

Title: The interplay between extinction intensity and selectivity:
2 correlation in trait effects on taxonomic survival

Running title: Variation in trait effects on taxonomic survival

4 **Author:** Peter D Smits, psmits@uchicago.edu, Committee on Evolutionary
Biology, University of Chicago, IL, USA.

6 **Keywords:** macroevolution, extinction, macroecology, Bayesian, brachiopods

Word count: approximately 5900.

8 **Table count:** 1.

Figure count: 5.

10 **Data archival location:** If accepted, all data and code necessary to duplicate
this analysis will be made available on DRYAD.

Abstract

While the effect of geographic range on extinction risk is well documented, how other traits may increase or decrease extinction risk is less well known. I analyze patterns of Paleozoic brachiopod genus durations and their relationship to geographic range, affinity for epicontinental seas versus open ocean environments, and body size. Additionally, I allow for environmental affinity to have a nonlinear effect on duration. Using a hierarchical Bayesian modeling approach, I also model the interaction between the effects of the biological traits and a taxon's time of origination. My analysis framework eschews the traditional distinction between background and mass extinction, instead the entire time period is analyzed where these "states" are part of a continuum. Contrary to previous work, I find evidence that as baseline extinction risk increases, the effects of biological traits (except body size) on extinction risk tend to increase. For parts of the Paleozoic I find support for a "survival of the generalists" scenario, there are times where this relationship is absent or even reversed. Importantly, I find that as baseline extinction risk increases, the steepness of this relationship is expected to increase as well.

1 Introduction

How do biological traits affect extinction risk? Jablonski (1986) observed that at a mass extinction event, the effects of biological traits on taxonomic survival decreased in size. However, this pattern was not the case for the effect of geographic range on survival (Jablonski, 1986). Biological traits are defined here as descriptors of a taxon's adaptive zone, which is the set of all biotic–biotic and biotic–abiotic interactions that a taxon can experience (Simpson, 1944). In effect, these are descriptors of a taxon's broad-sense ecology.

Jablonski (1986) phrased his conclusions in terms of background versus mass extinction, but this scenario is readily transferable to a continuous variation framework as there is no obvious distinction in terms of extinction rate between these two states (Wang, 2003). Additionally, the Jablonski (1986) scenario has strong model structure requirements in order to test its proposed macroevolutionary mechanism; not only do the taxon trait effects need to be modeled, but the correlation between trait effects need to be modeled as well.

There are two end-member macroevolutionary mechanisms which may underlie the pattern observed by Jablonski (1986): the effect of geographic range on predictive survival remains constant and those of other biological traits decrease, and the effect of geographic range in predicting survival increases and those of other biological traits stay constant. Reality, of course, may fall somewhere along this continuum.

I choose to model brachiopod taxon durations because trait based differences in extinction risk should manifest as differences in taxon durations. Namely, a taxon with a beneficial trait should persist for longer, on average, than a taxon without that beneficial trait. Conceptually, taxon survival can be considered an aspect of “taxon fitness” along with expected lineage specific branching/origination rate (Cooper, 1984, Palmer and Feldman, 2012).

Brachiopods are an ideal group for this study as they are well known for having an exceptionally complete fossil record (Foote, 2000). Specifically, I focus on the brachiopod record from most of the Paleozoic, from the start of the Ordovician (approximately 485 Mya) through the end Permian (approximately 252 Mya) as this represents the time of greatest global brachiopod diversity (Alroy, 2010).

The analysis of taxon durations, or time from origination to extinction, falls under the purview of survival analysis, a field of applied statistics commonly

- 66 used in health care (Klein and Moeschberger, 2003) but has a long history in
paleontology (Simpson, 1944, 1953, Van Valen, 1973, 1979).
- 68 Geographic range is widely considered the most important taxon trait for
estimating differences in extinction risk at nearly all times with large geographic
70 range associated with low extinction risk (Jablonski, 1986, 1987, Jablonski and
Roy, 2003, Payne and Finnegan, 2007), though Foote and Miller (2013) find
72 that this generalization does not hold in the Mesozoic. For the Paleozoic,
however, I expect this to hold true for the entire period analyzed.
- 74 Miller and Foote (2009) demonstrated that during several mass extinctions taxa
associated with open-ocean environments tend to have a greater extinction risk
76 than those taxa associated with epicontinental seas. During periods of
background extinction, however, they found no consistent difference between
78 taxa favoring either environment. These two environment types represent the
primary environmental dichotomy observed in ancient marine systems (Miller
80 and Foote, 2009, Peters, 2008, Sheehan, 2001).

Epicontinental seas are a shallow-marine environment where the ocean has
82 spread over the surface of a continental shelf with a depth typically less than
100m. In contrast, open-ocean coastline environments have much greater
84 variance in depth, do not cover the continental shelf, and can persist during
periods of low sea level. Because of this, it is strongly expected that taxa which
86 favor epicontinental seas would be at great risk during periods of low sea levels,
such as during glacial periods, where these seas are drained. During the
88 Paleozoic (approximately 541–252 My), epicontinental seas were widely spread
globally but declined over the Mesozoic (approximately 252–66 My) and
90 eventually nearly disappeared during the Cenozoic (approximately 66–0 My) as
open-ocean coastlines became the dominant shallow-marine setting (Johnson,
92 1974, Miller and Foote, 2009, Peters, 2008).

Given the above findings, I predict that as extinction risk increases, taxa
94 associated with open-ocean environments should generally increase in extinction
risk versus those that favor epicontinental seas. Additionally, there is a possible
96 nonlinear relationship between environmental preference and taxon duration. A
long standing hypothesis is that generalists or unspecialized taxa will have
98 greater survival than specialists (Baumiller, 1993, Liow, 2004, 2007, Nürnberg
and Aberhan, 2013, 2015, Simpson, 1944). In this analysis I allowed for
100 environmental preference to have a parabolic effect on taxon duration

Body size, measured as shell length (Payne et al., 2014), was also considered as
102 a potentially informative covariate. Body size is a proxy for metabolic activity
and other correlated life history traits (Payne et al., 2014). There is no strong
104 hypothesis of how body size effects extinction risk in brachiopods, such that a
positive, negative, or zero effect are all plausible.

106 I adopt a hierarchical Bayesian survival modeling approach, which represents
both a conceptual and statistical unification of the paleontological dynamic and
108 cohort survival analytic approaches (Baumiller, 1993, Foote, 1988, Raup, 1975,
1978, Simpson, 2006, Van Valen, 1973, 1979). By using a Bayesian framework I
110 am able to quantify the uncertainty inherent in the estimates of the effects of
biological traits on survival, especially in cases where the covariates of interest
112 (i.e. biological traits) are themselves known with error.

