When do traits matter? A hierarchical Bayesian modelling approach to Brachiopod survival

Peter D Smits

Committee on Evolutionary Biology, University of Chicago

May 13, 2015

Abstract

Write me.

1 Introduction

How do differences in taxon traits affect differences in extinction risk? Jablonski [16] hypothesizes that as baseline extinction risk increases, the effects of taxon traits on survival should decrease. The trait where this pattern should be weakest or absent is geographic range [16]. Taxon traits are defined here as descriptors of a taxon's adaptive zone which is the set of biotic–biotic and biotic–abiotic interactions that taxon can and does experience. In effect, these are descriptors of a taxon's broad-sense ecology.

Jablonski [16] phrased this hypothesis in terms of background versus mass extinction, but it is readily transferable to a continuous variation framework as there is no obvious distinction in terms of extinction rate between these two states [42]. I adopt a continuous variation framework as this is more amenable to actually modeling the long term relationships between taxon traits and extinction risk. Additionally, the Jablonski [16] hypothesis has strong model structure requirements as not only do the taxon trait effects need to be modeled but the relationships between these effects need to be modeled as well. This level of complexity is possible in a varying intercepts, varying slopes hierarchical model as discussed below.

I choose to model differences in taxon durations because trait based differences in extinction risk should manifest as differences in taxon durations. For example, species with a beneficial trait should survive longer, on average, than a species without that beneficial trait. Both conceptually and metaphorically, taxon survival can be considered an aspect of "taxon fitness" along with lineage specific branching/origination rate [3, 28]. The analysis of taxon durations, time from origination to extinction, falls under the purview of survival analysis, a field of

1

applied statistics commonly used in health care [20] but has a long history in paleontology [36, 37, 40, 41].

Geographic range is widely considered the most important taxon trait for estimating differences in extinction risk at nearly all times [16–18, 29]. This is logical because if mortality is randomly distributed spatially, a taxon with a large geographic range is less likely to be completely wiped out than a taxon with restricted range. This is a strong prediction based on strong prior evidence that should be incorporated into any model of extinction risk in order to best represent our knowledge.

Miller and Foote [25] demonstrated that during mass extinctions taxa associated with open ocean environments tend to have a greater extinction risk than those taxa associated with epicontinental seas. During periods of background extinction, however, they found no consistent difference between taxa favoring either environment. Because of this study, the following prediction for survival patterns can be made: as extinction risk increases, taxa associated with open ocean environments should generally increase in extinction risk versus those that favor epicontinental seas.

There is also a possible nonlinear relationship between environmental preference and taxon duration. A long standing hypothesis is that generalists or unspecialized taxa will have greater survival than specialists [2, 21, 22, 26, 27, 36] SMITS, IN PREP. A simple expectation for continuous traits is that this nonlinearity would manifest as a downward facing function, with taxa with intermediate trait values having a greater expected durations than taxa with high or low trait values. By utilizing a continuous measure of environmental preference, it is relatively straight forward to allow for a nonlinear effect in a model of taxon durations/extinction risk.

I adopt a hierarchical Bayesian survival modelling approach for multiple reasons, one being that it represents a conceptual and statistical unification of the paleontological dynamic and cohort survival analytic approaches [2, 5, 31, 32, 34, 40, 41]. By using a Bayesian framework I am best able to quantify the uncertainty inherent in the estimates of the effects of taxon traits on survival, especially in cases where the covariates of interest (taxon traits) are themselves known with error. Hierarchical modelling, sometimes called "mixed-effects modeling," is a statistical approach which explicitly takes into account the structure of the observed data in order to model both the with and between group variance [9, 10].

In this approach, the units of study (e.g. genera) belong to a single grouping (e.g. origination cohort) that are also considered draws from a shared probability distribution (e.g. all cohorts, observed and unobserved). The group level parameters are then estimated simultaneously as the other parameters of interest (e.g. covariate effects) [10]. The subsequent estimates are then partially pooled together, where parameters from groups with large samples or effects are allowed to remain large while those of groups with small samples or effects are pulled

towards the overall group mean. In this case, origination cohorts are the groups and the mean survival model corresponds to the dynamic survivorship model. This leads to simultaneous dynamic and cohort style analysis.

This partial pooling is one of the greatest advantages of hierarchical modeling. By letting the groups "support" each other, parameter estimates then better reflect our actual uncertainty. Additionally, this partial pooling helps control for multiple comparisons and spurious results as small effects with little empirical support are drawn towards the overall group mean [9, 10].

In this analysis, genera are structured as belonging to origination cohorts. All covariate effects (regression coefficients), as well as the intercept term (baseline extinction risk), were allowed to vary by group. The covariance/correlation between covariate effects was also modeled. This hierarchical structure allows inference for how covariates effects may change with respect to each other while simultaneously estimating the effects themselves, correctly propagating our uncertainty. Additionally, instead of relying on potentially biased point estimates of environmental affinity, I adopt a measurement error modeling approach where environmental affinity is measured as the difference in the taxon's probabilistic environmental occurrence pattern and the background probabilistic environmental occurrence pattern. This approach correctly propagates our uncertainty, which leads to more valid posterior inference.

2   Methods

2.1   Fossil occurrence information

The dataset analyzed here is derived from the a combination of the occurrence information from Foote and Miller [7] and the body size data from Payne et al. [30]. The Foote and Miller [7] dataset is based on the Paleobiology Database (http://www.paleodb.org); see Foote and Miller [7] for a full description of the inclusion criterion. Additionally, epicontinental versus open ocean assignemnts for occurrence information are based on Miller and Foote [25].

Sampled occurrences were restricted to those with latitude and longitude coordinates, assignment to either epicontinental or open-ocean environment, and belonging to a genus present in the body size dataset. Genus duration was calculated as the number of geologic stages from first appearance to last appearance, inclusive. Genera who's last appearance was in a stage preceding a mass extinction were right censored, and genera with a duration of only one stage and were left censored (see below for explanation of censoring). The covariates used to model genus duration were geographic range size (r), environmental preference (v), and body size (m).

Geographic range was calculated using an occupancy approach. First, all occurrences were projected onto an equal-area cylindrical map projection. Each

occurrence was then assigned to one of the cells from a 70×34 regular raster grid placed on the map. Each grid cell represents approximately 250,000 km^2. Following this, for each stage, the total number of grid cells occupied was calculated. The number of grid cells that each genus present occurs was also calculated and made relative by dividing by the total number of possible cells. Finally, mean relative genus occupancy was calculated as the mean of per stage relative occupancy. The map projection and regular lattice were made using shape files from http://www.naturalearthdata.com/ and the `raster` package for R [11].

Body size data was sourced directly from Payne et al. [30]. Because those measurements are presented without error, a measurement error model similar to the one for environmental affinity could not be implemented (below).

Prior to analysis, geographic range and body size were transformed and standardized in order to improve interpretability of the results. Geographic range size, which can only have between 0 and 1, was logit transformed. Body size, which is defined for all positive real values was natural log transformed. These covariates were then standardized by mean centering and dividing by two times their standard deviation following Gelman and Hill [9].

2.1.1 Uncertainty in environmental preference

The calculation and inclusion of environmental affinity in the survival model is a statistical procedure that takes into account our uncertainty based on where fossils tend to occur. Because we cannot directly observe if a fossil taxon had occurrences restricted to only a single environmental, instead we can only estimate its affinity with uncertainty. One advantage of using a Bayesian analytical approach is that both parameters and data are considered random samples from some underlying distribution, which means that is is possible to model the uncertainty in our covariates of interest [10]. My approach is conceptually similar to Simpson and Harnik [35] but instead of obtaining a single point estimate, an entire posterior distribution is estimated.

The first step is to determine the probability θ at which genus i occurs in an epicontinental settings based on its own pattern of occurrences. Define e_i as the number of occurrences of genus i in an epicontiental sea and o_i as the number of occurrences of genus i not in an epicontinental sea (e.g. open ocean). Because the value of interest is the probability of occurring in an epicontinental environment, given the observed fossil record, I assume that probability follows a binomial distribution. We can then define our sampling statement as

$$e_i \sim \text{Binomial}(e_i + o_i, \theta_i). \tag{1}$$

I used a flat prior of θ_i defined as $\theta_i \sim \text{Beta}(1,1)$. Because the beta distribution is the conjugate prior for the binomial distribution, the posterior is easy to compute in closed form. The posterior probability of θ is then

$$\theta_i \sim \text{Beta}(e_i + 1, o_i + 1) \tag{2}$$

It is extremely important, however, to take into account the overall environmental occurrence probability of all other genera present at the same time as genus i. This is incorporated as an additional probability Θ. Define E_i as the total number of other fossil occurrences (e.g. excepting for genus i) in epicontinental seas during stages where i occurs and O_i as the number of other fossil occurrences not on epicontinental seas. We can then define the sampling statement as

$$E_i \sim \text{Binomial}(E_i + O_i, \Theta_i). \tag{3}$$

Again, I used a flat prior of Θ_i defined as $\Theta_i \sim \text{Beta}(1,1)$. The posterior of Θ is then simply defined as

$$\Theta_i \sim \text{Beta}(E_i + 1, O_i + 1) \tag{4}$$

I then define the environmental affinity of genus i as $v_i = \theta_i - \Theta_i$. v_i is a value that can range between -1 and 1, where negative values indicate that genus i tends to occur more frequently in open ocean environments than background while positive values indicate that genus i tends to occur in epicontiental environments.

