Modeling for Dynamic Ordinal Regression Relationships: An Application to Estimating Maturity of Rockfish in California

Maria DeYoreo and Athanasios Kottas *

Abstract

We develop a Bayesian nonparametric framework for modeling ordinal regression relationships which evolve in discrete time. The motivating application involves a key problem in fisheries research on estimating dynamically evolving relationships between age, length and maturity, the latter recorded on an ordinal scale. The methodology builds from nonparametric mixture modeling for the joint stochastic mechanism of covariates and latent continuous responses. This approach yields highly flexible inference for ordinal regression functions while at the same time avoiding the computational challenges of parametric models that arise from estimation of cut-off points relating the latent continuous and ordinal responses. A novel dependent Dirichlet process prior for time-dependent mixing distributions extends the model to the dynamic setting. The methodology is used for a detailed study of relationships between maturity, age, and length for Chilipepper rockfish, using data collected over 15 years along the coast of California.

KEY WORDS: Chilipepper rockfish; dependent Dirichlet process; dynamic density estimation; growth curves; ordinal regression

^{*}M. DeYoreo (maria.deyoreo@stat.duke.edu) is postdoctoral researcher, Department of Statistical Science, Duke University, Durham, NC, 27708, USA, and A. Kottas (thanos@soe.ucsc.edu) is Professor of Statistics, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA, 95064, USA. The authors wish to thank Stephan Munch for providing the rockfish data and for several useful comments on the interpretation of the results, Alec MacCall, Don Pearson, and Marc Mangel for valuable input on data collection and on key aspects of the specific problem from fisheries research, as well as an Associate Editor and two reviewers for several comments that greatly improved the presentation of the material in the paper. This research is part of the Ph.D dissertation of M. DeYoreo, completed at University of California, Santa Cruz, and was supported in part by the National Science Foundation under award DMS 1310438.

1 Introduction

The motivating application for this work lies in estimating dynamic relationships between age, length and fish maturity, with maturity recorded on an ordinal scale. This is a key problem in fisheries science, one reason being that estimates of age at maturity play an important role in population estimates of sustainable harvest rates (Clark, 1991; Hannah et al., 2009). The specific data set comes from the National Marine Fisheries Service and consists of year of sampling, age recorded in years, length in millimeters, and maturity for female Chilipepper rockfish, with measurements collected over 15 years along the coast of California. The ordinal maturity scale involves values from 1 through 3, where 1 indicates immature and 2 and 3 represent pre- and post-spawning mature, respectively.

Exploratory data analysis suggests both symmetric, unimodal as well as less standard shapes for the marginal distributions of length and age. Bivariate data plots of age and length suggest similar shapes across years, with some differences in location and scale, and clear differences in sample size, as can be seen from Figure 1. Maturity level is also indicated in the plots, and to make them more readable, random noise has been added to age, which is recorded on a discretized scale. While data is only shown for three years, there are similarities across years, including the concentration of immature fish near the lower left quadrants, but also differences such as the lack of immature fish in years 1995 through 2000 as compared to the early and later years. More details on the data are provided in Section 3.

In addition to studying maturity as a function of age and length, inference for the age and length distributions is also important. This requires a joint model which treats age and length as random in addition to maturity. We are not aware of any existing modeling strategy for this problem which can handle multivariate mixed data collected over time. Compromising this important aspect of the problem, and assuming the regression of maturity on body characteristics is the sole inferential objective, a possible approach would be to use an ordered probit regression model. Empirical (data-based) estimates for the trend in maturity as a function of length or age indicate shapes which may not be captured well by a parametric



Figure 1: Bivariate plots of length versus age at three years of data, with symbols/colors indicating maturity level. Red plus represents level 1 (immature), green circles level 2 (pre-spawning mature), and blue triangles level 3 (post-spawning mature). Values of age have been jittered to make the plots more readable.

model. For instance, the probability a fish is immature (level 1) is generally decreasing with length, however, in some of the years, the probability a fish is post-spawning mature (level 3) is increasing up to a certain length value and then decreasing. This is not a trend that can be captured by parametric models for ordinal regression (Boes and Winkelmann, 2006, discuss some of these properties). One could include higher order and/or interaction terms, though it is not obvious which terms to include, and how to capture the different trends across years.

In practice, virtually all methods for studying maturity as a function of age and/or length use logistic regression or some variant, often collapsing maturity into two levels (immature and mature) and treating each covariate separately in the analysis (e.g., Hannah et al., 2009; Bobko and Berkeley, 2004). Bobko and Berkeley (2004) applied logistic regression with length as a covariate, and to obtain an estimate of age at 50% maturity (the age at which 50% of fish are mature), they used their estimate for length at 50% maturity and solved for the corresponding age given by the von Bertalanffy growth curve, which relates age to length using a particular parametric function. Others assume that maturity is independent of length after conditioning on age, leading to inaccurate estimates of the proportion mature at a particular age or length (Morgan and Hoenig, 1997). We develop a flexible model to study time-evolving relationships between maturation, length, and age. These three variables constitute a random vector, and although maturity is recorded on an ordinal scale, it is natural to conceptualize an underlying continuous maturation variable. Distinguishing features of our approach include the joint modeling for the stochastic mechanism of maturation and length and age, and the ability to obtain flexible time-dependent inference for multiple ordinal maturation categories. While estimating maturity as a function of length and age is of primary interest, the joint modeling framework provides inference for a variety of functionals involving the three body characteristics.

The proposed modeling approach for dynamic ordinal regression avoids restrictive parametric assumptions through use of Bayesian nonparametric mixture priors. The methodology is particularly well-suited to the fish maturity application, as well as to related evolutionary biology problems that involve studying natural selection characteristics (such as survival or maturation) in terms of phenotypic traits. However, the methodology is more generally applicable for modeling ordinal responses collected along with covariates over discrete time, with multiple observations recorded at each time point.

We build on previous work on ordinal regression not involving time (DeYoreo and Kottas, 2017), where the ordinal responses arise from latent continuous variables, and the joint latent response-covariate distribution is modeled using a Dirichlet process (DP) mixture of multi-variate normals (Müller et al., 1996). In the context of the rockfish data, we model maturity, length, and age jointly, using a DP mixture. This modeling approach is further developed here to handle ordinal regressions indexed in discrete time, using a new dependent Dirichlet process (DDP) prior (MacEachern, 1999, 2000), which estimates the regression relationship at each time point in a flexible way, while incorporating dependence across time.

We review the model for ordinal regression without the time component in Section 2.1. Section 2.2 introduces the DDP mixture model, and in Section 2.3, we develop a new method for incorporating dependence in the DP weights to handle distributions indexed in discrete time. We discuss related literature on dependent nonparametric priors in Section 2.4. In Section 3, the DDP mixture model is used for a comprehensive analysis of the rockfish data discussed above. Section 4 concludes with a discussion. Technical details on properties of the DDP prior model are included in the Appendix. The Supplementary Material contains additional details on prior specification, posterior inference, and model comparison results.

2 Modeling Framework

2.1 Bayesian Nonparametric Ordinal Regression

We first describe the regression model for a single distribution. Let $\{(y_i, x_i) : i = 1, ..., n\}$ denote the data, where each observation consists of an ordinal response y_i along with a vector of covariates $x_i = (x_{i1}, ..., x_{ip})$. The methodology is developed in DeYoreo and Kottas (2017) for multivariate ordinal responses, however our application involves a univariate response. We assume that the ordinal responses arise as discretized versions of latent continuous responses, which is particularly relevant for the fish maturity application, as maturation is a continuous variable recorded on a discrete scale. With C categories, introduce latent continuous responses $(Z_1, ..., Z_n)$ such that $Y_i = j$ if and only if $Z_i \in (\gamma_{j-1}, \gamma_j]$, for j = 1, ..., C, and cut-off points $-\infty = \gamma_0 < \gamma_1 < \cdots < \gamma_C = \infty$.

We consider settings in which the covariates may be treated as random, which is appropriate, indeed necessary, for many biological and environmental applications. In our application, the focus is on building a flexible model for the joint stochastic mechanism of maturation, length, and age. In particular, we model the joint density $f(z, \boldsymbol{x})$ with a DP mixture of multivariate normals: $(z_i, \boldsymbol{x}_i) \mid G \stackrel{iid}{\sim} \int N(z_i, \boldsymbol{x}_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) dG(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $G \mid \alpha, G_0 \sim DP(\alpha, G_0)$.

By the DP constructive definition (Sethuraman, 1994), a realization G from a DP (α, G_0) is almost surely of the form $G = \sum_{l=1}^{\infty} p_l \delta_{\theta_l}$. The locations, $\theta_l = (\mu_l, \Sigma_l)$, are independent and identically distributed (i.i.d.) realizations from the centering distribution G_0 , and the weights are determined through stick-breaking from beta distributed random variables. In particular, let $v_l \approx \text{beta}(1, \alpha), \ l = 1, 2, \dots$, independently of $\{\theta_l\}$, and define $p_1 = v_1$, and for $l = 2, 3, \dots, p_l = v_l \prod_{r=1}^{l-1} (1 - v_r)$. Therefore, the model for $f(z, \boldsymbol{x})$ is a countable mixture of multivariate normals, which implies the following model for the regression functions:

$$\Pr(Y = j \mid \boldsymbol{x}, G) = \sum_{r=1}^{\infty} w_r(\boldsymbol{x}) \int_{\gamma_{j-1}}^{\gamma_j} N(z \mid m_r(\boldsymbol{x}), s_r) \, \mathrm{d}z, \qquad j = 1, \dots, C$$
(1)

with covariate-dependent weights $w_r(\boldsymbol{x}) \propto p_r N(\boldsymbol{x} \mid \boldsymbol{\mu}_r^x, \boldsymbol{\Sigma}_r^{xx})$, covariate-dependent means $m_r(\boldsymbol{x}) = \mu_r^z + \boldsymbol{\Sigma}_r^{zx} (\boldsymbol{\Sigma}_r^{xx})^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_r^x)$, and variances $s_r = \boldsymbol{\Sigma}_r^{zz} - \boldsymbol{\Sigma}_r^{zx} (\boldsymbol{\Sigma}_r^{xx})^{-1} \boldsymbol{\Sigma}_r^{xz}$. Here, $\boldsymbol{\mu}_r$ is partitioned into μ_r^z and $\boldsymbol{\mu}_r^x$ according to Z and X, and $(\boldsymbol{\Sigma}_r^{zz}, \boldsymbol{\Sigma}_r^{xx}, \boldsymbol{\Sigma}_r^{zx}, \boldsymbol{\Sigma}_r^{xz})$ are the components of the corresponding partition of covariance matrix $\boldsymbol{\Sigma}_r$.

This modeling strategy allows for general regression relationships, and overcomes many limitations of standard parametric models. In addition, the cut-offs may be fixed to arbitrary increasing values (which we recommend to be equally spaced and centered at zero) without sacrificing model flexibility. In particular, it can be shown that the induced prior model on the space of mixed ordinal-continuous distributions assigns positive probability to all Kullback-Leibler neighborhoods of any distribution in this space. This represents a key computational advantage over parametric models. We refer to DeYoreo and Kottas (2017) for more details on model properties, and illustrations of the benefits afforded by the nonparametric joint model over standard methods. This discussion refers to ordinal responses with three or more categories. For the case of binary regression, i.e., when C = 2, additional restrictions are needed on the covariance matrix Σ to facilitate identifiability (DeYoreo and Kottas, 2015).

2.2 Discrete-Time Dependent Dirichlet Process Mixture Model

In developing a model for a collection of distributions indexed in discrete time, we seek to build on previous knowledge, retaining the powerful and well-studied DP mixture model marginally at each time $t \in \mathcal{T}$, with $\mathcal{T} = \{1, 2, ...\}$. We thus seek to extend the DP prior to model $G_{\mathcal{T}} = \{G_t : t \in \mathcal{T}\}$, a set of dependent distributions such that each G_t follows a DP marginally. The dynamic DP extension can be developed by introducing temporal dependence in the weights and atoms of the DP constructive definition, that is, extending the DP prior for the mixing distribution G to a DDP prior for the discrete-time indexed collection of distributions $G_t = \sum_{l=1}^{\infty} p_{l,t} \delta_{\theta_{l,t}}$, for $t \in \mathcal{T}$.

