

# Prior Sensitivity for SSD Model: Random Effect Variances

## Introduction

For the proposed Spatially Selected and Dependent (SSD) model, the data model is  $[\mathbf{y}|\boldsymbol{\theta}] \sim N_n(\boldsymbol{\theta}, \mathbf{D})$  where  $\mathbf{y}$  are the direct estimates from the survey and the covariance matrix  $\mathbf{D}$  is a diagonal matrix of the design variances. The small area means are modeled with  $\boldsymbol{\theta} = \mathbf{X}^\top \boldsymbol{\beta} + \mathbf{u}$ .

Each of the random effects are modeled as  $u_i = \delta_i \cdot (v_{1i} + v_{2i})$  where  $[\mathbf{v}_1|\sigma_1^2] \sim N_n(\mathbf{0}, \sigma_1^2 \mathbf{I})$  and  $[\mathbf{v}_2|\sigma_2^2] \sim N_n(\mathbf{0}, \sigma_2^2 \mathbf{Q}^-)$  where  $\mathbf{Q}^-$  is the scaled ICAR precision matrix. The selection indicators are modeled as  $[\delta_i|p_i] \stackrel{ind}{\sim} \text{Bernoulli}(p_i)$  where  $p_i$  are the selection probabilities. We model these selection probabilities via  $\text{logit}(\mathbf{p}) = \boldsymbol{\psi}_1 + \boldsymbol{\psi}_2$  through two logit effects  $[\boldsymbol{\psi}_1|s_1^2] \sim N_n(\mathbf{0}, s_1^2 \mathbf{I})$  and  $[\boldsymbol{\psi}_2|s_2^2] \sim N_n(\mathbf{0}, s_2^2 \mathbf{Q}^-)$ . For the full details of our model, please see the main manuscript.

The parameters that require priors are the regression coefficients  $\boldsymbol{\beta}$  and the two sets of variance parameters  $\sigma_1^2, \sigma_2^2$ , and  $s_1^2, s_2^2$ . We set a non-informative proper prior for the regression coefficients:  $\boldsymbol{\beta} \sim N_j(\mathbf{0}, 100^2 \mathbf{I})$ . We use inverse-gamma priors for both sets of variance parameters. The prior choice of  $s_1^2, s_2^2$  can affect convergence of the MCMC chains but they have minimal impact on the estimates themselves, especially given that these parameters are far down the model hierarchy.

We have observed that the choice of prior on the **random effect variances**  $\sigma_1^2, \sigma_2^2$  have the biggest impact on the estimates, as they influence the inclusion probabilities  $p_i$ . Thus, our prior **sensitivity analysis focuses entirely on these random effect variances**. We first start by discussing a key feature of the inverse-gamma distribution: the gap near the origin.

## The Gap Near the Origin for Different Inverse-Gamma Priors

Let  $X$  be an inverse-gamma prior with the density:

