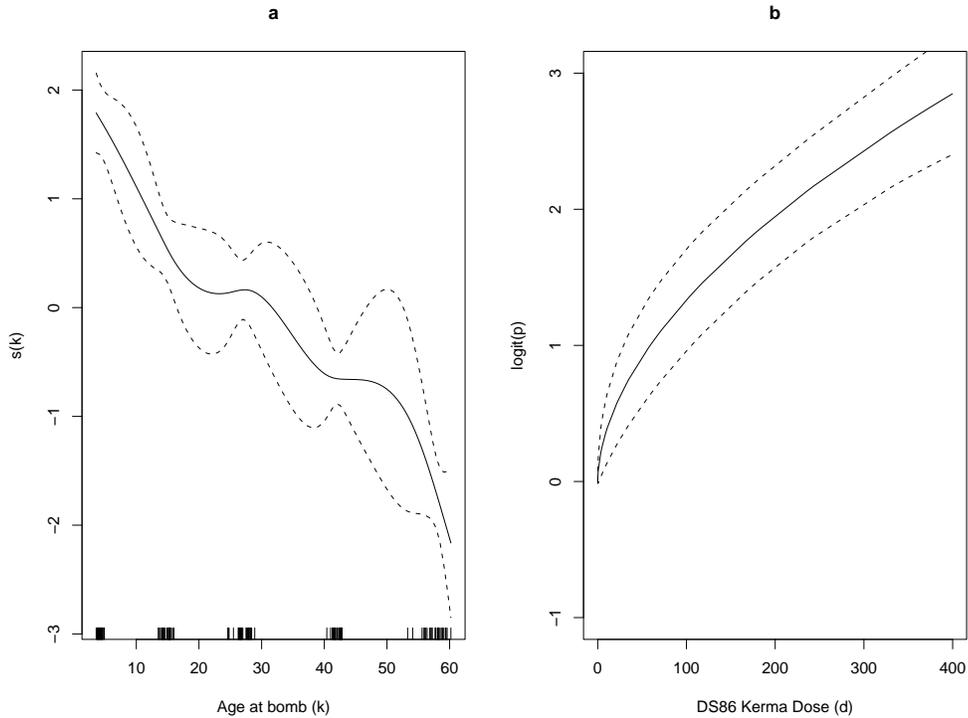
2.6 Proofs of the Theorems

In this section, we gives the proofs of Theorem 2.3.1 and Theorem 2.3.2 respectively. Before giving the proofs, we introduce some further notations. Recall that for a function $a(Y,T) : \Re \times [0,1]^2 \to \Re$, its expected value is denoted as $E(a(Y,T))$, the square of L_2 norm $\|a\|^2 = Ea^2(Y,T)$, and the square of the empirical L_2 norm

$\|a\|_n^2 = \frac{1}{n}\sum_{i=1}^n a^2(Y_i, T_i)$. For a function $a(T) : [0,1]^2 \to \Re$, similar norms can be defined: $\|a\|^2 = Ea^2(T)$, $\|a\|_n^2 = \frac{1}{n}\sum_{i=1}^n a^2(T_i)$, and $|a|_\infty = sup_{t\in[0,1]^2}|a(t)|$. For the parameter $\beta \in \Re^k$, $\|\beta\|^2 = \beta^T\beta$. For $h_\theta(x)$, its first and second derivatives w.r.t. θ are denoted as $\dot{h}_\theta = \frac{\partial h_\theta}{\partial \theta}$ and $\ddot{H}_\theta = \frac{\partial^2 h_\theta}{\partial\theta\partial\theta^T}$, respectively. Here $\theta \in \Re^r$, so $\dot{h}_\theta$ is a $r \times 1$ vector and $\ddot{H}_\theta$ is an $r \times r$ matrix. We write $\ddot{h}_{\theta q\cdot}^T$ for the qth row of $\ddot{H}_\theta$, hence, $\ddot{h}_{\theta q\cdot}^T = \frac{\partial^2 h_\theta}{\partial\theta_q\partial\theta^T}$. The (q^{th}, q'^{th}) entry of $\ddot{H}_\theta$ then equals $\ddot{h}_{\theta qq'} = \frac{\partial^2 h_\theta}{\partial\theta_q\partial\theta_{q'}}$. Below, we write $l_0 = l(g_0), \hat{l}_n = l(\hat{g}_n), f_0 = f(g_0), \hat{f}_n = f(\hat{g}_n), \dot{h}_0 = \frac{\partial h_\theta}{\partial\theta}|_{\theta=\theta_0}, \hat{\dot{h}}_n = \frac{\partial h_\theta}{\partial\theta}|_{\theta=\hat{\theta}_n}, \ddot{H}_0 = \frac{\partial^2 h_\theta}{\partial\theta\partial\theta^T}|_{\theta=\theta_0}, \hat{\ddot{H}}_n = \frac{\partial^2 h_\theta}{\partial\theta\partial\theta^T}|_{\theta=\hat{\theta}_n}, \ddot{h}_{0q\cdot}^T = \frac{\partial^2 h_\theta}{\partial\theta_q\partial\theta^T}|_{\theta=\theta_0}, \hat{\ddot{h}}_{nq\cdot}^T = \frac{\partial^2 h_\theta}{\partial\theta_q\partial\theta^T}|_{\theta=\hat{\theta}_n}, \ddot{h}_{0qq'} = \frac{\partial^2 h_\theta}{\partial\theta_q\partial\theta_{q'}}|_{\theta=\theta_0}$ and $\hat{\ddot{h}}_{nqq'} = \frac{\partial^2 h_\theta}{\partial\theta_q\partial\theta_{q'}}|_{\theta=\hat{\theta}_n}$. In the proofs, we write the smoothing parameter λ_n in (2.17) as λ_n^2, partly because it is positive and partly to match the notation used by Mammen and van de Geer (1997). Note that this implies that the orders specified for the smoothing parameter as stated in condition (A2) becomes $\lambda_n = o_p(n^{-1/4})$ and $1/\lambda_n = O_p(n^{k/(2k+1)})$.

2.6.1   Proof of Theorem 2.3.1

The main part of the proof is to demonstrate the following equation:
$$\frac{(1/n)\sum_{i=1}^n W_i(\hat{\gamma}_n(T_i) - \gamma_0(T_i))}{\|\hat{g}_n - g_0\|_n^{1-1/(2k)}(1 + J(\hat{g}_n))^{1/(2k)} \vee (1 + J(\hat{g}_n))n^{-(2k-1)/(2(2k+1))}} = O_p(n^{-1/2}).$$
$$(2.26)$$

It is noticed that (2.26) has the same form as (3.7) in Mammen and van de Geer (1997). Adapting the proof technique of Lemma 3.1 in Mammen and van de Geer (1997), it can be shown that $\|\hat{g}_n - g_0\| = O_p(\lambda_n)$ and $J(\hat{g}_n) = O_p(1)$, which are the consistency results shown in Theorem 2.3.1.

The rest of the proof is devoted to proving (2.26) and is rather technical. Readers willing to accept (2.26) can skip the rest of the proof.

We now present more technical details needed for verifying (2.26). Based on

the form of the quasi-(log) likelihood in Eqn. (2.3), we define γ_g for fixed y_0

$$\gamma_g = \int_{y_0}^{F(g)} \frac{1}{V(u)} du, \quad g \in G,$$

and $\hat{\gamma}_n = \gamma_{\hat{g}_n} = \int_{y_0}^{F(\hat{g}_n)} \frac{1}{V(u)} du$, $\gamma_0 = \gamma_{g_0} = \int_{y_0}^{F(g_0)} \frac{1}{V(u)} du$.

The function $\gamma = \int_{y_0}^{\mu} \frac{1}{V(u)} du$ has the following properties:

$$\frac{d}{d\gamma} \int_{\mu_0}^{\mu} \frac{u - \mu_0}{V(u)} du = \mu - \mu_0$$

$$\frac{d^2}{d\gamma^2} \int_{\mu_0}^{\mu} \frac{u - \mu_0}{V(u)} du = V(\mu), \tag{2.27}$$

which can be verified as follows. Note that

$$\frac{d\gamma}{d\mu} = \frac{d}{d\mu} \int_{y_0}^{\mu} V(u)^{-1} du = \frac{1}{V(\mu)}.$$

Since γ is monotone in μ, we get $\frac{d\mu}{d\gamma} = V(\mu)$. Then, $\frac{d}{d\gamma} \int_{\mu_0}^{\mu} \frac{u - \mu_0}{V(u)} du = \frac{\mu - \mu_0}{V(\mu)} V(\mu) = \mu - \mu_0$ and $\frac{d^2}{d\gamma^2} \int_{\mu_0}^{\mu} \frac{u - \mu_0}{V(u)} du = \frac{d(\mu - \mu_0)}{d\gamma} = V(\mu)$.

Next, we check the uniformly boundedness condition, $\frac{|\hat{\gamma}_n - \gamma_0|_\infty}{1 + J(\hat{g}_n)} = O_p(1)$, needed for applying Theorem 2.2 of Mammen and van de Geer (1997). This is done in the following steps. Recall from Eqn. (2.3)

$$\bar{Q}_n(\hat{\mu}_n) - \bar{Q}_n(\mu_0)$$

$$= \frac{1}{n} \sum_{i=1}^{n} (y_i - \mu_0(T_i)) \int_{F(g_0)}^{F(\hat{g}_n)} \frac{1}{V(u)} du - \frac{1}{n} \sum_{i=1}^{n} \int_{F(g_0)}^{F(\hat{g}_n)} \frac{u - \mu_0(T_i)}{V(u)} du$$

$$= \frac{1}{n} \sum_{i=1}^{n} W_i(\hat{\gamma}_n(T_i) - \gamma_0(T_i)) - \frac{1}{n} \sum_{i=1}^{n} \int_{\mu_0}^{\hat{\mu}_n} \frac{u - \mu_0(T_i)}{V(u)} du. \tag{2.28}$$

The 2nd term in Eqn. (2.28) equals,

$$\frac{1}{n}\sum_{i=1}^{n}\int_{\mu_0}^{\hat{\mu}_n}\frac{u-\mu_0(T_i)}{V(u)}du$$

$$=\frac{1}{n}\sum_{i=1}^{n}\{(\hat{\gamma}_n-\gamma_0)[\frac{d}{d\gamma}\int_{\mu_0}^{\mu}\frac{u-\mu_0}{V(u)}du|_{\gamma=\gamma_0}]+\frac{1}{2}(\hat{\gamma}_n-\gamma_0)^2[\frac{d^2}{d\gamma^2}\int_{\mu_0}^{\mu}\frac{u-\mu_0}{V(u)}du|_{\gamma=\gamma_n^*}]\}$$

$$\text{(where}\quad\gamma_n^*\quad\text{is between}\quad\gamma_0\quad\text{and}\quad\hat{\gamma}_n)$$

$$=\frac{1}{n}\sum_{i=1}^{n}[(\hat{\gamma}_n-\gamma_0)(\mu-\mu_0)|_{\gamma=\gamma_0}+\frac{1}{2}(\hat{\gamma}_n-\gamma_0)^2V(\mu)|_{\gamma=\gamma_n^*}]$$

$$=\frac{1}{2n}\sum_{i=1}^{n}(\hat{\gamma}_n-\gamma_0)^2V(\mu_n^*)$$

and hence it follows from condition (A6) that

$$\geq\frac{1}{2C_1}\|\hat{\gamma}_n-\gamma_0\|_n^2. \tag{2.29}$$

So, by the Cauchy-Schwarz inequality and the inequality (2.29),

$$\bar{Q}_n(\hat{\mu}_n)-\bar{Q}_n(\mu_0)\leq(\frac{1}{n}\sum_{i=1}^{n}W_i^2)^{1/2}\|\hat{\gamma}_n-\gamma_0\|_n-\frac{1}{2C_1}\|\hat{\gamma}_n-\gamma_0\|_n^2$$

$$\leq O(1)\|\hat{\gamma}_n-\gamma_0\|_n-\frac{1}{2C_1}\|\hat{\gamma}_n-\gamma_0\|_n^2. \tag{2.30}$$

Since $\hat{\mu}_n=F(\hat{g}_n)$ maximizes $\bar{Q}_n(F(g))-\lambda_n^2J^2(g)$, $\lambda_n=o_p(n^{-1/4})$ and $J(g)<\infty$

$$\bar{Q}_n(\hat{\mu}_n)-\bar{Q}_n(\mu_0)\geq\lambda_n^2(J^2(\hat{g}_n)-J^2(g_0))$$

$$\geq o_p(1). \tag{2.31}$$

Now, we claim that $\|\hat{\gamma}_n-\gamma_0\|_n$ and $\|\hat{g}_n\|_n$ are bounded in probability, which can be verified as follows. From (2.30) and (2.31), we have

$$o_p(1)\leq O(1)\|\hat{\gamma}_n-\gamma_0\|_n-\frac{1}{2C_1}\|\hat{\gamma}_n-\gamma_0\|_n^2$$

$$\Rightarrow\|\hat{\gamma}_n-\gamma_0\|_n=O_p(1). \tag{2.32}$$

From condition (A7), we get, $\forall g$ and $\tilde{g}\in\mathscr{G}$

$$\frac{1}{C_2}|g(t)-\tilde{g}(t)|\leq|\gamma_g(t)-\gamma_{\tilde{g}}(t)|\leq C_2|g(t)-\tilde{g}(t)|. \tag{2.33}$$

From (2.32) and (2.33), we have

$$\|\hat{g}_n-g_0\|_n=O_p(1), \tag{2.34}$$

$$\|\hat{g}_n\|_n=O_p(1).$$

We now show that $\frac{|\hat{\gamma}_n-\gamma_0|_\infty}{1+J(\hat{g}_n)}=O_p(1)$. Applying the Sobolev-embedding expression

for the estimator of s in (2.15), we have

$$\hat{g}_n = h_{\hat{\theta}_n}(x) + \hat{s}_1(z) + \hat{s}_2(z)$$

$$= \hat{g}_{1n}(x, z) + \hat{s}_2(z), \tag{2.35}$$

where $\quad |\hat{s}_2(z)| \leq J(\hat{g}_n) < \infty.$

$$\frac{\|\hat{g}_{1n}\|_n}{1 + J(\hat{g}_n)} \leq \frac{\|\hat{g}_n\|_n}{1 + J(\hat{g}_n)} + \frac{\|\hat{s}_2\|_n}{1 + J(\hat{g}_n)}$$

$$= O_p(1). \tag{2.36}$$

It follows from the fact that $s_1(z) = \beta^T \phi(z)$, assumption (A5), and inequality (2.36) that

$$O_p(1) \geq \|\hat{g}_{1n}\|_n \geq K(\|\hat{\theta}_n\| + \|\hat{\beta}_n\|) \quad \text{as} \quad n \to \infty$$

$$\Rightarrow \|\hat{\theta}_n\| = O_p(1) \Rightarrow \|\hat{\theta}_n - \theta_0\| = O_p(1)$$

$$\Rightarrow \|\hat{\beta}_n\| = O_p(1)$$

$$\Rightarrow \frac{\|\hat{\beta}_n\|}{1 + J(\hat{g}_n)} = O_p(1).$$

Since T is a bounded set,

$$\frac{|\hat{s}_{1n}|_\infty}{1 + J(\hat{g}_n)} = O_p(1). \tag{2.37}$$

Because $\|\hat{\theta}_n\| = O_p(1)$, we may, without loss of generality, assume that the parameter space of θ is a bounded subset henceforth in the proof. Thus, from assumption (A4), we have

$$\frac{|h_{\hat{\theta}_n} - h_{\theta_0}|_\infty}{1 + J(\hat{g}_n)} = O_p(1). \tag{2.38}$$

Upon noticing the equality $\frac{|\hat{s}_{2n}|_\infty}{1+J(\hat{g}_n)} = O_p(1)$, equations (2.37) and (2.38), we get

$$\frac{|\hat{g}_n - g_0|_\infty}{1 + J(\hat{g}_n)} = O_p(1), \tag{2.39}$$

and hence it follows from (2.33) that

$$\frac{|\hat{\gamma}_n - \gamma_0|_\infty}{1 + J(\hat{g}_n)} = O_p(1), \tag{2.40}$$

which is the uniformly boundedness condition required for applying Theorem 2.2 of Mammen and van de Geer (1997) in the proof of convergence rates.

We now verify the bracketing condition to apply Theorem 2.2 of Mammen and van de Geer (1997) in the proof of the consistency result.

We first introduce some notations related to entropy and bracketing. Let $\mathscr{A}$ be a subset of a (pseudo-)metric space $(\mathscr{L}, \rho)$, with ρ being a metric. The covering number $N(\delta, \mathscr{A}, \rho)$ is the minimum number of open balls of ρ-radius δ needed for covering the subset $\mathscr{A}$. A δ-bracket is a pair of functions $[a^L, a^U] \subset \mathscr{L}$ with $\rho(a^L, a^U) < \delta$. The bracketing number $N_B(\delta, \mathscr{A}, \rho)$ is the minimum number of δ-brackets needed to cover $\mathscr{A}$. Let $H(\delta, \mathscr{A}, \rho) = \log N(\delta, \mathscr{A}, \rho)$, which denotes the δ-entropy of subset $\mathscr{A}$, and $H_B(\delta, \mathscr{A}, \rho) = \log N_B(\delta, \mathscr{A}, \rho)$ is defined as the δ-entropy with bracketing of $\mathscr{A}$.

In the derivation of the bracketing condition, we need the following Lemmas as technical tools:

Lemma 2.6.1 *Suppose a_1 and a_2 belong to two uniformly bounded classes of functions, $\mathscr{A}_1$ and $\mathscr{A}_2$ respectively, and that for some $0 < \nu_1, \nu_2 < 2$ and $\forall \delta_1, \delta_2 > 0$,*

$$\sup \delta_1^{\nu_1} H(\delta_1, \mathscr{A}_1, |\cdot|_\infty) < \infty,$$

$$\sup \delta_2^{\nu_2} H(\delta_2, \mathscr{A}_2, |\cdot|_\infty) < \infty. \tag{2.41}$$

where the supremum is taken over all $\delta_1 > 0$ ($\delta_2 > 0$). Then $\sup \delta^{\nu_1 \vee \nu_2} H(\delta, \{a_1 \times a_2; a_1 \in \mathscr{A}_1, a_2 \in \mathscr{A}_2\}, |\cdot|_\infty) < \infty$

PROOF of Lemma 2.6.1: According to the definition of entropy, $\forall a_i \in \mathscr{A}_i, i = 1, 2$ and $\delta_1, \delta_2 > 0, \exists a_{ij_i} \in \{a_{ij_i}, j_i = 1, 2, \ldots, N_i\}$ such that $|a_i - a_{ij_i}|_\infty < \delta_i$. From the uniform boundedness condition of $\mathscr{A}_i$ and $|a_i - a_{ij_i}|_\infty < \delta_i$, it is easy to see that $\exists$ a constant C such that $|a_i|_\infty < C$ and $|a_{ij_i}|_\infty < C$.

Because of (2.41) and the inequality (with $\delta_1 = \delta/(2C)$ and $\delta_2 = \delta/(2C)$)

$$|a_1 a_2 - a_{1j_1} a_{2j_2}|_\infty$$

$$\leq |a_1 a_2 - a_{1j_1} a_2|_\infty + |a_{1j_1} a_2 - a_{1j_1} a_{2j_2}|_\infty$$

$$\leq C\delta_1 + C\delta_2$$

$$= \delta, \tag{2.42}$$

we have $\forall \delta > 0$, $\exists$ a constant K such that

$$H(\delta, \{a_1 \times a_2; a_1 \in \mathscr{A}_1, a_2 \in \mathscr{A}_2\}, |\cdot|_\infty) \leq K_1(\frac{2C}{\delta})^{\nu_1} + K_2(\frac{2C}{\delta})^{\nu_2}$$

$$\leq K(\frac{1}{\delta})^{\nu_1 \vee \nu_2}. \qquad (2.43)$$

This completes the proof of Lemma 2.6.1. $\square$

Lemma 2.6.2 *If $H = \{h_\theta; \theta \in \Theta^*\}$ is a collection of the functions such that:*

(i) Θ^* *is a bounded subset of* $\Re^r$

(ii) there exists a bounded function w such that

$$|h_{\theta_1}(x) - h_{\theta_2}(x)| \leq w(x)\|\theta_1 - \theta_2\| \quad \text{for any } \theta_1 \text{ and } \theta_2.$$

Then there exists a constant K, depending on Θ^* and r, such that the bracketing numbers satisfy $N_B(\delta|w|_\infty, H, |\cdot|_\infty) < K(\frac{diam(\Theta^*)}{\delta})^r$, $\forall \ 0 < \delta < diam(\Theta^*)$. Thus, the entropy of H is of order $(\frac{1}{\delta})^{1/k}$.

PROOF of Lemma 2.6.2: We denote $diam(\Theta^*)$ as the diameter of Θ^*, which is defined as the least upper bound of $\|\theta_1 - \theta_2\|$ for any θ_1 and θ_2 in Θ^*. We follow Example 19.7 of van der Vart (1998). Note that any bracket of the form $[h_\theta - \epsilon w, h_\theta + \epsilon w]$ is of bracketing size $2\epsilon|w(x)|_\infty$, in sup norm. If $\|\theta_1 - \theta_2\| \leq \epsilon$, then $h_{\theta_1} - \epsilon w \leq h_{\theta_2} \leq h_{\theta_1} + \epsilon w$. Thus, if the collection of open balls $B(\theta_j, \epsilon), j = 1, \ldots, N^*$ cover Θ^*, then the brackets $[h_{\theta_j} - \epsilon w, h_{\theta_j} + \epsilon w], (j = 1, \ldots, N^*)$ cover $\{h_\theta, \theta \in \Theta^*\}$. Note that Θ^* can be covered by the open balls with radius 2ϵ, the number of which is the order $(\frac{diam\Theta^*}{\delta})^r$. Therefore, $N_B(\delta|w|_\infty, H, |\cdot|_\infty) < K(\frac{diam(\Theta^*)}{\delta})^r$, $\forall \ 0 < \delta < diam(\Theta^*)$, and Lemma 2.6.2 is proved. $\square$

By Theorem 2.1 of Mammen and van de Geer (1997), for each $0 < C < \infty$,

$$\sup \delta^{1/k} H(\delta, \{s; |s|_\infty \leq C, J(s) \leq C\}, |\cdot|_\infty) < \infty. \qquad (2.44)$$

Thus, Eqn. (2.44) gives an entropy condition for the metric space $(\{s; |s|_\infty \leq C, J(s) \leq C\}, |\cdot|_\infty)$. In the proof of the convergence rates, we need to verify the

following bracketing condition for the metric space $(\{\frac{\gamma_\theta - \gamma_{\theta_0}}{1 + J(g_\theta)}; g_\theta \in G, \theta \in \Theta_1\}, \|\cdot\|_n)$,

$$\limsup_{n \to \infty} \sup_{\delta > 0} \delta^{1/k} H_B(\delta, \{\frac{\gamma_g - \gamma_{g_0}}{1 + J(g)}; g \in \mathscr{G}, \theta \in \Theta_1\}, \|\cdot\|_n) < \infty. \qquad (2.45)$$

The verification is done in the following steps.

We first claim that

$$\sup \delta^{1/k} H(\delta, \{\frac{s - s_0}{1 + J(g)}; |s - s_0|_\infty \le C, J(s) \le C\}, |\cdot|_\infty) < \infty. \qquad (2.46)$$

Proof of (2.46): By the entropy condition given in (2.44), $\exists$ a constant K, $\forall \delta > 0$,

such that the entropy numbers satisfy

$$\log N(\delta, \{s - s_0; |s - s_0|_\infty \le C, J(s) \le C\}, |\cdot|_\infty) \le K(\frac{1}{\delta})^{1/k}.$$

By the condition $0 \le J(s) = J(g) \le C$, it follows that $\frac{1}{1+J(g)} \in [\frac{1}{1+C}, 1]$. Consider

the grid of points, $\{\frac{1}{\kappa_p}, p = 1, \dots, v\} \subseteq [\frac{1}{1+C}, 1]$ with the property that $|\frac{1}{\kappa_p} - \frac{1}{\kappa_{p-1}}| \le \tau$

for each p. Clearly $v = O(\frac{1}{\tau})$. Moreover, $\exists$ a constant K^* such that $\forall \tau > 0$, the

entropy

$$\log N(\tau, \{\frac{1}{1 + J(g)}; J(g) \le C\}, |\cdot|_\infty) \le \log K^* + \log(\frac{1}{\tau}).$$

When τ decreases to zero, $\log(\frac{1}{\tau})$ grows slower than $(\frac{1}{\tau})^{1/k}$, for any fixed $k > 0$.

Applying Lemma 2.6.1, we get

$$\log N(\delta, \{\frac{s - s_0}{1 + J(g)}; |s - s_0|_\infty \le C, J(s) \le C\}, |\cdot|_\infty) \le K(\frac{1}{\delta})^{1/k},$$

hence the entropy condition in (2.46) is satisfied. $\square$

Next we claim that

$$\sup \delta^{1/k} H_B(\delta, \{\frac{s - s_0}{1 + J(g)}; |s - s_0|_\infty \le C, J(s) \le C\}, |\cdot|_\infty) < \infty. \qquad (2.47)$$

This claim follows from the observation that for any open ball, say, $B(\frac{s_j - s_0}{\kappa_p}, \delta)$, it

is a subset of the bracket $[a^L, a^U]$ where $a^L = \frac{s_j - s_0}{\kappa} - \delta$ and $a^U = \frac{s_j - s_0}{\kappa} + \delta$, and the

bracket is of size 2δ.

Thirdly, we claim that

$$\sup \delta^{1/k} H_B(\delta, \{\frac{h_\theta - h_{\theta_0}}{1 + J(g_\theta)}; g_\theta \in \mathscr{G}, \theta \in \Theta_1\}, |\cdot|_\infty) < \infty, \tag{2.48}$$

Proof of (2.48): Applying Lemma 2.6.2, and invoking the boundedness property of Θ_1 and assumption (A4), we get

$$\sup \delta^{1/k} H_B(\delta, \{h_\theta - h_{\theta_0}; g \in \mathscr{G}, \theta \in \Theta_1\}, |\cdot|_\infty) < \infty. \tag{2.49}$$

Similar to (2.46), the bracketing condition (2.48) can be derived from (2.49). This completes the proof of (2.48). $\square$

Combining (2.47) and (2.48), we get:

$$\sup \delta^{1/k} H_B(\delta, \{\frac{g - g_0}{1 + J(g)}; g \in \mathscr{G}, \theta \in \Theta_1\}, |\cdot|_\infty) < \infty. \tag{2.50}$$

Since $\|\cdot\|_n \leq |\cdot|_\infty$, a δ-brackets in the metric space $(\mathscr{A}, |\cdot|_\infty)$ is also a δ-bracket in the metric subset $(\mathscr{A}, \|\cdot\|_n)$, and $N_B(\delta, \mathscr{A}, \|\cdot\|_n) \leq N_B(\delta, \mathscr{A}, |\cdot|_\infty)$. Therefore, the following bracketing condition is satisfied:

$$\limsup_{n \to \infty} \sup_{\delta > 0} \delta^{1/k} H_B(\delta, \{\frac{g - g_0}{1 + J(g)}; g \in \mathscr{G}, \theta \in \Theta_1\}, \|\cdot\|_n) < \infty. \tag{2.51}$$

Hence, it follows from (2.33) that (2.45) holds.

It follows from (2.45), (2.40) and (A3) and an application of Theorem 2.2 of Mammen and van de Geer (1997) that (2.26) holds, which completes the proof of Theorem 2.3.1. $\square$

2.6.2 Proof of Theorem 2.3.2

The main task of the proof is to demonstrate the following set of equations:

$$0 = \frac{1}{n} \sum_{i=1}^{n} W_i l_0(T_i) e_{2q}(T_i)$$
$$- [E(f_0 l_0 e_{2q} \dot{h}_0^T) + o_p(1)](\hat{\theta}_n - \theta_0) + o_p(n^{-1/2}), \tag{2.52}$$

where i=1,2,...,r. Note that $E(f_0 l_0 e_2 \dot{h}_0^T) = E(f_0 l_0 e_2 e_2^T)$. Putting the r equations in (2.52) in matrix form and after some algebra, we obtain

$$[E(f_0 l_0 e_2 e_2^T) + o_p(1) 1_{r \times r}](\hat{\theta}_n - \theta_0) = \frac{1}{n} \sum_{i=1}^{n} W_i l_0(T_i) e_2(T_i) + o_p(n^{-1/2}). \qquad (2.53)$$

Applying (A12) and after some algebra, we obtain the asymptotic normality result stated in Theorem 2.3.2.

Now, we proceed to rigorously demonstrate (2.52). From Theorem 2.3.1 in Section 2.3, we get

$$\|\hat{g}_n - g_0\|_n = O_p(n^{-1/4}) = o_p(1). \qquad (2.54)$$

Using the same arguments as in the proof of Theorem 2.3.1, we can prove that $|\hat{g}_n - g_0|_\infty = O_p(1)$, and the bracketing condition $\sup_{\delta > 0} \delta^{1/k} H_B(\delta, \{g - g_0; g \in G, \theta \in \Theta_1\}, \|\cdot\|) < \infty$. Theorem 2.3 of Mammen and van de Geer (1997) then implies that

$$\|\hat{g}_n - g_0\| = o_p(1). \qquad (2.55)$$

Note that

$$\|\hat{g}_n - g_0\|^2 = E[(h_{\hat{\theta}_n}(\tilde{X}) - h_{\theta_0}(\tilde{X})) + (\hat{s}_n(\tilde{Z}) - s_0(\tilde{Z}))]^2,$$

where $(\tilde{X}, \tilde{Z})$ is independent of the data $(X_1, Z_1), \ldots, (X_n, Z_n)$, but shares the same distribution. Because

$$\|\hat{g}_n - g_0\|^2 = E[(\hat{\theta}_n - \theta_0)^T \dot{h}_{\theta_n^*} + (\hat{s}_n - s_0)]^2$$

$$= E[(\hat{\theta}_n - \theta_0)^T (\dot{h}_{\theta_n^*} - E[\dot{h}_{\theta_n^*} | \tilde{Z} = \tilde{z}]) + (\hat{\theta}_n - \theta_0)^T E[\dot{h}_{\theta_n^*} | \tilde{Z} = \tilde{z}]$$

$$+ (\hat{s}_n - s_0)]^2$$

$$= (\hat{\theta}_n - \theta_0)^T E[(\dot{h}_{\theta_n^*} - E[\dot{h}_{\theta_n^*} | \tilde{Z} = \tilde{z}])(\dot{h}_{\theta_n^*} - E[\dot{h}_{\theta_n^*} | \tilde{Z} = \tilde{z}])^T](\hat{\theta}_n - \theta_0)$$

$$+ E[(\hat{\theta}_n - \theta_0)^T E[\dot{h}_{\theta_n^*} | \tilde{Z} = \tilde{z}] + (\hat{s}_n - s_0)]^2,$$

(2.55), and assumption (A9), it follows that $\|\hat{\theta}_n - \theta_0\| = o_p(1)$.

Now,

$$\|\hat{s}_n - s_0\| \leq \|\hat{g}_n - g_0\| + \|(\hat{\theta}_n - \theta_0)^T \dot{h}_{\theta_n^*}\|$$

$$= \|\hat{g}_n - g_0\| + [(\hat{\theta}_n - \theta_0)^T E[\dot{h}_{\theta_n^*} \dot{h}_{\theta_n^*}^T](\hat{\theta}_n - \theta_0)]^{\frac{1}{2}}, \qquad (2.56)$$

where θ_n^* is some vector between $\hat{\theta}_n$ and θ_0. Because of (A13) and the fact that $\|\hat{\theta}_n - \theta_0\| = o_p(1)$, $E[\dot{h}_{\theta_n^*} \dot{h}_{\theta_n^*}^T]$ is bounded, so the second term of the right hand side of (2.56) is $o_p(1)$. Therefore, $\|\hat{s}_n - s_0\| = o_p(1)$. It follows from (A10) and the reproducing kernel Hilbert space theory (Gu, 2002) that

$$\sup_{z \in support(Z)} |\hat{s}_n(z) - s_0(z)| = o_p(1).$$

Then, (A4) and $\|\hat{\theta}_n - \theta_0\| = o_p(1)$ entails that $|h_{\hat{\theta}_n} - h_{\theta_0}|_\infty = o_p(1)$. We may assume that

$$|\hat{g}_n - g_0|_\infty \leq \eta_0, \qquad (2.57)$$

which is needed for using (A8).

