SuSiE L Sensitivity

Stuart Brabbs

6/1/2021

Our goal is to test the sensitivity of the SuSiE algorithm to the pre-specified number of single-effect vectors L. We do this using the minimal example included in susieR package documentation, with n = 1000 and p = 1000. We set 10 nonzero signals (all equal to 1).

```
set.seed(1)
n <- 1000
p <- 1000
beta <- rep(0,p)
beta[c(1,2,220,300,400,620,750,900,905,990)] <- 1
X <- matrix(rnorm(n*p), nrow=n, ncol=p)
y <- X %*% beta + rnorm(n)
res1 <- susie(X, y, L=10)
```

```
res2 <- susie(X, y, L=20)
```

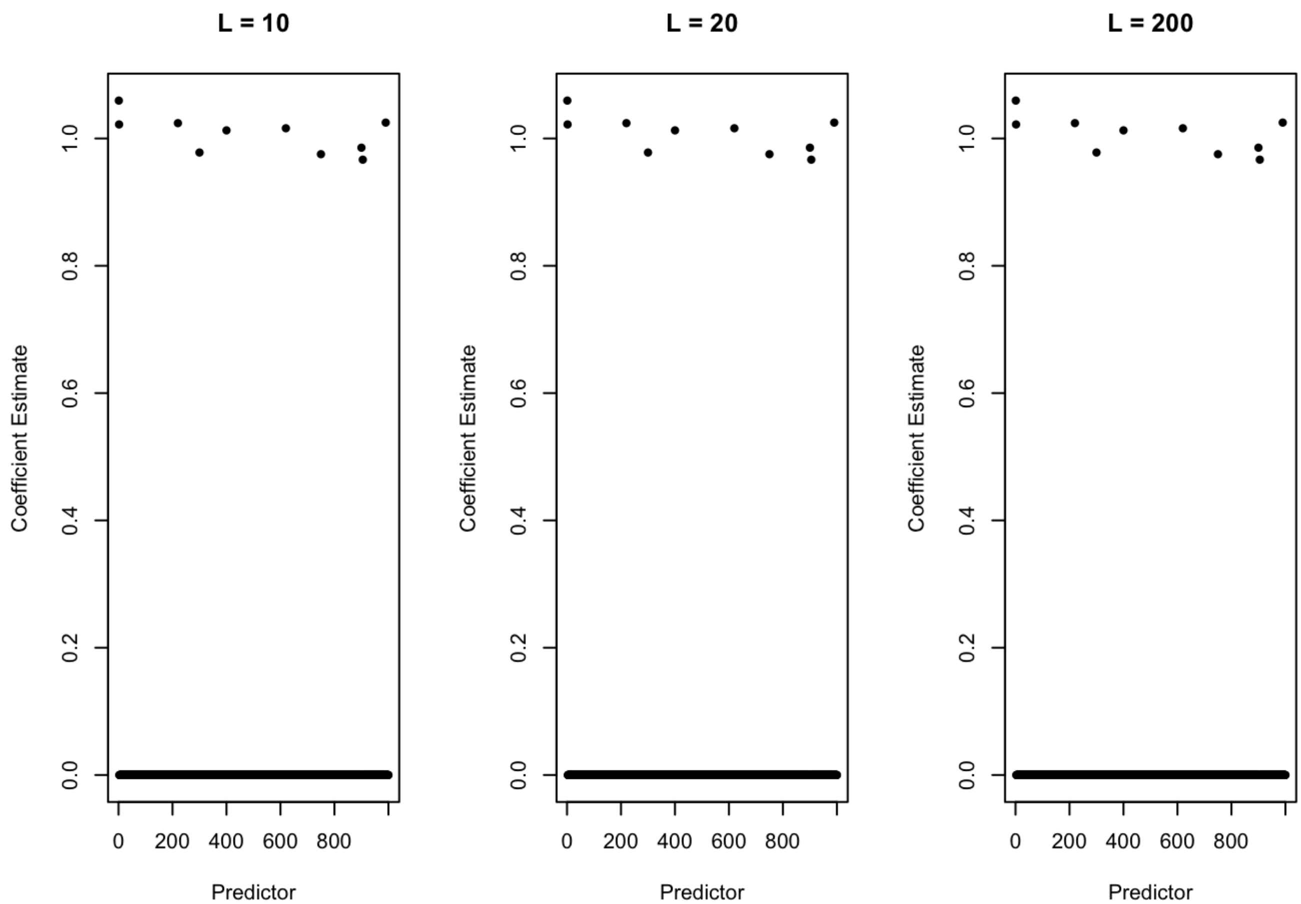
```
res3 <- susie(X, y, L = 200)
```

```
res4 <- susie(X, y, L = 8)
```

