

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 5800.

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 approach, I also model the interaction between the effects of 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 are part of the same continuum. For parts of the Paleozoic I find support for a "survival of the generalists" scenario, though there are times where this relationship is absent or even reversed. Importantly, I find evidence that as baseline extinction risk increases, the effect of geographic range increases but the effect of environmental preference tends to decrease. Additionally, I find strong evidence for correlation between the effects of geographic range and the non-linearity aspect of environmental preference which may help explain this pattern. These results support the hypothesis that as extinction intensity increases, overall extinction selectivity decreases.

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
40 effect, these are descriptors of a taxon’s broad-sense ecology.

Jablonski (1986) phrased his conclusions in terms of background versus mass
42 extinction, but this scenario is readily transferable to a continuous variation
framework as there is no obvious distinction in terms of extinction rate between
44 these two states (Wang, 2003). Additionally, the Jablonski (1986) scenario has
strong model structure requirements in order to test its proposed
46 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.

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

54 I choose to model brachiopod taxon durations because trait based differences in
extinction risk should manifest as differences in taxon durations. Namely, a
56 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
58 aspect of “taxon fitness” along with expected lineage specific
branching/origination rate (Cooper, 1984, Palmer and Feldman, 2012).

60 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
62 on the brachiopod record from most of the Paleozoic, from the start of the
Ordovician (approximately 485 Mya) through the end Permian (approximately
64 252 Mya) as this represents the time of greatest global brachiopod diversity
(Alroy, 2010).

66 The analysis of taxon durations, or time from origination to extinction, falls
under the purview of survival analysis, a field of applied statistics commonly
68 used in health care (Klein and Moeschberger, 2003) but has a long history in
paleontology (Simpson, 1944, 1953, Van Valen, 1973, 1979).

70 Geographic range is widely considered the most important taxon trait for
estimating differences in extinction risk at nearly all times with large geographic
72 range associated with low extinction risk (Jablonski, 1986, 1987, Jablonski and
Roy, 2003, Payne and Finnegan, 2007), though Foote and Miller (2013) find
74 that this generalization does not hold in the Mesozoic. For the Paleozoic,
however, I expect this to hold true for the entire period analyzed.

76 Miller and Foote (2009) demonstrated that during several mass extinctions taxa
associated with open-ocean environments tend to have a greater extinction risk
78 than those taxa associated with epicontinental seas. During periods of
background extinction, however, they found no consistent difference between
80 taxa favoring either environment. These two environment types represent the
primary environmental dichotomy observed in ancient marine systems (Miller
82 and Foote, 2009, Peters, 2008, Sheehan, 2001).

Epicontinental seas are a shallow-marine environment where the ocean has
84 spread over the surface of a continental shelf with a depth typically less than
100m. In contrast, open-ocean coastline environments have much greater
86 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
88 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
90 Paleozoic (approximately 541–252 My), epicontinental seas were widely spread
globally but declined over the Mesozoic (approximately 252–66 My) and
92 eventually nearly disappeared during the Cenozoic (approximately 66–0 My) as

open-ocean coastlines became the dominant shallow-marine setting (Johnson,
94 Miller and Foote, 2009, Peters, 2008).

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

Body size, measured as shell length (Payne et al., 2014), was also considered as
104 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
106 hypothesis of how body size effects extinction risk in brachiopods, such that a
positive, negative, or zero effect are all plausible.

108 I adopt a hierarchical Bayesian survival modeling approach, which represents
both a conceptual and statistical unification of the paleontological dynamic and
110 cohort survival analytic approaches (Baumiller, 1993, Foote, 1988, Raup, 1975,
1978, Simpson, 2006, Van Valen, 1973, 1979). By using a Bayesian framework I
112 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
114 (i.e. biological traits) are themselves known with error.

2 Materials and Methods

116 2.1 Fossil occurrence information

The dataset analyzed here was sourced from the Paleobiology Database
118 (<http://www.paleodb.org>) which was then filtered based on taxonomic,
temporal, stratigraphic, and other occurrence information that was necessary
120 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
122 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
124 (2009), with additional occurrences assigned similarly (Miller and Foote,
personal communication).

126 Fossil occurrences were analyzed at the genus level which is common for
paleobiological, macroevolution, or macroecological studies of marine
128 invertebrates (Alroy, 2010, Foote and Miller, 2013, Harnik et al., 2013, Kiessling
and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015,
130 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
132 of higher taxa, the nature of the fossil record makes accurate and precise
taxonomic assignments at the species level for all occurrences extremely difficult
134 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,
136 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.