$$f(x|c, d) = \frac{d^c}{\Gamma(c)} (1/x)^{c+1} \exp(-d/x)$$

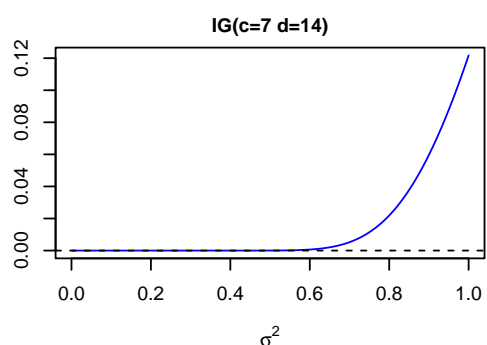
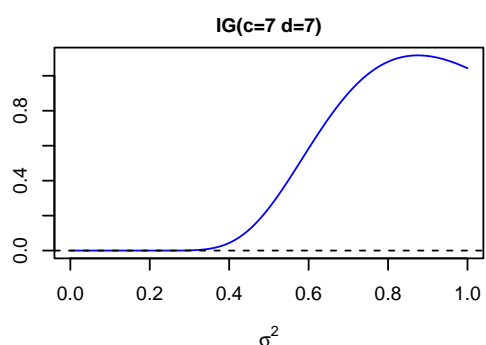
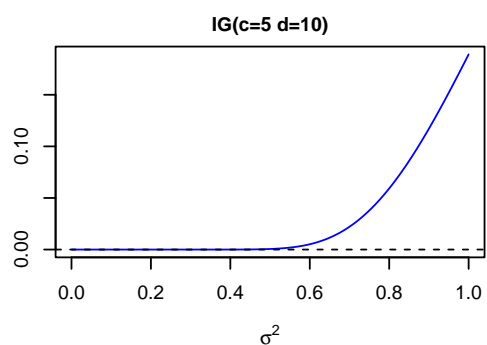
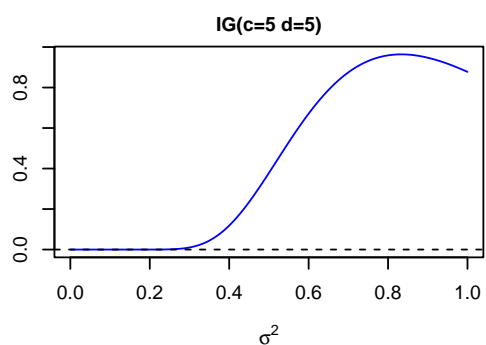
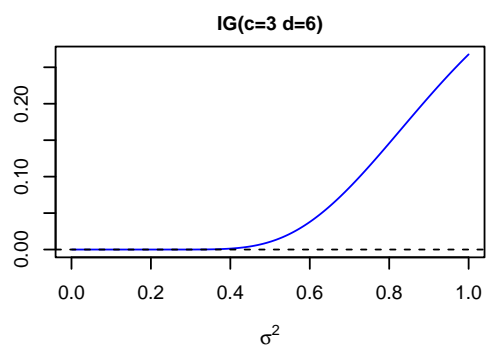
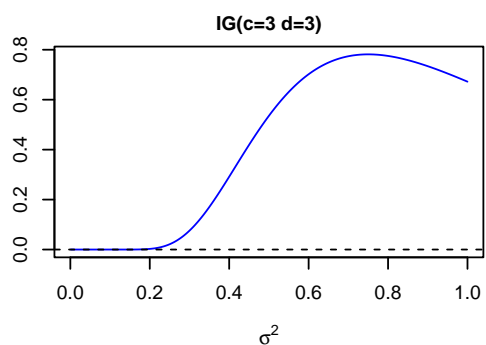
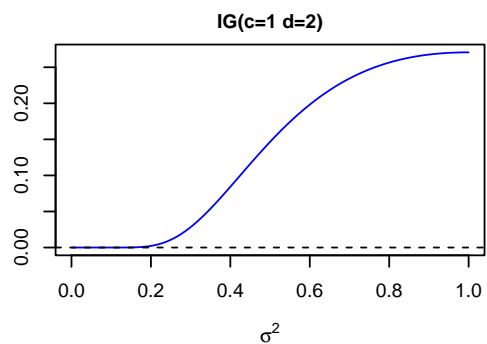
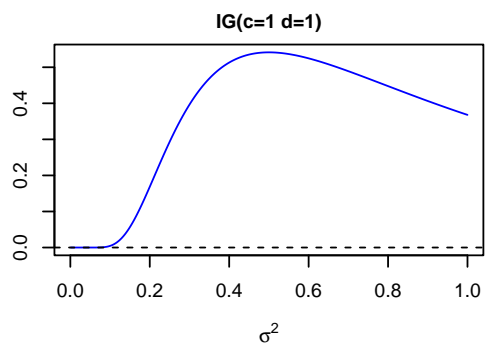
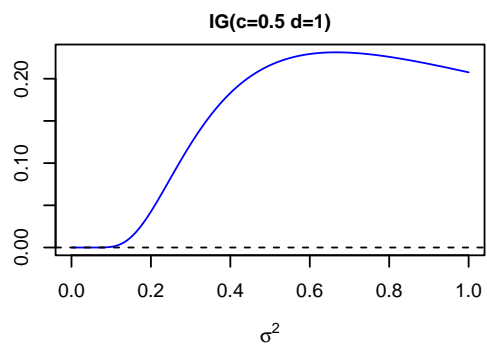
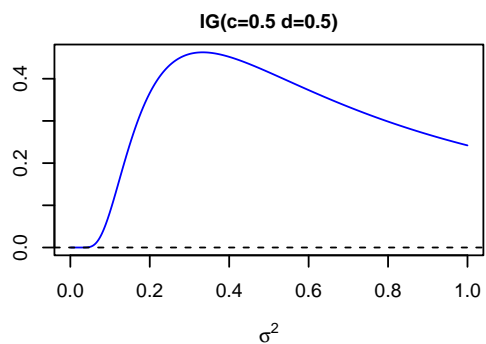
. We denote this as  $X \sim IG(c, d)$ .