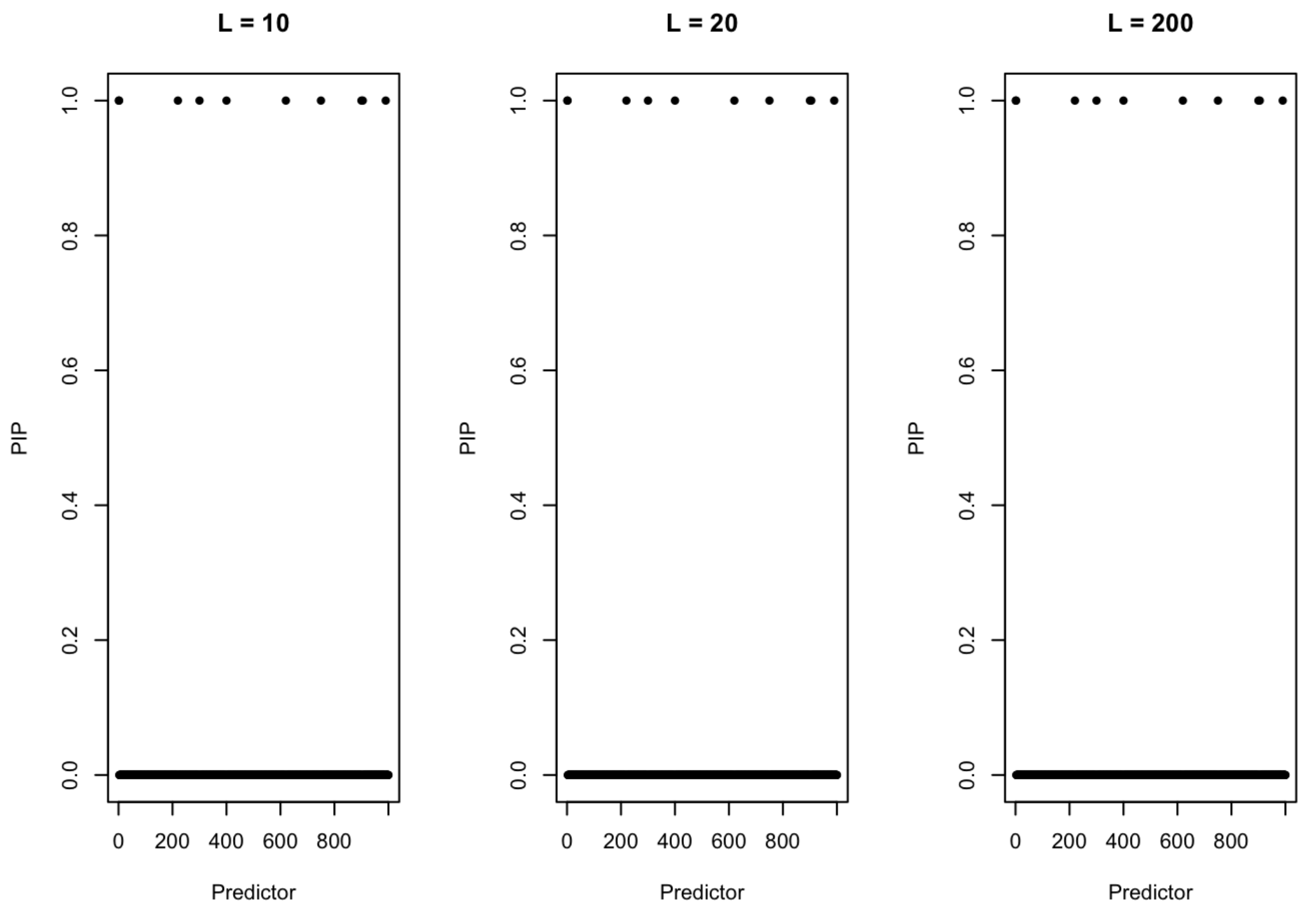
```
res5 <- susie(X, y, L = 4)
```

We first plot the coefficient estimates and posterior inclusion probabilities (PIPs) to compare increasing L:

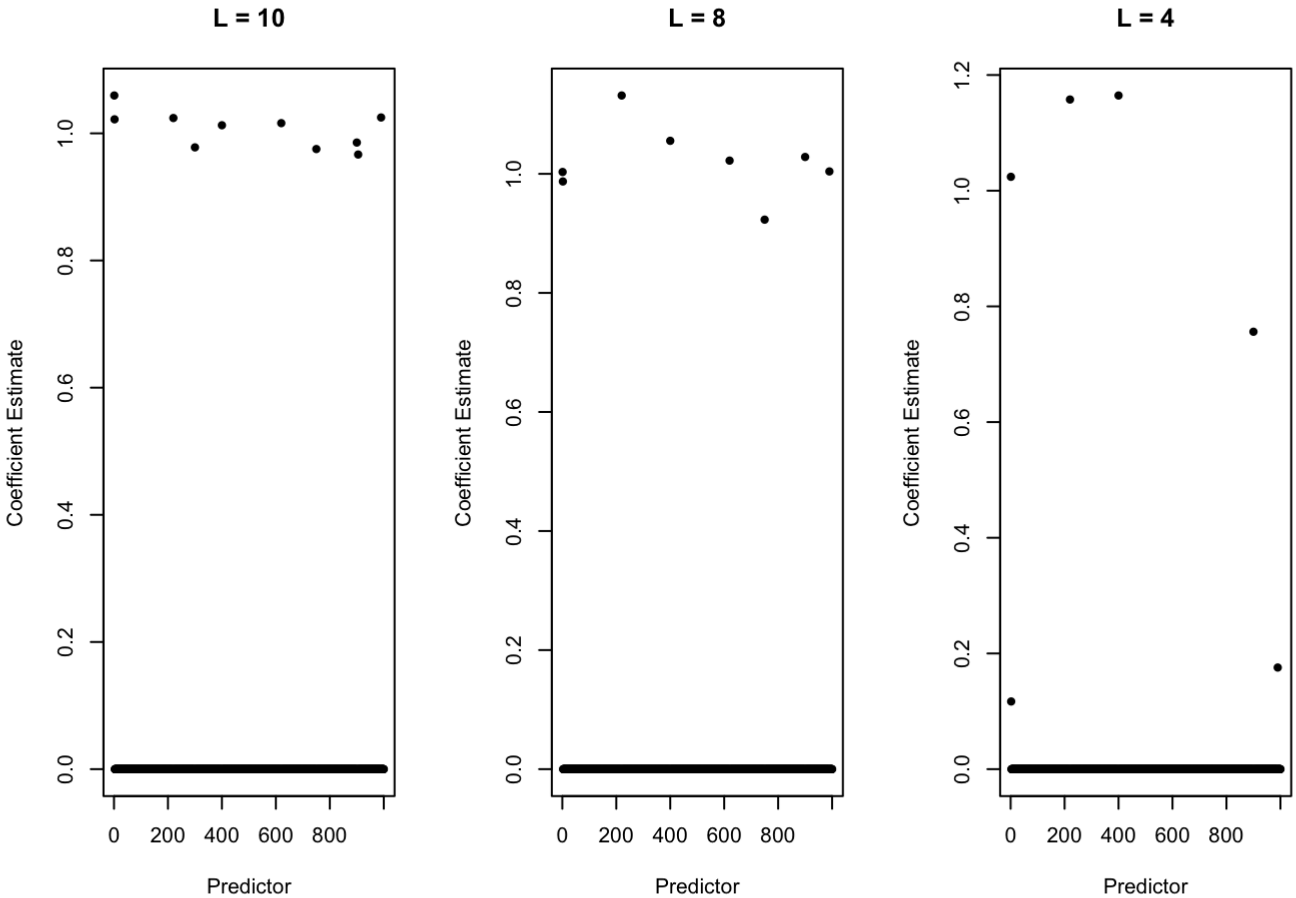
```
par(mfrow = c(1,3))
plot(coef(res1)[-1], pch = 20, xlab = "Predictor", ylab = "Coefficient Estimate", main =
"L = 10")
plot(coef(res2)[-1], pch = 20, xlab = "Predictor", ylab = "Coefficient Estimate", main =
"L = 20")
plot(coef(res3)[-1], pch = 20, xlab = "Predictor", ylab = "Coefficient Estimate", main =
"L = 200")
```