With assumption (A11), we consider the following final-dimensional sub-models passing through $\hat{g}_n$:

$$\hat{g}_{nm}(x, z) = h_{\hat{\theta}_n + m}(x) + [\hat{s}(z) - m^T e_1(z)], \qquad (2.58)$$

where $m = (m_1 \ldots m_q \ldots m_r)^T$. Thus,

$$\frac{d}{dm}[\bar{Q}_n(F(\hat{g}_{nm})) - \lambda_n^2 J^2(\hat{g}_{nm})]|_{m=0} = 0. \qquad (2.59)$$

Equation (2.59) is a system of r equations:

$$\frac{\partial}{\partial m_q}[\bar{Q}_n(F(\hat{g}_{nm})) - \lambda_n^2 J^2(\hat{g}_{nm})]|_{m=0} = 0 \quad \text{for} \quad q = 1 \ldots r. \qquad (2.60)$$

Consider the first term on the left hand side of (2.60)

$$\frac{\partial}{\partial m_q}\bar{Q}_n(F(\hat{g}_{nm}))\Big|_{m=0}$$

$$=\frac{\partial}{\partial m_q}\left[\frac{1}{n}\sum_{i=1}^{n}W_i\hat{\gamma}_{nm}(T_i)-\frac{1}{n}\sum_{i=1}^{n}\int_{\mu_0}^{\hat{\mu}_{nm}}\frac{u-\mu_0(T_i)}{V(u)}du\right]\Big|_{m=0}$$

$$=\frac{1}{n}\sum_{i=1}^{n}W_i\frac{\partial\hat{\gamma}_{nm}}{\partial\hat{g}_{nm}}\frac{d\hat{g}_{nm}}{ds_q}\Big|_{m=0}-\frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_n-\mu_0)\frac{\partial\hat{\gamma}_{nm}}{\partial\hat{g}_{nm}}\frac{d\hat{g}_{nm}}{ds_q}\Big|_{m=0}$$

$$=\frac{1}{n}\sum_{i=1}^{n}W_i\hat{l}_n(T_i)[\hat{\dot{h}}_{nq}(T_i)-e_{1q}(Z_i)]-\frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_n-\mu_0)\hat{l}_n(T_i)[\hat{\dot{h}}_{nq}(T_i)-e_{1q}(Z_i)]$$

$$=I_q-II_q.$$

where $\hat{\mu}_{nm}=F(\hat{g}_{nm})$ and $\hat{\gamma}_{nm}=\gamma_{\hat{g}_{nm}}$. We also define e_{1q} as the qth component of e_1, and similar defined are e_{2q}, $\dot{h}_{0q}$ and $\hat{\dot{h}}_{nq}$. Now,

$$I_q=\frac{1}{n}\sum_{i=1}^{n}W_i\hat{l}_n(T_i)[\hat{\dot{h}}_{nq}(T_i)-e_{1q}(Z_i)]$$

$$=\frac{1}{n}\sum_{i=1}^{n}W_i\hat{l}_n(T_i)[\hat{\dot{h}}_{nq}(T_i)-\dot{h}_{0q}(T_i)]+\frac{1}{n}\sum_{i=1}^{n}W_i\hat{l}_n(T_i)e_{2q}(T_i)$$

$$=I_{1q}+I_{2q}.$$

Consider

$$I_{2q}=\frac{1}{n}\sum_{i=1}^{n}W_i\hat{l}_n(T_i)e_{2q}(T_i)$$

$$=\frac{1}{n}\sum_{i=1}^{n}W_i l_0(T_i)e_{2q}(T_i)+\frac{1}{n}\sum_{i=1}^{n}W_i[\hat{l}_n(T_i)-l_0(T_i)]e_{2q}(T_i)$$

$$=\frac{1}{n}\sum_{i=1}^{n}W_i l_0(T_i)e_{2q}(T_i)+o_p(n^{-\frac{1}{2}}).$$

That the second term in the preceding equation is $o_p(n^{-\frac{1}{2}})$ can be proved by applying Theorem 2.4 of Mammen and van de Geer (1997). To apply this theorem, we need to show the bracketing condition

$$\sup\delta^{1/k}H_B(\delta,\{[y-\mu_0(t)]l(g(t))e_{2q}(t);|g-g_0|_\infty\leq\eta_0,g\in\mathscr{G},\theta\in\Theta_1\},$$

$$\|\cdot\|)<\infty.$$

The above bracketing condition can be derived by noting (i) the bracketing condition

for the class $\{l(g(t)); |g - g_0|_\infty \leq \eta_0, g \in \mathscr{G}, \theta \in \Theta_1\}$ (from (A8) and Theorem 2.1 in Mammen and van de Geer (1997)), (ii) $y - \mu_0(t)$ is a fixed P_0-square integrable function, and (iii) $e_{2q}(t)$ is a fixed bounded function. Additionally, A notice that $E[W(l(T) - l_0(T))e_{2q}(T)] = 0$. Next consider

$$I_{1q} = \frac{1}{n} \sum_{i=1}^{n} W_i \hat{l}_n(T_i)[\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)]$$

$$= \frac{1}{n} \sum_{i=1}^{n} W_i l_0(T_i)[\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)] + \frac{1}{n} \sum_{i=1}^{n} W_i[\hat{l}_n(T_i) - l_0(T_i)][\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)]$$

$$= I_{11q} + I_{12q}.$$

It follows from (A13) and Lemma 2.6.2 that the following bracketing condition holds:

$$\sup \delta^{1/k} H_B(\delta, \{\dot{h}_q(t) - \dot{h}_{0q}(t); |g - g_0|_\infty \leq \eta_0, g \in \mathscr{G}, \theta \in \Theta_1\}, \|\cdot\|) < \infty.$$

Assumption (A8) implies that, $l_0(t)$ is a fixed bounded function. Therefore, the following bracketing condition for applying Theorem 2.4 of Mammen and van de Geer (1997) is satisfied.

$$\sup \delta^{1/k} H_B(\delta, \{[y - \mu_0(t)]l_0(g(t))[\dot{h}_q(t) - \dot{h}_{0q}(t)]; |g - g_0|_\infty \leq \eta_0, g \in \mathscr{G}, \theta \in \Theta_1\},$$

$$\|\cdot\|) < \infty.$$

Using similar method for bounding the second term in I_{2q}, we get

$$I_{11q} = \frac{1}{n} \sum_{i=1}^{n} W_i l_0(T_i)[\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)]$$

$$= o_p(n^{-\frac{1}{2}}). \tag{2.61}$$

Applying Lemma 2.6.1 and noticing that $\|\cdot\| \leq |\cdot|_\infty$, the following bracketing condition is obtained:

$$\sup \delta^{1/k} H_B(\delta, \{[l(t) - l_0(t)][\dot{h}_q(t) - \dot{h}_{0q}(t)]; |g - g_0|_\infty \leq \eta_0, g \in G, \theta \in \Theta_1\},$$

$$\|\cdot\|) < \infty.$$

Following the same argument for bounding I_{11q}, we obtain

$$I_{12q} = o_p(n^{-\frac{1}{2}}).$$

Therefore,

$$I_q = \frac{1}{n} \sum_{i=1}^{n} W_i l_0(T_i) e_{2q}(T_i) + o_p(n^{-\frac{1}{2}}).　\qquad (2.62)$$

Now, consider the decomposition:

$$II_q = \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_n - \mu_0)\hat{l}_n(T_i)[\hat{\dot{h}}_{nq}(T_i) - e_{1q}(Z_i)]$$

$$= \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_n - \mu_0)\hat{l}_n(T_i)[\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)] + \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_n - \mu_0)\hat{l}_n(T_i)e_{2q}(T_i)$$

$$= II_{1q} + II_{2q},$$

where

$$II_{1q} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_n - \mu_0)\hat{l}_n(T_i)[\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)]$$

$$= \frac{1}{n} \sum_{i=1}^{n} [(\hat{g}_n - g_0)f_0(T_i)]l_0(T_i)[\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)]$$

$$+ \frac{1}{n} \sum_{i=1}^{n} [(\hat{\mu}_n - \mu_0) - (\hat{g}_n - g_0)f_0(T_i)]l_0(T_i)[\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)]$$

$$+ \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_n - \mu_0)(\hat{l}_n - l_0)[\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)]$$

$$= II_{11q} + II_{12q} + II_{13q}.$$

(A13) entails the Lipschitz condition: $\exists$ constant $C_1^* > 0$ such that $|\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)| \le C_1^* 1_r^T |\hat{\theta}_n - \theta_0|$. It follows from (A8), (A13) and the Cauchy-Schwarz inequality that

$$|II_{11q}| = \frac{1}{n} \sum_{i=1}^{n} |f_0(T_i)||l_0(T_i)||\hat{g}_n - g_0||\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)|$$

$$\le C_3 C_4 C_1^* \|(\hat{g}_n - g_0)\|_n 1_r^T |\hat{\theta}_n - \theta_0|$$

$$= o_p(1)|\hat{\theta}_n - \theta_0|.$$

$$|II_{12q}| = \frac{1}{n}\sum_{i=1}^{n}|(\hat{\mu}_n - \mu_0) - (\hat{g}_n - g_0)f_0(T_i)||l_0(T_i)||\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)|$$

$$= \frac{1}{n}\sum_{i=1}^{n}|(\hat{g}_n - g_0)(f_{g_n^*} - f_0)||l_0(T_i)||\hat{\dot{h}}_{nq}(T_i) - \dot{h}_{0q}(T_i)|$$

$$\leq C_3 C_4 C_1^* \|(\hat{g}_n - g_0)\|_n^2 1_r^T |\hat{\theta}_n - \theta_0|$$

$$= o_p(1)|\hat{\theta}_n - \theta_0|.$$

Similarly,

$$|II_{13q}| \leq o_p(1)|\hat{\theta}_n - \theta_0|.$$

Next consider

$$II_{2q} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_n - \mu_0)\hat{l}_n(T_i)e_{2q}(T_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n}[(\hat{g}_n - g_0)f_0(T_i)]l_0(T_i)e_{2q}(T_i)$$

$$+ \frac{1}{n}\sum_{i=1}^{n}[(\hat{\mu}_n - \mu_0) - (\hat{g}_n - g_0)f_0(T_i)]l_0(T_i)e_{2q}(T_i)$$

$$+ \frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_n - \mu_0)(\hat{l}_n - l_0)e_{2q}(T_i)$$

$$= II_{21q} + II_{22q} + II_{23q}.$$

Observe that

$$\hat{g}_n(x, z) - g_0(x, z)$$

$$= \hat{h}_n(x) - h_0(x) + \hat{s}_n(z) - s_0(z)$$

$$= \dot{h}_0^T(\hat{\theta}_n - \theta_0) + (\dot{h}_{\theta_n^*} - \dot{h}_0)^T(\hat{\theta}_n - \theta_0) + (\hat{s}_n(z) - s_0(z)).$$

Now,

$$II_{21q} = \frac{1}{n} \sum_{i=1}^{n} [(\hat{g}_n - g_0) f_0(T_i)] l_0(T_i) e_{2q}(T_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} f_0(T_i) l_0(T_i) e_{2q}(T_i) \dot{h}_0^T (\hat{\theta}_n - \theta_0)$$

$$+ \frac{1}{n} \sum_{i=1}^{n} f_0(T_i) l_0(T_i) e_{2q}(T_i) (\dot{h}_{\theta_n^*} - \dot{h}_0)^T (\hat{\theta}_n - \theta_0)$$

$$+ \frac{1}{n} \sum_{i=1}^{n} f_0(T_i) l_0(T_i) e_{2q}(T_i) (\hat{s}_n(Z_i) - s_0(Z_i))$$

$$= (E(f_0 l_0 e_{2q} \dot{h}_0^T) + o_p(1))(\hat{\theta}_n - \theta_0) + o_p(n^{1/2}).$$

Similar to the approach for bounding I_{2q}, the $o_p(n^{1/2})$ term above is obtained by applying Theorem 2.4 of Mammen and van de Geer (1997). Also,

$$|II_{22q}| \le C_4 \frac{1}{n} \sum_{i=1}^{n} (\hat{g}_n - g_0)^2 |l_0(T_i) e_{2q}(T_i)|$$

$$\le C_3 C_4 \|\hat{g}_n - g_0\|_n^2$$

$$= o_p(n^{-1/2}).$$

Similarly,

$$|II_{23}| \le o_p(n^{-1/2}).$$

Therefore,

$$II_q = (E(f_0 l_0 e_{2q} \dot{h}_0^T) + o_p(1))(\hat{\theta}_n - \theta_0) + o_p(n^{-1/2}). \tag{2.63}$$

The second term in equation(2.59),

$$\frac{d}{ds_q} \lambda_n^2 J^2(\hat{g}_{nm})|_{m=0} \le 2\lambda_n^2 J(\hat{g}_n) J(e_{1q}) = o_p(n^{-1/2}). \tag{2.64}$$

Taking (2.59), (2.62), (2.63) and (2.64) together yields (2.52). □

2.7 Assessing the Approximation of the Asymptotic Covariance

In Theorem 2.3.2, we show the asymptotic normality of the penalized estimator $\hat{\theta}_n$:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N(0, \Sigma),$$

where $\Sigma = [E(f_0 l_0 e_2 e_2^T)]^{-1}$, $f = \dot{F}(g)$, $l = f/V$, $e_2(x, z) = \dot{h}_0(x) - e_1(z)$ and $e_1(z) = E(\dot{h}_0(X) f_0(T) l_0(T)|Z = z)/E(f_0(T) l_0(T)|Z = z)$. As Σ depends on intractable conditional expectations, we provide an alternative approach to approximating the covariance matrix of the estimator based on the inverse of the observed Fisher information. The Fisher information matrix has the following form:

$$\mathcal{I} = \begin{bmatrix} -B^{*T} G_\delta B^* + n\lambda S & -B^{*T} G_\delta \dot{H}_\theta \\ -\dot{H}_\theta^T G_\delta B^* & -\dot{H}_\theta^T G_\delta \dot{H}_\theta - \ddot{H}_{\delta\theta} \end{bmatrix},$$

where $G_\delta = \text{Diag}\{\dot{\delta}_i/\dot{g}(\mu_i)\}$ with $\delta_i = (y - \mu_i)/[\phi V^*(\mu_i)\dot{g}(\mu_i)]$, $\dot{\delta}_i = \partial\delta_i/\partial\mu_i$ and the conditional variance function V is reparameterized as $\phi V^*(\mu_i)$; B^* is the design matrix of the spline basis for the smooth functions that are part of the conditional mean function, so the estimator of $s(z)$ can be written as a linear combination of the columns of B^*; $\dot{H}_\theta$ is an $n \times r$ matrix whose (i, q)th element equals $\dot{H}_{\theta iq} = \partial h(x_i)/\partial\theta_q$; and $\ddot{H}_{\delta\theta}$ represents an $r \times r$ matrix with its (q, q')th element equal to $\ddot{H}_{\delta\theta qq'} = \sum_{i=1}^n \delta_i[\partial^2 h(x_i)/\partial\theta_q \partial\theta_{q'}]$.

By assumption (A1), the smooth function s belongs to the Sobolev class $\{s : \int_0^1 (s^{(k)}(z))^2 dz < \infty\}$, which is a reproducing kernel Hilbert space (Gu, 2002). Therefore, the function s can be written as:

$$s(z) = \langle R_z(\cdot), s(\cdot)\rangle$$

$$= s_1(z) + s_2(z),$$

where $R_z(\cdot)$ is the reproducing kernel; $s_1(z) = \sum_{j=1}^k \beta_j z^{j-1}$ is the projection of $s(z)$ onto the space of polynomials of degree less than k, i.e. $\mathcal{H}_0 = \{s : s^{(k)} = 0\}$ with an inner product $\langle s, t\rangle_0 = \sum_{j=1}^k s^{(j-1)} t^{(j-1)}$; and $s_2(z)$ is the projection of $s(z)$ onto the orthogonal complement of $\mathcal{H}_0$, i.e.

$$\mathcal{H}_1 = \{s : s^{(j-1)}(0) = 0, j = 1, \ldots, k, \int_0^1 (s^{(k)}(z))^2 dz < \infty\}$$

with an inner product $\langle s, t\rangle_1 = \int_0^1 s^{(k)} t^{(k)} dz$. Further decomposing the $s_2(z)$, $s(z)$

can be written as

$$s(z) = \sum_{j=1}^{k} \beta_j z^{j-1} + \sum_{j=k+1}^{k^*} \beta_j R_1(z, z_j) + \varrho(z),$$

where $\varrho \in \mathcal{H}_1 \ominus \{s : s = \sum_{j=k+1}^{k^*} \beta_j R_1(\cdot, z_j)\}$, the orthogonal complement of the latter space in $\mathcal{H}_1$, and R_1 is the reproducing kernel for the space $\{s : s = \sum_{j=k+1}^{k^*} \beta_j R_1(\cdot, z_j)\}$. Additionally, we denote the coefficient $\beta = (\beta_1 \ldots \beta_k)^T$, $\beta_+ = (\beta_{k+1} \ldots \beta_{k^*})^T$, and $\beta^* = (\beta, \beta_+)^T$. Thus, s can be rewritten as:

$$s(z) = B^* \beta^* + \varrho(z), \tag{2.65}$$

where the first k columns of B^* correspond to the design matrix induced by the polynomial covariates z^{j-1} for $j = 1, \ldots, k$, and the other columns correspond to the design matrix induced by the covariates $R_1(z, z_j)$ for $j = (k+1), \ldots, k^*$. Thus B^* can be decomposed into the block matrix $[B, B_+]$, where B and B_+ have k and $k_+ = k^* - k$ columns respectively. The dimension of the space $\mathcal{H}_0 \oplus \{s : s = \sum_{j=k+1}^{k^*} \beta_j R_1(\cdot, z_j)\}$, k^*, increases with the sample size, n, and $k^* = O(n^\nu), 0 < \nu \leq 1$. Moreover, $\langle z^{j-1}, \varrho \rangle = 0, j = 1, \ldots, k$ and $\langle R_1(\cdot, z_j), \varrho \rangle = 0, j = (k+1), \ldots, k^*$, so $B^{*T} \varrho = 0$, and similarly, $B^T B_+ = 0$.

Below, we show the effectiveness of the approximation of the asymptotic variance-covariance matrix by inverting the Fisher information in a simple Gaussian case with i.i.d errors of zero mean and constant variance σ^2. Because of the identity link and Gaussian family, $f = 1$, $l = 1/\sigma^2$ and $e_1(z) = E(\dot{h}_0(X)|Z = z)$. Thus the asymptotic variance matrix can be simplified as:

$$\Sigma = \sigma^2 E^{-1}\{[\dot{h}_0(x) - E(\dot{h}_0(X)|Z = z)][\dot{h}_0(x) - E(\dot{h}_0(X)|Z = z)]^T\} \tag{2.66}$$

$$= \sigma^2 E^{-1}[Cov(\dot{h}_0(X)|Z = z)].$$

For this simple Gaussian case, $\delta_i = (y_i - \mu_i)/\sigma^2$ and $G_\delta = (-1/\sigma^2)I$ in the Fisher information expression. Therefore, the observed Fisher information matrix equals:

$$\mathcal{I} = \frac{1}{\sigma^2} \begin{bmatrix} B^{*T} B^* + n\lambda\sigma^2 S & B^{*T} \dot{H}_\theta \\ \dot{H}_\theta^{\,T} B^* & \dot{H}_\theta^{\,T} \dot{H}_\theta - \sigma^2 \ddot{H}_{\delta\theta} \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix},$$

evaluated at $\beta^* = \hat{\beta}_n^*$ and $\theta = \hat{\theta}_n$.

To derive the covariance approximation of the estimators based on the Fisher information, we apply the following matrix identities for the Fisher information matrix:

$$
\begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix}^{-1}
$$

$$
= \begin{bmatrix} \mathcal{I}_{11}^{-1} - \mathcal{I}_{11}^{-1}\mathcal{I}_{12}(\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1}\mathcal{I}_{21}\mathcal{I}_{11}^{-1} & -\mathcal{I}_{11}^{-1}\mathcal{I}_{12}(\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1} \\ -(\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1}\mathcal{I}_{21}\mathcal{I}_{11}^{-1} & (\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1} \end{bmatrix},
$$

$$
(\mathcal{I}_1 + auv^T)^{-1} = \mathcal{I}_1^{-1} - a\mathcal{I}_1^{-1}uv^T\mathcal{I}_1^{-1}/(1 + av^T\mathcal{I}_1^{-1}u), \tag{2.67}
$$

where $\mathcal{I}_{11}$ and $\mathcal{I}_1$ are nonsingular matrices, and $\mathcal{I}_{22}$ is possibly singular matrix; a is a scaler, u and v^T are column and row vectors, respectively.

The approximate variance-covariance matrix of $\hat{\theta}_n$ based on the Fisher information is the lower right block of $\mathcal{I}^{-1}$, which can be expressed as

$$
\sigma^2(\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1}
$$

$$
= \sigma^2[(\dot{H}_{\hat{\theta}_n}^T \dot{H}_{\hat{\theta}_n} - \sigma^2 \ddot{H}_{\delta\hat{\theta}_n}) - \dot{H}_{\hat{\theta}_n}^T B^*(B^{*T}B^* + n\lambda\sigma^2 S)^{-1} B^{*T}\dot{H}_{\hat{\theta}_n}]^{-1}
$$

$$
= \sigma^2\{\dot{H}_{\hat{\theta}_n}^T \dot{H}_{\hat{\theta}_n} - \sigma^2 \ddot{H}_{\delta\hat{\theta}_n} - \dot{H}_{\hat{\theta}_n}^T B(B^T B)^{-1}B^T \dot{H}_{\hat{\theta}_n} -
$$

$$
\dot{H}_{\hat{\theta}_n}^T B_+ n^{-1}[n^{-1}(B_+^T B_+) + \lambda\sigma^2 S_+]^{-1} B_+^T \dot{H}_{\hat{\theta}_n}\}^{-1}. \tag{2.68}
$$

Note that $\dot{H}_{\hat{\theta}_n}^T B^*$ in the first equality of Eqn. (2.68) can be decomposed into the block matrix $[\dot{H}_{\hat{\theta}_n}^T B, \dot{H}_{\hat{\theta}_n}^T B_+]$. Additionally, because of the orthogonality between B and B_+, we find that

$$
B^{*T}B^* = \begin{bmatrix} B^T B & 0 \\ 0 & B_+^T B_+ \end{bmatrix}.
$$

Since $\beta^{*T}S\beta^* = \beta_+^T S_+ \beta_+$ denotes the positive penalty component in the penalized likelihood, and in practice, we use

$$
S = \begin{bmatrix} 0 & 0 \\ 0 & S_+ \end{bmatrix}.
$$

Thus, the last equality in Eqn. (2.68) follows from the above decompositions of

$$\dot{H}_{\hat{\theta}_n}^T B^* (B^{*T} B^* + n\lambda\sigma^2 S)^{-1} B^{*T} \dot{H}_{\hat{\theta}_n}.$$

We apply a matrix inequality and the asymptotic equivalence of sequences of matrices to show that certain terms in Eqn. (2.68) are negligible. For matrix comparison, write $A \geq G$ if $A = G + G_+$, where A and G are two positive definite matrices, and G_+ is a non-negative definite matrix. By dual conjunctive diagonalization, there exists a matrix P such that $P^T G P = I$ and $P^T G_+ P = D$, where I is an identity matrix, and $D = \text{Diag}(d_i)$ with non-negative d_i for $i = 1, \ldots$, the dimension of G_+. It is obvious that $P^T A P = I + D$, and then $A^{-1} = P(I + D)^{-1} P^T = P\text{Diag}[1/(1 + d_i)]P^T$. Thus, $G^{-1} - A^{-1} = P\text{Diag}[d_i/(1 + d_i)]P^T$, which is non-negative definite, hence $A^{-1} \leq G^{-1}$. We also present the technical details related to the asymptotic behavior of matrices. Let α_i be the eigenvalues of a matrix U and $\eta_i \geq 0$ be the eigenvalues of the symmetric nonnegative definite matrix $U^T U$. The strong norm of matrix U is defined by:

$$\|U\| = \max_{z:z^T z=1} [z^T U^T U z]^{1/2},$$

and

$$\|U\|^2 = \max_i \eta_i.$$

The weak norm of $U = [a_{i,j}]_{n \times n}$ is defined by

$$|U| = \left(\frac{1}{n} \sum_{i=0}^{n-1} \sum_{i=0}^{n-1} |a_{i,j}|^2\right)^{1/2}$$

$$= \left(\frac{1}{n} \sum_{i=0}^{n-1} \eta_i\right)^{1/2}.$$

Two sequences of $n \times n$ matrices $\{U_n\}$ and $\{V_n\}$ are asymptotically equivalent if U_n and V_n are uniformly bounded in strong norm and $\lim_{n \to \infty} |U_n - V_n| = 0$. The asymptotic equivalence between U_n and V_n is denoted as $U_n \sim V_n$ in this section. Asymptotic equivalence of $\{U_n\}$ and $\{V_n\}$ implies that for any vector b_n of appropriate dimension,

$$b_n^T U_n b_n = b_n^T V_n b_n + o(\sqrt{n}|b_n|^2), \tag{2.69}$$

by Cauchy-Schwartz inequality. The latter property is made use below to simplify

(2.68).

In this section, we use $[a_{ij}]_{k_1 \times k_2}$ to denote a $k_1 \times k_2$ matrix with the (i,j)th entry equal to a_{ij}. As defined previously, B_+ is a $n \times k_+$ matrix $[R_1(z_i, z_j)]_{n \times k_+}$. With the inner product $\langle R_1(z_i, \cdot), R_1(\cdot, z_j) \rangle_1 = \int_0^1 R_1^{(k)}(z_i, z) R_1^{(k)}(z, z_j) dz = R_1(z_i, z_j)$, the term $B_+^T B_+ / n$ equals a $k_+ \times k_+$ matrix $[R_1(z_i, z_j) + \epsilon_{ij}]_{k_+ \times k_+}$, where each $\epsilon_{ij} = o(1/n)$. Since the penalty $\int_0^1 (s^{(k)}(z))^2 dz$ for the penalized likelihood estimate can be represented by $\beta_+^T S_+ \beta_+ / 2$, the penalty matrix S_+ is a $k_+ \times k_+$ matrix $[R_1(z_i, z_j)]_{k_+ \times k_+}$. Let $\epsilon^* = max|\epsilon_{ij}|$, and $\epsilon^* \to 0$ as $n \to \infty$. The term $n^{-1}(B_+^T B_+) + \lambda \sigma^2 S_+$ in the last equality of Eqn. (2.68) satisfies the following inequality:

$$n^{-1}(B_+^T B_+) + \lambda \sigma^2 n^{-1}(B_+^T B_+) - \lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T \leq n^{-1}(B_+^T B_+) + \lambda \sigma^2 S_+ \leq$$

$$n^{-1}(B_+^T B_+) + \lambda \sigma^2 n^{-1}(B_+^T B_+) + \lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T, \quad (2.70)$$

where 1_{k_+} is a $k_+ \times 1$ vector of 1, so $1_{k_+} 1_{k_+}^T$ is a $k_+ \times k_+$ matrix with all the elements equal to 1. Inverting (2.70) yields the following inequality:

$$[n^{-1}(B_+^T B_+) + \lambda \sigma^2 n^{-1}(B_+^T B_+) - \lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T]^{-1} \geq [n^{-1}(B_+^T B_+) + \lambda \sigma^2 S_+]^{-1} \geq$$

$$[n^{-1}(B_+^T B_+) + \lambda \sigma^2 n^{-1}(B_+^T B_+) + \lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T]^{-1}.$$

Thus, the last quadratic form in Eqn. (2.68), $\dot{H}_{\hat{\theta}_n}^T B_+ n^{-1}[n^{-1}(B_+^T B_+) + \lambda \sigma^2 S_+]^{-1} B_+^T \dot{H}_{\hat{\theta}_n}$, has the upper and lower bounds as follows:

$$\dot{H}_{\hat{\theta}_n}^T B_+ n^{-1}[n^{-1}(B_+^T B_+) + \lambda \sigma^2 n^{-1}(B_+^T B_+) \mp \lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T]^{-1} B_+^T \dot{H}_{\hat{\theta}_n}$$

$$= \dot{H}_{\hat{\theta}_n}^T B_+ n^{-1} \{ \frac{n(B_+^T B_+)^{-1}}{1 + \lambda \sigma^2} \pm$$

$$[\frac{n(B_+^T B_+)^{-1}}{1 + \lambda \sigma^2} \lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T \frac{n(B_+^T B_+)^{-1}}{1 + \lambda \sigma^2}] / [1 \mp 1_{k_+}^T \frac{\lambda \sigma^2 \epsilon^* n(B_+^T B_+)^{-1}}{1 + \lambda \sigma^2} 1_{k_+}] \} B_+^T \dot{H}_{\hat{\theta}_n}$$

$$= \dot{H}_{\hat{\theta}_n}^T B_+ \frac{(B_+^T B_+)^{-1}}{1 + \lambda \sigma^2} \{ n^{-1}(B_+^T B_+) + \lambda \sigma^2 n^{-1}(B_+^T B_+) \pm$$

$$(\lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T) / [1 \mp 1_{k_+}^T \frac{\lambda \sigma^2 \epsilon^* n(B_+^T B_+)^{-1}}{1 + \lambda \sigma^2} 1_{k_+}] \} \frac{n(B_+^T B_+)^{-1}}{1 + \lambda \sigma^2} B_+^T \dot{H}_{\hat{\theta}_n}. \quad (2.71)$$

The first equality in Eqn. (2.71) is derived by applying the inverting rule in Eqn. (2.67) to $[n^{-1}(B_+^T B_+) + \lambda \sigma^2 n^{-1}(B_+^T B_+) \mp \lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T]^{-1}$. According to assumption (A2), $\lambda = o_p(n^{-1/2})$, $\lambda \sigma^2$ is negligible relative to the constant 1, so the term $\lambda \sigma^2 n^{-1}(B_+^T B_+)$

in the last equality of Eqn. (2.71) is negligible relative to $n^{-1}(B_+^T B_+)$. The argument can be made more rigorous as follows. Recall that $n^{-1}(B_+^T B_+) = [R_1(z_i, z_j)]_{k_+ \times k_+} + [\epsilon_{ij}]_{k_+ \times k_+}$. We make use of the asymptotic equivalence between $[\epsilon_{ij}]_{k_+ \times k_+}$ and $\epsilon^* 1_{k_+} 1_{k_+}^T$ to drop $(\lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T)/[1 \mp 1_{k_+}^T \frac{\lambda \sigma^2 \epsilon^* n (B_+^T B_+)^{-1}}{1 + \lambda \sigma^2} 1_{k_+}]$ from the last equality of Eqn. (2.71). Since $\epsilon_{ij} = o(n^{-1})$ and $k_+ < n$, the strong norm $\|[\epsilon_{ij}]_{k_+ \times k_+}\| < \infty$. The only eigenvalue of $\epsilon^* 1_{k_+} 1_{k_+}^T \to 0$ as $n \to \infty$, so the strong norm $\|\epsilon^* 1_{k_+} 1_{k_+}^T\| < \infty$. Also, the weak norm of the difference satisfies $\lim_{n \to \infty} |[\epsilon_{ij} - \epsilon^*]_{k_+ \times k_+}| \to 0$. Then $[\epsilon_{ij}]_{k_+ \times k_+} \sim \epsilon^* 1_{k_+} 1_{k_+}^T$. Using the properties of the asymptotic equivalence, as well as the extra assumption that $1_{k_+}^T \frac{\lambda \sigma^2 \epsilon^* n (B_+^T B_+)^{-1}}{1 + \lambda \sigma^2} 1_{k_+}$ is bounded away from 1 as $n \to \infty$, $(\lambda \sigma^2 \epsilon^* 1_{k_+} 1_{k_+}^T)/[1 \mp 1_{k_+}^T \frac{\lambda \sigma^2 \epsilon^* n (B_+^T B_+)^{-1}}{1 + \lambda \sigma^2} 1_{k_+}]$ is negligible relative to $n^{-1}(B_+^T B_+)$, because $\lambda = o(n^{-1/2})$ and in view of (2.69). Therefore, the quadratic form $\dot{H}_{\hat{\theta}_n}^T B_+ n^{-1} [n^{-1}(B_+^T B_+) + \lambda \sigma^2 S_+]^{-1} B_+^T \dot{H}_{\hat{\theta}_n}$ can be simplified as $\dot{H}_{\hat{\theta}_n}^T B_+ (B_+^T B_+)^{-1} B_+^T \dot{H}_{\hat{\theta}_n}$ by dropping the negligible terms. Furthermore, by the law of large number, the term $\sigma^2 \ddot{H}_{\delta \hat{\theta}_n}$ in Eqn. (2.68) is also negligible with

$$E[\delta | (x, z)]|_{\beta_0^*, \theta_0} = E[(y - \mu)/\sigma^2 | (x, z)]|_{\beta_0^*, \theta_0} = 0.$$

After dropping the negligible terms in Eqn. (2.68), the covariance approximation of $\hat{\theta}_n$ can be simplified as:

$$\sigma^2 (\mathcal{I}_{22} - \mathcal{I}_{21} \mathcal{I}_{11}^{-1} \mathcal{I}_{12})^{-1} = \sigma^2 \{\dot{H}_{\hat{\theta}_n}^T \dot{H}_{\hat{\theta}_n} - \dot{H}_{\hat{\theta}_n}^T B^* (B^{*T} B^*)^{-1} B^{*T} \dot{H}_{\hat{\theta}_n}\}^{-1}. \tag{2.72}$$

Next, we consider the conditional expectation of $\dot{H}_{\hat{\theta}_n}$ given z. The conditional expectation $E[(\partial h_{\hat{\theta}_n}(X))/(\partial \theta)|Z = z]$ consists of $E[(\partial h_{\hat{\theta}_n}(X))/(\partial \theta_q)|Z = z]$ as its columns. We assume that $E[(\partial h_{\hat{\theta}_n}(X))/(\partial \theta_q)|Z = z]$ belongs to the Sobolev class, so this conditional expectation can be written as $B^* \beta_{h_q}^* + \varrho_{h_q}$ in the form of (2.65). Also, $B^{*T} \varrho_{h_q} = 0$. Recalling that $\dot{H}_{\hat{\theta}_n}$ comprises the first derivatives of $h_{\hat{\theta}_n}$, write

$(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q), q = 1, \ldots, r$ as follows:

$$(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q) = E[(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q)|Z = z]$$

$$+ \{(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q) - E[(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q)|Z = z]\}$$

$$= (B^*\beta^*_{h_q} + \varrho_{h_q}) + \{(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q) - E[(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q)|Z = z]\}.$$

The element $\{(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q) - E[(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q)|Z = z]\}$ is orthogonal to the space spanned by the columns of B^*. Since $B^{*T}\varrho_{h_q} = 0$,

$$B^*(B^{*T}B^*)^{-1}B^{*T}(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q) = B^*\beta^*_{h_q} = E[(\partial h_{\hat{\theta}_n}(X))/(\partial\theta_q)|Z = z] - \varrho_{h_q},$$

and Eqn.(2.72) can be expressed as

$$\sigma^2(\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1}$$

$$= \sigma^2[\dot{H}^T_{\hat{\theta}_n}\dot{H}_{\hat{\theta}_n} - \{E[(\partial h_{\hat{\theta}_n}(X))/(\partial\theta)|Z = z] - \varrho_h\}^T\{E[(\partial h_{\hat{\theta}_n}(X))/(\partial\theta)|Z = z] - \varrho_h\}]^{-1}$$

$$= \sigma^2\{\dot{H}^T_{\hat{\theta}_n}\dot{H}_{\hat{\theta}_n} - E[(\partial h_{\hat{\theta}_n}(X))/(\partial\theta)|Z = z]^T_{\hat{\theta}_n} E[(\partial h_{\hat{\theta}_n}(X))/(\partial\theta)|Z = z]_{\hat{\theta}_n} + \varrho_h^T\varrho_h\}^{-1},$$

where the matrix ϱ_h comprises columns ϱ_{h_q}'s.