While this approach is noticeably more complicated than previous ones [6, 19, 24, 35] there are some important benefits to both using a continuous measure of affinity as well directly modeling our uncertainty. In order to show case some of these benefits, I performed a simulation analysis of how modal/maximum *a posteriori* (MAP) estimates versus full posterior estimates.

In this simulation, I first defined the "background" epicontinental occurrence θ_b as 0.50 with a small amount of noise. This was represented as a beta distribution

$$\Theta_b = \text{Beta}(\alpha = 2500, \beta = 2500). \tag{5}$$

This choice of parameters for the distribution reflects the average number of background occurrences for either epicontinental or open ocean environments per genus.

Using this background occurrence ratio, randomly generated the occurrence patterns 1000 simulated taxa. This was done at multiple sample sizes (1, 2, 3, 4, 5, 10, 25, 50, 100) in order to demonstrate the effects of increasing sample size on the confidence of environmental affinity. For each simulated taxon I calculated the full posterior distribution while assuming a flat Beta prior ($\text{Beta}(1,1)$). Using the full posterior I calculated the MAP probability of occurring in epicontinental environments. The environmental affinity was calculated for each of the simulated taxa using both the full posterior and the MAP estimate. In this toy example, environmental affinity can range between -0.5 and 0.5.

As should be expected, as sample size increases the distribution of MAP estimates converge on the true value (Fig. 1). For taxa with less than 10 occurrences, the MAP estimate is biased towards extreme values. Note that the mode of the beta distribution is not defined for situations where there were 0 draws of one of the environmental conditions. Instead, the vertical line is based entirely on the

observed occurrences which are technically the modal estimates because they are the most frequently occurring/highest density.

In contrast, we can compare the true occurrence probability distribution versus the posterior estimate for a given sample (Fig. 2). When sample sizes are low, posterior estimates are flat and represent a compromise between the likelihood and the flat prior (Eq. 2). Because of this, estimates from small sizes are less likely to be overly biased towards the extremes. This is further emphasized by inspection of the estimates of environmental affinity for the simulated taxa (Fig. 3). Posterior estimates from simulated taxa with small sample size have a much broader distribution that both allows for the extreme observation but still captures the "true" value (0).

By defining environmental preference as the difference in full posterior estimates of occurrence probability, it is possible to include taxa with low sample sizes that are normal discarded [6, 19, 24, 35]. Additionally, 55+% of observed Paleozoic brachiopod genera have less than 10 occurrences which is the range of sample sizes where MAP (or ML) estimates would be potentially most biased. This is preferable to finding the difference between the MAP estimates (blue line; Fig. 3).

2.2   Survival model

Genus durations were assumed to follow either an exponential or Weibull distribution. Each of these distributions makes assumptions about how duration may effect extinction risk [20]. The exponential distribution assumes that extinction risk is independent of duration. In contrast, the Weibull distribution allows for age dependent extinction via the shape parameter α, though only as a monotonic function of duration. Importantly, the Weibull distribution is equivalent to the exponential distribution when $\alpha = 1$.

The simplest model of genus duration includes no covariate or structural information. Define y_i as the duration in stages of genus i, where $i = 1, \ldots, n$ and n is the number of observed genera. These two models are them simply defined as

$$
\begin{aligned}
y_i &\sim \text{Exponential}(\lambda) \\
y_i &\sim \text{Weibull}(\alpha, \sigma).
\end{aligned}
\tag{6}
$$

Note that λ is a "rate" or inverse-scale while σ is a scale parameter, meaning that $\frac{1}{\lambda} = \sigma$.

These simple models can then be expanded to include covariate information as predictors by reparameterizing λ or σ as a regression [20]. Each of the covariates of interest is given its own regression coefficient (e.g. β_r) along with an intercept term β_0. There are some additional complications to the parameterization of σ associated with the inclusion of α as well as for interpretability [20]. Both of

these are then written as

$$\lambda_i = \exp(\beta_0 + \beta_r r_i + \beta_v v_i + \beta_{v^2} v_i^2 + \beta_m m_i)$$
$$\sigma_i = \exp\left(\frac{-(\beta_0 + \beta_r r_i + \beta_v v_i + \beta_{v^2} v_i^2 + \beta_m m_i)}{\alpha}\right). \tag{7}$$

The regression equations are exponentiated because both λ and σ are only defined for positive reals. The quadratic term for environmental affinity v is to allow for the possible nonlinear relationship between environmental affinity and extinction risk.

The models which incorporate both equations 6 and 7 can then be further expanded to allow all of the β coefficients, including β_0, to vary with origination cohort while also modeling their covariance and correlation. This is called a varying-intercepts, varying-slopes model [9]. It is much easier to represent and explain how this is parameterized using matrix notation. First, define $\mathbf{B}$ as $k \times J$ matrix of the k coefficients including the intercept term ($k = 5$) for each of the J cohorts. Second, define $\mathbf{X}$ as a $n \times k$ matrix where each column is one of the covariates of interest. Importantly, $\mathbf{X}$ includes a columns of all 1s which correspond to the constant term β_0. Third, define $j[i]$ as the origination cohort of genus i, where $j = 1, \ldots, J$ and J is the total number of observed cohorts. We then rewrite λ and σ of equation 7 in matrix notation as

$$\lambda_i = \exp(\mathbf{X}_i B_{j[i]})$$
$$\sigma_i = \exp\left(\frac{-(\mathbf{X}_i B_{j[i]})}{\alpha}\right). \tag{8}$$

Because B is a matrix, I use a multivariate normal prior with unknown vector of means μ and covariance matrix Σ. This is written as

$$B \sim \mathrm{MVN}(\mu, \Sigma) \tag{9}$$

where μ is length k vector representing the overall mean of the distributions of β coefficients. Σ is a $k \times k$ covariance matrix of the β coefficients.

What remains is assigning priors the elements of $\vec{\mu}$ and the covariance matrix Σ. All elements of $\vec{\mu}$ except for μ_r were given weakly informative normal priors while μ_r was given an informative normal prior ($\mu_r \sim \mathcal{N}(-1, 1)$. The prior for Σ is a bit more complicated due to its multivariate nature. While the conditional conjugate prior distribution of the covariance matrix of a multivariate normal distribution is the inverse-Wishart [10], because I am using a variant for Hamiltonian Monte Carlo (HMC) called No U-Turn Sampling (NUTS) for posterior approximation as opposed to the more commonly used Gibbs sampling there is no benefit from using a conjugate prior Stan Development Team [39]. Additionally, the inverse-Wishart distribution strongly constrains the off-diagonal elements of the covariance matrix which is not desired. Instead, it is better to model the scale terms separate from the correlation structure of the coefficients. This is possible

because of the relationship between a covariance and a correlation matrix, defined as

$$\Sigma_B = \text{Diag}(\vec{\tau})\Omega\text{Diag}(\vec{\tau}) \tag{10}$$

where $\vec{\tau}$ is a length k vector of variances and $\text{Diag}(\tau)$ is a diagonal matrix.

I used a LKJ prior distribution for correlation matrix Ω as recommended by Stan Development Team [39]. The LKJ distribution is a single parameter multivariate distribution where values of the parameter η greater than 1 concentrate density at the unit correlation matrix, which corresponds to no correlation between the β coefficients. The scale parameters, $\vec{\tau}$, are given weakly informative half-Cauchy (C^+) priors following Gelman [8].

