

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

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. I analyze patterns of Paleozoic brachiopod genus durations and their relationship to geographic range, affinity for epicontinental seas versus open ocean environments, and body size. Additionally, I allow for environmental affinity to have a nonlinear effect on duration. Using a hierarchical Bayesian modeling approach, I also model the interaction between the effects of the biological traits and a taxon's time of origination. My analysis framework eschews the traditional distinction between background and mass extinction, instead the entire time period is analyzed where these "states" are part of a continuum. I find evidence that as extinction risk increases, the expected strength of the effects of biological traits (except body size) increase. 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. 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. I find weak evidence for a nonlinear relationship between environmental preference and extinction risk such that "generalists" have a lower expected extinction risk than either "specialists". Importantly, I find that as baseline 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 at
a mass extinction event, the effects of biological traits on taxonomic survival
40 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
42 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
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 trait effects need to be modeled as well.

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

58 I choose to model brachiopod taxon durations because trait based differences in
extinction risk should manifest as differences in taxon durations. Namely, a
60 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
62 aspect of “taxon fitness” along with expected lineage specific

branching/originatation rate (Cooper, 1984, Palmer and Feldman, 2012).

64 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
66 on the brachiopod record from most of the Paleozoic, from the start of the
Ordovician (approximately 485 Mya) through the end Permian (approximately
68 252 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 for the entire
78 period analyzed.

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

86 Epicontinental seas are a shallow-marine environment where the ocean has
spread over the surface of a continental shelf with a depth typically less than
88 100m. In contrast, open-ocean coastline environments have much greater

variance in depth, do not cover the continental shelf, and can persist during
90 periods of low sea level. Because of this, it is strongly expected that taxa which
favor epicontinental seas would be at great risk during periods of low sea levels,
92 such as during glacial periods, where these seas are drained. During the
Paleozoic (approximately 541–252 My), epicontinental seas were widely spread
94 globally but declined over the Mesozoic (approximately 252–66 My) and
eventually diminished disappearing during the Cenozoic (approximately 66–0
96 My) as open-ocean coastlines became the dominant shallow-marine setting
(Johnson, 1974, Miller and Foote, 2009, Peters, 2008).

98 Given the above findings, I predict that as extinction risk increases, taxa
associated with open-ocean environments should generally increase in extinction
100 risk versus those that favor epicontinental seas. Additionally, there is a possible
nonlinear relationship between environmental preference and taxon duration. A
102 long standing hypothesis is that generalists or unspecialized taxa will have
greater survival than specialists (Baumiller, 1993, Liow, 2004, 2007, Nürnberg
104 and Aberhan, 2013, 2015, Simpson, 1944). In this analysis I allowed for
environmental preference to have a parabolic effect on taxon duration
106 Body size, measured as shell length (Payne et al., 2014), was also considered as
a potentially informative covariate. Body size is a proxy for metabolic activity
108 and other correlated life history traits (Payne et al., 2014). There is no strong
hypothesis of how body size effects extinction risk in brachiopods, such that a
110 positive, negative, or zero effect are all plausible.

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

¹¹⁶ biological traits on survival, especially in cases where the covariates of interest
(i.e. biological traits) are themselves known with error.

¹¹⁸ **2 Materials and Methods**

2.1 Fossil occurrence information

¹²⁰ The dataset analyzed here was sourced from the Paleobiology Database
(<http://www.paleodb.org>) which was then filtered based on taxonomic,
¹²² temporal, stratigraphic, and other occurrence information that was necessary
for this analysis. These filtering criteria are very similar to those from Foote and
¹²⁴ Miller (2013) with an additional constraint of being present in the body size
data set from Payne et al. (2014). Epicontinental versus open-ocean assignments
¹²⁶ for each fossil occurrence are partially based on those from Miller and Foote
(2009), with additional occurrences assigned similarly (Miller and Foote,
¹²⁸ personal communication).