2 Materials and Methods

114 2.1 Fossil occurrence information

The dataset analyzed here was sourced from the Paleobiology Database
116 (<http://www.paleodb.org>) which was then filtered based on taxonomic,

temporal, stratigraphic, and other occurrence information that was necessary
118 for this analysis. These filtering criteria are very similar to those from Foote and
Miller (2013) with an additional constraint of being present in the body size
120 data set from Payne et al. (2014). Epicontinental versus open-ocean assignments
for each fossil occurrence are partially based on those from Miller and Foote
122 (2009), with additional occurrences assigned similarly (Miller and Foote,
personal communication).

124 Fossil occurrences were analyzed at the genus level which is common for
paleobiological, macroevolution, or macroecological studies of marine
126 invertebrates (Alroy, 2010, Foote and Miller, 2013, Harnik et al., 2013, Kiessling
and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015,
128 Payne and Finnegan, 2007, Simpson and Harnik, 2009, Vilhena et al., 2013).

While species diversity dynamics tend to be of much greater interest than those
130 of higher taxa, the nature of the fossil record makes accurate and precise
taxonomic assignments at the species level for all occurrences extremely difficult
132 if not impossible. Additionally, there is evidence of real differences in biological
patterns at the genus level versus the species level (Jablonski, 1987). As such,
134 the choice to analyze genera as opposed to species was in order to assure a
minimum level of confidence and accuracy in the data analyzed here.

136 Genus duration was calculated as the number of geologic stages from first
appearance to last appearance, inclusive. Durations were based on geologic
138 stages as opposed to millions of years because of the inherently discrete nature
of the fossil record; dates are not assigned to fossils themselves but instead
140 fossils are known from a geological interval which represents some temporal
range. Stages act as effectively irreducible globally consistent temporal intervals
142 in which taxa occur.

Genera with a last occurrence in or after Changhsingian stage were right

¹⁴⁴ censored at the Changhsingian. Genera with a duration of only one stage were
left censored (Appendix A). The covariates used to model genus duration were
¹⁴⁶ geographic range size (r), environmental preference (v, v^2), and body size (m).

¹⁴⁸ 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
¹⁵⁰ grid placed on the map. Each grid cell represents approximately 250,000 km².
The map projection and regular lattice were made using shape files from
¹⁵² <http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans,
2015).

¹⁵⁴ For each stage, the total number of occupied grid cells, or cells in which a fossil
occurs, was calculated. Then, for each genus, the number of grid cells occupied
¹⁵⁶ by that genus was calculated. Dividing the genus occupancy by the total
occupancy gives the relative occupancy of that genus. Mean relative genus
¹⁵⁸ occupancy was then calculated as the mean of the per stage relative occupancies
of that genus.

¹⁶⁰ Body size data was sourced directly from Payne et al. (2014). Because those
measurements are presented without error, a measurement error model similar
¹⁶² to the one for environmental affinity could not be implemented (Appendix A).

Prior to analysis, geographic range and body size were transformed and
¹⁶⁴ standardized in order to improve interpretability of the results. Geographic
range, which can only vary between 0 and 1, was logit transformed. Body size,
¹⁶⁶ 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
¹⁶⁸ their standard deviation following Gelman and Hill (2007).

2.2 Analytical approach

¹⁷⁰ Hierarchical modelling is a statistical approach which explicitly takes into account the structure of the observed data in order to model both the within ¹⁷² and between group variance (Gelman et al., 2013, Gelman and Hill, 2007). The units of study (e.g. genera) each belong to a single grouping (e.g. origination ¹⁷⁴ cohort). These groups are considered separate draws from a shared probability distribution (e.g. all cohorts, observed and unobserved). The group-level ¹⁷⁶ parameters are then estimated simultaneously as the other parameters of interest (e.g. covariate effects) (Gelman et al., 2013). The subsequent estimates ¹⁷⁸ are partially pooled together, where parameters from groups with large samples or effects remain large while those of groups with small samples or effects are ¹⁸⁰ pulled towards the overall group mean.

This partial pooling is one of the greatest advantages of hierarchical modeling. ¹⁸² By letting the groups “support” each other, parameter estimates then better reflect our statistical uncertainty. Additionally, this partial pooling helps control ¹⁸⁴ for multiple comparisons and possibly spurious results as effects with little support are drawn towards the overall group mean (Gelman et al., 2013, ¹⁸⁶ Gelman and Hill, 2007).

All covariate effects (regression coefficients), as well as the intercept term ¹⁸⁸ (baseline extinction risk), were allowed to vary by group (origination cohort). The covariance/correlation between covariate effects was also modeled. This ¹⁹⁰ hierarchical structure allows inference for how covariates effects may change with respect to each other while simultaneously estimating the effects ¹⁹² themselves, propagating our uncertainty through all estimates.

Additionally, instead of relying on point estimates of environmental affinity, I ¹⁹⁴ treat environmental affinity as a continuous measure of the difference between

the taxon's environmental occurrence pattern and the background occurrence
196 pattern (Appendix A).

2.3 Survival model

198 Genus durations were assumed to follow either an exponential or Weibull
distribution, both of which make different assumptions about how a taxon's
200 duration may effect its instantaneous extinction risk (Klein and Moeschberger,
2003). The exponential distribution assumes that extinction risk is independent
202 of duration. In contrast, the Weibull distribution allows for age dependent
extinction via the shape parameter α , though only as a monotonic function of
204 duration. Importantly, the Weibull distribution is equivalent to the exponential
distribution when $\alpha = 1$.

206 The following variables are here defined: y_i is the duration of genus i in geologic
stages, X is the matrix of covariates including a constant term, B_j is the vector
208 of regression coefficients for origination cohort j , Σ is the covariance matrix of
the regression coefficients, τ is the vector of scales the standard deviations of
210 the between-cohort variation in regression coefficient estimates, Ω is the
correlation matrix of the regression coefficients, and α_j is the shape parameter
212 for cohort j with a is the overall mean shape parameter and π is the variance
between estimates of α_j .