Given all the above, the exponential model is then defined as

$$
\begin{aligned}
y_i &\sim \text{Exponential}(\lambda) \\
\lambda_i &= \exp(\mathbf{X}_i B_{j[i]}) \\
B &\sim \text{MVN}(\vec{\mu}, \Sigma_B) \\
\Sigma_B &= \text{Diag}(\vec{\tau})\Omega_B\text{Diag}(\vec{\tau}) \\
\mu_\kappa &\sim \begin{cases} \mathcal{N}(0,5) & \text{if } k \neq r, or \\ \mathcal{N}(-1,1) & \text{if } k = r \end{cases} \\
\tau_\kappa &\sim \text{C}^+(1) \text{ for } \kappa \in 1:k \\
\Omega &\sim \text{LKJ}(2).
\end{aligned}
\tag{11}
$$

The Weibull model is then also defined as

$$
\begin{aligned}
y_i &\sim \text{Weibull}(\alpha, \sigma) \\
\sigma_i &= \exp\left(\frac{-(\mathbf{X}_i B_{j[i]})}{\alpha}\right) \\
B &\sim \text{MVN}(\vec{\mu}, \Sigma_B) \\
\Sigma_B &= \text{Diag}(\vec{\tau})\Omega_B\text{Diag}(\vec{\tau}) \\
\alpha &\sim \text{C}^+(2) \\
\mu_k &\sim \begin{cases} \mathcal{N}(0,5) & \text{if } k \neq r, or \\ \mathcal{N}(-1,1) & \text{if } k = r \end{cases} \\
\tau_k &\sim \text{C}^+(1) \\
\Omega &\sim \text{LKJ}(2).
\end{aligned}
\tag{12}
$$

Note that the above formulations of each model (Eq. 11, 12) does not include how the uncertainty in environmental affinity is included nor how censored observations are included. An explination of including censored observations follows.

2.3 Censored observations

A key aspect of survival analysis is the inclusion of censored, or incompletely observed, data points [15, 20]. The two classes of censored observations encountered in this study were right and left censored observations. Right censored genera are those that did not go extinct during the window of observation, or genera that are still extant. Left censored observations are those taxa that it is only known when it was extinct by. To put another way, this is a taxon that went extinct but the observed duration is an over estimate of the actual duration.

In the context of this study, I considered all genera that had a duration of only one geologic stage to be left censored as we do not have a finer degree of resolution. Conceptually, this is similar to if I was studying, say, survival patterns in rats and an individual had died between the start of the experiment and next time the rats were observed. We know the rat lived no more than that amount of time.

The key function for modeling censored observations is the survival function, or $S(t)$. $S(t)$ corresponds to the probability that a genus having existed for t stages will not have gone extinct while $h(t)$ corresponds to the instantaneous extinction rate at taxon age t [20]. For an exponential model, $S(t)$ is defined as

$$S(t) = \exp(-\lambda t), \tag{13}$$

and for the Weibull distribution $S(t)$ is defined as

$$S(t) = \exp\left(-\left(\frac{t}{\sigma}\right)^{\alpha}\right). \tag{14}$$

$S(t)$ is equivalent to the complementary cumulative distribution function, $1 - F(t)$ [20].

For right censored observations, instead of calculating the likelihood as normal (Eq. 8) the likelihood of an observation is evaluated using $S(t)$. Conceptually, this approach calculates the likelihood of observing a taxon that existed for at least that long. For left censored data, instead the likelihood is calculated using $1 - S(t)$ which corresponds to the likelihood of observing a taxon that existed no longer than t.

The full likelihood statements incorporating fully observed, right censored, and left censored observations are then

$$\begin{aligned}
\mathcal{L} &\propto \prod_{i \in C} \text{Exponential}(y_i | \lambda) \prod_{j \in R} S(y_j | \lambda) \prod_{k \in L} (1 - S(y_k | \lambda)) \\
\mathcal{L} &\propto \prod_{i \in C} \text{Weibull}(y_i | \alpha, \sigma) \prod_{j \in R} S(y_j | \alpha, \sigma) \prod_{k \in L} (1 - S(y_k | \alpha, \sigma))
\end{aligned} \tag{15}$$

where C is the set of all fully observed taxa, R the set of all right censored taxa, and L the set of all left-censored taxa.

9

2.4   Parameter estimation

The joint posterior was approximated using a Markov-chain Monte Carlo routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn Sampler [12] as implemented in the probabilistic programming language Stan [38]. The estimate of the posterior distribution were approximated from four parallel chains run for 10000 draws split half warm-up and half sampling thinned to every 10th sample for a total of 5000 posterior samples. Chain convergence was assessed via the scale reduction factor $\hat{R}$ where values close to 1 ($\hat{R} < 1.1$) indicate approximate convergence. Convergence means that the chains are approximately stationary and the samples are well mixed [10].

2.5   Model evaluation

Models were evaluated using both posterior predictive checks and an estimate of out-of-sample predictive accuracy. The motivation behind posterior predictive checks as tools for determining model adequacy is that replicated data sets using the fitted model should be similar to the original data [10]. Systematic differences between the simulations and observed indicate weaknesses of the currently fit model. An example of a technique that is very similar would be inspecting the residuals and Q-Q plots from a linear regression.

The strategy behind posterior predictive checks is to draw simulated values from the joint posterior predictive distribution, $p(y^{rep}|y)$, and then compare those draws to the empirically observed values [10]. To accomplish this, for each replicate, a single value is drawn from the marginal posterior distributions of each regression coefficient from the final model as well as α for the Weibull model (Eq. 11, 12). Then, given the covariate information $\mathbf{X}$, a new set of n genus durations are generated giving a single replicated data set y^{rep}. This is repeated 1000 times in order to provide a distribution of possible values that could have been observed given the model.

In order to compare the fitted model to the observed data, various graphical comparisons or test quantities need to be defined. The principal comparison used here is a comparison between non-parameteric approximation of the survival function $S(t)$ as estimated from both the observed data and each of the replicated data sets. The purpose of this comparison is to determine if the model approximates the same survival/extinction pattern as the original data.

The exponential and Weibull models were compared for out-of-sample predictive accuracy using the widely-applicable information criterion (WAIC) [43]. However, because the Weibull model reduces to the exponential model when $\alpha = 1$ my interest is not in choosing between these models. Instead, comparisons of WAIC values are useful for better understanding the effect of model complexity on out-of-sample predictive accuracy. The calculation of WAIC used here corresponds to the "WAIC 2" formulation recommended by Gelman et al. [10].

WAIC can be considered a fully Bayesian alternative to the Akaike information criterion, where WAIC acts as an approximation of leave-one-out cross-validation which acts as a measure of out-of-sample predictive accuracy. WAIC is calculated starting with the log pointwise posterior predictive density calculated as

$$\text{lppd} = \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i | \Theta^S) \right), \tag{16}$$

where n is sample size, S is the number posterior simulation draws, and Θ represents all of the estimated parameters of the model. This is similar to calculating the likelihood of each observation given the entire posterior. A correction for the effective number of parameters is then added to lppd to adjust for overfitting. The effective number of parameters is calculated, following the recommendations of Gelman et al. [10], as

$$p_{\text{WAIC}} = \sum_{i=1}^{n} V_{s=1}^{S} (\log p(y_i | \Theta^S)). \tag{17}$$

where V is the sample posterior variance of the log predictive density for each data point.

Given both equations 16 and 17, WAIC is then calculated

$$\text{WAIC} = \text{lppd} - p_{\text{WAIC}}. \tag{18}$$

When comparing two or more models, lower WAIC values indicate better out-of-sample predictive accuracy. Importantly, WAIC is just one way of comparing models. When combined with posterior predictive checks it is possible to get a more complete understanding of model fit.

3 Results

As stated above, posterior approximations for both the exponential and Weibull models achieved approximate stationarity after 10,000 steps, as all parameter estimates have an $\hat{R} < 1.1$.

Comparisons of the survival functions estimated from 1000 posterior predictive data sets to the estimated survival function of the observed genera demonstrates that both the exponential and Weibull models approximately capture the observed pattern of extinction (Fig. 4). The major difference in fit between the two models is that the Weibull model has a slightly better fit for longer lived taxa than the exponential model.

Additionally, the Weibull model is expected to have slightly better out-of-sample predictive accuracy when compared to the exponential model (3941 versus 3948, respectively). This is congruent with graphical comparisons of the survival
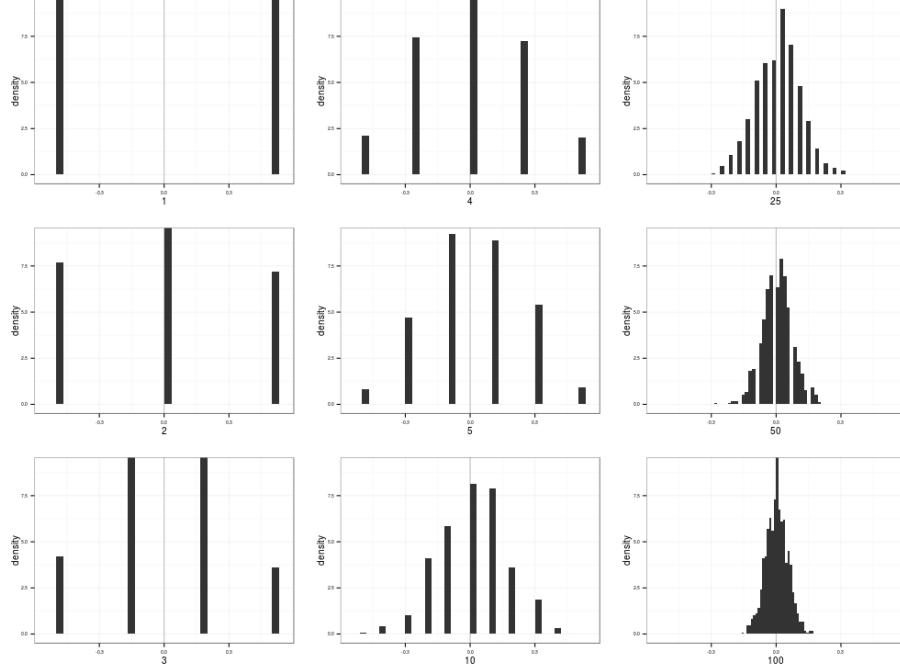
Figure 1: Histograms of the distributions of from the beta distribution defined in Eq. 5. As to be expected, as sample size increases the draws better resemble the underlying true distribution. Sample size is indicated as the label of the x-axis, increasing in column major order.

functions (Fig. 4). Because the difference between the WAIC scores is small, both results from the exponential and Weibull models will be analyzed.