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

¹³⁴ While species diversity dynamics are of much greater interest than those of
higher taxa, the nature of the fossil record makes accurate and precise
¹³⁶ taxonomic assignments at the species level for all occurrences extremely difficult
if not impossible. In particular, the simplicity of brachiopod external
¹³⁸ morphology and the quality of preservation makes it very difficult for
assignments below the genus level. As such, the choice to analyze genera as

¹⁴⁰ opposed to species was in order to assure a minimum level of confidence and accuracy in the data analyzed here.

¹⁴² Genus duration was calculated as the number of geologic stages from first appearance to last appearance, inclusive. Durations were based on geologic ¹⁴⁴ 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 ¹⁴⁶ fossils are known from a geological interval which represents some temporal range. Stages act as effectively irreducible globally consistent temporal intervals ¹⁴⁸ in which taxa 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 left censored (Appendix A). The covariates used to model genus duration were ¹⁵² 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 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 $250,000 \text{ km}^2$. ¹⁵⁶ The map projection and regular lattice were made using shape files from ¹⁵⁸ <http://www.naturalearthdata.com/> and the **raster** package for R (Hijmans, 2015).

¹⁶⁰ 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 by that genus was calculated. Dividing the genus occupancy by the total ¹⁶² occupancy gives the relative occupancy of that genus. Mean relative genus occupancy was then calculated as the mean of the per stage relative occupancies ¹⁶⁴ of that genus.

¹⁶⁶ Body size data was sourced directly from Payne et al. (2014). Because those
measurements are presented without error, a measurement error model similar
¹⁶⁸ to the one for environmental affinity could not be implemented (Appendix A).

Prior to analysis, geographic range and body size were transformed and
¹⁷⁰ standardized in order to improve interpretability of the results. Geographic
range, which can only vary between 0 and 1, was logit transformed. Body size,
¹⁷² 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
¹⁷⁴ their standard deviation following Gelman and Hill (2007).

2.2 Analytical approach

¹⁷⁶ Hierarchical modelling is a statistical approach which explicitly takes into
account the structure of the observed data in order to model both the within
¹⁷⁸ 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
¹⁸⁰ cohort). These groups are considered separate draws from a shared probability
distribution (e.g. all cohorts, observed and unobserved). The group-level
¹⁸² parameters are then estimated simultaneously as the other parameters of
interest (e.g. covariate effects) (Gelman et al., 2013). The subsequent estimates
¹⁸⁴ 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
¹⁸⁶ pulled towards the overall group mean.

This partial pooling is one of the greatest advantages of hierarchical modeling.
¹⁸⁸ By letting the groups “support” each other, parameter estimates then better
reflect our statistical uncertainty. Additionally, this partial pooling helps control
¹⁹⁰ for multiple comparisons and possibly spurious results as effects with little

support are drawn towards the overall group mean (Gelman et al., 2013,
192 Gelman and Hill, 2007).

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

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

2.3 Survival model

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

212 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
214 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

²¹⁶ the between-cohort variation in regression coefficient estimates, and Ω is the correlation matrix of the regression coefficients.

²¹⁸ 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

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

²²⁰ The principal difference between this model and the previous (Eq. 1) is the
²²¹ inclusion of the shape parameter α . Note that σ is approximately equivalent to
²²² $1/\lambda$.

²²³ For an explanation of how this model was developed, parameter explanations,
²²⁴ and choice of priors, please see Appendix B. Note that these models (Eq. 1, 2)
²²⁵ do not include how the uncertainty in environmental affinity is included nor how
²²⁶ censored observations are included. For an explanation of both of these aspects,
see Appendices A and C.

²²⁸ **2.4 Parameter estimation**

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

2.5 Model evaluation

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

The strategy behind posterior predictive checks is to draw simulated values from the joint posterior predictive distribution, $p(y^{rep}|y)$, and then compare those draws to the empirically observed values (Gelman et al., 2013). To accomplish this, for each replicate, a single value is drawn from the marginal posterior distributions of each regression coefficient from the final model as well as α for the Weibull model (Eq. 1, 2). Then, given the covariate information \mathbf{X} ,

a new set of n genus durations are generated giving a single replicated data set
254 y^{rep} . This is repeated 1000 times in order to provide a distribution of possible
values that could have been observed given the model.