In our context, the general DDP formulation expresses the atoms $\theta_{l\mathcal{T}} = \{\theta_{l,t} : t \in \mathcal{T}\}$, for $l = 1, 2, \ldots$, as i.i.d. sample paths from a time series model for the kernel parameters, with the stick-breaking weights $p_{l\mathcal{T}} = \{p_{l,t} : t \in \mathcal{T}\}$, for $l = 1, 2, \ldots$, arising through a latent time series with beta $(1, \alpha)$ (or beta $(1, \alpha_t)$) marginal distributions, independently of $\theta_{l\mathcal{T}}$. The construction of dependent weights requires dependent beta random variables, such that $p_{1,t} = v_{1,t}$, and $p_{l,t} = v_{l,t} \prod_{r=1}^{l-1} (1 - v_{r,t})$, for $l = 2, 3, \ldots$, with $\{v_{l,t} : t \in \mathcal{T}\}$, for $l = 1, 2, \ldots$, i.i.d. realizations from a time series with beta $(1, \alpha)$ marginals. Equivalently, we can write $p_{1,t} = 1 - \beta_{1,t}$, $p_{l,t} = (1 - \beta_{l,t}) \prod_{r=1}^{l-1} \beta_{r,t}$, for $l = 2, 3, \ldots$, with the $\{\beta_{l,t} : t \in \mathcal{T}\}$ i.i.d. realizations from a time series model with beta $(\alpha, 1)$ marginals.

The general DDP prior can be simplified by introducing temporal dependence only in the weights (common atoms) or only in the atoms (common weights). Our original intent was to use common atoms, however this model is not sufficiently flexible in predicting distributions for years without any data. The time window for our application spans year 1993 to 2007, but without data for years 2003, 2005, and 2006. Since the common atoms model forces the same values for the mixing parameters across years, it produces density estimates at missing years that tend to resemble an average across all years. This can adversely affect inference results when a temporal trend is supported by the data. We thus use the general DDP prior with a vector autoregressive (VAR) model for the atoms (detailed below), and a new construction for the DDP weights through transformation of a Gaussian time-series model (developed in Section 2.3).

We consider an observation time window spanning T time points (years), where data may be missing for some of the years. Denote by n_t the sample size in year t, and let \mathbf{Y}_{ti}^* , for $i = 1, \ldots, n_t$, be a (p+1)-dimensional continuous random vector consisting of the latent continuous random variable that determines the ordinal response, and the p covariates. Ordinal covariates can be accommodated in a similar way to the ordinal response, by introducing latent continuous covariates into \mathbf{Y}_{ti}^* . This is described further in Section 3.1 in the context of the rockfish data application which involves an ordinal and a continuous covariate. The DDP normal mixture model for the joint distribution of \boldsymbol{Y}_{ti}^* is expressed as

$$f(\boldsymbol{y}_{ti}^* \mid G_t) = \int \mathcal{N}(\boldsymbol{y}_{ti}^* \mid \boldsymbol{\theta}) \, \mathrm{d}G_t(\boldsymbol{\theta}) = \sum_{l=1}^{\infty} p_{l,t} \mathcal{N}(\boldsymbol{y}_{ti}^* \mid \boldsymbol{\mu}_{l,t}, \boldsymbol{\Sigma}_l), \quad t \in \boldsymbol{s}^c, \ i = 1, \dots, n_t$$
(2)

where temporal dependence in the atoms is introduced through the Gaussian kernel mean vectors, that is, $\boldsymbol{\theta}_{l,t} = (\boldsymbol{\mu}_{l,t}, \boldsymbol{\Sigma}_l)$. Here, \boldsymbol{s} collects the indexes for all years when data is missing, and $\boldsymbol{s}^c = \{1, \ldots, T\} \setminus \boldsymbol{s}$ represents all other years. The centering distribution for the $(p+1) \times (p+1)$ covariance matrices $\boldsymbol{\Sigma}_l$ is taken to be $\mathrm{IW}(\nu, \boldsymbol{D})$, an inverse-Wishart distribution with density proportional to $|\boldsymbol{\Sigma}_l|^{-(\nu+p+2)/2} \exp\{-0.5\mathrm{tr}(\boldsymbol{D}\boldsymbol{\Sigma}_l^{-1})\}$, where \boldsymbol{D} is a random hyperparameter and ν is fixed.

To build temporal dependence in the DDP atoms, we use a first-order VAR model for the time series that generates the $\{\mu_{l,t} : t \in \mathcal{T}\}$:

$$\boldsymbol{\mu}_{l,1} \mid \boldsymbol{m}_{0}, \boldsymbol{V}_{0} \sim \mathrm{N}(\boldsymbol{m}_{0}, \boldsymbol{V}_{0}), \qquad \boldsymbol{\mu}_{l,t} \mid \boldsymbol{\mu}_{l,t-1}, \boldsymbol{\Theta}, \boldsymbol{m}, \boldsymbol{V} \sim \mathrm{N}(\boldsymbol{m} + \boldsymbol{\Theta} \boldsymbol{\mu}_{l,t-1}, \boldsymbol{V}).$$
(3)

We take Θ to be diagonal, but allow a full covariance matrix for V. Hence, the mean of each element of $\mu_{l,t}$ depends only on the corresponding element of $\mu_{l,t-1}$, however dependence across elements of $\mu_{l,t}$ is allowed. A diagonal matrix Θ facilitates the selection of a stationary VAR model (it suffices to restrict the diagonal elements to lie in (-1, 1)) and the identification of the stationary distribution; see the Appendix. For the rockfish data application, we worked with uniform priors on (0, 1) for the diagonal elements of Θ .

The Supplementary Material includes details on prior specification for m, V, and D, as well as for the hyperparameters of the model for the DDP weights, which is discussed next.

2.3 A Time-Dependent Model for the DDP Weights

To obtain dependent stick-breaking weights, we define a stochastic process with $beta(\alpha, 1)$ marginal distributions as follows:

$$\mathcal{B} = \left\{ \beta_t = \exp\left(-\frac{\zeta^2 + \eta_t^2}{2\alpha}\right) : t \in \mathcal{T} \right\},\tag{4}$$

where $\zeta \sim N(0, 1)$ and, independently of ζ , $\eta_{\mathcal{T}} = \{\eta_t : t \in \mathcal{T}\}$ is a stochastic process with N(0, 1) marginal distributions. This transformation leads to marginal distributions $\beta_t \sim beta(\alpha, 1)$, since for two independent N(0, 1) random variables Y_1 and Y_2 , $W = (Y_1^2 + Y_2^2)/2$ follows an exponential distribution with mean 1, and thus $exp(-W/\alpha) \sim beta(\alpha, 1)$. To our knowledge, this is a novel construction for the weights in a DDP prior model. The practical utility of the transformation in (4) is that it facilitates building the temporal dependence through Gaussian time-series models, while maintaining the DP structure marginally.

Working with distributions indexed in discrete time, we take $\eta_{\mathcal{T}}$ to be a first-order autoregressive (AR) process. Therefore, the prior model for the DDP weights, $p_{1,t} = 1 - \beta_{1,t}$, $p_{l,t} = (1 - \beta_{l,t}) \prod_{r=1}^{l-1} \beta_{r,t}$, $l = 2, 3, \ldots$, is built from $\beta_{l,t} = \exp\{-(\zeta_l^2 + \eta_{l,t}^2)/2\alpha\}$, where

$$\zeta_{l} \stackrel{iid}{\sim} \mathcal{N}(0,1), \quad \eta_{l,1} \stackrel{iid}{\sim} \mathcal{N}(0,1), \quad \eta_{l,t} \mid \eta_{l,t-1}, \phi \sim \mathcal{N}(\phi \eta_{l,t-1}, 1 - \phi^{2}), \quad l = 1, 2, \dots, \quad t = 2, \dots, T$$
(5)

with $|\phi| < 1$. The restriction on the variance of the AR(1) model for $\eta_{\mathcal{T}}$, along with the assumption $\eta_{l,1} \sim N(0,1)$, results in the required N(0,1) marginals for the $\eta_{l,t}$.

The restriction $|\phi| < 1$ implies stationarity for stochastic process $\eta_{\mathcal{T}}$. Since \mathcal{B} is a transformation of a strongly stationary stochastic process, it is also strongly stationary. Note that the correlation in $(\beta_{l,t}, \beta_{l,t+k})$ is driven by the autocorrelation present in $\eta_{\mathcal{T}}$, and this induces dependence in the weights $(p_{l,t}, p_{l,t+k})$, which leads to dependent distributions (G_t, G_{t+k}) . We explore this dependence in the Appendix, deriving the expressions for $\operatorname{corr}(\beta_{l,t}, \beta_{l,t+k} \mid \alpha, \phi)$, $\operatorname{corr}(p_{l,t}, p_{l,t+k} \mid \alpha, \phi)$, and $\operatorname{corr}(G_t(A), G_{t+1}(A) \mid \alpha, \phi, G_{0,\mathcal{T}})$, for any specified measurable subset A in the support of the G_t . The last correlation is studied for both a generic DDP prior model with dependent atoms, as well as for the simplified model with common atoms. Note that $\operatorname{corr}(\beta_{l,t}, \beta_{l,t+k} \mid \alpha, \phi)$ depends on $\rho_k = \operatorname{corr}(\eta_{l,t}, \eta_{l,t+k} \mid \phi) = \phi^k$ only through ρ_k^2 , and thus it is natural to assume $\phi \in (0, 1)$.

2.4 Discussion of Related Literature

The literature includes several variations of the DDP model. The common weights version, with a Gaussian process used to generate dependent atoms, was first discussed by MacEachern (2000). Applications of common weights DDP mixtures include ANOVA models (De Iorio et al., 2004), spatial modeling (Gelfand et al., 2005; Kottas et al., 2012), survival regression (DeIorio et al., 2009), and dose-response modeling (Fronczyk and Kottas, 2014a,b).

For data indexed in discrete time, Rodriguez and ter Horst (2008) apply a common weights model, with atoms arising from a dynamic linear model. Di Lucca et al. (2013) develop a model for a time series of continuous or binary responses through a DDP with atoms dependent on lagged terms. Xiao et al. (2015) construct a dynamic DDP model for Poisson process intensities, using different types of autoregressive processes for the atoms.

Taddy (2010) assumes the alternative simplification of the DDP with common atoms, and models the stick-breaking proportions using the positive correlated autoregressive beta process from McKenzie (1985). Nieto-Barajas et al. (2012) also use the common atoms DDP version, modeling a time series of random distributions by linking the beta random variables through latent binomially distributed random variables. As discussed in Bassetti et al. (2014), the common atoms DDP prior structure may be a strong restriction when the distributions across time units exhibit a high degree of heterogeneity.

In the order-based DDP of Griffin and Steel (2006), covariates are used to sort the weights. Covariate dependence is incorporated in the weights in the kernel and probit stick-breaking models of Dunson and Park (2008) and Rodriguez and Dunson (2011), respectively, however these prior models do not retain the DP marginally. Leisen and Lijoi (2011) and Zhu and Leisen (2015) develop a dependent two-parameter Poisson-Dirichlet process, where the dependence is induced through a Lévy copula. There exist many related constructions for dependent distributions defined through correlated normalized completely random measures (e.g., Griffin et al., 2013; Lijoi et al., 2014).

Griffin and Steel (2011) and Bassetti et al. (2014) use different definitions of multivariate beta random variables to develop nonparametric prior models for multiple time series or repeated measures. While there exist many proposals for generating stick-breaking weights through multivariate beta random variables, an appealing feature of the construction of Section 2.3 is that temporal dependence is induced through a Gaussian time-series model. Moreover, the construction allows for extensions to modeling dependent distributions ranging over an uncountable set, such as spatially indexed distributions.