Estimates of the overall overall mean covariate effects μ can be considered time-invariant generalizations for brachiopod survival during the Paleozoic SMITS IN PREP (Fig. 5). Consistent with prior expectations, geographic range size has a negative effect on extinction risk where genera with large ranges having greater durations than genera with small ranges. I also find no time-invariant effect of body size on genus duration.

Interpretation of the effect of environmental preference v on duration is slightly more involved. Because a quadratic term is the equivalent of an interaction term, both μ_v and μ_{v^2} have to be interpreted together because it is illogical to change values of v without also changing values v^2. To determine the nature of the effect of v on duration I calculated the multiplicative effect of environmental preference on extinction risk.

Given mean estimated extinction risk $\tilde{\sigma}$, we can define the extinction risk
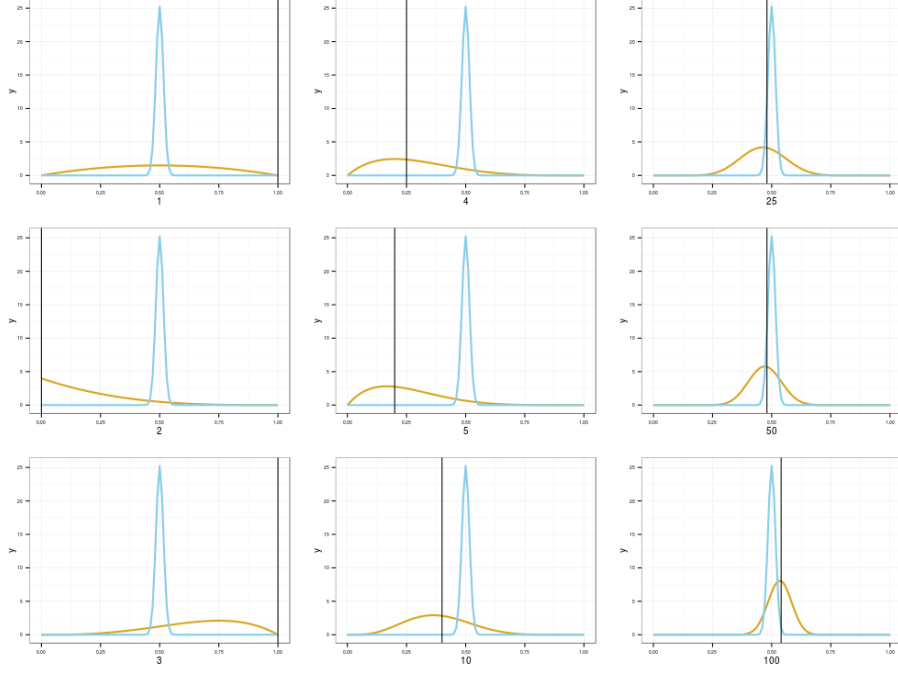
12

Figure 2: Comparisons of the underlying distribution (blue) to posterior estimates based on increasing sample size (gold). Each posterior estimate is represented for only a single realization of draws, each with sample size indicated as the x-axis label (increasing in column major order). Black vertical lines correspond to the MAP estimate of the simulated taxon's affinity. This stands in contrast to the posterior distribution of expected affinity in gold.

multiplier of an observation with environmental preference v_i as

$$\frac{\tilde{\sigma}_i}{\tilde{\sigma}} = f(v_i) = \exp\left(\frac{-(\mu_v v_i + \mu_{v^2} v^2)}{\alpha}\right). \tag{19}$$

This exponentiated quadratic function $f(v_i)$ has a y-intercept of $\exp(0)$ or 1 by definition as it does not have a specified non-zero intercept term. Equation 19 can be either upward or downward facing. A downward facing $f(v_i)$ indicates that genera of intermediate environmental preference have greater durations than either extreme, and *vice versa* for upward facing functions.

The expected effect of environmental preference as a multiplier of expected extinction risk can then be visualized (Fig. 6). This figure depicts 1000 posterior predictive estimates of Eq. 19 across all possible values of v. The number indicates the posterior probability that the function is downward facing, with generalists having lower extinction risk/greater duration than either type of specialist. Note that the inflection point/optimum of Fig. 6 is approximately $x = 0$, something
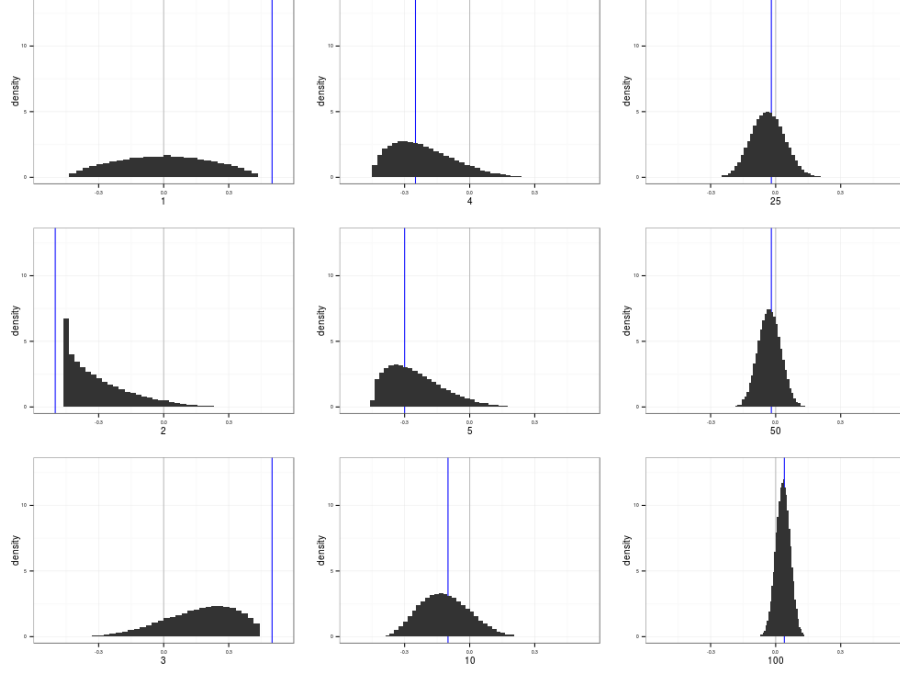
13

Figure 3: Histograms of the difference in the underlying occurrence distribution and the posterior distribution estimates from the previous graph (Fig. 2). The "true" value is included in all distributions of environmental affinities. Each affinity estimate is represented for only a single realization of draws, each with sample size indicated as the x-axis label (increasing in column major order). Blue vertical lines correspond to the difference in MAP estimates between the underlying distribution and the simulated taxon's draws. This stands in contrast to the distribution of the differences between the simulated taxon and background.

that is expected given the estimate of μ_v (Fig. 5).

The matrix Σ describing the covariance between the different coefficients describes how these coefficients might vary together across the origination cohorts. Similar to how this was modeled (Eq. 11, 12), for interpretation purposes Σ can be decomposed into a vector of standard deviations $\vec{\tau}$ and a correlation matrix Ω.

The estimates of the standard deviation of between cohort coefficient estimates τ vary greatly (Fig. 7). Coefficients with greater values of τ have greater between cohort variation. The individual covariate effects with the greatest between origination cohort variation are β_r, β_v, and β_{v^2}. β_m has negligible between cohort variation, as it has less have less between cohort variation than the between cohort variation in baseline extinction risk β_0. However the amount between cohort variation in estimates of β_{v^2} mean that it is possible for the

14

function describing the effect of environmental affinity can be upward facing for some cohorts (Eq. 19), which corresponds to environmental generalists being shorted lived than specialists in that cohort.

The correlation terms of Ω (Fig. 8) describe the relationship between the coefficients and how their estimates may vary together across cohorts. The correlations between the intercept term β_0 and the effects of the taxon traits are of particular interest for evaluating the Jablonski [16] hypothesis (Fig. 8 first column/last row). Keep in mind that when β_0 is low, extinction risk is low; and conversely, when β_0 is high, then extinction risk is high.

