Markov Switching Dirichlet Process Mixture Regression

Matthew Taddy and Athanasios Kottas *

February 21, 2009

Markov switching models can be used to study heterogeneous populations that are observed over time. This paper explores modeling the group characteristics nonparametrically, under both homogeneous and nonhomogeneous Markov switching for group probabilities. The model formulation involves a finite mixture of conditionally independent Dirichlet process mixtures, with a Markov chain defining the mixing distribution. The proposed methodology focuses on settings where the number of subpopulations is small and can be assumed to be known, and flexible modeling is required for group regressions. We develop Dirichlet process mixture prior probability models for the joint distribution of individual group responses and covariates. The implied conditional distribution of the response given the covariates is then used for inference. The modeling framework allows for both non-linearities in the resulting regression functions and non-standard shapes in the response distributions. We design a simulation-based model fitting method for full posterior inference. Furthermore, we propose a general approach for inclusion of external covariates dependent on the Markov chain but conditionally independent from the response. The methodology is applied to a problem from fisheries research involving analysis of stock-recruitment data under shifts in the ecosystem state.

KEY WORDS: Dirichlet process prior; hidden Markov model; Markov chain Monte Carlo; multivariate normal mixture; stock-recruitment relationship.

^{*}M. Taddy is Assistant Professor of Econometrics and Statistics, University of Chicago, Booth School of Business, Chicago, IL 60637 (E-mail: matt.taddy@chicagobooth.edu); and A. Kottas is Associate Professor, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064 (E-mail: thanos@ams.ucsc.edu). This research is part of the first author's Ph.D. dissertation, completed at University of California, Santa Cruz, and was supported in part by the National Science Foundation under Award DEB-0727543. The authors thank Alec MacCall, Steve Munch, and Marc Mangel for helpful discussions regarding the analysis of the Pacific sardine data.

1 INTRODUCTION

The focus of this work is to develop a flexible approach to nonparametric switching regression which combines Dirichlet process (DP) mixture nonparametric regression with a hidden Markov model. A modeling framework for data that has been drawn from a number of unobserved states (or regimes), where each state defines a different relationship between response and covariates, switching regression was originally developed in the context of econometrics (Goldfeld and Quandt, 1973; Quandt and Ramsey, 1978) and has primarily been approached through likelihood-based estimation. A hidden Markov mixture model in this context holds that the state vector constitutes a Markov chain, and thus introduces an underlying dependence into the data. In such models, the regression functions corresponding to individual population regimes are typically linear with additive error, and may or may not include an explicit time-series component (e.g., Hamilton, 1989; McCulloch and Tsay, 1994). The work presented here has a different focus: inference that is completely flexible and nonparametric within regimes, but guided by an informative parametric hidden Markov model for regime state switching. Such approaches reveal a baseline inference: the posterior distribution for individual regression functions when informed by little more than the state switching model. The proposed posterior simulation algorithms and approach to forward-backward sampling for stick-breaking prior models will also serve as a useful framework for more general inference about mixtures of conditionally independent nonparametric processes.

Bayesian nonparametrics, and DP mixtures in particular, provide highly flexible models for inference. Indeed, the practical implication of this flexibility is that, for inference based on small to moderate sample sizes, a certain amount of prior information must be provided to avoid a uselessly diffuse posterior. The DP hyperparameters provide the natural mechanism for introducing prior information, and this aspect of prior specification is discussed in Section 2.1. However, it is also possible to constrain inference by embedding the nonparametric component within a larger model. The typical semiparametric extension to linear regression – nonparametric modeling for the additive error distribution – is a familiar example of this approach. One can afford to be very noninformative about the error distribution only because linearity of the mean imposes a substantial constraint on model flexibility. This paper explores one such class of semiparametric inference settings: nonparametric density or regression estimation for heterogeneous populations, using a DP mixture framework, nested within an informative parametric model for the group membership, using an either homogeneous or nonhomogeneous hidden Markov switching model.

Although this framework applies generally to nonparametric density estimation, our particular focus is Markov switching nonparametric regression, specified in detail in Section 2. Section 2.1 presents the hidden Markov DP mixture model, and Section 2.2 develops an efficient Markov chain Monte Carlo (MCMC) algorithm for posterior simulation. Effective sampling of the hidden chain states is essential to success of the MCMC algorithm, and we thus propose a method based on forwardbackward sampling rather than direct Gibbs (see, e.g., Scott, 2002). An extension of the hidden Markov DP mixture model to include external variables that are correlated with the underlying Markov chain, but conditionally independent of the joint covariate-response distribution, is described in Section 2.3. In Section 3, the methods are illustrated with an application from fisheries research involving analysis of stock-recruitment data under shifts in the ecosystem state. Finally, Section 4 concludes with a summary and discussion of possible extensions. As a precursor to that work, we first introduce the two building blocks upon which the full model is based: Markov switching mixtures of DP mixtures in Section 1.1, and fully nonparametric implied conditional regression in Section 1.2.

1.1 Mixtures of Conditionally Independent Dirichlet Process Mixtures

Our methodological framework involves a known small number of states where prior information is available on the properties of the underlying state Markov chain, but there is a need for nonparametric modeling within each subpopulation. The assumption that the number of mixture states is known fits within the general premise of an informative state estimation coupled with flexible nonparametric modeling for regression estimation. In contrast with, say, switching linear regression, the state-specific DP mixture models do not generally provide significant information about the relative likelihood of different observation state allocations. Thus, while the methodology is not generally suitable for settings with little information about state membership, it offers a practical solution to switching regression problems that lack prior information about the shape of the individual regression functions.

The generic nonparametric DP mixture model is written as $f(z;G) = \int k(z;\theta) dG(\theta)$ for the density of z, with a parametric kernel density, $k(z;\theta)$, and a random mixing distribution G that is assigned a DP prior (Ferguson, 1973; Antoniak, 1974). In particular, $G \sim DP(\alpha, G_0)$, where α is the precision parameter, and G_0 is the centering distribution. A model for multiple heterogeneous populations may be built upon this platform. Assume R distinct random mixing distributions G_1, \ldots, G_R , each characterized as a DP in the prior, such that, for observations z_1, \ldots, z_n with population membership vector $\mathbf{h} = (h_1, \ldots, h_n)$, $f(z_i; G_{h_i}) = \int k(z_i; \theta) dG_{h_i}(\theta)$. This leads to the G_r being independent in the posterior full conditional (due, in particular, to conditioning on \mathbf{h}), which is both conceptually important and, in MCMC simulation, practically useful. Model specification is completed with a state probability vector, $\mathbf{p}_i = (p_{i,1}, \ldots, p_{i,R})$, defining the probability that the *i*-th observation was drawn from the DP mixture corresponding to each of the G_r . The goal of this framework is to introduce information into the model through the \mathbf{p}_i .

One way to inform $p_{i,r}$ is to incorporate temporal structure, and a natural way to do so is by assuming that the h_i constitute a Markov chain. Robert et al. (1993) and Chib (1996) discuss such hidden Markov models in the estimation of mixtures of parametric densities. Since, in our context, the G_r are modeled nonparametrically, this leads to inference that is driven primarily by state membership and, in particular, the Markov transition probabilities. Taking this approach further, the nonparametric switching regression methodology will be most effective when state membership probabilities are informed by external covariates. Hughes and Guttorp (1994) and Berliner and Lu (1999) have proposed nonhomogeneous hidden Markov models where each observation's state probability vector \mathbf{p}_i is regressed onto a set of external covariates, u_i . In Section 2.3, we obtain a similar model by assuming that the external u_i are randomly distributed according to a state dependent density function, $p_{h_i}(u_i)$. Conditioning on u_i then implies a nonhomogeneous hidden Markov model for **h**.

The general approach of having informative parametric modeling linked with nonparametric models through an underlying hidden stochastic process is both theoretically appealing and practically powerful. We believe that there is great potential for such models, since they provide an efficient way to bridge the difference in scale between two observed processes, and the MCMC algorithms presented in this paper can be the basis for extended techniques in other settings.

1.2 Implied Conditional Nonparametric Regression

In the Bayesian regression literature, two dominant trends have been to attempt to find increasingly flexible regression function models and to accompany these models with more comprehensive uncertainty quantification. Typically, Bayesian nonparametric modeling focuses on either the regression function or the error distribution. Müller and Quintana (2004) provide an overview of the respective methodologies.

The starting point for our approach is Bayesian nonparametric implied conditional regression, wherein DP mixtures are used to model the joint distribution of response and covariates, from which full inference is obtained for the desired conditional distribution for response given covariates. Both the response distribution and, implicitly, the regression function are modeled nonparametrically, thus providing a flexible framework for the general regression problem. In particular, working with (realvalued) continuous variables, we use DP mixtures of multivariate normal densities to model the joint density of the covariates, $\mathbf{X} = (X^1, ..., X^{d_x})$, and response Y (as in, e.g., Müller et al., 1996). This joint density, when divided by the implied marginal for covariates, allows for inference about the conditional density $f(y \mid \mathbf{x}) = f(\mathbf{x}, y)/f(\mathbf{x})$.