3 Estimating Maturity of Rockfish

3.1 Chilipepper Rockfish Data

Here, we provide details on the rockfish data pertaining to the application of the DDP mixture model developed in Section 2.

In the original rockfish data source, maturity is recorded on an ordinal scale from 1 to 6, representing immature (1), early and late vitellogenesis (2,3), eyed larvae (4), and post-spawning (5,6). Because scientists are not necessarily interested in differentiating between every one of these maturity levels, and to make the model output simpler and more interpretable, we collapse maturity into three ordinal levels, representing immature (1), prespawning mature (2,3,4), and post-spawning mature (5,6).

Many observations have age missing or maturity recorded as unknown. Exploratory analysis suggests there to be no systematic pattern in missingness. Further discussion with fisheries researchers having expertise in aging of rockfish and data collection revealed that the reason for missing age in a sample is that otoliths (ear stones used in aging) were not collected or have not yet been aged. Maturity may be recorded as unknown because it can be difficult to distinguish between stages, and samplers are told to record unknown unless they are reasonably sure of the stage. Therefore, there is no systematic reason that age or maturity is not present, and it is thus reasonable to assume that the data are missing completely at random. We therefore ignore the missing data mechanism, and base inferences only on the complete data (e.g., Rubin, 1976; Gelman et al., 2004). An alternative approach would be to impute the missing values. However, given the very large amount of missing data, and the

Table 1: Rockfish data organized by year, total sample size n_t , and sample sizes in each of the three maturity categories, labeled as $n_{1,t}$ (immature), $n_{2,t}$ (pre-spawning mature), and $n_{3,t}$ (post-spawning mature). Note that data is not available for years 2003, 2005 and 2006.

year	n_t	$n_{1,t}$	$n_{2,t}$	$n_{3,t}$
1993	305	35	232	38
1994	271	23	210	38
1995	256	9	160	87
1996	160	0	106	54
1997	184	3	144	37
1998	132	10	67	55
1999	396	7	317	72
2000	64	1	59	4
2001	146	9	89	48
2002	238	36	191	11
2004	37	6	30	1
2007	43	13	15	15

fact that we are not interested in the imputations, this approach would significantly increase the computational burden without providing any clear benefits.

Considering year of sampling as the index of dependence, observations occur in years 1993 through 2007, indexed by t = 1, ..., T = 15, with no observations in 2003, 2005, or 2006 (so $s = \{11, 13, 14\}$) and clear differences in sample size. The sample sizes, organized by year and maturity level, are provided in Table 1. This situation involving time points in which data is completely missing is not uncommon in these types of problems, and can be handled with our model for equally spaced time points. As discussed in Section 2.2, this aspect of the application is among the reasons for the use of the general DDP prior structure.

Age can not be treated as a continuous covariate, as there are approximately 25 distinct values of age in over 2,200 observations. Age is in fact recorded as an ordinal random variable, such that a recorded age k implies the fish was between k and k + 1 years of age. This relationship between discrete recorded age and continuous age is obtained by the following reasoning. Chilipepper rockfish are winter spawning, and the young are assumed to be born in early January. The annuli (rings) of the otiliths are counted in order to determine age, and these also form sometime around January. Thus, for each ring, there has been one year of growth. We therefore treat age much in the same way as maturity, using a latent continuous age variable. More specifically, Let U^* represent underlying continuous age, and assume, for k = 0, 1, ..., that we observe age k iff $U^* \in (k, k+1]$, or equivalently, iff $W = \log(U^*) \in (\log(k), \log(k+1)]$, so the support of the latent continuous random variable W corresponding to age is \mathbb{R} . Therefore, for the *i*-th fish at year *t* ordinal age *k* is observed iff $W_{ti} \in (\log(k), \log(k+1)]$, for k = 0, 1, ..., where W_{ti} is log-age on a continuous scale.

Denote also by X_{ti} and Z_{ti} the length and maturation on the continuous scale, respectively, for the *i*-th fish at year *t*, for $t \in s^c$ and $i = 1, ..., n_t$. Then, the DDP mixture model in (2) is applied to the $\mathbf{Y}_{ti}^* = (Z_{ti}, W_{ti}, X_{ti})$, where the observed ordinal maturity, Y_{ti} , arises from Z_{ti} through $Y_{ti} = j$ if and only if $\gamma_{j-1} < Z_{ti} \leq \gamma_j$, for j = 1, 2, 3.

3.2 Results

To sample from the posterior distribution of the hierarchical model for the data, we use blocked Gibbs sampling (Ishwaran and James, 2001), which involves a finite truncation approximation to the countable representation for each G_t . Details are included in the Supplementary Material, but we note here that a key aspect of the Markov chain Monte Carlo (MCMC) algorithm involves slice sampling steps for parameters $\{\zeta_l\}$ and $\{\eta_{l,t}\}$, which define the DDP weights. Moreover, we can estimate parameters $p_{l,t}$ and $\theta_{l,t}$ at every time point $t \in \{1, \ldots, T\}$, and therefore inference is available for G_t even if there is no data at year t.

There are several alternative MCMC methods that could be used, including a collapsed Gibbs sampler (e.g., Nieto-Barajas et al., 2012) or a slice sampler (e.g., Kalli et al., 2011). These approaches do not require truncation for the G_t from the outset. However, to obtain posterior samples for the G_t , which is essential for our inferential objectives, some form of truncation must be applied also under the alternative MCMC approaches; we refer to DeYoreo and Kottas (2017) for further details on this point.

Various simulation settings were developed to study both the common atoms version of the model and the more general version (DeYoreo, 2014, chapter 4). While we focus only on the fish maturity data application in this paper, our extensive simulation studies have revealed the inferential power of the model under different scenarios for the true latent



Figure 2: Posterior mean estimates for the bivariate density of age and length across all years. response distribution and ordinal regression relationships.

We first discuss inference results for functionals involving length and age, but not maturity. A feature of the modeling approach is that inference for the density of age can be obtained

over a continuous scale. The posterior mean surfaces for the bivariate density of age and length are shown in Figure 2 for all years, including the ones (years 2003, 2005 and 2006) for



Figure 3: Posterior mean and 95% interval bands for the expected value of length over (continuous) age, $E(\text{length} | \text{age}, G_t)$, across three years. Overlaid are the data (in blue) and the estimated von Bertalanffy growth curves (in red).

which data is not available. The model yields more smooth shapes for the density estimates in these years. An ellipse with a slight "banana" shape appears at each year, though some nonstandard features and differences across years are present. In particular, the estimated density in year 2002 extends down farther to smaller ages and lengths; this year is unique in that it contains a very large proportion of the young fish which are present in the data.

One can envision a curve going through the center of these densities, representing $E(X \mid U^* = u^*, G_t)$, for which we show posterior mean and 95% interval bands for three years in Figure 3. The estimates from our model are compared with the von Bertalanffy growth curves for length-at-age, which are based on a particular function of age and three parameters (estimated here using nonlinear least squares). It is noteworthy that the nonparametric mixture model for the joint distribution of length and age yields estimated growth curves which are overall similar to the von Bertalanffy parametric model, with some local differences especially in year 2002. The uncertainty quantification in the growth curves afforded by the nonparametric model is important, since the attainment of unique growth curves by group (i.e., by location or cohort) is often used to suggest that the groups differ in some way, and this type of analysis should clearly take into account the uncertainty in the estimated curves.

The last year 2007, in addition to containing few observations, is peculiar. There are no



Figure 4: Posterior mean (black lines) and 95% interval estimates (gray shaded regions) for the marginal maturation probability curves associated with length, for six years. Category 1 (immature) is given by the solid line, category 2 (pre-spawning mature) by the dashed line, and category 3 (post-spawning mature) by the dotted line.

fish that are younger than age 6 in this year, and most of the age 6 and 7 fish are recorded as immature, even though in all years combined, less than 10% of age 6 as well as age 7 fish are immature. This year appears to be an anomaly. As there are no observations in 2005 or 2006, and a small number of observations in 2007 which seem to contradict the other years of data, hereinafter, we report inferences only up to 2004.

Inference for the maturation probability curves is shown over length and age in Figures 4 and 5 for six years. The regression curve for maturity as a function of length, $\Pr(Y = j \mid x, G_t)$, can be obtained by marginalizing over w in the expression for $\Pr(Y = j, w, x \mid G_t)$ (which results from (2) by integrating over $Z \in (\gamma_{j-1}, \gamma_j]$) and dividing by $f(x \mid G_t)$. The expression for $\Pr(Y = j \mid w, G_t)$ is computed analogously, where now the marginalization is over length. The probability that a fish is immature is generally decreasing over length, reaching a value near 0 at around 350 mm in most years. There is a large change in this



Figure 5: Posterior mean (black lines) and 95% interval estimates (gray shaded regions) for the marginal maturation probability curves associated with age, for six years. Category 1 (immature) is given by the solid line, category 2 (pre-spawning mature) by the dashed line, and category 3 (post-spawning mature) by the dotted line.

probability over length in 2002 and 2004; these years suggest a probability of immaturity close to 1 for very small fish near 200 to 250 mm. Turning to age, the probability of immaturity is also decreasing with age, also showing differences in 2002 and 2004 in comparison to other years. There is no clear indication of a general trend in the probabilities associated with levels 2 or 3. Years 1995-1997 and 1999 display similar behavior, with a peak in probability of post-spawning mature for moderate length values near 350 mm, and ages 6-7, favoring pre-spawning mature fish at other lengths and ages. The last four years 2001-2004 suggest the probability of pre-spawning is favored most for large fish. Post-spawning appears to have a lower probability than pre-spawning mature for any age at all years, with the exception of 1998, for which the probability of post-spawning mature is high for older fish.

The Pacific States Marine Fisheries Commission states that all Chilipepper rockfish are



Figure 6: Posterior mean and 90% intervals for the smallest value of age above 2 years at which probability of maturity first exceeds 0.9 (left), and similar inference for length (right).

mature at around 4-5 years, and at size 304 to 330 mm. A stock assessment produced by the Pacific Fishery Management Council (Field, 2009) fitted a logistic regression to model maturity over length, from which it appears that 90% of fish are mature around 300-350mm. As our model does not enforce monotonicity on the probability of maturity across age. we obtain posterior distributions for the age at which the probability of maturity exceeds 0.9, given that it exceeds 0.9 at some point. That is, for each posterior sample we evaluate $\Pr(Y > 1 \mid u^*, G_t)$ over a grid in u^* beginning at 2 (since biologically all fish under 2 should be immature), and find the smallest value of u^* at which this probability exceeds 0.9. Note that there were very few posterior samples for which this probability did not exceed 0.9 for any age (only 4 samples in 1993 and 8 in 2003). The estimates for age at 90% maturity are shown in the left panel of Figure 6. The model uncovers a (weak) U-shaped trend across years. Also noteworthy are the very narrow interval bands in 2002. Recall that this year contains an abnormally large number of young fish. In this year, over half of fish age 2 (that is, of age 2-3) are immature, and over 90% of age 3 (that is, of age 3-4) fish are mature, so we would expect the age at 90% maturity to be above 3 but less than 4, which our estimate confirms. A similar analysis is performed for length (right panel of Figure 6) suggesting a trend over time which is consistent with the age analysis.



Figure 7: Left panel: boxplots of the proportion of age = 6 pre-spawning mature fish in the replicated data sets, with width proportional to the number of age 6 fish in each year. Middle panel: boxplots of the proportion of age \geq 7, and length > 400 mm pre-spawning mature fish in the replicated data sets, with width proportional to the number of fish of this age and length in each year. Right panel: boxplots of the sample correlation between length and age for pre-spawning mature fish in the replicated data sets, with width proportional to the number of pre-spawning mature fish in each year. The blue circles in the left and middle panels denote the actual data proportions, and in the right panel the data-based correlation.

