Supplementary Information for ‘Detecting’ Manuscript

Justin Pomeranz1,✉, James R. Junker2,3, Vojsava Gjoni4, and Jeff S. Wesner4

14 November, 2023

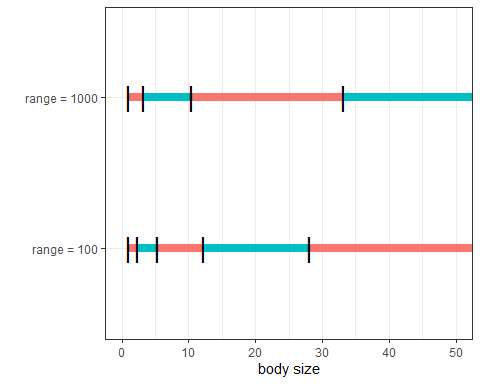
1 Colorado Mesa University, Grand Junction, CO, USA  
2 Great Lakes Research Center, Michigan Technological University, Houghton, MI USA  
3 Louisiana Universities Marine Consortium, Chauvin, LA USA  
4 Dept. of Biology, University of South Dakota, Vermillion, SD, USA

✉ Correspondence: [Justin Pomeranz <[jfpomeranz@gmail.com](mailto:jfpomeranz@gmail.com)>](mailto:jfpomeranz@gmail.com)

# Supplementary material

# ELBn bin widths

The ELBn method uses six bins of equal width on a logarithmic scale. To illustrate how the widths of these bins varies, we provide the following example when the body sizes range from 1 to 100 g vs 1 to 100g.

 Figure S1. The edges of the first 4 bins are shown when the body sizes range from 0-100 (bottom line) and 0-1000 (top line). The bins have alternating colors, and black segments are used to emphasize bin edges. As you can see, the range of the smallest bins are similar, but the edges quickly diverge in the large bin sizes. Only the first 4 bins are shown, and the image is cropped for visualization (i.e., the edges of the last bins are not shown).

# L2n bin widths

To contrast the variable bin widths, we show that the widths of the bins in the L2n method are always the same, but the number of bins changes to accomodate the range of body sizes present.