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

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

3 Results

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

Comparisons of the survival functions estimated from 1000 posterior predictive
278 data sets to the estimated survival function of the observed genera demonstrates
that both the exponential and Weibull models approximately capture the
280 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
282 taxa than the exponential model.

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

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

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

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

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

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

This function $f(v_i)$ has a y-intercept of $\exp(0)$ or 1 because it does not have a
 306 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
 308 preference have greater durations than either extreme, and *vice versa* for
 concave up function.

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

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

322 The estimates of the standard deviation of between-cohort coefficient estimates
 τ indicate that some effects can vary greatly between-cohorts (Fig. 1).
 324 Coefficients with greater values of τ have greater between-cohort variation. The
 covariate effects with the greatest between origination cohort variation are β_r ,
 326 β_v , and β_{v^2} . Estimates of β_m have negligible between cohort variation, as there

is less between cohort variation than the between cohort variation in baseline
328 extinction risk β_0 . However the amount of between cohort variation in estimates
of β_{v^2} means that it is possible for the function describing the effect of
330 environmental affinity to be upward facing for some cohorts (Eq. 3), which
corresponds to environmental generalists being shorter lived than specialists in
332 that cohort.

The correlation terms of Ω (Fig. 3a) describe the relationship between the
334 coefficients and how their estimates may vary together across cohorts. The
correlations between the intercept term β_0 and the effects of the taxon traits are
336 of particular interest for evaluating the Jablonski (1986) scenario (Fig. 3a first
column/last row). Keep in mind that when β_0 is low, extinction risk is low; and
338 conversely, when β_0 is high, then extinction risk is high.

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

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

352 While the overall group level estimates are of particular importance when

defining time-invariant differences in extinction risk, it is also important and
354 useful to analyze the individual level parameter estimates in order to better
understand how parameters actually vary across cohorts.

356 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
358 (Fig. 4a). A similar, if slightly greater, amount of variation is also observable in
cohort estimates of the effect of geographic range β_r (Fig. 4b). Again, smaller
360 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
362 geographic range

How the effect of environmental affinity varies between cohorts can be observed
364 by using the cohort specific coefficients estimates. Following the same procedure
used earlier (Fig. 1), but substituting cohort specific estimates of β_v and β_{v^2} for
366 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
368 model, though the observed pattern should be similar for the exponential model.

As expected based on the estimates of τ_v and τ_{v^2} , there is greater variation in
370 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
372 that generalists are potentially expected to be shorter lived than specialists,
though two of those cases have approximately a 50% probability of being either
374 concave up or down. This is congruent with the 0.72 posterior probability that
 μ_{v^2} is positive/ $f(v_i)$ is concave down.

376 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
378 primarily scenarios where one of the end member preferences is expected to

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

4 Discussion

388 My results demonstrate that both the effects of geographic range and the peakedness/concavity of environmental preference are both negatively correlated
389 with baseline extinction risk, meaning that as baseline extinction risk increases the effect sizes of both these traits are expected to increase (Fig. 3b). This
392 result supports neither of the two proposed end-member macroevolutionary mechanisms for how biological traits should correlate with extinction risk. The
394 observed correlation between the two effects as well as between the effects and baseline extinction risk instead implies that as baseline extinction risk increases, the strength of the total selection gradient on biological traits (except body size)
396 increases. This manifests as greater differences in extinction risk for each unit difference in the biological covariates during periods of high extinction risk,
398 while a relatively flatter selection gradient during periods of low extinction risk.
400 For the approximately 233 My period analyzed there is an approximate 75% posterior probability that brachiopod genera with intermediate environmental
402 preferences are expected to have a lower extinction risk than either end members. However, the overall curvature of $f(v_i)$ is not very peaked meaning

404 that when averaged over the entire Phanerozoic this relationship may not lead
to large differences in extinction risk (Fig. 2). Note that the duration of the
406 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
408 support for the universality of the hypothesis that environmental generalists
have greater survival than environmental specialists (Liow, 2004, 2007,
410 Nürnberg and Aberhan, 2013, 2015, Simpson, 1944).

