

Title: Temporal variation and correlation of effects of biological
2 traits on brachiopod taxonomic survival

Running title: Variation in effect of traits 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

Word count: ?

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.

12

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. Using a hierarchical Bayesian modeling approach, I also model the possible interaction between the effects of the biological traits and a taxon's time of origination. 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. 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. I find evidence that as extinction risk increases, the expected strength of the selection gradient on biological traits (except body size) increases. This manifests as greater expected differences in extinction risk for each unit change in geographic range and environmental preference during periods of high extinction risk, as opposed to a much flatter expected selection gradient during periods of low extinction risk. I find weak evidence for a universally nonlinear relationship between environmental preference and extinction risk such that "generalists" have a lower expected extinction risk than either "specialists". While for the many parts of the Paleozoic this hypothesis is supported, there are many times where this hypothesized relationship is absent or even reversed. Importantly, I find that as extinction risk increases, the steepness of this relationship is expected to increases as well.

1 Introduction

38 How do biological traits affect extinction risk? Jablonski (1986) observed that
during periods of high extinction risk, the effects of biological traits on survival
40 decreased in size. However, this pattern was weakest/absent in the effect of
geographic range on survival (Jablonski, 1986). Biological traits are defined here
42 as descriptors of a taxon's adaptive zone, which is the set of biotic–biotic and
biotic–abiotic interactions that a taxon can experience (Simpson, 1944). In
44 effect, these are descriptors of a taxon's broad-sense ecology.

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

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

I model brachiopod taxon durations because trait based differences in extinction
60 risk should manifest as differences in taxon durations. Namely, a taxon with a
beneficial trait should survive longer, on average, than a taxon without that
62 beneficial trait. Conceptually, taxon survival can be considered an aspect of

“taxon fitness” along with expected lineage specific branching/origination rate
64 (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
66 records (Foote, 2000). Specifically, I focus on the brachiopod record from most
of the Paleozoic, specifically from the start of the Ordovician (approximately
68 485 Mya) through the end Permian (approximately 252 Mya) as this represents
the time of greatest global brachiopod diversity (Alroy, 2010).

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

74 Geographic range is widely considered the most important taxon trait for
estimating differences in extinction risk at nearly all times with large geographic
76 range associated with low extinction risk (Jablonski, 1986, 1987, Jablonski and
Roy, 2003, Payne and Finnegan, 2007). I expect this to hold true nearly always.

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

Epicontinental seas are a shallow-marine environment where the ocean has
86 spread over the surface of a continental shelf with a depth typically less than
100m. In contrast, open-ocean coastline environments have much greater
88 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
90 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
92 Paleozoic (approximately 541–252 My), epicontinental seas were widely spread
globally but declined over the Mesozoic (approximately 252–66 My) and
94 eventually diminished disappearing during the Cenozoic (approximately 66–0
My) as open-ocean coastlines became the dominant shallow-marine setting
96 (Johnson, 1974, Miller and Foote, 2009, Peters, 2008).

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

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

110 I adopt a hierarchical Bayesian survival modeling approach, which represents a
conceptual and statistical unification of the paleontological dynamic and cohort
112 survival analytic approaches (Baumiller, 1993, Foote, 1988, Raup, 1975, 1978,
Simpson, 2006, Van Valen, 1973, 1979). By using a Bayesian framework I am
114 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

¹¹⁶ (i.e. biological traits) are themselves known with error.

2 Materials and Methods

¹¹⁸ 2.1 Fossil occurrence information

The dataset analyzed here was sourced from the Paleobiology Database
¹²⁰ (<http://www.paleodb.org>) which was then filtered based on taxonomic,
temporal, stratigraphic, and other occurrence information that was necessary
¹²² 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
¹²⁴ 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
¹²⁶ (2009), with additional occurrences assigned similarly (Miller and Foote,
personal communication).

¹²⁸ Fossil occurrences were analyzed at the genus level which is common for
paleobiological, macroevolution, or macroecological studies of marine
¹³⁰ invertebrates (Alroy, 2010, Foote and Miller, 2013, Harnik et al., 2013, Kiessling
and Aberhan, 2007, Miller and Foote, 2009, Nürnberg and Aberhan, 2013, 2015,
¹³² Payne and Finnegan, 2007, Simpson and Harnik, 2009, Vilhena et al., 2013).

While species diversity dynamics are of much greater interest than those of
¹³⁴ higher taxa, the nature of the fossil record makes accurate and precise
taxonomic assignments at the species level for all occurrences. In particular, the
¹³⁶ simplicity of brachiopod external morphology and the quality of preservation
makes it very difficult for assignments below the genus level. As such, the choice
¹³⁸ to analyze genera as opposed to species was in order to assure a minimum level
of confidence and accuracy in the data analyzed here.