In the $d = d_{\mathbf{x}} + 1$ dimensional setting, with data $\mathcal{D} = \{\mathbf{z}_i = (x_i^1, \dots, x_i^{d_{\mathbf{x}}}, y_i) : i = 1, \dots, n\}$, the

location-scale normal DP mixture model (model \mathcal{M}_1) can be described as follows:

$$\mathbf{z}_{i} \mid G \stackrel{ind}{\sim} f(\mathbf{z}_{i}; G) = \int \mathcal{N}(\mathbf{z}_{i}; \mu, \Sigma) dG(\mu, \Sigma), \quad G \mid \alpha, \psi \sim \mathcal{DP}(\alpha, G_{0}(\psi)),$$
(1)

with the DP centering distribution given by $G_0(\mu, \Sigma; \psi) = \mathcal{N}(\mu; m, V) \mathcal{W}_{\nu}(\Sigma^{-1}; S^{-1})$, where $\psi = (m, V, S)$, and $\mathcal{W}_v(\cdot; M)$ denotes the Wishart distribution with v degrees of freedom and expectation vM. This mixture specification provides also the prior model for the marginal density for \mathbf{X} , $f(\mathbf{x}; G) = \int \mathcal{N}(\mathbf{x}; \mu^{\mathbf{x}}, \Sigma^{\mathbf{x}\mathbf{x}}) dG(\mu, \Sigma)$, after the mean vector and covariance matrix of the normal kernel have been partitioned. In particular, μ comprises $(d_{\mathbf{x}} \times 1)$ vector $\mu^{\mathbf{x}}$ and scalar μ^{y} , and Σ is a square block matrix with diagonal elements given by $(d_{\mathbf{x}} \times d_{\mathbf{x}})$ covariance matrix $\Sigma^{\mathbf{x}\mathbf{x}}$ and scalar variance Σ^{y} , and above and below diagonal vectors $\Sigma^{\mathbf{x}y}$, and $\Sigma^{y\mathbf{x}}$, respectively.

Inference is based upon a truncation approximation of the infinite dimensional parameter G. Although Müller et al. (1996) use only posterior predictive densities to obtain an approximation to $E\{f(y \mid \mathbf{x}; G) \mid \mathcal{D}\}$, posterior dependence between regime specific densities and the underlying state chain necessitates full inference for $f(y \mid \mathbf{x}; G)$. This extension of the implied conditional regression framework follows the development in Taddy and Kottas (2009), where full inference about $f(y \mid \mathbf{x}; G)$ is required to estimate quantile regression functions. Draws of the truncated random mixing distribution are possible through the *stick-breaking* DP definition (Sethuraman, 1994). Based on this definition, a realization G from the DP($\alpha, G_0(\psi)$) is almost surely a discrete distribution with a countable number of possible values drawn i.i.d. from $G_0(\psi)$, and corresponding weights that are built from i.i.d. $\beta(1, \alpha)$ variables through stick-breaking. (We use $\beta(a, b)$ to denote the Beta distribution with mean a/(a+b).) Hence, a truncation approximation to G can be defined as follows,

$$G^{L}(\cdot) = \sum_{l=1}^{L} \omega_{l} \delta_{\tilde{\theta}_{l}}(\cdot) \quad \text{with} \quad \boldsymbol{\omega}, \tilde{\boldsymbol{\theta}} \sim \mathcal{P}_{L}(\boldsymbol{\omega} \mid 1, \alpha) \prod_{l=1}^{L} dG_{0}(\tilde{\theta}_{l}; \psi),$$
(2)

where $\boldsymbol{\omega} = (\omega_1, ..., \omega_L), \ \tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, ..., \tilde{\theta}_L)$, and the finite stick-breaking prior $\mathcal{P}_L(\boldsymbol{\omega} \mid a, b)$ is defined

constructively by

$$\zeta_1, \dots, \zeta_{L-1} \stackrel{iid}{\sim} \beta(a, b), \ \zeta_L = 1; \quad \omega_1 = \zeta_1, \text{ and for } l = 2, \dots, L: \ \omega_l = \zeta_l \prod_{s=1}^{l-1} (1 - \zeta_s),$$
(3)

(see, e.g., Ishwaran and James, 2001). Given a posterior draw of the truncated DP parameters, $G^{L} = \{(\omega_{l}, (\tilde{\mu}_{l}, \tilde{\Sigma}_{l})) : l = 1, ..., L\}$, posterior realizations for the joint and marginal densities are readily available through $f(\mathbf{x}, y; G^{L}) = \sum_{l=1}^{L} \omega_{l} N(\mathbf{x}, y; \tilde{\mu}_{l}, \tilde{\Sigma}_{l})$ and $f(\mathbf{x}; G^{L}) = \sum_{l=1}^{L} \omega_{l} N(\mathbf{x}; \tilde{\mu}_{l}^{\mathbf{x}}, \tilde{\Sigma}_{l}^{\mathbf{xx}})$. Then, the posterior realization for the conditional response density at any value (\mathbf{x}, y) is given by $f(y \mid \mathbf{x}; G^{L}) = f(\mathbf{x}, y; G^{L}) / f(\mathbf{x}; G^{L})$. In addition, the structure of conditional moments for the normal mixture kernel enables posterior sampling of the conditional mean regression function without having to compute the conditional density. Specifically,

$$\mathbb{E}\left[Y \mid \mathbf{x}; G^{L}\right] = \frac{1}{f(\mathbf{x}; G^{L})} \sum_{l=1}^{L} \omega_{l} \mathcal{N}(\mathbf{x}; \tilde{\mu}_{l}^{\mathbf{x}}, \tilde{\Sigma}_{l}^{\mathbf{xx}}) \left[\tilde{\mu}_{l}^{y} + \tilde{\Sigma}_{l}^{y\mathbf{x}} (\tilde{\Sigma}_{l}^{\mathbf{xx}})^{-1} (\mathbf{x} - \tilde{\mu}_{l}^{\mathbf{x}})\right],$$
(4)

which, evaluated over a grid in \mathbf{x} , yields posterior realizations of the conditional mean function.

2 MARKOV SWITCHING NONPARAMETRIC REGRESSION

Mixtures of regressions are used to study multiple populations each of which involves a different conditional relationship between response and covariates. The basic switching regression model defines distinct regression functions for data that have been drawn from populations corresponding to a number of unobserved states. Following the early work of Goldfeld and Quandt (1973) and Quandt and Ramsey (1978), the more recent literature includes, for instance, approaches for switching dynamic linear models (Shumway and Stoffer, 1991) and switching ARMA models (Billio et al., 1999). Moreover, Hurn et al. (2003) describe a Bayesian decision theoretic approach to estimation for mixtures of linear regressions, whereas the approach of Shi et al. (2005) offers a departure from the linear regression assumption through a mixture of Gaussian process regressions. The generic mixtures of regressions setting holds that the response Y given covariates **X** has been drawn from a member of a heterogeneous set of R conditional distributions defined by the densities $f_1(y | \mathbf{x}), \ldots, f_R(y | \mathbf{x})$, and hence that $\Pr(y | \mathbf{x}) = p_1 f_1(y | \mathbf{x}) + \ldots + p_R f_R(y | \mathbf{x})$, where $\sum_{r=1}^{R} p_r = 1$. We propose a departure from this standard form, wherein the response and covariates are jointly distributed according to one of the densities $f_1(\mathbf{x}, y), \ldots, f_R(\mathbf{x}, y) - \text{i.e.}$, now $\Pr(\mathbf{x}, y) = p_1 f_1(\mathbf{x}, y) + \ldots + p_R f_R(\mathbf{x}, y)$ – and therefore $\Pr(y | \mathbf{x}) = \rho_1 f_1(\mathbf{x}, y) + \ldots + \rho_R f_R(\mathbf{x}, y)$, where $\rho_r = p_r / \sum_{\ell=1}^{R} p_\ell f_\ell(\mathbf{x})$. Thus, the approach is particularly appropriate whenever mixture component probabilities for a given **x** and y should be dependent upon the joint distribution for response and covariates, even though primary interest is in the regression relationship for response given covariates.

Section 2.1 describes our Bayesian nonparametric modeling approach to hidden Markov switching regression. Section 2.2 elaborates on the model formulation to develop methodology for posterior simulation. Finally, Section 2.3 provides a modeling extension for the inclusion of external covariates.

2.1 Model Specification for Hidden Markov Nonparametric Switching Regression

Here, we develop the extension of DP mixture implied conditional regression to the context of time dependent switching regression. The data consist of a set of covariate vectors \mathbf{x}_t and corresponding responses y_t observed at times t = 1, ..., T. The data from each time point are associated with a hidden state variable, $h_t \in \{1, ..., R\}$, such that, given h_t , the response-covariate joint distribution is defined by a state-specific density $f_{h_t}(\mathbf{x}_t, y_t)$. As in Section 1.2, we begin by describing density estimation in the $d = d_{\mathbf{x}} + 1$ dimensional setting, with data $\mathcal{D} = \{\mathbf{z}_t = (x_t^1, \ldots, x_t^{d_{\mathbf{x}}}, y_t) : t = 1, \ldots, T\}$. Now, however, the successive observations \mathbf{z}_t are correlated through dependence in state membership $\mathbf{h} = (h_1, \ldots, h_T)$, which constitutes a stationary Markov chain defined by an $R \times R$ transition matrix \mathbf{Q} . Although we consider only first-order dependence in the Markov chain, the model and posterior simulation methods can be extended to handle higher order Markov chains.