138 Genus duration was calculated as the number of geologic stages from first
appearance to last appearance, inclusive. Durations were based on geologic

¹⁴⁰ 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
¹⁴² fossils are known from a geological interval which represents some temporal
range. Stages act as effectively irreducible globally consistent temporal intervals
¹⁴⁴ in which taxa occur.

¹⁴⁶ Genera with a last occurrence in or after Changhsingian stage were right
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
¹⁴⁸ geographic range size (r), environmental preference (v, v^2), and body size (m).

¹⁵⁰ 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
¹⁵² grid placed on the map. Each grid cell represents approximately $250,000 \text{ km}^2$.
The map projection and regular lattice were made using shape files from
¹⁵⁴ <http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans,
2015).

¹⁵⁶ 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
¹⁵⁸ by that genus was calculated. Dividing the genus occupancy by the total
occupancy gives the relative occupancy of that genus. Mean relative genus
¹⁶⁰ occupancy was then calculated as the mean of the per stage relative occupancies
of that genus.

¹⁶² Body size data was sourced directly from Payne et al. (2014). Because those
measurements are presented without error, a measurement error model similar
¹⁶⁴ to the one for environmental affinity could not be implemented (Appendix A).

Prior to analysis, geographic range and body size were transformed and

¹⁶⁶ standardized in order to improve interpretability of the results. Geographic range, which can only vary between 0 and 1, was logit transformed. Body size, ¹⁶⁸ 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 ¹⁷⁰ their standard deviation following Gelman and Hill (2007).

2.2 Analytical approach

¹⁷² Hierarchical modelling is a statistical approach which explicitly takes into account the structure of the observed data in order to model both the within ¹⁷⁴ 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 ¹⁷⁶ cohort). These groups are considered separate 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 ¹⁷⁸ interest (e.g. covariate effects) (Gelman et al., 2013). The subsequent estimates ¹⁸⁰ 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 ¹⁸² pulled towards the overall group mean.

This partial pooling is one of the greatest advantages of hierarchical modeling. ¹⁸⁴ By letting the groups “support” each other, parameter estimates then better reflect our statistical uncertainty. Additionally, this partial pooling helps control ¹⁸⁶ for multiple comparisons and possibly spurious results as effects with little support are drawn towards the overall group mean (Gelman et al., 2013, ¹⁸⁸ Gelman and Hill, 2007).

All covariate effects (regression coefficients), as well as the intercept term ¹⁹⁰ (baseline extinction risk), were allowed to vary by group (origination cohort).

The covariance/correlation between covariate effects was also modeled. This
192 hierarchical structure allows inference for how covariates effects may change
with respect to each other while simultaneously estimating the effects
194 themselves, propagating our uncertainty through all estimates.

Additionally, instead of relying on point estimates of environmental affinity, I
196 treat environmental affinity as a continuous measure of the difference between
the taxon's environmental occurrence pattern and the background occurrence
198 pattern (Appendix A).

2.3 Survival model

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

208 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
210 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
212 the between-cohort variation in regression coefficient estimates, Ω is the
correlation matrix of the regression coefficients, and α_j is the shape parameter
214 for cohort j with a is the overall mean shape parameter and π is the variance
between estimates of α_j .

²¹⁶ 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

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

²¹⁸ 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 ²²⁰ $1/\lambda$.

²²² For an explanation of how this model was developed, parameter explanations, 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 ²²⁴ censored observations are included. For an explanation of both of these aspects, see Appendices A and C.

²²⁶ **2.4 Parameter estimation**

The joint posterior was approximated using a Markov chain Monte Carlo
²²⁸ routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn
Sampler (Hoffman and Gelman, 2014) as implemented in the probabilistic
²³⁰ programming language Stan (Stan Development Team, 2014). The posterior
distribution was approximated from four parallel chains run for 10,000 draws
²³² 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
²³⁴ 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
²³⁶ samples are well mixed (Gelman et al., 2013).

2.5 Model evaluation

²³⁸ Models were evaluated using both posterior predictive checks and an estimate of
out-of-sample predictive accuracy. The motivation behind posterior predictive
²⁴⁰ 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.,
²⁴² 2013). Systematic differences between the simulations and observations indicate
weaknesses of the model fit. An example of a technique that is very similar
²⁴⁴ would be inspecting the residuals from a linear regression.

The strategy behind posterior predictive checks is to draw simulated values
²⁴⁶ 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
²⁴⁸ 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
²⁵⁰ 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
252 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.