¹⁴⁰ Sampled occurrences were restricted to those with paleolatitude and
¹⁴¹ paleolongitude coordinates, assignment to either epicontinental or open-ocean
¹⁴² environment, and belonging to a genus present in the body size dataset (Payne
¹⁴³ et al., 2014). Genus duration was calculated as the number of geologic stages
¹⁴⁴ from first appearance to last appearance, inclusive. Durations were based on
¹⁴⁵ geologic stages as opposed to millions of years because of the inherently discrete
¹⁴⁶ nature of the fossil record; dates are not assigned to fossils themselves but
¹⁴⁷ instead fossils are known from a geological interval which represents some
¹⁴⁸ temporal range. Stages are effectively irreducible temporal intervals in which
taxa may occur.

¹⁵⁰ 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
¹⁵² left censored (Appendix A). The covariates used to model genus duration were
geographic range size (r), environmental preference (v, v^2), and body size (m).

¹⁵⁴ Geographic range was calculated using an occupancy approach. First, all
occurrences were projected onto an equal-area cylindrical map projection. Each
¹⁵⁶ occurrence was then assigned to one of the cells from a 70×34 regular raster
grid placed on the map. Each grid cell represents approximately $250,000 \text{ km}^2$.
¹⁵⁸ The map projection and regular lattice were made using shape files from
<http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans,
¹⁶⁰ 2015).

For each stage, the total number of occupied grid cells, or cells in which a fossil
¹⁶² occurs, was calculated. Then, for each genus, the number of grid cells occupied
by that genus was calculated. Dividing the genus occupancy by the total
¹⁶⁴ occupancy gives the relative occupancy of that genus. Mean relative genus
occupancy was then calculated as the mean of the per stage relative occupancies
¹⁶⁶ of that genus.

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

170 Prior to analysis, geographic range and body size were transformed and
standardized in order to improve interpretability of the results. Geographic
172 range, which can only vary between 0 and 1, was logit transformed. Body size,
which is defined for all positive real values, was natural log transformed. These
174 covariates were then standardized by mean centering and dividing by two times
their standard deviation following Gelman and Hill (2007).

176 **2.2 Analytical approach**

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

This partial pooling is one of the greatest advantages of hierarchical modeling.
190 By letting the groups “support” each other, parameter estimates then better
reflect our statistical uncertainty. Additionally, this partial pooling helps control

¹⁹² for multiple comparisons and possibly spurious results as effects with little
¹⁹³ support are drawn towards the overall group mean (Gelman et al., 2013,
¹⁹⁴ Gelman and Hill, 2007).

All covariate effects (regression coefficients), as well as the intercept term
¹⁹⁶ (baseline extinction risk), were allowed to vary by group (origination cohort).
The covariance/correlation between covariate effects was also modeled. This
¹⁹⁸ hierarchical structure allows inference for how covariates effects may change
with respect to each other while simultaneously estimating the effects
²⁰⁰ themselves, propagating our uncertainty through all estimates.

Additionally, instead of relying on point estimates of environmental affinity, I
²⁰² treat environmental affinity as a continuous measure of the difference between
the taxon's environmental occurrence pattern and the background occurrence
²⁰⁴ pattern (Appendix A).

2.3 Survival model

²⁰⁶ Genus durations were modeled as time-till-event data (Klein and Moeschberger,
²⁰⁷ 2003), with covariate information used in estimates of extinction risk as a
²⁰⁸ hierarchical regression model. Genus durations were assumed to follow either an
exponential or Weibull distribution. THe use of either of these distributions
²¹⁰ makes assumptions about how duration may effect extinction risk (Klein and
Moeschberger, 2003). The exponential distribution assumes that extinction risk
²¹² is independent of duration. In contrast, the Weibull distribution allows for age
dependent extinction via the shape parameter α , though only as a monotonic
²¹⁴ function of duration. Importantly, the Weibull distribution is equivalent to the
exponential distribution when $\alpha = 1$.

²¹⁶ The following variables are 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
 218 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
 220 the between-cohort variation in regression coefficient estimates, and Ω is the
 correlation matrix of the regression coefficients.

222 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

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

²²⁴ 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 ²²⁶ $1/\lambda$.

²²⁸ 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 ²³⁰ censored observations are included. For an explanation of both of these aspects, see Appendices A and C.

²³² **2.4 Parameter estimation**

The joint posterior was approximated using a Markov-chain Monte Carlo
²³⁴ routine that is a variant of Hamiltonian Monte Carlo called the No-U-Turn
Sampler (Hoffman and Gelman, 2014) as implemented in the probabilistic
²³⁶ programming language Stan (Stan Development Team, 2014). The posterior
distribution was approximated from four parallel chains run for 10,000 draws
²³⁸ 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
²⁴⁰ reduction factor \hat{R} where values close to 1 ($\hat{R} < 1.1$) indicate approximate
convergence. Convergence means that the chains are approximately stationary
²⁴² and the samples are well mixed (Gelman et al., 2013).

2.5 Model evaluation

²⁴⁴ Models were evaluated using both posterior predictive checks and an estimate of
out-of-sample predictive accuracy. The motivation behind posterior predictive
²⁴⁶ 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.,
²⁴⁸ 2013). Systematic differences between the simulations and observations indicate
weaknesses of the model fit. An example of a technique that is very similar
²⁵⁰ would be inspecting the residuals from a linear regression.