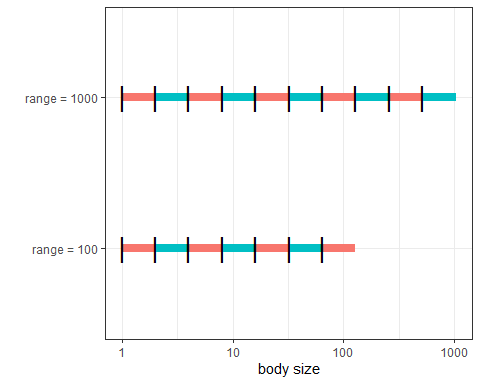
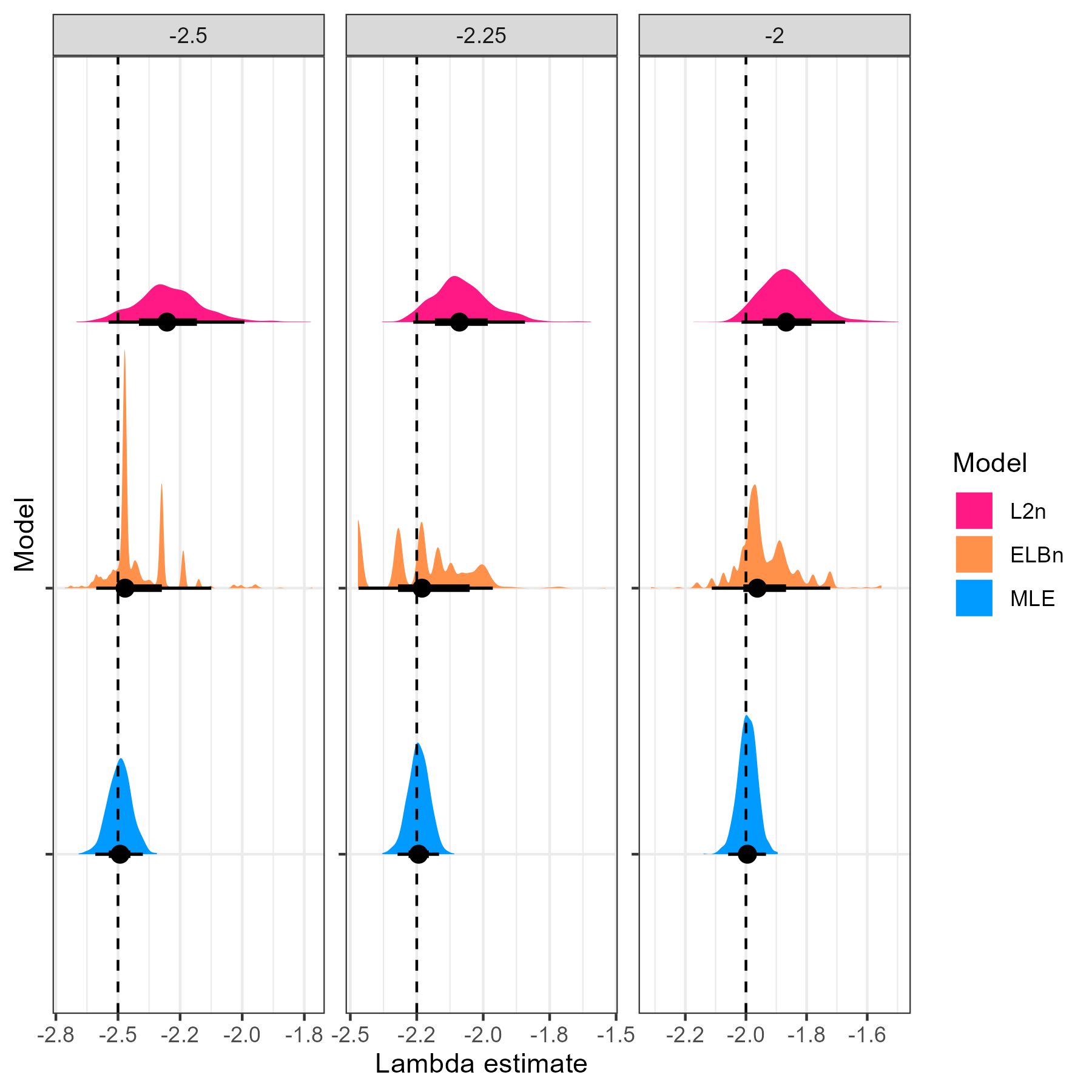


Figure S2. The edges of all bins for the L2n method are shown when body sizes range from 0-100 (bottom line) and 0-1000 (top line). The bins have alternating colors, and black segments are used to emphasize bin edges. As you can see, the bin widths are identical, but the number of bins increases with the range of body sizes in the data. The plot is presented on a log10 scale for visualization.

# Top panel of Figure 2, main text

 Figure S3. The variation in the distributions of estimates for “steep” values in Figure 2 in the main text makes it difficult to compare them with the “shallower” distirbutions. Here, we repeat just the top panel of Figure 2 from the main text to show the variation in estimates across methods. As in the other values of , MLE is the most accurate and precise. The L2N method performs slightly worse, and the ELBn method is much more variable, likely due to the fact that the presence/absence of even a single large individual could drastically alter the widths of the bins present within an individual simulation replicate (See main text).

# and relationship estimates

In the main analysis, we sampled body size values from bounded power law distributions while varying the exponent which describes the distribution. For the results presented in the main text, we held the number of body size values constant at n=999, held the number of sites at 5, scaled the environmental gradient from -1 to 1, and set the minimum and maximum body sizes to and respectively. Here, we plot the mean and standard deviation of the absolute value of deviations of the estimates from the known values for the main results, and when varying the number of body sizes sampled (n = 200, 500, 5000, 10000), varying the number of sites (3, 10), increasing the scale of the environmental gradient (-100 to 100) and decreasing the range of body sizes (min = 1, max = 100). Generally, the results reported in the main manuscript are robust to changing these parameters: the MLE estimate is nearly always closer to the known parameters, and the variation in these estimates is usually smaller than the binning methods.

## Sample size,

The number of observations in our simulations may bias the results. Therefore, we repeated the simulations described in the main, but varied the number of body size values sampled. We tested values of . All other parameters of our simulations were kept the same as the simulations presented in the main text.