Marginal posterior probabilities of the correlation terms between the level of baseline extinction risk β_0 and the effects of the taxon traits indicate that the correlation between between expected extinction risk and geographic range β_r is of particular note (Fig. 9). The negative correlation between β_0 and β_r implies that as extinction risk increases, the effect/importance of geographic range on genus duration increases. This translates to mean that increases in baseline extinction rate are correlated with an increased importance of geographic range size.

While the overall group level estimates are of particular importance when defining time-invariant differences in extinction risk, it is also important and useful to analyze the individual level parameter estimates in order to better understand how parameters actually vary across cohorts.

In comparison to the overall mean extinction risk $\mu_{intercept}$, cohort level estimates β_0 show some amount of variation through time as expected by estimates of $\tau_{intercept}$ (Fig. 10). A similar, if slightly greater, amount of variation is also observable in cohort estimates of the effect of geographic range β_r (Fig. 11). Again, smaller values of β_0 correspond to lower expected extinction risk. Similarly, smaller values of β_r correspond to greater decrease in extinction risk with increasing geographic range

How the effect of environmental affinity varies between cohorts can be observed by using the cohort specific coefficients estimates. Following the exact same procedure used earlier (Fig. 7), but substituting cohort specific estimates of β_v and β_{v^2} for μ_v and μ_{v^2}, the cohort specific effect of environmental preference as a multiplier of mean extinction risk can be calculated. This was done only for the Weibull model, though the observed pattern should be similar for the exponential model.

As expected based on the estimates of τ_v and τ_{v^2}, there is greater variation in the peakedness of $f(v_i$ than there is variation between upward and downward facing functions (Fig. 12). Only 9 of the 31 cohorts have less than 50% posterior probability that generalists are shorter lived than specialists, though 2 of those cases have approximately a 50% probability of being either upward or downward facing. This is congruent with the 0.70+ posterior probability that μ_{v^2} is positive/$f(v_i)$ is downward facing, which corresponds to approximately 8 out of 31 cohorts.

15

Additionally, a quite striking pattern emerges when the inflection point of the function is either far away from the y-intercept (x = 0, y = 1) or when there is little evidence of non-linearity (Fig. 12). Cohort 21 and 20, for example, have almost linear relationships between environmental preference and the duration multiplier. This type of relationship occurs when β_{v^2} approaches 0, flattening the non-linear curvature of the relationship.
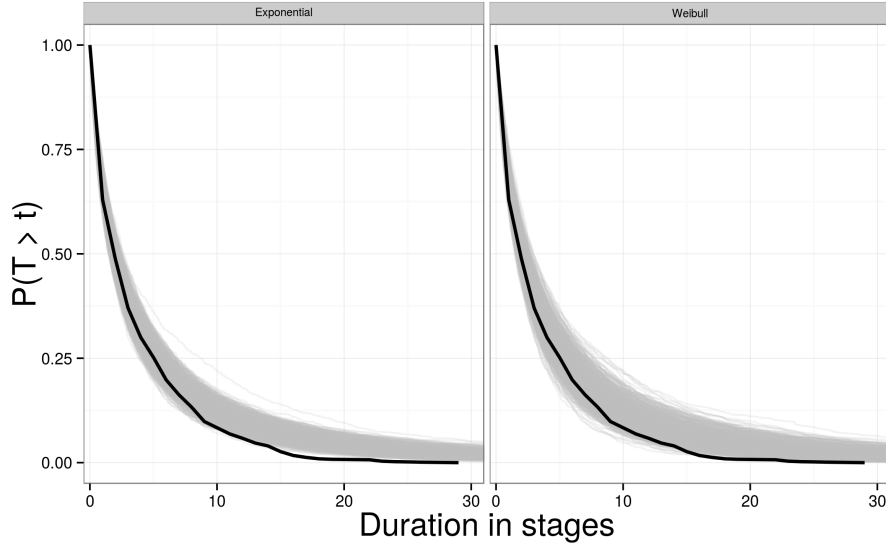

Figure 4: Comparison of empirical estimates of $S(t)$ versus estimates from 1000 posterior predictive data sets. $S(t)$ corresponds to $P(T > t)$ as it is the probability that a given genus observed at age t will continue to live. This is equivalent to the probability that t is less than the genus' ultimate duration T. Note that the Weibull (left) model has noticeably better fir to the data than the exponential (right).

4   Discussion

The results presented here are in an interesting combination of congruence and contrast with those of Jablonski [16]. These results are partially consistent with Jablonski [16] because during periods of higher baseline extinction risk the effect of geographic range is greater than the effects of other taxon traits, though the macroevolutionary mechanism proposed by Jablonski [16] is not supported by my results. Jablonski [16] hypothesized that as extinction risk increases, the effects of all taxon traits except for geographic range size decrease in size. My results instead demonstrate that the only effect that has strong correlation with baseline extinction risk is that of geographic range (Fig. 9). Because of this,
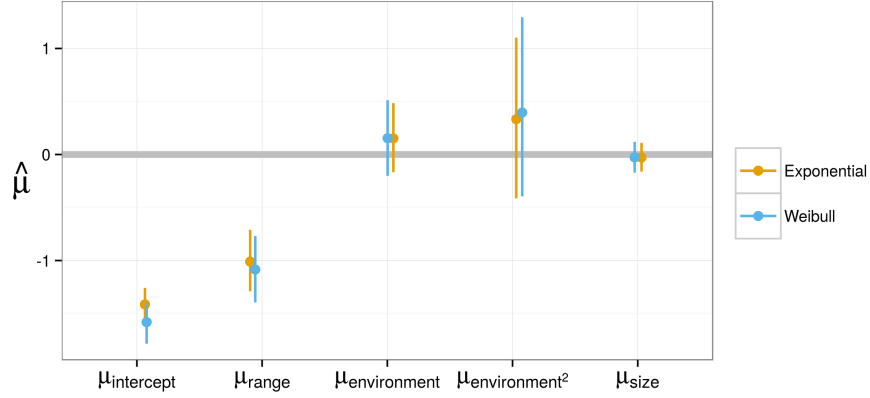
Figure 5: Estimates of the overall effects of the covariates on extinction risk. Included is also the estimate of μ_0 which corresponds to the intercept term or, because of standardization, the overall mean expected extinction risk. Estimates are presented for both the exponential (gold) and Weibull (blue) models. The point corresponds to the median of the posterior distribution, while the error bars correspond to the 80% credible intervals of the estimates.

instead of the effects of other taxon traits decreasing with increasing baseline extinction risk, the effect of geographic range increases while the effects of other taxon traits have no strong correlation with baseline extinction risk. Keep in mind, however, that this is a model of continuous rates and not a model of the differences between background and mass extinction.

There are two mass extinction events that are captured within the time frame considered here: the end Ordovician, and the end Devonian. The nature of the analysis used here means that the effect of the mass extinction is only captured with any clarity by the survival patterns of the origination cohort that immediately proceeds the event. These two cohorts are Hirnantian and Frasnian for the end Ordovician and end Devonian, respectively.

For the Hirnantian, there is a 57% posterior probability that $f(v_i)$ is nonlinear and that taxa which tend to favor epicontinental seas are expected to survive longer than those favoring open ocean environments. Similarly, for the Frasnian there is a 47% posterior probability that $f(v_i)$ is nonlinear and taxa which tend to favor epicontinental seas are expected to survive longer than those favoring open ocean environments. These probabilities were calculated as the percent of posterior draws of $f(v_i)$ which have their optima greater than 0, or towards favoring epicontinental.