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

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

3 Results

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

Comparisons of the survival functions estimated from 1000 posterior predictive
276 data sets to the estimated survival function of the observed genera demonstrates
that Weibull models approximately capture the observed pattern of extinction
278 (Fig. 1). This is partially expected given that the unit of analysis is genus species
duration as opposed species level which can alter the shape of $S(t)$ (Foote, 2001,
280 Raup, 1975, 1978, 1985). Additionally, the Weibull model is expected to have
slightly better out-of-sample predictive accuracy than the exponential model
282 (WAIC 4534 versus 4604). Because the difference in WAIC between these two
models is large, only those results from the Weibull model will be discussed.

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

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

Interpretation of the effect of environmental preference v on duration is slightly
294 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
296 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
298 preference on extinction risk.

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

³⁰⁰ 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
³⁰² 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
³⁰⁴ preference have greater durations than either extreme, and *vice versa* for
concave up function.

³⁰⁶ The expected effect of environmental preference as a multiplier of expected
extinction risk can then be visualized (Fig. 2). This figure depicts 1000 posterior
³⁰⁸ predictive estimates of Eq. 3 across all possible values of v . The number
indicates the posterior probability that the function is concave down, with
³¹⁰ generalists having lower extinction risk/greater duration than either type of
specialist. Note that the inflection point of Figure 2 is approximately $x = 0$,
³¹² something that is expected given the estimate of μ_v (Table 1).

Coefficients with greater values of τ have greater between-cohort variation
³¹⁴ (Table 1). The covariate effects with the greatest between origination cohort
variation are β_r , β_v , and β_{v^2} . Estimates of β_m have negligible between cohort
³¹⁶ variation, as there is less between cohort variation than the between cohort
variation in baseline extinction risk β_0 .

³¹⁸ The correlations between the intercept term β_0 and the effects of the taxon
traits are of particular interest for evaluating the Jablonski (1986) scenario (Fig.
³²⁰ 3a first column/last row). The correlation terms of Ω (Fig. 3a) describe the
relationship between the coefficients and how their estimates may vary together
³²² across cohorts. 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.

³²⁴ Marginal posterior probabilities of the correlations between the level of baseline extinction risk β_0 and the effects of the taxon traits indicate that the correlation
³²⁶ between expected extinction risk and both geographic range β_r and β_{v^2} are of particular note (Fig. 3b).

³²⁸ There is an approximately 85% probability that β_0 and β_r are negatively correlated (Fig. 3b), meaning that as extinction risk increases, the
³³⁰ effect/importance of geographic range on genus duration may increase. There is a 93% probability that β_0 and β_{v^2} are negatively correlated (Fig. 3b), meaning
³³² that as extinction risk decreases, the peakedness of $f(v_i)$ may increase and the relationship tends towards concave down. Additionally, there is a 94%
³³⁴ probability that values of β_r and β_{v^2} are positively correlated (Mean correlation 0.45, standard deviation 0.26).

³³⁶ While the overall group level estimates are of particular importance when defining time-invariant differences in extinction risk, it is also important and
³³⁸ useful to analyze the individual level parameter estimates in order to better understand how parameters actually vary across cohorts.

³⁴⁰ In comparison to the overall mean extinction risk μ_0 , cohort level estimates β_0 show some amount of variation through time as expected by estimates of τ_0
³⁴² (Fig. 4). A similar, if slightly greater, amount of variation is also observable in cohort estimates of the effect of geographic range β_r (Fig. 4). Again, smaller
³⁴⁴ 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
³⁴⁸ by using the cohort specific coefficients estimates. Following the same procedure used earlier (Eq. 3), but substituting cohort specific estimates of β_v and β_{v^2} for

³⁵⁰ 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
³⁵² model, though the observed pattern should be similar for the exponential model.

³⁵⁴ 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
functions (Fig. 5). 12 of the 33 cohorts have less than 50% posterior probability
³⁵⁶ that generalists are potentially expected to be shorter lived than specialists,
though two of those cases have approximately a 50% probability of being either
³⁵⁸ concave up or down. This is congruent with the 0.74 posterior probability that
 μ_{v^2} is positive/ $f(v_i)$ is concave down.

³⁶⁰ Additionally, for some cohorts there is a quite striking pattern where the effect
of environmental preference v has a nearly-linear relationship (Fig. 5). These are
³⁶² scenarios where one of the end member preferences is expected to have a greater
duration than either intermediate or the opposite end member preference.

³⁶⁴ 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
³⁶⁶ between the Givetian through the Viséan, inclusive, intermediate preferences are
of intermediate extinction risk when compared with epicontinental specialists
³⁶⁸ (lowest risk) or open-ocean specialists (highest risk). This time period represents
most of the Devonian through the early Carboniferous.