All methods improved with increasing sample size (Fig. S1). However, the accuracy of the MLE method was far superior. For example, the average absolute deviation for the MLE method with only 200 body sizes values was similar to the ELBn method with 5,000 body size values and smaller than the L2n method with 10,000 body size values (Table S1).

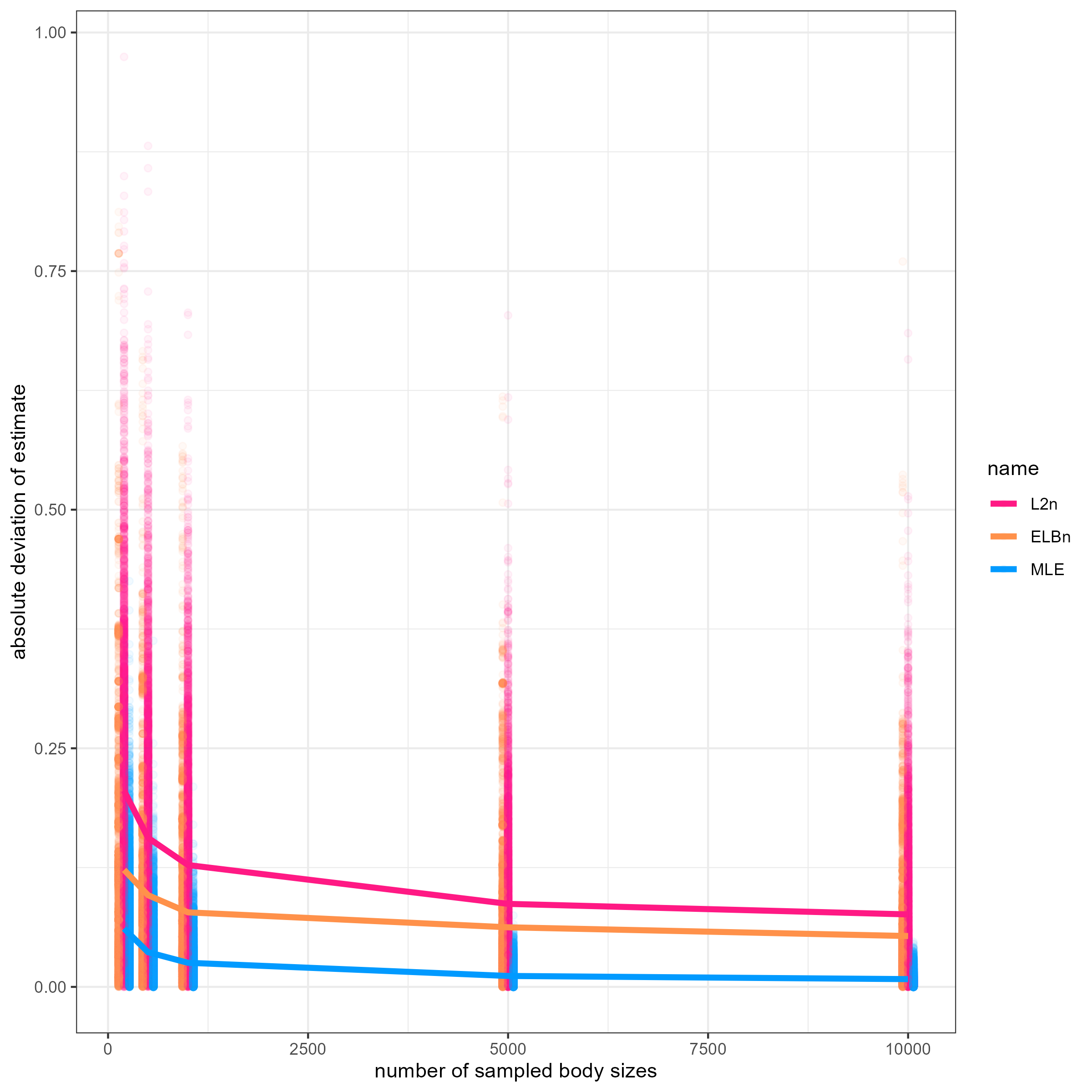


Figure S4. Absolute deviation in estimate from known value of . Points are jittered horizontally for visualization, and are deviations for each simulated replicate (n = 1000) across varying sample sizes of body size values. Lines are the average absolute deviation from the estimate. Colors indicate which method was used to estimate .