For every inverse-gamma distribution, there is threshold value  $t$  at which

$$P(X < t) = \alpha$$

for some small probability  $\alpha$ . For example for  $c = 1, d = 1$ , and  $\alpha = 0.001$  the threshold  $t$  is  $\approx 0.145$

Thus an  $IG(c = 1, d = 1)$  prior has a ‘gap’ in the density between zero and 0.145. When considering inverse-gamma priors for the random effects variance parameter models like the Datta-Mandal or the SSD model, this gap is important for to make sure that the “spike” and “slab” are distinguishable. Below are the density plots different inverse-gamma priors we compare in this analysis. Each plot is zoomed in to illustrate the difference in the gap near the origin.



For the SSD model, we would like a prior that is non-informative but leaves enough space around the origin. Setting the inverse-gamma hyperparameters  $IG(c, d)$  smaller results in a non-informative prior. However, doing so results in a very small threshold and not enough gap near the origin. So what threshold is optimal to balance these two factors? In our experience, we found what makes a suitable prior is dependent on the scale of the data. This is why we recommend scaling  $(y_i - \bar{y})/s_{\mathbf{y}}$  where  $\bar{y}, s_{\mathbf{y}}$  is the sample mean and standard deviation of the direct estimates to aid prior specification. Below are the different  $t$  values at  $\alpha = 0.001$  for different inverse-gamma priors. A larger value of  $t$  indicates more gap near the origin.

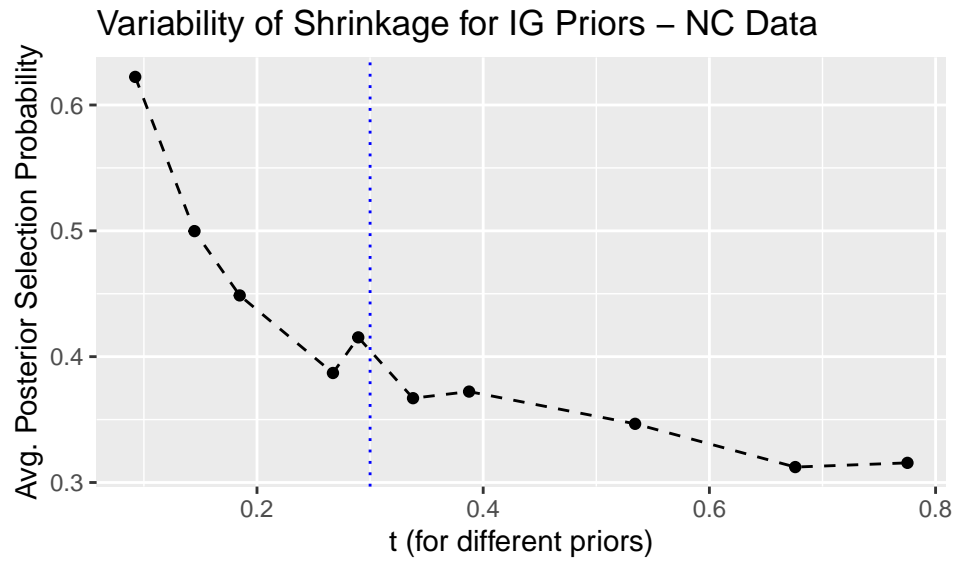
c (shape)	d (scale)	t : P(X<t)=0.001
0.5	0.5	0.0923366
1.0	1.0	0.1447512
0.5	1.0	0.1847212
3.0	3.0	0.2671574
1.0	2.0	0.2895404
5.0	5.0	0.3379632
7.0	7.0	0.3875382
3.0	6.0	0.5343250
5.0	10.0	0.6759267
7.0	14.0	0.7751258

## Impact on Shrinkage

The **selection probabilities**  $p_i$  are key parameters to assess *shrinkage* in the random effects; it tells us the probability that a random effect is needed in area  $i$ . To get a sense of the overall level of shrinkage, we examine the average selection probability  $\bar{p}$  across all areas. Below is a table of the different  $t$  and average *posterior* average selection probabilities resulting from the  $IG(c, d)$  priors for the rent burden dataset in North Carolina (used for the paper). We can see that higher  $t$  leads to lower selection probabilities (and higher shrinkage).

c (shape)	d (scale)	t : P(X<t)=0.001	Avg Selection Probability
0.5	0.5	0.0923366	0.6223148
1.0	1.0	0.1447512	0.4997786
0.5	1.0	0.1847212	0.4486652
3.0	3.0	0.2671574	0.3870384
1.0	2.0	0.2895404	0.4153408
5.0	5.0	0.3379632	0.3669949
7.0	7.0	0.3875382	0.3722475
3.0	6.0	0.5343250	0.3466417
5.0	10.0	0.6759267	0.3122663
7.0	14.0	0.7751258	0.3156045

Here we show the  $t$  values plotted against the average *posterior* selection probabilities. We can see that there is larger values of  $t$  lead to more random effects being reduced to zero (lower selection probability).