In the Frasnian, however, there is evidence that $f(v_i)$ is effectively linear and taxa which favor epicontinental seas will be longer lived than those which favor
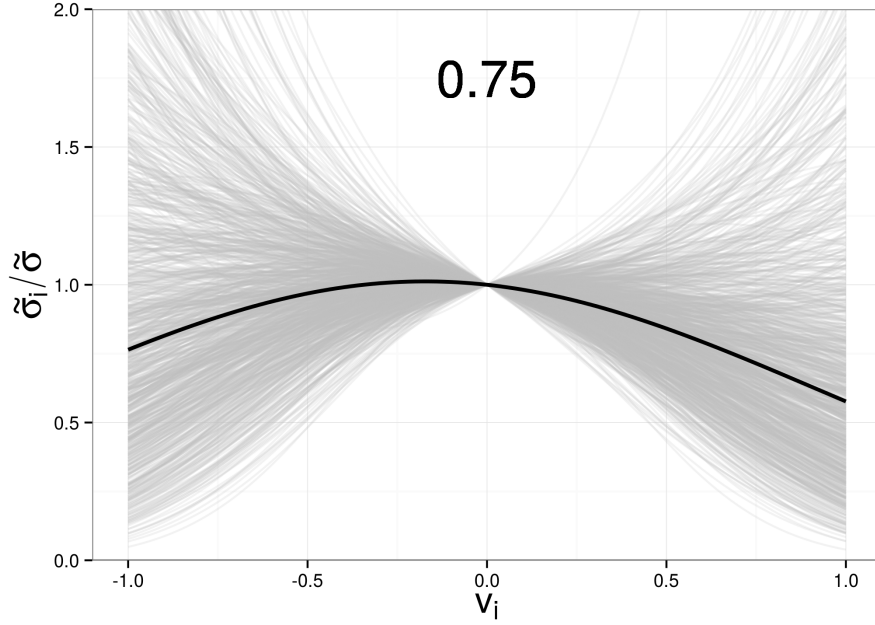
Figure 6: The overall expected relationship $f(v_i)$ between environmental affinity v_i and a multiplier of extinction risk (Eq. 19). Each grey line corresponds to a single draw posterior predictive distribution, while the black corresponds to the median of the posterior predictive distribution. The overall shape of $f(v_i)$ is concave down with an optimum of close 0, which corresponds to affinity approximately equal to the expectation based on background environmental occurrence rates.

open ocean environements (18% prosterior probability). This probability was calculated as the probability that the inflection point/optima of $f(v_i$ will be between -1 and 1 which is a genera desciription of if there is any non-linearity around values corresponding "generalists."

In the stages following the mass extinction, however, this story is slightly different. The Rhuddanian, which just follows the end Ordovician extinction, there is a noticible non-linear relationship, where taxa which favor open open environments tend to have a greater duration multiplier than those which favor epicontinental seas. This is consistent with the proposed mechanism of the end Ordovician mass extinction where the sea level dropped and the epicontinental seas were drained CITATIONS.

For the Fammanian, which follows the end Devonian mass extinction, there is an almost linear relationship where taxa which favor epicontinental seas have a much greater duration multiplier than any other environmental affinity. This
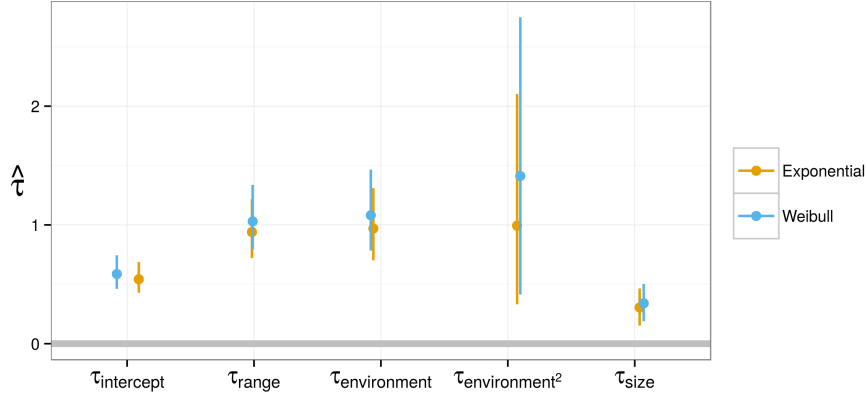
Figure 7: Estimates of the scale parameters describing the expected differences in the effect of the covariates, and of the intercept/baseline extinction risk, between cohorts. Higher values of τ correspond to greater expected differences between cohorts. Estimates are presented for both the exponential (gold) and Weibull (blue) models. The point corresponds to the median of the posterior distribution, while the error bars correspond to the 80% credible intervals of the estimates.

means that the observation of Miller and Foote [25] is a consistent descriptor of the long term extinction risk of taxa originating in stages just following a mass extinction.

These results indicate weak evidence for the observation of Miller and Foote [25] that epicontinental seas are favored at mass extinction boundaries they analyzed. For stages proceeding mass extinction, both affinity end members are approximately equal in extinction risk. For stages following mass extinctions, however, the results observed here for the Fammanian are consistent with those of Miller and Foote [25].

This analysis, however, may not be capturing the patterns analyzed in Miller and Foote [25] as my focus is on long term taxonomic duration between and within origination cohorts and not patterns of instantaenous extinction rates between stages. Additionally, my censoring criteria are specifically for smoothing out this kind of disruption, regardless of cohort. This censoring pattern may be the reason for the difference in relationshi of the stage preceeding a mass extinction and the stage following it. Also, as said before, this is not a model of background versus mass extinction, but instead it is a model of differences in long term extinction risk. Given of these caveats, it is hard to make strong conclusions directly about the results of Miller and Foote [25].

There is an approximate 75% posterior probability that taxa with intermediate environmental preferences have a lesser extinction risk than either end members,
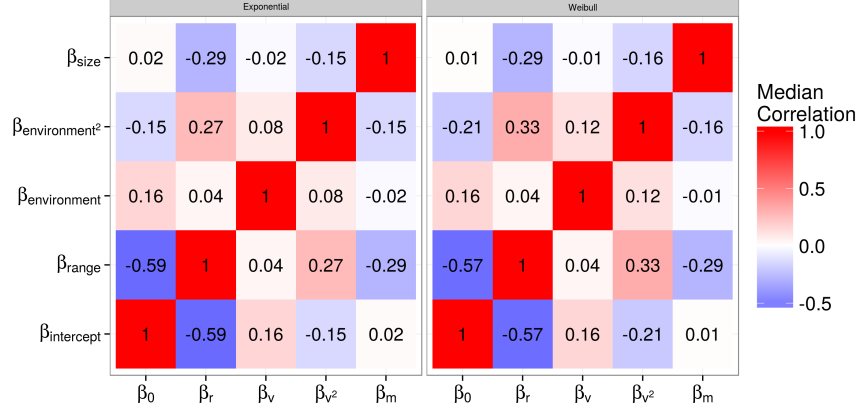
Figure 8: Heatmap for the median estimates of the terms of the correlation matrix Ω between cohort-level covariate effects. Both the exponential (left) and Weibull (right) models are presented. The off-diagonal terms are the correlation between the estimates of the cohort-level estimates of the effects of covariates, along with intercept/baseline extinction risk.

the over all curvature of $f(v_i)$ is not very peaked meaning that this relationship does not lead to very strong differences in extinction risk (Fig. 6). This result supports the hypothesis that, in general, environmental generalists survive greater than environmental specialists [21, 22, 26, 27, 36].

The variance in estimate of the overall $f(v_i)$ reflects the large between cohort variance in cohort specific estimates of $f(v_i)$ (Fig. 12). Given that there is only a 75% posterior probability that the expected overall estimate of $f(v_i)$, it is not surprising that there are some stages where the theorized relationship is in fact reversed. Additionally, as discussed earlier, there are some stages where $f(v_i)$ does not resemble the theorized nonlinear relation with the optimum in the middle but instead is highly skewed or effectively linear (Fig. 12).

These results do not necessarily refute "survival of the unspecialized" as a time-invariant generalization, but instead demonstrate how while the expected group-level estimate of $f(v_i)$ might favor one hypothesis there is still enough variability between cohorts so that in some realizations this pattern may not hold or can even be reversed. These results are also consistent with aspects of Miller and Foote [25] who found that the effect of environmental preference on extinction risk was quite variable and without obvious patterning during times of background extinction.

This model can be improved through either increasing the number of analyzed taxon traits, expanding the hierarchical structure of the model to include other major taxonomic groups of interest, and the inclusion of explicit phylogenetic
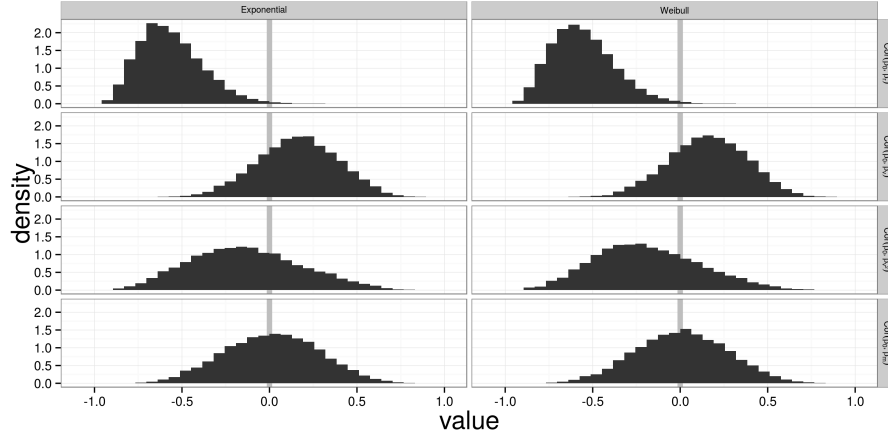
Figure 9: Marginal posterior distributions of the correlations between intercept terms/baseline extinction risk and the effects of each of the covariates. These are presented for both the exponential (left) and Weibull (right) models.

relationships between the taxa in the model as an additional hierarchical effect.