Since the dimension of $\mathcal{H}_0 \oplus \{s : s = \sum_{j=k+1}^{k^*} \beta_j R_1(\cdot, z_j)\}$, k^*, increases with n, the last term in the previous equation, $\varrho_h^T\varrho_h$, is negligible for large samples. Using the consistency of $\hat{\theta}_n$ derived in Theorem 2.3.1 and the uniform law of large number, we find that the covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ based on the Fisher information converges to

$$\sigma^2 E^{-1}\{[\dot{h}_0(x) - E(\dot{h}_0(X)|z)][\dot{h}_0(x) - E(\dot{h}_0(X)|z)]^T\}, \qquad (2.73)$$

as $n \to \infty$, which completes the justification of computing the limiting covariance matrix of $\hat{\theta}$ via inverting the Fisher information.

2.8   Justification of the Laplace Approximation of the Marginal Likelihood

As discussed in Section 2.2.3, the parameter in the PPGAM, ζ^T, can be further decomposed into $(\beta^T, \beta_+^T, \theta^T)$, where β and β_+ are the parameters corresponding to the bases of the $s(z)$ with zero penalty and positive penalty, respectively, and θ is

the parameter in the parametric nonlinear component. Using the prior

$$P(\zeta|M_j) = \frac{|n\lambda S_+|^{1/2}}{(2\pi)^{k_+/2}} \exp(-\frac{1}{2}n\lambda\beta_+^T S_+\beta_+)P(\beta,\theta|M_j),$$

the marginal likelihood can be written as

$$P(D|M_j) = \frac{|n\lambda S_+|^{1/2}}{(2\pi)^{k_+/2}} \int \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta.$$

We apply Laplace approximation to rewrite the integral in (2.11) for an explicit marginal likelihood expression, and the approximation approach needs a few steps for justification.

We assume the well-separated condition for the maximum of $\exp\{\bar{l}_p(\zeta)\}$, i.e. with probability approaching 1 as $n \to \infty$, $\forall\ \epsilon > 0$, $\exists 0 < \varrho < \exp\{\bar{l}_p(\zeta)\}$, such that $\hat{\Gamma} = \{\zeta; \exp\{\bar{l}_p(\zeta)\} > \varrho\} \subseteq D_\epsilon = \{\zeta; \|\zeta - \hat{\zeta}\| \leq \epsilon\}$. Define $\exp_{\hat{\Gamma}}\{\bar{l}_p(\zeta)\}$ as a function agreeing with $\exp\{\bar{l}_p(\zeta)\}$ on the subset $\hat{\Gamma}$, and zero's otherwise. By the well-separated maximum condition,

$$\exp\{n\bar{l}_p(\zeta)\} - \exp_{\hat{\Gamma}}\{n\bar{l}_p(\zeta)\} \leq \varrho^n.$$

So, taking the integral

$$\int_{\hat{\Gamma}} \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta$$

$$\leq \int \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta \leq \int_{\hat{\Gamma}} \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta + \varrho^n$$

$$\int \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta$$

$$\leq \int_{\hat{\Gamma}} \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta[1 + \varrho^n/\int_{\hat{\Gamma}} \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta] \qquad (2.74)$$

Since

$$[\int_{\hat{\Gamma}} \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta]^{1/n} \geq E_{\hat{\Gamma}}^{1/n}[\exp\{n\bar{l}_p(\zeta)\}] \to \sup_{\hat{\Gamma}} \exp\{\bar{l}_p(\zeta)\}$$

$$= \sup \exp\{\bar{l}_p(\zeta)\}$$

$$> \varrho,$$

we get

$$[1 + \varrho^n/\int_{\hat{\Gamma}} \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta] \to 1. \qquad (2.75)$$

Using (2.74) and (2.75), we obtain

$$\log \int \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta = \log \int_{\hat{\Gamma}} \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta + o_p(1)$$

Applying Taylor expansion of $\bar{l}_p(\zeta)$ over D_ϵ around the penalized likelihood estimator $\hat{\zeta}$, $\bar{l}_p(\zeta) = \bar{l}_p(\hat{\zeta}) - 1/2(\zeta - \hat{\zeta})^T H_{\bar{l}_p}(\tilde{\zeta})(\zeta - \hat{\zeta})$, where $H_{\bar{l}_p}$ is the negative Hessian matrix of the normalized penalized likelihood, $\bar{l}_p$; and $\tilde{\zeta}$ is between ζ and $\hat{\zeta}$, so $\|\tilde{\zeta} - \hat{\zeta}\| \le \epsilon$. Since ϵ is arbitrary, so we can find an $\tilde{\zeta}$ to satisfy $\log H_{\bar{l}_p}(\tilde{\zeta}) = \log H_{\bar{l}_p}(\hat{\zeta}) + o_p(1)$. Using the similar prior assumption for $P(\beta,\theta|M_j)$ as Schwarz (1978), we find that

$$\log \int_{\hat{\Gamma}} \exp\{n\bar{l}_p(\zeta)\}P(\beta,\theta|M_j)d\zeta$$
$$= n\bar{l}_p(\hat{\zeta}) - \frac{1}{2}\log|H_{\bar{l}_p}(\hat{\zeta})| + \frac{k^* + r}{2}[\log(2\pi) - \log(n)] + O_p(1).$$

Thus,

$$\log P(D|M_j) = l_p(\hat{\zeta}) + \frac{k^* + r - k_+}{2}\log(2\pi) - \frac{k^* + r}{2}\log(n) - \frac{1}{2}\log\{|H_{\bar{l}_p}(\hat{\zeta})|\}$$
$$+ \frac{1}{2}\log|n\lambda S_+| + O_p(1)$$
$$= l_p(\hat{\zeta}) + \frac{k^* + r - k_+}{2}\log(2\pi) - \frac{1}{2}\log\{|H(\hat{\zeta})|\} + \frac{1}{2}\log|n\lambda S_+| + O_p(1),$$

where H is the negative Hessian matrix of the penalized likelihood, l_p.

CHAPTER 3
HATCHDATE ANALYSIS OF POLLOCK LARVAE

3.1   Background

Marine fish populations are generally subject to severe annual fluctuations, due to strong environmental effects on spawning and survival processes (Bailey and Macklin, 1994). Marine environmental conditions can be highly variable, with occasional drastic changes brought about by, for example, (climate) regime shifts, as well as natural and/or man-made interventions such as algae blooms or oil spills (Chan, 2003). Regime shifts and interventions, however, provide opportunities for assessing the resilience of fish populations to severe perturbations (Chan, 2003). Gaining quantitative understanding of environmental effects or intervention effects on larval fish survival is, moreover, important in stock forecasting with regards to harvest versus environment effects. Key elements to understanding these effects can be found in historical records of patterns in birth(hatch)date distributions of fish larvae collected in fisheries monitoring surveys (Bailey and Macklin, 1994), but lack of suitable analytical methodologies has left this cache of information largely unexplored. The shape and height of the annual hatchdate frequency distribution may be dynamically altered by events influencing the processes of spawning, hatching and larval survival, which form the basis for inference on how environmental conditions shape the early dynamics of larval fish. Here, we demonstrate a general statistical framework for untangling various environmental effects and intervention effects on spawning and larval survival using hatchdate data and the PPGAM.

Walleye pollock (Theragra chalcogramma) is a gadid species that is widely distributed across the North Pacific Ocean; it is currently the world's second largest fishery and a key species of the Gulf of Alaska (GOA) and other ecosystems, both as

prey and predator (Ciannelli et al., 2005). The GOA pollock population experienced a rapid increase in the 1970s and early 1980s, but recently has been near collapse, having reached about 22% of its estimated unfished biomass (Dorn et al., 2008a). In the western GOA pollock spawning is localized at specific spawning grounds during restricted periods. The majority of spawning occurs in a deep sea valley, Shelikof Strait, during the first two weeks in April. Eggs take about 2 weeks to hatch and larvae drift in the Alaska Coastal Current, where they may be retained by eddies, transported onto the continental shelf, or swept offshore into the swift-flowing Alaskan Stream. Shelikof Strait is at the center of storm activity in the GOA, and considerable year-to-year variability of environmental conditions is experienced there. Larvae have been sampled in surveys commencing in late May since 1981. Based on counts of daily growth increments on their otoliths (ear stones), the ages of pollock larvae collected in late May have been determined and are used here to estimate the annual hatchdate frequency of the population.

The proposed methodology applied to 19 years of annual hatchdate abundance data enabled us to reveal rich hidden information in these hatchdate data about the environmental effects on larval dynamics (Fig. 3.1), similar to dendrochronolgy used in analyzing the role of events and climate in tree and forest dynamics (Abrams et al., 2000). Some intertwined intervention effects may have influenced pollock in the GOA in 1989/90, as acoustic survey estimates of the Shelikof spawning population differed markedly from those of age structured stock assessment analysis models in those years (Dorn et al., 2008a) and both the Exxon Valdez oil spill and a shift in the Pacific Decadal Oscillation (PDO) were observed, whose intervention effects are assessed below.

Figure 3.1: Log abundance of caught pollock larvae hatched over the hatch dates. Logarithm of pollock abundances hatched over various hatchdates (in Julian dates), with the fitted log abundance curves derived from the fitted model defined by (3.1) (black curves) superimposed on the diagrams. Outliers are marked as x's. Also imposed on the diagrams are the fitted values with the SSTB effects omitted (SST effects on spawning and hatching; red curves) and the fitted values with SSTA effects omitted (SST effects on larval survival; blue curves). The yearly average SSTB and SSTA are pictorially represented by two thermometers where the red bar is proportional to the average annual SSTB whereas the blue bar is proportional to the yearly average SSTA.
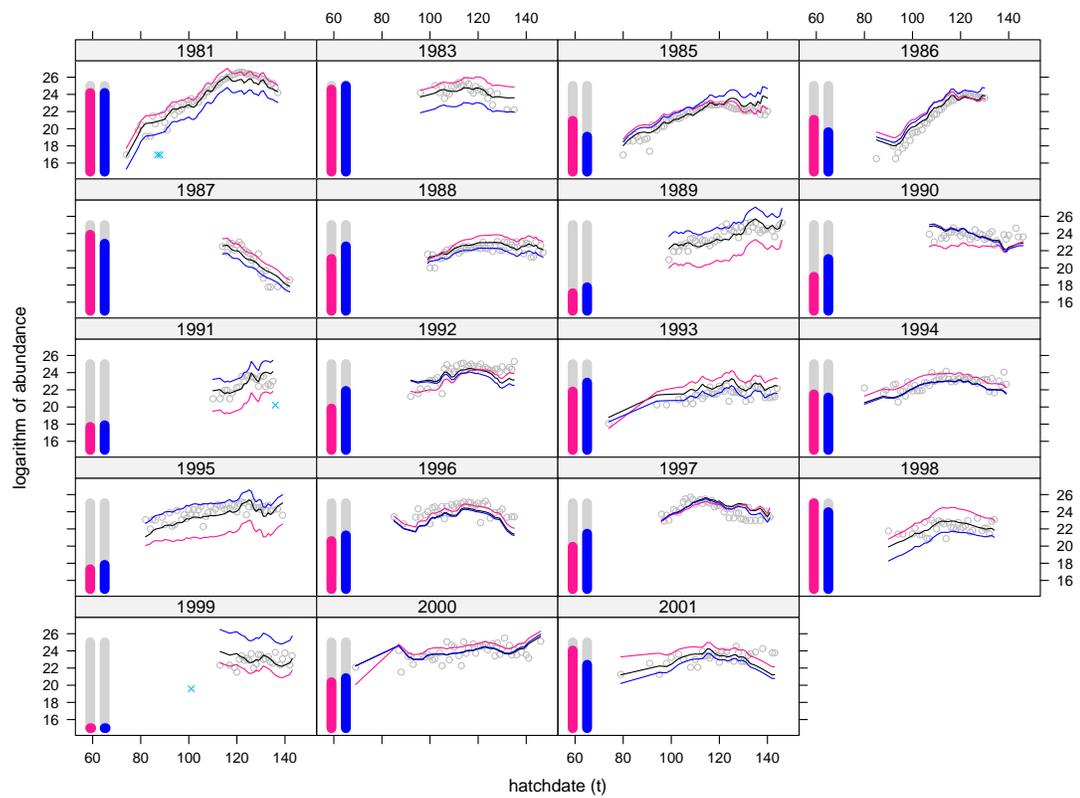
3.2 Data

Larval pollock abundance and hatchdate data were obtained from annual surveys of Shelikof Strait in late May-early June from 1981 to 2001. Generally sampling began near Unimak Pass and continued along the southern side of the Alaska Peninsula to northeast of Kodiak Island. Sampling was done with bongo frames equipped with a 500 m mesh plankton netting from 100 m to surface. Otoliths were dissected out and daily increments counted under a microscope following protocols described in Yoklavich and Bailey (1990). Age frequencies were determined and used to estimate population hatchdate distribution (hatchdate = survey date-age) (see Hinckley et al. (1993)). Age frequencies from each sample were standardized to the mean survey date (Bailey et al., 1996). The log abundances of pollock larvae were plotted against their hatchdates year by year in Fig. 3.1, with the raw data denoted by open circles. Several covarites were derived from 4 environmental variables, namely, drift out of GOA (Transport), (anomalous) sea surface temperature (SST), (anomalous) surface wind speed in Shelikof Strait (WindS) and the PDO. For untangling the environmental effects on spawning/hatching and larval survival, we generally computed 2 auxiliary variables from each of Transport, SST and WindS by averaging each over (i) thirty days before the hatchdate, and (ii) the period between the hatchdate and the catchdate. For assessing how large-scale climatic pattern may alter the shape and/or height of the pollock's hatchdate frequency distribution, we also computed the mean anomolous PDO levels from January to March, the three-month period before the general pollock spawning. The biological factors consisted of the pollock spawning biomass and its age structure.

3.3   The Model

The baseline hatchdate distribution (as determined by the underlying spawn-ing distribution) may be specified as some discrete probability distribution denoted by $\exp(s(t))$ which equals the probability that a larva is hatched on day t. As the functional form of $s(t)$ is unknown, it is specified as a nonparametric, smooth func-tion in t. Assuming a log-logistic distribution for the life-time distribution of a larva, the baseline probability that a larva survives k or more days equals $1/[1 + (\lambda k)^\gamma]$, where $\gamma > 0$ is the shape parameter of the distribution and the reciprocal of $\lambda > 0$ is the scale parameter, specifically the median life expectancy. An important property of the log-logistic distribution is its ability to capture non-monotone hazard rate. Indeed, for the log-logistic distribution with $\gamma > 1$, its hazard function, whose value at $k \geq 0$ equals the probability of instantaneous death given survival for at least k days, is unimodal. On the other hand, for $\gamma \leq 1$, the hazard function is mono-tone decreasing. Let $\exp(\beta_0)$ be the catch probability, assumed constant over the study period. These baseline probabilities describe the situation in a typical year in which the probability that a larva, spawned on day t survived for more than k days and caught on day $t + k$ equals $\exp(s(t)) \exp(\beta_0)/[1 + (\lambda k)^\gamma]$. However, both the hatching and survival processes are influenced by environmental conditions, so modifications are needed. First, young and older spawners may have different egg productivity and spawning time distributions resulting in a mixture of hatchdate distributions. Let $\alpha_{y,a}$ be the proportion of group-a in year y (only 2 groups are considered with the group-1 consisting of young (4 to 7 years old) spawners and the group-2 the 8 years or older fish). We assume that, on the log-scale, the baseline probability that a larva is hatched on day t of year y is (approximately) propor-tional to $\exp(\alpha_{y,1}s_1(t) + \alpha_{y,2}s_2(t))$ where the hatchdate distribution of eggs from the group-a equals $s_a(t), a = 1, 2$ on the log-scale, up to some additive constant.

Second, environmental effects on spawning and hatching can alter the height uniformly and/or the shape of the hatching distribution on the log-scale so that the probability of hatching on day t of year y is proportional to:

$$exp[\alpha_{y,1}s_1(t) + \alpha_{y,2}s_2(t) + \sum_{j=1}^{p_1} B_{y,j}s_{Bj}(t) + \sum_{j=p_1+1}^{p} s(B_{t,y,j})],$$

where $B_{y,1}, \ldots, B_{y,p_1}$ are p_1 yearly environmental factors that alter the shape (and height) of the hatching distribution through $B_{y,j}s_{Bj}(t), j = 1, \ldots, p_1$, and $B_{t,y,p_1+1}, \ldots, B_{t,y,p}$ are $p - p_1$ environmental covariates over a period of thirty days prior to day t of year y that alter the height of the hatching distribution. The function $s_{Bj}(t)$ can be interpreted as the (additive) change in the (log) hatchdate frequency distribution per unit increase in the covariate B_j. The functions s, with distinct arguments, are generally distinct smooth functions. Because the functional forms of the environmental effects are generally unknown, they are specified as smooth functions. Similarly, we nonparametrically model the environmental effects on survival multiplicatively so that the (conditional) probability that a larva, hatched on day t of year y lived for k or more days equals

$$exp[\sum_{l=1}^{q} s(A_{t,y,l})]/[1 + (\lambda k)^{\gamma}],$$

where $A_{t,y,1}, \ldots, A_{t,y,q}$ are q environmental variables measured over the period between day t when the larva was hatched and day $t + k$ of year y when it was caught. The effects of an intervention in year 1989 on the survival of larvae can be modeled by keeping the shape parameter of the log-logistic distribution fixed but changing the parameter λ. The parameter λ in the survival function for 1989 may differ from the baseline by the multiplicative constants $(1 + \psi_{1989})$. For $-1 < \psi_{1989} \leq 0$, the hazard rates in 1989 are lower than the baseline, whereas ψ_{1989} corresponds to elevated hazard rates in 1989. For $\psi_{1989} = 0$, there is no intervention effects. We postulate that the intervention effect after 1989 may diminish geometrically so that the λ parameter in the mth year after 1989 is modified by the multiplicative

coefficient $(1 + \phi^m \psi_{1989})$. Note that for $0 < \phi < 1$, the intervention effects decay geometrically, whereas the case $\phi = 0$ implies no after-effects and the case $\phi = 1$ signifies permanent, constant intervention effects. Altogether, based on the preceding consideration, we formulate the following model:

$$l_{t,y,k} = n_y + \alpha_{y,1} s_1(t) + \alpha_{y,2} s_2(t) + \sum_{j=1}^{p_1} B_{y,j} s_{Bj}(t) + \sum_{j=p_1+1}^{p} s(B_{t,y,j})$$

$$- \log(1 + [(1 + \phi^{y-1989} \psi_{1989} 1_{(y \geq 1989)}) \lambda k]^{\gamma}) + \sum_{l=1}^{q} s(A_{t,y,l}) + \epsilon_{y,t,k}, \qquad (3.1)$$

where $l_{t,y,k}$ is the log abundance of the captured larval pollock that were hatched in day t of year y, and caught on day $t + k$, and n_y is the log spawning biomass in year y; $1_{(\cdot)}$ is the dummy variable for the expression within the parentheses; $\epsilon_{y,t,k}$ are uncorrelated, normally distributed errors of zero mean and identical variance. Note that the catching probability parameter β_0 and the age-specific productivity factors are absorbed into the hatchdate functions s_a. All other smooth functions without subscripts are centered to have zero mean over the data.

The model defined by Eqn. (3.1) and its variants are instances of the PPGAM can be fitted by the method of penalized maximum likelihood. Based on marginal likelihood, the following model is found to be most consistent with the data:

$$l_{y,t,k} = n_y + \alpha_{y,1} s_1(t) + \alpha_{y,2} s_2(t) + PDOB_y s_{PDOB}(t) + s(SSTB_{t,y}) + s(WindSB_{t,y})$$

$$- \log(1 + [(1 + \phi^{y-1989} \psi_{1989} 1_{(y \geq 1989)}) \lambda k]^{\gamma}) + s(TransA_{t,k,y}) + s(SSTA_{t,k,y})$$

$$+ s(WindSA_{t,k,y}) + \epsilon_{y,t,k}, \qquad (3.2)$$

where $PDOB_y$ equals the mean anomalous PDO level from January to March of year y; $SSTB$ and $SSTA$ are the 30 day average sea surface temperature anomalies before the hatchdate and the average sea surface temperature anomalies from the hatchdate to the catchdate respectively. $WindSB$, $WindSA$ and $TransA$ are defined similarly. Generally, the last letter B (A) in a variable name stands for before (after) hatching. Model diagnostics suggest that the fitted model defined by Eqn. (3.2) provides a

good fit to the data after removing 4 outliers, marked as "x" in Fig. 3.1.

3.4 Model Interpretation and Model Diagnostics

Fig. 3.2 plots the additive effects of the environmental factors on the spawning, hatching and survival of pollock larvae, whose interpretations are given below:

- Fig. 3.2a shows the baseline hatching distribution (on the log-scale) of eggs from the older spawners and Fig. 3.2b that of the younger spawners. The mean (log) hatching abundance of the old spawners (20.4) is significantly higher than that of the young spawners (19.1) with p-value being equal to 0.001. The hatchdate distribution for the older group has a peak around (Julian) day 82 and a secondary peak near day 112 suggesting that spawning from the older group peaks around days 68 and 98. On the other hand, the hatchdate distribution for the younger group is rather flat, suggesting a more variable spawning time distribution of the young spawners than their older counterpart.

- Fig. 3.2c shows the change in the (log) hatchdate frequency distribution per unit increase in PDO. The curve in Fig. 3.2c is entirely below zero, so higher PDO tends to be associated with lower spawning and hatching. The unimode shape and the drop-down on the left tail of the curve in Fig. 3.2c indicate that higher PDO is associated with more intensive spawning during the middle of the spawning period.

- Fig. 3.2d portrays the nonlinear "transport effect" on survival. The positive "transport effect" on larval abundance as the drift out of GOA (transport) increasing from a low to a moderate level may be explained by some beneficial effects of moderate transport related to nutrient input and enhanced prey production. However, when the transport out of GOA is strong, it flushes

pollock larvae out of the area and reduces larval abundance.

- Fig. 3.2e shows that higher sea surface temperature before hatching is associated with lower larval abundance, which can be explained as follows: during the spawning months, SST is strongly correlated with bottom temperature due to a well-mixed water column. In relatively warm temperatures spawning occurs earlier and the turnover of eggs is faster and hence larvae have a longer duration period in the water before capture; they therefore experience a higher cumulative mortality (although not necessarily a higher instantaneous mortality) prior to capture. The opposite may hold in colder environments as eggs would have a longer duration time and spawning occurs later than an average year, and therefore larvae would have a shorter duration before capture.

- Fig. 3.2f indicates that higher sea surface temperature after spawning results in higher survival rates for larval pollock. This effect is probably due to lower instantaneous mortality rates of larvae in warmer conditions (Bailey, 2000).

- Based on the data intensive parts of Figs. 3.2g and 3.2h, it can be inferred that strong sea surface winds depress pollock's spawning and reduce larval survival in the GOA. The data on the right tail of the curve in Fig. 3.2h correspond mainly to the late hatchdates; the unexpected positive relationship between the wind speed and survival during the late hatching period may be caused by the confounding effects of water stratification late in the season. The effect of winds on spawning is difficult to explain, but may be related to linkages of winter storminess, currents and termperature and their effects on spawning location and timing.

Fig. 3.1 provides an alternative approach for visualizing the annual SST anomaly effects on the hatching and larval survival processes, by comparing the fitted values

from the final model with those having the SSTB (SSTA) effects suppressed. Before 1989, SST tended to depress the pollock's spawning and hatching. However, SST reversed the trend and instead greatly enhanced the pollock's spawning and hatching, for 3 years starting from 1989, while 1989 saw a sharp drop in SST, and SST remained low until 1991; these climatic changes are consistent with the climate regime shift discussed by (Ciannelli et al., 2007) and (Hare and Mantua, 2000).

Table 3.1 summarizes the parameter estimates of the final model. The baseline survival probability for the larval pollock to survive k or more days is estimated to equal $1/[1 + (0.08k)^{2.18}]$ (Fig. 3.3). The 1989 intervention depressed the survival probability function in 1989 to $1/[1 + (0.23k)^{2.18}]$. The estimates of ψ_{1989} and ϕ are all significantly different from 0 (2-sided p-values < 0.05), suggesting significant, depressing changes in the survival rate due to the intervention in 1989. Based on the estimated $\phi(0.47)$, the intervention left a weak memory on the pollock populations, with the 1989 intervention effects having a half-life of about 11 months (Cryer and Chan (2008); Box and Tiao (1975)). More specifically, by 1990, the change in hazard rate is about 47% of the change in 1989 while by 1992, the change is only 10% of the change in 1989; see Fig. 3.3. Such a quickly decaying rate of the 1989 intervention effects indicates a relatively quick recovery from the 1989 intervention.

For the fit of the model defined by Eqn. (3.2) using the full dataset, there are 4 outliers, with standardized residuals smaller than -4. Examination of the data indicates that the outliers correspond to the data on (Julian) day 101 in 1999, on day 136 in 1991, on days 87 and 88 in 1981. The log abundances of pollock larvae on these 4 specific days are marked as "x" in Fig. 3.1. Day 101 in 1999 is the earliest hatching date, and there is no pollock abundance information in the following 11 days. Moreover, the larval abundance estimate on day 101 in 1999 is very low, suggesting that the abundance estimate was based on relatively few age samples.

Day 136 is the last hatching date in 1991 and the larval abundance estimate on this day is much lower than the other late-hatching larval abundance in the same year. Upon checking the data in 1981, the larval abundances on days 87 and 88 are found to be exactly the same as that on day 74, which is the lowest larval abundance in that year. Consequently, we decide to drop these 4 outlying data cases from subsequent analyses.