The strategy behind posterior predictive checks is to draw simulated values
²⁵² 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
²⁵⁴ 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
²⁵⁶ as α 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 replicated data set
258 y^{rep} . This is repeated 1000 times in order to provide a distribution of possible
values that could have been observed given the model.

260 In order to compare the fitted model to the observed data, various graphical
comparisons or test quantities need to be defined. The principal comparison
262 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
264 replicated data sets. The purpose of this comparison is to determine if the
model approximates the same survival/extinction pattern as the original data.

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

3 Results

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

Comparisons of the survival functions estimated from 1000 posterior predictive
282 data sets to the estimated survival function of the observed genera demonstrates
that both the exponential and Weibull models approximately capture the
284 observed pattern of extinction (Fig. 1). The major difference in fit between the
two models is that the Weibull model has a slightly better fit for longer lived
286 taxa than the exponential model.

Additionally, the Weibull model is expected to have slightly better out-of-sample
288 predictive accuracy when compared to the exponential model (WAIC 4576
versus 4604, respectively). 1). Because the difference in WAIC between these
290 two models is large, while results from both the exponential and Weibull models
will be presented, only those from the Weibull model will be discussed.

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

I find that while the mean estimate of the effect of body size on extinction risk
298 is negative, implying that increasing body size decreases extinction risk, this
estimate is within 2 standard deviations of 0 (mean $\mu_m = -0.09$, standard
300 deviation 0.09; Fig. 1). Because of this, I infer that body size has no
distinguishable effect on brachiopod taxonomic survival.

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

preference on extinction risk.

³⁰⁸ Given mean estimated extinction risk $\tilde{\sigma}$, we can define the extinction risk multiplier of an observation with environmental preference v_i as

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

³¹⁰ This function $f(v_i)$ has a y-intercept of $\exp(0)$ or 1 because it does not have a non-zero intercept term. Equation 3 can be either concave up or down. A ³¹² concave down $f(v_i)$ may indicate that genera of intermediate environmental preference have greater durations than either extreme, and *vice versa* for ³¹⁴ concave up function.

The expected effect of environmental preference as a multiplier of expected ³¹⁶ extinction risk can then be visualized (Fig. 2). This figure depicts 1000 posterior predictive estimates of Eq. 3 across all possible values of v . The number ³¹⁸ indicates the posterior probability that the function is concave down, with generalists having lower extinction risk/greater duration than either type of ³²⁰ specialist. Note that the inflection point/optimum of Fig. 2 is approximately $x = 0$, something that is expected given the estimate of μ_v (Fig. 1).

³²² The matrix Σ describing the covariance between the different coefficients describes how these coefficients might vary together across the origination ³²⁴ cohorts. Similar to how this was modeled (Eq. 1, 2), for interpretation purposes Σ can be decomposed into a vector of standard deviations $\vec{\tau}$ and a correlation ³²⁶ matrix Ω .

The estimates of the standard deviation of between-cohort coefficient estimates ³²⁸ τ indicate that some effects can vary greatly between-cohorts (Fig. 1). Coefficients with greater values of τ have greater between-cohort variation. The ³³⁰ 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 variation, as there
332 is less between cohort variation than the between cohort variation in baseline
extinction risk β_0 . However the amount of between cohort variation in estimates
334 of β_{v^2} means that it is possible for the function describing the effect of
environmental affinity to be upward facing for some cohorts (Eq. 3), which
336 corresponds to environmental generalists being shorter lived than specialists in
that cohort.

338 The correlation terms of Ω (Fig. 3a) describe the relationship between the
coefficients and how their estimates may vary together across cohorts. The
340 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. 3a first
342 column/last row). Keep in mind that when β_0 is low, extinction risk is low; and
conversely, when β_0 is high, then extinction risk is high.

344 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
346 between expected extinction risk and both geographic range β_r and β_{v^2} are of
particular note (Fig. 3b).

348 There is approximately a 98% probability that β_0 and β_r are negatively
correlated (Fig. 3b), meaning that as extinction risk increases, the
350 effect/importance of geographic range on genus duration increases. This means
that increases in baseline extinction rate are correlated with an increased
352 importance of geographic range size. There is a 93% probability that β_0 and β_{v^2}
are negatively correlated (Fig. 3b), meaning that as extinction risk increases,
354 the peakedness of $f(v_i)$ increases and the relationship tends towards concave
down. Additionally, there is a 97% probability that values of β_r and β_{v^2} are
356 positively correlated (Mean correlation 0.51, standard deviation 0.23).

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

In comparison to the overall mean extinction risk μ_0 , cohort level estimates β_0
362 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
364 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
366 values of β_r correspond to greater decrease in extinction risk with increasing
geographic range

368 How the effect of environmental affinity varies between cohorts can be observed
by using the cohort specific coefficients estimates. Following the same procedure
370 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
372 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.

374 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
376 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,
378 though two of those cases have approximately a 50% probability of being either
concave up or down. This is congruent with the 0.72 posterior probability that
380 μ_{v^2} is positive/ $f(v_i)$ is concave down.