The first-order hidden Markov location-scale normal DP mixture model (model \mathcal{M}_2) can then be

expressed as follows,

$$\mathbf{z}_{t} \mid h_{t}, G_{h_{t}} \stackrel{ind}{\sim} f_{h_{t}}(\mathbf{z}_{t}) \equiv f(\mathbf{z}_{t}; G_{h_{t}}) = \int \mathcal{N}(\mathbf{z}_{t}; \mu, \Sigma) dG_{h_{t}}(\mu, \Sigma), \quad t = 1, \dots, T$$

$$G_{r} \mid \alpha_{r}, \psi_{r} \stackrel{ind}{\sim} \mathcal{DP}(\alpha_{r}, G_{0}(\psi_{r})), \quad r = 1, \dots, R$$

$$\mathbf{h} \mid \mathbf{Q} \sim \operatorname{Pr}(\mathbf{h} \mid \mathbf{Q}) = \prod_{t=2}^{T} Q_{h_{t-1}, h_{t}}, \qquad (5)$$

where we denote the r-th row of \mathbf{Q} by $Q_r = (Q_{r,1}, \ldots, Q_{r,R})$, with $Q_{r,s} = \Pr(h_t = s \mid h_{t-1} = r)$, for $r, s = 1, \ldots, R$. Applying the regression approach of Section 1.2, the joint response-covariate density specification in (5) yields our proposed hidden Markov switching regression model.

We assume that, in the prior, each state is equally likely for h_1 . Moreover, the DP centering distributions, $G_0(\mu, \Sigma; \psi_r) = \mathcal{N}(\mu; m_r, V_r) \mathcal{W}_{\nu_r}(\Sigma^{-1}; S_r^{-1})$, with $\psi_r = (m_r, V_r, S_r)$. For $r = 1, \ldots, R$, we place hyperpriors on ψ_r and α_r such that $\pi(\psi_r) = \mathcal{N}(m_r; a_{m_r}, B_{m_r}) \mathcal{W}_{a_{V_r}}(V_r^{-1}; B_{V_r}^{-1}) \mathcal{W}_{a_{S_r}}(S_r; B_{S_r})$, and $\pi(\alpha_r) = \Gamma(\alpha_r; a_{\alpha_r}, b_{\alpha_r})$. The prior for \mathbf{Q} is built from independent Dirichlet distributions, $\pi(Q_r) = \operatorname{Dir}(Q_r; \lambda_r)$, where $\operatorname{Dir}(Q_r; \lambda_r)$, with $\lambda_r = (\lambda_{r,1}, \ldots, \lambda_{r,R})$, denotes the Dirichlet distribution such that $\mathbb{E}[Q_{r,s}] = \lambda_{r,s}/(\sum_{i=1}^R \lambda_{r,i})$.

In practice, the hyperparameters for the α_r , ψ_r and for \mathbf{Q} need to be carefully chosen. We are motivated by a setting where prior information is available on the state vector \mathbf{h} , and the λ_r parameters of $\pi(Q_r)$ are chosen based on prior expectation for the probabilities of moving from state r to each state in a single time step. However, this prior information pertains only to the transition probabilities between states and does not fully identify the state components. Thus, we need to provide enough information to facilitate identification of the mixture components and ensure that the transition probabilities defined by \mathbf{Q} refer to the intended states. On the other hand, the nonparametric regression is motivated by a desire to be noninformative about each regression component and we thus seek a more automatic prior specification for each ψ_r .

Within the framework of our DP mixture implied conditional regression, it is possible to have each state-specific centering distribution, $G_0(\psi_r)$, associate the densities $\int N(\mathbf{z}; \mu, \Sigma) dG_r(\mu, \Sigma)$ with

specific regions of the joint response-covariate space, without putting prior information on the shape of the conditional response density or regression curve within each region. Since the prior parameters m_r and V_r control the location of the normal kernels, the hyperparameters a_{m_r} , B_{m_r} , a_{V_r} , and B_{V_r} can be used to express prior belief about the state-specific joint response-covariate distributions. Specifically, assume a prior guess for the mean and covariance matrix corresponding to the population for state r, where prior information for the covariance may only be available in the form of a diagonal matrix. Then, we can set a_{m_r} equal to the prior mean, B_{m_r} to the prior covariance, and choose a_{V_r} and B_{V_r} such that $\mathbb{E}[V_r]$ is equal to the prior covariance (alternatively, $\mathbb{E}[V_r^{-1}]$ can be set equal to the inverse of the prior covariance matrix and we have observed the method to be robust to either specification). In the absence of such prior information, one can use a data-dependent prior specification technique. Given a prior allocation of observations expressed as the state vector $\mathbf{h}^{\pi} =$ $(h_1^{\pi}, ..., h_T^{\pi})$, each set $\{a_{m_r}, B_{m_r}, B_{V_r}\}$ can be specified through the mean and covariance of the data subset $\{\mathbf{z}_t : h_t^{\pi} = r\}$. In particular, a_{m_r} is set to the state-specific data mean and both B_{m_r} and $\mathbb{E}[V_r] = (a_{V_r} - d - 1)^{-1} B_{V_r}$ are set to the state-specific data covariance. With care taken to ensure that it does not overly restrict the component locations, this approach provides an automatic prior specification that combines strong state allocation beliefs with weak information about the statespecific regression functions.

For the S_r we seek only to scale the mixture components to the data, and thus we set all the $\mathbb{E}(S_r) = a_{S_r}B_{S_r}$ equal to a diagonal matrix with each diagonal entry a quarter of the full data range for the respective dimension. The precision parameters a_{V_r} , a_{S_r} , and ν_r , for $r = 1, \ldots, R$, are set to values slightly larger than d+2; in practice, we have found 2(d+1) to work well. Working with various data sets, including the one in Section 3, we have observed results to be insensitive to reasonable changes in this specification. In particular, experimentation with a variety of choices for the matrices B_{S_r} , indicating prior expectation of either more or less diffuse normal kernel components, resulted in robust posterior inference.

Specification of the hyperpriors on DP precision parameters is facilitated by the role that each α_r plays in the prior distribution for the number of unique components in the set of n_r latent mixing parameters $\theta_t = (\mu_t, \Sigma_t)$ corresponding to state r. For a given n_r (i.e., conditional on **h**), we can use results from Antoniak (1974) to explore properties of this prior for different α_r values. For instance, the prior expected number of unique components in the set $\{\theta_t : h_t = r\}$ is approximately $\alpha_r \log[(n_r + \alpha_r)/\alpha_r]$, and this expression may be used to guide prior intuition about the α_r .

2.2 Efficient Posterior Simulation

Here, we present MCMC methods for posterior inference under the model developed in Section 2.1. To obtain the full probability model, we introduce latent parameters $\boldsymbol{\theta} = \{\theta_t = (\mu_t, \Sigma_t) : t = 1, ..., T\}$ such that the first stage in (5) is replaced with $\mathbf{z}_t \mid \theta_t \stackrel{ind}{\sim} N(\mathbf{z}_t; \theta_t)$ and $\theta_t \mid h_t, G_{h_t} \stackrel{ind}{\sim} G_{h_t}$, for t = 1, ..., T. The standard approach to posterior simulation from DP-based hierarchical models involves marginalization of the random mixing distributions G_r in (5) over their DP priors. Conditionally on \mathbf{h} , the vector of latent mixing parameters breaks down into state-specific subvectors $\boldsymbol{\theta}_r = \{\theta_t : h_t = r\}, r = 1, ..., R$, such that the distribution of each $\boldsymbol{\theta}_r$ is built from independent G_r distributions for the θ_t corresponding to state r. Thus, the full posterior can be written as $\Pr(\mathbf{h} \mid \mathbf{Q}) \prod_{r=1}^R \pi(\alpha_r) \pi(\psi_r) \pi(Q_r) \Pr(\boldsymbol{\theta}_r \mid \mathbf{h}, \alpha_r, \psi_r) DP(G_r; \alpha_r^*, G_{r0}^*) \prod_{t=1}^T N(\mathbf{z}_t; \theta_t)$, using results from Blackwell and MacQueen (1973) and Antoniak (1974). Here, $\Pr(\boldsymbol{\theta}_r \mid \mathbf{h}, \alpha_r, \psi_r)$ is the Pólya urn marginal prior for $\boldsymbol{\theta}_r; \alpha_r^* = \alpha_r + n_r$ (where $n_r = |\{t : h_t = r\}|$); and $G_{r0}^*(\cdot) \equiv G_{r0}^*(\cdot \mid \mathbf{h}, \theta_r, \alpha_r, \psi_r) =$ $(\alpha_r + n_r)^{-1} [\alpha_r dG_0(\cdot; \psi_r) + \sum_{\{t:h_t = r\}} \delta_{\theta_t}(\cdot)].$

This posterior can be sampled extending standard MCMC techniques for DP mixtures (see, e.g., Neal, 2000). However, marginalization over the G_r requires that each pair (θ_t, h_t) must be sampled jointly, conditional on the remaining paramaters $(\theta_{t'}, h_{t'})$, for all $t' \neq t$. This is possible, but inefficient, through use of a Metropolis-Hastings step with proposal distribution built from a marginal $\Pr(h_t = r) \propto Q_{h_{t-1},r}Q_{r,h_{t+1}}, r = 1, ..., R$, and a conditional for $\theta_t | h_t = r$ given by the Pólya urn prior full conditional arising from $\Pr(\theta_r | \mathbf{h}, \alpha_r, \psi_r)$.