Next, we check whether the fitted model provides a good fit to the data after dropping the 4 outliers. The upper left quantile-quantile normal score plot and the bottom left histogram plot in Fig. 3.4 suggest that the errors are approximately normal. The plot of residuals against fitted values at the upper right diagram in Fig. 3.4 indicates that the variance of the errors is approximately constant. The bottom right plot shows that the responses have a positive linear relationship with the fitted values. Altogether these model diagnostics suggest that the final model provides good fit to the data

Table 3.1: Parameter estimation on pollock larvae's survival

Parameter	Estiamte	Std. Error	z value	$Pr(> z)$
λ	0.08	0.019	4.21	< 0.001
γ	2.18	0.351	6.21	< 0.001
ψ_{1989}	1.89	0.679	2.78	0.005
ϕ	0.47	0.061	7.70	< 0.001

3.5 Model Selection

Above, we consider the situation that the intervention effect began in 1989 and it might decline geometrically afterwards. Another senario is that the intervention

Figure 3.2: Smooth function estimates of the covariate effects on hatching and survival of larval pollock. Solid lines portray the estimated additive covariate effects. Dashed lines encompass the 95% confidence bands. Dots in the diagrams are the partial residuals.
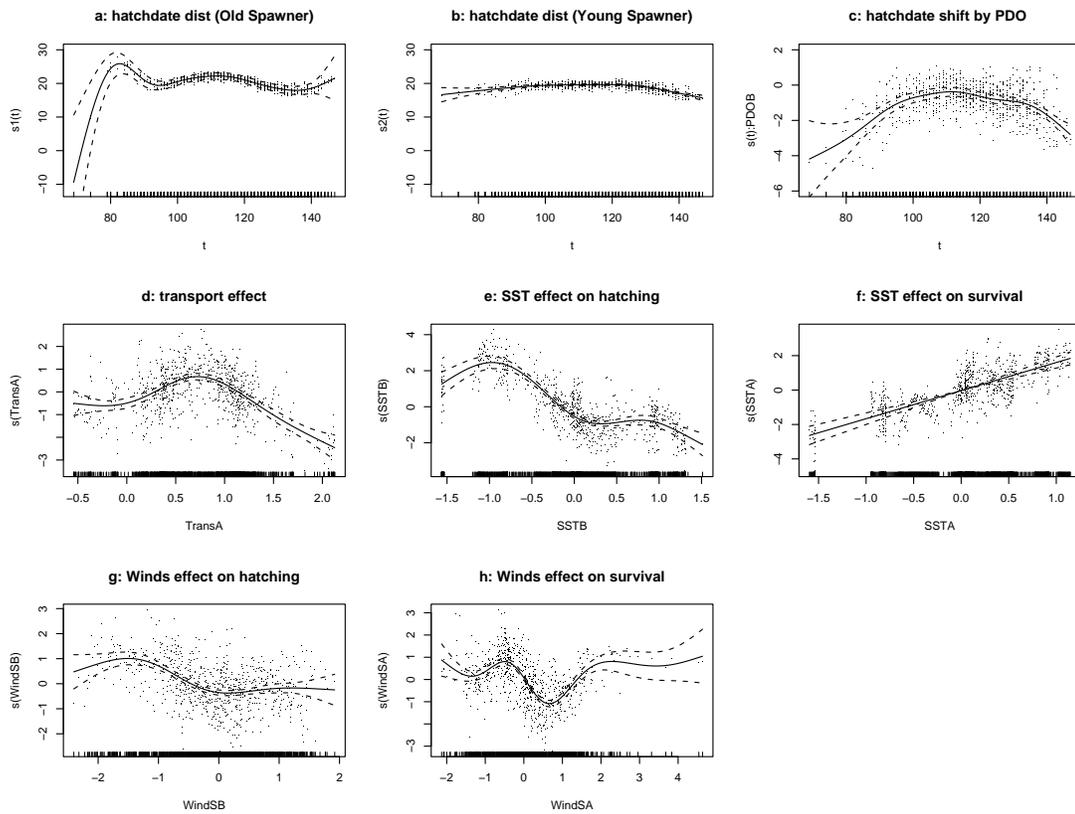
Figure 3.3: Estimates of the survial probability curves of larval pollock (age in days).

larval survival
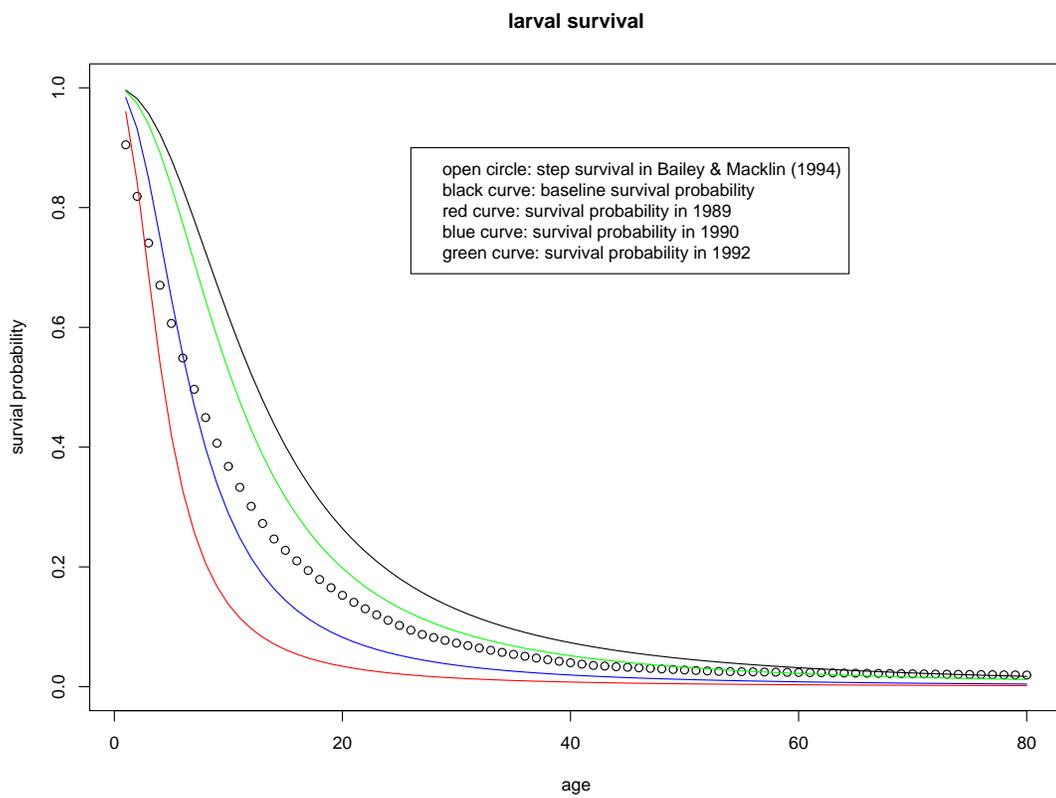
Figure 3.4: Residual diagnostic checks of Model (3.2)
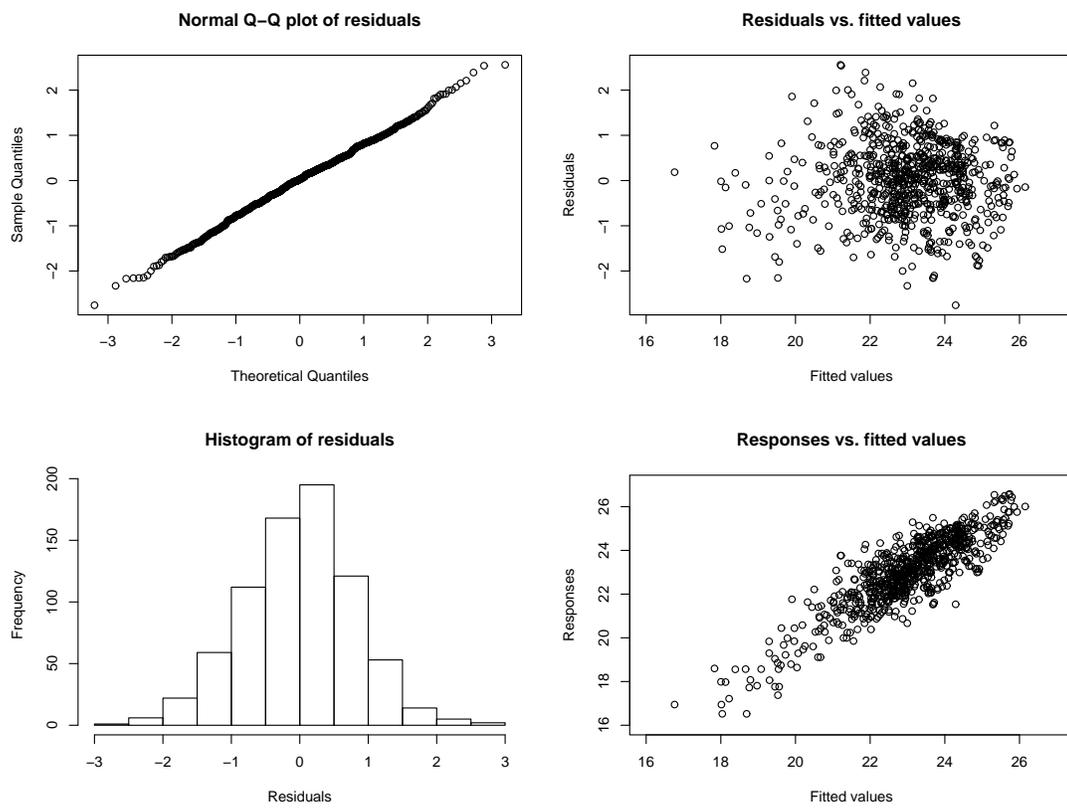
was only effective in 1989 and left no memory afterwards, i.e.

$$l_{t,y,k} = n_y + \alpha_{y,1}s_1(t) + \alpha_{y,2}s_2(t) + \sum_{j=1}^{p_1} B_{y,j}s_{Bj}(t) + \sum_{j=p_1+1}^{p} s(B_{t,y,j})$$

$$- \log(1 + [(1 + \psi_{1989}1_{(y=1989)})\lambda k]^{\gamma}) + \sum_{l=1}^{q} s(A_{t,y,l}) + \epsilon_{y,t,k}. \qquad (3.3)$$

These scenarios with the intervention effects modeled by Eqn. (3.1) and Eqn. (3.3) were compared to the null case that there is no intervention effects, i.e. ψ_{1989} in the model defined by Eqn. (3.3) equals zero. For selecting among the various models, we employ the (Laplace approximation of) log marginal likelihood of a model as a model selection criterion. Models with higher log marginal likelihoods are hence preferred. We used the log marginal likelihood to determine (i) which covariates will be included as environmental factors and (ii) the significance of the 1989 intervention effects on larval survival, and if so, in what manner. Table 3.2 summarizes the model comparison results for a number of models with 3 different specifications of intervention effects for the intertwining intervention in 1989 and 2 sets of covariates, based on which the model defined by Eqn. (3.2) is selected.

3.6 Conclusion

We applied the PPGAM for extracting the rich ecological information from a panel of otolith time series, which is useful for revealing the dynamics of fish larvae. Environmental effects had a significant influence on the timing of spawning and on survival of pollock larvae in GOA. Cold temperatures prior to spawning tended to increase later larval abundance, possibly because of delayed spawning and egg duration, resulting in a shorter period between hatching and capture. Moderate drift out of GOA benefited larval pollock, while strong drift tended to reduce the larval stock. Pollock larvae had higher survival in a warm environment during larval life, while strong winds decreased larval survival. High PDO conditions depressed spawning and hatching of pollock, and also intensified the spawning process. The fitted model

Table 3.2: Log marginal likelihood of various models

Environmental Covariates	Intervention Specification	Log Marginal Likelihood
Transport, SST and Wind Speed	Intervention effects began in 1989 and diminished geometrically afterwards	-1122.4
Transport, SST and Wind Speed	Intervention was only effective in 1989	-1125.5
Transport, SST and Wind Speed	No intervention effects on larval survival	-1127.7
Transport, SST, Wind Speed and PDO	Intervention effects began in 1989 and diminished geometrically afterwards	-1032.7
Transport, SST, Wind Speed and PDO	Intervention was only effective in 1989	-1075.8
Transport, SST, Wind Speed and PDO	No intervention effects on larval survival	-1065.8

depicts the baseline shape of the spawning time distribution as dome-shaped with a higher peak for the old spawner group than the young group. Therefore, harvesting of the older age classes may increase variability in the timing of spawning. Natural (regime shift) and/or human interventions (e.g. the Exxon Valdez oil spill) reduced the pollock larvae's survival rate in 1989. Because these events happened simultaneously, it is not possible to distinguish their effects. However, the post-1989 intervention effects on larval survival decayed very fast at a geometric rate after 1989, hence the system has a rather weak memory of the intervention.

CHAPTER 4
FORECASTING POLLOCK RECRUITMENT

In this chapter, we aim at developing two sets of models for forecasting pollock recruitment based on the abundances of their early life stages, which is either the late larval abundance or the abundance of 1 year old juvenile pollock. Pollock are caught as adults beginning at 4 to 5 years of age, so it is essential to fisheries for studying the recruitment of young pollock into age-4 pollock, which is known as the recruitment process. Many ecological studies on marine fishes point to the importance of environmental factors and/or abundances of their early life stages in predicting recruitment (e.g. Bailey et al., 1986; van der Veer, 1986; Meekan et al., 1993; Helle et al., 2000; Platt et al., 2003; Axentrot and Hansson, 2003; Svendsen et al., 2007). During the late 1980s to the early 1990s, the trends between the abundances of young pollock and the recruitment levels of age-4 pollock switched, which indicates that certain factors involved in pollock's recruitment changed during that period. The abundance of arrowtooth flounder (ATF), which is a major predator of the juvenile pollock, has increased substantially since early 1990s, and there was a greater overlap between the ATF active areas and the juvenile pollock nursery areas during the same period. These changes coincided with changes in the recruitment of pollock, which suggests a threshold structure in the predation effects to be studied below.

The study of the recruitment of age-4 pollock from their early life stages need to incorporate the effects of the environmental factors, the predators, and the community structure changes. The factor effects in the models are estimated nonparametrically, and the threshold generalized additive mixed models (threshold GAMM) are used to forecast the recruitment of age-4 pollock. With further analyses, some

factor effects are found to have linear forms. In a specific case, all covariates are found to affect the recruitment linearly except that the abundance of ATF has a threshold predation effect on age-1 pollock, resulting in the threshold generalized linear mixed model (threshold GLMM).

In the following two sections of this chapter, we discuss the age-4 recruitment forecasts from the late larval abundance and the abundance of age-1 juvenile pollock, respectively. In general, we introduce the background and motivation for the recruitment forecast of marine fishes, describe the structures and fitting of the forecast models, and evaluate the forecasting precision. Specifically, we discuss the selection of the environmental covariates and the forecast based on wild bootstrap. For the recruitment forecast from age-1 pollock abundance, we demonstrate the mechanisms for the autocorrelation structures in the fitted models, and present the likelihood ratio test for the validity of a threshold structure.

4.1 The Recruitment Forecast from Late Larval Abundance

4.1.1 Background and Methods

Study on the recruitment of marine fish is important for fishery, and recruitment forecast is useful especially for species with a short lag between birth and recruitment, or stocks where harvests rely on young fish due to depletion of older age classes. Forecasting the recruitment of young adult pollock (i.e. fish of 4 years old) from the early larval abundance is challenging, because the early larvae have not passed a critical period of high mortality. The survival rates of pollock larvae during this high mortality period vary a lot from year to year, so the relationship between the amounts of young adult pollock and their early larvae is not consistent. The recruitment forecast from the late larval abundance was applicable when the methods to estimate late larval abundance with the adjustment for size and

temperature-dependent mortality were developed.

Pollock spawn in the Shelikof Strait region of the western Gulf of Alaska mostly from late March to late April. At ambient temperatures eggs take about 2 weeks to hatch. The larval survey has been conducted from late May to early June each year in the Strait and downstream waters since 1979. Abundance was calculated from the mean number per 10 m^2 per station within the survey grid, multiplied by the number of 10 m^2 units in the grid. Because the raw abundance does not reflect the age structure of the population and older larvae have already passed through a critical period of high mortality, a weighting system for size and variability in the timing of the survey to give relative value to a larva's potential for recruitment was developed. This weighting system is an algorithm to estimate the number of larvae that will pass through a critical size of 15 mm based on mean historical growth and mortality rates (Bailey and Spring, 1992). Mean mortalities at different ages were determined from a regression of mortality against age for accumulated years in the GOA. Temperature is an important factor in mortality (Bailey et al., 1996) and hence a mortality schedule was developed to account for potential effect of SST on recruitment. From information presented in Bailey et al. (1996), we assumed as a crude adjustment such that for temperatures above the historical median, mortality was 0.5 times the average value, and for temperatures below the median, the mortality was 1.5 times the average value. Historical average temperatures for May in the region were obtained from the NCEP Reanalysis data set provided by A. Macklin (Pacific Marine Environmental Laboratory, NOAA, pers. comm.; http://www.cdc.noaa.gov/cdc/reanalysis/). Mean SST was interpolated across a longitude band in the GOA from $155.6°W$ to $157.5°W$ centered at latitude $56.2°N$. For 2007 and 2008 CTD casts were taken at each station and the mean SST for late May was calculated.

The abundance of age-4 recruits was obtained from Dorn et al. (2006, 2008b and 2009) which was estimated from age-structured assessment models using catch-at-age data in the fishery. The age-4 recruits from 1983 to 2006, corresponding to the year classes from 1979 to 2002, are used to estimate the recruitment forecast model. The updated recruitment data in the following years from 2007 to 2009 (year classes from 2003 to 2005) are applied to evaluate the model's forecasting precision. The abundance of arrowtooth flounder (ATF), as a main predator of the juvenile pollock, was obtained from Turnock and Wilderbuer (2007).

Data on pollock larval abundance, pollock and ATF biomass are partly missing as they are unavailable in 1980 and 1984. We imputed the missing values as follows. The missing late larval abundances were computed as the product of the age-1 abundance times the geometric mean ratio of late larval abundance to age-1 abundance, with the latter estimated with the complete data set from 1982 to 2005. For the missing pollock and ATF biomass data, they were computed by linear interpolation on the time series data.

Potentially important environmental factors that may affect the survival of a late larval cohort to age-4, include the drift out of Shelikof Stait (Transport), sea surface temperature (T), surface wind speed in Shelikof Strait (W), and the Pacific Decadal Oscillation (PDO) (SST data described above; the wind speed covariates were computed from sea level pressure data collected twice per day in Shelikof Strait ($56°N$, $156°W$, data source; transport was calculated from "line 8", a transect of stations across Shelikof Strait: data provided by A. Macklin and M. Spillane, Pacific Marine Environmental Laboratory). Sea surface temperature and PDO are monthly data. Based on values for the daily Transport and daily W, we calculated their monthly averages respectively. An issue is over which period within a year a particular environmental covariate affected survival. Thus, for each environmental

covariate, we computed their twelve monthly means, five seasonal means and yearly means, thereby creating eighteen variables per environmental factor. Later, we adopt some model selection criterion to determine which of these mean covariates enter into the recruitment model. For example, the twelve monthly average sea surface temperature are denoted by $T.Jan$ to $T.Dec$; the five seasonal averages as the mean level from January to March (JFM as the extension to the variable name), from April to June (AMJ), from July to September (JAS), from January to April (JFMA) and from May to June (MJ), plus the annual mean (Year). The environmental data used for the recruitment model estimation and forecasting are available from 1979 to 2005 except for wind speed. We only had wind speed data through 2002, which is fine for the phase of model estimation, because the model is built based on the data from year classes 1979 to 2002. However, wind speed data from 2003 to 2005 are needed to compute the age-4 recruitment forecast. In this study, we built a time series regression model to impute the wind speed information after 2002; see the model subsection.

Abundance data were log-transformed to normalize the distribution and reduce conditional heteroscedasticity. We compared several different model formulations involving inclusion of different independent variables and thresholds (continuous and noncontinuous, or abrupt changes in, recruitment patterns) using the varying coefficient GAMM formulations in the **mgcv** package of R (Wood, 2001). All models were fitted by maximizing the penalized likelihood with penalty on the roughness of the function estimates (Wood, 2006). Although we compared several model structures, we only report the best model here based on the lowest generalized cross validation scores (GCV); the GCV of a GAMM is a proxy for the model's mean squared out-of-sample prediction error (Wood, 2006). We also used model diagnostic checks such as a Quantile-Quantile normal score plot of residuals and Shapiro-Wilks normality

test, as well as checking of constant variance and test for autocorrelation of residuals (ACF and Ljung-Box tests); see Cryer and Chan (2008, Chapter 8).

In the final model, we consider the wind speed effect, larval abundance effect and ATF predation effects at different sea surface temperature (SST) levels. Considering a possible shift of the ATF spatial distribution since late 1980s to early 1990s (see below), we propose a model with different ATF predation effects before and after a threshold year during that period. However, the final fitted model shows that the ATF predation is only effective after a threshold year. The mechanism for this threshold predation effect assumes that before the ATF increased to the threshold level in late 1980s or early 1990s, there existed a relatively smaller spatial overlap between the ATF population and the juvenile pollock. Therefore, the predation effects from ATF before the early 1990s are trivial in pollock recruitment, and not considered in the model. Based on the forecasting model, we use a wild bootstrap method to forecast the number of age-4 recruits in 2007-2009 and point out several factors that may influence forecasting success.

4.1.2 Models

There is a generally positive relationship between the abundance of the pollock's late larvae and their recruitment 4 years later (Fig. 4.1). However, their positive relationship after 1988 is much weaker than that before, suggesting a change in the recruitment process starting some time from late 1980s to early 1990s. Indeed, one factor that may have brought about the aforementioned change is the increasing ATF biomass and the accompanied more overlap between the ATF active areas and the juvenile pollock's nursery areas since the late 1980s. In turn, the extent of overlap between the predators (ATF) and their prey (early stage pollock) is likely associated with temperature. In our case, sea surface temperature was used as a proxy for the bottom temperature (BT) because BT was not available for all years.

In the Bering Sea, ATF avoid cold water (Spencer 2008). In the GOA, ATF also tend to avoid cold water as in colder La Nina years they are found in warmer areas (Speckman et al. 2005). Therefore in some years ATF may avoid cold water over the shelf thus influencing their overlap with pollock prey. Additionally, preliminary analysis also suggests that wind speed is significant for the pollock recruitment model, while the pollock biomass, Transport and PDO are not.

Figure 4.1: Scatter plot of (log) late larval abundance versus (log) age-4 recruitment. (○) observations through 1988; (■) observations after 1988.
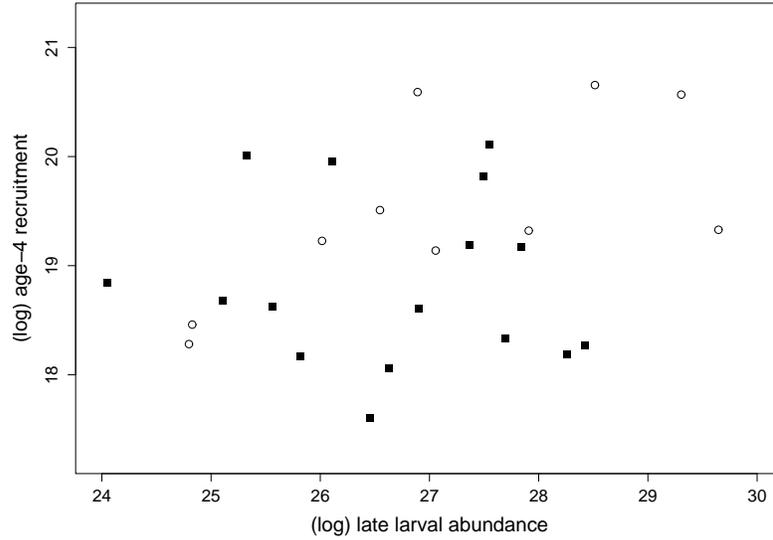
The fitted model includes late larval abundance, wind speed and an interaction term including ATF biomass and sea surface temperature. Moreover, it is found that the ATF predation effects took place after a threshold year, probably due to the fact that the spatial overlap between ATF and early-stage pollock was significant only starting at some time in late 1980s or early 1990s. Specifically, the recruitment model is of the following form:

$$y_{4,t+4} = \beta_0 + \beta_1 l_t + \beta_2 W_t + s(T_t)a_t 1_{(t>t_c)} + \xi_{t+4} \tag{4.1}$$

where the subscript t denotes year, $y_{4,t+4}$ denotes the log recruitment level of age-4 pollock spawned in year t and observed in year $t+4$, l_t denotes the late larval abundance of pollock in logarithm, and $y_{4,t+4}$ and l_t belong to the same spawning year class. Let W_t denote the wind speed, T_t the sea surface temperature, a_t the log ATF biomass, and $s(T_t)$ is a smooth function of T_t. Also, t_c is the threshold year to be estimated from the data, and the variable $1_{(t>t_c)}$ equals 1 in the years after t_c, and zero otherwise. The term $s(T_t)a_t 1_{(t>t_c)}$ represents the threshold ATF interaction with temperature, where $s(T_t)$ can be interpreted as the percent of juvenile pollock consumed per unit increase in ATF biomass at temperature T_t. The error terms $\{\xi_t\}$ are of zero mean and finite variance, but they may correlated over time with the auto-correlation structure modeled as some autoregressive (AR) process, e.g. AR(2). More specifically, $\xi_t = \phi_1 \xi_{t-1} + \phi_2 \xi_{t-2} + \varepsilon_t$, where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables of zero mean and finite variance, and ϕ_1 and ϕ_2 are the autoregressive parameters. Furthermore, the structure of Model (4.1) specifies that the late larval abundance (l_t) and wind speed (W_t) are linearly correlated with the recruitment level ($y_{4,t+4}$), which is supported by preliminary data analysis.

The mean structure of Model (4.1) contains a smooth function with unknown functional form, $s(T_t)$, while the regression errors are autocorrelated. Such a model structure can be fit within the framework of the generalized additive mixed models (GAMMs) (Lin and Zhang, 1999; Wood, 2006). Model (4.1) can be estimated by the method of penalized log-likelihood; the penalized log-likelihood is obtained by adding to the log-likelihood function a roughness penalty term on the smooth function so that maximizing the penalized log-likelihood function amounts to finding a model with good fit to the data and yet the function estimate is smooth. In practice, the penalty is proportional to the integrated squared second derivative

of the smooth function, i.e. $\lambda \int \{\ddot{s}(\cdot)\}^2$, where the smoothing parameter λ is a tuning parameter that controls the trade off between model fit and smoothness of the function estimate. All models reported herein were fitted in R via the **mgcv** package (Wood, 2006).

Model (4.1) contains two environmental covariates, W_t and T_t. For each of them, there are 18 choices representing the monthly, seasonal and yearly average levels respectively, as discussed in the data subsection. We employed the Akaike information criterion (AIC) for selecting the combination of their averages. We fitted Model (4.1) with each of the combinations of W_t and T_t, with the threshold year searched from 1978 to 2000, and compared their AIC values. Models with lower AIC values are preferred as they attain good fit to the data as well as model parsimony. We would also need to perform model diagnostics check on a selected model. If that model can pass the diagnostics check, its choice of the environmental factors and threshold year is accepted; otherwise, we consider the model with the next best AIC value, etc. The solid curve in Fig. 4.2 shows the profile of the best AIC among the combinations of W_t and T_t, as a function of the candidate threshold year, and the dashed curve in Fig. 4.2 represents the second best AIC profile. These two curves clearly show the best four AIC values obtained from all the choices of the environmental factors and threshold year, and these four choices comprise the first pool for the model selection discussed above. Additionally, since the lowest overall AIC is from a model with threshold year 1988 (Fig. 4.2), we studied the heat map of AIC for the combinations of environmental factors with that particular threshold year (Fig. 4.3). The darkest square in Fig. 4.3 suggests that Model (4.1) with the environmental covariates $W.MJ_t$ and $T.Year_t$ has the lowest AIC. Since this model can successfully pass the diagnostics check, it becomes the final fitted model:

$$y_{4,t+4} = \beta_0 + \beta_1 l_t + \beta_2 W.MJ_t + s(T.Year_t)a_t 1_{(t>t_c)} + \xi_{t+4} \qquad (4.2)$$

Figure 4.2: AIC profiles with different threshold years in Model (4.1). Solid curve: the best AIC among the environmental combinations at each threshold year; dashed curve: the second best AIC among the environmental combinations at each threshold year.
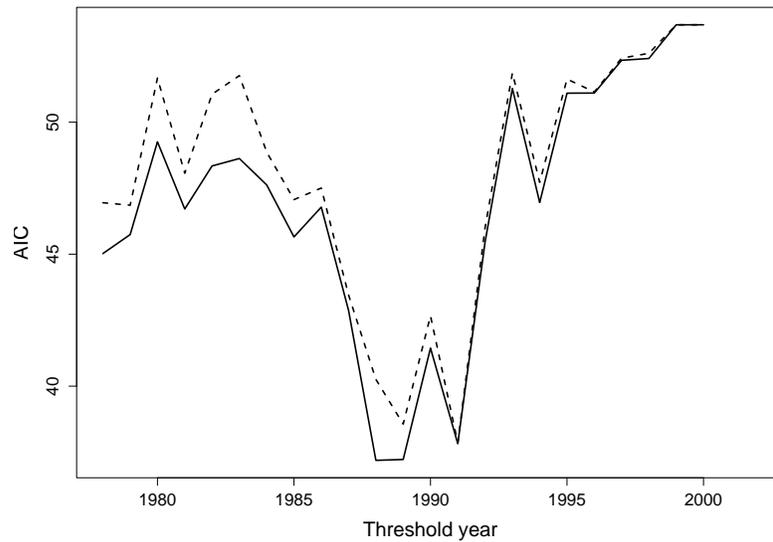
Figure 4.3: AIC heat map of Model (4.2) with various combinations of the environmental factors when the threshold year is 1988. Lower AIC is darker, and higher AIC is brighter; the white color represents NA value in AIC.
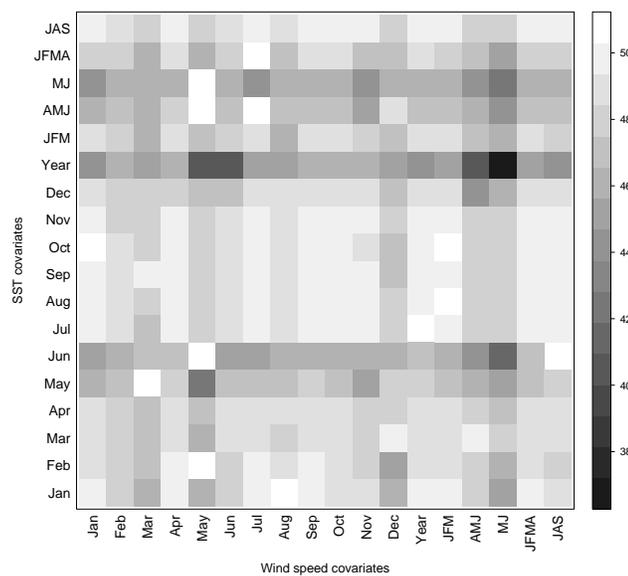
Furthermore, the five best environmental combinations in Fig. 4.3 are related to the wind speed in May, June or both. Thus, there is strong evidence that the wind speed in late spring or early summer is correlated with the age-4 recruitment from their late larvae.

As discussed in Section 4.1.1, the wind speed data are not available after 2002, so we develop a regression model for imputing the missing $W.MJ$, which is the wind speed covariate in the final fitted Model (4.2). Wind speed may be related to other environmental factors and their lags. Preliminary analysis suggests the inclusion of the lag-1 of $W.MJ$ term in the wind speed model. Using an approach similar to the one used in developing the recruitment model, the following model is found to be most appropriate for predicting $W.MJ$:

$$W.MJ_t = \gamma_0 + s(T.Jun_t) + \gamma_1 PDO.Jul_t + \gamma_2 W.MJ_{t-1} + \zeta_t, \qquad (4.3)$$

where $T.Jun_t$ is the sea surface temperature in June, $PDO.Jul_t$ denotes the PDO level in July. Greater flexibility is allowed in assessing the association between $W.MJ_t$ and $T.Jun_t$, by using a nonparametric smooth function $s(T.Jun_t)$, and $PDO.Jul_t$ and the lag-1 term $W.MJ_{t-1}$ are found to be linearly related to wind speed. The error terms $\{\zeta_t\}$ are assumed to be independent and identically distributed with zero mean and constant variance.

4.1.3 Forecast

In this study, the age-4 recruitment forecasts are obtained by bootstrapping. The forecasts for the age-4 recruitment after 2006 are computed as follows: (i) drawing a "future" realization of $W.MJ$ for 2003-2005 by bootstrapping the residuals, (ii) drawing a future realization of the age-4 recruitments for 2007-2009 by treating the $W.MJ$ realization computed in (i) as the true wind speed and bootstrapping the white noise process, and (iii) repeating (i) and (ii), say 10000 times. The average k-step ahead future recruitment forecasts, $k = 1, 2, 3$, are obtained by averaging the

k-step ahead recruitments computed in step (iii); their associated prediction intervals comprise of the 2.5 and 97.5 percentiles of the k-step ahead future recruitments. In the bootstrap approach, we simulate the future error terms by resampling the white noise residuals from the fitted model. Such a bootstrap sampling approach can provide valid predictions and associated prediction confidence intervals, even with unknown error distribution.

Since the error terms in Model (4.3) are independent and identically distributed with zero mean and constant variance, it is straightforward to get a bootstrap "future" realization of $W.MJ$ by the following formula:

$$W.\hat{M}J^*_{n+k} = \hat{\gamma}_0 + \hat{s}(T.Jun_{n+k}) + \hat{\gamma}_1 PDO.Jul_{n+k} + \hat{\gamma}_2 W.\hat{M}J^*_{n+k-1} + \hat{\zeta}^*_{n+k},$$

for $k = 1, 2, 3$, where n represents the last year with available wind speed data, which is 2002 in this study; the estimator $W.\hat{M}J^*_n = W.MJ^*_n$; and $\hat{\zeta}^*_{n+k}$ are generated by sampling with replacement from the residual $\{\hat{\zeta}_t\}$ of Model (4.3).

A wild bootstrap method is applied in the recruitment forecast to assure that the simulated white noise errors are symmetric around zero. Since the error terms in the recruitment Model (4.2) follow an AR(2) process, the bootstrap realizations of future recruits are computed by the following equations:

$$\hat{y}^*_{n+4+k} = \hat{\beta}_0 + \hat{\beta}_1 l_{n+k} + \hat{\beta}_2 W.\hat{M}J^*_{n+k} + \hat{s}(T.Year_{n+k})a_{n+k}1_{(n+k>\hat{t}_c)} + \hat{\xi}^*_{n+4+k},$$

$$\hat{\xi}^*_{n+4+k} = \hat{\phi}_1\hat{\xi}^*_{n+3+k} + \hat{\phi}_2\hat{\xi}^*_{n+2+k} + \hat{\varepsilon}^*_{n+4+k}, \text{ for } k = 1, 2, 3,$$

where $W.\hat{M}J^*_{n+k}$ is the bootstrap wind speed, $\{\hat{\xi}^*_{n+4+k}\}$ is randomly sampled from the white noise residuals $\{\hat{\varepsilon}_t\}$ in Model (4.2) but with their signs randomly flipped with probability $1/2$. Randomly flipping the signs of the residuals ensures that the bootstrap white noise has a distribution that is symmetric about zero. The error term $\hat{\xi}^*_t$ are generated dynamically by the above equation for $t > n + 4$, while $\hat{\xi}^*_t = \hat{\xi}_t = \hat{y}_t - [\hat{\beta}_0 + \hat{\beta}_1 l_{t-4} + \hat{\beta}_2 W.MJ_{t-4} + \hat{s}(T.Year_{t-4})a_{t-4}1_{(t-4>\hat{t}_c)}]$ are the regression

residuals from Model (4.2) for $t \leq n + 4$.

The above bootstrap procedure is repeated B times, with $B = 10000$ in our applications, to get the replicates of $\{\hat{y}^*_{n+4+k}, k = 1, 2, 3\}$. The mean or median trajectory of these future realizations provides the point predictions, with their associated 95% prediction bands enclosed by the 2.5 and 97.5 percentiles of the bootstrap future realizations.