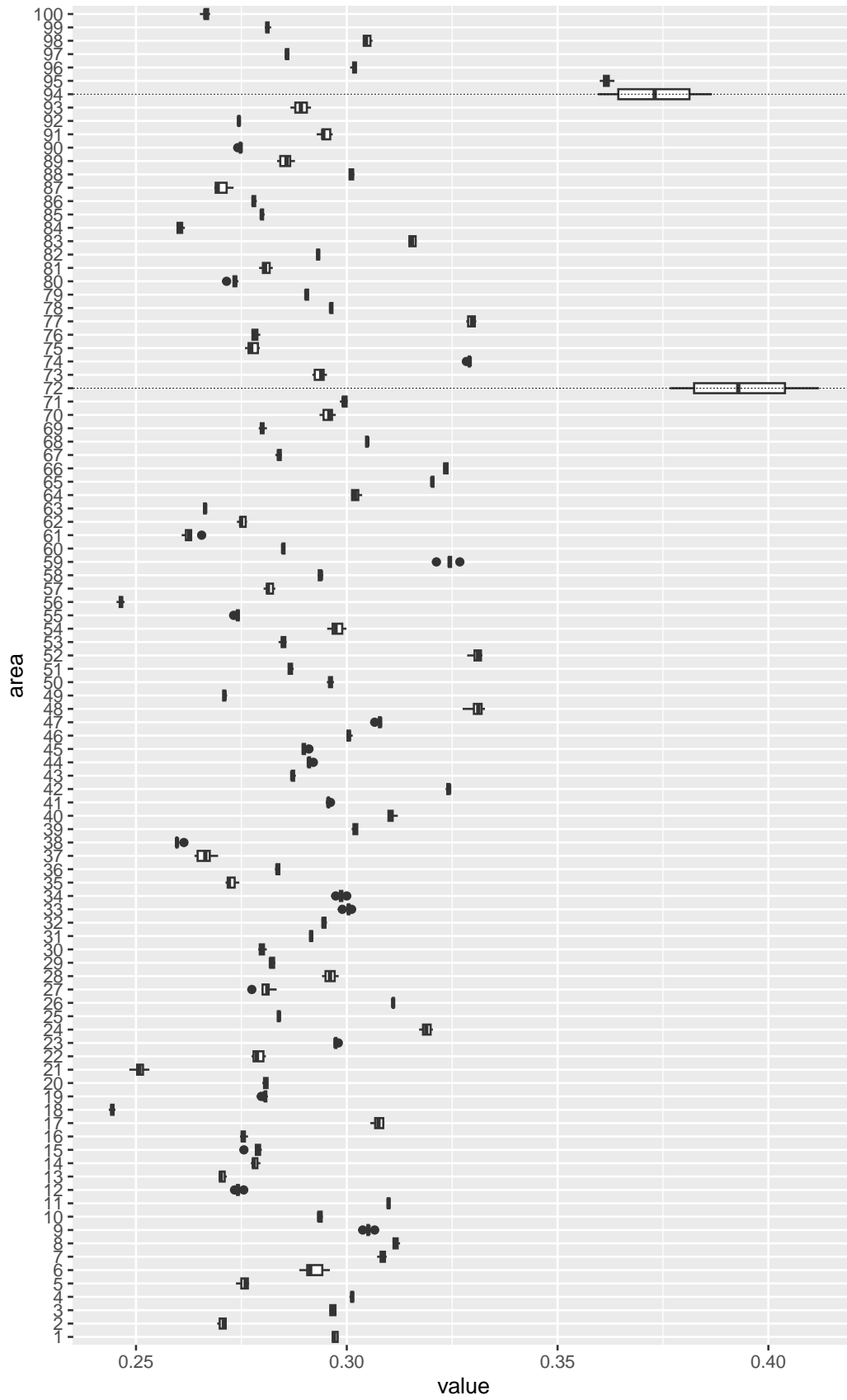


Note that the drop in average selection probability is quite large for  $t < 0.3$  but there is an inflection point near 0.3. Thus we recommend any Inverse-Gamma prior with  $t > 0.3$  with  $\alpha = 0.001$ .