³⁷⁰ 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
³⁷² (Fig. 4). Also, for the majority of the time analyzed there is evidence for
accelerating extinction with taxon age. If this is due to biological causes
³⁷⁴ (Rosindell et al., 2015, Wagner and Estabrook, 2014) or an artifact of
preservation/the minimum observable duration of a genus (Sepkoski, 1975) is
³⁷⁶ 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
378 taxon age as opposed to accelerating (Foote, 2001, Raup, 1975, 1978, 1985).
This uncertainty remains an open issue for future analysis.

380 4 Discussion

My results demonstrate that both the effects of geographic range and the
382 peakedness/concavity of environmental preference are both negatively
correlated with baseline extinction risk, meaning that as baseline extinction risk
384 increases the effect size of geographic range increases but the curvature of the
effect of environmental preference decreases (Fig. 3b).

386 However, the interaction between geographic range and origination cohort,
while weak, indicates that neither of the two proposed end-member
388 macroevolutionary mechanisms is strictly right. The evidence for a correlation
between changes in the effect of geographic range and the curvature of the effect
390 of environmental preference may help explain this result. I hypothesize that
because taxa with large geographic ranges encompass more possible
392 environments, when extinction intensity is high there is little coherent difference
in the environmental preference among the surviving taxa. The intensity
394 decreases the selectivity such that the effect of environmental preference is
effectively washed out by the strength of the effect of geographic range; this is
396 consistent with similar hypotheses presented in Raup (1991).

For the approximately 233 My period analyzed there is an approximate 74%
398 posterior probability that brachiopod genera with intermediate environmental
preferences are expected to have a lower extinction risk than either end
400 members. However, the overall curvature of $f(v_i)$ is not very peaked meaning
that when averaged over the entire Phanerozoic this relationship may not lead

⁴⁰² 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.
⁴⁰⁴ time since the extinction of the non-avian dinosaurs). This result gives very weak support for the universality of the hypothesis that environmental
⁴⁰⁶ generalists have greater survival than environmental specialists (Liow, 2004, 2007, Nürnberg and Aberhan, 2013, 2015, Simpson, 1944).

⁴⁰⁸ 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
⁴¹⁰ only a 74% 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
⁴¹² 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
⁴¹⁴ expectation of a concave down nonlinear relation are instead is highly skewed and effectively linear (Fig. 5). These results demonstrate that, while the
⁴¹⁶ group-level estimate may only weakly support one hypothesis, the cohort-level estimates may exhibit very different characteristics. These results are also
⁴¹⁸ consistent with aspects of Miller and Foote (2009) who found that the effect of environmental preference on extinction risk was quite variable and without
⁴²⁰ obvious patterning during times of background extinction.

⁴²² There are two mass extinction events that are captured within the time frame considered here: the Ordovician-Silurian and the Frasnian-Famennian. The cohorts bracketing these events are worth considering in more detail.

⁴²⁴ 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
⁴²⁶ (Johnson, 1974, Sheehan, 2001). My results are broadly consistent with this scenario with both epicontinental and open-ocean specialists having a much
⁴²⁸ lower expected duration than intermediate taxa (Fig. 5). All of the stages

between the Darriwillian and the Llandovery, except the Hirnantian, have a
430 high probability (90+%) that $f(v)$ is concave down. The pattern for the
Darriwillian, which proceeds the supposed start of Ordovician glacial activity,
432 demonstrates that taxa tend to favor open-ocean environments are expected to
have a greater duration than either intermediate or epicontinental specialists, in
434 decreasing order.

For nearly the entire Devonian estimates of $f(v)$ indicate that one of the
436 environmental end members is favored over the other end member of
intermediate preference (Fig. 5). For almost the entirety the Givetian though
438 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
440 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 result,
442 however, seems to reflect the intensity of the seemingly nearly-linear difference
in expected duration across the range of v as opposed to an inversion of the
444 weakly expected curvilinear pattern.

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

⁴⁵⁶ actually affect extinction risk.

The model used here could be improved through either increasing the number of
⁴⁵⁸ analyzed taxon traits, expanding the hierarchical structure of the model to
include other major taxonomic groups of interest, and the inclusion of explicit
⁴⁶⁰ phylogenetic relationships between the taxa in the model as an additional
hierarchical effect.

⁴⁶² 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
⁴⁶⁴ been found to be related to brachiopod survival (Alexander, 1977) so its
inclusion may be of particular interest.