4.1.4   Results

Since imputing the missing wind speed data after 2002 is a preliminary step in the recruitment forecast, we will first discuss the estimation and diagnostics of the wind speed model denoted by Eqn. (4.3).

A nonlinear association between $J.Jun$ and $W.MJ$ is shown by the curve in Fig. 4.4. When the sea surface temperature in June is at normal level, it is negatively related to the wind speed in May and June. However, when the temperature is anomalously low or high, there is a positive association between the wind speed and sea surface temperature. The positive estimate $\hat{\gamma}_1 = 0.24$ (with s.e. 0.07) indicates that stronger wind in May and June is always accompanied by a higher PDO level in July (Table 4.1). According to the estimate $\hat{\gamma}_2 = 0.5$ (with s.e. 0.11), the average wind speed in May and June is positively correlated with its lag-1.

Table 4.1: Parameter estimates of Model (4.3)

	Value	Std. Error	t-value	p-value
γ_0	2.46	0.59	4.18	<0.001
γ_1	0.238	0.066	3.60	0.0025
γ_2	0.498	0.11	4.48	<0.001

The residual diagnostics of Model (4.3) suggest that this wind speed model

Figure 4.4: Association between the wind speed and sea surface temperature in Model (4.3).
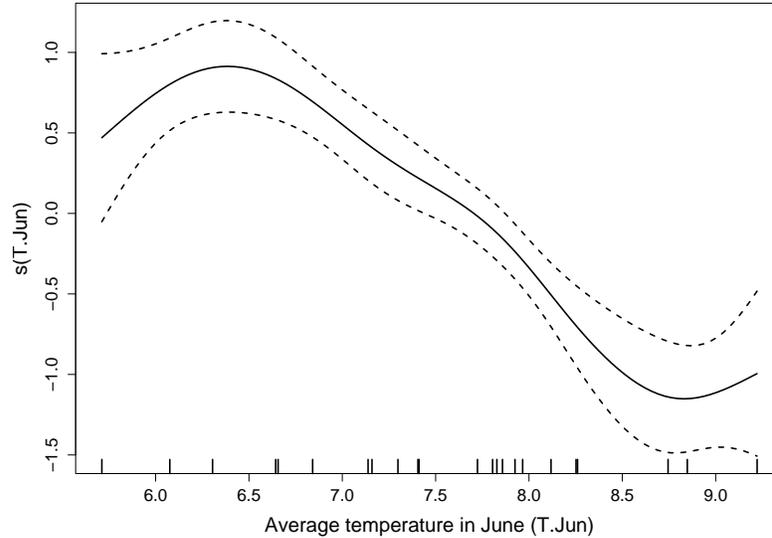
fits the data well. According to the Shapiro-Wilk normality test result (p-value = 0.497), the residuals of Model (4.3) follow a normal distribution approximately. The Ljung-Box test results indicate that there is no autocorrelation up to 10 lags among the residuals, which is also supported by the ACF plot. By checking the scatter plot of the residuals against the fitted values, we find that the variance of the errors is approximately constant. Hence, Model (4.3) is expected to provide reasonable estimates for the late spring wind speed from 2002 to 2005, which will be used in the recruitment forecast.

As discussed in the methods section, Model (4.2) has a better fit to the data than the other models in the framework of Model (4.1) with a similar model structure but different choices of the environmental factors. Additionally, Model (4.2) also shows its advantage over the other recruitment model structures according to

the model selection criterion. First, an AR(2) error process in Model (4.2) provides a better fit with lower AIC (37.2) than other autoregressive structures in $\{\xi_t\}$ (Table 4.2). Second, through comparing the fitting results from Model (4.2) and its constrained alternatives (Table 4.3), we found that all terms in Model (4.2) are "significant". The first alternative model for comparison does not take into account the ATF predation effect, and has the following form:

$$y_{4,t+4} = \beta_0 + \beta_1 l_t + \beta_2 W.MJ_t + \beta_3 T.Year_t + \xi_{t+4} \tag{4.4}$$

Table 4.2: Model comparison with various stochastic error processes in Model (4.2)

Autocorrelation Structure	No autoregressive	AR(1) error process	AR(2) error process	AR(3) error process
AIC	54.4	54.2	37.2	37.6
edf	6	7	8	9

Table 4.3: Comparison of Model (4.2) with its constrained models

Fitted Model	AIC	Adjusted R^2	edf
model (4.2)	37.2	74.3%	8
no ATF predation effect	54.5	41.9%	6
no environmental effects	44.3	59.5%	6
with only late larval abundance	50.6	47.6%	4
simple linear regression on late larval abundance	61.6	5.8%	2

Strictly speaking, Model (4.4) is not a constrained version of Model (4.2), since it includes a linear temperature effect denoted by $\beta_3 T.Year_t$. The second constrained model does not include any environmental factors, i.e. β_2 in Model

(4.2) equals to zero and $s(T.Year_t)$ in the predation effect is restricted to be a slope parameter. The third constrained model only considers the larval abundance effect, but retains the AR(2) structure in the error terms. The fourth model is only a simple linear regression on the (log) late larval abundance. Overall, Model (4.2) has a much better AIC(37.2) and adjusted R^2 (74.3%) than the other four simplified recruitment models. Furthermore, another model structure assessment concerns the threshold in Model (4.2). Applying a similar likelihood ratio test for the threshold structure in a GAMM, which is discussed in Zhang et al. (2010) (see, also, Section 4.2.2 below), the threshold structure in the ATF predation appears to be significant with p-value being equal to 0.048.

From the estimated $\hat{\beta}_1 = 0.076$ (with s.e. 0.042) in Table 4.4, the late larval abundance is positively related to the recruitment of age-4 pollock. The association between the late larval abundance and age-4 recruitment is weaker than the one between the age-1 pollock and their age-4 recruits (Zhang et al., 2010). It suggests that the survival of pollock's late larvae into the age-1 introduces more uncertainty in the recruitment process of age-4 pollock. It confirms the expectation that the recruitment forecast from late larval abundance will be more challenged than the forecast from age-1 abundance. The slope of the linear wind effect is -0.43 (with s.e. 0.13), which indicates that strong wind impedes the recruitment of pollock. Fig. 4.5 reveals the predation effect on pollock recruitment in two aspects. First, the curve is mostly below zero, which indicates that higher ATF biomass tends to reduce the recruitment of age-4 pollock. Second, the decreasing shape of the curve shows that the ATF predation effect is stronger in warmer environment, which results in a lower recruitment of age-4 pollock. The found ATF and temperature interaction is reasonable, because the ATF population tends to shift toward the nursery area of juvenile pollock in warmer environment.

Table 4.4: Parameter estimates of Model (4.2)

	Value	s.e.	t-value	p-value
β_0	19.9	1.6	12.7	<0.001
β_1	0.076	0.042	1.82	0.084
β_2	−0.428	0.13	−3.32	0.0036
ϕ_1	0.540			
ϕ_2	−0.821			

Figure 4.5: Temperature interaction with the ATF predation in model (4.2)
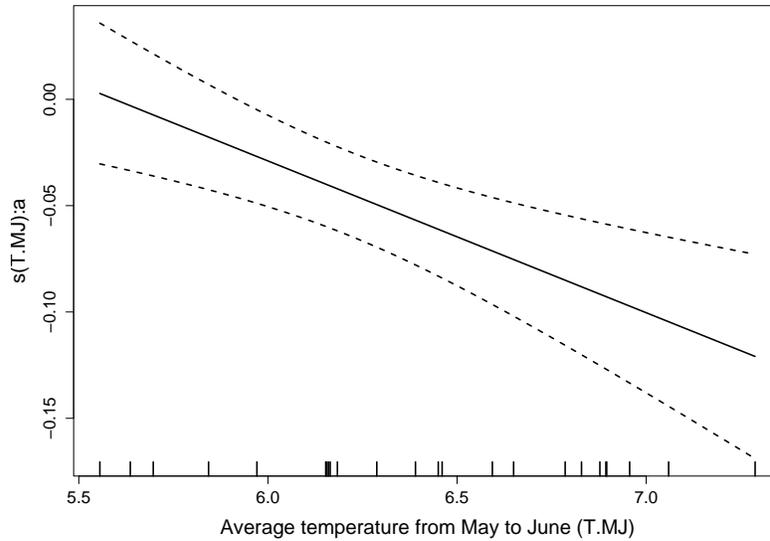
Additionally, more information in the pollock recruitment and their population's dynamics are revealed by the estimated autoregressive parameters, $\hat{\phi}_1 = 0.54$ and $\hat{\phi}_2 = -0.82$. The error process $\{\xi_t\}$ in Model (4.2) is stationary (Cryer and Chan, 2008) and has a stochastic cycle close to 5 years, which is consistent with stochastic structure in the age-4 pollock. The inter-group competition and cannibalism of the pollock provides a mechanism for the 5-year quasi-periodicity, see

Zhang et al. (2010).

The normalized residuals equal the standardized residuals pre-multiplied by the inverse square-root of the estimated error correlation matrix, and are assumed to follow an independent normal distribution with zero mean and constant variance. The linear trend in the Q-Q normal plot of the normalized residuals (upper left plot in Fig. 4.6) suggests that the normal distribution assumption for the errors is satisfied, which is also supported by the Shapiro-Wilk normality test result (with p-value 0.89). The scatter plot of the normalized residuals against fitted values (upper right plot in Fig. 4.6) indicates that the variance of the normalized residuals is approximately constant. Both the ACF plot and Ljung-Box test suggest no autocorrelation among the normalized residuals. Therefore, these diagnostics suggest that the fitted Model (4.2) provides a good fit to the data.

Figure 4.6: Residual checks for Model (4.2)
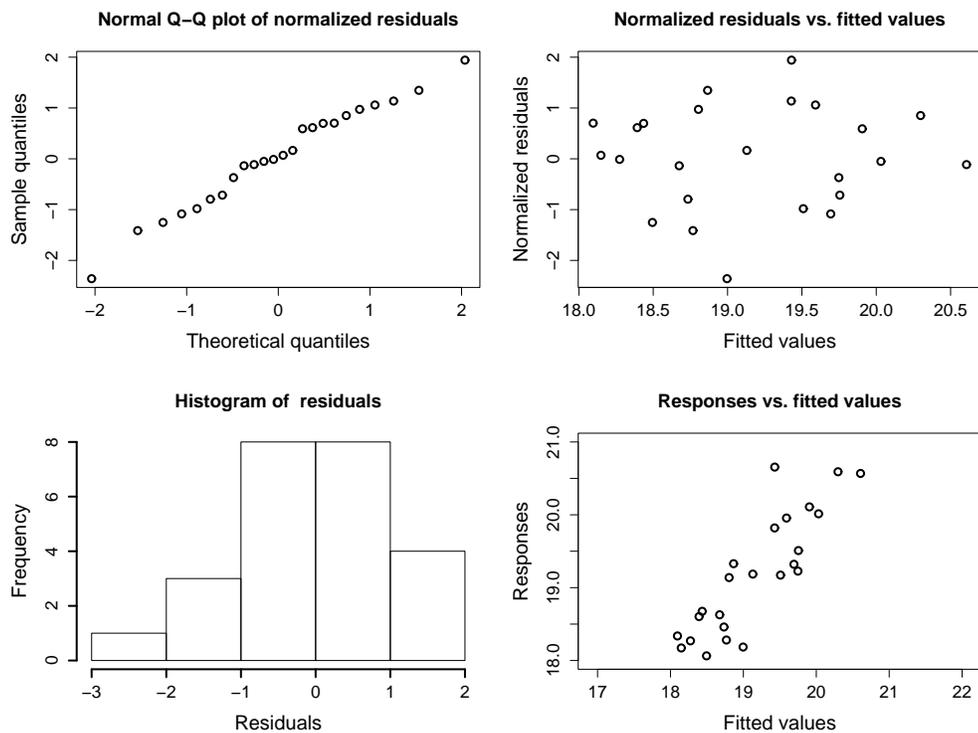
Applying the bootstrap forecast discussed in the methods section, 10000 bootstrap forecast trajectories of the age-4 recruitment has been generated over the period from 2007 to 2009. The median forecast trajectory gives the point forecasts for future recruitments (dotted curve in Fig. 4.7). We also constructed the 95% prediction band (gray shaded area) based on the 2.5% and 97.5% quantiles of the empirical distribution of the bootstrap forecast paths. It is found that the 95% prediction band successfully covers the actual age-4 recruitment levels in 2007 and 2008, and just misses the 2009 recruitment observation. Therefore, the out-of-sample forecasts from Model (4.2) match well with the actual age-4 recruitment from late larval abundance, especially for the near-term recruitment forecast. Additionally, the dynamic trend of the estimated pollock recruitment levels (dashed curve) is consistent with the observed values (gray solid curve), so Model (4.2) clearly provides a good fit to the recruitment data through 2002.

4.1.5 Discussion

We propose a forecasting strategy based on larval surveys that accounts for several factors: (1) weighting to adjust larval abundance for size composition and temperature-dependent mortality, (2) accounting for potential predation occurring during juvenile life and interactions with environmental factors, and (3) recognizing that shifts have occurred in factors influencing recruitment. There are only a few forecasting schemes that have used combined monitoring and environmental statistical models as we have done. One example attained relatively good results using larval abundance, spawning stock biomass and environmental factors occurring during larval life (Axentrot and Hansson, 2003), however in that study, since the environmental factors selected overlapped with the larval period they served to adjust the fit of larval abundance with recruitment. The difference in our strategy is an approach that incorporates factors that account for variability in survival after

Figure 4.7: Observed recruitment and predicted recruitment from Model (4.2). Numbers are abundance of 4 year olds occurring in each year. Gray solid line: observed recruitment over the study period; dashed line: estimated recruitment; dotted line: out-of-sample recruitment forecasts; 95% prediction band of the recruitment forecasts is shaded in gray; gray circles: recent observed recruitment in 2007 to 2009.
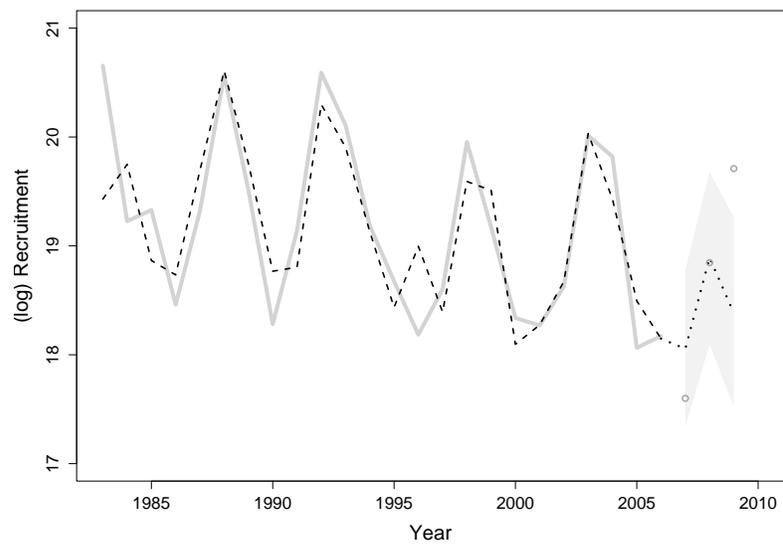
the larval monitoring survey. Our results support the forecast model of Zhang et al. (2010) based on age-1 pollock abundance coupled with ATF biomass, but push back the forecast to the larval stage.

Recruitment is a complex process, which impacts the strategy to forecast recruitment (Bailey et al., 2005; Houde, 2008). When dynamics are complex due to nonlinear systems and interactions, predictions become more uncertain as the time between initial conditions and the prediction target increases. Forecasts are most difficult to make years in advance strictly from environmental conditions because of the changes that occur in the relative abundance of a cohort as it develops. For pollock there is a significant trend of abundance with recruitment by the age-1 juvenile stage (Bailey et al., 2005; Zhang et al., 2010). Because of interannual variability of mortality during juvenile life, it is expected that larval abundance, taken as a whole, is not an accurate or precise predictor of recruitment as we found in this study. In fact, total larval abundance reflected the spawning stock abundance, even nearly 2 months after peak spawning (Bailey, unpublished data). However, weighting the abundances of larvae for the parameters of size- and temperature-dependent mortality improved the forecast value during the period when larval survival and recruitment were coupled. We did not make an attempt to optimize the fit, which might be useful. We also believe that the larval abundance estimates could be improved, most significantly by expanding the area surveyed. Areal coverage outside of the main grid is inconsistent and in some years relatively large numbers of larvae have been caught outside the main survey grid. We believe this may have been the case in the outlier years when the main larval distribution was observed to extend beyond the boundaries of our truncated grid, and perhaps for the forecast year 2005, when larval abundance was extremely low, but preliminary indications of recruitment indicate a weak to moderate year class.

Changes occurring in the ecosystem structure in the late 1980s to early 90s had a remarkable effect on recruitment of pollock (Bailey, 2000; Ciannelli et al., 2005; Zhang et al., 2010). The decoupling of larval abundance and recruitment that occurred in the 1990s could be explained by the gradual buildup in the abundance of ATF, finally reaching a threshold level of predation. This threshold occurred at about the same time that ATF surpassed pollock as the dominant biomass of groundfish in the GOA. We hypothesize somewhat more complicated mechanism than a simple buildup of biomass, such that along with the increase of ATF, a slow shift in their distribution may have also occurred so they have become concentrated in the juvenile pollock nursery region. The arrowtooth biomass has accumulated to an extent where it may consume a large portion of the pollock year class; however, whereas previously it was distributed across the western Gulf, it appears that in the early 1990s a large portion of the population shifted towards the Shelikof Strait region nursery area. Such "swarming" is a powerful density-dependent control mechanism. It is furthermore intriguing that 1999, the one pollock year class that was relatively strong during this period and where we have information on arrowtooth distribution, the predator was relatively diminished in the Shelikof region, possibly blocked from the nursery by a cold water mass. The interaction of arrowtooth biomasss and ocean temperature with pollock recruitment is supported by our exploratory threshold GAMM analyses. A better understanding of these relationships will be difficult to unravel from historical data alone, given that the trawl surveys are triennial, and the groundfish food habits database is lacking good information on arrowtooth feeding in autumn and winter, the likely seasons when they are feeding heavily on age-0 pollock. In summer when most stomachs are collected for examination, age-0 pollock are mostly pelagic and have limited vulnerability to arrowtooth predation. However, these results do stress the importance of

understanding the complex mechanisms involved in the recruitment process before expecting an accurate and resilient forecasting capability.

Alternative explanations for the shift in recruitment conditions should be considered. One potential factor is a shift in the location of spawning or a density-dependent Allee effect, such that Shelikof recruits are spawning elsewhere (Ciannelli et al., 2007). By this mechanism recruitment and larval abundance have become decoupled. Another is a change in physical environment (Hollowed et al., 2001). Examination of these processes also merits further examination.

There are significant obstacles to making recruitment forecasts from statistical models incorporating environmental factors; over time, forecasts based on environmental factors fail (Walters and Collie, 1988), and relationships with environmental factors change in significance. In some cases, especially when a large number of indirect factors are included, the probability of a false positive (Type I error) increases, contributing to the forecast's downfall in future years. In other cases, a large-scale atmospheric effect may be implicated, such as the PDO (Pacific Decadal Oscillation), but the dominance of these factors over local conditions may shift over time. For example in the Gulf of Alaska and Bering Sea the PDO's dominant influence on oceanic conditions was supplanted by another climate pattern (Bond et al., 2003), now known as the Victoria pattern (McKinnell, 2004). Shifting environmental conditions have been implicated in complex dynamics of walleye pollock recruitment in the ocean off Japan as well (Shida et al., 2007) and a changing baseline makes forecasting from environmental factors very difficult. In such conditions, short term bases for forecasts might be a better approach rather than attempting to utilize a longer time series. We also believe that significant advances in utilizing data in our statistical models are possible, such as improved understanding of processes influencing larval mortality.

Better accounting for stock structure in the Gulf of Alaska is another area where forecasting improvements could be made. Currently the recruitment level is estimated from the fishery catch-at-age data (VPA) for the whole western Gulf of Alaska, but forecasting schemes all assume that all recruits are derived from the Shelikof Strait spawning population. Beginning in the late 1990s the Shumagin Islands population became a significant component of the western Gulf of Alaska population and the inclusion of these fish in the recruitment estimate confounds comparisons of recruitment and larval numbers (and environmental conditions) from Shelikof. Therefore, knowledge of stock structure relative to the population sampled by the larval census and that estimated from the age-structured fisheries models is critical information needed for a meaningful forecast.

We believe that reasonable forecasts of recruitment are possible but should be tuned each season as new information becomes available. Starting with environmental conditions, at various abundance checkpoints along the way, observed deviations may provide valuable information on changing conditions. If the larval rough count is continued as a recruitment forecasting parameter, we suggest a forecasting scheme based on the concept that forecasting accuracy will improve as new information becomes available during the development of a cohort, using a combined monitoring and modeling effort. Managers need to decide whether a forecast at a stage earlier than age-1 is cost effective; but regardless of the management requirements, there is value in attempting to forecast recruitment from larval abundance as it provides feedback to our understanding of the recruitment process, and *vice versa*.

4.2 The Recruitment Forecast from Age-1 Juvenile Pollock

4.2.1   Introduction

Marine scientists searched for ways to forecast the abundance of commercial fish stocks since the early 1900s (Hjort, 1914, p. 227; NOAA, 2008), especially for stocks where harvests rely heavily on recruitment of young fish due to depletion of older age groups. However, forecasting recruitment from environmental conditions (Walters and Collie, 1988), or from abundances of early life stages (Mukhina et al., 2003) has proven difficult due to the complexity of multiple interacting factors (Bailey et al., 2005, Houde, 2008). Forecasting the abundance of walleye pollock *Theragra chalcogramma* in the Gulf of Alaska (GOA) would be especially useful because the fishery is largely dependent on the recruitment of strong year classes, and has recently been near collapse, having declined to about 30% of its estimated unfished biomass (Dorn et al., 2009).

Walleye pollock is an important component of North Pacific ecosystems, both as a predator and competitor of other fishes, and as prey to seabirds and marine mammals, and it is currently the world's second largest fishery. Walleye pollock abundance in the GOA increased dramatically in the late 1970s with a series of strong year classes, and began a long downward trend in the late 1980s; both periods coincided with changes in ocean conditions. However, community structure in the GOA has also changed (Anderson and Piatt, 1999; Litzow and Ciannelli, 2007), and an important source of predation mortality of juvenile pollock, arrowtooth flounder *Atheresthes stomias* (ATF), has dramatically increased in abundance over the past 2 decades. In fact, in the 1990s ATF surpassed pollock as the dominant groundfish species (by biomass) in the GOA.

Numerous attempts have been made to forecast pollock recruitment from environmental effects on eggs and larvae (e.g. Lee et al., 2009). In this study we examine juvenile pollock because, generally speaking, recruitment prediction from the abundance of older cohorts should be more accurate than that from egg or larval abundances (Bradford, 1992; Helle et al., 2000) or environmental factors alone (Axentrot and Hansson, 2003). Here we use statistical models to link juvenile survey data to predictor variables that influence their later survival to forecast the recruitment of pollock to the fishery. By testing and comparing several statistical models we examine whether (1) an increasing abundance of predators, specifically ATF, in the GOA strongly affects pollock recruitment, (2) the weight of predictor variables shift with phase (or regime) shifts in the marine environment, and (3) intercohort interactions are important. The results presented here indicate that forecasting recruitment of marine fishes needs to account for changes in community structure, rather than just environmental correlates with egg and larval survival.

4.2.2 Methods

Walleye pollock spawn in Shelikof Strait, Gulf of Alaska from late March to early May. Echo integration-trawl (EIT) surveys have been conducted annually by the Alaska Fisheries Science Center since 1981 (except 1982 and 1999) to assess the biomass of pollock before spawning in the Shelikof Strait area. In early surveys (1981-91) biomass estimates were obtained with a Biosonics acoustic system and from 1992 onward data were obtained with a Simrad EK echosounder. An index of age-1 juveniles was estimated as the abundance of 9 to 16 cm fish in the Shelikof Strait EIT surveys (McKelvey, 1996) and was reported in Dorn et al. (2007). The abundance of age-4 recruits was taken from Dorn et al. (2006, 2008b and 2009), estimated from age-structured population models of catch-at-age data in the fishery. Pollock are caught as adults beginning at 4 to 5 years of age. The number of age-4

recruits is from the western GOA; we assume that the number of age-1 juveniles in the Shelikof region is a relative index of juvenile abundance in the western GOA, and that variability in their survival reflects larger-scale patterns. In the early years of the pollock fishery most of the spawning in the GOA occurred in the Shelikof Strait, but over the years, other spawning areas have varied in their contribution. The biomass of age 3+ ATF was taken from Turnock and Wilderbuer (2007).

The models proposed in this paper are used to study the recruitment of young adult pollock of age-4 from early juvenile fish of age-1, incorporating the effects of predators and environmental factors. Additionally, since the age-4 recruits are auto-correlated, we depict this autocorrelation using one of the following 2 model structures:

$$y_{4,t+3} = f_1(y_{1,t}, a_t, e_t) + \xi_{t+3}, \tag{4.5}$$

and

$$y_{4,t+3} = f_1(y_{1,t}, a_t, e_t) + \sum_{j=m_1}^{m_2} g_j(y_{4,t+j}) + \varepsilon_{t+3}, \tag{4.6}$$

where t denotes year, m_1 and m_1 are 2 integers (for example, $m_1 = 1$ and $m_2 = 2$) $y_{k,t}$ represents the log abundance of the k-year old pollock in Year t, $\{\xi_t\}$ are colored noise with zero mean and finite variance, i.e. auto-correlated, whereas $\{\varepsilon_t\}$are independent and identically distributed noises with zero mean and finite variance. Thus, $y_{1,t}$ denotes the log age-1 abundance index in year t and $y_{4,t+3}$ is the log abundance of age-4 pollock 3 years after t; hence, $y_{1,t}$ and $y_{4,t+3}$ equal the density of the cohort of the pollock spawned in Year $t - 1$, when they are 1 and 4 years old, respectively. Additionally, a_t denotes the log abundance of ATF, which is the main predator of the juvenile pollock in the western GOA, and e_t represents a vector of environmental factors in year t, which includes the spring sea surface temperature (SST, i.e. the average SST from April to June), the fall SST (the average SST from July to September), the spring sea surface wind speed, the fall wind speed, and the

mean annual wind speed. The SST variables were derived from average monthly SST interpolated across a longitude band in the GOA from $155.6°W$ to $157.5°W$ centered at latitude $56.2°N$ (data source: A. Macklin, Pacific Marine Environmental Laboratory, pers. comm. http://www.cdc.noaa.gov/cdc/reanalysis/), The wind speed covariates were computed from sea level pressure data collected twice per day in Shelikof Strait ($56°N$, $156°W$, data source: A. Macklin & M. Spillane, Pacific Marine Environmental Laboratory pers. comm.). The term $f_1(y_{1,t}, a_t, e_t)$ summarizes the main effects on the recruitment of age-4 pollock. We tried several systems of notation and the preceding format was easiest to follow. A notation system formed around pollock year classes was complicated because of the 2 time sequences for pollock and ATF, and it was difficult to use pollock year class to indicate the time subscript of ATF. In the model comparison and selection, we employed Akaike's information criterion (AIC) for model comparison and selection. We also used model diagnostic checks such as the residuals' normality test, constant variance and autocorrelation checks among the residuals (ACF and Ljung-Box tests) (see Chapter 8, Cryer and Chan, 2008), and outlier detection methods.

For Model (4.5), any autocorrelation in the age-4 recruits beyond that induced by the main regression effects $f_1(y_{1,t}, a_t, e_t)$ is modeled by the auto-correlated error, $\{\xi_{t+3}\}$, which, in practice, is specified as some autoregressive (AR) error process. On the other hand, Model (4.6) uses the lagged recruitment $(y_{4,t+j}, j = 2, 1, 0)$ to account for any such autocorrelation in the recruitment data. Below, we refer to the autocorrelation in the recruitment beyond that induced by the main effects $f_1(y_{1,t}, a_t, e_t)$ as the extra autocorrelation. Because the response variable, $y_{4,t+3}$, is the recruitment of age-4 pollock in Year $t+3$, its lag-1 equals $y_{4,t+2}$. Similarly, $y_{4,t+j}$ in the term $g_j(y_{4,t+j})$ is the lag-$(3-j)$ recruitment.

The autocorrelation pattern in the recruitment models may result from inter-cohort interactions in the pollock population. Specifically, we demonstrate 3 such mechanisms in the simple case of linear and noise-free dynamics. One mechanism results in the AR error process as specified in Model (4.5), and the other 2 mechanisms introduce different lagged recruitment into Model (4.6).

First, the autocorrelation structure in Model (4.5) can be justified by cannibalism and/or competition from adult pollock, specified as the age-4+ (age-4 and older) group, in which case the (linear) recruitment dynamics is driven by the following system of equations:

$$y_{4,t+3} = \beta_0 + \beta_1 a_t + \beta_2 y_{1,t} + \omega y_{4+,t}, \tag{4.7}$$

$$y_{4+,t+1} = \gamma_4 y_{3,t} + \delta y_{4+,t}, \tag{4.8}$$

$$y_{3,t+1} = \gamma_3 y_{2,t} + \tau_3 y_{4+,t}. \tag{4.9}$$

Eqn. (4.7) models the recruitment from age-1 pollock to age-4, with the term $\omega y_{4+,t}$ representing the effects of cannibalism of the age-4+ group on the juvenile pollock and/or competition effects from the older pollock in the recruitment process. It is expected that ω is negative. Eqn. (4.8) accounts for the survival of the 4+ group from the previous year and new members from the age-3 group. Eqn. (4.9) accounts for the survival of age-2 fish to age-3 and competition from age-4+ pollock. Note that the γ's are expected to be between 0 and 1 and so is δ. Eqn. (4.7) is a special case of Eqn. (4.5) where f_1 is a linear function and the error term ξ_{t+3} equals $\omega y_{4+,t}$ which we now show to be an autoregressive process, hence the error term in Eqn. (4.5) is autocorrelated. Indeed, Eqs. (4.8-4.9) imply that

$$y_{4+,t+1} = \delta y_{4+,t} + \gamma_4 \tau_3 y_{4+,t-1} + \gamma_4 \gamma_3 y_{2,t-1}. \tag{4.10}$$

Since the values for γ's are expected to be between 0 and 1, the coefficient $\gamma_4 \gamma_3$ is probably negligible; therefore, the preceding equation can be approximated by

$$y_{4+,t+1} = \delta y_{4+,t} + \gamma_4 \tau_3 y_{4+,t-1} \tag{4.11}$$

.

In practice, the preceding relationship holds only on the average, so that a stochastic error term has to be added to the right side of the equation; hence, $\{y_{4+,t}\}$ is approximately an AR(2) process, with δ being the AR(1) coefficient and $\gamma_4\tau_3$ being the AR(2) coefficient. In particular, δ can be interpreted as the survival rate of the age-4+ pollock. The interpretation of the AR(2) coefficient estimate is more complex as it equals the product $\gamma_4\tau_3$, which is expected to be negative because of the assumed positivity of γ_4 and the negativity of τ_3 .