The posterior variance in the estimate of overall $f(v_i)$ reflects the large between
412 cohort variance in cohort specific estimates of $f(v_i)$ (Fig. 5). Given that there is
only a 75% posterior probability that the expected overall estimate of $f(v_i)$ is
414 concave down, it is not surprising that there are some stages where the
estimated relationship is in fact the reverse of the prior expectation.

416 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
418 and effectively linear (Fig. 5). These results demonstrate that, while the
group-level estimate may only weakly support one hypothesis, the cohort-level
420 estimates may exhibit very different characteristics. These results are also
consistent with aspects of Miller and Foote (2009) who found that the effect of
422 environmental preference on extinction risk was quite variable and without
obvious patterning during times of background extinction.

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

The proposed mechanism for the end Ordovician mass extinction is a decrease
428 in sea level and the draining of epicontinental seas due to protracted glaciation
(Johnson, 1974, Sheehan, 2001). My results are broadly consistent with this
430 scenario with both epicontinental and open-ocean specialists having a much

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

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

450 The use of genera as the unit of the study and how to exactly interpret the
effects of the biological traits is a remaining concern. For example, if any of the
452 traits analyzed here are associated with increases in speciation rates, this might
“artificially” increase the duration of genera through self-renewal (Raup, 1991,
454 1994). This could lead to a trait appearing to decrease generic level extinction
risk by increasing species level origination rate instead of decreasing species
456 level extinction risk. However, given the nature of the brachiopod fossil record
and the difficulty of identifying individual specimens to the species level, there

458 is no simple solution to decreasing this uncertainty in the interpretations of how
the biological traits studied here actually affect extinction risk.

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

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

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

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

⁴⁸⁴ actual differences associated with these taxonomic traits.

In summary, patterns of Paleozoic brachiopod survival were analyzed using a
⁴⁸⁶ fully Bayesian hierarchical survival modelling approach while also eschewing the
traditional separation between background and mass extinction. I modeled both
⁴⁸⁸ the overall mean effect of biological covariates on extinction risk while also
modeling the correlation between cohort-specific estimates of covariate effects. I
⁴⁹⁰ find that as baseline extinction risk increases, the strength of the selection
gradient on biological traits (except body size) increases. This manifests as
⁴⁹² greater differences in extinction risk for each unit difference in the biological
covariates during periods of high extinction risk, while a much flatter total
⁴⁹⁴ selection gradient during periods of low extinction risk. I also find very weak for
“survival of the unspecialized” (Liow, 2004, 2007, Nürnberg and Aberhan, 2013,
⁴⁹⁶ 2015, Simpson, 1944) as a general characterization of the effect of environmental
preference on extinction risk (Fig. 2), though there is heterogeneity between
⁴⁹⁸ origination cohorts with most periods of time conforming to this hypothesis
(Fig. 5).

⁵⁰⁰ Acknowledgements

I would like to thank K. Angielczyk, M. Foote, P. D. Polly, and R. Ree for
⁵⁰² helpful discussion and advice. Additionally, thank you A. Miller for the
epicontinental versus open-ocean assignments. This entire study would
⁵⁰⁴ not have been possible without the Herculean effort of the many contributors to
the Paleobiology Database. In particular, I would like to thank J. Alroy, M.
⁵⁰⁶ Aberhan, D. Bottjer, M. Clapham, F. Fürsich, N. Heim, A. Hendy, S. Holland,
L. Ivany, W. Kiessling, B. Kröger, A. McGowan, T. Olszewski, P.
⁵⁰⁸ Novack-Gottshall, M. Patzkowsky, M. Uhen, L. Villier, and P. Wager. This work

was supported by a NASA Exobiology grant (NNX10AQ446) to A. Miller and
510 M. Foote. I declare no conflicts of interest. This is Paleobiology Database
publication XXX.