2.2.1 Blocked Gibbs with forward-backward sampling

The posterior simulation approach discussed above requires updating each h_t one at a time, whereas forward-backward sampling for the entire state vector **h** is a substantially more efficient method for exploring the state space (see, e.g., Scott, 2002). To implement forward-backward sampling, we need to evaluate the joint probability mass function for states (h_{t-1}, h_t) conditional on the incomplete data vector $\{\mathbf{z}_1, ..., \mathbf{z}_t\}$ and relevant model parameters, which include the random mixing distributions $\{G_1, ..., G_R\}$. It is thus necessary to compute state probabilities approximating each G_r with the finite stick-breaking version in (2). The blocked Gibbs sampling method for DP mixture models (Ishwaran and James, 2001) provides a natural approach wherein the entire MCMC algorithm is based on a finite stick-breaking approximation of the DP. As well as being the consistent choice if the truncated distributions are used in state vector draws, blocked Gibbs can lead to very efficient sampling for the complete posterior.

Using the DP stick-breaking representation, we replace each G_r in model $\mathcal{M}2$ with a truncation approximation of the form in (2). Specifically, for specified (finite) L, we work with $G_r^L(\cdot) = \sum_{l=1}^L \omega_{l,r} \delta_{\tilde{\theta}_{l,r}}(\cdot)$, where the $\tilde{\theta}_{l,r} = (\tilde{\mu}_{l,r}, \tilde{\Sigma}_{l,r}), l = 1, ..., L$, are i.i.d. $G_0(\psi_r)$, and $\omega_r = (\omega_{1,r}, ..., \omega_{L,r})$ has distribution $\mathcal{P}_L(\omega_r \mid 1, \alpha_r)$ defined in (3). Hence, each G_r^L is defined by the set of L location-scale parameters $\tilde{\theta}_r = (\tilde{\theta}_{1,r}, ..., \tilde{\theta}_{L,r})$ and weights ω_r . Guidelines to choose the truncation level L, up to any desired accuracy, can be obtained, e.g., from Ishwaran and Zarepour (2000).

The first stage of model (5) is replaced with $\mathbf{z}_t \mid h_t$, $(\boldsymbol{\omega}_{h_t}, \tilde{\boldsymbol{\theta}}_{h_t}) \stackrel{ind}{\sim} \sum_{l=1}^L \omega_{l,h_t} N(\mathbf{z}_t; \tilde{\theta}_{l,h_t})$, t = 1, ..., T. The limiting case of this finite mixture model (as $L \to \infty$) is the countable DP mixture model $f(\mathbf{z}_t; G_{h_t}) = \int N(\mathbf{z}_t; \theta) dG_{h_t}(\theta)$ in (5). Again, we can introduce latent parameters $\theta_t = (\mu_t, \Sigma_t)$ to expand the first stage specification to $\mathbf{z}_t \mid \theta_t \stackrel{ind}{\sim} N(\mathbf{z}_t; \theta_t)$ and $\theta_t \mid h_t, (\boldsymbol{\omega}_{h_t}, \tilde{\boldsymbol{\theta}}_{h_t}) \stackrel{ind}{\sim} G_{h_t}^L$, for t = 1, ..., T. Alternatively, since $\theta_t = \tilde{\theta}_{l,h_t}$ with probability ω_{l,h_t} , we can work with configuration variables $\mathbf{k} = (k_1, ..., k_T)$, where each k_t takes values in $\{1, ..., L\}$, such that, conditionally on h_t , $k_t = l$ if and only if $\theta_t = \tilde{\theta}_{l,h_t}$. Hence, model $\mathcal{M}2$ with the DP truncation approximation can be expressed in the following hierarchical form

$$\mathbf{z}_{t} \mid \tilde{\boldsymbol{\theta}}_{h_{t}}, k_{t} \stackrel{ind}{\sim} \mathrm{N}(\mathbf{z}_{t}; \tilde{\boldsymbol{\theta}}_{k_{t}, h_{t}}), \quad t = 1, ..., T$$

$$k_{t} \mid h_{t}, \boldsymbol{\omega}_{h_{t}} \stackrel{ind}{\sim} \sum_{l=1}^{L} \omega_{l, h_{t}} \delta_{l}(k_{t}), \quad t = 1, ..., T$$

$$\boldsymbol{\omega}_{r}, \tilde{\boldsymbol{\theta}}_{r} \mid \alpha_{r}, \psi_{r} \stackrel{ind}{\sim} \mathcal{P}_{L}(\boldsymbol{\omega}_{r} \mid 1, \alpha_{r}) \prod_{l=1}^{L} dG_{0}(\tilde{\boldsymbol{\theta}}_{l, r}; \psi_{r}), \quad r = 1, ..., R$$

$$(6)$$

with $\mathbf{h} \mid \mathbf{Q} \sim \Pr(\mathbf{h} \mid \mathbf{Q}) = \prod_{t=2}^{T} Q_{h_{t-1},h_t}$, and the hyperpriors for $\boldsymbol{\alpha}, \boldsymbol{\psi}$, and \mathbf{Q} given in Section 2.1.

Denote by ϕ the vector comprising model parameters α , ψ , \mathbf{k} , \mathbf{Q} , and $\{(\boldsymbol{\omega}_r, \tilde{\boldsymbol{\theta}}_r) : r = 1, ..., R\}$. The full posterior, $\Pr(\phi, \mathbf{h} \mid \mathcal{D})$, corresponding to model (6) is now proportional to

$$\Pr(\mathbf{h} \mid \mathbf{Q}) \prod_{r=1}^{R} \left[\pi(\alpha_r) \pi(\psi_r) \pi(Q_r) \mathcal{P}_L(\boldsymbol{\omega}_r \mid 1, \alpha_r) \prod_{l=1}^{L} dG_0(\tilde{\theta}_{l,r}; \psi_r) \prod_{\{t:h_t=r\}} \left(N(\mathbf{z}_t; \tilde{\theta}_{k_t, r}) \sum_{l=1}^{L} \omega_{l, r} \delta_l(k_t) \right) \right].$$

Here, the key observation is that, conditionally on **h**, the first two stages of model (6), $\Pi_{t=1}^{T} \Pr(\mathbf{z}_{t}, k_{t} \mid h_{t}, (\boldsymbol{\omega}_{h_{t}}, \tilde{\boldsymbol{\theta}}_{h_{t}})) = \Pi_{t=1}^{T} \operatorname{N}(\mathbf{z}_{t}; \tilde{\boldsymbol{\theta}}_{k_{t}, h_{t}}) \left\{ \sum_{l=1}^{L} \omega_{l, h_{t}} \delta_{l}(k_{t}) \right\}, \text{ can be expressed in the state-specific form, } \Pi_{r=1}^{R} \left\{ \prod_{\{t:h_{t}=r\}} \operatorname{N}(\mathbf{z}_{t}; \tilde{\boldsymbol{\theta}}_{k_{t}, r}) \left\{ \sum_{l=1}^{L} \omega_{l, r} \delta_{l}(k_{t}) \right\} \right\}.$ To explore the full posterior, we develop an MCMC approach that combines Gibbs sampling steps for parameters in $\boldsymbol{\phi}$ with forward-backward sampling for the state vector **h**. We discuss the latter next, deferring to the Appendix the details of the Gibbs sampler for all other parameters.

As discussed above, sampling the truncated random mixing distribution $G_r^L \equiv (\boldsymbol{\omega}_r, \tilde{\boldsymbol{\theta}}_r)$ for each state r, enables use of forward-backward recursive sampling for the posterior full conditional distribution, $\Pr(\mathbf{h} \mid \boldsymbol{\phi}, \mathcal{D})$. Note that this conditional distribution can be written, in general, as $\Pr(h_T \mid \boldsymbol{\phi}, \mathcal{D}) \prod_{t=1}^{T-1} \Pr(h_{T-t} \mid \{h_{T-t+1}, ..., h_T\}, \boldsymbol{\phi}, \mathcal{D})$, whereas under the hidden Markov model structure it simplifies to

$$\Pr(\mathbf{h} \mid \boldsymbol{\phi}, \mathcal{D}) = \Pr(h_T \mid \boldsymbol{\phi}, \mathcal{D}) \prod_{t=1}^{T-1} \Pr(h_{T-t} \mid h_{T-t+1}, \boldsymbol{\phi}, \{\mathbf{z}_1, ..., \mathbf{z}_{T-t+1}\}).$$
(7)

Hence, the state vector can be updated as a block in each MCMC iteration by sampling from each component in (7).

To this end, the forward-backward sampling scheme begins by recursively calculating the *forward* matrices $F^{(t)}$, for t = 2, ..., T, where $F_{r,s}^{(t)} = \Pr(h_{t-1} = r, h_t = s \mid \phi, \{\mathbf{z}_1, ..., \mathbf{z}_t\})$, for r, s = 1, ..., R. Thus, $F^{(t)}$ defines the joint distribution for (h_{t-1}, h_t) given model parameters and data up to time t. For t = 3, ..., T, $F^{(t)}$ is obtained from $F^{(t-1)}$ through the following recursive calculation:

$$F_{r,s}^{(t)} \propto \Pr(h_{t-1} = r, h_t = s, \mathbf{z}_t \mid \phi, \{\mathbf{z}_1, ..., \mathbf{z}_{t-1}\})$$

$$= \Pr(h_t = s \mid h_{t-1} = r, \phi) \Pr(\mathbf{z}_t \mid h_t = s, \phi) \Pr(h_{t-1} = r \mid \phi, \{\mathbf{z}_1, ..., \mathbf{z}_{t-1}\})$$

$$= Q_{r,s} \sum_{l=1}^{L} \omega_{l,s} N(\mathbf{z}_t; \tilde{\theta}_{l,s}) \sum_{i=1}^{R} \Pr(h_{t-2} = i, h_{t-1} = r \mid \phi, \{\mathbf{z}_1, ..., \mathbf{z}_{t-1}\})$$

$$= Q_{r,s} \sum_{l=1}^{L} \omega_{l,s} N(\mathbf{z}_t; \tilde{\theta}_{l,s}) \sum_{i=1}^{R} F_{i,r}^{(t-1)}$$
(8)

where the proportionality constant is obtained from $\sum_{r=1}^{R} \sum_{s=1}^{R} F_{r,s}^{(t)} = 1$. For t = 2, a similar calculation yields $F_{r,s}^{(2)} \propto Q_{r,s} \sum_{l=1}^{L} \omega_{l,s} N(\mathbf{z}_2; \tilde{\theta}_{l,s}) \sum_{l=1}^{L} \omega_{l,r} N(\mathbf{z}_1; \tilde{\theta}_{l,r})$, where, again, the proportionality constant results from $\sum_{r=1}^{R} \sum_{s=1}^{R} F_{r,s}^{(2)} = 1$.