²¹⁴ The exponential model is defined

$$y_i \sim \text{Exponential}(\lambda)$$

$$\lambda_i = \exp(\mathbf{X}_i B_{j[i]})$$

$$B \sim \text{MVN}(\vec{\mu}, \Sigma)$$

$$\Sigma = \text{Diag}(\vec{\tau}) \Omega \text{Diag}(\vec{\tau})$$

$$\mu_k \sim \begin{cases} \mathcal{N}(0, \psi_k \nu) & \text{if } k \neq r, \text{ or} \\ \mathcal{N}(-1, 1) & \text{if } k = r \end{cases} \quad (1)$$

$$\tau_k \sim C^+(1)$$

$$\psi_k \sim C^+(1) \text{ if } k \neq r$$

$$\nu \sim C^+(1)$$

$$\Omega \sim \text{LKJ}(2).$$

Similarly, the Weibull model is defined

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

²¹⁶ The principal difference between this model and the previous (Eq. 1) is the inclusion of the shape parameter α . Note that σ is approximately equivalent to ²¹⁸ $1/\lambda$.

For an explanation of how this model was developed, parameter explanations, ²²⁰ and choice of priors, please see Appendix B. Note that these models (Eq. 1, 2) do not include how the uncertainty in environmental affinity is included nor how ²²² censored observations are included. For an explanation of both of these aspects, see Appendices A and C.

²²⁴ **2.4 Parameter estimation**

The joint posterior was approximated using a Markov chain Monte Carlo
²²⁶ routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn
Sampler (Hoffman and Gelman, 2014) as implemented in the probabilistic
²²⁸ programming language Stan (Stan Development Team, 2014). The posterior
distribution was approximated from four parallel chains run for 10,000 draws
²³⁰ each, split half warm-up and half sampling and thinned to every 10th sample for
a total of 5000 posterior samples. Chain convergence was assessed via the scale
²³² reduction factor \hat{R} where values close to 1 ($\hat{R} < 1.1$) indicate approximate
convergence, meaning that the chains are approximately stationary and the
²³⁴ samples are well mixed (Gelman et al., 2013).

2.5 Model evaluation

²³⁶ Models were evaluated using both posterior predictive checks and an estimate of
out-of-sample predictive accuracy. The motivation behind posterior predictive
²³⁸ checks as tools for determining model adequacy is that replicated data sets
using the fitted model should be similar to the original data (Gelman et al.,
²⁴⁰ 2013). Systematic differences between the simulations and observations indicate
weaknesses of the model fit. An example of a technique that is very similar
²⁴² would be inspecting the residuals from a linear regression.

The strategy behind posterior predictive checks is to draw simulated values
²⁴⁴ from the joint posterior predictive distribution, $p(y^{rep}|y)$, and then compare
those draws to the empirically observed values (Gelman et al., 2013). To
²⁴⁶ 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
²⁴⁸ as estimates of α_j for the Weibull model (Eq. 1, 2). Then, given the covariate

information \mathbf{X} , a new set of n genus durations are generated giving a single
250 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.

252 In order to compare the fitted model to the observed data, various graphical comparisons or test quantities need to be defined. The principal comparison
254 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
256 replicated data sets. The purpose of this comparison is to determine if the model approximates the same survival/extinction pattern as the original data.

258 The exponential and Weibull models were compared for out-of-sample predictive accuracy using the widely-applicable information criterion (WAIC) (Watanabe,
260 2010). Out-of-sample predictive accuracy is a measure of the expected fit of the model to new data. However, because the Weibull distribution reduces to the
262 exponential distribution when $\alpha = 1$, my interest is not in choosing between these models. Instead, comparisons of WAIC values are useful for better
264 understanding the effect of model complexity on out-of-sample predictive accuracy. The calculation of WAIC used here corresponds to the “WAIC 2”
266 formulation recommended by Gelman et al. (2013). For an explanation of how
268 WAIC is calculated, see Appendix D. Lower values of WAIC indicate greater expected out-of-sample predictive accuracy than higher values.

3 Results

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

Comparisons of the survival functions estimated from 1000 posterior predictive
274 data sets to the estimated survival function of the observed genera demonstrates
that both the exponential and Weibull models approximately capture the
276 observed pattern of extinction (Fig. 1). This is partially expected given that the
unit of analysis is genus species duration as opposed species level which can
278 alter the shape of $S(t)$ (Foote, 2001, Raup, 1975, 1978, 1985). The major
difference in fit between the two models is that the Weibull model has a slightly
280 better fit for longer lived taxa than the exponential model.

Additionally, the Weibull model is expected to have slightly better
282 out-of-sample predictive accuracy when compared to the exponential model
(WAIC 4534 versus 4604). Because the difference in WAIC between these two
284 models is large, only those results from the Weibull model will be discussed.

Estimates of the overall mean covariate effects μ can be considered
286 time-invariant generalizations for brachiopod survival during the Paleozoic (Fig.
1). Consistent with prior expectations, geographic range size has a negative
288 effect on extinction risk, where genera with large ranges having greater
durations than genera with small ranges.

290 I infer that body size has no distinguishable effect on brachiopod taxonomic
survival. The mean estimate of the effect of body size on extinction risk is
292 negative, implying that increasing body size decreases extinction risk, however
this estimate is within 2 standard deviations of 0 (mean $\mu_m = -0.07$, standard
294 deviation 0.08; Fig. 1).

Interpretation of the effect of environmental preference v on duration is slightly
296 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
298 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
300 preference on extinction risk.

Given mean estimated extinction risk $\tilde{\sigma}$, we can define the extinction risk
302 multiplier of an observation with environmental preference v_i as

$$f(v_i) = \exp\left(\frac{-(\mu_v v_i + \mu_{v^2} v^2)}{\exp(a)}\right). \quad (3)$$

This function $f(v_i)$ has a y-intercept of $\exp(0)$ or 1 because it does not have a
304 non-zero intercept term. Equation 3 can be either concave up or down. A
concave down $f(v_i)$ may indicate that genera of intermediate environmental
306 preference have greater durations than either extreme, and *vice versa* for
concave up function.

308 The expected effect of environmental preference as a multiplier of expected
extinction risk can then be visualized (Fig. 2). This figure depicts 1000 posterior
310 predictive estimates of Eq. 3 across all possible values of v . The number
indicates the posterior probability that the function is concave down, with
312 generalists having lower extinction risk/greater duration than either type of
specialist. Note that the inflection point of Fig. 2 is approximately $x = 0$,
314 something that is expected given the estimate of μ_v (Fig. 1).

The matrix Σ describing the covariance between the different coefficients
316 describes how these coefficients might vary together across the origination
cohorts. Similar to how this was modeled (Eq. 1, 2), for interpretation purposes
318 Σ is decomposed into a vector of standard deviations $\vec{\tau}$ and a correlation matrix
 Ω .