512 References

- Alexander, R. R., 1977. Generic longevity of articulate brachiopods in relation
514 to the mode of stabilization on the substrate. *Palaeogeography, Palaeoclimatology, Palaeoecology* 21:209–226.
- 516 Alroy, J., 2010. The Shifting Balance of Diversity Among Major Marine Animal Groups. *Science* 329:1191–1194.
- 518 Baumiller, T. K., 1993. Survivorship analysis of Paleozoic Crinoidea: effect of filter morphology on evolutionary rates. *Paleobiology* 19:304–321.
- 520 Cooper, W. S., 1984. Expected time to extinction and the concept of fundamental fitness. *Journal of Theoretical Biology* 107:603–629.
- 522 Felsenstein, J., 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- 524 Foote, M., 1988. Survivorship analysis of Cambrian and Ordovician Trilobites. *Paleobiology* 14:258–271.
- 526 ———, 2000. Origination and extinction components of taxonomic diversity: Paleozoic and post-Paleozoic dynamics. *Paleobiology* 26:578–605.
- 528 Foote, M. and A. I. Miller, 2013. Determinants of early survival in marine animal genera. *Paleobiology* 39:171–192.
- 530 Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013. Bayesian data analysis. 3 ed. Chapman and Hall, Boca Raton, FL.
- Gelman, A. and J. Hill, 2007. Data Analysis using Regression and
534 Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.
- Harnik, P. G., C. Simpson, and J. L. Payne, 2013. Long-term differences in

- 536 extinction risk among the seven forms of rarity. Proceedings of the Royal
Society B: Biological Sciences 282.
- 538 Hijmans, R. J., 2015. raster: Geographic data analysis and modeling. URL
<http://CRAN.R-project.org/package=raster>. R package version 2.3-24.
- 540 Hoffman, M. D. and A. Gelman, 2014. The No-U-Turn Sampler: Adaptively
Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine
542 Learning Research 15:1351–1381.
- Housworth, E. A., P. Martins, and M. Lynch, 2004. The Phylogenetic Mixed
544 Model. The American Naturalist 163:84–96.
- Hunt, G., K. Roy, and D. Jablonski, 2005. Species-level heritability reaffirmed: a
546 comment on "On the heritability of geographic range sizes". American
Naturalist 166:129–135.
- 548 Jablonski, D., 1986. Background and mass extincitons: the alternation of
macroevolutionary regimes. Science 231:129–133.
- 550 ———, 1987. Heritability at the species level: analysis of geographic ranges of
cretaceous mollusks. Science 238:360–363.
- 552 Jablonski, D. and K. Roy, 2003. Geographical range and speciation in fossil and
living molluscs. Proceedings. Biological sciences / The Royal Society
554 270:401–6.
- Johnson, J. G., 1974. Extinction of Perched Faunas. Geology 2:479–482.
- 556 Kiessling, W. and M. Aberhan, 2007. Environmental determinants of marine
benthic biodiversity dynamics through Triassic Jurassic time. Paleobiology
558 33:414–434.

- Klein, J. P. and M. L. Moeschberger, 2003. Survival Analysis: Techniques for
560 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
562 unspecialized" using fossil crinoids. *The American naturalist* 164:431–43.
- , 2007. Does versatility as measured by geographic range, bathymetric
564 range and morphological variability contribute to taxon longevity? *Global
Ecology and Biogeography* 16:117–128.
- Lynch, M., 1991. Methods for the analysis of comparative data in evolutionary
566 biology. *Evolution* 45:1065–1080.
- Miller, A. I. and M. Foote, 2009. Epicontinental seas versus open-ocean settings:
568 the kinetics of mass extinction and origination. *Science* 326:1106–9.
- Nürnberg, S. and M. Aberhan, 2013. Habitat breadth and geographic range
570 predict diversity dynamics in marine Mesozoic bivalves. *Paleobiology*
572 39:360–372.
- , 2015. Interdependence of specialization and biodiversity in Phanerozoic
574 marine invertebrates. *Nature communications* 6:6602.
- Palmer, M. E. and M. W. Feldman, 2012. Survivability is more fundamental
576 than evolvability. *PloS one* 7:e38025.
- Payne, J. L. and S. Finnegan, 2007. The effect of geographic range on
578 extinction risk during background and mass extinction. *Proceedings of the
National Academy of Sciences* 104:10506–11.
- Payne, J. L., N. A. Heim, M. L. Knope, and C. R. McClain, 2014. Metabolic
580 dominance of bivalves predates brachiopod diversity decline by more than 150
582 million years. *Proceedings of the Royal Society B* 281:20133122.

