**MSiA 400**

**Lab Assignment 3**

**Problem 1**

**a.** It is difficult to sample points from complex distributions with simply Monte Carlo (for example the Poisson posterior distribution with regressors). Markov Chain Monte Carlo’s (MCMC) purpose is to overcome this difficulty and sample points from complex distribution. MCMC accomplishes this by constructing a Markov Chain stationary distribution that is exactly equal to the desired distribution.

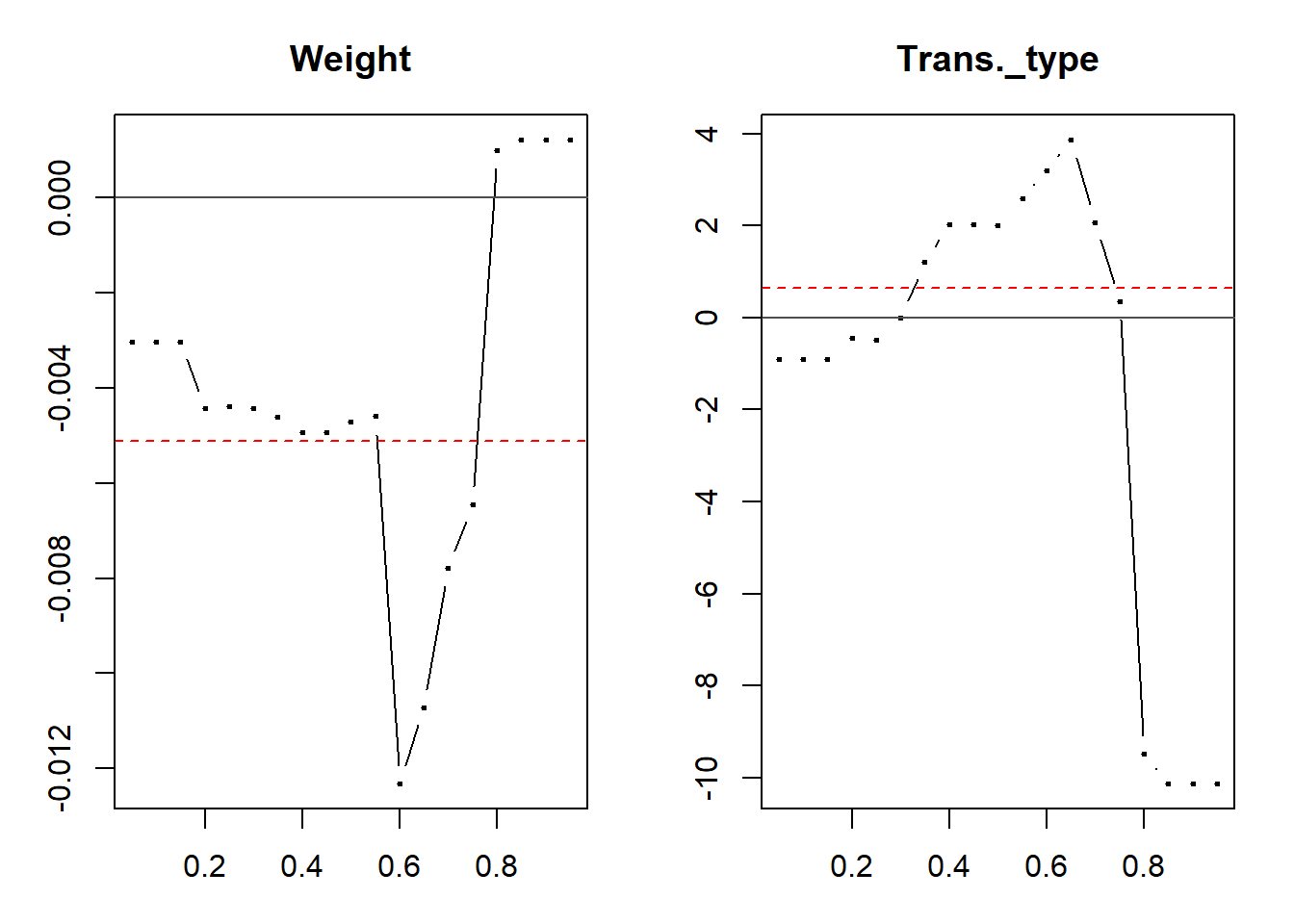
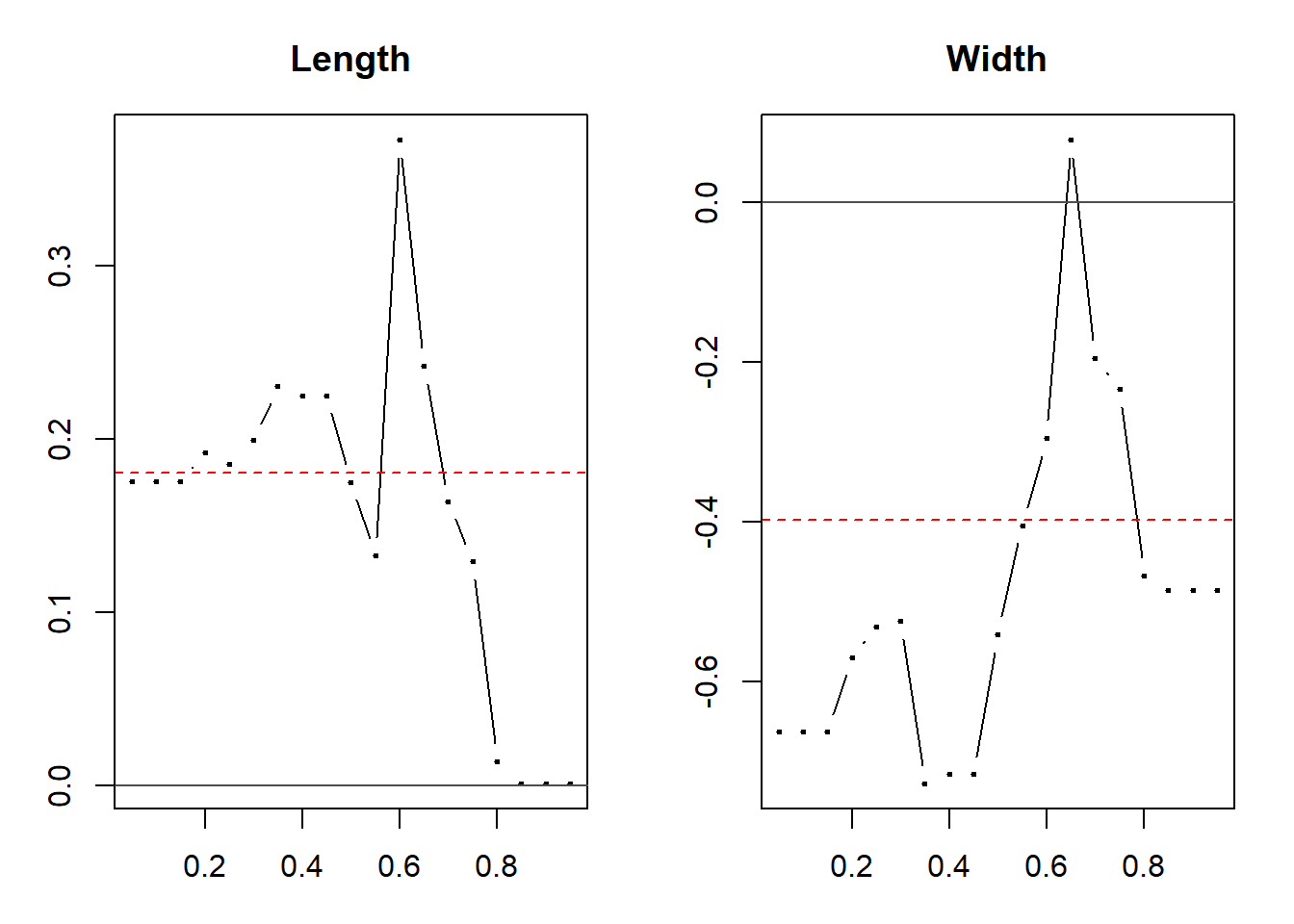