320 The estimates of the standard deviation of between-cohort coefficient estimates
 τ indicate that some effects can vary greatly between-cohorts (Fig. 1).
322 Coefficients with greater values of τ have greater between-cohort variation. The

covariate effects with the greatest between origination cohort variation are β_r ,
324 β_v , and β_{v^2} . Estimates of β_m have negligible between cohort variation, as there
is less between cohort variation than the between cohort variation in baseline
326 extinction risk β_0 . However the amount of between cohort variation in estimates
of β_{v^2} means that it is possible for the function describing the effect of
328 environmental affinity to be upward facing for some cohorts (Eq. 3), which
corresponds to environmental generalists being shorter lived than specialists in
330 that cohort.

The correlation terms of Ω (Fig. 3a) describe the relationship between the
332 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
334 of particular interest for evaluating the Jablonski (1986) scenario (Fig. 3a first
column/last row). Keep in mind that when β_0 is low, extinction risk is low; and
336 conversely, when β_0 is high, then extinction risk is high.

Marginal posterior probabilities of the correlations between the level of baseline
338 extinction risk β_0 and the effects of the taxon traits indicate that the correlation
between expected extinction risk and both geographic range β_r and β_{v^2} are of
340 particular note (Fig. 3b).

There is only an approximately 85% probability that β_0 and β_r are negatively
342 correlated (Fig. 3b), meaning that as extinction risk increases, the
effect/importance of geographic range on genus duration may increases. This
344 means that increases in baseline extinction rate are weakly correlated with an
increased importance of geographic range size. There is a 93% probability that
346 β_0 and β_{v^2} are negatively correlated (Fig. 3b), meaning that as extinction risk
increases, the peakedness of $f(v_i)$ may increases and the relationship would
348 then tend towards concave down. Additionally, there is a 94% probability that
values of β_r and β_{v^2} are positively correlated (Mean correlation 0.45, standard

350 deviation 0.26).

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

In comparison to the overall mean extinction risk μ_0 , cohort level estimates β_0
356 show some amount of variation through time as expected by estimates of τ_0
(Fig. 4a). A similar, if slightly greater, amount of variation is also observable in
358 cohort estimates of the effect of geographic range β_r (Fig. 4b). Again, smaller
values of β_0 correspond to lower expected extinction risk. Similarly, smaller
360 values of β_r correspond to greater decrease in extinction risk with increasing
geographic range

362 How the effect of environmental affinity varies between cohorts can be observed
by using the cohort specific coefficients estimates. Following the same procedure
364 used earlier (Fig. 1), 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
366 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.

368 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 convave up and down
370 functions (Fig. 5). 12 of the 33 cohorts have less than 50% posterior probability
that generalists are potentially expected to be shorter lived than specialists,
372 though two of those cases have approximately a 50% probability of being either
concave up or down. This is congruent with the 0.74 posterior probability that
374 μ_{v^2} is positive/ $f(v_i)$ is concave down.

Additionally, for some cohorts there is a quite striking pattern where the effect

³⁷⁶ of environmental preference v has a nearly-linear relationship (Fig. 5). These are primarily scenarios where one of the end member preferences is expected to have a greater duration than either intermediate or the opposite end member preference. Whatever curvature is present in these nearly-linear cases is due to the definition of $f(v)$ as it is not defined for non-negative values of σ (Eq. 3). For all stages between the Emsian through the Viséan, inclusive, intermediate preferences are of intermediate extinction risk when compared with epicontinental specialists (lowest risk) or open-ocean specialists (highest risk). This time period represents most of the Devonian through the early Carboniferous.

³⁸⁴ Interestingly, the estimates of α_j have low between cohort variation with most cohorts having approximately the same estimate as the overall estimate of α (Fig. 4c). Also, for the majority of the time analyzed there is evidence for accelerating extinction with taxon age. If this is due to biological causes (Rosindell et al., 2015, Wagner and Estabrook, 2014) or an artifact of preservation/the minimum observable duration of a genus (Sepkoski, 1975) is unknown. While it is expected that a non-exponential model will better fit genus-level data, this would be reflected as decelerating extinction risk with taxon age as opposed to accelerating (Foote, 2001, Raup, 1975, 1978, 1985). This uncertainty remains an open issue for future analysis.

³⁹⁶ 4 Discussion

My results demonstrate that both the effects of geographic range and the peakedness/concavity of environmental preference are both negatively correlated with baseline extinction risk, meaning that as baseline extinction risk increases the effect sizes of both these traits are expected to increase (Fig. 3b). This

result supports neither of the two proposed end-member macroevolutionary
402 mechanisms for how biological traits should correlate with extinction risk. The
observed correlation between the two effects as well as between the effects and
404 baseline extinction risk instead implies that as baseline extinction risk increases,
the strength of the total selection gradient on biological traits (except body size)
406 increases. This manifests as greater differences in extinction risk for each unit
difference in the biological covariates during periods of high extinction risk,
408 while a relatively flatter selection gradient during periods of low extinction risk.

For the approximately 233 My period analyzed there is an approximate 75%
410 posterior probability that brachiopod genera with intermediate environmental
preferences are expected to have a lower extinction risk than either end
412 members. However, the over all curvature of $f(v_i)$ is not very peaked meaning
that when averaged over the entire Phanerozoic this relationship may not lead
414 to large differences in extinction risk (Fig. 2). Note that the duration of the
period analyzed is approximately four times then length of the Cenozoic (e.g.
416 time since the extinction of the non-avian dinosaurs). This result gives weak
support for the universality of the hypothesis that environmental generalists
418 have greater survival than environmental specialists (Liow, 2004, 2007,
Nürnberg and Aberhan, 2013, 2015, Simpson, 1944).

420 The posterior variance in the estimate of overall $f(v_i)$ reflects the large between
cohort variance in cohort specific estimates of $f(v_i)$ (Fig. 5). Given that there is
422 only a 75% posterior probability that the expected overall estimate of $f(v_i)$ is
concave down, it is not surprising that there are some stages where the
424 estimated relationship is in fact the reverse of the prior expectation.
Additionally, some of those same stages where $f(v_i)$ does not resemble the prior
426 expectation of a concave down nonlinear relation are instead is highly skewed
and effectively linear (Fig. 5). These results demonstrate that, while the

428 group-level estimate may only weakly support one hypothesis, the cohort-level
estimates may exhibit very different characteristics. These results are also
430 consistent with aspects of Miller and Foote (2009) who found that the effect of
environmental preference on extinction risk was quite variable and without
432 obvious patterning during times of background extinction.

There are two mass extinction events that are captured within the time frame
434 considered here: the Ordovician-Silurian and the Frasnian-Famennian. The
cohorts bracketing these events are worth considering in more detail.