3.3 Model Checking

Here, we discuss results from posterior predictive model checking. In particular, we generated replicate data sets from the posterior predictive distribution, and compared to the real data using specific test quantities (Gelman et al., 2004). We avoid using posterior predictive p–values because these predictive probabilities are not well calibrated, tending to have non-uniform distributions under the null hypothesis that the model is correct (e.g., Bayarri and Berger, 2000; Gelman, 2013).

As an illustration, Figure 7 shows results for years 1993 to 2004 (excluding year 2003 for which data is not available) and for three test quantities: the proportion of age 6 pre-spawning mature fish; the proportion of pre-spawning mature fish of at least 7 years of age and length larger than 400 mm; and the sample correlation between length and age for pre-spawning mature fish. The results, a subset of which is shown in Figure 7, suggest that the model is predicting data which is similar to the observed data in terms of practically important inferences.

Although results are not shown here, we also studied residuals with cross-validation, randomly selecting 20% of the observations in each year and refitting the model, leaving out these observations. We obtained residuals $\tilde{y}_{ti} - E(Y | W = \tilde{w}_{ti}, X = \tilde{x}_{ti}, G_t)$ for each observation $(\tilde{y}_{ti}, \tilde{w}_{ti}, \tilde{x}_{ti})$ which was left out. There was no apparent trend in the residuals across covariate values, that is, no indication that we are systematically under or overestimating fish maturity of a particular length and/or age.

Allowing for both time-dependent weights and atoms proved to be important for the application, as evidenced by formal comparison with the common atoms model. Specifically, the posterior predictive criterion of Gelfand and Ghosh (1998) was calculated for each year of data between 1993 and 2004, and indicated that the more general DDP model provided a better fit to the data at most years. Refer to the Supplementary Material for more details.

4 Discussion

The methods developed for dynamic ordinal regression are widely applicable to modeling mixed ordinal-continuous distributions indexed in discrete time. At any particular point in time, the DP mixture representation for the latent response-covariate distribution is retained, enabling flexible inference for a variety of functionals, and allowing standard posterior simulation techniques for DP mixture models to be utilized.

In contrast to standard approaches to ordinal regression, the model does not force specific trends, such as monotonicity, in the regression functions. We view this as an attribute in most settings. Nevertheless, in situations in which it is believed that monotonicity exists, we must realize that the data will determine the model output, and may not produce strictly monotonic relationships. Referring to the fish maturity application, it is generally accepted that monotonicity exists in the probability of maturity as a function of age or length. Although our model does not enforce this, the inferences generally agree with what is expected to be true biologically. Specifically, the probability that a fish is immature is generally decreasing over length and age. Our model is also relevant to this setting, as the covariates age and length are treated as random, and the ordinal nature of recorded age is accounted for using variables which represent underlying continuous age. The set of inferences that are provided under this framework, including estimates for length as a function of age, make this modeling approach powerful for the particular application considered, as well as related problems.

While year of sampling was considered to be the index of dependence in this analysis, an alternative is to consider cohort as an index of dependence. All fish born in the same year, or the same age in a given year, represent one cohort. Grouping fish by cohort rather than year of record should lead to more homogeneity within a group, however there are also some possible issues since fish will generally be younger as cohort index increases. This is a consequence of having a particular set of years for which data is collected, i.e., the cohort of fish born in 2006 can not be older than 4 if data collection stopped in 2009. Due to complications such as these, combined with exploration of the relationships within each cohort, we decided to treat year of data collection as the index of dependence, but cohort indexing could be more appropriate in other analyses of similar data structures.

The proposed modeling approach could also be useful in applications in finance. One such example arises in the analysis of price changes of stocks. In the past, stocks traded on the New York Stock Exchange were priced in eighths, later moved to sixteenths, and corporate bonds still trade in eighths. In analyzing price changes of stocks which are traded in fractions, it is inappropriate to treat the measurements as continuous, particularly if the range of values is not very large (e.g., Müller and Czado, 2009). The price changes should be treated with a discrete response model, and the possible responses are ordered, ranging from a large negative return to a large positive return. One possible analysis may involve modeling the monthly returns as a function of covariates, such as trade volume, taking into account the ordinal nature of the responses. In addition, the distribution of returns in a particular month is likely correlated with the previous month, and the regression relationships must be allowed to be related from one month to the next. In finance as well as environmental science, empirical distributions may exhibit non-standard features which require more general methods, such as the nonparametric mixture model developed here.

Appendix: Properties of the DDP Prior Model

Here we provide derivations of various correlations associated with the DDP prior model presented in Section 2.3.

Autocorrelation of (β_t, β_{t+k})

First, consider the correlation of the beta random variables used to define the dynamic stickbreaking weights. Let $\rho_k = \operatorname{corr}(\eta_t, \eta_{t+k})$, which is equal to ϕ^k under the assumption of an AR(1) process for $\eta_{\mathcal{T}}$. The autocorrelation function associated with \mathcal{B} is

$$\operatorname{corr}(\beta_t, \beta_{t+k} \mid \alpha, \phi) = \frac{\alpha^{1/2} (1 - \rho_k^2)^{1/2} (\alpha + 1)^2 (\alpha + 2)^{1/2}}{\left\{ (1 - \rho_k^2 + \alpha)^2 - \alpha^2 \rho_k^2 \right\}^{1/2}} - \alpha(\alpha + 2).$$
(A.1)

The expectations required to derive expression (A.1) are obtained as follows. Since the process is stationary with $\beta_t \sim \text{beta}(\alpha, 1)$ at any time t, $E(\beta_t \mid \alpha) = \alpha/(\alpha + 1)$ and $\text{var}(\beta_t \mid \alpha) = \alpha/\{(\alpha + 1)^2(\alpha + 2)\}$. Also, using the definition of the \mathcal{B} process in (4),

$$\mathbf{E}(\beta_t \beta_{t+k} \mid \alpha, \phi) = \mathbf{E}\left\{\exp(-\zeta^2/\alpha)\right\} \mathbf{E}\left\{\exp(-(\eta_t^2 + \eta_{t+k}^2)/2\alpha)\right\}.$$
 (A.2)

The first expectation can be obtained from the moment generating function of $\zeta^2 \sim \chi_1^2$, given by $\mathrm{E}(e^{t\zeta^2}) = (1-2t)^{-1/2}$, for t < 1/2. Hence, for $t = -1/\alpha$, we obtain $\mathrm{E}\left\{\exp(-\zeta^2/\alpha)\right\} = \alpha^{1/2}/(2+\alpha)^{1/2}$. Regarding the second expectation, note that $(\eta_t, \eta_{t+k}) \sim \mathrm{N}(0, C_k)$, with C_k a covariance matrix with diagonal elements equal to 1 and off-diagonal element equal to ρ_k . Integration results in $\mathrm{E}\left\{\exp(-(\eta_t^2+\eta_{t+k}^2)/2\alpha)\right\} = \alpha(1-\rho_k^2)^{1/2}/\{(1-\rho_k^2+\alpha)^2-\alpha^2\rho_k^2\}^{1/2}$.

Figure 8 plots the autocorrelation function in (A.1), for k ranging from 1 to 50, and for various values of α and ϕ . Smaller values for α lead to smaller correlations for any fixed ϕ at a particular lag, and ϕ controls the strength of correlation, with large ϕ producing large correlations which decay slowly. Parameters ϕ and α combined can lead to a range of correlations, however $\alpha \geq 1$ implies a lower bound near 0.5 for any lag k. In the limit, as $\alpha \rightarrow 0^+$, $\operatorname{corr}(\beta_t, \beta_{t+k} \mid \alpha, \phi) \rightarrow 0$, and as $\alpha \rightarrow \infty$, $\operatorname{corr}(\beta_t, \beta_{t+k} \mid \alpha, \phi)$ tends towards 0.5 as



Figure 8: Autocorrelation function $\operatorname{corr}(\beta_t, \beta_{t+k} \mid \alpha, \phi)$ for a range of α values (indicated in the title of each panel) and values of ϕ of 0.99 (solid lines), 0.9 (dashed lines), 0.5 (dotted lines), and 0.3 (dashed/dotted lines). The lag k on the x-axis runs from 1 to 50.

 $\rho_k \to 0^+, \text{ and } 1 \text{ as } \rho_k \to 1^-.$ For $\rho_k = \phi^k$, we obtain $\lim_{\phi \to 1^-} \operatorname{corr}(\beta_t, \beta_{t+k} \mid \alpha, \phi) = 1$ and $\lim_{\phi \to 0^+} \operatorname{corr}(\beta_t, \beta_{t+k} \mid \alpha, \phi) = \alpha^{1/2} (\alpha+1)(\alpha+2)^{1/2} - \alpha(\alpha+2)$, which tends to 0.5 as $\alpha \to \infty$.

Autocorrelation of DDP weights

We next study the dependence induced in the DDP weights at consecutive time points. First, $E(p_{l,t} \mid \alpha) = E\{(1 - \beta_{l,t}) \prod_{r=1}^{l-1} \beta_{r,t} \mid \alpha\}$. Since the $\beta_{l,t}$ are independent across l, and $E(\beta_{l,t} \mid \alpha) = \alpha/(\alpha + 1)$, we obtain $E(p_{l,t} \mid \alpha) = \alpha^{l-1}/(1 + \alpha)^{l}$. Similarly, $E(p_{l,t}^{2} \mid \alpha) = E\{(1 - \beta_{l,t})^{2} \mid \alpha\} \prod_{r=1}^{l-1} E(\beta_{r,t}^{2} \mid \alpha) = 2\alpha^{l-1}/\{(\alpha + 1)(\alpha + 2)^{l}\}$, from which $var(p_{l,t} \mid \alpha)$ obtains.

Since $p_{l,t}p_{l,t+1} = (1 - \beta_{l,t})(1 - \beta_{l,t+1}) \prod_{r=1}^{l-1} \beta_{r,t}\beta_{r,t+1}$, and $(\beta_{l,t}, \beta_{l,t+1})$ is independent of $(\beta_{m,t}, \beta_{m,t+1})$, for any $l \neq m$, we can write $\mathbb{E}(p_{l,t}p_{l,t+1} \mid \alpha, \phi) = \mathbb{E}\{(1 - \beta_{l,t})(1 - \beta_{l,t+1}) \mid \alpha, \phi\} \prod_{r=1}^{l-1} \mathbb{E}(\beta_{r,t}\beta_{r,t+1} \mid \alpha, \phi)$. The required expectations in the above equation can be obtained from (A.2) for k = 1, such that $\rho_1 = \phi$. Combining these expressions yields the



Figure 9: Autocorrelation function $\operatorname{corr}(p_{l,t}, p_{l,t+1} \mid \alpha, \phi)$ for a range of α values (indicated in the title of each panel) and values of ϕ of 0.99 (solid lines), 0.9 (dashed lines), 0.5 (dotted lines), and 0.3 (dashed/dotted lines). The weight index l on the x-axis runs from 1 to 100.

covariance

$$\operatorname{cov}(p_{l,t}, p_{l,t+1} \mid \alpha, \phi) = \left\{ \frac{\alpha^{3/2} (1 - \phi^2)^{1/2}}{(2 + \alpha)^{1/2} \left\{ (1 - \phi^2 + \alpha)^2 - \alpha^2 \phi^2 \right\}^{1/2}} \right\}^{l-1} \\ \left\{ 1 - \frac{2\alpha}{\alpha + 1} + \frac{\alpha^{3/2} (1 - \phi^2)^{1/2}}{(2 + \alpha)^{1/2} \left\{ (1 - \phi^2 + \alpha)^2 - \alpha^2 \phi^2 \right\}^{1/2}} \right\} - \frac{\alpha^{2l-2}}{(1 + \alpha)^{2l}}.$$
(A.3)

This can be divided by $\operatorname{var}(p_{l,t} \mid \alpha) = \{2\alpha^{l-1}/((1+\alpha)(2+\alpha)^l)\} - \{\alpha^{2l-2}/(1+\alpha)^{2l}\}$ to yield $\operatorname{corr}(p_{l,t}, p_{l,t+1} \mid \alpha, \phi)$; note that $\operatorname{E}(p_{l,t} \mid \alpha) = \operatorname{E}(p_{l,t+1} \mid \alpha)$ and $\operatorname{var}(p_{l,t} \mid \alpha) = \operatorname{var}(p_{l,t+1} \mid \alpha)$.