⁴⁶⁶ It is theoretically possible to expand this model to allow for comparisons within
and between major taxonomic groups. This approach would better constrain the
⁴⁶⁸ brachiopod estimates while also allowing for estimation of similarities and
differences in cross-taxonomic patterns. The major issue surrounding this
⁴⁷⁰ particular expansion involves finding a similarly well sampled taxonomic group
that is present during the Paleozoic. Example groups include Crinoidea,
⁴⁷² Ostracoda, and other members of the “Paleozoic fauna” (Sepkoski Jr., 1981).

Taxon traits like environmental preference or geographic range (Hunt et al.,
⁴⁷⁴ 2005, Jablonski, 1987) are most likely heritable, at least phylogenetically
(Housworth et al., 2004, Lynch, 1991). Without phylogenetic context, this
⁴⁷⁶ analysis assumes that differences in extinction risk between taxa are
independent of those taxa’s shared evolutionary history (Felsenstein, 1985). In
⁴⁷⁸ contrast, the origination cohorts only capture shared temporal context. The
inclusion of phylogenetic context as an additional individual-level hierarchical
⁴⁸⁰ effect independent of origination cohort would allow for determining how much
of the observed variability is due to shared evolutionary history versus actual

482 differences associated with these taxonomic traits.

In summary, patterns of Paleozoic brachiopod survival were analyzed using a
484 fully Bayesian hierarchical survival modelling approach while also eschewing the
traditional separation between background and mass extinction. I modeled both
486 the overall mean effect of biological covariates on extinction risk while also
modeling the correlation between cohort-specific estimates of covariate effects. I
488 find that as baseline extinction risk increases, the form of the selectivity of
extinction changes such that during periods of low extinction the effect
490 environmental preference has a more nonlinear relationship than during times of
high extinction intensity. In particular, the correlation between the effect of
492 geographic range and the curvature of the effect of environmental preference on
taxon survival supports the idea that during periods of high extinction intensity
494 the size of effect of geographic range effectively washes out the effects of other
biological traits (Jablonski, 1987, Raup, 1991). Finally, I find very weak for
496 “survival of the unspecialized” (Liow, 2004, 2007, Nürnberg and Aberhan, 2013,
2015, Simpson, 1944) as a general characterization of the effect of environmental
498 preference on extinction risk (Fig. 2), though there is heterogeneity between
origination cohorts with most periods of time conforming to this hypothesis
500 (Fig. 5).

Acknowledgements

502 I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for
helpful and Marities Marites Villarosa Garcia for draft comments. Additionally,
504 thank you A. Miller for the epicontinental versus open-ocean assignments. This
entire study would have not been possible without the Herculean effort of
506 the many contributors to the Paleobiology Database. In particular, I would like

to thank J. Alroy, M. Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N. Heim,
508 A. Hendy, S. Holland, L. Ivany, W. Kiessling, B. Kröger, A. McGowan, T.
Olszewski, P. Novack-Gottshall, M. Patzkowsky, M. Uhen, L. Villier, and P.
510 Wager. This work was supported by a NASA Exobiology grant (NNX10AQ446)
to A. Miller and M. Foote. I declare no conflicts of interest. This is Paleobiology
512 Database publication XXX.

References

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

- 536 Rubin, 2013. Bayesian data analysis. 3 ed. Chapman and Hall, Boca Raton,
FL.
- 538 Gelman, A. and J. Hill, 2007. Data Analysis using Regression and
Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.
- 540 Harnik, P. G., C. Simpson, and J. L. Payne, 2013. Long-term differences in
extinction risk among the seven forms of rarity. Proceedings of the Royal
542 Society B: Biological Sciences 282.
- Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL
544 <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
546 Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine
Learning Research 15:1351–1381.
- 548 Housworth, E. A., P. Martins, and M. Lynch, 2004. The Phylogenetic Mixed
Model. The American Naturalist 163:84–96.
- 550 Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a
comment on "On the heritability of geographic range sizes". American
552 Naturalist 166:129–135.
- Jablonski, D., 1986. Background and mass extinctions: the alternation of
554 macroevolutionary regimes. Science 231:129–133.
- , 1987. Heritability at the species level: analysis of geographic ranges of
556 cretaceous mollusks. Science 238:360–363.
- , 2007. Scale and hierarchy in macroevolution. Palaeontology 50:87–109.
- 558 ———, 2008. Species Selection: Theory and Data. Annual Review of Ecology,
Evolution, and Systematics 39:501–524.