436 The proposed mechanism for the end Ordovician mass extinction is a decrease
in sea level and the draining of epicontinental seas due to protracted glaciation
438 (Johnson, 1974, Sheehan, 2001). My results are broadly consistent with this
scenario with both epicontinental and open-ocean specialists having a much
440 lower expected duration than intermediate taxa (Fig. 5). All of the stages
between the Darriwillian and the Llandovery, except the Hirnantian, have a
442 high probability (90+%) that $f(v)$ is concave down. The pattern for the
Darriwillian, which proceeds the supposed start of Ordovician glacial activity,
444 demonstrates that taxa tend to favor open-ocean environments are expected to
have a greater duration than either intermediate or epicontinental specialists, in
446 decreasing order.

For nearly the entire Devonian estimates of $f(v)$ indicate that one of the
448 environmental end members is favored over the other end member of
intermediate preference (Fig. 5). For almost the entirety the Givetian though
450 the end of the Devonian and into the Viséan, I find that epicontinental favoring
taxa are expected to have a greater duration than either intermediate or
452 open-ocean specialists. Additionally, for nearly the entire Devonian and through
to the Visean, the cohort-specific estimates of $f(v)$ are concave-up. This is the
454 opposite pattern than what is expected (Fig. 2). This result, however, seems to

reflect the intensity of the seemingly nearly-linear difference in expected
456 duration across the range of v as opposed to an inversion of the weakly expected
curvilinear pattern.

458 The use of genera as the unit of the study and how to exactly interpret the
effects of the biological traits is a remaining concern. For example, if any of the
460 traits analyzed here are associated with increases in speciation rates, this might
increase the duration of genera through self-renewal (Raup, 1991, 1994), which
462 would be an example of the difference in biological pattern between species and
genera (Jablonski, 1987, 2007, 2008). This could lead to a trait appearing to
464 decrease generic level extinction risk by increasing species level origination rate
instead of decreasing species level extinction risk. However, given the nature of
466 the brachiopod fossil record and the difficulty of identifying individual
specimens to the species level, there is no simple solution to decreasing this
468 uncertainty in the interpretations of how the biological traits studied here
actually affect extinction risk.

470 This model could be improved through either increasing the number of analyzed
taxon traits, expanding the hierarchical structure of the model to include other
472 major taxonomic groups of interest, and the inclusion of explicit phylogenetic
relationships between the taxa in the model as an additional hierarchical effect.

474 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
476 been found to be related to brachiopod survival (Alexander, 1977) so its
inclusion may be of particular interest.

478 It is theoretically possible to expand this model to allow for comparisons within
and between major taxonomic groups. This approach would better constrain the
480 brachiopod estimates while also allowing for estimation of similarities and

differences in cross-taxonomic patterns. The major issue surrounding this
482 particular expansion involves finding an similarly well sampled taxonomic group
that is present during the Paleozoic. Example groups include Crinoidea,
484 Ostracoda, and other members of the “Paleozoic fauna” (Sepkoski Jr., 1981).

Taxon traits like environmental preference or geographic range (Hunt et al.,
486 2005, Jablonski, 1987) are most likely heritable, at least phylogenetically
(Housworth et al., 2004, Lynch, 1991). Without phylogenetic context, this
488 analysis assumes that differences in extinction risk between taxa are
independent of those taxa's shared evolutionary history (Felsenstein, 1985). In
490 contrast, the origination cohorts only capture shared temporal context. The
inclusion of phylogenetic context as an addition individual level hierarchical
492 structure independent of origination cohort would allow for determining how
much of the observed variability is due to shared evolutionary history versus
494 actual differences associated with these taxonomic traits.

In summary, patterns of Paleozoic brachiopod survival were analyzed using a
496 fully Bayesian hierarchical survival modelling approach while also eschewing the
traditional separation between background and mass extinction. I modeled both
498 the overall mean effect of biological covariates on extinction risk while also
modeling the correlation between cohort-specific estimates of covariate effects. I
500 find that as baseline extinction risk increases, the strength of the selection
gradient on biological traits (except body size) increases. This manifests as
502 greater differences in extinction risk for each unit difference in the biological
covariates during periods of high extinction risk, while a much flatter total
504 selection gradient during periods of low extinction risk. I also find very weak for
“survival of the unspecialized” (Liow, 2004, 2007, Nürnberg and Aberhan, 2013,
506 2015, Simpson, 1944) as a general characterization of the effect of environmental
preference on extinction risk (Fig. 2), though there is heterogeneity between

⁵⁰⁸ origination cohorts with most periods of time conforming to this hypothesis
(Fig. 5).

⁵¹⁰ **Acknowledgements**

I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for
⁵¹² helpful and Marities Marites Villarosa Garcia for draft comments. Additionally,
thank you A. Miller for the epicontinental versus open-ocean assignments. This
⁵¹⁴ entire study would not have been possible without the Herculean effort of
the many contributors to the Paleobiology Database. In particular, I would like
⁵¹⁶ to thank J. Alroy, M. Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N. Heim,
A. Hendy, S. Holland, L. Ivany, W. Kiessling, B. Kröger, A. McGowan, T.
⁵¹⁸ Olszewski, P. Novack-Gottshall, M. Patzkowsky, M. Uhen, L. Villier, and P.
Wager. This work was supported by a NASA Exobiology grant (NNX10AQ446)
⁵²⁰ to A. Miller and M. Foote. I declare no conflicts of interest. This is Paleobiology
Database publication XXX.

522 References

- Alexander, R. R., 1977. Generic longevity of articulate brachiopods in relation
524 to the mode of stabilization on the substrate. *Palaeogeography, Palaeoclimatology, Palaeoecology* 21:209–226.
- Alroy, J., 2010. The Shifting Balance of Diversity Among Major Marine Animal Groups. *Science* 329:1191–1194.
- Baumiller, T. K., 1993. Survivorship analysis of Paleozoic Crinoidea: effect of filter morphology on evolutionary rates. *Paleobiology* 19:304–321.
- Cooper, W. S., 1984. Expected time to extinction and the concept of fundamental fitness. *Journal of Theoretical Biology* 107:603–629.
- Felsenstein, J., 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- Foote, M., 1988. Survivorship analysis of Cambrian and Ordovician Trilobites. *Paleobiology* 14:258–271.
- , 2000. Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology* 26:578–605.
- , 2001. Evolutionary Rates and the Age Distributions of Living and Extinct Taxa. Pp. 245–294, *in* J. B. Jackson, S. Lidgard, and F. K. McKinney, eds. *Evolutionary Patterns: Growth, Form, and Tempo in the Fossil Record*. University of Chicago Press, Chicago.
- Foote, M. and A. I. Miller, 2013. Determinants of early survival in marine animal genera. *Paleobiology* 39:171–192.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B.