Next, exploiting the form in (7), the (stochastic) backward sampling step begins by drawing h_T according to $\Pr(h_T = r \mid \phi, D) = \sum_{i=1}^R \Pr(h_{T-1} = i, h_T = r \mid \phi, D) = \sum_{i=1}^R F_{i,r}^{(T)}$, for r = 1, ..., R. Sampling from (7) is then completed by drawing for each t = T - 1, T - 2, ..., 1 from $\Pr(h_t = r \mid h_{t+1}, \phi, \{\mathbf{z}_1, ..., \mathbf{z}_{t+1}\}) \propto \Pr(h_t = r, h_{t+1} \mid \phi, \{\mathbf{z}_1, ..., \mathbf{z}_{t+1}\}) = F_{r,h_{t+1}}^{(t+1)}$, for r = 1, ..., R, where the proportionality constant arises from $\sum_{r=1}^R F_{r,h_{t+1}}^{(t+1)}$.

2.2.2 Inference and forecasting for regression relationships

Inference for the state-specific regressions is based on an extension of the approach discussed in Section 1.2. In particular, conditional on the posterior draw for the state-specific mixing distribution, G_r^L , the posterior realization for the conditional response density, $f(y \mid \mathbf{x}; G_r)$, corresponding to state r is

$$f(y \mid \mathbf{x}; G_r^L) = \frac{f(\mathbf{x}, y; G_r^L)}{f(\mathbf{x}; G_r^L)} = \frac{\sum_{l=1}^L \omega_{l,r} \mathcal{N}(\mathbf{x}, y; \tilde{\mu}_{l,r}, \tilde{\Sigma}_{l,r})}{\sum_{l=1}^L \omega_{l,r} \mathcal{N}(\mathbf{x}; \tilde{\mu}_{l,r}^{\mathbf{x}}, \tilde{\Sigma}_{l,r}^{\mathbf{xx}})}.$$
(9)

The conditional mean regression for state r, $\mathbb{E}[Y | \mathbf{x}; G_r]$, can be estimated through a state-specific application of equation (4).

Moreover, of interest is prediction in future years (forecasting) for the joint response-covariate distribution and the corresponding implied conditional regression relationship. Illustrating with year T+1, the full model that includes the future covariate-response vector ($\mathbf{x}_{T+1}, y_{T+1}$) and corresponding regime state h_{T+1} , can be expressed as

$$\Pr((\mathbf{x}_{T+1}, y_{T+1}), h_{T+1}, \phi, \mathbf{h} \mid \mathcal{D}) = \Pr(\phi, \mathbf{h} \mid \mathcal{D}) Q_{h_T, h_{T+1}} \sum_{l=1}^{L} \omega_{l, h_{T+1}} N(\mathbf{x}_{T+1}, y_{T+1}; \tilde{\theta}_{l, h_{T+1}}).$$

Hence, the posterior samples for (ϕ, \mathbf{h}) along with draws for the new regime state h_{T+1} , driven by \mathbf{Q} and h_T , can be used to estimate the joint posterior forecast density $\Pr(\mathbf{x}_{T+1}, y_{T+1} \mid \mathcal{D})$. More generally, using the posterior samples for (ϕ, \mathbf{h}) and h_{T+1} , we obtain posterior realizations for the conditional response density in year T + 1 through $f(y \mid \mathbf{x}; G_{h_{T+1}}^L) = f(\mathbf{x}, y; G_{h_{T+1}}^L)/f(\mathbf{x}; G_{h_{T+1}}^L)$. Note that, in contrast to (9), these realizations incorporate posterior uncertainty in h_{T+1} . This type of inference is illustrated with the data example of Section 3.

2.3 Extension to Semiparametric Modeling with External Covariates

In the spirit of allowing the switching probabilities to drive the nonparametric regression, we extend here model $\mathcal{M}2$ to include additional information about the state vector in the form of an external covariate, U, with values $\boldsymbol{u} = \{u_1, ..., u_T\}$. (Although we present the methodology for a single covariate, the work can be readily extended to the setting with multiple external covariates.) The modeling extension involves a non-homogeneous Markov mixture where the hidden state provides a link between the joint covariate-response random variable and the external covariate.

The standard non-homogeneous hidden Markov model holds that the transition probabilities are dependent upon the external covariates, such that $\Pr(h_t \mid h_1, \ldots, h_{t-1}, u) = \Pr(h_t \mid h_{t-1}, u_t)$. Berliner and Lu (1999) present a Bayesian parametric approach to non-homogeneous hidden Markov models in which $\Pr(h_t \mid h_{t-1}, u_t)$ is estimated through probit regression. Also related is the likelihood analysis of Hughes and Guttorp (1994), wherein a heuristic argument, using Bayes theorem, is proposed to justify the model $\Pr(h_t \mid h_{t-1}, u_t) \propto \Pr(h_t \mid h_{t-1})\mathcal{L}(h_t; u_t)$, where the likelihood $\mathcal{L}(h_t; u_t)$ in their example is normal with state dependent mean.

Treating each u_t as the realization of a random variable yields a natural modeling framework in the context of our approach. Hence, we obtain a semiparametric extension of model $\mathcal{M}2$ (referred to as model $\mathcal{M}3$) by adding a further stage, $u_t \mid h_t \stackrel{ind}{\sim} p(u_t \mid \gamma_{h_t})$, to the hierarchical model specification in (6), along with hyperpriors for $\gamma = \{\gamma_r : r = 1, ..., R\}$, the state-specific parameters of the distribution for the external covariate. Moreover, we assume that u is conditionally independent of $\{\mathbf{z}_1, ..., \mathbf{z}_T\}$ given **h**. Thus, utilizing again the DP truncation approximation, the first stage of model $\mathcal{M}3$ is given by

$$\mathbf{z}_t, u_t \mid h_t, (\boldsymbol{\omega}_{h_t}, \tilde{\boldsymbol{\theta}}_{h_t}), \boldsymbol{\gamma} \stackrel{ind}{\sim} p(u_t \mid \gamma_{h_t}) \sum_{l=1}^L \omega_{l,h_t} N(\mathbf{z}_t; \tilde{\theta}_{l,h_t}), \quad t = 1, ..., T.$$
(10)

Clearly, the formulation of model $\mathcal{M}3$ implies that the hidden Markov chain is non-homogeneous conditional on \boldsymbol{u} . However, unconditionally in the prior, it is more accurate to say that $\{\mathbf{z}_1, ..., \mathbf{z}_T\}$ and \boldsymbol{u} are dependent upon a shared homogeneous Markov chain, and that they are conditionally independent given \mathbf{h} . In Section 3, we illustrate with a Gaussian form for $p(u_t | \gamma_{h_t})$. More general examples, with multiple external covariates, could incorporate dependence relationships, or even model some subset of the vector of external covariates as a function of the others.

Posterior inference under model $\mathcal{M}3$ can be implemented with a straightforward extension of the MCMC algorithm of Section 2.2. The parameters γ can be sampled conditional on only u and the state vector **h**. Regarding the other model parameters, only the MCMC draws that involve **h** need to be altered. In particular, the starting point is again an expression analogous to (7) for the posterior full conditional for **h**. Specifically, $Pr(\mathbf{h} \mid \phi, \gamma, D) = Pr(h_T \mid \phi, \gamma, D) \prod_{t=1}^{T-1} Pr(h_{T-t} \mid h_{T-t+1}, \phi, \gamma, \{(\mathbf{z}_{\ell}, u_{\ell}) : \ell = 1, ..., T - t + 1\})$. Note that now the data vector D comprises $\{(\mathbf{z}_t, u_t) : t = 1, ..., T\}$. For t = 3, ..., T, the recursive calculation of (8) for the forward matrices becomes

$$F_{r,s}^{(t)} \propto Q_{r,s} \mathbf{p}(u_t \mid \gamma_s) \sum_{l=1}^{L} \omega_{l,s} \mathbf{N}(\mathbf{z}_t; \tilde{\theta}_{l,s}) \sum_{i=1}^{R} F_{i,r}^{(t-1)},$$

with the proportionality constant obtained from $\sum_{r=1}^{R} \sum_{s=1}^{R} F_{r,s}^{(t)} = 1$. Moreover, $F_{r,s}^{(2)} \propto Q_{r,s} p(u_2 | \gamma_s) p(u_1 | \gamma_r) \sum_{l=1}^{L} \omega_{l,s} N(\mathbf{z}_2; \tilde{\theta}_{l,s}) \sum_{l=1}^{L} \omega_{l,r} N(\mathbf{z}_1; \tilde{\theta}_{l,r})$, where $\sum_{r=1}^{R} \sum_{s=1}^{R} F_{r,s}^{(2)} = 1$. Finally, the backward sampling step proceeds as described in Section 2.2 using probabilities from the forward matrices $F^{(T)}, F^{(T-1)}, ..., F^{(2)}$.