Additionally, for some cohorts there is a quite striking pattern where the effect
382 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
384 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
386 the defintion 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
388 preferences are of intermediate extinction risk when compared with
epicontinental specialists (lowest risk) or open-ocean specialists (highest risk).
390 This time period represents most of the Devonian through the early
Carboniferous.

392 4 Discussion

My results demonstrate that both the effects of geographic range and the
394 peakedness/concavity of environmental preference are both negatively
correlated with baseline extinction risk, meaning that as baseline extinction risk
396 increases the effect sizes of both these traits are expected to increase (Fig. 3b).
This result supports neither of the two proposed macroevolutionary mechanisms
398 for how biological traits should correlate with extinction risk. The observed
correlation between the two effects as well as between the effects and baseline
400 extinction risk instead implies that as baseline extinction risk increases, the
strength of the total selection gradient on biological traits (except body size)
402 increases. This manifests as greater differences in extinction risk for each unit
difference in the biological covariates during periods of high extinction risk,
404 while a relatively flatter selection gradient during periods of low extinction risk.

For the approximately 233 My period analyzed there is an approximate 75%
406 posterior probability that brachiopod genera with intermediate environmental
preferences are expected to have a lower extinction risk than either end

⁴⁰⁸ members. However, the over all curvature of $f(v_i)$ is not very peaked meaning
⁴¹⁰ that when averaged over the entire Phanerozoic this relationship may not lead
⁴¹² to large differences in extinction risk (Fig. 2). Note that the duration of the
⁴¹⁴ period analyzed is approximately four times then length of the Cenozoic (e.g.
time since the extinction of the non-avian dinosaurs). This result gives weak
support for the universality of the hypothesis that environmental generalists
have greater survival than environmental specialists (Liow, 2004, 2007,
Nürnberg and Aberhan, 2013, 2015, Simpson, 1944).

⁴¹⁶ 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
⁴¹⁸ 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
⁴²⁰ estimated relationship is in fact the reverse of the prior expectation.

⁴²² Additionally, some of those same stages where $f(v_i)$ does not resemble the prior
expectation of a concave down nonlinear relation are instead is highly skewed
and effectively linear (Fig. 5). These results demonstrate that, while the
⁴²⁴ group-level estimate may only weakly support one hypothesis, the cohort-level
estimates may exhibit very different characteristics. These results are also
⁴²⁶ consistent with aspects of Miller and Foote (2009) who found that the effect of
environmental preference on extinction risk was quite variable and without
⁴²⁸ obvious patterning during times of background extinction.

⁴³⁰ 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.

⁴³² 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
⁴³⁴ (Johnson, 1974, Sheehan, 2001). My results are broadly consistent with this

scenario with both epicontinental and open-ocean specialists having a much
436 lower expected duration than intermediate taxa (Fig. 5). All of the stages
between the Darriwillian and the Llandovery, except the Hirnantian, have a
438 high probability (90%) that $f(v)$ is concave down. The pattern for the
Darriwillian, which proceeds the supposed start of Ordovician glacial activity,
440 demonstrates that taxa tend to favor open-ocean environments are expected to
have a greater duration than either intermediate or epicontinental specialists, in
442 decreasing order.

For nearly the entire Devonian estimates of $f(v)$ indicate that one of the
444 environmental end members is favored over the other end member of
intermediate preference (Fig. 5). This is consistent with the predictions of Miller
446 and Foote (2009). For almost the entirety the Givetian though the end of the
Devonian and into the Viséan, I find that epicontinental favoring taxa are
448 expected to have a greater duration than either intermediate or open-ocean
specialists. Additionally, for nearly the entire Devonian and through to the
450 Visean, the cohort-specific estimates of $f(v)$ are concave-up. This is the opposite
pattern than what is expected (Fig. 2). This result, however, seems to reflect
452 the intensity of the seemingly nearly-linear difference in expected duration
across the range of v) as opposed to an inversion of the weakly expected
454 curvilinear pattern.

Of concern is the use of genera as the unit of the study and how to exactly
456 interpret the effects of the biological traits. For example, if any of the traits
analyzed here are associated with increases in speciation rates, this might
458 “artificially” increase the duration of genera through self-renewal (Raup, 1991,
1994). This could lead to a trait appearing to decrease generic level extinction
460 risk by increasing species level origination rate instead of decreasing species
level extinction risk. However, given the nature of the brachiopod fossil record

⁴⁶² and the difficulty of identifying individual specimens to the species level, there
⁴⁶³ is no simple solution to decreasing this uncertainty in the interpretations of how
⁴⁶⁴ the biological traits studied here actually affect extinction risk.

⁴⁶⁵ This model could be improved through either increasing the number of analyzed
⁴⁶⁶ taxon traits, expanding the hierarchical structure of the model to include other
⁴⁶⁷ major taxonomic groups of interest, and the inclusion of explicit phylogenetic
⁴⁶⁸ relationships between the taxa in the model as an additional hierarchical effect.

⁴⁷⁰ An example taxon trait that may be of particular interest is the affixing
⁴⁷¹ strategy or method of interaction with the substrate of the taxon. This trait has
⁴⁷² been found to be related to brachiopod survival (Alexander, 1977) so its
inclusion may be of particular interest.