- Rubin, 2013. Bayesian data analysis. 3 ed. Chapman and Hall, Boca Raton,
546 FL.
- Gelman, A. and J. Hill, 2007. Data Analysis using Regression and
548 Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.
- Harnik, P. G., C. Simpson, and J. L. Payne, 2013. Long-term differences in
550 extinction risk among the seven forms of rarity. Proceedings of the Royal
Society B: Biological Sciences 282.
- Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL
552 <http://CRAN.R-project.org/package=raster>. R package version 2.3-24.
- Hoffman, M. D. and A. Gelman, 2014. The No-U-Turn Sampler: Adaptively
554 Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine
Learning Research 15:1351–1381.
- Housworth, E. A., P. Martins, and M. Lynch, 2004. The Phylogenetic Mixed
556 Model. The American Naturalist 163:84–96.
- Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a
560 comment on "On the heritability of geographic range sizes". American
Naturalist 166:129–135.
- Jablonski, D., 1986. Background and mass extinctions: the alternation of
562 macroevolutionary regimes. Science 231:129–133.
- , 1987. Heritability at the species level: analysis of geographic ranges of
564 cretaceous mollusks. Science 238:360–363.
- , 2007. Scale and hierarchy in macroevolution. Palaeontology 50:87–109.
566 ———, 2008. Species Selection: Theory and Data. Annual Review of Ecology,
Evolution, and Systematics 39:501–524.

- Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and
570 living molluscs. *Proceedings. Biological sciences / The Royal Society*
270:401–6.
- 572 Johnson, J. G., 1974. Extinction of Perched Faunas. *Geology* 2:479–482.
- Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine
574 benthic biodiversity dynamics through Triassic–Jurassic time. *Paleobiology*
33:414–434.
- 576 Klein, J. P. and M. L. Moeschberger, 2003. *Survival Analysis: Techniques for
Censored and Truncated Data*. 2nd ed. Springer, New York.
- 578 Liow, L. H., 2004. A test of Simpson's "rule of the survival of the relatively
unspecialized" using fossil crinoids. *The American naturalist* 164:431–43.
- 580 ———, 2007. Does versatility as measured by geographic range, bathymetric
range and morphological variability contribute to taxon longevity? *Global
Ecology and Biogeography* 16:117–128.
- 582 Lynch, M., 1991. Methods for the analysis of comparative data in evolutionary
biology. *Evolution* 45:1065–1080.
- 584 Miller, A. I. and M. Foote, 2009. Epicontinental seas versus open-ocean settings:
the kinetics of mass extinction and origination. *Science* 326:1106–9.
- Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range
586 predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*
39:360–372.
- 590 ———, 2015. Interdependence of specialization and biodiversity in Phanerozoic
marine invertebrates. *Nature communications* 6:6602.

- 592 Palmer, M. E. and M. W. Feldman, 2012. Survivability is more fundamental
than evolvability. *PLoS one* 7:e38025.
- 594 Payne, J. L. and S. Finnegan, 2007. The effect of geographic range on
extinction risk during background and mass extinction. *Proceedings of the
National Academy of Sciences* 104:10506–11.
- 596 Payne, J. L., N. A. Heim, M. L. Knope, and C. R. McClain, 2014. Metabolic
598 dominance of bivalves predates brachiopod diversity decline by more than 150
million years. *Proceedings of the Royal Society B* 281:20133122.
- 600 Peters, S. E., 2008. Environmental determinants of extinction selectivity in the
fossil record. *Nature* 454:626–9.
- 602 Raup, D. M., 1975. Taxonomic survivorship curves and Van Valen's Law.
Paleobiology 1:82–96.
- 604 ———, 1978. Cohort Analysis of generic survivorship. *Paleobiology* 4:1–15.
- , 1985. Mathematical models of cladogenesis. *Paleobiology* 11:42–52.
- 606 ———, 1991. *Extinction: Bad Genes or Bad Luck?* Norton, New York.
- , 1994. The role of extinction in evolution. *Proceedings of the National
608 Academy of Sciences* 91:6758–6763.
- Rosindell, J., L. J. Harmon, and R. S. Etienne, 2015. Unifying ecology and
610 macroevolution with individual-based theory. *Ecology letters* 18:472–82.
- Sepkoski, J. J., 1975. Stratigraphic biases in the analysis of taxonomic
612 survivorship. *Paleobiology* 1:343–355.
- Sepkoski Jr., J. J., 1981. A factor analytic description of the Phanerozoic
614 marine fossil record. *Paleobiology* 7:36–53.

- Sheehan, P., 2001. The late Ordovician mass extinction. *Annual Review of Earth and Planetary Sciences* 29:331–364.
- Simpson, C., 2006. Levels of selection and large-scale morphological trends. Ph.D. thesis, University of Chicago.
- Simpson, C. and P. G. Harnik, 2009. Assessing the role of abundance in marine bivalve extinction over the post-Paleozoic. *Paleobiology* 35:631–647.
- Simpson, G. G., 1944. *Tempo and Mode in Evolution*. Columbia University Press, New York.
- , 1953. *The Major Features of Evolution*. Columbia University Press, New York.
- Stan Development Team, 2014. Stan: A c++ library for probability and sampling, version 2.5.0. URL <http://mc-stan.org/>.
- Van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.
- , 1979. Taxonomic survivorship curves. *Evolutionary Theory* 4:129–142.
- Vilhena, D. A., E. B. Harris, C. T. Bergstrom, M. E. Maliska, P. D. Ward, C. A. Sidor, C. A. E. Strömberg, and G. P. Wilson, 2013. Bivalve network reveals latitudinal selectivity gradient at the end-Cretaceous mass extinction. *Scientific Reports* 3:1790.
- Wagner, P. J. and G. F. Estabrook, 2014. Trait-based diversification shifts reflect differential extinction among fossil taxa. *Proceedings of the National Academy of Sciences* 111:16419–16424.
- Wang, S. C., 2003. On the continuity of background and mass extinction. *Paleobiology* 29:455–467.
- Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and

Widely Applicable Information Criterion in Singular Learning Theory.
640 Journal of Machine Learning Research 11:3571–3594.