## Impact on Estimates

Here we show the variability of the point estimates for Inverse-Gamma priors with  $t > 0.3$ .

Variability of Point Estimates for IG Priors ( $t > 0.3$ ) – NC Data



We can see that the prior choice has minimal impact outside of a few areas. There are two areas with some variability (areas 72, 94). These correspond to estimates for Perquimans and Washington counties that have the highest design variance. Thus the prior choice has a significant impact on the estimates for these counties. Note that this is not a problem for all datasets, as seen in the next example.

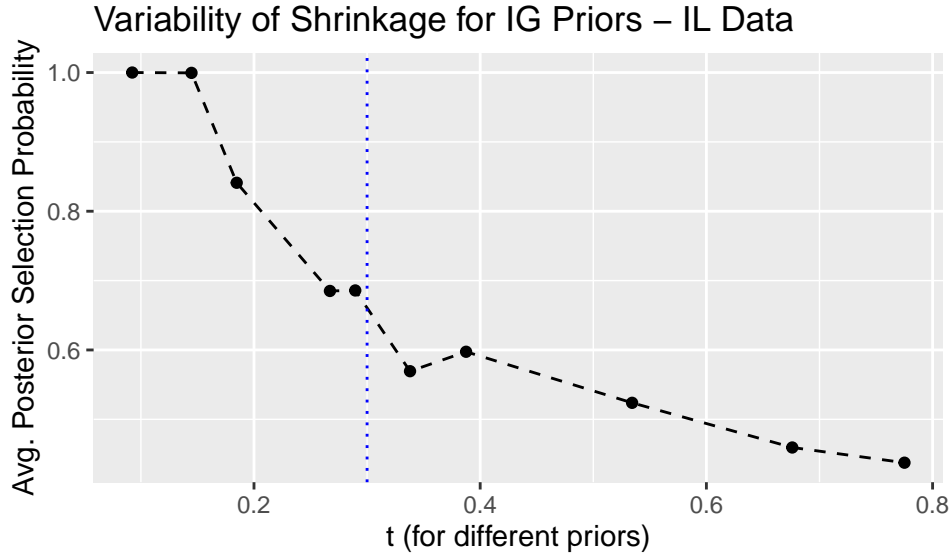
## Repeat with different data (Illinois)

We repeat this exercise again with Illinois  $n = 102$  and estimating poverty rates.

Here is a table of the different  $t$  and average *posterior* selection probabilities resulting from the  $IG(c, d)$  priors for another dataset from Illinois, where we estimate poverty rates.