Figure 9 displays $\operatorname{corr}(p_{l,t}, p_{l,t+1} \mid \alpha, \phi)$ for different values of α and ϕ . The correlations are decreasing in index l, and larger values of ϕ lead to larger correlations in the weights at any particular l. Moreover, the decay in correlations with weight index is faster for small α and small ϕ . As $\alpha \to 0^+$, $\operatorname{corr}(p_{1,t}, p_{1,t+1} \mid \alpha, \phi) \to 1$ for any value of ϕ , and as $\alpha \to \infty$, $\operatorname{corr}(p_{1,t}, p_{1,t+1} \mid \alpha, \phi)$ is contained in (0.5, 1), with values closer to 1 for larger ϕ . Note that $\operatorname{corr}(p_{l,t}, p_{l,t+k} \mid \alpha, \phi)$ has the same expression as $\operatorname{corr}(p_{l,t}, p_{l,t+1} \mid \alpha, \phi)$, but with ϕ replaced by ϕ^k ; it is thus decreasing with the lag k, with the speed of decay controlled by ϕ .

Autocorrelation of consecutive distributions

Denote by S the support of distributions $G_t = \sum_{l=1}^{\infty} p_{l,t} \delta_{\boldsymbol{\theta}_{l,t}}$ (say, $\mathbb{S} \subseteq \mathbb{R}^M$) and consider a specific measurable set $A \subset \mathbb{S}$. We study the correlation between $G_t(A)$ and $G_{t+1}(A)$ for the general DDP prior based on the model of Section 2.3 for the weights and a generic time series model $G_{0\mathcal{T}} = \{G_{0,t} : t \in \mathcal{T}\}$ for the atoms $\boldsymbol{\theta}_{l,t}$, as well as for the common atoms DDP prior.

The DDP prior model implies at any t a $DP(\alpha, G_{0,t})$ prior for G_t . Hence, we have $E(G_t(A) \mid \alpha, G_{0T}) = G_{0,t}(A)$, and $var(G_t(A) \mid \alpha, G_{0T}) = G_{0,t}(A)(1 - G_{0,t}(A))/(\alpha + 1)$.

The additional expectation needed in order to obtain $\operatorname{corr}(G_t(A), G_{t+1}(A) \mid \alpha, \phi, G_{0\mathcal{T}})$ is

$$E(G_t(A)G_{t+1}(A) \mid \alpha, \phi, G_{0\mathcal{T}}) = E\left(\sum_{l=1}^{\infty} p_{l,t}p_{l,t+1}\delta_{\boldsymbol{\theta}_{l,t}}(A)\delta_{\boldsymbol{\theta}_{l,t+1}}(A) \mid \alpha, \phi, G_{0\mathcal{T}}\right) + \\ E\left(\sum_{l=1}^{\infty} \sum_{m \neq l} p_{l,t}p_{m,t+1}\delta_{\boldsymbol{\theta}_{l,t}}(A)\delta_{\boldsymbol{\theta}_{m,t+1}}(A) \mid \alpha, \phi, G_{0\mathcal{T}}\right).$$

The first expectation is equal to $G_{0,t,t+1}(A) \sum_{l=1}^{\infty} E(p_{l,t}p_{l,t+1} \mid \alpha, \phi)$, where $G_{0,t,t+1}(A) = \Pr(\boldsymbol{\theta}_{l,t} \in A, \boldsymbol{\theta}_{l,t+1} \in A \mid G_{0\mathcal{T}})$, and the term $E(p_{l,t}p_{l,t+1} \mid \alpha, \phi) = H_l(\alpha, \phi)$ is given as part of expression (A.3). The second expectation becomes $G_{0,t}(A)G_{0,t+1}(A) \sum_{l=1}^{\infty} \sum_{m\neq l} E(p_{l,t}p_{m,t+1} \mid \alpha, \phi)$. Here, we have used the independence between the random variables defining the weights and the atoms, as well as the independence between $\boldsymbol{\theta}_{l,t}$ and $\boldsymbol{\theta}_{m,t+1}$, for $m \neq l$.

The final step of the derivation therefore involves $\mathcal{E}(p_{l,t}p_{m,t+1} \mid \alpha, \phi)$, for $m \neq l$. Note that, if l < m, $p_{l,t}p_{m,t+1} = (\prod_{r=1}^{l-1} \beta_{r,t}\beta_{r,t+1}) \beta_{l,t+1}(1 - \beta_{l,t}) (\prod_{r=l+1}^{m-1} \beta_{r,t+1}) (1 - \beta_{m,t+1})$, and thus $D_{l < m}(\alpha, \phi) \equiv \mathcal{E}(p_{l,t}p_{m,t+1} \mid \alpha, \phi) = C(\alpha, \phi)^{l-1} \{\alpha(\alpha+1)^{-1} - C(\alpha, \phi)\} \alpha^{m-l-1}(\alpha+1)^{l-m}$, where $C(\alpha, \phi) = \mathcal{E}(\beta_{r,t}\beta_{r,t+1} \mid \alpha, \phi)$. Analogously, for l > m, $D_{l > m}(\alpha, \phi) \equiv \mathcal{E}(p_{l,t}p_{m,t+1} \mid \alpha, \phi) = C(\alpha, \phi)^{m-1} \{\alpha(\alpha+1)^{-1} - C(\alpha, \phi)\} \alpha^{l-m-1}(\alpha+1)^{m-l}$.

Putting this all together, $E(G_t(A)G_{t+1}(A) \mid \alpha, \phi, G_{0\mathcal{T}})$ is equal to

$$G_{0,t,t+1}(A)\sum_{l=1}^{\infty}H_{l}(\alpha,\phi) + G_{0,t}(A)G_{0,t+1}(A)\sum_{l=1}^{\infty}\left(\sum_{mm}(\alpha,\phi) + \sum_{m>l}D_{l< m}(\alpha,\phi)\right),$$

and $\operatorname{corr}(G_t(A), G_{t+1}(A) \mid \alpha, \phi, G_{0\mathcal{T}})$ is equal to

$$\left(\frac{\alpha+1}{\{G_{0,t}(A)G_{0,t+1}(A)(1-G_{0,t}(A))(1-G_{0,t+1}(A))\}^{1/2}}\right)\times G_{0,t,t+1}(A)\sum_{l=1}^{\infty}H_{l}(\alpha,\phi)+G_{0,t}(A)G_{0,t+1}(A)\left\{-1+\sum_{l=1}^{\infty}\left(\sum_{mm}(\alpha,\phi)+\sum_{m>l}D_{l< m}(\alpha,\phi)\right)\right\}.$$
(A.4)

Under the simplified version of the model with common atoms, where $\boldsymbol{\theta}_{l,t} \equiv \boldsymbol{\theta}_l$ arise i.i.d. from G_0 , corr $(G_t(A), G_{t+1}(A) \mid \alpha, \phi, G_0)$ becomes

$$\frac{(\alpha+1)\left\{-G_0(A) + \sum_{l=1}^{\infty} H_l(\alpha,\phi) + G_0(A) \sum_{l=1}^{\infty} \left(\sum_{m < l} D_{l > m}(\alpha,\phi) + \sum_{m > l} D_{l < m}(\alpha,\phi)\right)\right\}}{1 - G_0(A)}.$$
(A.5)

The complex nature of expressions (A.4) and (A.5) does not allow ready identification of correlation trends in terms of the model parameters. Moreover, since the expressions are obtained through differences of expectations, they are numerically unstable to compute for certain combinations of parameter values. However, we can alternatively estimate the correlation through the sample correlation of a set of draws from the prior distribution of $(G_t(A), G_{t+1}(A))$. We discuss next some results from this prior simulation approach.

We first consider the common atoms simplification of the model. To focus on the effects of parameters α and ϕ , we take $\mathbb{S} = \mathbb{R}$ and set $G_0 = \mathcal{N}(0,1)$, with A = (-10,0), which fixes $G_0(A) \approx 0.5$. Note that we did not observe a significant effect in the correlation patterns for different sets A. Figure 10 displays $\operatorname{corr}(G_t(A), G_{t+1}(A) \mid \alpha, \phi)$ for different values of α and ϕ . The correlations are increasing with ϕ for fixed α , and decreasing with α for fixed ϕ . However, for practically all combinations of α and ϕ values, the correlation is fairly large as a consequence of the common atoms restriction.

Next, consider the more general model, involving time-dependent atoms $\theta_{l,t}$. Simulating from the prior requires sampling $(\theta_{l,t}, \theta_{l,t+1})$ from the bivariate distribution implied by the



Figure 10: Correlations between $G_t(A)$ and $G_{t+1}(A)$, for A = (-10, 0), under the common atoms DDP prior with $G_0 = N(0, 1)$. Left panel: Correlations as a function of ϕ for $\alpha = 0.5$ (circles), $\alpha = 1$ (triangles), and $\alpha = 4$ (+ symbols). Right panel: Correlations as a function of α for $\phi = 0.1$ (circles), $\phi = 0.5$ (triangles), and $\phi = 0.9$ (+ symbols).

time series model for $G_{0\mathcal{T}}$. Under the particular model for the rockfish data application, $\boldsymbol{\theta}_{l,t} = (\boldsymbol{\mu}_{l,t}, \boldsymbol{\Sigma}_l)$. Evidently, it is simpler to study the correlation of consecutive distributions under a stationary model for the $\boldsymbol{\mu}_{l,\mathcal{T}}$. The VAR process we use to define the $\boldsymbol{\mu}_{l,\mathcal{T}}$ is a stationary model under a diagonal matrix Θ with diagonal elements taking values in (-1, 1).

To facilitate comparison with the common atoms simplification and focus on the effect of the AR coefficient, we take $\mathbb{S} = \mathbb{R}$ and $\theta_{l,t} = \mu_{l,t}$. In this case, the bivariate stationary distribution for $(\theta_{l,t}, \theta_{l,t+1})$ under the univariate version of the autoregressive model in (3) is bivariate normal, with means $m/(1-\Theta)$, variances $V/(1-\Theta^2)$, and covariance $\Theta V/(1-\Theta^2)$. We fix m = 0, V = 1, and A = (-10,0), and focus on the effect of the AR parameter Θ . Figure 11 displays correlations over Θ for different ϕ and α values. The correlations are increasing with Θ , and increasing to a lesser degree with ϕ . In contrast to the common atoms model, different combinations of Θ , α and ϕ values result in correlations that span the entire interval (0, 1).



Figure 11: Correlations between $G_t(A)$ and $G_{t+1}(A)$, for A = (-10, 0), under the general DDP prior with a stationary AR model for the atoms with AR parameter Θ and marginal distribution $G_0 = N(0, 1)$. The left plot fixes $\alpha = 0.5$, and the right fixes $\alpha = 4$. In both plots, correlations are shown for $\phi = 0.1$ (circles), $\phi = 0.5$ (triangles), and $\phi = 0.9$ (+ symbols).

References

- Bassetti, F., Casarin, R., and Leisen, F. (2014), "Beta-product dependent Pitman-Yor processes for Bayesian inference," *Journal of Econometrics*, 180, 49 72.
- Bayarri, S. and Berger, J. (2000), "P-values for Composite Null Models," Journal of the American Statistical Association, 95, 1127–1142.
- Bobko, S. and Berkeley, S. (2004), "Maturity, ovarian cycle, fecundity, and age-specific parturition of black rockfish (Sebastes melanops)," *Fisheries Bulletin*, 102, 418–429.
- Boes, S. and Winkelmann, R. (2006), "Ordered Response Models," Advances in Statistical Analysis, 90, 165–179.
- Clark, W. (1991), "Groundfish exploitation rates based on life history parameters," Canadian Journal of Fisheries and Aquatic Sciences, 48, 734–750.
- De Iorio, M., Müller, P., Rosner, G., and MacEachern, S. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205– 215.
- DeIorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009), "Bayesian nonparametric non-proportional hazards survival modelling," *Biometrics*, 63, 762–771.
- DeYoreo, M. (2014), "A Bayesian Framework for Fully Nonparametric Ordinal Regression," Ph.D. thesis, University of California, Santa Cruz.
- DeYoreo, M. and Kottas, A. (2015), "A Fully Nonparametric Modeling Approach to Binary Regression," *Bayesian Analysis*, 10, 821–847.