3 ANALYSIS OF STOCK-RECRUITMENT RELATIONSHIPS SUBJECT TO ENVIRONMENTAL REGIME SHIFTS

The relationship between the number of mature individuals of a species (stock) and the production of offspring (recruitment) is fundamental to the behavior of any ecological system. This has special relevance in fisheries research, where the stock-recruitment relationship applies directly to decision problems of fishery management with serious policy implications (e.g., Quinn and Derisio, 1999). A standard ecological modeling assumption holds that as stock abundance increases, successful recruitment per individual (reproductive success) decreases. However, a wide variety of factors will influence this reproductive relationship and there are many competing models for the influence of biological and physical mechanisms. Munch et al. (2005) present an overview of the literature on parametric modeling for stock-recruitment functions, arguing for the utility of standard semiparametric Gaussian process regression modeling. In the same spirit, albeit under the more general DP mixture modeling framework developed in Sections 1.2 and 2, our focus is to allow flexible regression to capture the nature of recruitment dependence upon stock without making parametric assumptions for either the stock-recruitment function or the errors around it.

An added complexity in studying stock-recruitment relationships is introduced by ecosystem *regime switching*. It has been observed that rapid shifts in the ecosystem state can occur, during which biological relationships, such as that between stock and recruitment, will undergo major change. This has been observed in the North Pacific in particular (McGowan et al., 1998; Hare and Mantua, 2000). Although empirical evidence of regime shifts is well documented and there have been attempts to establish mechanisms for the effect of this switching on stock-recruitment (e.g., Jacobson et al., 2005), the relationship between the physical effects of regime shifts and their biological manifestation is still unclear. This presents an ideal setting for Markov-dependent switching regression models due to their ability to link observed processes that occur on different scales (in this case, biological and physical) and are correlated in an undetermined manner.

To illustrate our Markov switching regression models, we use data on annual stock and recruitment for Japanese sardine from years 1951 to 1991. Wada and Jacobson (1998) use modeling of catch abundance and egg count samples to estimate R, the successful recruits of age less than one (in multiples of 10⁶ fish). With estimated annual egg production E (in multiples of 10¹² eggs) used as a proxy for stock abundance, they investigate the relationship between $\log(E)$ and log reproductive success, $\log(R/E)$. Japanese sardine have been observed to switch between favorable and unfavorablefeeding regime states related to the North Pacific environmental regime switching discussed above. Based upon a predetermined regime allocation (see Figure 1), Wada and Jacobson (1998) fit a linear regression relationship for $\log(E)$ vs $\log(R/E)$ within each regime.

Figure 1 about here.

We consider an analysis of the Japanese sardine data using the modeling framework developed in Section 2, which relaxes parametric (linear) regression assumptions and allows for simultaneous estimation of regime state allocation and regime-specific stock-recruitment relationships. As in the original analysis by Wada and Jacobson (1998), this model formulation does not take into account temporal dependence between successive observations from the same regime. This suits the purposes of our application, but one can envision many settings where a structured time series model is more appropriate than the fully nonparametric approach. Although the low dimensionality of this example is useful for illustrative purposes, the techniques will perhaps be most powerful in the exploration of higher dimensional datasets where such temporal structure is not assumed (an example of implied conditional regression in higher dimensions is studied in Taddy and Kottas, 2009).

We first apply model \mathcal{M}_2 in (6) to the sardine data, $\mathbf{z}_t = (\log(E_t), \log(R_t/E_t))$, available for T = 41 years from 1951 to 1991, with the underlying states h_t defined by either the unfavorable or favorable feeding regime (with values 1 or 2, respectively). A (conservative) truncation of L = 100was used in the stick-breaking priors. Regarding the prior hyperparameters, we set $a_{\alpha} = 2$ and $b_{\alpha} = 0.2$ in the gamma prior for α . The prior for ψ_r is specified as outlined in Section 2.1 such that, conditional on the prior regime allocation taken from Wada and Jacobson (1998), a_{m_1} and a_{m_2} are set to data means (5,3) and (5,5) for the unfavorable and favorable regime observations, respectively, while B_{m_1} and $(a_{V_1}-3)^{-1}B_{V_1}$, with diagonal (5.3, 2.6) and off-diagonal -3.1, and B_{m_2} and $(a_{V_2}-3)^{-1}B_{V_2}$, with diagonal (4.5, 1.4) and off-diagonal -2.0, is the observed covariance matrix for each regime. The B_{S_r} , for r = 1, 2, are diagonal matrices and are specified by setting the diagonal entries of $a_{S_r}B_{S_r}$ equal to (7.8, 7.7), which defines one quarter of the data range. Finally, we set $\nu_1 =$ $\nu_2 = a_{V_1} = a_{V_2} = a_{S_1} = a_{S_2} = 2(d+1) = 6$. The prior for **Q** is induced by a $\beta(3, 1.5)$ prior for the probability of staying in the same state, which reflects the relative rarity of regime shifts. The data and prior allocation are shown in Figure 1 along with bivariate normal draws based on the marginal mean and covariance matrix for the location, μ_r , of a single component of the DP mixture, for each of the two regimes. Hence, the right panel of Figure 1 shows draws from the prior expectation of the random mixing distribution for the μ_r (i.e., from state-specific normal distributions with means

 $\mathbb{E}[\mu_r] = a_{m_r}$ and variance $\operatorname{var}(\mu_r) = \operatorname{var}(m_r) + \mathbb{E}[V_r] = B_{m_r} + (a_{V_r} - 3)^{-1}B_{V_r})$. Noting that this does not include prior uncertainty in the μ_r due to the DP mixture, clearly shows that the prior specification has not overly restricted mixture components.

As described above, the sardine feeding regime is part of a larger ecosystem state for this region of the North Pacific. The physical variables that are linked to the ecosystem state switching can be used as external covariates for the hidden Markov chain. Hence, to illustrate the modeling approach of Section 2.3, we choose a physical variable as the single external covariate, specifically, the winter average Pacific decadal oscillation (PDO) index, which is highly correlated with biological regime switching (Hare and Mantua, 2000). The PDO index provides the first principle component of an aggregate of North Pacific sea surface temperatures. Although not directly responsible, sea surface temperature is believed to be a proxy for mechanisms such as current flow that control the regime switching (MacCall, 2002). Therefore, with vector \boldsymbol{u} comprising winter average PDO values from 1951 to 1991, we apply model $\mathcal{M}3$ to the sardine data working with a normal PDO distribution with state-specific mean. Hence, we assume $u_t \mid h_t \stackrel{ind}{\sim} N(u_t; \gamma_{h_t}, \tau^{-2})$, with (independent) normal priors for $\gamma = \{\gamma_1, \gamma_2\}$ and a gamma prior for τ^2 , in particular, $\gamma_1 \sim N(-0.44, 0.26), \gamma_2 \sim N(0.73, 0.26),$ and $\tau^2 \sim \Gamma(0.5, 0.125)$. The γ_r prior mean values are average winter PDO for two ten year periods that are generally accepted to fall within each ecosystem regime (Hare and Mantua, 2000); the common γ_r prior variance is the pooled variance for these mean estimates, and the prior median for τ^{-2} is chosen to provide some overlap between prior PDO densities for each regime. Extending the MCMC algorithm of Section 2.2 to sample the γ_r and τ^2 is straightforward, since their posterior full conditionals, conditional on u and \mathbf{h} , are given by normal and gamma distributions, respectively. The posterior means for γ_1 and γ_2 are given by -0.65 and 0.69, with 90% posterior intervals of (-0.89, -0.40) and (0.30, 1.10), respectively, and τ^{-2} has posterior mean 0.68 with a 90% posterior interval of (0.45, 1.00).



Results from the analyses under the two models are presented in Figures 2 – 4. The regimespecific posterior mean implied conditional densities, $\mathbb{E}\left[f(\log(R/E) \mid \log(E); G_r^L) \mid \mathcal{D}\right]$, evaluated over a 50×50 grid, are shown in Figure 2. These provide point estimates of the conditional relationship between stock and recruitment for each regime. Figure 3 shows the posterior mean for the state vector **h** as well as posterior point and interval estimates for mean regression functions, $\mathbb{E}[\log(R/E) \mid$ $\log(E); G_r^L]$, for each regime. The impact of inclusion of PDO as an external variable is evident. In the absence of such information, the observations for years 1988 - 1991 are more likely to be allocated in the favorable regime due to the rarity of regime shifting (i.e., due to posterior realizations of **Q** which put a high probability on staying in the same state). However, with the inclusion of PDO, these years are more probably associated with the unfavorable regime. Also, the posterior estimates for the regime-specific mean regression curves do not exclude the possibility of a linear mean relationship between log egg production and log reproductive success. Hence, it is interesting to note that the more general DP mixture switching regression modeling framework provides a certain level of support to the original assumptions of Wada and Jacobson (1998).

Figure 4 about here.