Table S1. Deviation of estimates across methods for different sample sizes of body size values. The mean and standard deviation for each row (mean abs diff and sd abs diff, respectively) in the table are calculated across all replicates for each value of lambda (1000 reps \* 5 lambda values = 5000 values.

| method | Body sample size | mean abs diff | SD abs diff |
| --- | --- | --- | --- |
| ELBn | 200 | 0.123 | 0.112 |
| ELBn | 500 | 0.096 | 0.087 |
| ELBn | 999 | 0.078 | 0.083 |
| ELBn | 5000 | 0.062 | 0.068 |
| ELBn | 10000 | 0.053 | 0.059 |
| L2n | 200 | 0.206 | 0.135 |
| L2n | 500 | 0.156 | 0.118 |
| L2n | 999 | 0.128 | 0.103 |
| L2n | 5000 | 0.087 | 0.079 |
| L2n | 10000 | 0.076 | 0.073 |
| MLE | 200 | 0.061 | 0.053 |
| MLE | 500 | 0.036 | 0.032 |
| MLE | 999 | 0.025 | 0.023 |
| MLE | 5000 | 0.011 | 0.010 |
| MLE | 10000 | 0.008 | 0.007 |

## Range of values for hypothetical environmental gradient (x) and body sizes (m)

We also wanted to ensure that our results were not sensitive to the values for the hypothetical gradient or the range of body sizes sampled. For this, we varied the values of the hypothetical gradient from -100 to 100, instead of from -1 to 1 as in the main text while keeping all other parameters of the simulation as described in the main text. We found no differences in the deviation of estimates when the hypothetical gradient was changed across any of the methods (Table S2).

When the range of body sizes sampled was reduced to be from 1 to 100, the ELBn method had slightly greater absolute deviations (0.02) from the known values of , although the variation in the estimates was reduced by ~0.016 (Table S2). In contrast, the L2n method had generally lower absolute deviations (0.05) and lower variability (0.04) with smaller body size ranges. Finally, the MLE method had slightly higher absolute deviations (0.004) and variability (0.001) when the range of body sizes was reduced, but was more similar to estimated values presented in the main results (Table S2).

Table S2. Deviation of estimates across methods when the values of the hypothetical gradient (large x) and the range of body sizes (small m) are changed compared with the results presented in the manuscript (main).

| method | id | mean abs diff | SD abs diff |
| --- | --- | --- | --- |
| ELBn | large\_x | 0.078 | 0.083 |
| ELBn | main | 0.078 | 0.083 |
| ELBn | small\_m | 0.098 | 0.067 |
| L2n | large\_x | 0.128 | 0.103 |
| L2n | main | 0.128 | 0.103 |
| L2n | small\_m | 0.077 | 0.060 |
| MLE | large\_x | 0.025 | 0.023 |
| MLE | main | 0.025 | 0.023 |
| MLE | small\_m | 0.029 | 0.024 |

## Varying number of sites

The results presented in the main text used five sites equally spaced across a hypothetical environmental gradient. We also ran the simulations using 3 and 10 sites, while keeping all other parameters as described in the main text. Generally, the average absolute deviation in estimates increased with the number of sites, whereas the variation in the deviations decreased, although the magnitude of these changes was slight (Table S3). The change in both of these metrics was the smallest for the MLE method (both = 0.001) and was approximately 0.007 to 0.012 for the absolute deviations and variation in the binning methods.

Table S3. Deviation of estimates across methods when the number of sites was changed. The results presented in the main text use 5 sites.

| method | id | mean abs diff | SD abs diff |
| --- | --- | --- | --- |
| ELBn | sites\_03 | 0.071 | 0.083 |
| ELBn | sites\_05 | 0.078 | 0.083 |
| ELBn | sites\_10 | 0.083 | 0.079 |
| L2n | sites\_03 | 0.124 | 0.114 |
| L2n | sites\_05 | 0.128 | 0.103 |
| L2n | sites\_10 | 0.131 | 0.101 |
| MLE | sites\_03 | 0.025 | 0.023 |
| MLE | sites\_05 | 0.025 | 0.023 |
| MLE | sites\_10 | 0.026 | 0.022 |

# Altering the number of replicates

The results presented in the main text for experiment 2.1 use 1000 replicates for each scenario. This number was chosen arbitrarily - we wanted a large enough number to capture the variability, but a small enough number to make repeated simulations tractable. Here, we alter the number of replicates to show how the results presented in the main text were largely invariant to this decision.