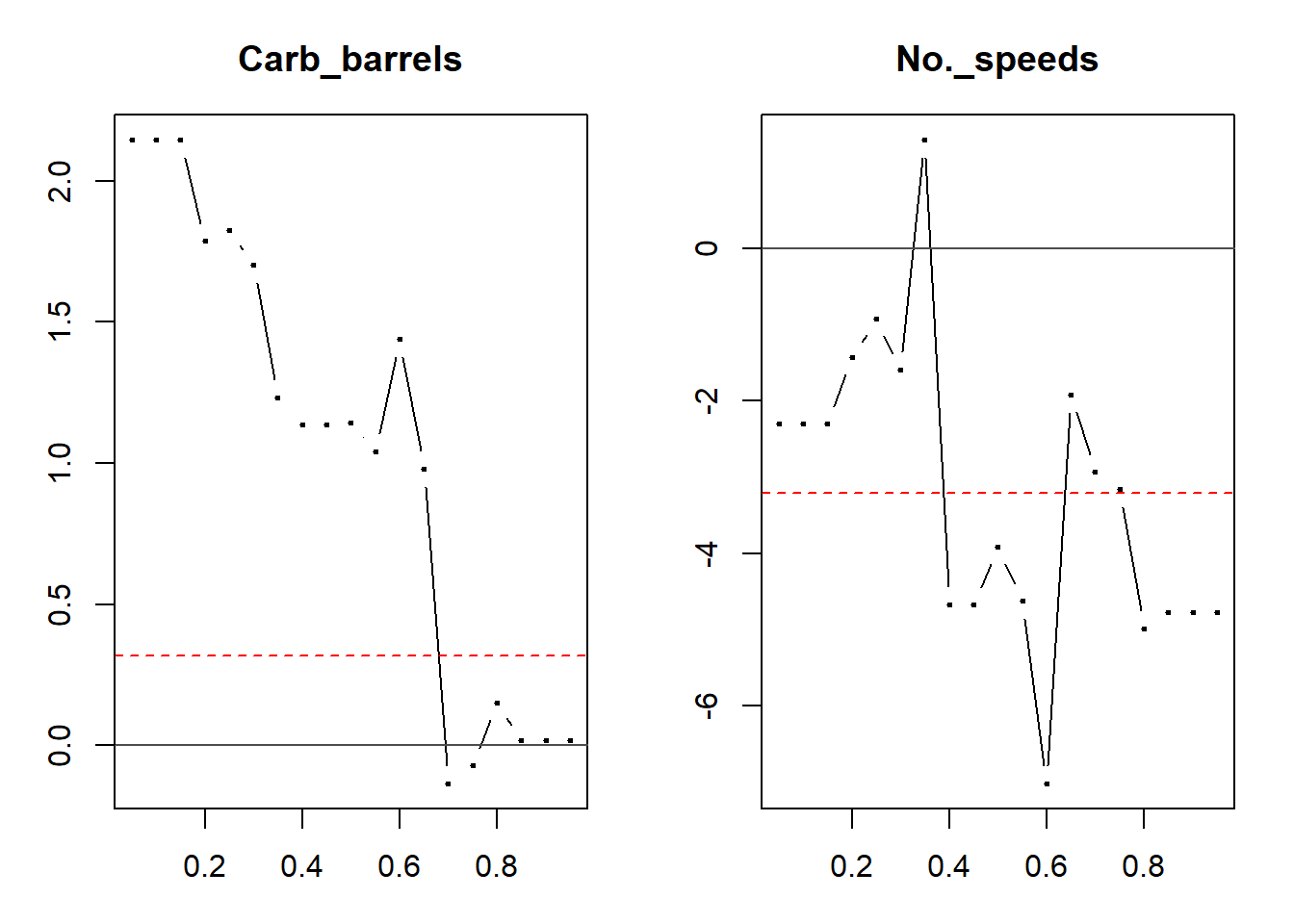
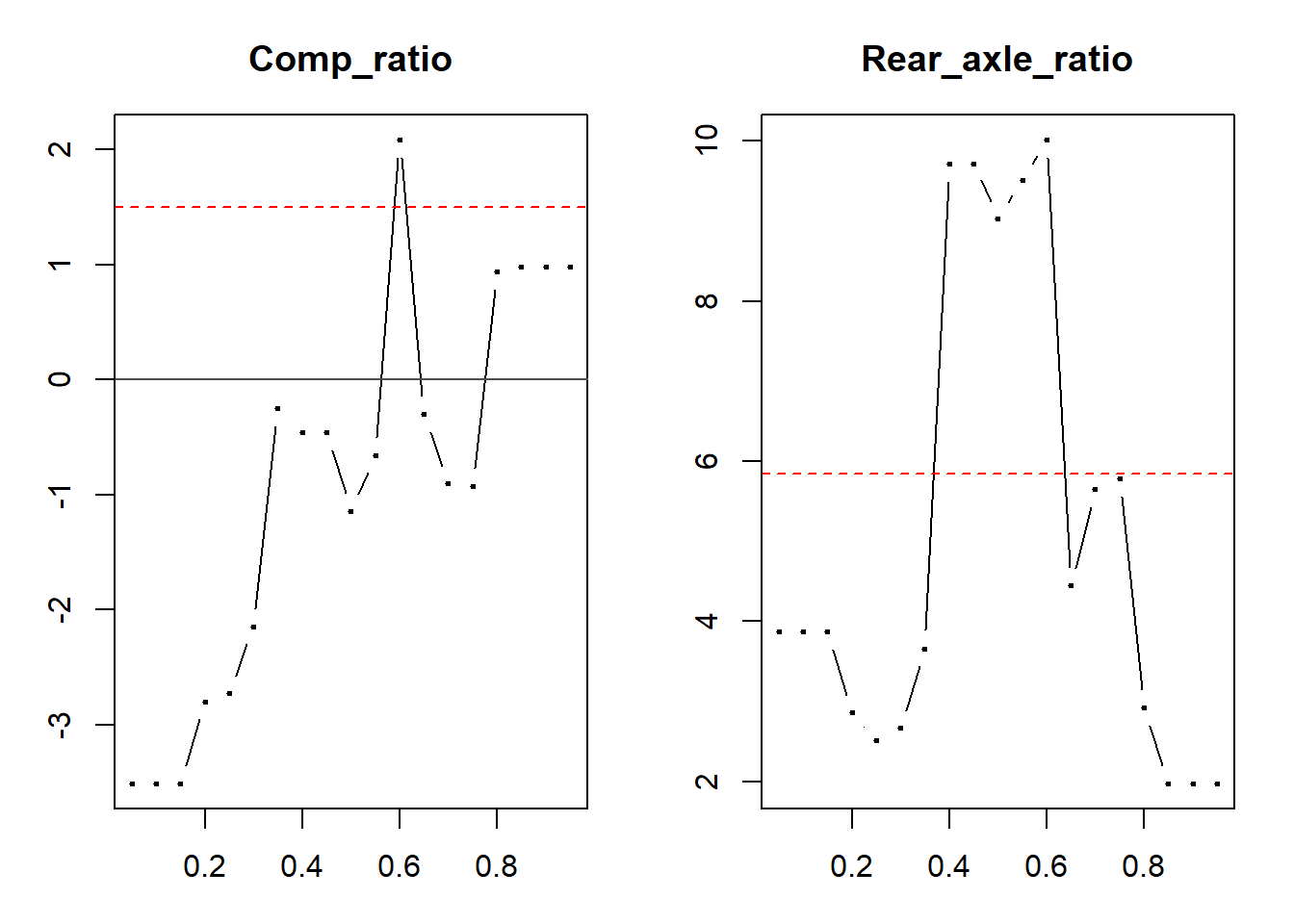
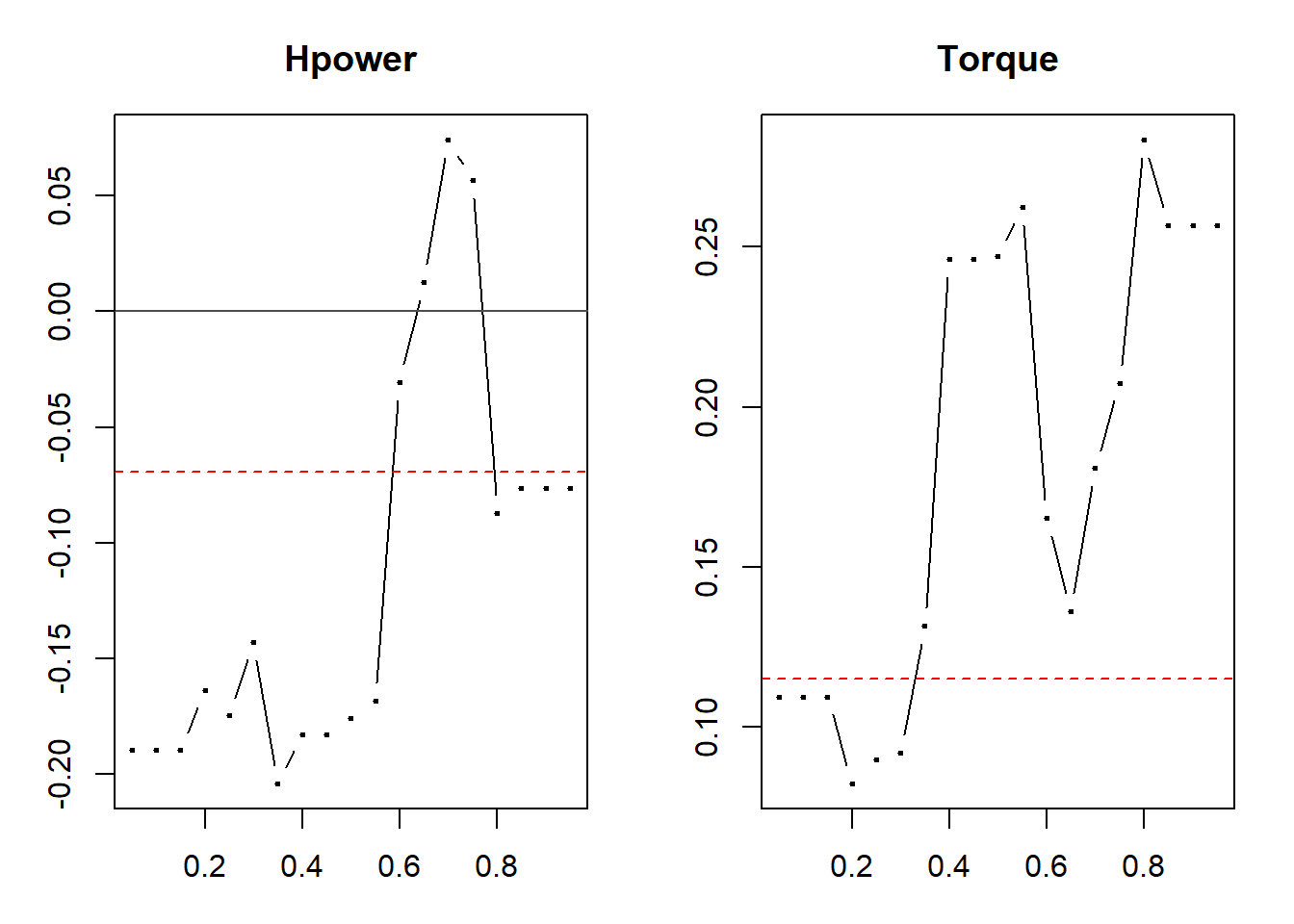
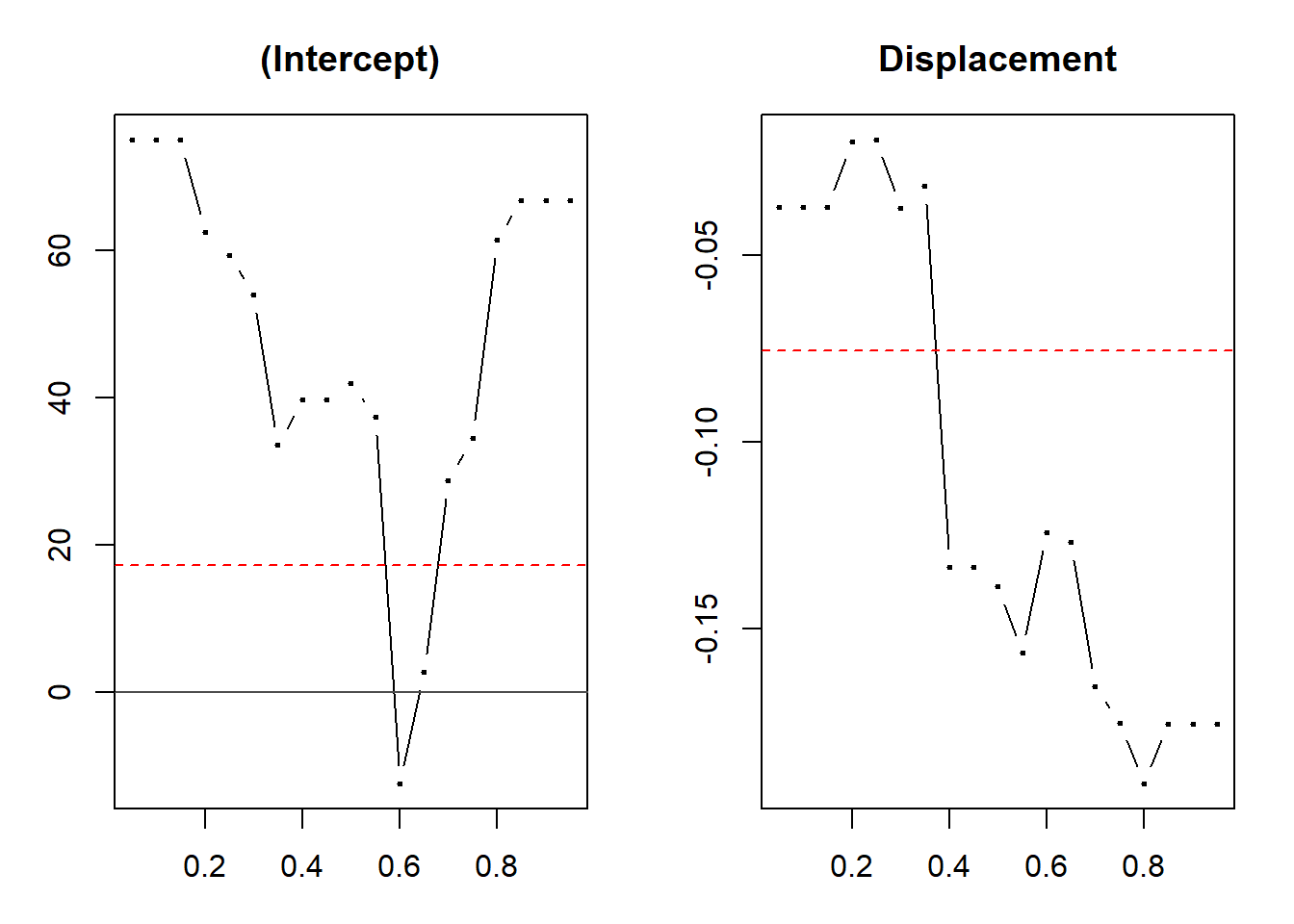
**b.** Metropolis algorithm is a special case of Metropolis-Hastings algorithm. In Metropolis algorithm the proposal distribution, q, has to be symmetric (. Since the distribution is symmetric the probability of accepting , is given by simply . Metropolis-Hastings is more generalized and thus, q does not have to be symmetric. The probability of accepting for MH is given by .