⁴⁷⁴ It is theoretically possible to expand this model to allow for comparisons within
⁴⁷⁵ and between major taxonomic groups. This approach would better constrain the
⁴⁷⁶ brachiopod estimates while also allowing for estimation of similarities and
differences in cross-taxonomic patterns. The major issue surrounding this
particular expansion involves finding a similarly well sampled taxonomic group
⁴⁷⁸ that is present during the Paleozoic. Example groups include Crinoidea,
Ostracoda, and other “Paleozoic” groups (Sepkoski Jr., 1981).

⁴⁸⁰ Taxon traits like environmental preference or geographic range (Hunt et al.,
2005, Jablonski, 1987) are most likely heritable, at least phylogenetically
⁴⁸² (Housworth et al., 2004, Lynch, 1991). Without phylogenetic context, this
analysis assumes that differences in extinction risk between taxa are
⁴⁸⁴ independent of those taxa's shared evolutionary history (Felsenstein, 1985). In
contrast, the origination cohorts only capture shared temporal context. The
⁴⁸⁶ inclusion of phylogenetic context as an addition individual level hierarchical
structure independent of origination cohort would allow for determining how

⁴⁸⁸ much of the observed variability is due to shared evolutionary history versus
⁴⁸⁹ actual differences associated with these taxonomic traits.

⁴⁹⁰ In summary, patterns of Paleozoic brachiopod survival were analyzed using a
⁴⁹¹ fully Bayesian hierarchical survival modelling approach while also eschewing the
⁴⁹² traditional separation between background and mass extinction. I modeled both
⁴⁹³ the overall mean effect of biological covariates on extinction risk while also
⁴⁹⁴ modeling the correlation between cohort-specific estimates of covariate effects. I
⁴⁹⁵ find that as baseline extinction risk increases, the strength of the selection
⁴⁹⁶ gradient on biological traits (except body size) increases. This manifests as
⁴⁹⁷ greater differences in extinction risk for each unit difference in the biological
⁴⁹⁸ covariates during periods of high extinction risk, while a much flatter total
⁴⁹⁹ selection gradient during periods of low extinction risk. I also find some support
⁵⁰⁰ for “survival of the unspecialized” (Liow, 2004, 2007, Nürnberg and Aberhan,
⁵⁰¹ 2013, 2015, Simpson, 1944) as a general characterization of the effect of
⁵⁰² environmental preference on extinction risk (Fig. 2), though there is
⁵⁰³ heterogeneity between origination cohorts with most periods of time conforming
⁵⁰⁴ to this hypothesis (Fig. 5).

Acknowledgements

⁵⁰⁶ I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for
⁵⁰⁷ helpful discussion and advice. Additionally, thank you A. Miller for the
⁵⁰⁸ epicontinental versus open-ocean assignments. This entire study would
⁵⁰⁹ not have been possible without the Herculean effort of the many contributors to
⁵¹⁰ the Paleobiology Database. In particular, I would like to thank J. Alroy, M.
⁵¹¹ Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N. Heim, A. Hendy, S. Holland,
⁵¹² L. Ivany, W. Kiessling, B. Kröger, A. McGowan, T. Olszewski, P.

Novack-Gottshall, M. Patzkowsky, M. Uhen, L. Villier, and P. Wager. This work
514 was supported by a NASA Exobiology grant (NNX10AQ446) to A. Miller and
M. Foote. I declare no conflicts of interest. This is Paleobiology Database
516 publication XXX.

References

- 518 Alexander, R. R., 1977. Generic longevity of articulate brachiopods in relation
to the mode of stabilization on the substrate. *Palaeogeography,*
520 *Palaeoclimatology, Palaeoecology* 21:209–226.
- Alroy, J., 2010. The Shifting Balance of Diversity Among Major Marine Animal
522 Groups. *Science* 329:1191–1194.
- Baumiller, T. K., 1993. Survivorship analysis of Paleozoic Crinoidea: effect of
524 filter morphology on evolutionary rates. *Paleobiology* 19:304–321.
- Cooper, W. S., 1984. Expected time to extinction and the concept of
526 fundamental fitness. *Journal of Theoretical Biology* 107:603–629.
- Felsenstein, J., 1985. Phylogenies and the comparative method. *American
528 Naturalist* 125:1–15.
- Foote, M., 1988. Survivorship analysis of Cambrian and Ordovician Trilobites.
530 *Paleobiology* 14:258–271.
- , 2000. Origination and extinction components of taxonomic diversity:
532 Paleozoic and post-Paleozoic dynamics. *Paleobiology* 26:578–605.
- 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 Society B: Biological Sciences* 282.
- Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL <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 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 Model. *The American Naturalist* 163:84–96.
- Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a comment on "On the heritability of geographic range sizes". *American Naturalist* 166:129–135.
- Jablonski, D., 1986. Background and mass extinctions: the alternation of macroevolutionary regimes. *Science* 231:129–133.
- , 1987. Heritability at the species level: analysis of geographic ranges of cretaceous mollusks. *Science* 238:360–363.
- Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and living molluscs. *Proceedings. Biological sciences / The Royal Society* 270:401–6.
- Johnson, J. G., 1974. Extinction of Perched Faunas. *Geology* 2:479–482.
- Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine benthic biodiversity dynamics through Triassic–Jurassic time. *Paleobiology* 33:414–434.