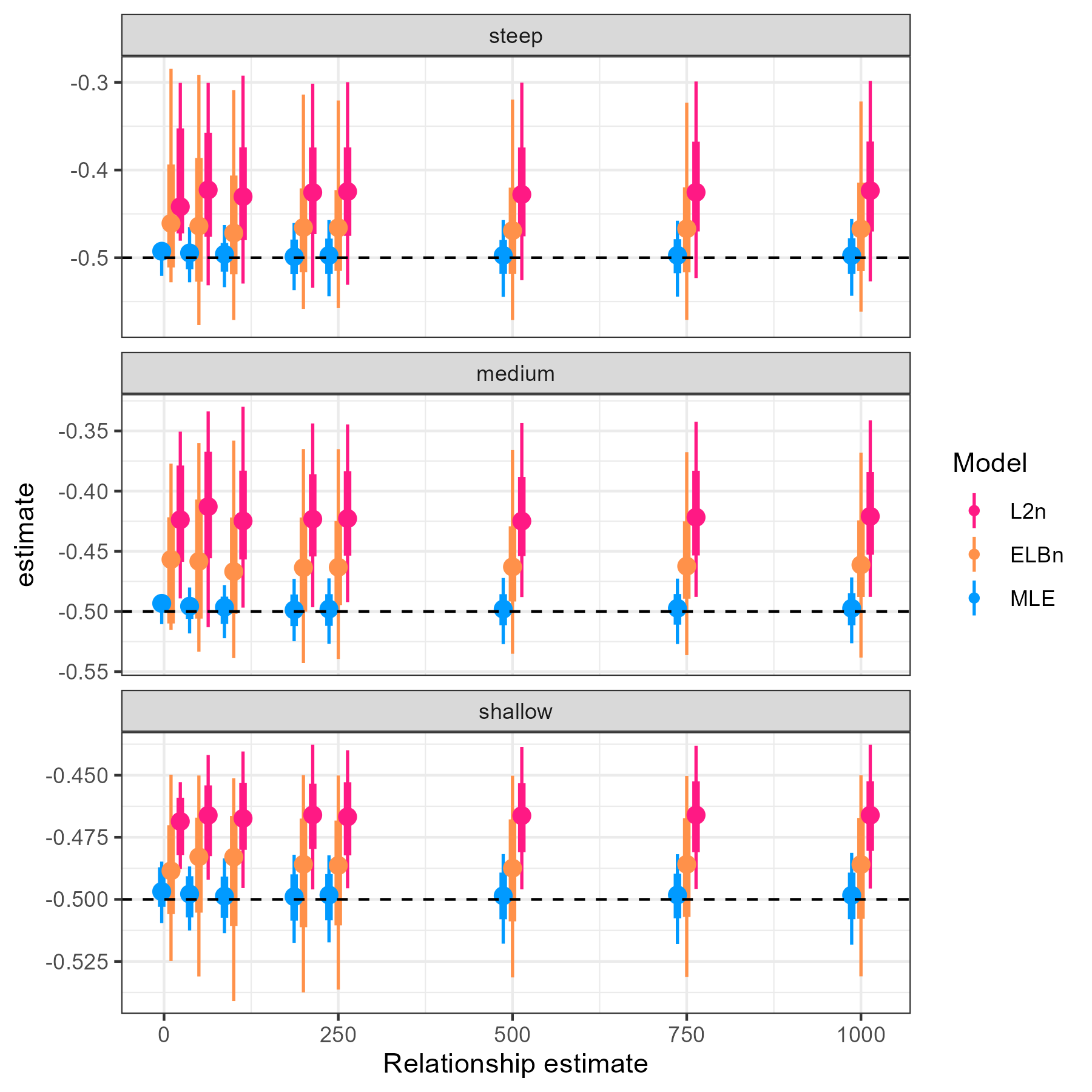


Figure S5. We varied the number of replicates (n = 10, 50, 100, 200, 250, 500, 750, 1000) and plotted the distribution of estimates across the three scenarios. The point represents the median of the distribution, and the bold and narrow bars represent the 66th and 95th percentile of the distributions, respectively. There was some variation in estimates when n < ~250, but generally the results were surprisingly robust to the number of replicates chosen. Note that the y-axis varies across rows.