An example taxon trait that may be of particular interest is the affixing strategy or method of interaction with the substrate of the taxon. This trait has been found to be related to brachiopod survival [1] so it's inclusion may be of particular interest.

It is theoretically possible to expand this model to allow for comparisons within and between major taxonomic groups. This approach would better constrain the brachiopod estimates while also allowing for estimation of similarities and differences in cross-taxonomic patterns. The major issue surrounding this particular expansion involves finding an similarly well sampled taxonomic group that is present during the Paleozoic. Example groups include Crinoidea, Ostracoda, and other "Paleozoic" groups [33].

Taxon traits like environmental preference or geographic range [14, 17] are most likely heritable, at least phylogenetically [13, 23]. Without phylogenetic context, this analysis assumes that differences in extinction risk between taxa are independent of those taxa's shared evolutionary history [4]. In contrast, the origination cohorts only capture shared temporal context. The inclusion of phylogenetic context as an addition individual level hierarchical structure independent of origination cohort would allow for determining how much of the observed variability is due to shared evolutionary history versus actual differences associated with these taxonomic traits. It has been shown that phylogeny contribute non-trivially to differences in mammal species durations SMITS IN PREP.

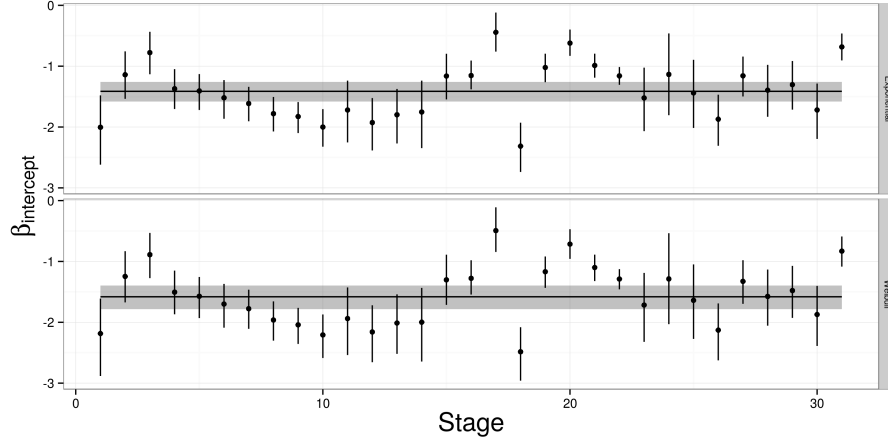In summary, patterns of Paleozoic brachiopod survival were analyzed using a

Figure 10: Comparison of cohort-specific estimates of β_0 presented along with the estimate for the overall baseline extinction risk. Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. The horizontal line is the median estimate of the overall baseline extinction risk along with 80% credible intervals. Results are presented for the exponential (top) and Weibull (bottom) models.

fully Bayesian hierarchical survival modelling approach. Using a varying-slopes, varying-intercepts approach I am able to model both the overall mean effect of biological covariates on extinction risk while also modeling the correlation between origination cohort-specific estimates of covariate effects. I find support for the observation of Jablonski [16] that, for origination cohorts with greater baseline extinction risk, the effect of geographic range on extinction risk is greatest. However, I do not find support for the macroevolutionary mechanism proposed by Jablonski [16] that of the effects of other covariates decreasing with increases in extinction risk. Instead, I find support for the effect of geographic range increasing with increases in baseline extinction risk while the effects of other covariates are uncorrelated with baseline extinction risk. I also find some support for "survival of the unspecialized" Liow [21, 22], Nürnberg and Aberhan [26, 27], Simpson [36] as a general characterization of the effect of environmental preference on extinction risk (Fig. 6), though there is heterogeneity between origination cohorts (Fig. 12). Generally, this study demonstrates the advantages of a hierarchical Bayesian framework for taking into account the structured nature of the data. Future studies of structured data should adopt similar strategies in order to best model our knowledge instead of ignoring that structure which can lead to poor and/or incorrect inference.
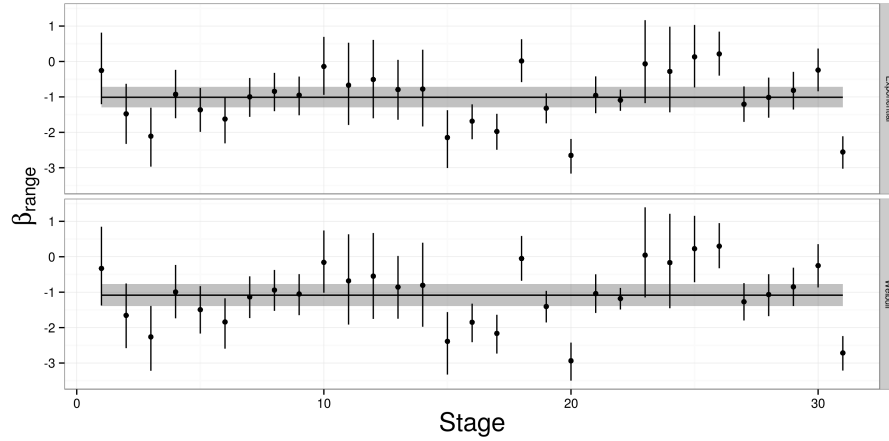
.

Figure 11: Comparison of cohort-specific estimates of the effect of geographic range on extinction risk β_r presented along with the estimate for the overall effect of geographic range. Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. The horizontal line is the median estimate of the overall baseline extinction risk along with 80% credible intervals. Results are presented for the exponential (top) and Weibull (bottom) models.

References

[1] R. R. Alexander. Generic longevity of articulate brachiopods in relation to the mode of stabilization on the substrate. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 21:209–226, 1977.

[2] T. K. Baumiller. Survivorship analysis of Paleozoic Crinoidea: effect of filter morphology on evolutionary rates. *Paleobiology*, 19(3):304–321, 1993.

[3] W. S. Cooper. Expected time to extinction and the concept of fundamental fitness. *Journal of Theoretical Biology*, 107:603–629, 1984.

[4] J. Felsenstein. Phylogenies and the comparative method. *American Naturalist*, 125(1):1–15, 1985. URL http://www.jstor.org/stable/2461605.

[5] M. Foote. Survivorship analysis of Cambrian and Ordovician Trilobites. *Paleobiology*, 14(3):258–271, 1988.

[6] M. Foote. Substrate affinity and diversity dynamics of Paleozoic marine animals. *Paleobiology*, 32(3):345–366, Sept. 2006. ISSN 0094-8373. doi: 10.1666/05062.1. URL http://www.bioone.org/doi/abs/10.1666/05062.1.

[7] M. Foote and A. I. Miller. Determinants of early survival in marine animal genera. *Paleobiology*, 39(2):171–192, Mar. 2013. ISSN 0094-8373. doi: 10.1666/12028. URL http://www.bioone.org/doi/abs/10.1666/12028.

[8] A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533, 2006.

[9] A. Gelman and J. Hill. *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY, 2007.

[10] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall, Boca Raton, FL, 3 edition, 2013.

[11] R. J. Hijmans. *raster: Geographic data analysis and modeling*, 2015. URL `http://CRAN.R-project.org/package=raster`. R package version 2.3-24.

[12] M. D. Hoffman and A. Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1351–1381, 2014.

[13] E. A. Housworth, P. Martins, and M. Lynch. The Phylogenetic Mixed Model. *The American Naturalist*, 163(1):84–96, 2004.

[14] G. Hunt, K. Roy, and D. Jablonski. Species-level heritability reaffirmed: a comment on "On the heritability of geographic range sizes". *American Naturalist*, 166(1):129–135, 2005.

[15] J. G. Ibrahim, M.-H. Chen, and D. Sinha. *Bayesian Survival Analysis*. Springer, New York, 2001.

[16] D. Jablonski. Background and mass extincitons: the alternation of macroevolutionary regimes. *Science*, 231(4734):129–133, 1986.

[17] D. Jablonski. Heritability at the species level: analysis of geographic ranges of cretaceous mollusks. *Science*, 238(4825):360–363, Oct. 1987. ISSN 0036-8075. doi: 10.1126/science.238.4825.360. URL `http://www.ncbi.nlm.nih.gov/pubmed/17837117`.

[18] D. Jablonski and K. Roy. Geographical range and speciation in fossil and living molluscs. *Proceedings. Biological sciences / The Royal Society*, 270 (1513):401–6, Feb. 2003. ISSN 0962-8452. doi: 10.1098/rspb.2002.2243. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1691247&tool=pmcentrez&rendert`

[19] W. Kiessling and M. Aberhan. Environmental determinants of marine benthic biodiversity dynamics through Triassic Jurassic time. *Paleobiology*, 33(3):414–434, 2007.