As for Model (4.6), the extra autocorrelation is assumed to be captured by the lagged recruits. Such an autocorrelation structure can be attributed to interactions between the young adult pollock (age-4) and the juvenile pollock, in which case the pollock population dynamics follow the following system of equations:

$$y_{4,t+3} = \beta_{0,4} + \beta_{2,4}y_{3,t+2} + \tau_4^* y_{4,t+2}, \tag{4.12}$$

$$y_{3,t+2} = \beta_{0,3} + \beta_{1,3}a_{t+1} + \beta_{2,3}y_{2,t+1} + \tau_3 y_{4,t+1}, \tag{4.13}$$

$$y_{2,t+1} = \beta_{0,2} + \beta_{1,2}a_t + \beta_{2,2}y_{1,t} + \tau_2 y_{4,t}, \tag{4.14}$$

$$a_{t+1} = \varphi a_t + \vartheta y_{1,t}. \tag{4.15}$$

Eqn. (4.12) shows the recruiting process of age-4 pollock from age-3, which includes the survival of age-3 fish given by $\beta_{2,4}y_{3,t+2}$ and the interactions between the age-4 pollock and the age-3 in year $t+2$ in terms of $\tau_4^* y_{4,t+2}$. The coefficient τ_4^* reflects the intergroup interactions in 2 aspects: (1) it measures the competition and/or cannibalism between age-4 and age-3 pollock, and (2) it accounts for misclassification between the age-4 group and its neighboring age groups, which commonly occurs when a certain portion of a strong year class are incorrectly aged and overflow into adjacent year classes. The competition and/or cannibalistic effects tend to reduce the recruitment of age-4 pollock, and misclassification probably results in positive association between the age-4 recruitment and its lag-1. Since these 2 interactions

are opposite in direction, the sign of τ_4^* is undetermined in Eqn. (4.12). Eqn. (4.13) accounts for the survival of age-3 pollock from the previous year's age-2 fish in the term $\beta_{2,3}y_{2,t+1}$, the predation from ATF represented by $\beta_{1,3}a_{t+1}$, and the competition and/or cannibalism between the age-4 and age-2 pollock denoted by $\tau_3 y_{4,t+1}$. In Eqn. (4.13), τ_3 is expected to be < 0, because it mainly assesses the intergroup competition and/or cannibalism. Similarly, the coefficient τ_2 in Eqn. (4.14) is expected to be negative. Eqn. (4.15) indicates that the abundance of the ATF predators is related to their last year's abundance and the corresponding age-1 pollock abundance, as ATF mainly eats age-0 and age-1 pollock, as well as some age-2 fish. The coefficients φ and ϑ are expected to be positive. Substituting Eqns. (4.13-4.15) into Eqn. (4.12) yields

$$
\begin{aligned}
y_{4,t+3} =\ & [\beta_{0,4} + \beta_{2,4}\beta_{0,3} + \beta_{2,4}\beta_{2,3}\beta_{0,2}] + [\beta_{2,4}\beta_{1,3}\varphi + \beta_{2,4}\beta_{2,3}\beta_{1,2}]a_t \\
& + [\beta_{2,4}\beta_{1,3}\vartheta + \beta_{2,4}\beta_{2,3}\beta_{2,2}]y_{1,t} + [\tau_4^* y_{4,t+2} + \beta_{2,4}\tau_3 y_{4,t+1} + \beta_{2,4}\beta_{2,3}\tau_2 y_{4,t}] \qquad (4.16) \\
=\ & \beta_0 + \beta_1 a_t + \beta_2 y_{1,t} + [\tau_4^* y_{4,t+2} + \beta_{2,4}\tau_3 y_{4,t+1} + \beta_{2,4}\beta_{2,3}\tau_2 y_{4,t}].
\end{aligned}
$$

The model represented by Eqn. (4.16) contains the lag-1 to lag-3 of the recruitment that generates the extra autocorrelation in the recruits of age-4 pollock. The parameters $\beta_{2,j}, j = 2, 3, 4$ are survival rates that are likely to fall between 0 and 1. If we further assume weak competition and/or cannibalism between the age-4 and the age-1 pollock (small τ_2 in magnitude), the coefficient of $y_{4,t}$ in Eqn. (4.16) is negligible compared with the coefficients of $y_{4,t+2}$ and $y_{4,t+1}$. Therefore, the recruitment equation can be simplified as:

$$
y_{4,t+3} = \beta_0 + \beta_1 a_t + \beta_2 y_{1,t} + [\tau_4^* y_{4,t+2} + \beta_{2,4}\tau_3 y_{4,t+1}]. \qquad (4.17)
$$

The terms within the square brackets of Eqn. (4.17) measures the lagged recruitment effects on the recruitment of age-4 pollock. Since the sign of the parameter τ_4^* is unclear, we cannot determine the sign of the lag-1 recruitment effect. However, we expect a negative lag-2 effect in the recruiting model with the negative τ_3 and

positive $\beta_{2,4}$. We can further generalize Eqn. (4.17) by replacing the linear lagged recruitment by nonlinear lag-1 and lag-2 effects in terms of $g_j(y_{4,t+j})$, j = 1,2. Similarly, we have constructed a mechanistic model that results in explaining the extra autocorrelation in terms of the lag-2 and lag-3 recruitment effects. However, as that model is discredited by the data (see Section 4.2.3); hence, for simplicity, we do not elaborate on the third model.

Although we mainly assume linear effects in the above derivation, the models may be rendered more flexible by allowing nonlinear effects in terms of some unknown smooth functions, and including stochastic errors in the models. Indeed, these models then fall into the general framework of the generalized additive mixed models (GAMMs) (Lin and Zhang, 1999; Wood, 2006). The recruitment models in the form of Eqn. (4.5) and Eqn. (4.6) can be represented as a GAMM. For the simple case of single covariate and Gaussian errors, the GAMM takes the following form:

$$w_t = s(u_t) + \xi_t, t = 1, 2, ...n, \tag{4.18}$$

where the response w_t bears a nonlinear relationship with the covariate u_t and the noise terms $\{\xi_t\}$ are auto-correlated. For the simple case that $\{\xi_t\}$ has a multivariate normal distribution with zero mean vector and covariance matrix Λ, the log-likelihood of Model (4.18) is given by $-0.5 \times \log|\Lambda| - \{W - s(U)\}^T \Lambda^{-1} \{W - s(U)\}/2$, up to some additive constant, where W is the vector of response values, $s(U)$ the vector of the smooth function s evaluated at the covariate values, $|\Lambda|$ is the determinant of the covariance matrix Λ. For estimating the unknown smooth function, we use the penalized likelihood approach, which tries to find the function estimate that provides good fit to the data and yet assures that the function is not too rough. The penalty term is a multiple of the integrated squared second derivative of the smooth function,

i.e.$\lambda \int \{\ddot{s}(u)\}^2 du$ where the non-negative parameter λ is known as the smoothing parameter that describes the trade-off between goodness of fit and smoothness of the function estimate. Altogether, the GAMM can be estimated by maximizing the penalized log-likelihood: $-0.5 \times \log|\Lambda| - \{W - s(U)\}^T \Lambda^{-1} \{W - s(U)\}/2 - \lambda \int \{\ddot{s}(u)\}^2 du$. It should be noted that if the noise terms are uncorrelated over time, then the GAMM becomes the generalized additive model (GAM); see Wood (2006) for details and other estimation methods.

Since a threshold structure was introduced for fitting the main effects in $f_1(y_{1,t}, a_t, e_t)$, we needed to develop a test for the validity of the proposed threshold structure. Considering a possible shift of the ATF spatial distribution from the late 1980s to early 1990s and the occurrence of an environmental regime shift about the same time (see below), we further proposed different ATF predation effects before and after a threshold year during that period. A threshold year effect in the models was assessed by testing the null hypothesis (H_0) that the ATF predation effect was present for all years versus the alternative hypothesis (H_a) that it started to become important to pollock survival after a threshold year. A likelihood ratio test was employed to justify the threshold structure in ATF predation. Let θ be the parameter vector, including the threshold year, $\log L(\hat{\theta}_{0,n})$ be the log likelihood function evaluated at the maximum likelihood estimators under the null hypothesis, and $\log L(\hat{\theta}_n)$ be the maximum log likelihood function under the general hypotheses. The log likelihood ratio, lr, can be written as $lr = \log L(\hat{\theta}_n) - \log L(\hat{\theta}_{0,n})$, which is used as the test statistic. For Model (4.5) with stochastic error process, the empirical distribution of the test statistic under the null hypothesis is obtained by the following bootstrap approach. Based on the residual vector (ξ) and the correlation matrix (Λ) estimated under the null hypothesis, we calculate the normalized residuals, $\hat{\varepsilon}_0 = \hat{\Lambda}_0^{-1/2} \hat{\xi}_0$. For each $k = 1, \ldots K$, we randomly permute the elements in $\hat{\varepsilon}_0$

to get $\hat{\varepsilon}_0^{(k)}$. Using the parameter estimator $\hat{\theta}_{0,n}$ and the residuals $\hat{\xi}_0^{(k)} = \hat{\Lambda}_0^{1/2}\hat{\varepsilon}_0^{(k)}$, we generate the new recruitment level data $\hat{y}_{4,t+3}^{(k)} = f_{1,\hat{\theta}_{0,n}}(y_{1,t}, a_t, e_t) + \hat{\xi}_{(t+3),0}^{(k)}$. Based on the generated data, $\hat{y}_{4,t+3}^{(k)}$, we calculate $lr^{(k)}$. The empirical distribution of the test statistic is formed from the $lr^{(k)}$ values. Additionally, to form such an empirical distribution based on Model (4.6), whose error terms are independent and identically distributed, we can generate the new data by bootstrapping the residuals directly. Finally, the p-value of the likelihood ratio test is calculated as the proportion of $lr^{(k)}$ values that are higher than the observed lr.

4.2.3   Results

There was generally a positive relationship between the age-1 abundance index and the recruitment of pollock to the fishery 3 years later (Fig. 4.8). However, the relationship appeared to have shifted downward after 1991, suggesting the presence of emerging factors on age-1 pollock survival. At around the same time, ATF abundance was increasing and surpassed pollock to become the dominant groundfish species in the GOA (Fig. 4.9). We therefore selected these factors (age-1 pollock abundance, ATF abundance, a threshold effect) and some environmental factors (e.g. temperature) to test and compare in statistical recruitment prediction models.

In the Methods, we discussed 3 mechanisms to derive the autocorrelation structures in recruitment models. For each of the 3 autocorrelation structures, we selected one fitted model based on AIC and performed model diagnostic results. As shown in Table 4.5, the selected model with the AR(2) error process has slightly better AIC score (22.5) and adjusted R^2 (81.5%) than the model including lag-1 and lag-2 recruitment effects with $AIC = 24.0$ and adjusted $R^2 = 80.7\%$; the model-fitting results from both models are discussed in detail below. The AIC of the model containing lag-2 and lag-3 recruitment (28.3) is much higher than the AICs from the other 2 selected models (Table 4.5), which indicates that this model is discredited

Figure 4.8: Scatter plot of (log) cohortspecific age-1 walleye pollock abundance versus (log) age-4 recruitment. (○) observations before 1992; (■) observations since 1992.
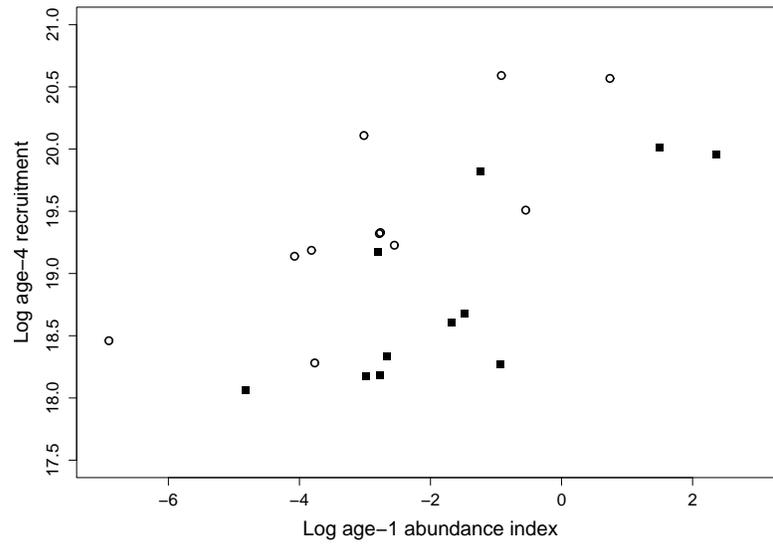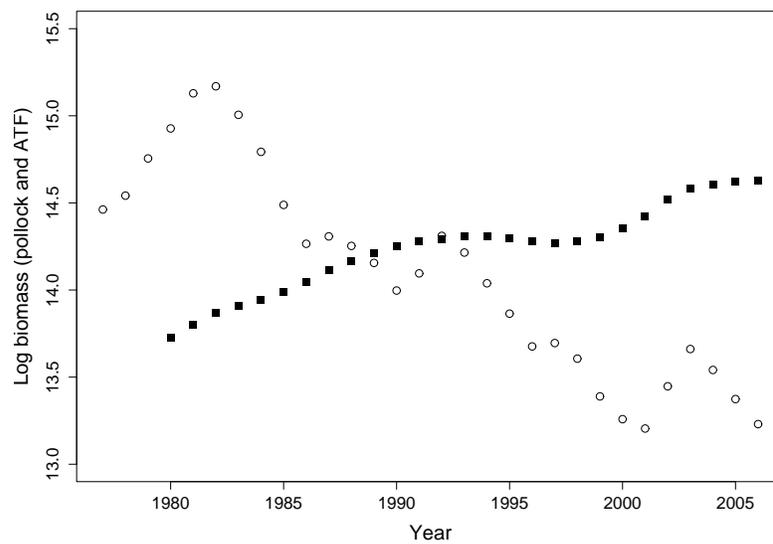

Figure 4.9: Time plot of (log) walleye pollock biomass and (log) arrowtooth flounder biomass. (○) (log) biomass of age 3+ pollock; (■) (log) flounder biomass.

by the data. Temperature was not significant in any of the models and was dropped from consideration. Including spring SST as an additive covariate results in an increase in the AIC for each of the 3 models reported in Table 4.5, e.g. an increase in AIC from 22.5 to 24.2 for Model (4.19) with AR(2) error structure; similarly, other environmental factors, including fall SST and various wind speeds, were found to be inconsequential. We also fitted Model (4.19) with SST as the threshold variable but that fitted model was deemed unacceptable based on model diagnostics and interpretation. Consequently, we shall confine our discussion to the first 2 models.

Table 4.5: Akaikes information criterion (AIC) of the fitted models to predict walleye pollock recruitment in the Gulf of Alaska

Fitted Model	4.19	4.20	Variant of 4.20
Autocorrelation Mechanism	(4.7-4.9)	(4.12-4.15)	
Autocorrelation Structure	AR(2) error process	lag-1 and s(lag-2)	lag-2 and monotone s(lag-3)
AIC	22.5	24.0	28.3

Table 4.6: AIC of the models with various stochastic error process in structure (4.5)

Autocorrelation Structure	No autoregressive	AR(1) error process	AR(2) error process	AR(3) error process
AIC	35.0	34.8	22.5	23.9

The first fitted model contains an auto-correlated error process $\{\xi_t\}$. Fitting results with different autoregressive structures in $\{\xi_t\}$ suggested that an AR(2) error process provided the best fit for the data with the lowest AIC values (Table 4.6).

The mechanism for this AR(2) error process is shown by Eqns. (4.7-4.9). A preliminary analysis indicated that the log age-1 abundance ($y_{1,t}$) and log ATF abundance (a_t) after the threshold year (t_c) were linearly correlated with the recruitment level ($y_{4,t+3}$). Additionally, no environmental factors enter into the model, probably because any change in environmental factors that strongly influenced juvenile survival was incorporated in the threshold shift term. Therefore, the fitted model with AR(2) error process has the following structure:

$$y_{4,t+3} = \beta_0 + \beta_1 a_t 1_{(t>t_c)} + \beta_2 y_{1,t} + \xi_{t+3}, \qquad (4.19)$$

where the dummy variable $1_{(t>t_c)}$ equals 1 in the years after the threshold year t_c, and 0 otherwise, so $\beta_1 a_t 1_{(t>t_c)}$ accounts for the threshold ATF predation effect on pollock recruitment after year t_c. The errors form a stationary Gaussian AR(2) process: $\xi_t = \phi_1 \xi_{t-1} + \phi_2 \xi_{t-2} + \varepsilon_t$ where ϕ_1 and ϕ_2 are the autoregressive parameters and the ε_t are independent and identically normally distributed errors, so that $\xi = (\xi_{t_0+3}, \xi_{t_0+4}, ..., \xi_{t_L+3})^T$ follows a multivariate normal distribution $N(0, \sigma^2 \Lambda)$, where $(t_0 + 3)$ and $(t_L + 3)$ denote the first and last recruiting years of age-4 pollock in the study period respectively, and Λ is the correlation matrix with an AR(2) structure. More specifically, $\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}, j \geq 2$, where ρ_j is the correlation between ξ_t and ξ_{t-j}, with the initial conditions $\rho_0 = 1$, $\rho_1 = \phi_1/(1-\phi_2)$ and σ^2 is the stationary variance of ξ_t.

To check whether the threshold structure of the ATF predation effect in Model (4.19) was appropriate for the data, we employed the likelihood ratio test discussed in the Methods. The p-value of the likelihood ratio test for the threshold structure in Model (4.19) is 0.024. Thus, there is strong evidence that the ATF predation affected pollock recruitment after a threshold year t_c.

According to the profile AIC of the threshold years (Fig. 4.10), 1991 was estimated to be the threshold year for Model (4.19). Model (4.19) has 6 parameters,

namely, 3 β values, 2 autoregressive coefficient parameters and 1 threshold param-eter. Thus, with a sample size of 22, the residuals have 16 degrees of freedom. As shown in Table 4.7, The estimated β_1 equals -0.052 (with s.e. 9.17×10^{-3}), which indicates that ATF significantly decreases the recruitment of age-4 pollock, such that a 1% increase in ATF results, on average, in 0.052% decrease in recruitment. Additionally, the age-1 pollock abundance index is positively related with the re-cruitment of age-4 pollock ($\hat{\beta}_2 = 0.209$, with s.e. 0.039), which indicates that age-1 abundance is an important factor explaining the variability in the recruitment of age-4.

Figure 4.10: Akaikes information criterion (AIC) levels with different threshold years in Model (4.19)
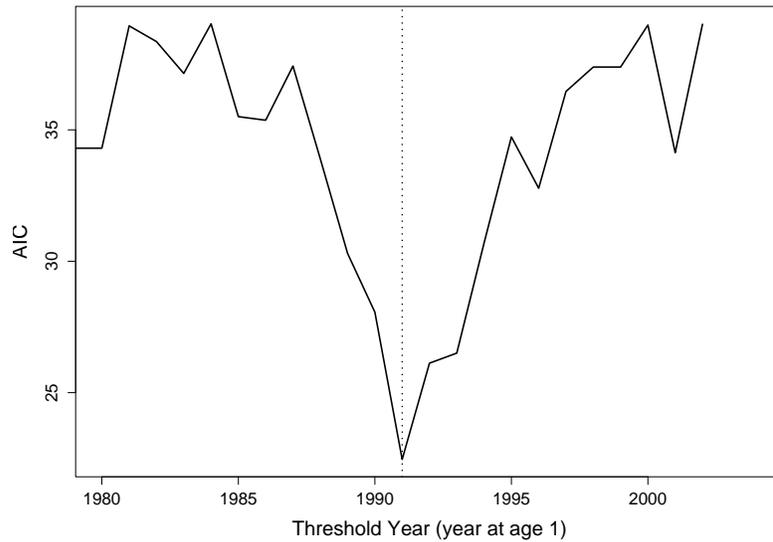
Based on the estimated autoregressive parameters, $\hat{\phi}_1 = 0.546$ and $\hat{\phi}_2 = -0.719$ (Table 4.7), the error process $\{\xi_t\}$ is, indeed, stationary (see Cryer and Chan 2008, p. 72). The average length of the stochastic cycle, $2\pi / \cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]$, is approximately 5 years (Fig. 4.11). The competition and cannibalism between the

cohorts of different ages, especially the cannibalism of the adult pollock on the juveniles, as detailed in Eqns. (4.7-4.9), are instrumental for the 5-year quasi-periodicity. Additionally, according to the discussion of the stochastic mechanism for this model in the previous section, the AR(1) parameter ϕ_1 can be interpreted as the average survival rate of pollock of ages 4+, which is estimated to be 0.546 (= 54.6%), although with considerable uncertainty as the 95% confidence interval ranges from 28.5% to 63.5%.

Table 4.7: Estimates of Model (4.19) for predicting walleye pollock recruitment using 1991 as the threshold year

	Value	SE	t-value	p-value
β_0	20.0	1.33×10^{-1}	149.78	<0.001
β_1	-0.0520	9.17×10^{-3}	-5.67	<0.001
β_2	0.209	3.86×10^{-2}	5.41	<0.001
ϕ_1	0.546	1.43×10^{-1}	3.82	1.52×10^{-3}
ϕ_2	-0.719	1.31×10^{-1}	-5.50	<0.001

The normalized residuals equal the standardized residuals pre-multiplied by the inverse square-root of the estimated error correlation matrix. In the model assumptions, the normalized errors follow the independent normal distribution with zero mean and constant variance. The linear trend in the Q-Q normal plot of the normalized residuals (upper left plot in Fig. 4.12) and the Shapiro-Wilk normality test result (p-value = 0.58) suggest that the normal distribution assumption in the errors is satisfied. The scatter plot of the normalized residuals (upper right plot in Fig. 4.12) shows that the constant variance assumption for the normalized errors appears appropriate. The normalized residuals appear to be uncorrelated over time, which is supported by the ACF and Ljung-Box test. Therefore, the assumptions of

Figure 4.11: Theoretical autocorrelation function (ACF) of the AR(2) error process from the fitted Model (4.19), showing the quasiperiodicity of the error process.
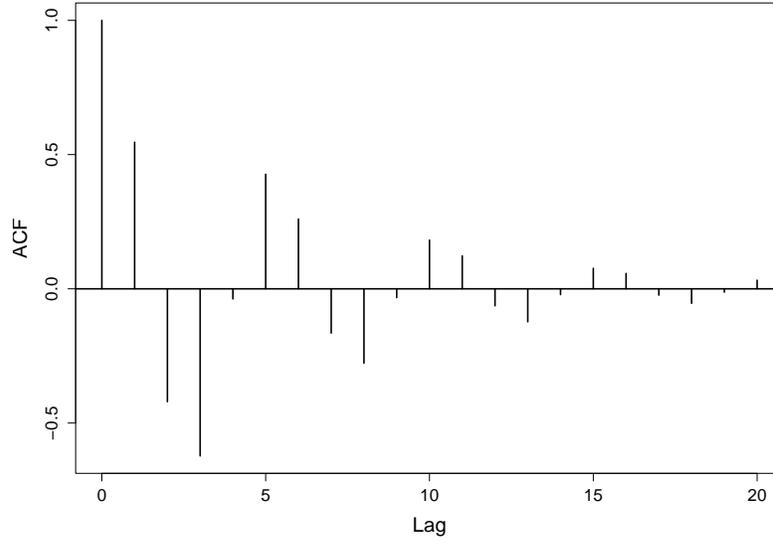
the error terms are satisfied approximately for the fitted Model (4.19), suggesting that it provides a good fit to the data.

Owing to the AR(2) correlation structure in the error term, $\{\xi_t\}$ and that the covariates in Model (4.19) lag the recruitment by three years, we can compute out-of-sample k years ahead forecasts for k=1,2 and 3 years into the future by the following formula:

$$\hat{y}_{4,n+k} = \hat{\beta}_0 + \hat{\beta}_1 a_{n+k-3} 1_{(n+k-3>\hat{t}_c)} + \hat{\beta}_2 y_{1,n+k-3} + \hat{\xi}_{n+k}$$

where n is the last year of the study period, being 2006, and $\hat{\xi}_{n+k}$ are computed recursively by the formula:

$$\hat{\xi}_{n+k} = \hat{\phi}_1 \hat{\xi}_{n+k-1} + \hat{\phi}_2 \hat{\xi}_{n+k-2}$$

with $\hat{\xi}_t = y_{4,t} - (\hat{\beta}_0 + \hat{\beta}_1 a_{t-3} 1_{(t-3>\hat{t}_c)} + \hat{\beta}_2 y_{1,t-3})$ being the regression residuals from

Figure 4.12: Residual checks of Model (4.19)
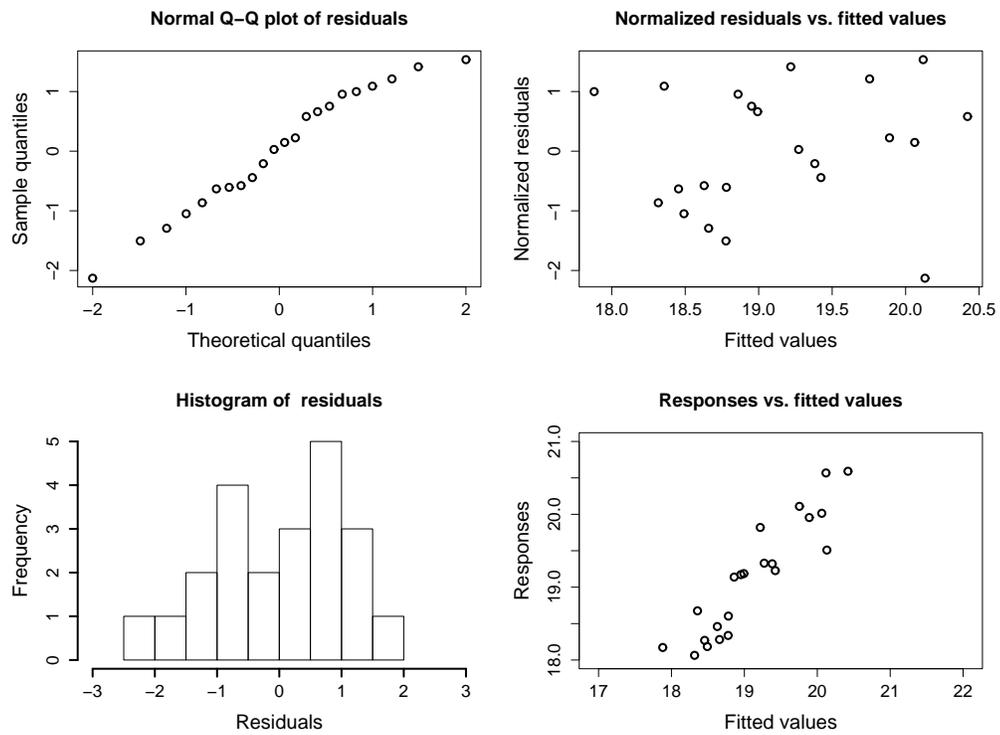
Model (4.19) for $t \leq n$. For formulas for computing 95% prediction intervals see Chapter 9 of Cryer and Chan (2008). For computing forecasts for 4 years or longer, we needed to compute out-of-sample forecasts for ATF abundance and that of 1-year old pollock, which required the development of joint modeling for these covariate processes with the recruitment. Note that $y_{1,t}$ in year 1999 was a missing value, and we used the nave scheme of averaging the $y_{1,t}$'s in 1998 and 2000 to impute this missing log age-1 abundance index. The fitted recruitment of age-4 pollock in 2002 was calculated based on the imputed age-1 abundance in 1999.

Recruitment estimates and out-of-sample (point) forecasts, i.e. conditional means of the future values, from Model (4.19) for 3 years are shown in Fig. 4.13. The dynamic trend shown by the fitted pollock recruitment levels (dashed curve) follows the observed recruitment trend (gray curve) in the study period, which indicates that the recruitment Model (4.19) fits the data well. The (log) recruitment forecasts of the age-4 from 2007 to 2009 are 18.5, 19.5 and 19.0, respectively, resulting in the forecast errors (forecasts minus observations) being 0.93, 0.67 and -0.68, respectively. The dotted curve in Fig. 4.13 plots the recruitment forecasts of the age-4 from 2007 to 2009, with the 95% pointwise prediction band shaded in gray; the point forecast for 2007 is just outside the 95% prediction interval (but still successfully predicting a weak year class), while the point forecasts for 2008 and 2009 are both inside their 95% prediction intervals.

The second fitted recruitment model uses lag-1 and lag-2 recruitment to account for the (extra) autocorrelation among the age-4 recruitment. The autocorrelation structure can be explained by the population dynamics described by Eqns. (4.12-4.15), with the greater flexibility in a nonlinear lag-2 effect. The threshold ATF predation effect and age-1 abundance effect are still found to be linear. The formulation of this fitted model is shown as follows:

Figure 4.13: Observed recruitment and predicted recruitment from Model (4.19). Numbers are abundance of 4 year old fish occurring in each year. Gray solid line: observed recruitment over the study period; dashed line: estimated recruitment; dotted line: out-of-sample recruitment forecasts; gray shaded area: 95% point-wise prediction band of the recruitment forecasts; ∘: recent observed recruitment in 2007 to 2009.
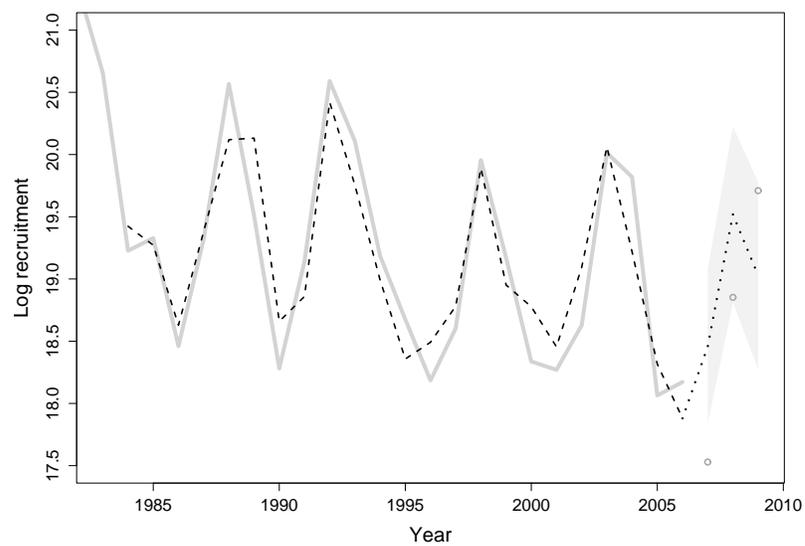
$$y_{4,t+3} = \beta_0 + \beta_1 a_t 1_{(t>t_c)} + \beta_2 y_{1,t} + \beta_3 y_{4,t+2} + s(y_{4,t+1}) + \varepsilon_{t+3}, \qquad (4.20)$$

where $\beta_3 y_{4,t+2}$ represents a linear lag-1 recruitment effect, and $s(y_{4,t+1})$ assesses the lag-2 recruitment effects nonparametrically. The term $\{\varepsilon_{t+3}\}$ is a sequence of uncorrelated error terms that are of zero mean and constant variance.

Model (4.20) considers that the strong ATF predation effects take place only after a threshold year. Before accepting the fit of Model (4.20), we determine whether the threshold effect is significant in the recruitment model. Using the likelihood ratio test, we find strong evidence for the threshold structure embodied in Model (4.20) with p-value being equal to 0.002. Additionally, according to the AIC values under different threshold choices, the estimated threshold year is 1991.

The estimation results of Model (4.20) are given in Table 4.8. The slope of the linear ATF predation effect after 1991 is -0.059 (with s.e. $1.14. \times 10^{-2}$), which indicates that higher density of ATF tends to reduce the recruitment of age-4 pollock. Meanwhile, the positive slope ($\hat{\beta}_2 = 0.179$ with s.e. 4.89×10^{-2}) of the linear age-1 abundance effect indicates that the higher age-1 abundance is associated with the higher recruitment level. The positive estimate $\hat{\beta}_3 = 0.252$ (with s.e. 0.104) suggests that the misclassification effect overwhelms the competition and/or cannibalism from the lag-1 recruitment. The decreasing pattern in the "Lag 2 recruitment effects" (Fig. 4.14) shows that the age-4 recruitment is negatively correlated with the lag-2 recruitment, which is consistent with the mechanism underlying Eqns. (4.12-4.15). The smooth function estimate $s(y_{4,t+1})$ has 1.64 degrees of freedom, so Model (4.20) has about 6.64 parameters, slightly more than that of Model (4.19), and its residuals have 15.36 degrees of freedom.