# Confidence intervals

We examined the proportion of 95% Confidence Intervals which had the true value for site specific ’s (S6) as well as the CI’s for the across scenarios (S7) as well as when the effect size of varies.

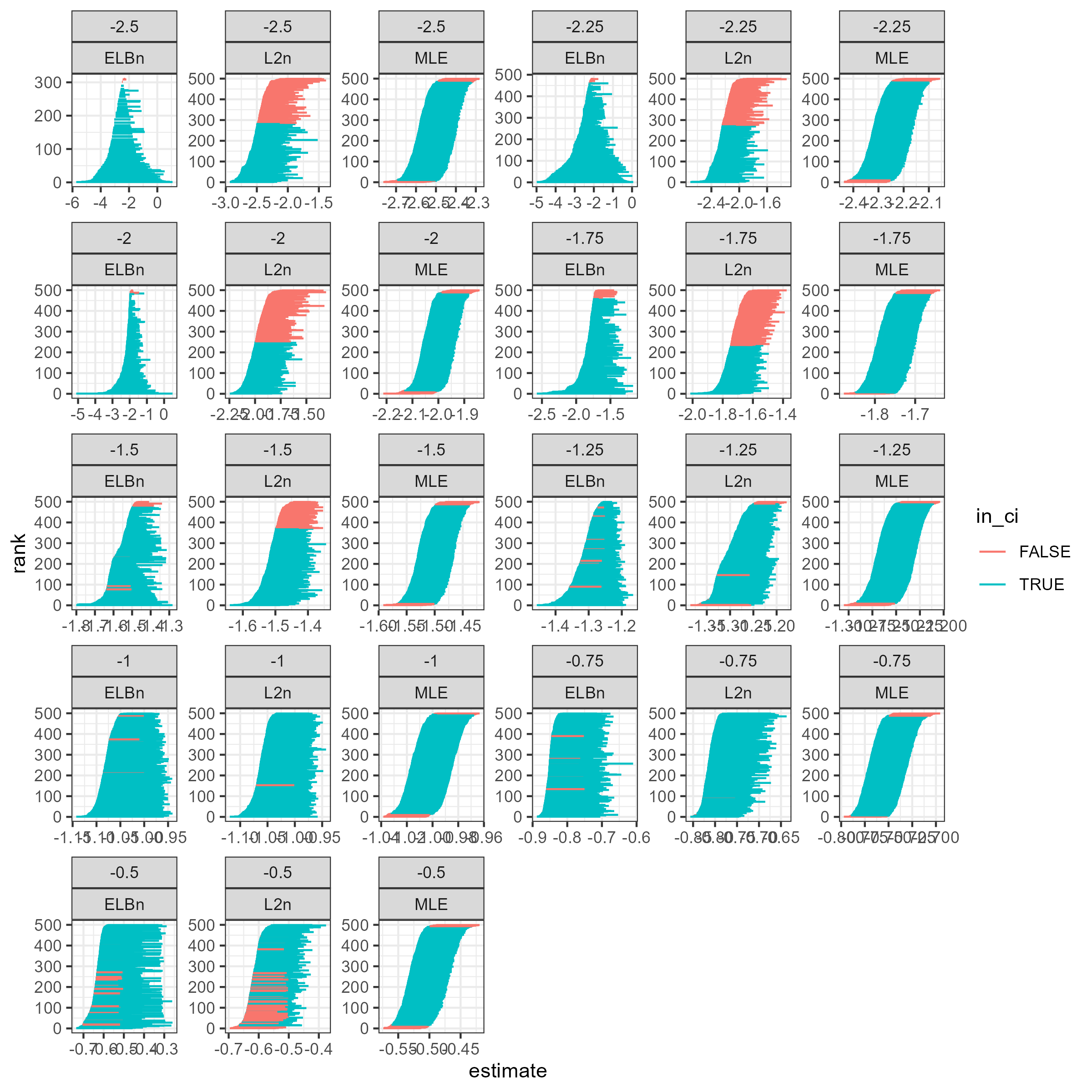


Figure S6. CI’s and estimates. CI’s are colored if they contain the true value (blue) or if they do not (red). Only the first 500 replicates are displayed for visualization purposes. They are ordered by the rank of the lower 95% CI bound