- 560 Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and
living molluscs. *Proceedings. Biological sciences / The Royal Society*
562 270:401–6.
- Johnson, J. G., 1974. Extinction of Perched Faunas. *Geology* 2:479–482.
- 564 Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine
benthic biodiversity dynamics through Triassic–Jurassic time. *Paleobiology*
566 33:414–434.
- Klein, J. P. and M. L. Moeschberger, 2003. *Survival Analysis: Techniques for
568 Censored and Truncated Data*. 2nd ed. Springer, New York.
- Liow, L. H., 2004. A test of Simpson's "rule of the survival of the relatively
570 unspecialized" using fossil crinoids. *The American Naturalist* 164:431–43.
- , 2007. Does versatility as measured by geographic range, bathymetric
572 range and morphological variability contribute to taxon longevity? *Global
Ecology and Biogeography* 16:117–128.
- 574 Lynch, M., 1991. Methods for the analysis of comparative data in evolutionary
biology. *Evolution* 45:1065–1080.
- 576 Miller, A. I. and M. Foote, 2009. Epicontinental seas versus open-ocean settings:
the kinetics of mass extinction and origination. *Science* 326:1106–9.
- 578 Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range
predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*
580 39:360–372.
- , 2015. Interdependence of specialization and biodiversity in Phanerozoic
582 marine invertebrates. *Nature communications* 6:6602.

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

- 606 Sheehan, P., 2001. The late Ordovician mass extinction. *Annual Review of Earth and Planetary Sciences* 29:331–364.
- 608 Simpson, C., 2006. Levels of selection and large-scale morphological trends. Ph.D. thesis, University of Chicago.
- 610 Simpson, C. and P. G. Harnik, 2009. Assessing the role of abundance in marine bivalve extinction over the post-Paleozoic. *Paleobiology* 35:631–647.
- 612 Simpson, G. G., 1944. *Tempo and Mode in Evolution*. Columbia University Press, New York.
- 614 ———, 1953. *The Major Features of Evolution*. Columbia University Press, New York.
- 616 Stan Development Team, 2014. Stan: A c++ library for probability and sampling, version 2.5.0. URL <http://mc-stan.org/>.
- 618 Van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.
- , 1979. Taxonomic survivorship curves. *Evolutionary Theory* 4:129–142.
- 620 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
- 622 latitudinal selectivity gradient at the end-Cretaceous mass extinction. *Scientific Reports* 3:1790.
- 624 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.
- 626 Wang, S. C., 2003. On the continuity of background and mass extinction. *Paleobiology* 29:455–467.
- 628 Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and

- ⁶³⁰ Widely Applicable Information Criterion in Singular Learning Theory.
Journal of Machine Learning Research 11:3571–3594.

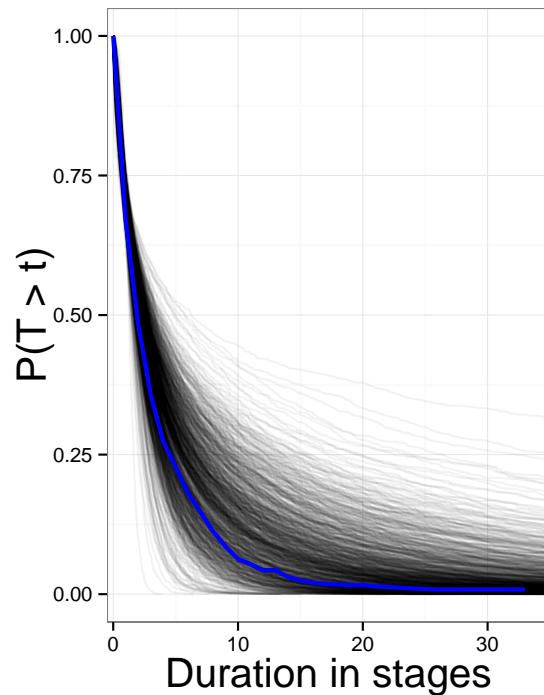


Figure 1: Comparison of empirical estimates of $S(t)$ versus estimates from 1000 posterior predictive data sets. $S(t)$ corresponds to $P(T > t)$ as it is the probability that a given genus observed at age t will continue to live. This is equivalent to the probability that t is less than the genus' ultimate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