Through the model diagnostics, we find that the residuals of Model (4.20) marginally satisfied the Shapiro-Wilk normality test with p-value being equal to

Table 4.8: Estimates of Model (4.20) for predicting walleye pollock recruitment using 1991 as the threshold year

	Value	SE	t-value	p-value
β_0	15.1	2.02	7.49	<0.001
β_1	-0.0594	1.14×10^{-2}	-5.21	<0.001
β_2	0.179	4.89×10^{-2}	3.66	0.002
β_3	0.252	1.04×10^{-1}	2.43	0.028

Figure 4.14: Additive lag-2 recruitment effects of Model (4.20).
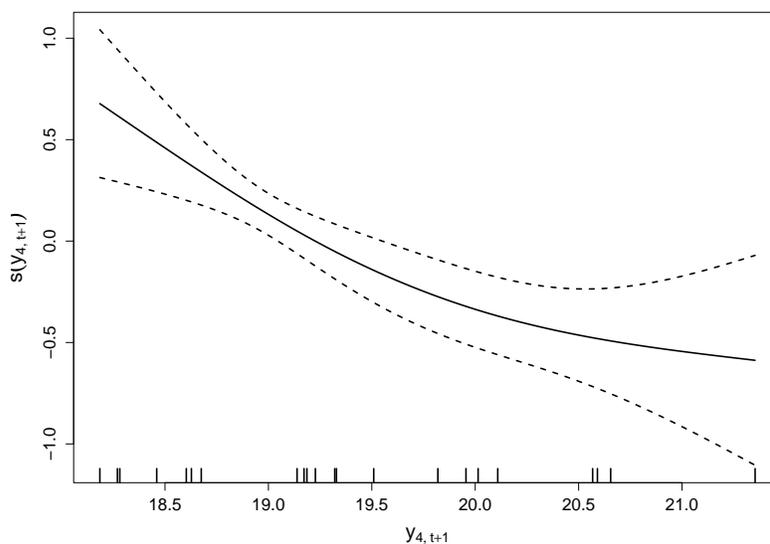
0.07, which was slightly higher than the 0.05 significance level. All the other assumptions of the error terms, such as the constant variance and independence, are approximately satisfied. Overall, the residual diagnostics of Model (4.20) are not as quite as good as those of Model (4.19), although they look similar to Fig. 4.12 so they are not shown here. Additionally, the out-of-sample forecasts from Model (4.20) (dotted curve in Fig. 4.15) are less accurate than those from Model (4.19)

as all of the forecasts fall outside the 95% prediction intervals (although the general predictions of weak or moderate year classes are accurate). The forecast errors (forecasts minus observations) in 2007, 2008 and 2009 equal 1.52, 0.95 and -0.71, respectively, which are larger in magnitude than their counterparts from Model (4.19). However, the within-sample recruitment estimation from Model (4.20) fits the observed recruitment data well (dashed curve in Fig. 4.15).

Figure 4.15: Observed recruitment and predicted recruitment from Model (4.20). Gray solid line: observed recruitment over the study period; dashed line: estimated recruitment; dotted line: out-of-sample recruitment forecasts; gray shaded area: 95% point-wise prediction band of the recruitment forecasts; ○: recent observed recruitment (of age-4 pollock) in 2007 to 2009.
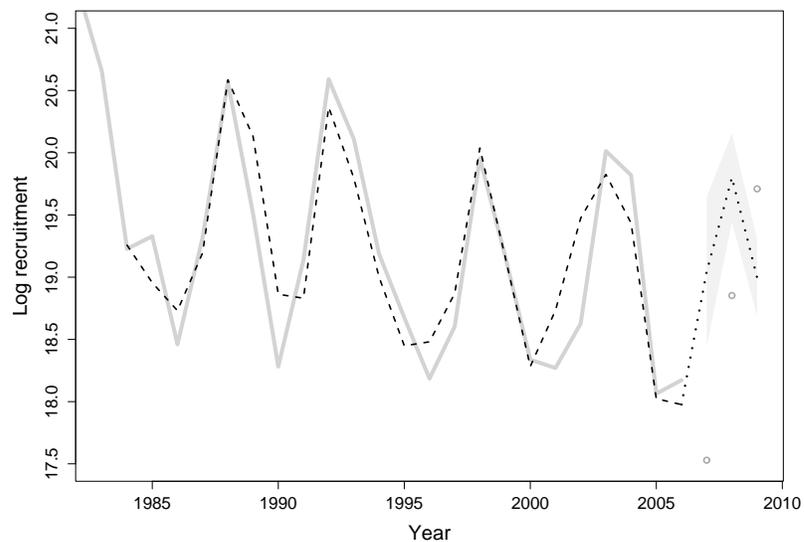
4.2.4 Discussion

The dynamics of marine fish populations respond to many factors, including commercial harvesting, climate change and shifts in community structure. One of the world's largest fisheries, walleye pollock, has experienced sequential and lasting

declines of populations in Puget Sound, Shelikof Strait, the Aleutian Basin, and other regions, most probably due to the combined effects of climate change (Ciannelli et al., 2005), shifts in community structure (Anderson and Piatt, 1999; Litzow and Ciannelli, 2007), and possibly harvesting activity. The Gulf of Alaska population is about 22% of its pre-fishing biomass. Under such circumstances, forecasting recruitment to the population is important as the commercial harvest takes increasingly younger members of the population. The traditional methods of forecasting marine fisheries recruitment from environmental conditions during egg and larval stages have been marginally successful due to the complexity of interacting biological and environmental conditions coupled with the high and variable mortality of these early life history stages (Bailey et al., 2005; Houde, 2008), although there are exceptions (Svendsen et al., 2007). Here we demonstrate some novel forecasting models that are fairly successful. In our forecasting models we start with 1 year-old juvenile pollock abundance in survey catches and examine factors that may influence their survival over the next 3 to 4 years until recruitment. We find that a change in the predator community favoring an increase in the abundance of ATF, (a voracious predator of juvenile pollock that has come to dominate the groundfish biomass in the Gulf of Alaska in the past decade; Turnock and Wilderbuer, 2007), a strong autocorrelation effect probably caused by inter year-class interactions, and a threshold effect in predation that is linked with a phase shift in environment factors are closely coupled with recruitment predictability. These findings represent a new approach in forecasting recruitment success of marine fisheries and further demonstrate the importance of predation during the juvenile stage influencing the dynamics of marine populations.

Recruitment of pollock became decoupled from larval mortality in the early 1990s and control shifted to juvenile survival (Bailey, 2000); a shift in community

structure also occurred around the same time that included a marked increase in the abundance of flatfishes (Anderson and Piatt, 1999), particularly ATF. In general, larval dynamics are thought to activate variability in year- class strength, whereas predation on juveniles is thought to dampen variability (van der Veer, 1986; Bailey et al., 2005). However, ATF has become an important force in the recruitment of pollock, as ATF is currently the dominant groundfish species and predator in the Gulf of Alaska ecosystem. ATF biomass dwarfs that of other potential predators such as cod, by about an order of magnitude (North Pacific Fisheries Management Council 2009). Generally about 40-50% of the diet of adult ATF comprises juvenile pollock although the exact composition and lengths of fish preyed upon depends on availability (Yang, 1993; Yang et al., 2006; Knoth and Foy, 2008). Shifts in community structure resulting in changes in abundance of top predators are recognized to have major cascading effects on lower trophic levels (Hunt et al., 2002; Frank et al., 2005). We perceive that if and/or when ATF abundance in the GOA declines, other predators may become important factors in juvenile survival, leading to a forecasting strategy that includes adaptable models.

A threshold effect on the importance of ATF predation on pollock seems to have occurred around the same time that control of pollock recruitment shifted to juvenile survival. In this sense inclusion of the threshold in the model is consonant with the known biology. We suggest the possibility that an increasing ATF population may have been conditioned over time to locate predictable hotspots on age-1 juvenile pollock, resulting in a phase transition, or their distribution otherwise expanded at that time to overlap more with juvenile pollock. Recent studies have shown that whereas age-0 pollock distributions are variable from year-to-year, age-1 (Wilson 2009) and older fish distributions (Shima et al., 2002) are relatively consistent. Alternatively an environmental phase shift occurring around the same

time (i.e. 1989; Hollowed et al., 2001) and potentially causing a shift in the over-lapping distributions of predator and prey, could be a factor (Ciannelli et al., 2005). An environmental shift that started around 1989 was associated with changes in a broad array of biological and climate factors, including enhanced summer warming in the coastal waters of the GOA (Hare and Mantua, 2000). Yet another alternative is that the acoustic gear changed in 1992, near our threshold year, and this gear change may have altered the log-transformed data by some additive constant after 1992, resulting in a jump in the intercept term; however this effect cannot produce the piecewise linear threshold effects described in Models (4.19) and (4.20).

Our initial modeling efforts included an interaction effect of temperature and ATF abundance on pollock recruitment. Although the final model dropped the temperature interaction, it is likely to be an important consideration in the overlap of predators and prey; in our case its precision may have been affected by using SST as a proxy for bottom temperature (BT). In the Bering Sea ATF avoid cold water (Spencer, 2008). In the GOA, there is some support that ATF tend to avoid cold water, for example in colder La Nia years they are found in warmer areas (Speckman et al., 2005). We suggest that in some years ATF may avoid cold water over the shelf, influencing their overlap with juvenile pollock prey. More studies, better understanding of temperature interactions and availability of bottom temperature data may indicate whether temperature effects should be included in future models.

The autocorrelation among the pollock recruits included in our models reflects the pollock's population dynamics, which is difficult to explain, but is biologically plausible. The sometimes cyclic nature of population dynamics is well-established (e.g. Kendall et al., 1999; Stenseth et al., 2003). Cyclic variations in recruitment of marine fishes can occur on many scales, from lunar (Meekan et al., 1993) to decadal and longer cycles (Southward et al., 1988; Ravier and Fromentin, 2001).

Marine fish recruitment cycles with a period approximating a generation time (4 to 5 years for pollock) may result from intra-population interactions (Bjornstad et al., 1999; Bailey et al., 2003). In the fitted models with stochastic error terms, the dynamic cycle may be due to the competition or cannibalism effects from older pollock groups (Bjornstad et al., 1999). Walleye pollock in the Bering Sea are highly cannibalistic and there is a strong seasonality in the process (Dwyer et al., 1987) with up to 50 to 90% of the diet of adults in autumn and winter comprised of juveniles, mainly age-0s. There is also a high degree of cannibalism on age-1 pollock in the eastern Hokkaido Island stock of pollock (Yamamura et al., 2001). In the GOA there is little published information on seasonal changes in the diet of pollock, but in summer around 10% of the diet of adults consisted of juvenile pollock (Yang, 1993; Yang et al., 2006). Competition between year classes and between adults and juveniles is also viable since a large component of the diet of both age-1 pollock and adults is comprised of copepods and euphausids. Competition between year classes is also thought to be important in recruitment of pollock off eastern Hokkaido Island (Shida et al., 2007). Autocorrelation in juvenile survival rates may also be linked to autocorrelation in environmental variables, such as zooplankton biomass in the GOA and/or competitor effects on prey (Brodeur et al., 1996; Shiomoto et al., 1997). Consequently, we believe that the autocorrelation in pollock dynamics is important to capture in forecasting models, and a better understanding of the phenomenon is needed.

The models we propose for pollock recruitment forecasting are novel in the sense that they are based on indices after the complex egg and larval period, starting with survey estimates of age-1 juveniles. The models account for changes in community structure, such as an increasing trend in the predatory capacity of the community, and for biological causes of the observed periodicity in recruitment. In

the hindcast mode, Model (4.19) in particular provides a very close fit to observed recruitment levels, and the forecasts for 4 year olds recruiting in 2007-09 appear to be relatively accurate. This model, which has a slightly better fit than Model (4.20) based on AIC and adjusted R^2 as well as model diagnostics, accounts for the effects of age-1 abundance, the threshold effect of ATF abundance and auto-correlated error terms due to as yet unidentified covariates on recruitment 3 years later. Furthermore, Model (4.19) provides a far superior fit (adjusted $R^2 = 81.5\%$) than the simple linear regression with the log age-1 abundance as the only covariate (adjusted $R^2 = 31.2\%$). Model (4.20) is more specific and provided insight to the possible missing covariates, including potential misclassifying ages of adult pollock (especially spillover effects of strong year classes) and predation/competition inter-actions among cohorts. There may be other co-variates involving stock structure and spawning behavior, or competition and cannibalism (and their representation) underlying the AR(2) error structure that are not presently recognized.

CHAPTER 5
AN R PACKAGE FOR FITTING PPGAMS

In this chapter, we illustrate the use of an R package **PPGAM** which contains some R functions that implement the new methodologies detailed in Chapter 2. The function `ppgam()` in the package implements the quasi-likelihood estimation of a PPGAM. The quasi-likelihood estimation is implemented by an iterative scheme that alternately updates the nonparametric functions and the parametric estimator of the nonlinear component of the mean function. With the parametric nonlinear component fixed, the additive nonparametric function estimates are updated by the method of penalized iteratively re-weighted least squares (P-IRLS) via the `gam()` function of the package **mgcv** (Wood, 2006). With the additive nonparametric functions fixed, the unknown parameter in the parametric nonlinear component of the mean function is updated by the method of weighted least square based on some linearization technique. The update of the nonlinear parametric estimates is obtained by solving the equation that sets zero to the corresponding first derivative of the penalized (log-)likelihood, as specified by Eqn. (2.6). Numerically, Eqn. (2.6) can be solved by an approximate weighted least squares problem asa follows. Given the current parametric estimate ($\tilde{\theta}^{\langle m-1\rangle}$), Eqn. (2.6), in vector form, can be approximated as follows:

$$\frac{\partial l_p}{\partial \theta} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\phi V^*(\mu_i)} \frac{\partial h_\theta(z_i)}{\partial \theta} \frac{1}{\dot{g}(\mu_i)}$$
$$\approx \sum_{i=1}^{n} \{y_i - \tilde{\mu}_i^{\langle m-1\rangle} - \frac{1}{\dot{g}(\tilde{\mu}_i^{\langle m-1\rangle})} \frac{\partial h_{\tilde{\theta}^{\langle m-1\rangle}}(z_i)}{\partial \theta^T}(\theta - \tilde{\theta}^{\langle m-1\rangle})\}/[\phi V^*(\tilde{\mu}_i^{\langle m-1\rangle})]$$
$$\frac{\partial h_{\tilde{\theta}^{\langle m-1\rangle}}(z_i)}{\partial \theta} \frac{1}{\dot{g}(\tilde{\mu}_i^{\langle m-1\rangle})}. \tag{5.1}$$

The approximation denoted by Eqn. (5.1) is obtained by replacing μ_i in Eqn. (2.6) with its first order Taylor approximation $\tilde{\mu}_i^{\langle m-1\rangle} + \frac{1}{\dot{g}(\tilde{\mu}_i^{\langle m-1\rangle})} \frac{\partial h_{\tilde{\theta}^{\langle m-1\rangle}}(z_i)}{\partial \theta^T}(\theta - \tilde{\theta}^{\langle m-1\rangle})$

given the estimates in the $\langle m-1 \rangle$th iteration. Thus, the first order condition in Eqn. (5.1) can be fitted by the method of weighted least square, i.e. regressing $y - \tilde{\mu}^{\langle m-1 \rangle} + \frac{1}{\dot{g}(\tilde{\mu}_i^{\langle m-1 \rangle})} \frac{\partial h_{\tilde{\theta}\langle m-1 \rangle}(z)}{\partial \theta^T} \tilde{\theta}^{\langle m-1 \rangle}$ on $\frac{1}{\dot{g}(\tilde{\mu}^{\langle m-1 \rangle})} \frac{\partial h_{\tilde{\theta}\langle m-1 \rangle}(z)}{\partial \theta^T}$ with weights $1/[\phi V^*(\tilde{\mu}^{\langle m-1 \rangle})]$.

The package **PPGAM** includes several R functions for (i) fitting a PPGAM, (ii) summarizing a fitted PPGAM, (iii) graphing the fitted functions and (iv) doing model diagnostics. We use two simulated data and one real analysis to illustrate the capabilities of the package **PPGAM**.

5.1 A Conditionally Poisson Model

As the first example, we simulated Poisson distributed counts. We use this example to illustrate how to (1) fit a PPGAM, (2) summarize and visualize the model fit and (3) do predictions given new covariate values. Counts were simulated from the Poisson distribution with a logarithmic link function so that

$$\log(\mu(z,x)) = s(z) + h_\theta(x), \tag{5.2}$$

$$s(z) = z^{11}(7(1-z))^6 + 7(7z)^3(1-z)^{10}, \tag{5.3}$$

$$h_\theta(x) = \alpha x^\gamma = 2x^{0.75} \tag{5.4}$$

where μ denotes the conditional mean response variable; $s(z)$ is a 1-dimensional smooth function to be estimated nonparametrically; $h_\theta(x)$ is a power function indexed by the parameter $\theta = (\alpha, \gamma)$; the covariates (z, x) are assumed to be uniformly distributed over $(0,1)^2$. N independent and identically distributed $(y_i, x_i, z_i), i = 1, \ldots, N$ were then simulated.

The simulated Poisson data are generated using the function `data.genpoisson1()`, and are saved in an R object named `data1`:

```
> data.genpoisson1 <- function(N) {
  x <- runif(N,0,1)
  z <- runif(N,0,1)
  s <- h <- mu <- y <- numeric(N)
```

```
    out.data <- matrix(0, ncol=3, nrow=N)

    s <- z^11*(7*(1-z))^6+7*(7*z)^3*(1-z)^10

    h <- 2*x^0.75

    mu <- exp(s+h)

    y <- rpois(N,mu)

    out.data[,1] <- y

    out.data[,2] <- x

    out.data[,3] <- z

    out.data

}

> set.seed(100)

> N.samp <- 1000

> data1 <- data.genpoisson1(N.samp)

> colnames(data1) <- c("y", "x", "z")

> data1 <- data.frame(data1)
```

We can fit a PPGAM using the `ppgam()` function. The model formula is split into two parts, namely the nonparametric additive part and the parametric (nonlinear) part. The nonparmetric part is specified via the `formula.gam` argument by supplying a formula using the same convention used by the R package **mgcv**. For example, for this Poisson example, the model formula may be written as `y~s(z)+alpha*x^gamma`. Its nonparametric part is supplied to the `ppgam()` function as `formula.gam=y~s(z)` whereas the parametric part is specified via the formula.nl argument as `formula.nl=h~alpha*x^gamma`. Here the symbol `h` is reserved for denoting the parametric nonlinear part. Which variables stand for the unknown parameters in the parametric part (and their initial values) have to be supplied to the `ppgam()` function via the `start` argument. For example, in the

Poisson example, `alpha` and `gamma` are parameters. We can supply their initial values as follows: `start=c(alpha=1,gamma=0.3)`. Here their initial values are set to be 1 and 0.3, respectively. Another approach for specifying the initial values is to do a grid search over a supplied range; this option will be illustrated in second example below. (Further options available for the `ppgam()` function are described later.) Below is the R code for fitting the above PPGAM to the simulated Poisson data, with the fitted model output saved in the R object "out.poisson1".

```
> out.poisson1 <- ppgam(formula.gam = y~s(z),
+ formula.nl = h~alpha*x^gamma, start=c(alpha=1, gamma=0.3),
+ family="poisson", data=data1)
```

The argument `family` specifies the distribution and its corresponding link function to use in the modeling; the default distribution is Gaussian with identity link. Other available distributions include binomial and gamma. The specification rules for the `family` argument of the `ppgam()` is similar to those of the `glm()` for fitting a generalized linear model (GLM). The dataset containing the response variable and covariates are supplied via the `data` argument.

The `ppgam()` function returns a list containing the estimation results as well as some information concerning the convergence of the iterative estimation scheme. Applying the function `names()` to a fitted PPGAM object would display a list of elements in the object. A direct way to retrieve useful fitting information is to list the corresponding element in the fitted PPGAM object. However, it is simpler to summarize the fitted model via the function `summary.ppgam()`:

```
> summary.ppgam(out.poisson1)
```

```
Family: poisson Link function: log
```

```
Formula: y ~ s(z) + alpha * x^gamma


Parametric coefficients:

            Estimate Std. Error z value Pr(>|z|)

(Intercept)  1.14666    0.07222   15.88   <2e-16 ***

alpha        1.97179    0.06708   29.39   <2e-16 ***

gamma        0.76885    0.05272   14.58   <2e-16 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Approximate significance of smooth terms:

      edf Est.rank Chi.sq p-value

s(z) 8.672        9   4047   <2e-16 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Scale est. = 1          n = 1000


Log marginal likelihood: -2532.128,   AIC: 5007.989
```

The summary includes information on the distribution family and displays to the right of "Formula" the conditional mean structure on the link scale. Note that the model automatically includes an intercept term. If an intercept term is not needed, this should be specified in the option `formula.gam`, which, for this example, can be done by specifying `formula.gam=y~s(z)-1`; the `-1` in the formula instructs that an intercept term is not needed. The above summary displays both

the intercept and other parameters indexing the parametric terms as well as a table listing the estimates, their standard errors and p-vlaues of the Wald tests for testing the null hypotheses that each individual parameter is zero. For the nonparametric function estimates (smooth terms), their effective degrees of freedom (edf) and significance test results are presented in a table. Furthermore, `Scale est.` lists the estimated scale parameter or its true value when the scale parameter is known. For example, the scale parameter of the Poisson distribution equals 1, if over-dispersion is not allowed. Finally, values of two model selection criteria, (approximate) log marginal likelihood and Akaikes information criterion (AIC), are reported in the summary of the fitted PPGAM. AIC of a fitted PPGAM equals `2(degree of freedom - log likelihood)`, where the degree of freedom is the sum of the number of the estimated parameters and the effective degrees of freedom of the nonparametric function estimates. Generally, the fitted PPGAM with the highest log marginal likelihood (lowest AIC) value provides the best fit to the data.

The smooth function estimate of $s(z)$ can be plotted by the function `plot.ppgam()`.

```
> plot.ppgam(out.poisson1, residuals=T)
```

For the Poisson example, the nonparametric part contains one 1-dimensional smooth function. The solid black curve in Fig. 5.1 displays the function estimate and the dashed black curves encompass its (individual) 95% confidence band. Setting the option `residuals=T`, partial residuals (black dots) are added to the plot. The partial residuals for a smooth function are obtained by adding the function estimate to the working residuals. The partial residual plot graphically describes the relationship between a particular covariate that is the argument of the smooth function under study and the response variable with other covariates kept fixed. The partial residuals are helpful for visualizing the validity of the nonparametric fit, the presence of outliers and/or influential observations. Note that the estimated smooth

Figure 5.1: Fit of the smooth function for the Poisson example.
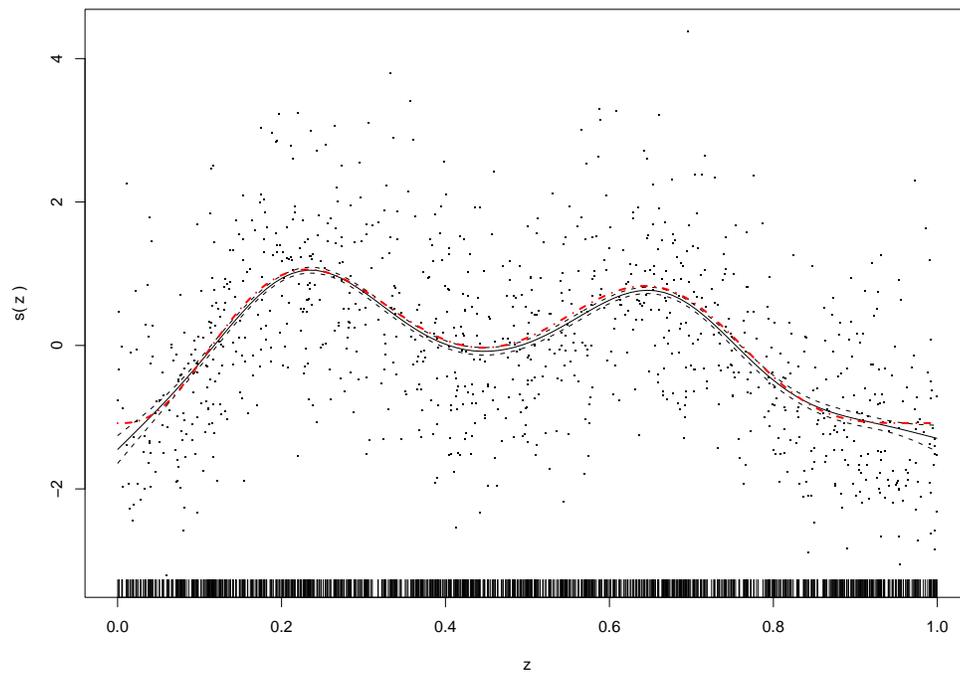
function is generally centered to have zero mean over the data in order to ensure model identification (Wood, 2006, p. 134). The ticks at the bottom of the plot mark the observed values of the covariate, which is the default option `Rug=TRUE` in the function `ppgam()`. (Setting `Rug=FALSE` will not add the ticks.) For comparison, we superimpose in Fig. 5.1 the true smooth function (red dash-dot curve), which has been adjusted by subtracting from it the intercept estimate stored in `out.poisson1`. Note the solid black curve tracks closely the dash-dot red curve, hence graphically illustrating the good PPGAM fit to the simulated Poisson data.

```
> N.p <- 200
> z.p <- seq(0,1,length.out=N.p)
> s.p <- z.p^11*(7*(1-z.p))^6+7*(7*z.p)^3*(1-z.p)^10
> s.p.adj <- s.p - out.poisson1$coefficient[1]
> lines(s.p.adj~z.p, ylab="s(z)", xlab="z", ylim=c(-3,4), lty=3,
+ col="red", lwd=2)
```

The function `predict.ppgam()` computes the point predictors of the future responses (and their prediction standard deviations) given a new set of covariates. Setting the option `type="response"`, predictions on the scale of the response variable are returned by the `predict.ppgam()` function. The option `type="link"` results in the predictions on the link scale. When `type="terms"`, the predictions, component by component on the link scale, are listed separately, including any linear parametric component and smooth function components specified in the formula.gam option, as well as the parametric nonlinear part. With the option `se.fit=TRUE`, the prediction function also returns the prediction standard deviations.

Based on the fitted object `out.poisson1`, the function `predict.ppgam()` is called to make predictions with new sets of covariate values in the dataset `newdata1`. In the new dataset for prediction, only the values of the covariates in the fitted model

are required, but the names of the covariates need to match exactly those in the fitted model.

```
> newdata1 <- data.frame(x=c(data1$x[1], 0.2), z=c(data1$z[1], 0.6))
> pred.poisson1 <- predict.ppgam(out.poisson1, newdata=newdata1,
+ type="response", se.fit=TRUE)
> pred.poisson1

      fit       se
1  4.036799 0.1498030 2 10.714724 0.2812875
```

Note that, in this example, `newdata1` contains the first data case of the simulated Poisson data, `data1`, so the corresponding point predictor should be same as the first fitted value in the fitted object `out.poisson1`, which is, indeed, the case as shown below.

```
> out.poisson1$fitted.values[1]

[1] 4.036799
```

5.2   A Conditionally Gaussian Model

In this section, we illustrate the capabilities of the package **PPGAM** regarding (1) an estimation approach using the initial values from a grid search, (2) visualization of an estimated two-dimensional smooth function, (3) residuals calculation and model diagnostics, (4) model selection among several potential models and (5) tracking the convergence behavior of the nonlinear parameter estimation. The example in this section is based on a simulated data set with conditionally Gaussian response, whose conditional mean is the sum of the following two additive components:

$$s(z_1, z_2) = 0.8 \times 0.6\pi(1.2e^{-(z_1-0.2)^2/0.6^2-(z_2-0.3)^2} + 0.8e^{-(z_1-0.7)^2/0.4^2-(z_2-0.8)^2/0.6^2}),$$
$$(5.5)$$

$$h(x) = -\log(1 + \alpha x^\gamma) = -\log(1 + 2x^{0.7}). \tag{5.6}$$

The `identity` link is used for the Gaussian data generation, so the conditional mean $\mu(z, x) = s(z) + h(x)$, where $z = (z_1, z_2)$ and $s(z) = s(z_1, z_2)$ is a 2-dimensional smooth function. (The covariates (Z_1, Z_2) are assumed to be uniformly distributed over $[0, 1]^2$, and X uniformly distributed over $[0, 10]$.) Independent data were simulated using the function, `data.gengaus1()`:

```
> data.gengaus1 <- function(N) {
  x <- runif(N,0,10)
  z1 <- runif(N,0,1)
  z2 <- runif(N,0,1)
  s <- h <- mu <- y <- numeric(N)
  out.data <- matrix(0, ncol=4, nrow=N)
  #sigma <- 0.1
  sigma <- 0.3
  s <- 0.8*0.6*pi*(1.2*exp(-(z1-0.2)^2/0.6^2-(z2-0.3)^2)+
    0.8*exp(-(z1-0.7)^2/0.4^2-(z2-0.8)^2/0.6^2))
  h <- -log(1+2*x^0.7)
  mu <- s+h
  y <- rnorm(N,mu,sigma)
  out.data[,1] <- y
  out.data[,2] <- x
  out.data[,3] <- z1
  out.data[,4] <- z2
  out.data
}
```

The simulated data contains 500 data cases, and are saved in `data2`.

```
> set.seed(200)
```

```
> N.samp <- 500

> data2 <- data.gengaus1(N.samp)

> colnames(data2) <- c("y", "x", "z1", "z2")

> data2 <- data.frame(data2)
```

The function `ppgam()` is called to fit the simulated Gaussian data, and the estimation results are saved in the object `out.gaus1`. The nonparametric part of the model is specified by setting the argument `formula.gam = y~s(z1,z2)` while the parametric nonlinear part is given by setting `formula.nl=h~-log(1+alpha*x^gamma)` in the function call. Instead of setting fixed initial values for the parameters, we illustrate a second option to get the initial values for the nonlinear parameters. The function `ppgam()` can compute initial values for the nonlinear parameters using a grid search. The user then has to supply the lower bound and upper bound for each parameter to do the grid search for more refined initial values. Specifically, the lower and upper end points for each parameter have to be supplied as a vector via the `lower` and `upper` arguments, respectively. Also, the user has to specify the number of segments into which each interval is to be equally partitioned via the `n.grid` argument. The function `ppgam()` then fits a GAM with the nonlinear parameters fixed at each set of values prescribed by the grid, and picks the set of initial nonlinear parameter values for which the model has the smallest GCV. The following R codes specify that the initial values are to be determined by a grid search over the rectangle $(0.0001, 5) \times (0.0001, 2)$ with a grid that divides each side of the rectangle into three equal sub-intervals, so altogether 16 evenly spaced grid points are used in the grid search for the initial values. (For the simulated Gaussian example, the grid search yields the more refined initial values of (α, γ) that equal (`3.3333`,`0.6667`), which are saved in the `nlpar.ini` object of the list returned by the function `ppgam()`.) Furthermore, with the option `family="gaussian"`, the model is fitted under the

assumption that the response variable conditionally follows the normal distribution with default identity link.

```
> out.gaus1 <- ppgam(formula.gam = y~s(z1,z2),
+ formula.nl = h~-log(1+alpha*x^gamma),
+ lower=c(alpha=0.0001,gamma=0.0001), upper=c(alpha=5,gamma=2),
+ n.grid=c(3,3), family="gaussian", data=data2)
```

For the simulated Gaussian example, the main estimation results can be summarized by calling the function `summary.ppgam()`. The lay-out of the summary from `summary.ppgam(out.gaus1)` is similar to that from `summary.ppgam(out.poisson1)` for the simulated Poisson example.

```
Family: gaussian Link function: identity


Formula: y ~ s(z1, z2) + -log(1 + alpha * x^gamma)


Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.63278    0.16941   9.638  < 2e-16 ***
alpha        1.88302    0.50485   3.730 0.000215 ***
gamma        0.70885    0.04689  15.117  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1


Approximate significance of smooth terms:
           edf Est.rank     F p-value
s(z1,z2) 16.21       29 21.79  <2e-16 ***
---
```

```
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
Scale est. = 0.095195  n = 500
```

```
Log marginal likelihood: -165.0187,  AIC: 266.2358
```

The conditional mean function of the simulated Gaussian example contains a 2-dimensional smooth function $s(z_1, z_2)$; the corresponding smooth fit can be described by either a contour plot or a perspective plot. Before illustrating the estimated 2-dimensional smooth function for the Gaussian example, we draw the contour plot of the "centered" true smooth function as defined by (5.5) (the left diagram in Fig. 5.2) for comparison; the centering was done by subtracting the intercept estimate from the true function. Recall that for model identifiability each smooth functions is centered. For the simulated Gaussian example, the estimated intercept is approximately equal to the estimated mean of the only smooth function, $s(z_1, z_2)$. Hence, we adjust the true smooth function by subtracting the intercept estimate from it to ensure that it is comparable to the estimated smooth function from the PPGAM fit. The right diagram in Fig. 5.2 depicts the estimated smooth function, which is generated by the following R command.

```
> plot.ppgam(out.gaus1, xlab=expression(z[1]), ylab=expression(z[2]))
```

When applying the function `plot.ppgam()` to draw a 2-dimensional smooth function estimate, the default is a contour plot, i.e. the default option is `plot.2d="contour"`. With the default option `Rug=TRUE`, the covariate values are superimposed on the diagram as dots in a contour plot, as shown by the right diagram in Fig. 5.2. Overall, these contour plots display a close fit of the adjusted true smooth function by the estimated smooth function (Fig. 5.2). In particular, we find that the PPGAM estimation successfully describes the shape of the true smooth function for the Gaussian

example.