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

- Vilhena, D. A., E. B. Harris, C. T. Bergstrom, M. E. Maliska, P. D. Ward, C. A.
608 Sidor, C. A. E. Strömberg, and G. P. Wilson, 2013. Bivalve network reveals
latitudinal selectivity gradient at the end-Cretaceous mass extinction.
610 Scientific Reports 3:1790.
- Wang, S. C., 2003. On the continuity of background and mass extinction.
612 Paleobiology 29:455–467.
- Watanabe, S., 2010. Asymptotic Equivalence of Bayes Cross Validation and
614 Widely Applicable Information Criterion in Singular Learning Theory.
Journal of Machine Learning Research 11:3571–3594.

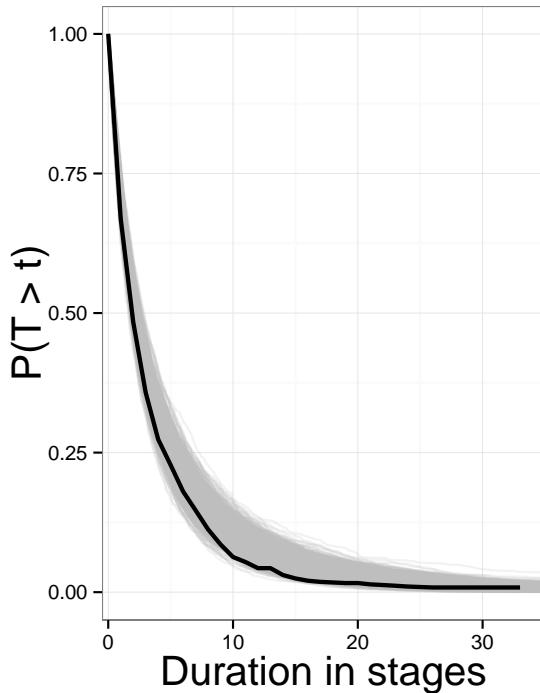


Figure 1: Comparison of empirical estimates of $S(t)$ versus estimates from 1000 posterior predictive data sets. $S(t)$ corresponds to $P(T > t)$ as it is the probability that a given genus observed at age t will continue to live. This is equivalent to the probability that t is less than the genus' ultimate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