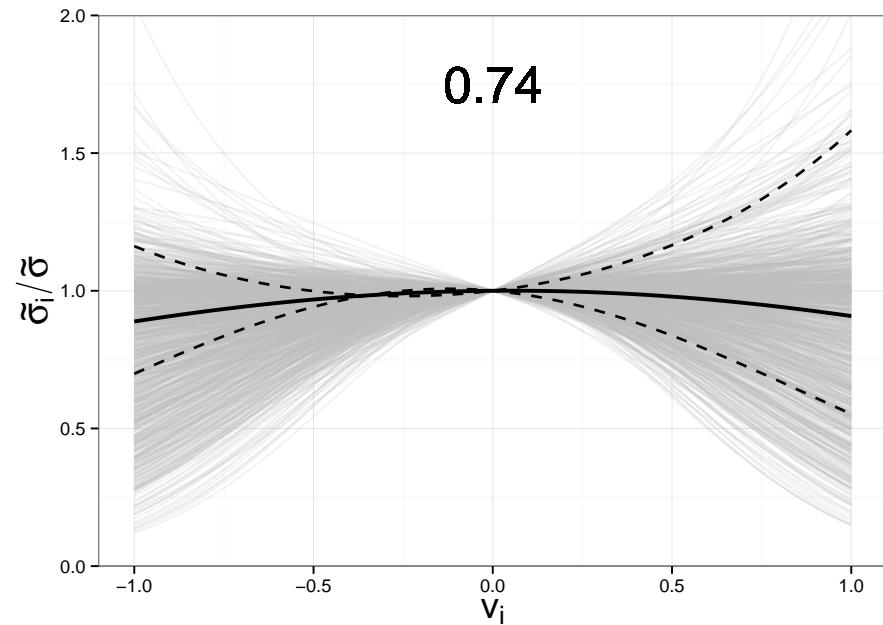


Figure 2: The overall expected relationship $f(v_i)$ between environmental affinity v_i and a multiplier of extinction risk (Eq. 3). Each grey line corresponds to a single draw from the posterior predictive distribution, the solid black line corresponds to the median of the posterior predictive distribution,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

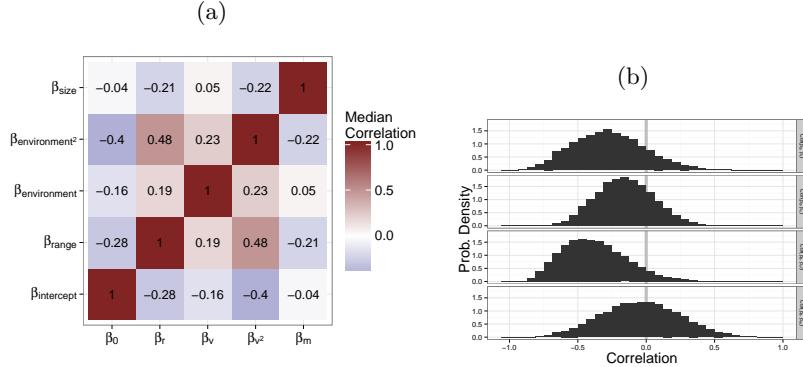


Figure 3: **A:** Heatmap for the median estimates of the terms of the correlation matrix Ω between cohort-level covariate effects. Both the exponential (left) and Weibull (right) models are presented. The off-diagonal terms are the correlation between the estimates of the cohort-level estimates of the 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.