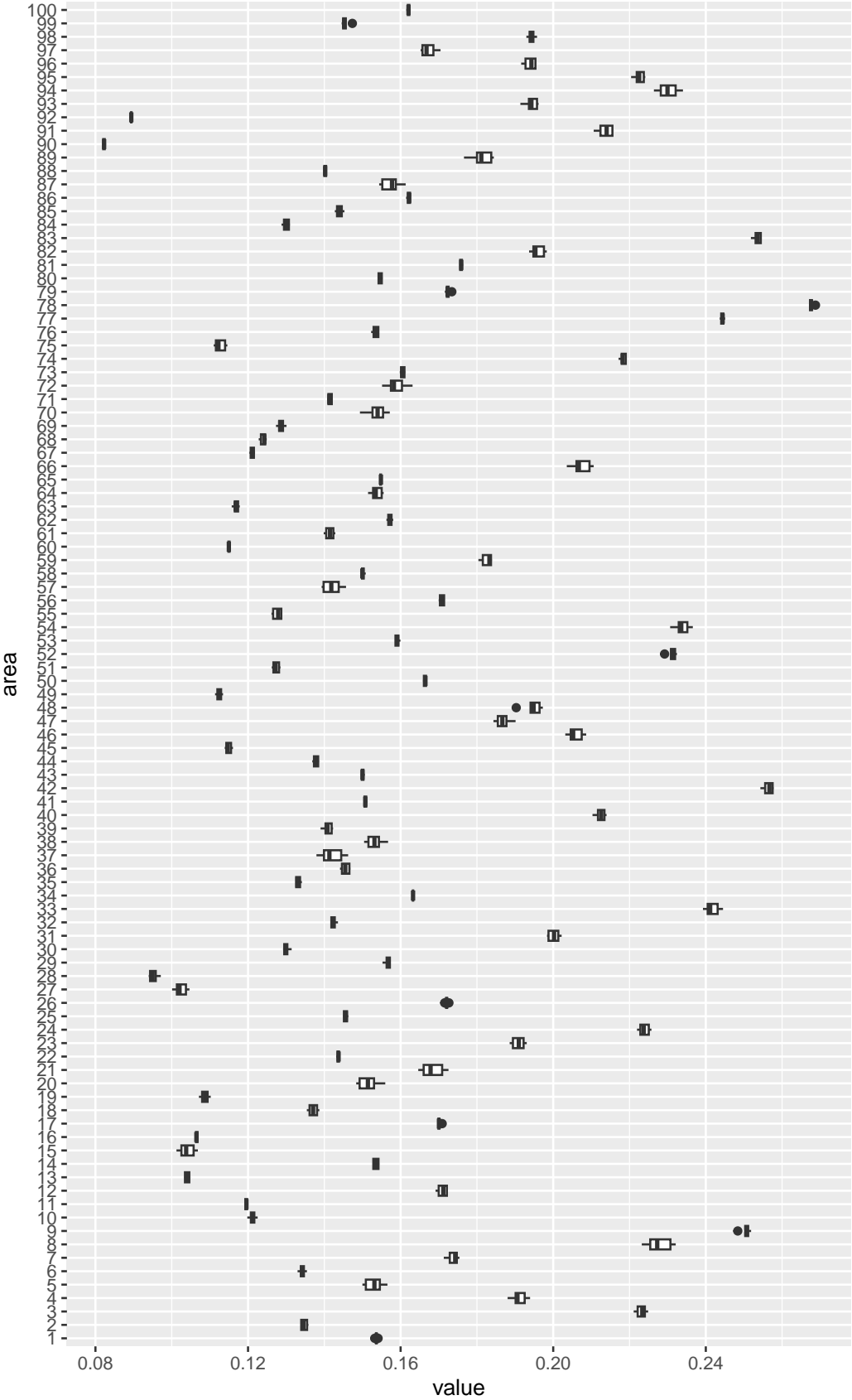
c (shape)	d (scale)	t : $P(X < t) = 0.001$	Avg Selection Probability
0.5	0.5	0.0923366	1.0000000
1.0	1.0	0.1447512	0.9995303
0.5	1.0	0.1847212	0.8410624
3.0	3.0	0.2671574	0.6850771
1.0	2.0	0.2895404	0.6858159
5.0	5.0	0.3379632	0.5693626
7.0	7.0	0.3875382	0.5973950
3.0	6.0	0.5343250	0.5237839
5.0	10.0	0.6759267	0.4593800
7.0	14.0	0.7751258	0.4374525

Here we show the  $t$  values plotted against the average *posterior* selection probabilities. Note that priors with smaller values of  $t$  result in very little to no shrinkage (almost all random effects are included). Again, we see that selection probabilities generally decrease as  $t$  increases. Finally we observe that priors with  $t > 0.3$  again results in a smaller range of *posterior* selection probabilities.



Here we show the variability of the point estimates for Inverse-Gamma priors with  $t > 0.3$ .

Variability of Point Estimates for IG Priors ( $t>0.3$ ) – IL Data



We can see that for inverse-gamma priors with  $t > 0.3$ , the estimates are barely different from one another for all areas. For this dataset, the prior choice has minimal impact on the estimates for inverse-gamma priors with  $t > 0.3$ .

## Conclusion & Recommendation

We conducted a prior sensitivity analysis for the random effect variance parameters for the SSD model. The prior choice of these parameters can influence the level of shrinkage and the value of estimates, especially in areas with high design variance.

We discussed that there for every inverse-gamma distribution, there is value  $t$  at which  $P(X < t) = \alpha$  for some small probability  $\alpha$ . We showed the differing values for  $t$  at  $\alpha = 0.001$  for different inverse-gamma priors. To aid prior specification, we recommended that the data be scaled  $(y_i - \bar{y})/s_y$  where  $\bar{y}, s_y$  is the sample mean and sample standard deviation respectively.

We tested the model on two different datasets. We saw that choosing an inverse-gamma prior with  $t$  value that is too small may force the random effects for most areas to be selected. On the other hand, having  $t$  that is very large can force higher shrinkage and lower selection probabilities. We recommend choosing a inverse-gamma prior with  $t > 0.3$  for  $\alpha = 0.001$  that balances both of these factors. For the work done for our main manuscript, we chose the prior  $IG(5, 5)$  where  $t \approx 0.34$ .