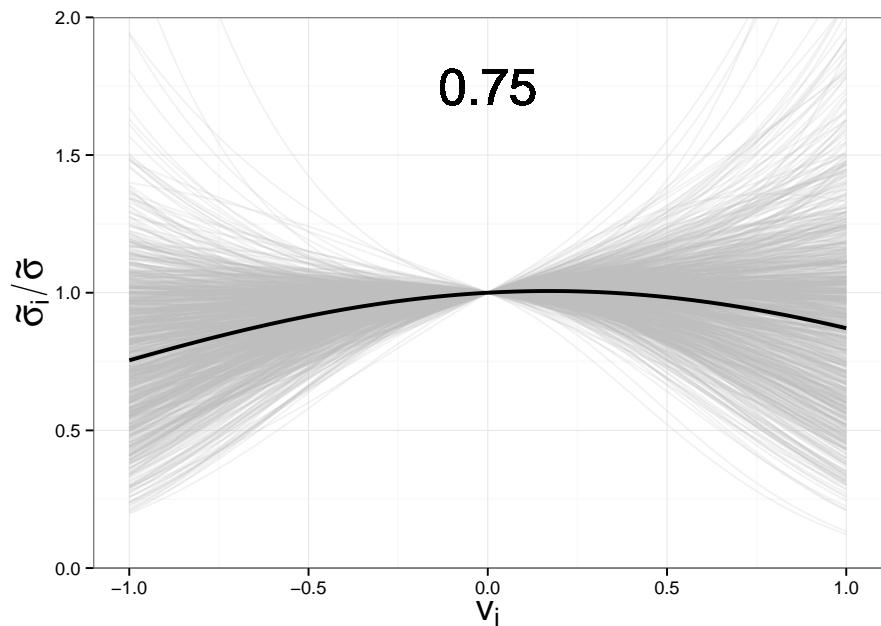


Figure 2: The overall expected relationship $f(v_i)$ between environmental affinity v_i and a multiplier of extinction risk (Eq. 3). Each grey line corresponds to a single draw from the posterior predictive distribution, while the black corresponds to the median of the posterior predictive distribution.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

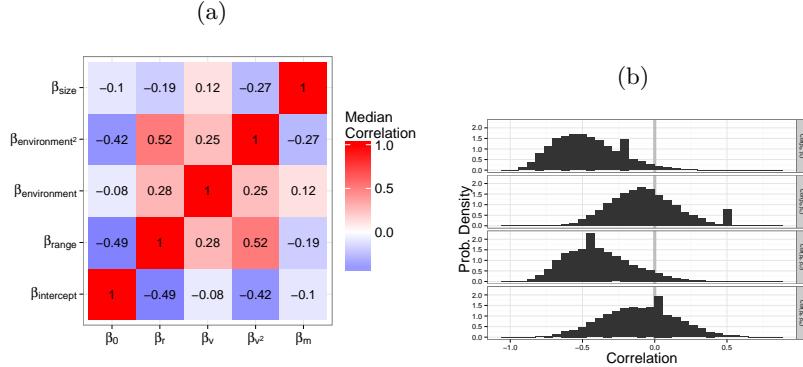


Figure 3: **A:** Heatmap for the median estimates of the terms of the correlation matrix Ω between cohort-level covariate effects. Both the exponential (left) and Weibull (right) models are presented. The off-diagonal terms are the correlation between the estimates of the cohort-level estimates of the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

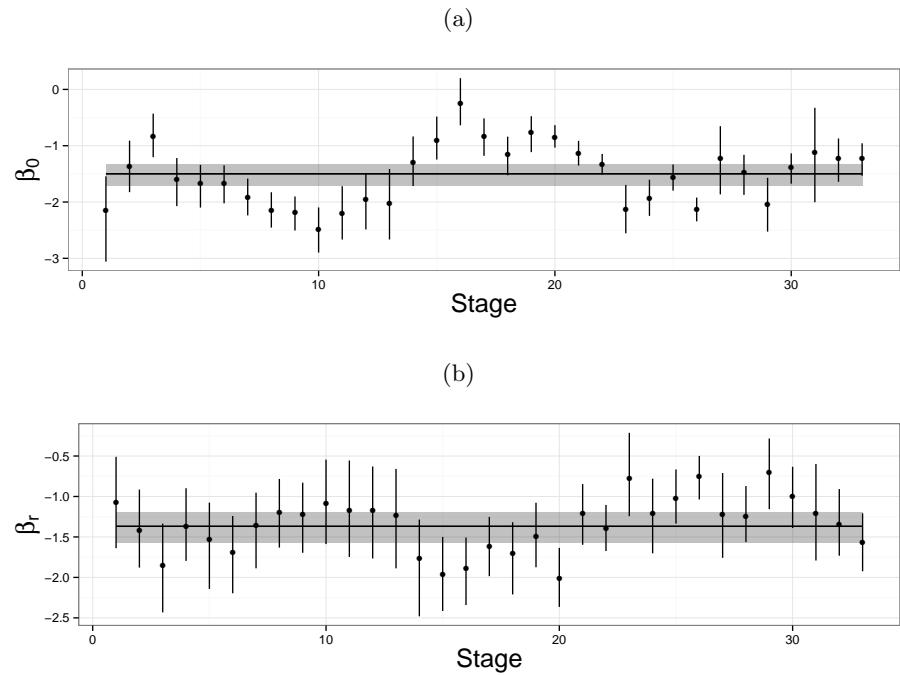


Figure 4: Comparison of cohort-specific estimates of β_0 presented along with the estimate for the overall baseline extinction risk. Points correspond to the median of the cohort-specific estimate, along with 80% credible intervals. The horizontal line is the median estimate of the overall baseline extinction 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.