**c.** Multicollinearity in the X matrix causes it to be close to singular, and thus the   (predicted coefficients of regression) vector tends to be too long from the actual (actual coefficients of regression) vector. Ridge regression ameliorates this multicollinearity by reducing the level singularity in X. Ridge regression shrinks the coefficients. Lasso is similar to ridge, but lasso actually zeros out the coefficients of certain variables, effectively removing them from the model. In short, ridge and lasso regressions are used to address the ill-effects in data due to multicollinearity or too many predictors.

**d.** The ratio of the probabilities of choosing two alternatives is independent of the presence or attributes of any other alternative.

**Problem 2**

**a.** Quantile Regression Plots ( the plot function did not work for summary of the fit (plot(summary(fit)), so I just plotted fit (plot(fit)))

****

**c.** For displacements the coefficients are very negative for vehicles with Mpg above the 40th quantile. For vehicles with Mpg in the lower quantiles (below 40) the displacement coefficient magnitude is much lower and displacement plays a less significant role in determining fuel economy for lower Mpg vehicles.

For Horsepower, there is a peak for the coefficient around the 70th percentile. Horsepower has a positive effect for fuel economy when mpg is around the 65-75th percentile. For vehicles with mpg in other percentiles, horsepower has a negative effect. Horsepower has a significantly more negative effect for vehicles with mpg in the lower quantiles (<60th quantile) than vehicles with mpg >80th quantile).

Torque has positive coefficients for all the quantiles, and thus always has a positive effect on fuel economy. Torque has the most effect on fuel economy for vehicles with mpg between 40-50th percentile and >80 percentile. The effect of torque decreases around the 70th percentile for some reason.

The red line in all the plots indicates the coefficients for the mean mpg. In other words, it marks the coefficient for ordinary linear regression.

**d.** Summary of coefficients and their standard error using bootstrap method for conditional median.

Call: rq(formula = Mpg ~ ., tau = 0.5, data = gasdata)

tau: [1] 0.5

Coefficients:

Value Std. Error t value Pr(>|t|)

(Intercept) 41.98707 56.27002 0.74617 0.46520

Displacement -0.13873 0.11928 -1.16307 0.25999

Hpower -0.17596 0.19346 -0.90954 0.37509

Torque 0.24692 0.17393 1.41966 0.17279

Comp\_ratio -1.14223 5.03548 -0.22684 0.82311

Rear\_axle\_ratio 9.03682 7.18608 1.25755 0.22463

Carb\_barrels 1.14349 2.59236 0.44110 0.66439

No.\_speeds -3.91968 8.49951 -0.46117 0.65020

Length 0.17526 0.31399 0.55819 0.58359

Width -0.54095 0.63342 -0.85402 0.40432

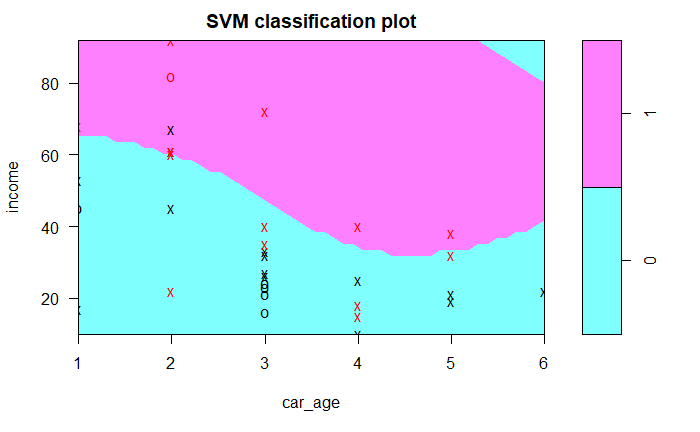
Weight -0.00472 0.01088 -0.43344 0.66985

Trans.\_type 1.99845 8.87414 0.22520 0.82436

**Problem 3**

**a.**

**b.** SVM classification plot for car data. As car age between 4 and 5 seem to have a higher probability of having 1 as a response, as the pink area (which denotes 1) increases here. Income above 60, seem to have a higher probability of having 1 as response. Car age of 1 seems to only occur for income above 60.



**c.** The predicted response for family with income 50 and car age 5 was **1**.

**R Code Appendix**

SabCMSiA400Lab3Assign

*Sabarish Chockalingam*

*December 5, 2018*

Problem 2

**library**(quantreg)

## Loading required package: SparseM

##

## Attaching package: 'SparseM'

## The following object is masked from 'package:base':

##

## backsolve

gasdata <- read.csv('gas\_mileage.csv')