[20] J. P. Klein and M. L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, 2nd edition, 2003.

[21] L. H. Liow. A test of Simpson's "rule of the survival of the relatively unspecialized" using fossil crinoids. *The American naturalist*, 164 (4):431–43, Oct. 2004. ISSN 1537-5323. doi: 10.1086/423673. URL `http://www.ncbi.nlm.nih.gov/pubmed/15459876`.

[22] L. H. Liow. Does versatility as measured by geographic range, bathymetric range and morphological variability contribute to taxon longevity? *Global Ecology and Biogeography*, 16(1):117–128, 2007. URL `http://doi.wiley.com/10.1111/j.1466-8238.2006.00269.x` `papers2://publication/doi/10.1111/j.1466-8238.2006.00269.x`.

[23] M. Lynch. Methods for the analysis of comparative data in evolutionary biology. *Evolution*, 45(5):1065–1080, 1991.

[24] A. I. Miller and S. R. Connolly. Substrate affinities of higher taxa and the Ordovician Radiation. *Paleobiology*, 27(4):768–778, Dec. 2001. ISSN 0094-8373. doi: 10.1666/0094-8373(2001)027¡0768:SAOHTA¿2.0.CO;2. URL `http://www.bioone.org/doi/abs/10.1666/0094-8373%282001%29027%3C0768%3ASAOHTA%3E2.0.CO%3E`

[25] A. I. Miller and M. Foote. Epicontinental seas versus open-ocean settings: the kinetics of mass extinction and origination. *Science*, 326(5956): 1106–9, Nov. 2009. ISSN 1095-9203. doi: 10.1126/science.1180061. URL `http://www.ncbi.nlm.nih.gov/pubmed/19965428`.

[26] S. Nürnberg and M. Aberhan. Habitat breadth and geographic range predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*, 39(3):360–372, Apr. 2013. ISSN 0094-8373. doi: 10.1666/12047. URL `http://www.bioone.org/doi/abs/10.1666/12047`.

[27] S. Nürnberg and M. Aberhan. Interdependence of specialization and biodiversity in Phanerozoic marine invertebrates. *Nature communications*, 6:6602, Jan. 2015. ISSN 2041-1723. doi: 10.1038/ncomms7602. URL `http://www.ncbi.nlm.nih.gov/pubmed/25779979`.

[28] M. E. Palmer and M. W. Feldman. Survivability is more fundamental than evolvability. *PloS one*, 7(6):e38025, Jan. 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0038025. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3377627&tool=pmcentrez&rendert`

[29] J. L. Payne and S. Finnegan. The effect of geographic range on extinction risk during background and mass extinction. *Proceedings of the National Academy of Sciences*, 104:10506–11, June 2007. ISSN 0027-8424. doi: 10.1073/pnas.0701257104. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1890565&tool=pmcentrez&rendert`

[30] J. L. Payne, N. A. Heim, M. L. Knope, and C. R. Mcclain. Metabolic dominance of bivalves predates brachiopod diversity decline by more than 150 million years. *Proceedings of the Royal Society B*, 281:20133122, 2014.

[31] D. M. Raup. Taxonomic survivorship curves and Van Valen's Law. *Paleobiology*, 1(1):82–96, Jan. 1975. ISSN 0036-8075. doi: 10.1126/science.49.1254.50. URL `http://www.ncbi.nlm.nih.gov/pubmed/17777225`.

[32] D. M. Raup. Cohort Analysis of generic survivorship. *Paleobiology*, 4(1): 1–15, 1978.

[33] J. J. Sepkoski Jr. A factor analytic description of the Phanerozoic marine fossil record. *Paleobiology*, 7(1):36–53, 1981.

[34] C. Simpson. *Levels of selection and large-scale morphological trends*. PhD thesis, University of Chicago, 2006.

[35] C. Simpson and P. G. Harnik. Assessing the role of abundance in marine bivalve extinction over the post-Paleozoic. *Paleobiology*, 35(4):631–647, Dec. 2009. ISSN 0094-8373. doi: 10.1666/0094-8373-35.4.631. URL `http://www.bioone.org/doi/abs/10.1666/0094-8373-35.4.631`.

[36] G. G. Simpson. *Tempo and Mode in Evolution*. Columbia University Press, New York, 1944.

[37] G. G. Simpson. *The Major Features of Evolution*. Columbia University Press, New York, 1953.

[38] Stan Development Team. Stan: A c++ library for probability and sampling, version 2.5.0, 2014. URL `http://mc-stan.org/`.

[39] Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0*, 2014. URL `http://mc-stan.org/`.

[40] L. Van Valen. A new evolutionary law. *Evolutionary Theory*, 1:1–30, 1973. URL `http://ci.nii.ac.jp/naid/10011264287/`.

[41] L. Van Valen. Taxonomic survivorship curves. *Evolutionary Theory*, 4: 129–142, 1979.

[42] S. C. Wang. On the continuity of background and mass extinction. *Paleobiology*, 29(4):455–467, Dec. 2003. ISSN 0094-8373. doi: 10.1666/0094-8373(2003)029¡0455:OTCOBA¿2.0.CO;2. URL `http://www.bioone.org/doi/abs/10.1666/0094-8373%282003%29029%3C0455%3AOTCOBA%3E2.0.CO%3B`

[43] S. Watanabe. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, 11:3571–3594, 2010.
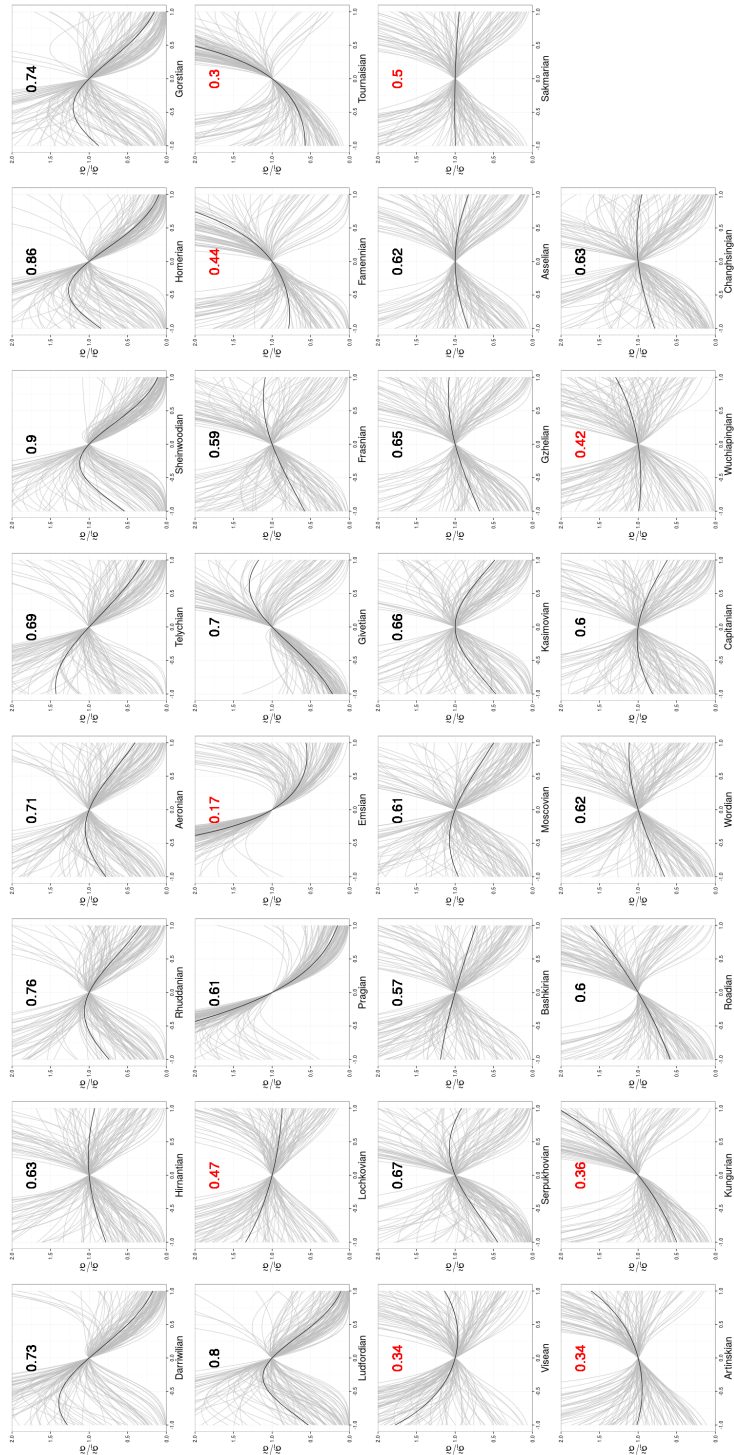
Figure 12: ¡+caption text+¿