- (2017), "Bayesian Nonparametric Modeling for Multivariate Ordinal Regression," Journal of Computational and Graphical Statistics, To appear.
- Di Lucca, M., Guglielmi, A., Muller, P., and Quintana, F. (2013), "Bayesian Dynamic Density Estimation," A Simple Class of Bayesian Nonparametric Autoregressive Models, 8, 63–88.
- Dunson, D. and Park, J. (2008), "Kernel Stick-breaking Processes," Biometrika, 95, 307–323.
- Field, J. (2009), "Status of the Chilipepper rockfish, Sebastes Goodei, in 2007," Tech. rep., Southwest Fisheries Science Center.
- Fronczyk, K. and Kottas, A. (2014a), "A Bayesian approach to the analysis of quantal bioassay studies using nonparametric mixture models," *Biometrics*, 70, 95–102.
- (2014b), "A Bayesian Nonparametric Modeling Framework for Developmental Toxicity Studies (with discussion)," Journal of the American Statistical Association, 109, 873–893.
- Gelfand, A. and Ghosh, S. (1998), "Model choice: A minimum posterior predictive loss approach," *Biometrika*, 85, 1–11.
- Gelfand, A., Kottas, A., and MacEachern, S. (2005), "Bayesian Nonparametric Spatial Modeling with Dirichlet Process Mixing," *Journal of the American Statistical Association*, 100, 1021–1035.
- Gelman, A. (2013), "Two simple examples for understanding posterior p-values whose distributions are far from uniform," *Electronical Journal of Statistics*, 7, 2595–2602.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004), Bayesian Data Analysis, Chapman and Hall/CRC.
- Griffin, J. and Steel, M. (2006), "Order-based Dependent Dirichlet Processes," Journal of the American Statistical Association, 101, 179–194.
- (2011), "Stick-breaking autoregressive processes," Journal of Econometrics, 162, 383 396.
- Griffin, J. E., Kolossiatis, M., and Steel, M. F. J. (2013), "Comparing distributions by using dependent normalized random-measure mixtures," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 499–529.
- Hannah, R., Blume, M., and Thompson, J. (2009), "Length and age at maturity of female yelloweye rockfish (Sebastes rubberimus) and cabezon (Scorpaenichthys marmoratus) from Oregon waters based on histological evaluation of maturity," Tech. rep., Oregon Department of Fish and Wildlife.
- Ishwaran, H. and James, L. (2001), "Gibbs Sampling Methods for Stick-breaking Priors," Journal of the American Statistical Association, 96, 161–173.
- Kalli, M., Griffin, J., and Walker, S. (2011), "Slice sampling mixture models," Statistics and Computing, 21, 93–105.

- Kottas, A., Wang, Z., and Rodríguez, A. (2012), "Spatial modeling for risk assessment of extreme values from environmental time series: A Bayesian nonparametric approach," *Environmetrics*, 23, 649–662.
- Leisen, F. and Lijoi, A. (2011), "Vectors of two-parameter Poisson-Dirichlet processes," Journal of Multivariate Analysis, 102, 482 – 495.
- Lijoi, A., Nipoti, B., and Prünster, I. (2014), "Bayesian inference with dependent normalized completely random measures," *Bernoulli*, 20, 1260–1291.
- MacEachern, S. (1999), "Dependent Nonparametric Processes," in ASA Proceedings of the Section on Bayesian Statistical Science, pp. 50–55.
- (2000), "Dependent Dirichlet processes," Tech. rep., The Ohio State University Department of Statistics.
- McKenzie, E. (1985), "An Autoregressive Process for Beta Random Variables," Management Science, 31, 988–997.
- Morgan, M. and Hoenig, J. (1997), "Estimating Maturity-at-Age from Length Stratified Sampling," Journal of Northwest Atlantic Fisheries Science, 21, 51–63.
- Müller, G. and Czado, C. (2009), "Stochastic Volatility Models for Ordinal Valued Time Series with Application to Finance," *Statistical Modeling*, 9, 69–95.
- Müller, P., Erkanli, A., and West, M. (1996), "Bayesian Curve Fitting Using Multivariate Normal Mixtures," *Biometrika*, 83, 67–79.
- Nieto-Barajas, L., Müller, P., Ji, Y., Lu, Y., and Mills, G. (2012), "A Time-Series DDP for Functional Proteomics Profiles," *Biometrics*, 68, 859–868.
- Rodriguez, A. and Dunson, D. (2011), "Nonparametric Bayesian models through probit stickbreaking processes," *Bayesian Analysis*, 6, 145–177.
- Rodriguez, A. and ter Horst, E. (2008), "Bayesian Dynamic Density Estimation," *Bayesian Analysis*, 3, 339–366.
- Rubin, D. (1976), "Inference and Missing Data," Biometrika, 63, 581–592.
- Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," Statistica Sinica, 4, 639–650.
- Taddy, M. (2010), "Autoregressive Mixture Models for Dynamic Spatial Poisson Processes: Application to Tracking the Intensity of Violent Crime," *Journal of the American Statistical Association*, 105, 1403–1417.
- Xiao, S., Kottas, A., and Sansó, B. (2015), "Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences," *The Annals of Applied Statistics*, 9, 353–382.
- Zhu, W. and Leisen, F. (2015), "A multivariate extension of a vector of two-parameter Poisson-Dirichlet processes," *Journal of Nonparametric Statistics*, 27, 89–105.

Supplementary Material for "Modeling for Dynamic Ordinal Regression Relationships: An Application to Estimating Maturity of Rockfish in California"

Maria DeYoreo and Athanasios Kottas

Section 1 discusses posterior inference under the DDP mixture model, as well as prior specification and sensitivity to the prior choice. Details on model comparison for the common atoms DDP model versus the more general model are presented in Section 2.

1 Prior Specification and Posterior Simulation Details

1.1 MCMC Posterior Simulation Method

We implement posterior simulation using blocked Gibbs sampling (Ishwaran and James, 2001), which involves truncation of the countable representation for G_t to a finite level N. Hence, the stick-breaking weights are given by $p_{1,t} = 1 - \beta_{1,t}$, $p_{l,t} = (1 - \beta_{l,t}) \prod_{r=1}^{l-1} \beta_{r,t}$, for $l = 2, \ldots, N-1$, and $p_{N,t} = \prod_{l=1}^{N-1} \beta_{l,t}$, ensuring $\sum_{l=1}^{N} p_{l,t} = 1$. Since α is not a function of t, the same truncation level is applied for all mixing distributions. In choosing the truncation level, we use the prior expectation for the sum of the first N DP stick-breaking weights w_1, \ldots, w_N , that is, $E(\sum_{j=1}^{N} w_j \mid \alpha) = 1 - \{\alpha/(\alpha+1)\}^N$. This can be averaged over the prior for α to obtain $E(\sum_{j=1}^{N} w_j)$, with N chosen such that this expectation is close to 1 up to the desired level of tolerance for the approximation. For the rockfish data application, we used N = 50, which corresponds to $E(\sum_{j=1}^{N} w_j) > 0.999999$, under the prior for α given below.

Regarding the model hyperparameters, α and ϕ are given priors IG (a_{α}, b_{α}) and uniform on (0, 1), respectively. We choose $a_{\alpha} = 5$ and $b_{\alpha} = 3$ to favor small values for α , hence relatively

few clusters. We assume that Θ is diagonal, with elements $(\theta_1, \theta_2, \theta_3)$, however we advocate for a full covariance matrix V. We assume uniform priors on (0, 1) for each element of Θ and for ϕ . We use conditionally conjugate priors for the hyperparameters $\boldsymbol{m}, \boldsymbol{V}$ and \boldsymbol{D} , given by $\boldsymbol{m} \sim N(\boldsymbol{a}_{\boldsymbol{m}}, \boldsymbol{B}_{\boldsymbol{m}}), \boldsymbol{V} \sim IW(\boldsymbol{a}_{\boldsymbol{V}}, \boldsymbol{B}_{\boldsymbol{V}}), \text{ and } \boldsymbol{D} \sim W(\boldsymbol{a}_{\boldsymbol{D}}, \boldsymbol{B}_{\boldsymbol{D}})$. The inverse-Wishart density for \boldsymbol{V} is proportional to $|\boldsymbol{V}|^{-(a_{\boldsymbol{V}}+p+2)/2} \exp\{-0.5 \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{V}}\boldsymbol{V}^{-1})\}$, and the Wishart density for \boldsymbol{D} is proportional to $|\boldsymbol{D}|^{(a_{\boldsymbol{D}}-p-2)/2} \exp\{-0.5 \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{D}}^{-1}\boldsymbol{D})\}$.

The hierarchical model for the data is written by introducing a set of mixture configuration variables $L_{t,i}$, which indicate the mixture component for observation i at time t. Conditional on $L_{t,i}$, the mixture model given in Equation (2) of the paper implies a normal distribution for $\boldsymbol{y}_{ti}^* = (z_{ti}, w_{ti}, x_{ti})$ with mean vector $\boldsymbol{\mu}_{L_{t,i},t}$ and covariance matrix $\boldsymbol{\Sigma}_{L_{t,i}}$. The posterior full conditional for each $L_{t,i}$ is a discrete distribution on $\{1, ..., N\}$ with probabilities proportional to $p_{l,t} N(\boldsymbol{y}_{ti}^* \mid \boldsymbol{\mu}_{l,t}, \boldsymbol{\Sigma}_{l})$.

Since $(z_{ti}, w_{ti}, x_{ti}) | L_{t,i}, (\boldsymbol{\mu}_{L_{t,i},t}, \boldsymbol{\Sigma}_{L_{t,i}}) \sim N(\boldsymbol{\mu}_{L_{t,i},t}, \boldsymbol{\Sigma}_{L_{t,i}})$, the posterior full conditional for each latent continuous z_{ti} is proportional to the conditional normal distribution for $z_{ti} | (w_{ti}, x_{ti})$, with lower and upper truncation points given by $\gamma_{y_{ti}-1}$ and $\gamma_{y_{ti}}$. The full conditionals for the latent continuous w_{ti} follow analogously, and each latent continuous random variable is sampled from a truncated normal distribution.

The conjugate priors for the hyperparameters m, V and D yield closed-form full conditional distributions via standard conjugate updating. Below we derive the posterior full conditionals and provide updating strategies for the remaining model parameters that are more complicated to sample and have not yet been described.