Figure 5.2: True smooth function and its estimate from the PPGAM fit to the simulated Gaussian data set. The left panel plots the "centered" true smooth function, and the right panel depicts the estimated function. The dots on the right panel are the observed values of (z_1, z_2).
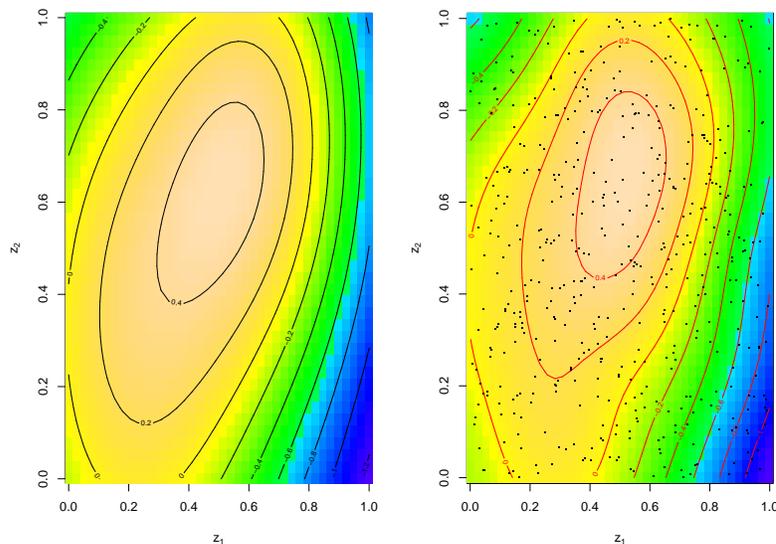

With the option `plot.2d = "persp"`, the function `plot.ppgam()` can generate a perspective plot. The perspective plot of the smooth function estimate in the Gaussian example (Fig. 5.3) is drawn by the following R command.

```
> plot.ppgam(out.gaus1, plot.2d = "persp",
+ xlab=expression(z[1]), ylab=expression(z[2]))
```

The **PPGAM** package provides a quick model diagnostic approach based on some basic residual plots. Various types of residuals in the `ppgam` fit can be retrieved by calling the function `residuals.ppgam()` with different `type` options, which include "response", "Pearson", "scaled.Pearson", "working" and "deviance". The response residual is the raw residual on the scale of the response variable; the scaled

Figure 5.3: Perspective plot of the estimated smooth function for the Gaussian example.
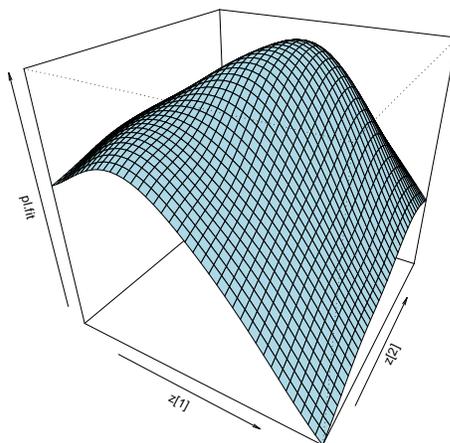
Pearson residual is the raw residuals scaled by the standard deviation of the corresponding datum; the Pearson residual equals the scaled Pearson residual multiplied by the square root of the scale parameter; the working residual is the residual on the link scale; and the deviance residual is based on the deviance contributed by each datum. The convention for the `type` argument is similar to that adopted by the **mgcv** package for a GAM fit. The first 5 scaled Pearson residuals of the PPGAM fit for the Gaussian example are listed by the following R code.

```
> resgaus1.std <- residuals.ppgam(out.gaus1, type="scaled.pearson")
> resgaus1.std[1:5]

        1         2         3         4         5
0.4910554 1.2863246 1.2417204 0.2575857 0.5180664
```

Model diagnostics can be carried out by the function `ppgam.check()`. The model diagnostics are displayed in a panel of four figures, including the normal QQ

plot, the histogram of the residuals, the scatter plot of the residuals against the fitted values, and the scatter plot of the observed response values against the fitted values. The diagnostic checks of the `ppgam` fit for the Gaussian example as saved in the object `out.gaus1` are performed by the R command reproduced below, and the diagnostic results are shown in Fig. 5.4. (The default residuals are the deviance residuals, which are also the raw residuals for normal data.) As shown by Fig. 5.4, the residuals appear to be uncorrelated and normally distributed, suggesting that the fitted PPGAM provides a good fit to the data.

```
> ppgam.check(out.gaus1)
```

Figure 5.4: Diagnostic checks of the PPGAM fit to the simulated Gaussian data.
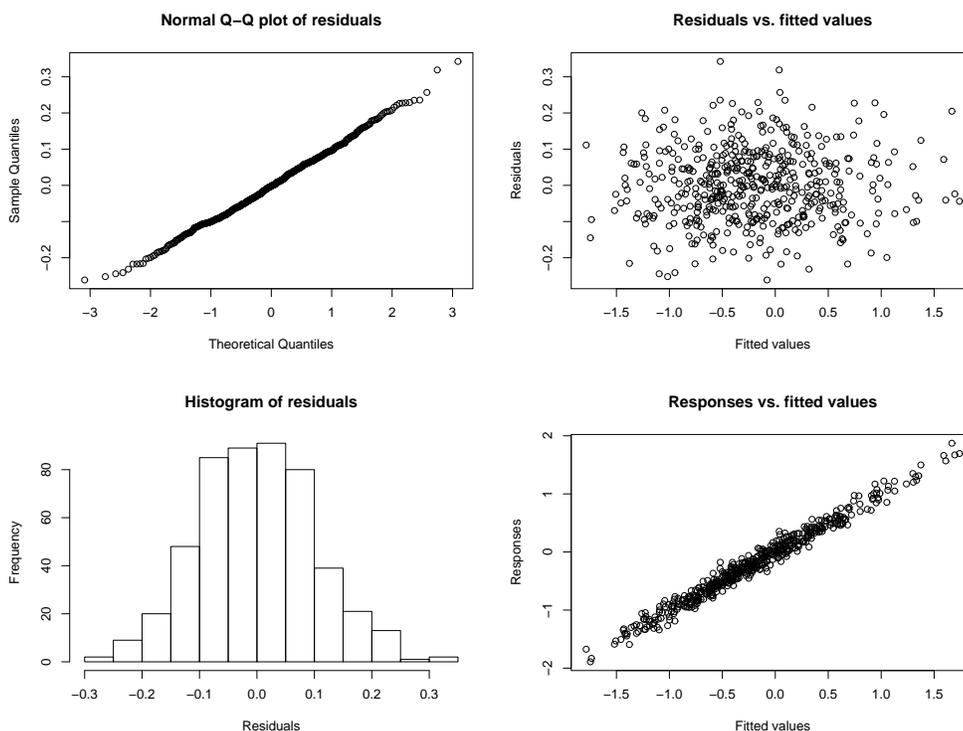
Two model selection criteria, log marginal likelihood and AIC, are provided by the **PPGAM** package. We use an example to show the usage of the model selection

criteria. Based on the simulated Gaussian data, a second model are fitted with the smooth function $s(z_1, z_2)$ now approximated as a sum of two additive functions, $s(z_1) + s(z_2)$.

```
> out.gaus2 <- ppgam(formula.gam = y~s(z1)+s(z2),
+formula.nl = h~-log(1+alpha*x^gamma),
+lower=c(alpha=0.0001,gamma=0.0001), upper=c(alpha=5,gamma=2),
+n.grid=c(3,3), family="gaussian", data=data2)
```

The values of the model selection criteria for the above model fit can be displayed using the `summary.ppgam()` function, or they can be retrieved directly from the fitted object `out.gaus2` using the following R code. The first element is the log marginal likelihood and the second is the AIC value of the model fit using $s(z_1) + s(z_2)$.

```
> print(c(out.gaus2$lmarglik, out.gaus2$aic))

[1] -188.9215  345.0711
```

In comparison, the log marginal likelihood and the AIC of the former model which allows for interactions in z_1 and z_2, equal

```
Log marginal likelihood: -165.0187,  AIC: 266.2358,
```

hence modeling the nonparametric part of the mean function additively in terms of $s(z_1) + s(z_2)$ results in a lower log marginal likelihood (-188.92) and a higher AIC (345.07) than the model with $s(z_1, z_2)$ modeling the possibly interactive covariate effects of z_1 and z_2 (log marginal likelihood $= -165.02$, AIC $= 266.24$). Thus, among the models containing the true parametric nonlinear component, the nonparametric term $s(z_1, z_2)$ results in a better fit to the simulated data than the term $s(z_1) + s(z_2)$, which is expected since the true nonparametric term is of the form $s(z_1, z_2)$.

The `ppgam()` function returns a list one of its object is named nlpar.path which is a vector obtained by concatenating the parameter vector iterates. Recall the estimation of a PPGAM is done iteratively by alternately updating the smooth function estimates and the nonlinear parametric estimates. The following R code illustrates how to extract the information saved in `out.gaus1$nlpar.path`, so that we can examine the convergence paths (Fig. 5.5) of the two nonlinear parameters for the Gaussian example. Fig. 5.5 shows that the algorithm converged successfully.

```
> nlpar1.path <- nlpar2.path <- numeric()

> n.rec <- length(out.gaus1$nlpar.path)/out.gaus1$n.nlpar

> for (i in 1:n.rec) {

+ nlpar1.path[i] <- out.gaus1$nlpar.path[(i-1)*out.gaus1$n.nlpar+1]

+ nlpar2.path[i] <- out.gaus1$nlpar.path[(i-1)*out.gaus1$n.nlpar+2]

+ }

> par(mfrow=c(2,1))

> plot(nlpar1.path, ylim=c(1.5,4),

+ xlab="Number of iteration", ylab=expression(alpha))

> plot(nlpar2.path, ylim=c(0.6,0.75),

+ xlab="Number of iteration", ylab=expression(gamma))
```

5.3 Revisit the Dose-Response Study on the Leukemia Risk Due to Radiation from the Atomic Bomb

In Section 2.5, we fit a PPGAM to assess the leukemia risk of the atomic bomb survivors in Hiroshima and Nagasaki from 1950 to 1982. The data for this study were obtained from the Atomic Bomb Casualty Commission (ABCC); see reference `http://www.rerf.jp/index_e.html`. The response variable (`death.dose01`) is a two-column matrix with the first column containing the numbers of death

Figure 5.5: Convergence paths of the two parameters for the Gaussian example.
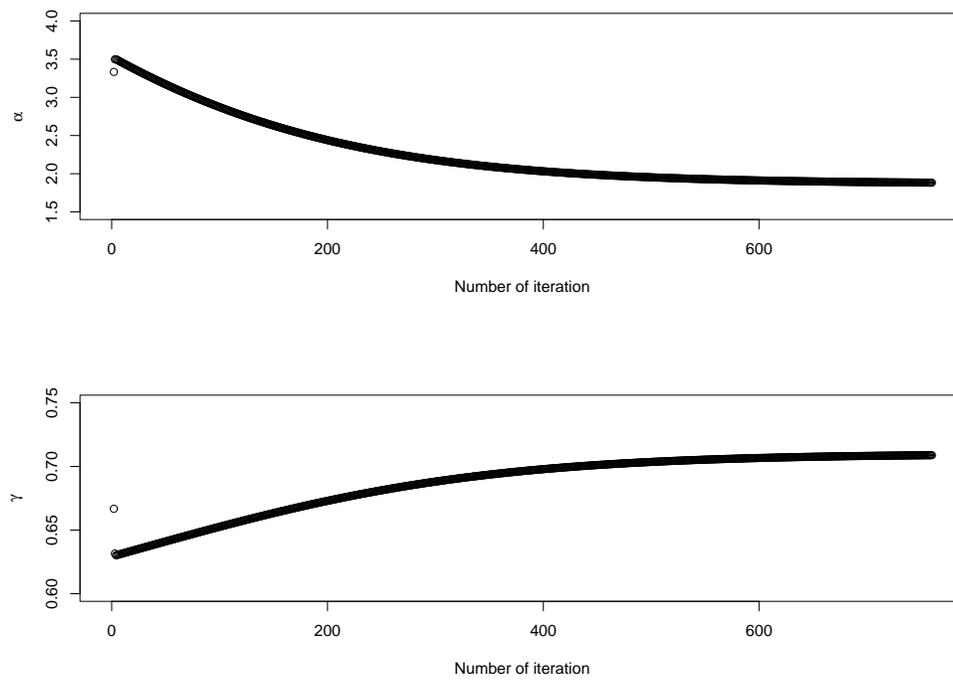
from leukemia and the second containing those from disease other than leukemia for various age groups. The covariates in the study include the radiation levels (`Dose.ds86true.adj101`) the survivors were exposed to, city where they were in at the time of explosion (`City.dummy01`) and two characteristics of the survivors, namely sex (`Gender01`) and age at the time of the atomic bomb explosion (`Age.true01`). The following R codes loaded the dataset into R.

```
> dose.abomb01<- read.delim("dose-abomb01.dat", header=T)

> dose.abomb01 <- data.frame(dose.abomb01)

> attach(dose.abomb01)

> death.dose01 <- cbind(death.leuk01, death.dexcl01)

> dose.abomb01$death.dose01 <- death.dose01
```

The model is formulated for studying how the propensity of the atomic bomb survivors to die from the leukemia was affected by their level of exposure to radiation from the atomic bomb. The model accounts for effects of two confounding factors, namely the cities where the survivors lived and their ages at the explosion, with the age effects modeled nonparametrically and the dose-response effects modeled by a power function. The dose-response relationship is as specified by Eqn. (2.24). To relax the constraint due to the positive restriction on the parameter α indexing the power dosage function, the dosage function is re-parameterized as specified by (2.25). The re-parameterized model can be fitted using the following R command:

```
> dose.ab13ppgam <- ppgam(formula.gam = death.dose01 ~

+ City.dummy01 + s(Age.true01),

+ formula.nl = h~exp(beta+gamma*log(Dose.ds86true.adj101)),

+ start = c(beta=-1.5, gamma=0.31), family="binomial")
```

where the argument `formula.gam = death.dose01~City.dummy01+s(Age.true01)` specifies the linear and nonparametric terms in Model (2.25); and the argument

`formula.nl = h~exp(beta+gamma*log(Dose.ds86true.adj101))` specifies the parametric nonlinear dosage function, which is also the last term in the mean function stated in (2.25).

The estimation results of Model (2.25) is summarized as follows:

```
> summary.ppgam(dose.ab13ppgam)
```

```
Family: binomial Link function: logit
```

```
Formula: death.dose01 ~ City.dummy01 + s(Age.true01) + exp(beta +
gamma * log(Dose.ds86true.adj101))
```

```
Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -4.67965    0.13776  -33.970  < 2e-16  ***
City.dummy01  -0.44972    0.16275   -2.763  0.00646  **
beta          -2.23675    0.53982   -4.144  5.75e-05 ***
gamma          0.54807    0.09015    6.079  9.89e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
              edf Est.rank    F p-value
s(Age.true01) 4.843       9 24.08  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Scale est. = 1          n = 158
```

```
Log marginal likelihood: -206.8401 ,  AIC: 400.3082
```

From the PPGAM fit, $\hat{\beta} = -2.24(0.54)$ and $\hat{\gamma} = 0.55(0.09)$, which demonstrate a significant positive dosage effect on the risk of dying from leukemia for the atomic bomb survivors. Additionally, the estimated age effects are depicted by Fig. 2.3a, using the following R code:

```
> plot.ppgam(dose.ab13ppgam, xlab="Age at bomb (k)", ylab="s(k)")
```

CHAPTER 6
CONCLUSION AND SOME DIRECTIONS OF FUTURE WORK

The PPGAM extends the generalized nonlinear regression model by allowing the addtive confounding covariate effects to be modeled nonparametrically. On the other hand the PPGAM can be regarded as a generalization of GAM by incorporating a parametric nonlinear term in the mean function. Like the GAM, the PPGAM allows both discrete and continuous responses from an exponential family, and the estimation can be performed using the penalized likelihood scheme. We derived some asymptotic properties of the penalized likelihood estimator, including consistency and asymptotic normality for the nonlinear parametric estimator. The derived asymptotic results are valuable for statistical and scientific inference. The proposed model selection criteria (log marginal likelihood and AIC) are shown to be effective in the simulations and two real case studies.

The application of the PPGAM for the hatchdate analysis of the pollock larvae data reveals valuable information about the pollock's spawning, hatching pattern and their larvae's survival process, which are important for understanding the pollock's population dynamics in their early stage. The dose-response example illustrates the use of the PPGAM for non-normal responses.

In the most general form, the partly parametric nonlinear regression model entertains interaction between the parametric nonlinear component and the nonparametric components. Thus the partly parametric generalized nonlinear regression (PPGNR) model provides more flexibility than its simpler additive version, PPGAM, but also brings more challenges in the model estimation and its theoretical properties, for example, deriving conditions for the model identifiability. Considering the complexity of the nonlinear functions, it is generally difficult to develop a

simple and universal constraint for model identifiability. The further development of the partly parametric nonlinear regression model and its associated problems are some interesting future research directions.

REFERENCES

Abrams, M. D., van de Gevel, S., Dodson, R. C., and Copenheaver, C. A. (2000), "The dendroecology and climatic impacts for old-growth white pine and hemlock on the extreme slopes of the Berkshire Hills, Massachusetts, U.S.A," *Can. J. Bot.*, 78, 851–861.

Anderson, P. and Piatt, J. F. (1999), "Community reorganization in the Gulf of Alaska following ocean climate regime shift," *Mar Ecol Prog Ser*, 189, 117–123.

Axentrot, T. and Hansson, S. (2003), "Predicting herring recruitment from young-of-the-year densities, spawning stock biomass, and climate," *Limnol Oceanogr*, 48, 1716–1720.

Bailey, K. M. (2000), "Shifting control of recruitment of walleye pollock (*Theragra chalcogramma*) after a major climate and ecosystem change," *Mar Ecol Prog Ser*, 198, 215–224.

Bailey, K. M., Ciannelli, L., and Agostini, V. (2003), "Complexity and constraints combined in simple models of recruitment," in *The big fish bang. Proceedings of the 26th Annual Larval Fish Conference*, eds. Browman, H. I. and Skiftesvik, A. B., Institute of Marine Research, Bergen, Norway, ISBN 82-7461-059-8, pp. 293–301.

Bailey, K. M., Ciannelli, L., Bond, N., Belgrano, A., and Stenseth, N. C. (2005), "Recruitment of walleye pollock in a complex physical and biological ecosystem," *Prog Oceanogr*, 67, 24–42.

Bailey, K. M., Francis, R. C., and Mais, K. (1986), "Evaluating incidental catches of 0-age Pacific hake to forecast recruitment," *CalCOFI Repts*, 27, 109–112.

Bailey, K. M. and Macklin, S. A. (1994), "Analysis of patterns in larval walleye pollock Theragra chalcogramma survival and wind mixing events in Shelikof Strait, Gulf of Alaska," *Mar. Ecol. Prog. Ser.*, 113, 1–12.

Bailey, K. M., Picquelle, S. J., and Spring, S. M. (1996), "Mortality of larval walleye pollock Theragra chalcogramma in the western Gulf of Alaska," *Fisheries Oceanography*, 5, 124–136.

Bailey, K. M. and Spring, S. (1992), "Comparison of larval, age-0 juvenile and age-2 recruit abundance indices of walleye pollock *Theragra chalcogramma* in the western Gulf of Alaska," *ICES J Mar Sci*, 49, 297–304.

Bates, D. M. and Watts, D. G. (1988), *Nonlinear Regression Analysis and Its Applications*, New York: Wiley.

Bjornstad, O. N., Fromentin, J.-M., Stenseth, N. C., and Gjosetter, J. (1999), "Cycles and trends in cod populations," *Proc Nat Acad Sci*, 96, 5066–5071.

Bond, N. A., Overland, J. E., Spillane, M., and Stabeno, P. (2003), "Recent shifts in the state of the North Pacific," *Geophys Res Letters*, 30(23),2183, 1–4.

Box, G. and Tiao, G. C. (1975), "Intervention analysis with applications to economic and environmental problems," *J. Am. Stat. Assoc.*, 70, 70–79.

Bradford, M. J. (1992), "Precision of recruitment predictions from early life stages of marine fishes," *Fish Bull US*, 90, 439–453.

Brodeur, R. D., Frost, B. W., Hare, S. R., Francis, R. C., and Ingraham, W. J. (1996), "Interannual variations in zooplankton biomass in the Gulf of Alaska, and covariation with California Current zooplankton biomass," *CalCOFI Rep*, 37, 80–98.

Chan, K. S. (2003), "Modeling pulse disturbance impact on cod population dynamics: the 1988 algal bloom of Skagerrak, Norway," *Ecol. ogical Mon.ographs*, 73, 151–171.

Ciannelli, L., Bailey, K. M., Chan, K. S., Belgrano, A., and Stenseth, N. C. (2005), "Climate change causing phase transitions of walleye pollock (Theragra chalcogramma) recruitment dynamics," *Proc. Biol. Sci.*, 272, 1735–1743.

Ciannelli, L., Bailey, K. M., Chan, K. S., and Stenseth, N. C. (2007), "Phenological and geographical patterns of walleye pollock spawning in the western Gulf of Alaska," *Can J Fish Aquat Sci*, 64, 713–722.

Craven, P. and Wahba, G. (1979), "Smoothing noisy data wih spline functions," *Numerische Mathematik*, 31, 377–403.

Cryer, J. D. and Chan, K. S. (2008), *Time Series Analysis with Applications in R*, New York: Springer Verlag.

Dorn, M., Aydin, K., Barbeaux, S., Guttormsen, M., Megrey, B., and Spalinger, K. (2008a), "Stock Assessment and Fishery Evaluation Report: Gulf of Alaska Walleye Pollock." *Can. J. Fish. Aquat. Sci.*, 64, 713–722.

Dorn, M., Aydin, K., Barbeaux, S., Guttormsen, M., Megrey, B., Spalinger, K., and Wilkins, M. (2006), "Stock assessment and fishery evaluation report: assessment of walleye pollock in the Gulf of Alaska," `http://www.afsc.noaa.gov/REFM/docs/2006/GOApollock.pdf`.

— (2007), "Stock Assessment and Fishery Evaluation Report: Gulf of Alaska Walleye Pollock," `http://www.afsc.noaa.gov/REFM/docs/2007/GOApollock.pdf`.

— (2008b), "Stock assessment and fishery evaluation report: assessment of walleye pollock in the Gulf of Alaska," `http://www.afsc.noaa.gov/REFM/docs/2008/GOApollock.pdf`.

— (2009), "Stock assessment and fishery evaluation report: Gulf of Alaska walleye pollock," `http://www.afsc.noaa.gov/REFM/docs/2009/GOApollock.pdf`.

Duchon, J. (1977), "Spline minimizing rotation-invariant semi-norms in solober spaces," in *Construction Theory of Functions of Several Variables*, eds. Schemp, W. and Zeller, K., Berlin: Springer, pp. 85–100.

Dwyer, D. A., Bailey, K. M., and Livingston, P. A. (1987), "Feeding habits and daily ration of walleye pollock (*Theragra chalcogramma*) in the eastern Bering Sea, with special reference to cannibalism," *Can J Fish Aquat Sci*, 44, 1972–1984.

Frank, K. T., B, B. P., Choi, J. S., and Leggett, W. C. (2005), "Trophic cascades in a formerly cod-dominated ecosystem," *Science*, 308, 1621–1623.

Golub, G. H., Heath, M., and Wahba, G. (1979), "Generalized cross validation as a method for choosing a good ridge parameter," *Technometrics*, 21, 215–223.

Gu, C. (2002), *Smoothing Spline ANOVA Model*, New York: Springer.

Hare, S. R. and Mantua, N. J. (2000), "Empirical evidence for North Pacific regime shifts in 1977 and 1989," *Prog Oceanogr*, 47, 103–145.

Hastie, T. and Tibshirani, R. (1986), "Generalized additive models," *Statistical Science*, 1, 297–318.

— (1990), *Generalized Additive Models*, Chapman & Hall.

Helle, K., Bogstad, B., Marshall, T. T., Michalsen, K., Ottersen, G., and Pennington, M. (2000), "An evaluation of recruitment indices for Arcto-Norwegian cod (Gadus morhua L.)," *Fish Res*, 48, 55–67.

Hinckley, S., Bailey, K. M., Picquelle, S., Yoklavich, M., and Stabeno, P. (1993), "Age-specific mortality and transport of larval walleye pollock (Theragra chalcogramma) in the western Gulf of Alaska." *Mar. Ecol. Prog. Ser.*, 98, 17–29.

Hjort, J. (1914), "Fluctuations in the great fisheries of northern Europe, viewed in the light of biological research," *Rap P -V Reun Conseil Perm Int Explor Mer*, 20, 1–228.

Hollowed, A. B., Hare, S. R., and Wooster, W. S. (2001), "Pacific basin climate variability and patterns of Northeast Pacific marine fish production," *Prog Oceanogr*, 49, 257–282.

Houde, E. D. (2008), "Emerging from Hjort's Shadow," *J Northw Atl Fish Sci*, 41, 53–70.

Huang, T.-M. and Chen, H. (2008), "Estimating the Parametric Component of Nonlinear Partial Spline Model," *Journal of Multivariate Analysis*, 99, 1665–1680.

Hunt, G. J., Stabeno, P., Walters, G., Sinclair, E., and R. D. Brodeur, Napp JM, B. N. (2002), "Climate change and control of the southeastern Bering Sea pelagic ecosystem," *Deep-Sea Res. II*, 49, 5821–5854.

Kendall, B. E., Briggs, C. J., Murdoch, W. W., Turchin, P., Ellner, S. P., Mc-Cauley, E., Nisbet, R. M., and Wood, S. N. (1999), "Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches," *Ecology*, 86, 1789–1805.

Knoth, B. A. and Foy, R. J. (2008), "Temporal variability in the food habits of arrowtooth flounder (*Atheresthes stomias*) in the Western Gulf of Alaska," *U. S. Dep. Commer., NOAA Tech. Memo.*, NMFS-AFSC-184, 30.

Lee, Y. W., Megrey, B. A., and Mackin, S. A. (2009), "Evaluating the performance of Gulf of Alaska walleye pollock (*Theragra chalcogramma*) recruitment forecasting models models using a Monte Carlo resampling strategy," *Can J Fish Aquat Sci*, 66, 367–381.

Lin, X. and Zhang, D. (1999), "Inference in generalized additive mixed models using smoothing splines," *Journal of the Royal Statistical Society, Series B*, 61, 381–400.

Litzow, M. A. and Ciannelli, L. (2007), "Oscillating trophic control induces community reorganization in a marine ecosystem," *Ecol Letters*, 10, 1–11.

Liu, H. and Chan, K. S. (2009), "Constrained Generalized Additive Models for Zero-Inflated Data," *Technical Report 388 of Department of Statistics & Actuarial Science, University of Iowa*.

Mammen, E. and van de Geer, S. (1997), "Penalized Quasi-Likelihood Estimation in Partial Linear Models," *The Annals of Statistics*, 25, 1014–1035.

Marx, B. D. and Eilers, P. H. (1998), "Direct generalized additive modeling with penalized likelihood," *Computational Statistics and Data Analysis*, 28, 193–209.

McKelvey, D. R. (1996), "Juvenile pollock distribution and abundance in Shelikof Strait-What can we learn from acoustic survey results?" *NOAA Tech Rep*, 126, 25–34.

McKinnell, S. (2004), "The precautionary approach to the PDO," *Pices Press*, 12, 16–17.

Meekan, M. G., Milicich, M. J., and Doherty, P. J. (1993), "Larval production drives temporal patterns of larval supply and recruitmen of a coral reef damselfish," *Mar Ecol Prog Ser*, 93, 217–225.

Mukhina, N. V., Marshall, C. T., and Yaragina, N. A. (2003), "Tracking the signal in year-class strength of northeast Arctic cod through multiple survey estimates of egg, larval and juvenile abundance," *J Sea Res*, 50, 57–75.

NOAA (2008), "NOAA 200th Visions: Ocean research," `celebrating200years.noaa.gov/visions/ocean_research/welcome.html`.

Platt, T., Fuentes-Yaco, C., and Frank, K. (2003), "Spring algal bloom and larval fish survival," *Nature*, 423, 398–399.

Ravier, C. and Fromentin, J.-M. (2001), "Long-term fluctuations in the eastern Atlantic and Mediterranean bluefin tuna population," *ICES J Mar Sci*, 58, 1299–1317.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Shida, O., Hamatsu, T., Nishimura, A., Suzaki, A., Yamamoto, J., Miyashita, K., and Sakurai, Y. (2007), "Interannual fluctuations in recruitment of walleye pollock in the Oyashio region related to environmental changes," *Deep-Sea Res II*, 54, 2822–2831.

Shima, M., Hollowed, A. B., and VanBlaricom, G. R. (2002), "Changes over time in the sptial distribution of walleye pollock (*Theragra chalcogramma*) in the Gulf of Alaska, 1984-1996," *Fish Bull US*, 100, 307–323.

Shiomoto, A., Tadokoro, K., Nagasawa, K., and Ishida, Y. (1997), "Trophic relations in the subarctic North Pacific ecosystem: possible feeding effect from pink salmon," *Mar Ecol Prog Ser*, 150, 75–85.

Southward, A. J., Boalch, G. T., and Maddock, L. (1988), "Fluctuations in the herring and pilchard fisheries of Devon and Cornwall linked to change in climate since the 16th century," *J Mar Biol Assoc UK*, 68, 423–445.

Speckman, S. G., Piatt, J. F., Minte-Vera, C. V., and Parrish, J. K. (2005), "Parallel structure among environmental gradients and three trophic levels in a subarctic estuary," *Prog Oceanogr*, 66, 25–65.

Spencer, P. D. (2008), "Density-independent and density-dependent factors affecting temporal changes in spatial distributions of eastern Bering Sea flatfish," *Fish Oceanogr*, 17, 396–410.

Stenseth, N. C., Viljugrein, H., Saitoh, T., Hansen, T. F., Kittilsen, M. O., Bolviken, E., and Glockner, F. (2003), "Seasonality, density dependence, and population cycles in Hokkaido voles," *Proc Natl Acad Sci*, 100, 11478–11483.

Svendsen, E., Skogen, M., Budgell, P., Huse, G., Stiansen, J., Adlandsvik, B., Vikebo, F., Asplin, L., and Sundby, S. (2007), "An ecosystem modeling approach to predicting cod recruitment," *Deep-Sea Res II*, 54, 2810–2821.

Turnock, B. J. and Wilderbuer, T. K. (2007), "Stock Assessment and Fishery Evaluation Report: Gulf of Alaska Arrowtooth Flounder Stock Assessment," `http://www.afsc.noaa.gov/refm/docs/2007/GOAatf.pdf`.

van der Vart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.

van der Veer, H. W. (1986), "Immigration, settlement, and density-dependent mortality of a larval and early postlarval 0-group plaice (Pleuronectes platessa) population in the western Wadden Sea," *Mar Ecol Prog.Ser*, 29, 23–236.

Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia: SIAM.

Walters, C. J. and Collie, J. S. (1988), "Is research on environmental factors useful to fisheries management?" *Can J Fish Aquat Sci*, 45, 1848–1854.

Wang, Y. and Ke, C. (2009), "Smoothing Spline Semiparametric Nonlinear Regression Models," *Journal of Computational and Graphical Statistics*, 18, 165–183.

Wijesinha, M. C. and Piantadosi, S. (1995), "Dose-Response Models with Co-variate," *Biometrics*, 51, 977–987.

Wood, S. (2001), "mgcv: GAMs and Generalized Ridge Regression for R," *R News*, 1(2), 20–25.

Wood, S. N. (2006), *Generalized Additive Models - an Introduction with R*, Chapman & Hall.

Yamamura, O., Yabuki, K., Shida, O., Watanabe, K., and Honda, S. (2001), "Spring cannibalism on 1 year walleye pollock in the Doto area, northern Japan: is it density dependent?" *J Fish Biol*, 59, 645–656.

Yang, M. (1993), "Food habits of the commercially important groundfishes in the Gulf of Alaska in 1990," *U.S. Dep. Commer., NOAA Tech. Memo.*, NMFS-AFSC-22, 1–150.

Yang, M.-S., Dodd, K., Hibpshman, R., and Whitehouse, A. (2006), "Food habits of groundfishes in the Gulf of Alaska in 1999 and 2001. U.S. Dep. Commer." *NOAA Tech. Memo.*, NMFS-AFSC-164, 1–199.

Yoklavich, M. M. and Bailey, K. M. (1990), "Hatching period, growth and survival of larval walleye pollock as determined from otolith analysis," *Mar. Ecol. Prog. Ser.*, 64, 13–23.

Zhang, T., Bailey, K. M., and Chan, K. S. (2010), "Recruitment forecast models of walleye pollock (Theragra chalcogramma) fine-tuned from juvenile survey data and predator community structure and environmental phase shifts (in press)," *Mar Ecol Prog Ser.*