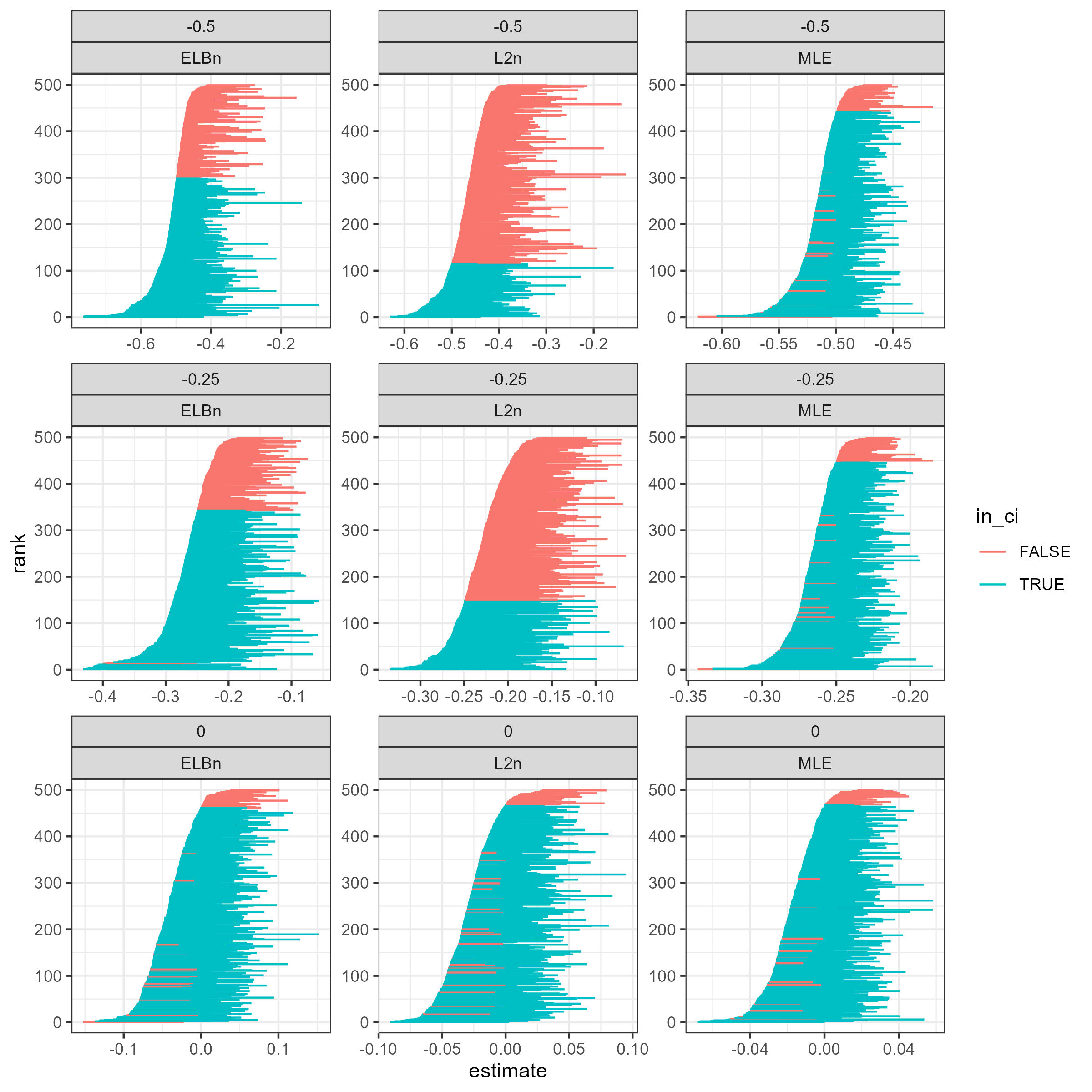


Figure S7. CI’s and estimates when varying the known value of . CI’s are colored if they contain the true value (blue) or if they do not (red). Only the first 500 replicates are displayed for visualization purposes. They are ordered by the rank of the lower 95% CI bound