Updating the weights

The full conditional for $(\{\zeta_l\}, \{\eta_{l,t}\})$ is given by $p(\{\zeta_l\}, \{\eta_{l,t}\} \mid \dots, \text{data}) \propto$

$$\prod_{l=1}^{N-1} \mathcal{N}(\zeta_l \mid 0, 1) \mathcal{N}(\eta_{l,1} \mid 0, 1) \prod_{t=2}^{T} \prod_{l=1}^{N-1} \mathcal{N}(\eta_{l,t} \mid \phi \eta_{l,t-1}, 1-\phi^2) \prod_{t=1}^{T} \prod_{i=1}^{n_t} \sum_{l=1}^{N} p_{l,t} \delta_l(L_{t,i}).$$

Write $\prod_{i=1}^{n_t} \sum_{l=1}^{N} p_{l,t} \delta_l(L_{t,i}) = \prod_{l=1}^{N} p_{l,t}^{M_{l,t}}$, where $M_{l,t} = |\{(t,i) : L_{t,i} = l\}|$, i.e., the number of observations at time t assigned to component l. Filling in the form for $\{p_{l,t}\}$ gives

$$\prod_{i=1}^{n_t} \sum_{l=1}^{N} p_{l,t} \delta_l(L_{t,i}) = \left(1 - \exp\left(-\frac{\zeta_1^2 + \eta_{1,t}^2}{2\alpha}\right)\right)^{M_{1,t}} \exp\left(-\frac{M_{N,t} \sum_{l=1}^{N-1} (\zeta_l^2 + \eta_{l,t}^2)}{2\alpha}\right)$$
$$\prod_{l=2}^{N-1} \left\{ \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,t}^2}{2\alpha}\right)\right)^{M_{l,t}} \exp\left(-\frac{M_{l,t} \sum_{r=1}^{l-1} (\zeta_r^2 + \eta_{r,t}^2)}{2\alpha}\right)\right\}.$$

The full conditional for each ζ_l , $l = 1, \ldots, N - 1$, is therefore

$$p(\zeta_l \mid \dots, \text{data}) \propto \exp\left(-\frac{\zeta_l^2}{2}\right) \exp\left(\frac{-\zeta_l^2 \sum_{t=1}^T \sum_{r=l+1}^N M_{r,t}}{2\alpha}\right) \prod_{t=1}^T \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,t}^2}{2\alpha}\right)\right)^{M_{l,t}}$$

giving

$$p(\zeta_l \mid \dots, \text{data}) \propto N(\zeta_l \mid 0, (1 + \alpha^{-1} \sum_{t=1}^T \sum_{r=l+1}^N M_{r,t})^{-1}) \prod_{t=1}^T \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,t}^2}{2\alpha}\right)\right)^{M_{l,t}}$$

We use a slice sampler to update ζ_l , with the following steps:

- Draw $u_t \sim \text{uniform}\left(0, \left(1 \exp\left(-\frac{\zeta_l^2 + \eta_{l,t}^2}{2\alpha}\right)\right)^{M_{l,t}}\right), \text{ for } t = 1, \dots, T.$
- Draw $\zeta_l \sim \mathcal{N}(0, (1 + \alpha^{-1} \sum_{t=1}^T \sum_{r=l+1}^N M_{r,t})^{-1})$, restricted to the lie in the interval $\left\{\zeta_l : u_t < \left(1 \exp\left(-\frac{\zeta_l^2 + \eta_{l,t}^2}{2\alpha}\right)\right)^{M_{l,t}}, t = 1, \ldots, T\right\}$. Solving for ζ_l in each of these T equations gives $\zeta_l^2 > -\eta_{l,t}^2 2\alpha \log(1 u_t^{1/M_{l,t}})$, for $t = 1, \ldots, T$. Therefore, if $-\eta_{l,t}^2 2\alpha \log(1 u_t^{1/M_{l,t}}) < 0$ for all t, then ζ_l has no restrictions, and is therefore sampled from a normal distribution. Otherwise, if $-\eta_{l,t}^2 2\alpha \log(1 u_t^{1/M_{l,t}}) > 0$ for some t, then $|\zeta_l| > \max_t \{(-\eta_{l,t}^2 2\alpha \log(1 u_t^{1/M_{l,t}}))^{1/2}\}$. This then requires sampling ζ_l from a normal distribution, restricted to the intervals $(-\infty, -\max_t\{(-\eta_{l,t}^2 2\alpha \log(1 u_t^{1/M_{l,t}}))^{1/2}\})$, and $(\max_t\{(-\eta_{l,t}^2 2\alpha \log(1 u_t^{1/M_{l,t}}))^{1/2}\}, \infty)$.

In the second step above, we may have to sample from a normal distribution, restricted to two disjoint intervals. The resulting distribution is therefore a mixture of two truncated normals,

with probabilities determined by the (normalized) probability the normal assigns to each interval. These truncated normals both have mean 0 and variance $(1 + \sum_{t=1}^{T} \sum_{r=l+1}^{N} M_{r,t}/\alpha)^{-1}$, and each mixture component has equal probability.

The full conditional for each $\eta_{l,t}$, l = 1, ..., N - 1, t = 2, ..., T - 1, is proportional to

$$N\left(\eta_{l,t} \mid 0, \frac{\alpha}{\sum_{r=l+1}^{N} M_{r,t}}\right) N(\eta_{l,t} \mid \phi \eta_{l,t-1}, 1 - \phi^2) N(\eta_{l,t+1} \mid \phi \eta_{l,t}, 1 - \phi^2) \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,t}^2}{2\alpha}\right)\right)^{M_{l,t}}$$

$$\propto \mathcal{N}\left(\eta_{l,t} \mid \frac{\phi\alpha(\eta_{l,t-1} + \eta_{l,t+1})}{\phi^2(\alpha - \sum_{r=l+1}^N M_{r,t}) + \alpha + \sum_{r=l+1}^N M_{r,t}}, \frac{\alpha(1-\phi^2)}{\phi^2(\alpha - \sum_{r=l+1}^N M_{r,t}) + \alpha + \sum_{r=l+1}^N M_{r,t}}\right) \\ \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,t}^2}{2\alpha}\right)\right)^{M_{l,t}}$$

Each $\eta_{l,t}$, l = 1, ..., N - 1, and t = 2, ..., T - 1, can therefore be sampled with a slice sampler:

• Draw
$$u \sim \operatorname{Unif}\left(0, \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,t}^2}{2\alpha}\right)\right)^{M_{l,t}}\right).$$

• Draw $\eta_{l,t} \sim \operatorname{N}\left(\eta_{l,t} \mid \frac{\phi \alpha(\eta_{l,t-1} + \eta_{l,t+1})}{\phi^2(\alpha - \sum_{r=l+1}^N M_{r,t}) + \alpha + \sum_{r=l+1}^N M_{r,t}}, \frac{\alpha(1 - \phi^2)}{\phi^2(\alpha - \sum_{r=l+1}^N M_{r,t}) + \alpha + \sum_{r=l+1}^N M_{r,t}}\right),$
restricted to $\left\{\eta_{l,t} : \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,t}^2}{2\alpha}\right)\right)^{M_{l,t}} > u\right\}, \operatorname{giving} \eta_{l,t}^2 > -2\alpha \log(1 - u^{1/M_{l,t}}) - \zeta_l^2.$

In the second step above, we will again either sample from a single normal or a mixture of truncated normals, where each normal has the same mean and variance, but the truncation intervals differ. Since the mean of this normal is not zero, the weights assigned to each truncated normal are not the same. The unnormalized weight assigned to the normal which places positive probability on $((-2\alpha \log(1 - u^{1/M_{l,t}}) - \zeta_l^2)^{1/2}, \infty)$ is given by $1-F((-2\alpha \log(1-u^{1/M_{l,t}})-\zeta_l^2)^{1/2})$, where F is the CDF of the normal for $\eta_{l,t}$ given in the second step. The unnormalized weight given to the component which places positive probability on $(-\infty, -(-2\alpha \log(1-u^{1/M_{l,t}}) - \zeta_l^2)^{1/2})$ is given by $F(-(-2\alpha \log(1-u^{1/M_{l,t}}) - \zeta_l^2)^{1/2})$. The full conditionals for $\eta_{l,1}$ and $\eta_{l,T}$ are slightly different. The full conditional for $\eta_{l,1}$ is

$$p(\eta_{l,1} \mid \dots, \text{data}) \propto \mathcal{N}(\eta_{l,1} \mid 0, \frac{\alpha}{\sum_{r=l+1}^{N} M_{r,1}}) \mathcal{N}(\eta_{l,1} \mid 0, 1) \mathcal{N}(\eta_{l,2} \mid \phi \eta_{l,1}, 1 - \phi^2) \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,1}^2}{2\alpha}\right)\right)^{M_{l,1}},$$

$$\propto \mathcal{N}\left(\eta_{l,1} \mid \frac{\phi \alpha \eta_{l,2}}{\alpha + \sum_{r=l+1}^{N} M_{r,1} - \phi^2 \sum_{r=l+1}^{N} M_{r,1}}, \frac{\alpha(1-\phi^2)}{\alpha + \sum_{r=l+1}^{N} M_{r,1} - \phi^2 \sum_{r=l+1}^{N} M_{r,1}}\right) \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,1}^2}{2\alpha}\right)\right)^{M_{l,1}}.$$

For $\eta_{l,T}$, we have:

$$p(\eta_{l,T} \mid \dots, \text{data}) \propto \mathcal{N}(\eta_{l,T} \mid 0, \frac{\alpha}{\sum_{r=l+1}^{N} M_{r,T}}) \mathcal{N}(\eta_{l,T} \mid \phi \eta_{l,T-1}, 1-\phi^2) \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,T}^2}{2\alpha}\right)\right)^{M_{l,T}}$$

,

which is proportional to

$$N\left(\eta_{l,T} \mid \frac{\phi \alpha \eta_{l,T-1}}{\alpha + \sum_{r=l+1}^{N} M_{r,T} - \phi^2 \sum_{r=l+1}^{N} M_{r,T}}, \frac{\alpha (1-\phi^2)}{\alpha + \sum_{r=l+1}^{N} M_{r,T} - \phi^2 \sum_{r=l+1}^{N} M_{r,T}}\right) \left(1 - \exp\left(-\frac{\zeta_l^2 + \eta_{l,T}^2}{2\alpha}\right)\right)^{M_{l,T}}$$

The slice samplers for $\eta_{l,1}$ and $\eta_{l,T}$ are therefore implemented in the same way as for $\eta_{l,t}$, except the normals which are sampled from have different means and variances.

Note that if there is no data at some time point r, there are no $L_{r,i}$ variables, so the full conditional for $\eta_{l,r}$ simply arises from the distribution for $f(\eta_{l,r} \mid \eta_{l,r-1})$ and $f(\eta_{l,r+1} \mid \eta_{l,r})$ in Equation (5) of the paper.

Updating α

The posterior full conditional for α is proportional to

$$p(\alpha) \exp\left(-\frac{\sum_{t=1}^{T} M_{N,t} \sum_{l=1}^{N-1} (\zeta_{l}^{2} + \eta_{l,t}^{2})}{2\alpha}\right) \exp\left(-\frac{\sum_{t=1}^{T} \sum_{l=2}^{N-1} M_{l,t} \sum_{r=1}^{l-1} (\zeta_{r}^{2} + \eta_{r,t}^{2})}{2\alpha}\right) \prod_{t=1}^{T} \prod_{l=1}^{N-1} \left(1 - \exp\left(-\frac{\zeta_{l}^{2} + \eta_{l,t}^{2}}{2\alpha}\right)\right)^{M_{l,t}}$$

Therefore, with $p(\alpha) = IG(a_{\alpha}, b_{\alpha})$, we have

$$p(\alpha \mid \dots, \text{data}) = \text{IG}\left(\alpha \mid a_{\alpha}, b_{\alpha} + \frac{1}{2} \sum_{t=1}^{T} \left(M_{N,t} \sum_{l=1}^{N-1} (\zeta_{l}^{2} + \eta_{l,t}^{2}) + \sum_{l=2}^{N-1} M_{l,t} \sum_{r=1}^{l-1} (\zeta_{r}^{2} + \eta_{r,t}^{2}) \right) \right)$$
$$\prod_{t=1}^{T} \prod_{l=1}^{N-1} \left(1 - \exp\left(-\frac{\zeta_{l}^{2} + \eta_{l,t}^{2}}{2\alpha}\right) \right)^{M_{l,t}}$$

The parameter α can be sampled using a Metropolis-Hastings algorithm. In particular we work with $\log(\alpha)$, and use a normal proposal distribution centered at the log of the current value of α .

Updating ϕ and Θ

The full conditional for the AR parameter ϕ is $p(\phi \mid \dots, \text{data}) \propto p(\phi) \prod_{t=2}^{T} \prod_{l=1}^{N-1} N(\eta_{l,t} \mid \phi \eta_{l,t-1}, 1 - \phi^2)$, where $p(\phi)$ is the uniform(0, 1) prior for ϕ . We use a Metropolis-Hastings step to sample $\log{\{\phi/(1-\phi)\}}$ based on a normal proposal distribution.