- 564 Klein, J. P. and M. L. Moeschberger, 2003. Survival Analysis: Techniques for
Censored and Truncated Data. 2nd ed. Springer, New York.
- 566 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.
- 568 ———, 2007. Does versatility as measured by geographic range, bathymetric
range and morphological variability contribute to taxon longevity? *Global
Ecology and Biogeography* 16:117–128.
- 570 Lynch, M., 1991. Methods for the analysis of comparative data in evolutionary
biology. *Evolution* 45:1065–1080.
- 572 Miller, A. I. and M. Foote, 2009. Epicontinental seas versus open-ocean settings:
the kinetics of mass extinction and origination. *Science* 326:1106–9.
- 574 Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range
predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*
39:360–372.
- 578 ———, 2015. Interdependence of specialization and biodiversity in Phanerozoic
marine invertebrates. *Nature communications* 6:6602.
- 580 Palmer, M. E. and M. W. Feldman, 2012. Survivability is more fundamental
than evolvability. *PloS one* 7:e38025.
- 582 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.
- 584 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.

- 588 Peters, S. E., 2008. Environmental determinants of extinction selectivity in the fossil record. *Nature* 454:626–9.
- 590 Raup, D. M., 1975. Taxonomic survivorship curves and Van Valen's Law. *Paleobiology* 1:82–96.
- 592 ———, 1978. Cohort Analysis of generic survivorship. *Paleobiology* 4:1–15.
- , 1991. *Extinction: Bad Genes or Bad Luck?* Norton, New York.
- 594 ———, 1994. The role of extinction in evolution. *Proceedings of the National Academy of Sciences* 91:6758–6763.
- 596 Sepkoski Jr., J. J., 1981. A factor analytic description of the Phanerozoic marine fossil record. *Paleobiology* 7:36–53.
- 598 Sheehan, P., 2001. The late Ordovician mass extinction. *Annual Review of Earth and Planetary Sciences* 29:331–364.
- 600 Simpson, C., 2006. Levels of selection and large-scale morphological trends. Ph.D. thesis, University of Chicago.
- 602 Simpson, C. and P. G. Harnik, 2009. Assessing the role of abundance in marine bivalve extinction over the post-Paleozoic. *Paleobiology* 35:631–647.
- 604 Simpson, G. G., 1944. *Tempo and Mode in Evolution*. Columbia University Press, New York.
- 606 ———, 1953. *The Major Features of Evolution*. Columbia University Press, New York.
- 608 Stan Development Team, 2014. Stan: A c++ library for probability and sampling, version 2.5.0. URL <http://mc-stan.org/>.
- 610 Van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1:1–30.
- , 1979. Taxonomic survivorship curves. *Evolutionary Theory* 4:129–142.

- 612 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
614 latitudinal selectivity gradient at the end-Cretaceous mass extinction.
Scientific Reports 3:1790.
- 616 Wang, S. C., 2003. On the continuity of background and mass extinction.
Paleobiology 29:455–467.
- 618 Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and
Widely Applicable Information Criterion in Singular Learning Theory.
620 Journal of Machine Learning Research 11:3571–3594.

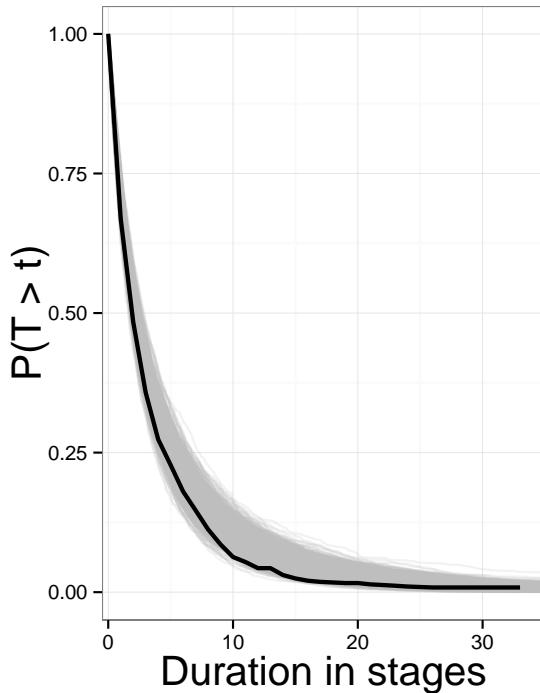


Figure 1: Comparison of empirical estimates of $S(t)$ versus estimates from 1000 posterior predictive data sets. $S(t)$ corresponds to $P(T > t)$ as it is the probability that a given genus observed at age t will continue to live. This is equivalent to the probability that t is less than the genus' ultimate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