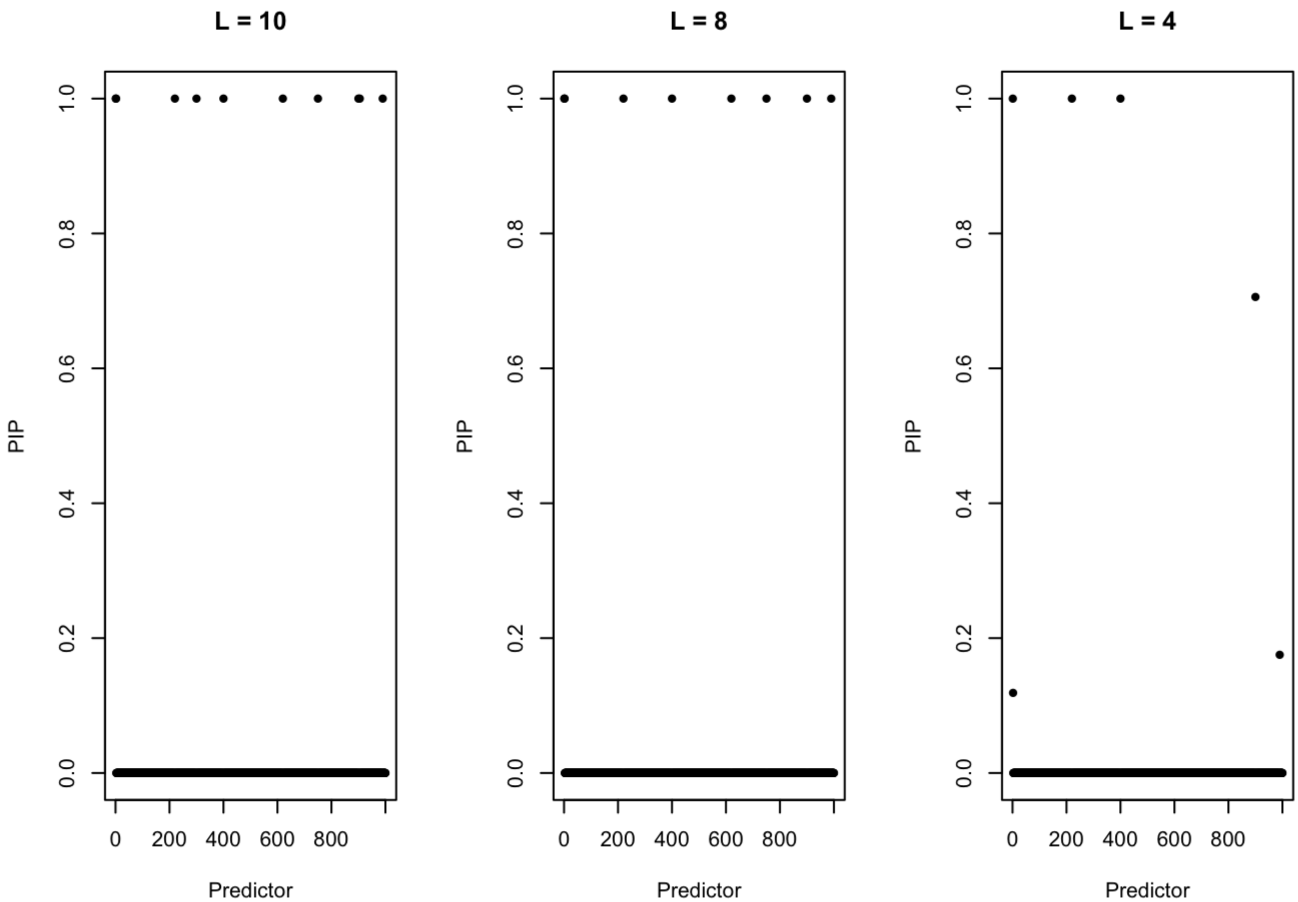

```
par(mfrow = c(1,3))
plot(res1$pip, pch = 20, xlab = "Predictor", ylab = "PIP", main = "L = 10")
plot(res2$pip, pch = 20, xlab = "Predictor", ylab = "PIP", main = "L = 20")
plot(res3$pip, pch = 20, xlab = "Predictor", ylab = "PIP", main = "L = 200")
```




```
par(mfrow = c(1,3))
plot(coef(res1)[-1], pch = 20, xlab = "Predictor", ylab = "Coefficient Estimate", main =
"L = 10")
plot(coef(res4)[-1], pch = 20, xlab = "Predictor", ylab = "Coefficient Estimate", main =
"L = 8")
plot(coef(res5)[-1], pch = 20, xlab = "Predictor", ylab = "Coefficient Estimate", main =
"L = 4")
```




```
par(mfrow = c(1,3))
plot(res1$pip, pch = 20, xlab = "Predictor", ylab = "PIP", main = "L = 10")
plot(res4$pip, pch = 20, xlab = "Predictor", ylab = "PIP", main = "L = 8")
plot(res5$pip, pch = 20, xlab = "Predictor", ylab = "PIP", main = "L = 4")
```



The authors' assertion that overstating L has little effect appears correct - the plots for L = 10, 20, and 200 are essentially the same, though the L = 200 case takes considerably longer to complete. When understating L, however, the outcome changes considerably. The signals with the largest estimated coefficients at L = 10 remain largely intact with lower L, but others are significantly dampened. When only slightly understated, at L = 8, the effects are not too dramatic, and most estimates and PIPs remain near their values at L = 10. For a much lower value of L = 4, however, most estimates and PIPs are pulled downward significantly.