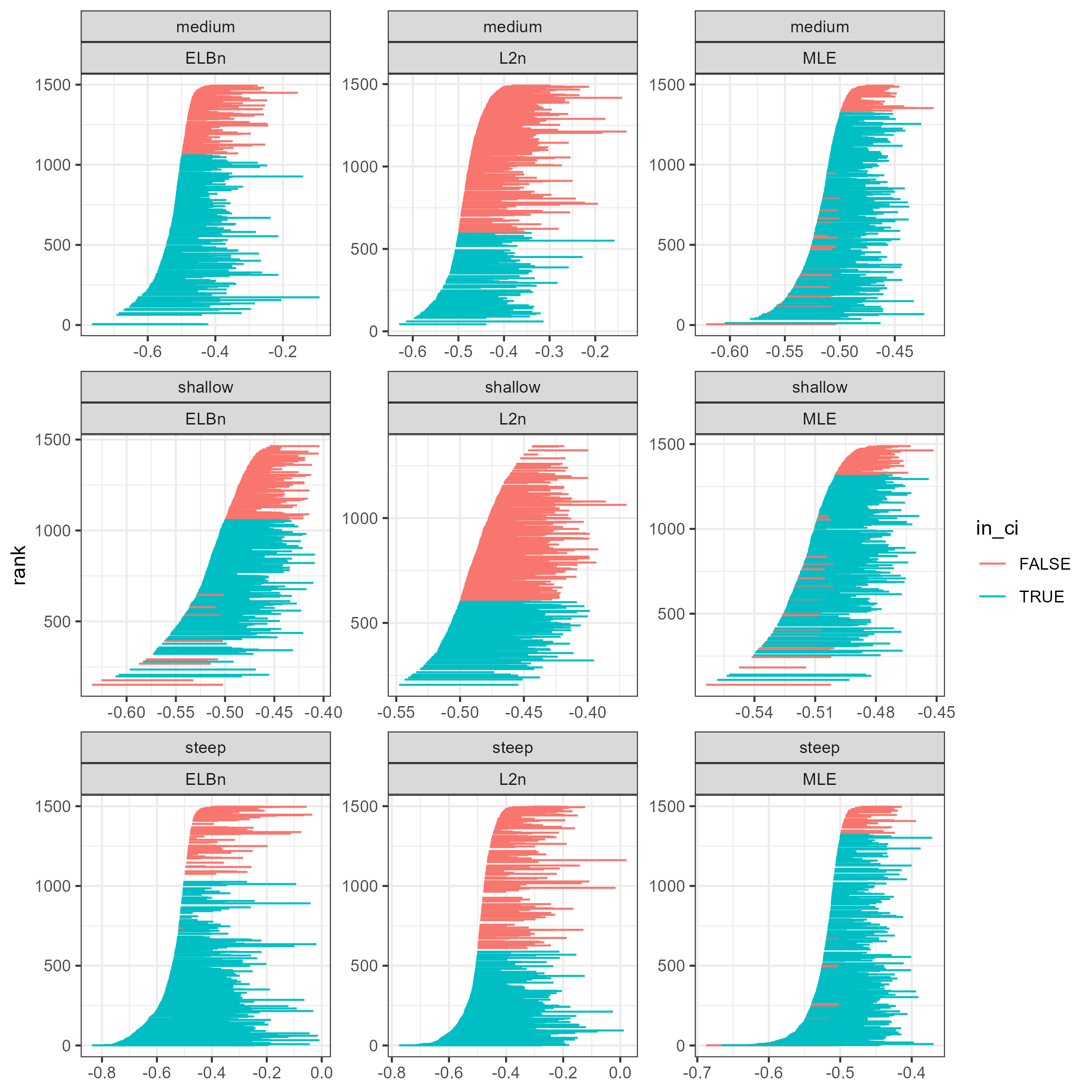


Figure S8. CI’s and estimates when varying the scenario of parameter space. CI’s are colored if they contain the true value (blue) or if they do not (red). Only the first 500 replicates are displayed for visualization purposes. They are ordered by the rank of the lower 95% CI bound