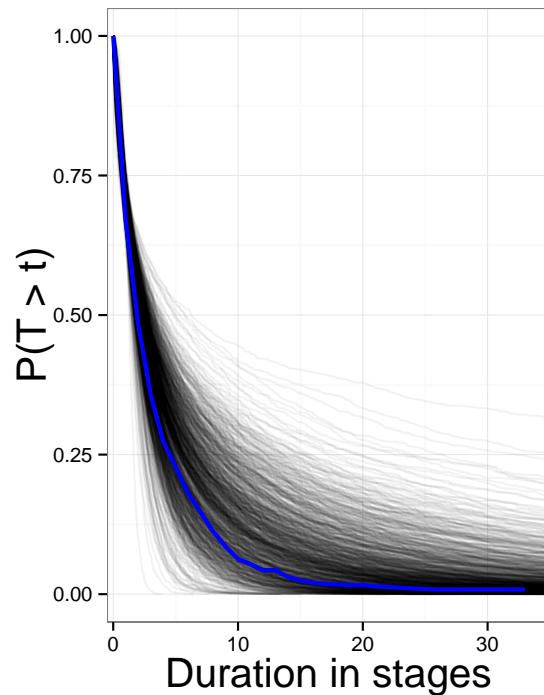


Figure 1: Comparison of empirical estimates of $S(t)$ versus estimates from 1000 posterior predictive data sets. $S(t)$ corresponds to $P(T > t)$ as it is the probability that a given genus observed at age t will continue to live. This is equivalent to the probability that t is less than the genus' ultimate duration T .

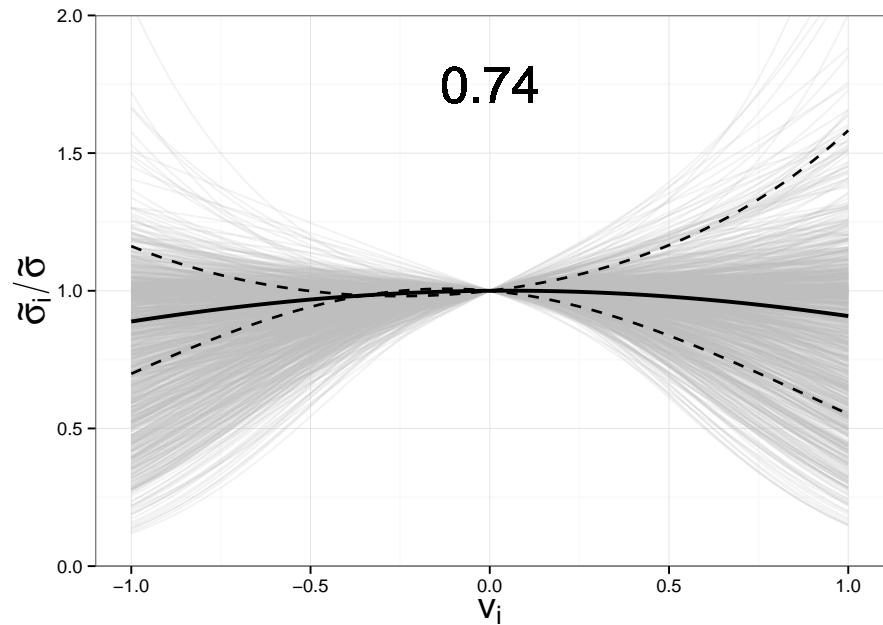


Figure 2: The overall expected relationship $f(v_i)$ between environmental affinity v_i and a multiplier of extinction risk (Eq. 3). Each grey line corresponds to a single draw from the posterior predictive distribution, the solid black line corresponds to the median of the posterior predictive distribution, and the dashed black lines correspond to the median relationship plus or minus one standard deviation. 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.

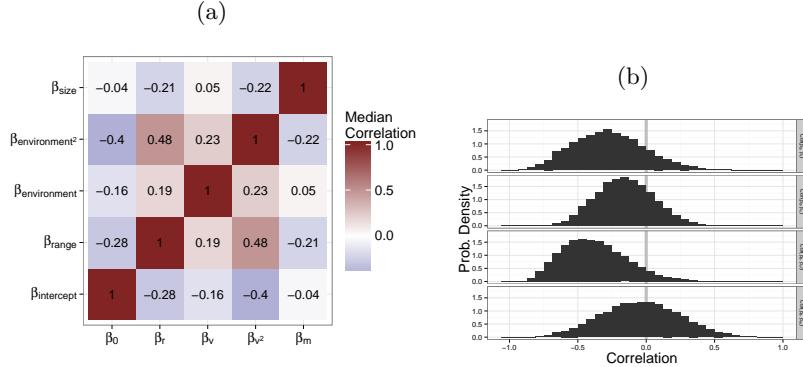


Figure 3: **A:** Heatmap for the median estimates of the terms of the correlation matrix Ω between cohort-level covariate effects. Both the exponential (left) and Weibull (right) models are presented. The off-diagonal terms are the correlation between the estimates of the cohort-level estimates of the effects of covariates, along with intercept/baseline extinction risk. **B:** 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.

parameter	mean	standard deviation
μ_i	-1.52	0.16
μ_r	-1.39	0.13
μ_v	-0.04	0.16
μ_{v^2}	0.30	0.45
μ_m	-0.07	0.08
τ_i	0.77	0.14
τ_r	0.40	0.13
τ_v	1.05	0.23
τ_{v^2}	1.87	0.64
τ_m	0.24	0.13

Table 1: Group-level estimates of the intercept terms the effects of biological traits on brachiopod generic survival from equations 1 and 2, presented as means and standard deviations. μ values are the location parameters of the effects, while τ values are the scale terms describing the variation between cohorts. The subscripts correspond to the following: i intercept, r geographic range, v environmental affinity, v^2 environmental affinity squared, m body size.