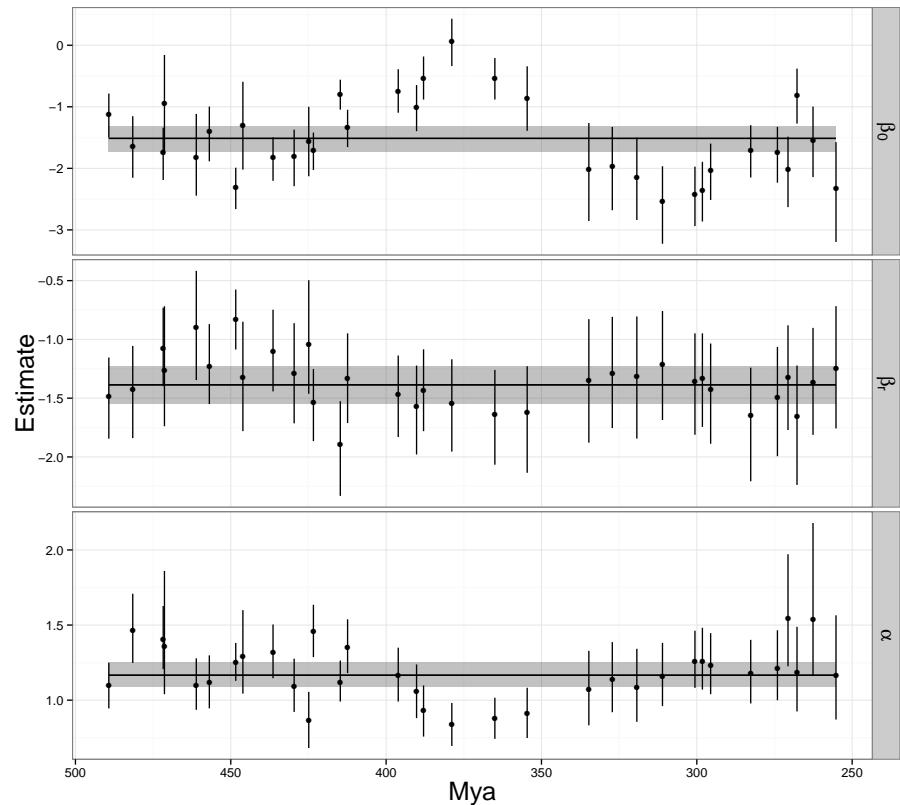


Figure 4: Comparison of cohort-specific estimates of β_0 (first row), cohort-specific estimates of the effect of geographic range on extinction risk β_r (second row), and cohort-specific estimates of the Weibull shape parameter α where values greater than 1 correspond to accelerating extinction with 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.

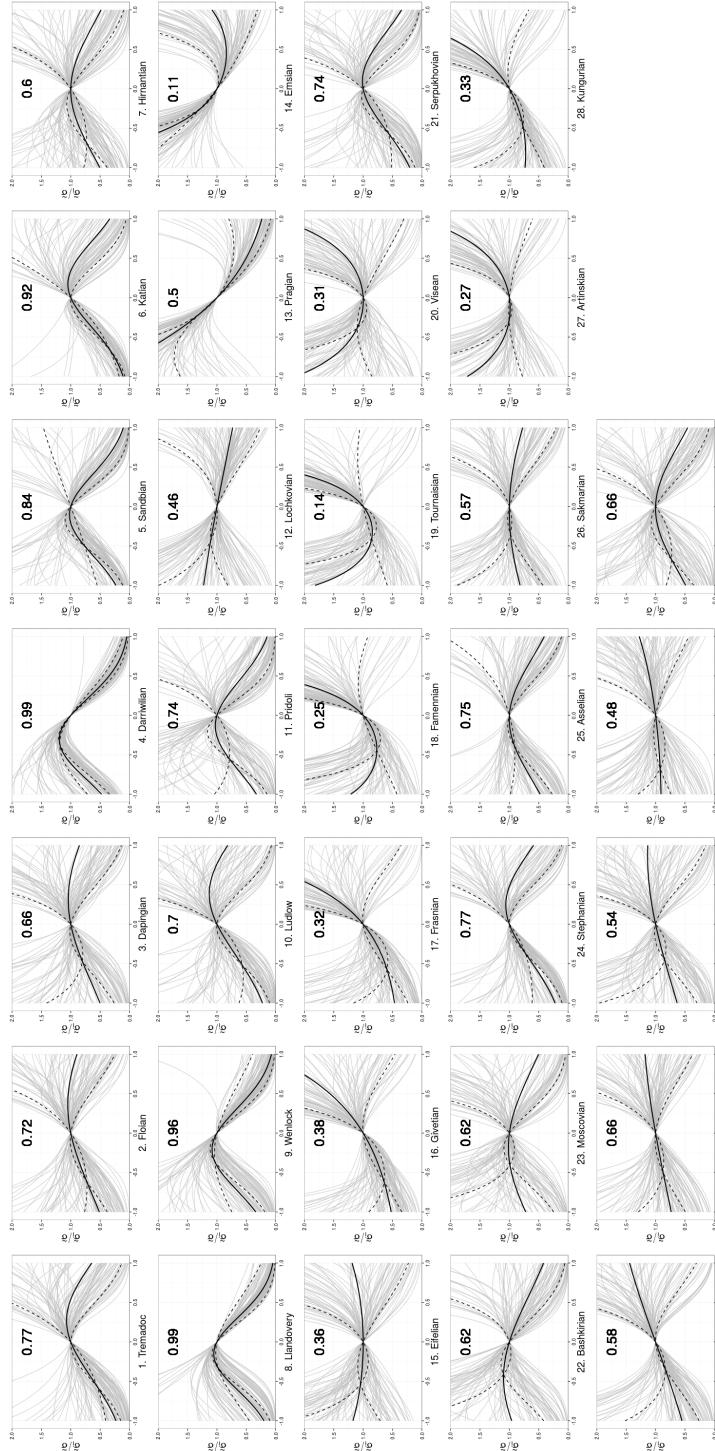


Figure 5: Comparison of the cohort-specific estimates of $f(v_i)$ (Eq. 3) for the 33 analyzed origination cohorts. The stage of origination is labeled on the x-axis of each panel. The oldest stage is in the upper left, while the youngest is in the lower left. The number in each panel corresponds to the 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).