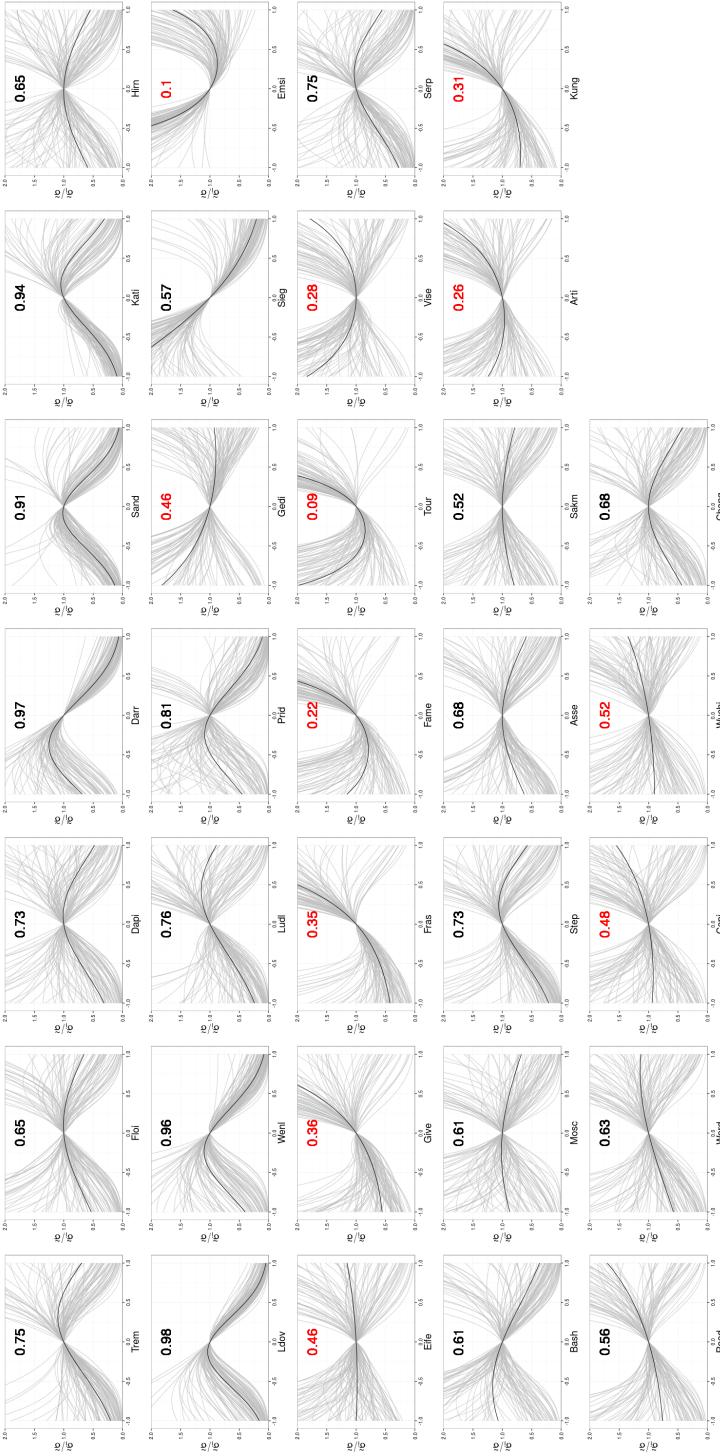


Figure 5: Comparison of the cohort-specific estimates of $f(v_i)$ (Eq. 3) for the 33 analyzed origination cohorts. The stage of origination is labeled on the x-axis of each panel. The oldest stage is in the upper left, while the youngest is in the lower left. The number in each panel corresponds to the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