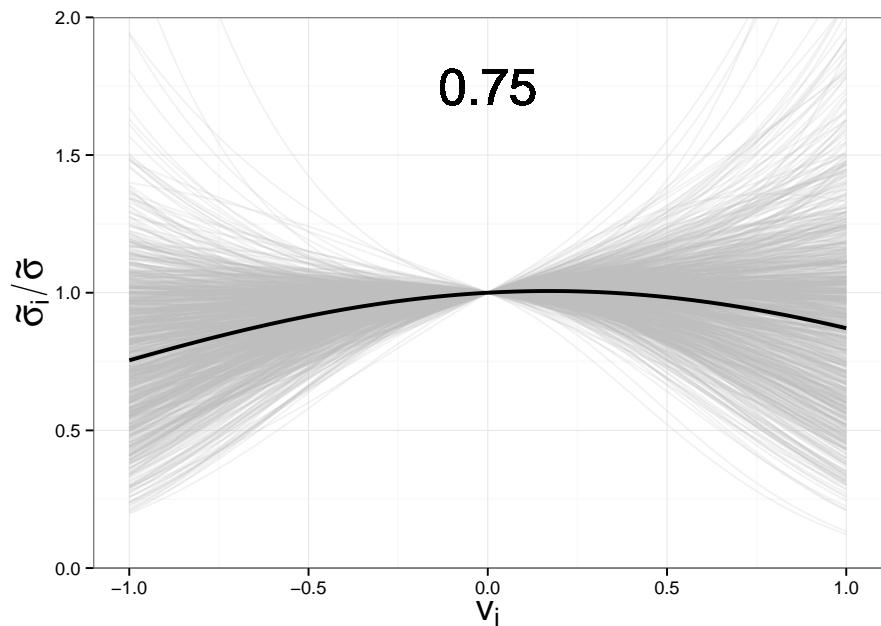


Figure 2: The overall expected relationship $f(v_i)$ between environmental affinity v_i and a multiplier of extinction risk (Eq. 3). Each grey line corresponds to a single draw from the posterior predictive distribution, while the black corresponds to the median of the posterior predictive distribution. The overall shape of $f(v_i)$ is concave down with an optimum of close 0, which corresponds to affinity approximately equal to the expectation based on background environmental occurrence rates.

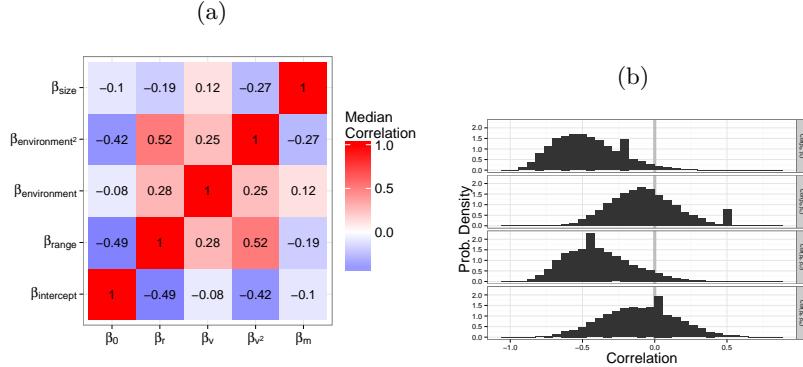


Figure 3: **A:** Heatmap for the median estimates of the terms of the correlation matrix Ω between cohort-level covariate effects. Both the exponential (left) and Weibull (right) models are presented. The off-diagonal terms are the correlation between the estimates of the cohort-level estimates of the 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	-1.51	0.15
μ_r	-1.38	0.14
μ_e	-0.08	0.18
μ_{e2}	0.25	0.43
μ_m	-0.09	0.09
τ_i	0.63	0.11
τ_r	0.48	0.12
τ_e	1.07	0.23
τ_{e^2}	1.88	0.66
τ_m	0.32	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, e environmental affinity, e^2 environmental affinity squared, m body size.