Finally, using the approach discussed in Section 2.2.2, we obtain prediction for 1992, the year following the end of our dataset. Posterior mean estimates of $f(\log(R/E) | \log(E); G_{h_{1992}}^L)$, the conditional response density in year 1992, are shown in the left panels of Figure 4, using the model with and without PDO as an external covariate (in the former case, based on $u_{1992} = 0.26$). In addition, the data include egg production estimates for the years 1992 to 1995 with $\log(E) = \log(675) = 6.515$ in 1992. Wada and Jacobson (1998) found a recruitment estimate based on partial stock assessment for 1992 of R = 20591 ($\log(20591/675) \approx 3.4$), but they were unable to quantify uncertainty about this value. The right panels of Figure 4 provide posterior point and interval estimates for the density $f(\log(R/E) | \log(E) = 6.515; G_{h_{1992}}^L)$. These results show the considerable change in year 1992 predicted response distribution (conditional on $\log(E) = 6.515$) from the inclusion of winter PDO in the model. The predictive uncertainty is actually increased with the inclusion of PDO, since it is now more likely that the regime state will be unfavorable in 1992 and there is little information from the data to inform the unfavorable regime regression curve around $\log(E) = 6.515$. The inference results reported in Figure 4 illustrate the posterior variability and non-standard shape of the predicted conditional response density. The quantification of this variability as well as the capacity of the DP mixture switching regression models to capture non-standard features of the response distribution are important aspects of the proposed nonparametric modeling framework.

4 CONCLUSION

We have presented a general framework for fully nonparametric regression based on Dirichlet process mixture models, and extended this approach with novel modeling and posterior simulation techniques for semiparametric hidden Markov switching regression. While the basic switching DP mixture regression methodology provides a powerful modeling technique in its own right, we feel that it is most practically important when combined with further parametric modeling for the effect of external covariates on state membership. Both modeling techniques, with or without the influence of external covariates, have been illustrated with the analysis of stock-recruitment data.

We have focused on models for switching regression, but the methodology is applicable in more general settings involving hidden Markov model structure. In particular, since the switching occurs at the level of the joint distribution for response and covariates, the modeling approach is directly applicable to nonparametric density estimation through DP mixtures of multivariate normals for heterogeneous populations where switching between subpopulations occurs as a Markov chain. Furthermore, the modeling framework can be elaborated for problems where the multivariate normal is not a suitable choice for the DP mixture kernel. For instance, categorical covariates can be accommodated through mixed continuous-discrete kernels. Finally, our work in the development of the MCMC algorithm can be extended to incorporate stick-breaking priors other than the DP.

APPENDIX: MCMC POSTERIOR SIMULATION

Here, we develop the approach to MCMC posterior simulation discussed in Section 2.2. Recall that the key to the finite stick-breaking algorithm is that we are able to use forward-backward recursive sampling of the posterior conditional distribution for \mathbf{h} as described in Section 2.2. Gibbs sampling details for all other parameters of model (6) are provided below.

First, for each t = 1, ..., T, k_t has a discrete posterior full conditional distribution with values in $\{1, ..., L\}$ and corresponding probabilities $\omega_{l,h_t} N(\mathbf{z}_t; \tilde{\theta}_{l,h_t}) / \{\sum_{m=1}^L \omega_{m,h_t} N(\mathbf{z}_t; \tilde{\theta}_{m,h_t})\}$, for l = 1, ..., L.

For each r = 1, ..., R, the posterior full conditional distribution for $\boldsymbol{\omega}_r$, is proportional to $\mathcal{P}_L(\boldsymbol{\omega}_r \mid 1, \alpha_r) \prod_{\{t:h_t=r\}} \left(\sum_{l=1}^L \omega_{l,r} \delta_l(k_t) \right) = \mathcal{P}_L(\boldsymbol{\omega}_r \mid 1, \alpha_r) \prod_{l=1}^L \omega_{l,r}^{M_{l,r}}$, where $M_{l,r} = |\{t : h_t = r, k_t = l\}|$. Note that the $\mathcal{P}_L(\boldsymbol{\omega}_r \mid 1, \alpha_r)$ prior for $\boldsymbol{\omega}_r$, defined constructively in (3), is given by

$$\mathcal{P}_{L}(\boldsymbol{\omega}_{r} \mid 1, \alpha_{r}) = \alpha_{r}^{L-1} \omega_{L,r}^{\alpha_{r}-1} (1 - \omega_{1,r})^{-1} (1 - (\omega_{1,r} + \omega_{2,r}))^{-1} \dots \left(1 - \sum_{l=1}^{L-2} \omega_{l,r}\right)^{-1}.$$
 (A.1)

Recall the generalized Dirichlet distribution $\mathcal{GD}(\mathbf{p}; \mathbf{a}, \mathbf{b})$ (Connor and Mosimann, 1969) for random vector $\mathbf{p} = (p_1, ..., p_L)$, supported on the L dimensional simplex, with density proportional to $p_1^{a_1-1} \dots p_{L-1}^{a_{L-1}-1} p_L^{b_L-1-1} (1-p_1)^{b_1-(a_2+b_2)} \dots (1-(p_1+...+p_{L-2}))^{b_{L-2}-(a_{L-1}+b_{L-1})}$, where the parameters are $\mathbf{a} = (a_1, ..., a_{L-1})$ and $\mathbf{b} = (b_1, ..., b_{L-1})$. Then, $\mathcal{P}_L(\boldsymbol{\omega}_r \mid 1, \alpha_r) \equiv \mathcal{GD}(\boldsymbol{\omega}_r; \mathbf{a}, \mathbf{b})$ with $\mathbf{a} = (1, ..., 1)$ and $\mathbf{b} = (\alpha_r, ..., \alpha_r)$. Moreover, the $\prod_{l=1}^{L} \omega_{l,r}^{M_{l,r}}$ form is also proportional to a $\mathcal{GD}(\boldsymbol{\omega}_r; \mathbf{a}, \mathbf{b})$ distribution with $\mathbf{a} = (M_{1,r} + 1, ..., M_{L-1,r} + 1)$ and $\mathbf{b} = ((L-1) + \sum_{l=2}^{L} M_{l,r}, ..., 2 + M_{L-1,r} + M_{L,r}, 1 + M_{L,r})$. Hence, the posterior full conditional for $\boldsymbol{\omega}_r$ can be completed to a generalized Dirichlet distribution with parameters $\mathbf{a} = (M_{1,r} + 1, ..., M_{L-1,r} + 1)$ and $\mathbf{b} = (\alpha_r + \sum_{l=2}^{L} M_{l,r}, \alpha_r + \sum_{l=3}^{L} M_{l,r}, ..., \alpha_r + M_{L,r})$. This distribution can be sampled constructively by first drawing independent $\zeta_l \sim \beta(1 + M_{l,r}, \alpha_r + \sum_{s=l+1}^{L} M_{s,r})$, for l = 1, ..., L - 1, and then setting $\omega_{1,r} = \zeta_1; \omega_{l,r} = \zeta_l \prod_{s=1}^{l-1} (1 - \zeta_s), l = 2, ..., L - 1;$ and $\boldsymbol{\omega}_{L,r} = 1 - \sum_{l=1}^{L-1} \omega_{l,r}$.

Next, for each r = 1, ..., R, the posterior full conditional distribution for $\tilde{\theta}_r$ is proportional to

 $\prod_{l=1}^{L} dG_0(\tilde{\theta}_{l,r};\psi_r) \prod_{j=1}^{n_r^*} \prod_{\{t:h_t=r,k_t=k_j^*\}} N(\mathbf{z}_t;\tilde{\theta}_{k_j^*,r}).$ Here, n_r^* is the number of distinct values of the k_t that correspond to the r-th state, i.e., the number of distinct k_t for $t \in \{t: h_t = r\}$. These distinct values are denoted by k_j^* , $j = 1, ..., n_r^*$. Now, for all $l \notin \{k_j^*: j = 1, ..., n_r^*\}$, we can draw $\tilde{\theta}_{l,r}$ i.i.d. $G_0(\psi_r)$. Otherwise, the posterior full conditional for $\tilde{\theta}_{k_j^*,r} \equiv (\tilde{\mu}_{j,r}^*, \tilde{\Sigma}_{j,r}^*)$ is proportional to

$$N(\tilde{\mu}_{j,r}^{*}; m_{r}, V_{r}) W_{\nu_{r}}(\tilde{\Sigma}_{j,r}^{*-1}; S_{r}^{-1}) \prod_{\{t: h_{t}=r, k_{t}=k_{j}^{*}\}} N(\mathbf{z}_{t}; \tilde{\mu}_{j,r}^{*}, \tilde{\Sigma}_{j,r}^{*}),$$

and can be sampled by extending the Gibbs sampler to draw from the full conditional for $\tilde{\mu}_{j,r}^*$ and for $\tilde{\Sigma}_{j,r}^{*-1}$. The former is normal with covariance matrix $T_j = (V_r^{-1} + M_{j,r}^* \tilde{\Sigma}_{j,r}^{*-1})^{-1}$, where $M_{j,r}^* = |\{t : h_t = r, k_t = k_j^*\}|$, and mean vector $T_j(V_r^{-1}m_r + \tilde{\Sigma}_{j,r}^{*-1} \sum_{\{t:h_t=r,k_t=k_j^*\}} \mathbf{z}_t)$. The latter is $W_{\nu_r+M_{j,r}^*}(\cdot; (S_r + \sum_{\{t:h_t=r,k_t=k_j^*\}} (\mathbf{z}_t - \tilde{\mu}_{j,r}^*)(\mathbf{z}_t - \tilde{\mu}_{j,r}^*)^T)^{-1})$.

