STAT600 HW1

Muyang Shi

2024-01-25

Problem 2

To install my package, please download the `SimpLin_1.0.tar.gz` file, github link here: https://github.com/muyangshi/STAT600Homework/tree/main/HW1.

Problem 3

Simulate Dataset

```r
set.seed(1)
Dataset100 <- foreach(i = 1:100) %do% {
  x <- rnorm(100)
  e <- rnorm(100)
  y <- 1 - x + e
  list(x = x, y = y)
}
```

Base R `lm()`

```r
# Base lm()
lm_base <- foreach(i = 1:100) %dopar% {
  x <- Dataset100[[i]]$x
  y <- Dataset100[[i]]$y
  elapsed <- system.time({mod.base <- lm(y~x)})[3]
  coef <- c(mod.base$coefficients)
  beta0_ci <- confint(mod.base)[1,]
  beta1_ci <- confint(mod.base)[2,]
  yhat <- mod.base$fitted.values
  resids <- mod.base$residuals
  list(time=elapsed,coef=coef,beta0_ci=beta0_ci,beta1_ci=beta1_ci,yhat=yhat,resids=resids)
}
stopImplicitCluster()
```

SimpLinCpp

```
lm_cpp <- foreach(i = 1:100) %dopar% {
  x <- Dataset100[[i]]$x
  y <- Dataset100[[i]]$y
  elapsed <- system.time({mod.cpp <- SimpLinR(x,y)})[3]
  coef <- c(mod.cpp$Coefficients)
  beta0_ci <- mod.cpp$`95% CI`[1,]
  beta1_ci <- mod.cpp$`95% CI`[2,]
  yhat <- mod.cpp$`Predicted Values`
  resids <- mod.cpp$Residuals
  list(time=elapsed,coef=coef,beta0_ci=beta0_ci,beta1_ci=beta1_ci,yhat=yhat,resids=resids)
}
stopImplicitCluster()
```

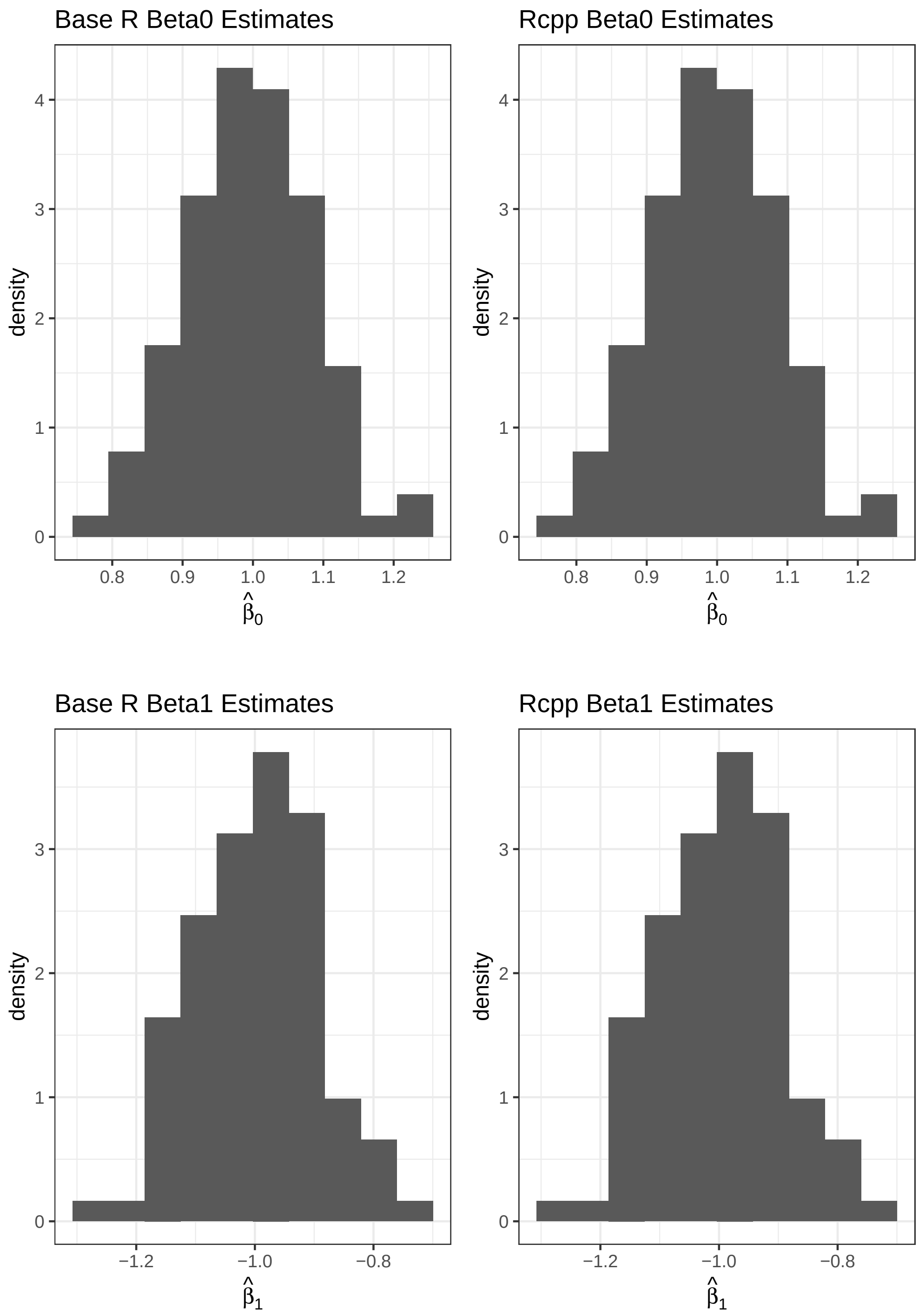
Problem 4

Both the base R `lm()` and the Cpp `SimpLin` produce same result (in terms of bias, coverage probability, mean squared error for the regression coefficients as well as predictive mse for $\hat{y}$) like they should be. Time-wise, the cpp is faster than the base R `lm()`.

A table comparing the two methods' performance is included below, as well as histogram of the estimated coefficients.

Table 1: Performace Comparison between Base R and Rcpp

	Base R	Rcpp
Avg Time (secs)	0.0007500	0.0002200
Bias: b0	-0.0066561	-0.0066561
Bias: b1	0.0013153	0.0013153
Empirical Coverage: b0	0.9600000	0.9600000
Empirical Coverage: b1	0.9500000	0.9500000
MSE: b0	0.0076562	0.0076562
MSE: b1	0.0104577	0.0104577
Prediction MSE	0.9663379	0.9663379



Base R Beta0 Estimates

Rcpp Beta0 Estimates

Base R Beta1 Estimates

Rcpp Beta1 Estimates