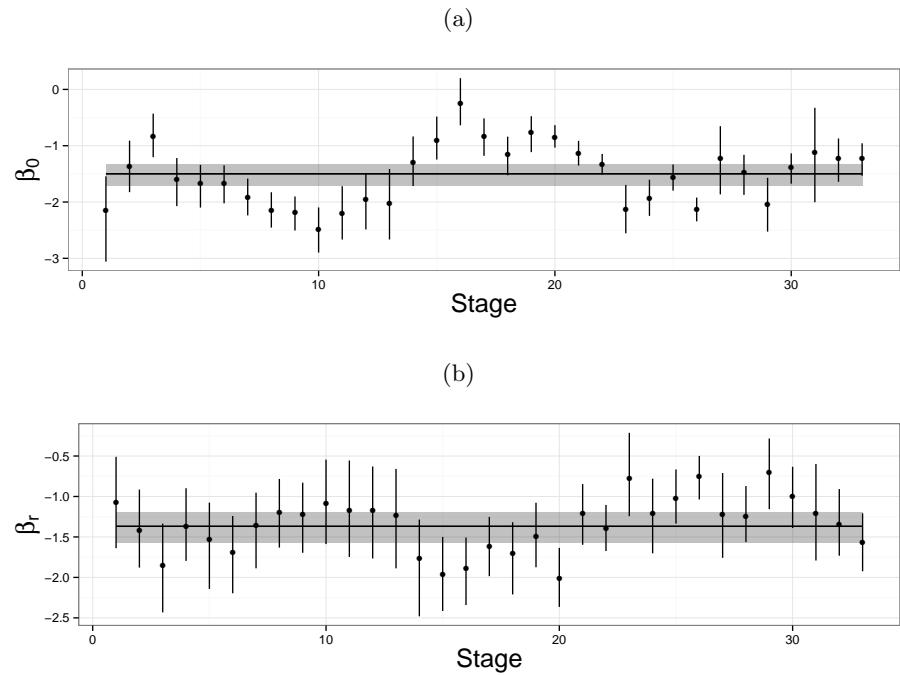


Figure 4: Comparison of cohort-specific estimates of β_0 presented along with the estimate for the overall baseline extinction risk. Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. The horizontal line is the median estimate of the overall baseline extinction risk along with 80% credible intervals. Results are presented for the exponential (top) and Weibull (bottom) models. Comparison of cohort-specific estimates of the effect of geographic range on extinction risk β_r presented along with the estimate for the overall effect of geographic range. Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. The horizontal line is the median estimate of the overall baseline extinction risk along with 80% credible intervals. Results are presented for the exponential (top) and Weibull (bottom) models.

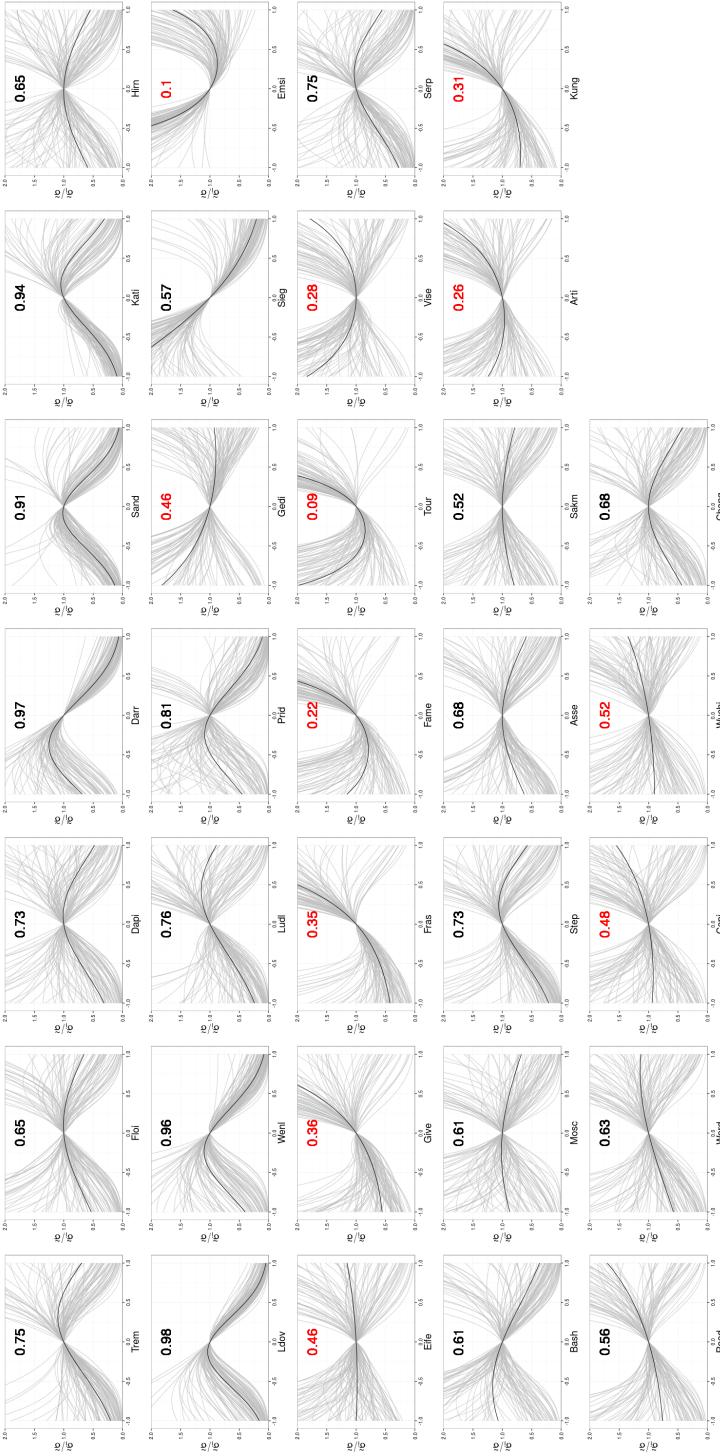


Figure 5: Comparison of the cohort-specific estimates of $f(v_i)$ (Eq. 3) for the 33 analyzed origination cohorts. The stage of origination is labeled on the x-axis of each panel. The oldest stage is in the upper left, while the youngest is in the lower left. The number in each panel corresponds to the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