The posterior full conditional for the hyperparameters, $\psi_r = (m_r, V_r, S_r)$, can be simplified by marginalizing the joint posterior full conditional for $\tilde{\theta}_r$ and ψ_r over all the $\tilde{\theta}_{l,r}$ for $l \notin \{k_j^* : j = 1, ..., n_r^*\}$. Thus, for each r = 1, ..., R, the full conditional for ψ_r is proportional to

$$N(m_r; a_{m_r}, B_{m_r}) W_{a_{V_r}}(V_r^{-1}; B_{V_r}^{-1}) W_{a_{S_r}}(S_r; B_{S_r}) \prod_{j=1}^{n_r^*} N(\tilde{\mu}_{j,r}^*; m_r, V_r) W_{\nu_r}(\tilde{\Sigma}_{j,r}^{*-1}; S_r^{-1})$$

Hence, ψ_r can be updated by separate draws from the posterior full conditionals for m_r , V_r , and S_r . The full conditional for m_r is normal with covariance matrix $B'_{m_r} = (B_{m_r}^{-1} + n_r^* V_r^{-1})^{-1}$ and mean vector $B'_{m_r}(B_{m_r}^{-1}a_{m_r} + V_r^{-1}\sum_{j=1}^{n_r^*}\tilde{\mu}_{j,r}^*)$. The full conditional for V_r^{-1} is $W_{n_r^*+a_{V_r}}(\cdot; (B_{V_r} + \sum_{j=1}^{n_r^*}(\tilde{\mu}_{j,r}^* - m_r)(\tilde{\mu}_{j,r}^* - m_r)^T)^{-1})$, and the full conditional for S_r is $W_{\nu_r n_r^*+a_{S_r}}(\cdot; (B_{S_r}^{-1} + \sum_{j=1}^{n_r^*}\tilde{\Sigma}_{j,r}^{*-1})^{-1})$.

Regarding the DP precision parameters, combining the $\Gamma(a_{\alpha_r}, b_{\alpha_r})$ prior for α_r with the relevant terms from (A.1), we obtain that, for each r = 1, ..., R, the posterior full conditional for α_r is a $\Gamma(a_{\alpha_r} + L - 1, -\log(\omega_{L,r}) + b_{\alpha_r})$ distribution.

Finally, with the $\text{Dir}(Q_r; \lambda_r)$ prior on each row Q_r of the transition matrix \mathbf{Q} , the posterior full conditional for Q_r is $\text{Dir}(Q_r; \lambda_r + J_r)$, where $J_r = (J_{r,1}, ..., J_{r,R})$ with $J_{r,s}$ denoting the number of transitions from state r to state s, which are defined by the currently imputed state vector \mathbf{h} .

References

- Antoniak, C. (1974), "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," *The Annals of Statistics*, 2, 1152 1174.
- Berliner, L. M. and Lu, Z.-Q. (1999), "Markov switching time series models with application to a daily runoff series," *Water Resources Research*, 35, 523–534.
- Billio, M., Monfort, A., and Robert, C. P. (1999), "Bayesian estimation of switching ARMA models," Journal of Econometrics, 93, 229–255.
- Blackwell, D. and MacQueen, J. (1973), "Ferguson distributions via Pólya urn schemes," The Annals of Statistics, 1, 353–355.
- Chib, S. (1996), "Calculating posterior distributions and modal estimates in Markov mixture models," Journal of Econometrics, 75, 79–97.
- Connor, R. and Mosimann, J. (1969), "Concepts of independence for proportions with a generalization of the Dirichlet distribution," *Journal of the American Statistical Association*, 64, 194–206.
- Ferguson, T. (1973), "A Bayesian analysis of some nonparametric problems," The Annals of Statistics, 1, 209–230.
- Goldfeld, S. M. and Quandt, R. E. (1973), "A Markov model for switching regression," Journal of Econometrics, 1, 3–16.
- Hamilton, J. D. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–384.
- Hare, S. R. and Mantua, N. J. (2000), "Empirical evidence for North Pacific regime shifts in 1977 and 1989," *Progress in Oceanography*, 47, 103–145.

- Hughes, J. P. and Guttorp, P. (1994), "A class of stochastic model for relating synoptic atmospheric patters to regional hydrologic phenomena," *Water Resources Research*, 30, 1535–1546.
- Hurn, M., Justel, A., and Robert, C. P. (2003), "Estimating mixtures of regressions," Journal of Computational and Graphical Statistics, 12, 55–79.
- Ishwaran, H. and James, L. (2001), "Gibbs sampling methods for stick-breaking priors," Journal of the American Statistical Association, 96, 161–173.
- Ishwaran, H. and Zarepour, M. (2000), "Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models," *Biometrika*, 87, 371–390.
- Jacobson, L. D., Bograd, S. J., Parrish, R. H., Mendelssohn, R., and Schwing, F. B. (2005), "An ecosystem-based hypothesis for climatic effects on surplus production in california sardine and environmentally dependent surplus production models," *Canadian Journal of Fisheries and Aquatic Sciences*, 62, 1782–1796.
- MacCall, A. D. (2002), "An hypothesis explaining biological regimes in sardine-producing pacific boundary current systems," in *Climate and fisheries: interacting paradigms, scales, and policy approaches: the IRI-IPRC Pacific Climate-Fisheries Workshop*, pp. 39–42, International Research Institute for Climate Prediction, Columbia University.
- McCulloch, R. E. and Tsay, R. S. (1994), "Statistical analysis of economic time series via markov switching models," *Journal of Time Series Analysis*, 15, 523–539.
- McGowan, J. A., Cayan, D. R., and Dorman, L. M. (1998), "Climate-ocean variability and ecosystem response in the Northeast Pacific," *Science*, 281, 210–217.
- Müller, P. and Quintana, F. A. (2004), "Nonparametric Bayesian data analysis," *Statistical Science*, 19, 95–110.

- Müller, P., Erkanli, A., and West, M. (1996), "Bayesian curve fitting using multivariate normal mixtures," *Biometrika*, 83, 67–79.
- Munch, S. B., Kottas, A., and Mangel, M. (2005), "Bayesian nonparametric analysis of stockrecruitment relationships," *Canadian Journal of Fisheries and Aquatic Sciences*, 62, 1808–1821.
- Neal, R. (2000), "Markov chain sampling methods for Dirichlet process mixture models," Journal of Computational and Graphical Statistics, 9, 249–265.
- Quandt, R. E. and Ramsey, J. B. (1978), "Estimating mixtures of normal distributions and switching regressions (with discussion)," *Journal of the American Statistical Association*, 73, 730–752.
- Quinn, T. J. I. and Derisio, R. B. (1999), Quantitative Fish Dynamics, Oxford University Press.
- Robert, C., Celeux, G., and Diebolt, J. (1993), "Bayesian estimation of hidden Markov chains: a stochastic implementation," *Statistics & Probability Letters*, 16, 77–83.
- Scott, S. (2002), "Bayesian methods for hidden Markov models: recursive computing for the 21st century," *Journal of the American Statistical Association*, 97, 337–351.
- Sethuraman, J. (1994), "A constructive definition of Dirichlet priors," Statistica Sinica, 4, 639–650.
- Shi, J. Q., Murray-Smith, R., and Titterington, D. M. (2005), "Hierarchical Gaussian process mixtures for regression," *Statistics and Computing*, 15, 31–41.
- Shumway, R. H. and Stoffer, D. S. (1991), "Dynamic linear models with switching," Journal of the American Statistical Association, 86, 763–769.
- Taddy, M. A. and Kottas, A. (2009), "A Bayesian nonparametric approach to inference for quantile regression," *Journal of Business and Economic Statistics*, to appear.
- Wada, T. and Jacobson, L. D. (1998), "Regimes and stock-recruitment relationships in Japanese sardine, 1951-1995," Canadian Journal of Fisheries and Aquatic Sciences, 55, 2455–2463.



Figure 1: The left panel plots the data with the regime allocation from Wada and Jacobson (1998). The right panel includes draws from the bivariate normal distribution, which, under each regime, is defined by the marginal mean and covariance matrix for the location of a single DP mixture component (see Section 3 for details). In both panels, black and grey color indicate the unfavorable and favorable regime, respectively.



Figure 2: Mean posterior conditional density surface for each regime. The unfavorable regime is plotted on the left panels and the favorable on the right panels. The top row corresponds to the analysis from model $\mathcal{M}2$ and the bottom row to model $\mathcal{M}3$, which includes PDO as an external covariate.



Figure 3: The left panels show the posterior mean regime membership by year, where 0 corresponds to the unfavorable regime. The right panels include posterior point and 90% interval estimates for the conditional mean regression function under each regime (interval estimates are denoted by dashed lines for the favorable regime, and by dotted lines for the unfavorable). The top row corresponds to model $\mathcal{M}2$ and the bottom row to model $\mathcal{M}3$, which includes PDO as an external covariate.



Figure 4: Predictive inference for year 1992. The left panels include posterior mean estimates of the conditional response density in 1992. The right panels show posterior point estimates (posterior means given by solid lines) and 90% interval estimates (dashed-dotted lines) for the conditional response density given the specified egg count for year 1992 of $\log(E) = 6.515$. The top row provides results under model $\mathcal{M}2$. The bottom row corresponds to the analysis with model $\mathcal{M}3$ with PDO as the external covariate, where prediction is conditional on a winter average PDO of 0.26 in 1992.