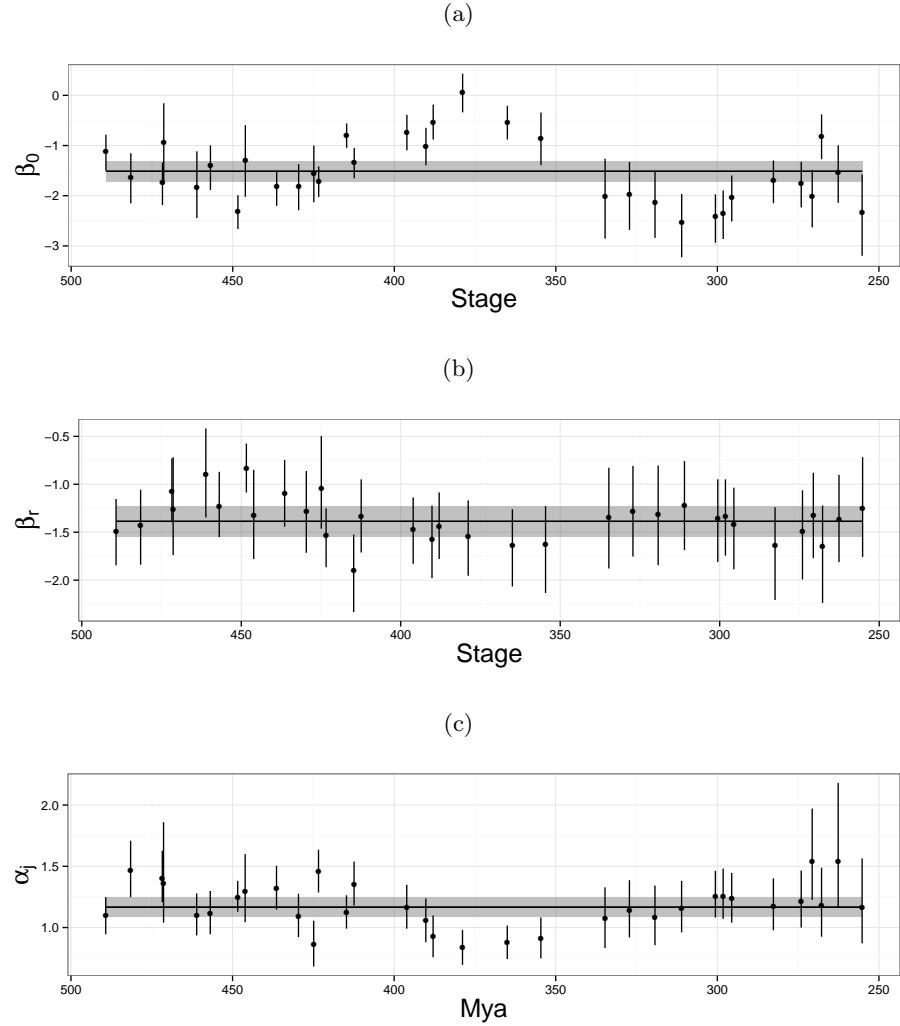


Figure 4: A) Comparison of cohort-specific estimates of β_0 presented along with the estimate for the overall baseline extinction risk. B) Comparison of cohort-specific estimates of the effect of geographic range on extinction risk β_r presented along with the estimate for the overall effect of geographic range. C) Comparison of cohort-specific estimates of the Weibull shape parameter α . Values greater than 1 correspond to accelerating extinction with age, and those below 1 to decelerating extinction with age. 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.

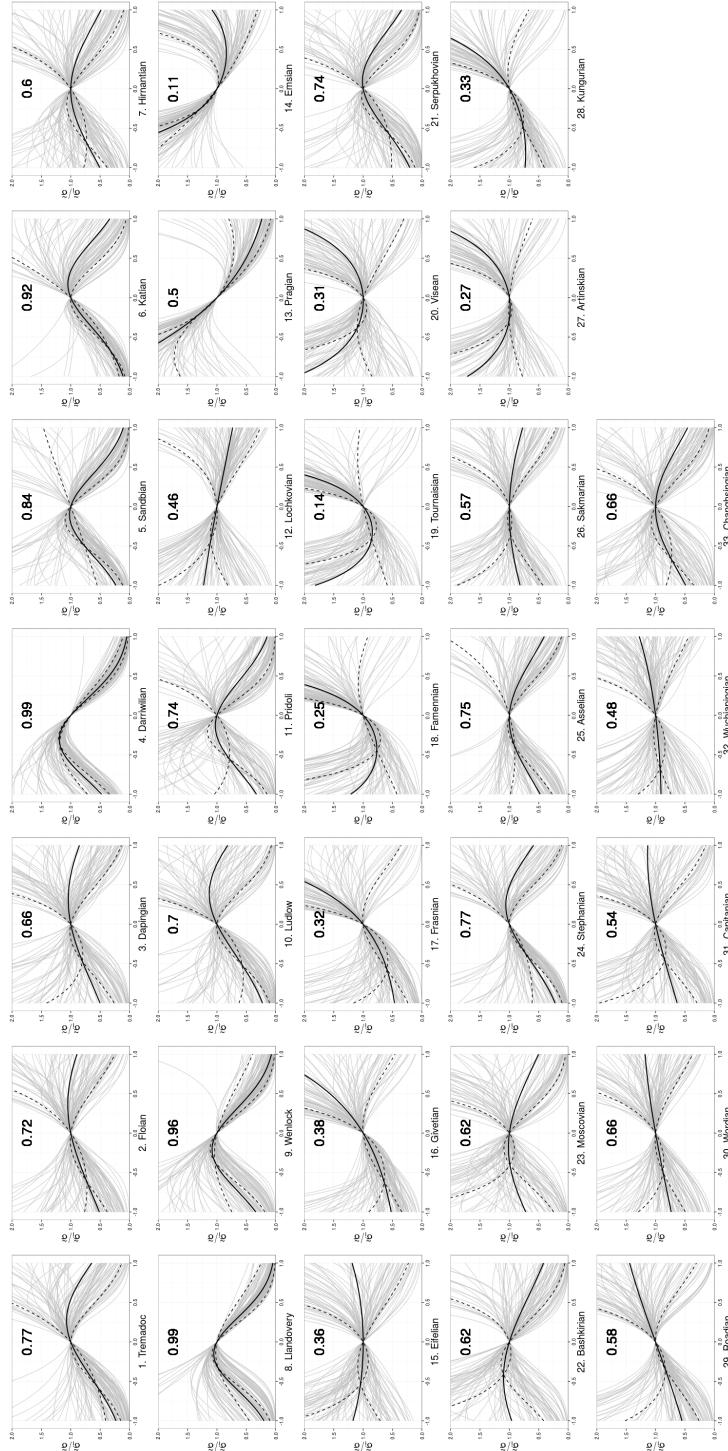


Figure 5: Comparison of the cohort-specific estimates of $f(v_i)$ (Eq. 3) for the 33 analyzed origination cohorts. The stage of origination is labeled on the x-axis of each panel. The oldest stage is in the upper left, while the youngest is in the lower left. The number in each panel corresponds to the posterior probability that $f(v_i)$ is concave down. Those that are highlighted in red have less than 51% posterior predictive probability that $f(v_i)$ is concave down. Each grey line corresponds to a single draw from the posterior predictive distribution, the solid black line corresponds to the median of the posterior predictive distribution, and the dashed black lines correspond to the median relationship plus or minus one standard deviation. Note that all estimates must pass through $y = 1$ when $x = 0$ (Eq. 3).