The full conditional for Θ is proportional to $\prod_{t=2}^{T} \prod_{l=1}^{N} N(\boldsymbol{m} + \Theta \boldsymbol{\mu}_{l,t-1}, \boldsymbol{V})$. With the assumption that Θ is diagonal with elements θ_j , and with Unif(0, 1) priors on each element, the Metropolis-Hastings algorithm can be used to update each θ_j individually by sampling $\log\{\theta_j/(1-\theta_j)\}$ from a normal proposal distribution. Alternatively, the θ_j can be updated jointly with a multivariate normal proposal distribution on the logit scale.

Updating $\{\mu_{l,t}\}$

The updates for $\mu_{l,t}$ are based on $N(\boldsymbol{m}^*, \boldsymbol{V}^*)$ distributions, with \boldsymbol{m}^* and \boldsymbol{V}^* given by:

- For t = 2, ..., T-1, if $M_{l,t} = 0$, then the update for $\mu_{l,t}$ has $V^* = (V^{-1} + (\Theta^{-1}V\Theta^{-T})^{-1})^{-1}$ and $m^* = V^*(V^{-1}(m + \Theta\mu_{l,t-1}) + (\Theta^{-1}V\Theta^{-T})^{-1}\Theta^{-1}(\mu_{l,t+1} - m))$
- For t = 2, ..., T-1, if $M_{l,t} \neq 0$, then the update for $\mu_{l,t}$ has $V^* = (V^{-1} + (\Theta^{-1}V\Theta^{-T})^{-1} + M_{l,t}\Sigma_l^{-1})^{-1}$ and $m^* = V^*(V^{-1}(m + \Theta\mu_{l,t-1}) + (\Theta^{-1}V\Theta^{-T})^{-1}\Theta^{-1}(\mu_{l,t+1} m) + \Sigma_l^{-1}\sum_{\{i:L_{t,i}=l\}} y_{t,i}^*)$

- for t = 1, if $M_{l,1} = 0$, then the update for $\mu_{l,1}$ has $V^* = ((\Theta^{-1}V\Theta^{-T})^{-1} + V_0^{-1})^{-1}$, and $m^* = V^*((\Theta^{-1}V\Theta^{-T})^{-1}\Theta^{-1}(\mu_{l,2} - m) + V_0^{-1}m_0)$
- for t = 1, if $M_{l,1} \neq 0$, then the update for $\mu_{l,1}$ has $V^* = (M_{l,1}\boldsymbol{\Sigma}_l^{-1} + (\Theta^{-1}V\Theta^{-T})^{-1} + V_0^{-1})^{-1}$ and $\boldsymbol{m}^* = V^*(\boldsymbol{\Sigma}_l^{-1}\sum_{\{i:L_{1,i}=l\}}\boldsymbol{y}_{1,i}^* + (\Theta^{-1}V\Theta^{-T})^{-1}\Theta^{-1}(\mu_{l,2} \boldsymbol{m}) + V_0^{-1}\boldsymbol{m}_0)$
- for t = T, if $M_{l,T} = 0$, then the update for $\boldsymbol{\mu}_{l,T}$ has $\boldsymbol{V}^* = \boldsymbol{V}$, and $\boldsymbol{m}^* = \boldsymbol{m} + \Theta \boldsymbol{\mu}_{l,T-1}$
- for t = T, if $M_{l,T} \neq 0$, then the update for $\mu_{l,T}$ has $V^* = (M_{l,T} \Sigma_l^{-1} + V^{-1})^{-1}$ and $m^* = V^* (\Sigma_l^{-1} \sum_{\{i:L_{T,i}=l\}} y_{t,i}^* + V^{-1} (m + \Theta \mu_{l,T-1}))$

Updating $\{\Sigma_l\}$

Let $M_l = \sum_{t=1}^T M_{l,t}$. The posterior full conditional for $\boldsymbol{\Sigma}_l$ is proportional to $\mathrm{IW}(\nu + M_l, \boldsymbol{D} + \sum_{\{(t,i):L_{ti}=l\}} (\boldsymbol{y}_{t,i}^* - \boldsymbol{\mu}_{l,t}) (\boldsymbol{y}_{t,i}^* - \boldsymbol{\mu}_{l,t})^T)$. When $|\{(t,i): L_{t,i} = l\}| = 0$, $\boldsymbol{\Sigma}_l$ is drawn from $\mathrm{IW}(\nu, \boldsymbol{D})$.

1.2 Prior Specification

To implement the model, we must specify the parameters of the hyperpriors on ψ . A default specification strategy is developed by considering the limiting case of the model as $\alpha \rightarrow 0^+$ and $\Theta \rightarrow 0$, which results in a single normal distribution for Y_t^* . In the limit, with $Y_t^* \mid \mu_t, \Sigma \sim N(\mu_t, \Sigma)$ and $\mu_t \mid m, V \sim N(m, V)$, we find $E(Y_t^*) = a_m$ and $Cov(Y_t^*) = B_m + B_V(a_V - d - 1)^{-1} + a_D B_D(\nu - d - 1)^{-1}$, where d is the response-covariate dimension, here d = 3. The only covariate information we require is an approximate center and range, denoted by c^x and r^x for X, and analogously for W. A widely used default choice uses the midpoint and range of the data. Alternatively, these can be specified through expert opinion. In the rockfish application, there is sufficient knowledge of the averages and ranges of length and age for Chilipepper rockfish to create a scientifically informed but vague prior. We use c^x and $r^x/4$ as proxies for the marginal mean and standard deviation of X. We also seek to scale the latent variables appropriately. The centers and ranges of observed age provide approximate centers c^w and ranges r^w of latent log-age W. Since Y is supported on $\{1, \ldots, C\}$, latent continuous Z must be supported on values slightly below γ_1 up to slightly above γ_{C-1} , so that $r^z/4$ is a proxy for the standard deviation of Z, where $r^z = (\gamma_{C-1} - \gamma_1)$. Using these mean and variance proxies, we fix $a_{\mathbf{m}} = (0, c^w, c^x)$. Each of the three terms in $\operatorname{Cov}(\mathbf{Y}_t^*)$ can be assigned an equal part of the total covariance; we set this to $3^{-1}\operatorname{diag}\{(r^z/4)^2, (r^w/4)^2, (r^x/4)^2\}$. For dispersed but proper priors, ν , a_V and a_D can be fixed to small values. In the application, we fix $a_V = a_D = \nu = d + 4$, and calculate \mathbf{B}_m , \mathbf{B}_V , and \mathbf{B}_D so that $\operatorname{Cov}(\mathbf{Y}_t^*) = 3^{-1}\operatorname{diag}\{(r^z/4)^2, (r^w/4)^2, (r^x/4)^2\}$

It remains to specify \mathbf{m}_0 and \mathbf{V}_0 , the mean and covariance for the initial distributions $\boldsymbol{\mu}_{l,1}$. We propose a fairly conservative specification, noting that in the limit, $\mathbf{E}(\mathbf{Y}_1^*) = \mathbf{m}_0$, and $\operatorname{Cov}(\mathbf{Y}_1^*) = a_{\mathbf{D}} \mathbf{B}_{\mathbf{D}} (\nu - d - 1)^{-1} + \mathbf{V}_0$. Therefore, \mathbf{m}_0 can be specified in the same way as $\mathbf{a}_{\mathbf{m}}$ but using only the subset of data at t = 1. We set $\mathbf{m}_0 = (0, c_1^w, c_1^x)$ and \mathbf{V}_0 to diag $\{(r_1^z/4)^2, (r_1^w/4)^2, (r_1^x/4)^2\} - a_{\mathbf{D}} \mathbf{B}_{\mathbf{D}} (\nu - d - 1)^{-1}$, where $c_1^w, r_1^w, c_1^x, r_1^x$, denote the midpoints and ranges of W and X at time 1. We set \mathbf{V}_0 to diag $\{(r_1^z/4)^2, (r_1^w/4)^2, (r_1^x/4)^2\} - a_{\mathbf{D}} \mathbf{B}_{\mathbf{D}} (\nu - d - 1)^{-1}$, where the subscript 1 indicates the subset of data at t = 1.

1.3 Prior Sensitivity

In simulation studies and the rockfish application, we observed a moderate to large amount of learning for all hyperparameters. For instance, for the rockfish data, the posterior distribution for ϕ was concentrated on values close to 1, indicating the DDP weights are strongly correlated across time. There was also moderate learning for α as its posterior distribution was concentrated around 0.5, with small variance, shifted down slightly relative to the prior which had expectation .75. The posterior distribution for each element of m was reduced in variance and concentrated on values not far from those indicated by the prior mean. The posterior samples for the covariance matrices V and D supported smaller variance components than suggested by the prior. Figures 1–3 illustrate prior and posterior inference for some hyperparameters of the DDP mixture model in the context of the application.



Figure 1: Posterior density for each component of Θ (solid) compared with prior density (dashed).



Figure 2: Posterior density for α (solid) compared with prior density (dashed).



Figure 3: Posterior density for each component of m (solid) compared with the prior density (dashed).

2 Model Comparison Results

We fit the common atoms DDP and the more general DDP model to the data and computed the posterior predictive criterion of Gelfand and Ghosh (1998), which is composed of a sum of squares goodness of fit term, and sum of predictive variances penalty term. For data (y_1, \ldots, y_n) and model m, the goodness of fit term G(m) is given by $\sum_{i=1}^{n} (E(y_{rep,i} \mid data, m) - y_i)^2$, and the penalty term P(m) by $\sum_{i=1}^{n} var(y_{rep,i} \mid data, m)$. An overall comparison criterion $D_k(m)$ can be obtained as $P(m) + k(k+1)^{-1}G(m)$, where $k \ge 0$, and as $k \to \infty$, the goodness of fit and penalty terms are equally weighted. We compare the models in terms of their ability to explain maturity as a function of age and length, and thus estimate posterior expectations and variances for maturity, conditional on age and length. These expressions require estimation of conditional expectations, which can be written as integrals and efficiently estimated via Monte Carlo integration based on posterior samples for model parameters.

The goodness of fit and penalty terms are computed for each year of data between 1993 and 2004, and we find that the goodness of fit term is lower under the general DDP model at every time point by a small amount. The penalty terms show larger differences, and are lower under the general model for all years except 1996, 2002, and 2004. These values are shown in

Table 1: Goodness of fit and penalty terms for each year of data under the general model (g) and the simpler model (s). The overall values for comparison $D_k(g)$ and $D_k(s)$ are shown for $k \to \infty$.

year	n_t	G(g)	G(s)	P(g)	P(s)	$D_{\infty}(g)$	$D_{\infty}(s)$
1993	305	64.9	65.5	61.8	71.7	126.8	137.2
1994	271	54.2	56.9	55.5	57.8	109.6	114.6
1995	256	60.3	61.7	61.0	64.7	121.3	126.4
1996	160	30.0	31.4	35.0	33.3	65.0	64.7
1997	184	28.1	29.7	31.9	34.9	60.0	64.6
1998	132	37.2	37.3	36.6	37.6	73.8	74.9
1999	396	62.4	64.0	65.5	71.9	127.9	135.9
2000	64	2.7	3.4	4.9	8.9	7.6	12.3
2001	146	42.2	44.1	26.4	38.3	68.6	82.4
2002	238	21.2	21.6	41.3	28.3	62.5	49.9
2004	37	2.7	3.0	7.6	6.0	10.2	9.0

Table 1. This suggests that the general model is providing a better fit to the data with less uncertainty, except in the few cases in which the posterior predictive variance is larger under the more complex model. We also obtain the corresponding comparison criterion $D_k(m)$ for $k \to \infty$, which yield the same conclusions. Other values of k give effectively the same results. Overall, the general model is preferred to the common atoms model using this criterion.

References

- Gelfand, A. and Ghosh, S. (1998), "Model choice: A minimum posterior predictive loss approach," *Biometrika*, 85, 1–11.
- Ishwaran, H. and James, L. (2001), "Gibbs Sampling Methods for Stick-breaking Priors," Journal of the American Statistical Association, 96, 161–173.