gasdata<-gasdata[!is.na(gasdata$Torque),] *## removing na's*

fit1 <- rq(Mpg~.,tau=seq(0.05,0.95,0.05), data=gasdata) *## rqfit*

## Warning in rq.fit.br(x, y, tau = tau, ...): Solution may be nonunique

sfm <- summary(fit1) *## summary*

## Warning in rq.fit.br(x, y, tau = tau, ci = TRUE, ...): Solution may be

## nonunique

## Warning in rq.fit.br(x, y, tau = tau, ci = TRUE, ...): Solution may be

## nonunique

## Warning in rq.fit.br(x, y, tau = tau, ci = TRUE, ...): Solution may be

## nonunique

## Warning in rq.fit.br(x, y, tau = tau, ci = TRUE, ...): Solution may be

## nonunique

## Warning in rq.fit.br(x, y, tau = tau, ci = TRUE, ...): Solution may be

## nonunique

## Warning in rq.fit.br(x, y, tau = tau, ci = TRUE, ...): Solution may be

## nonunique

## Warning in rq.fit.br(x, y, tau = tau, ci = TRUE, ...): Solution may be

## nonunique

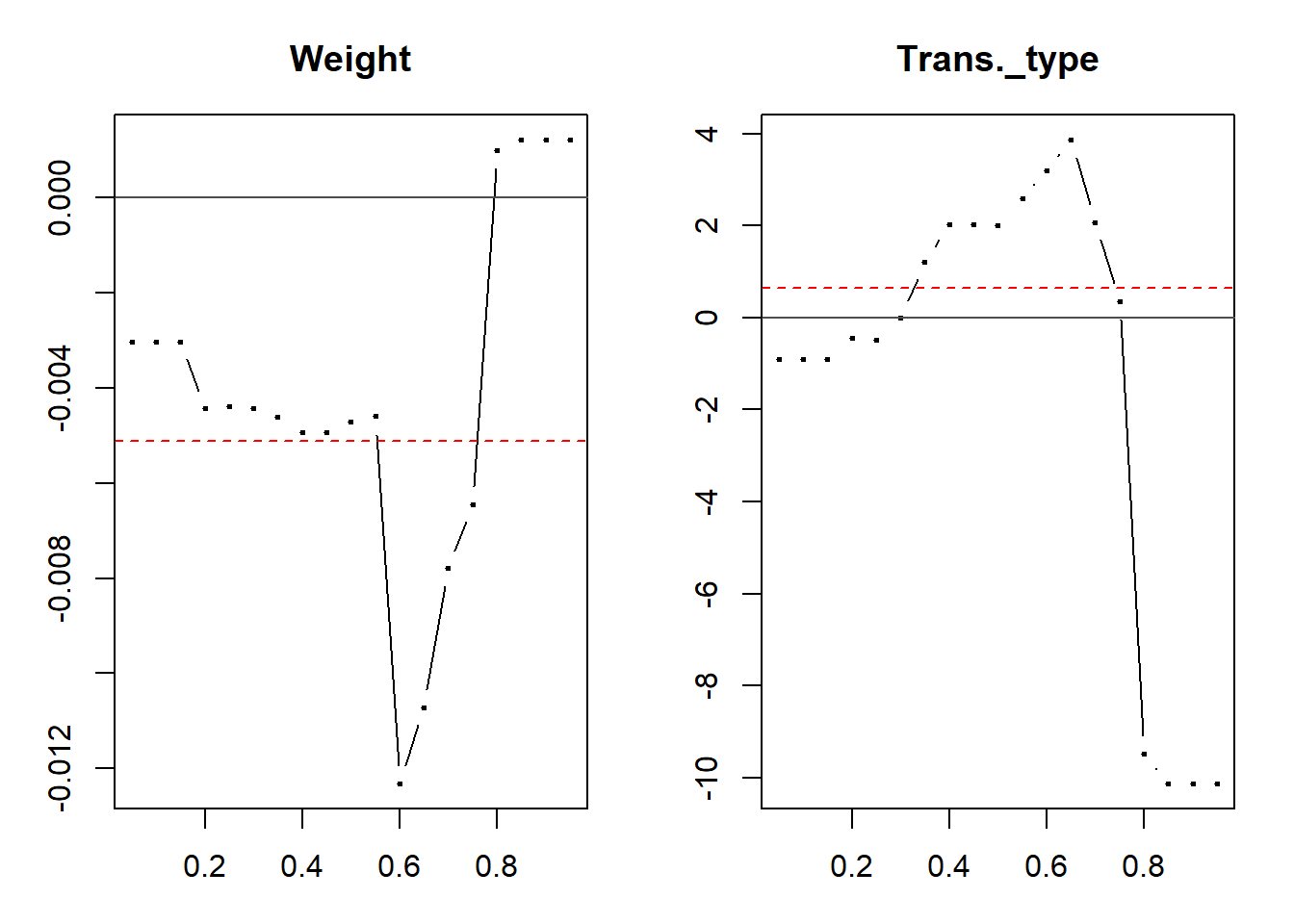
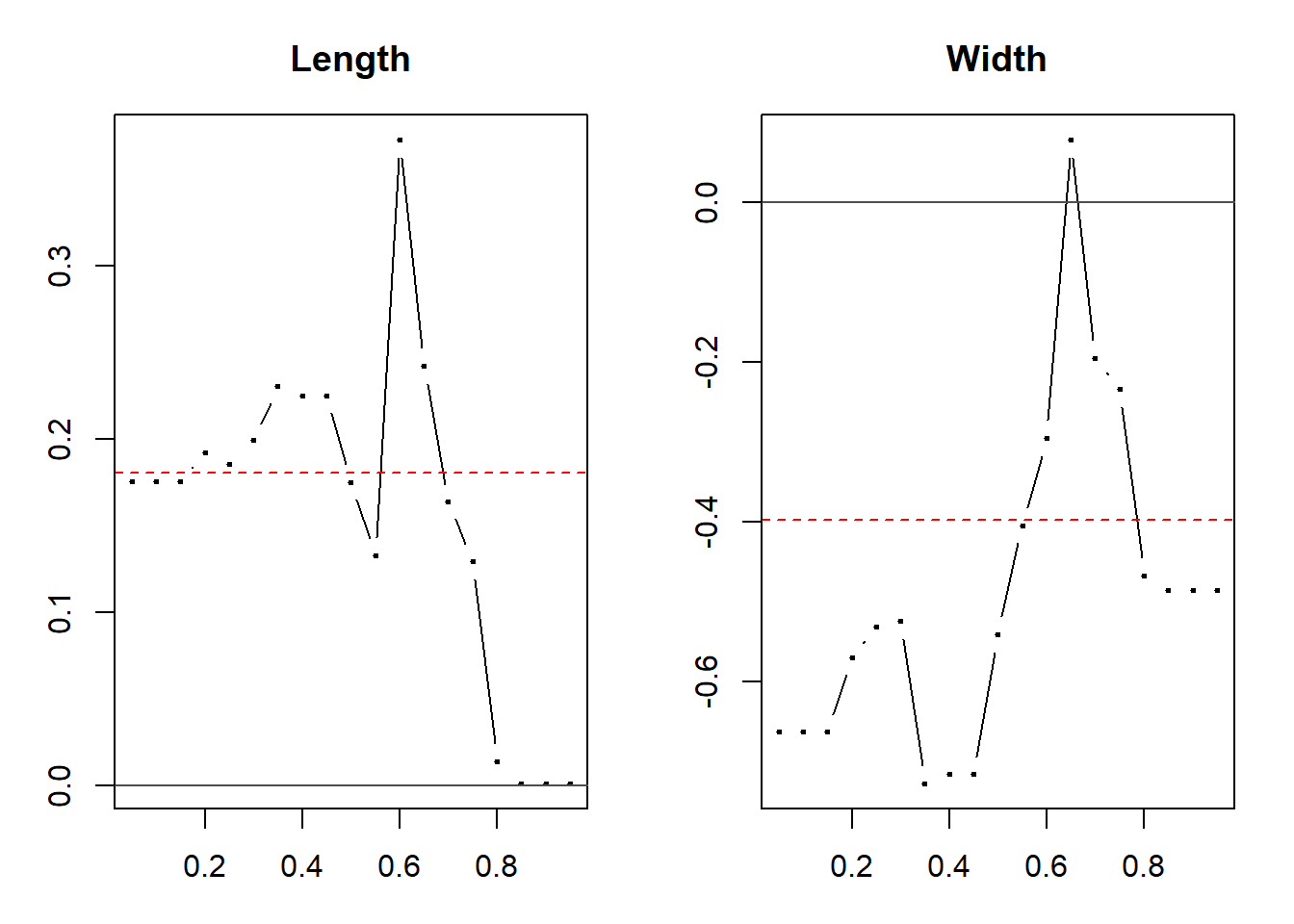
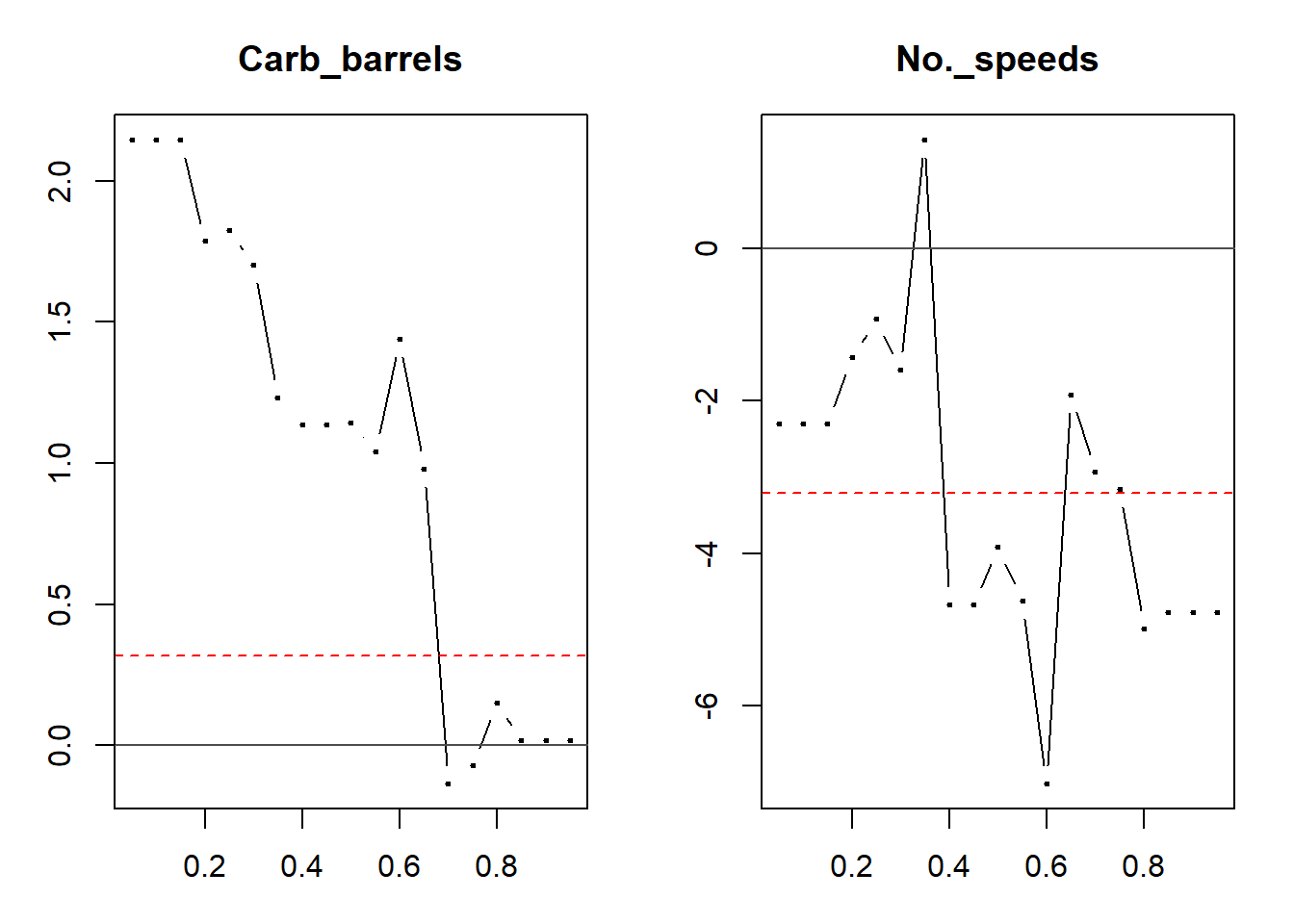
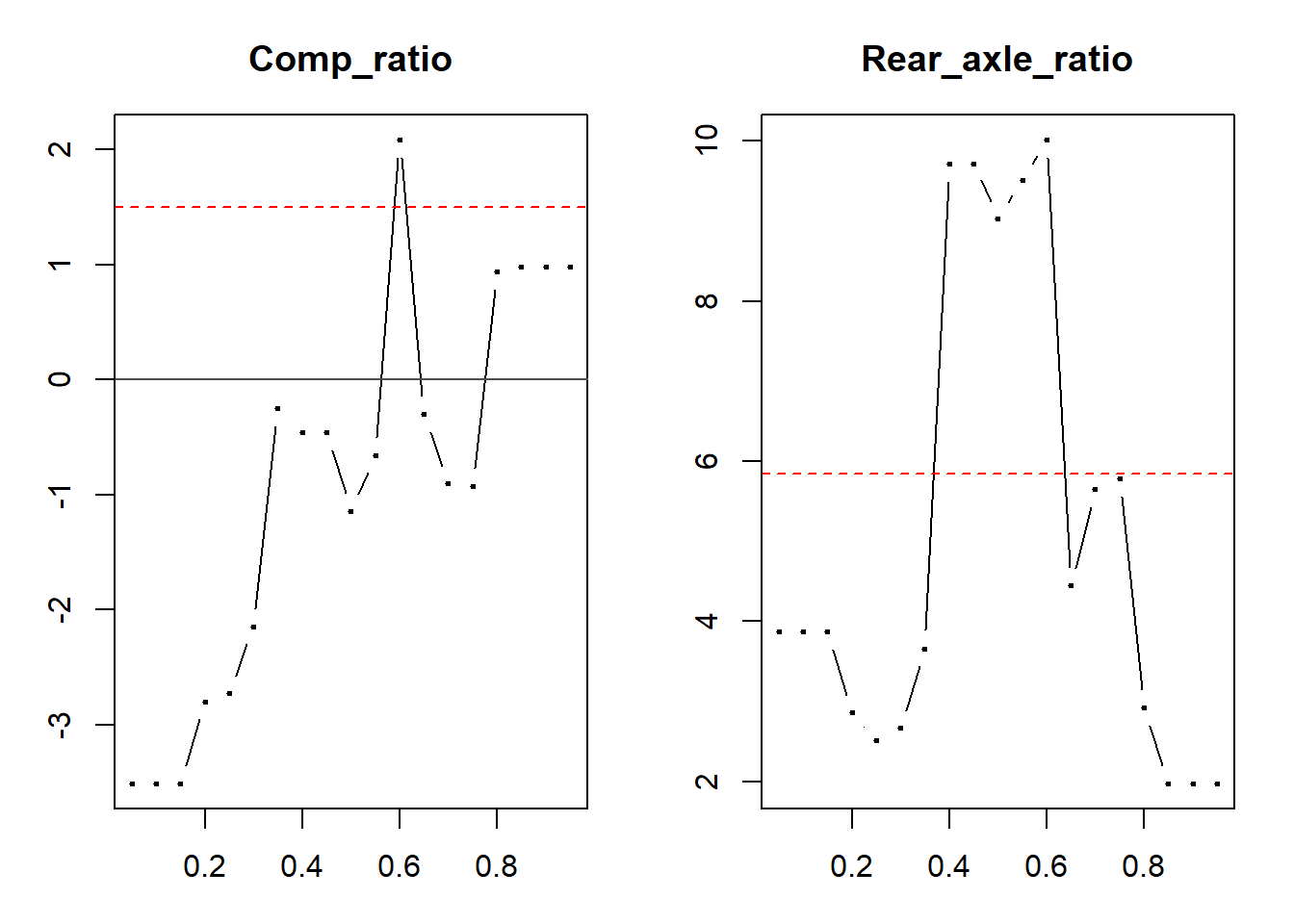
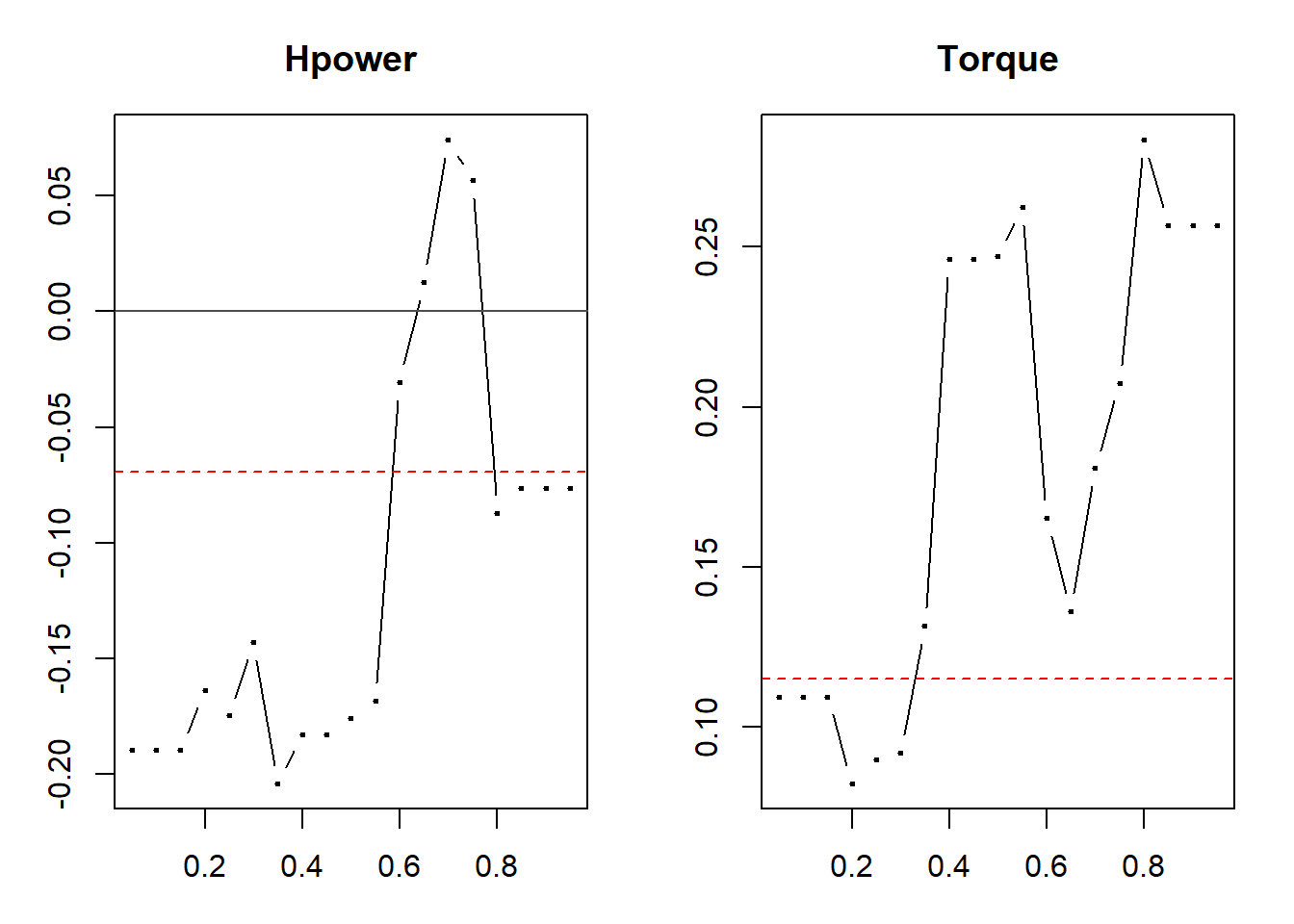
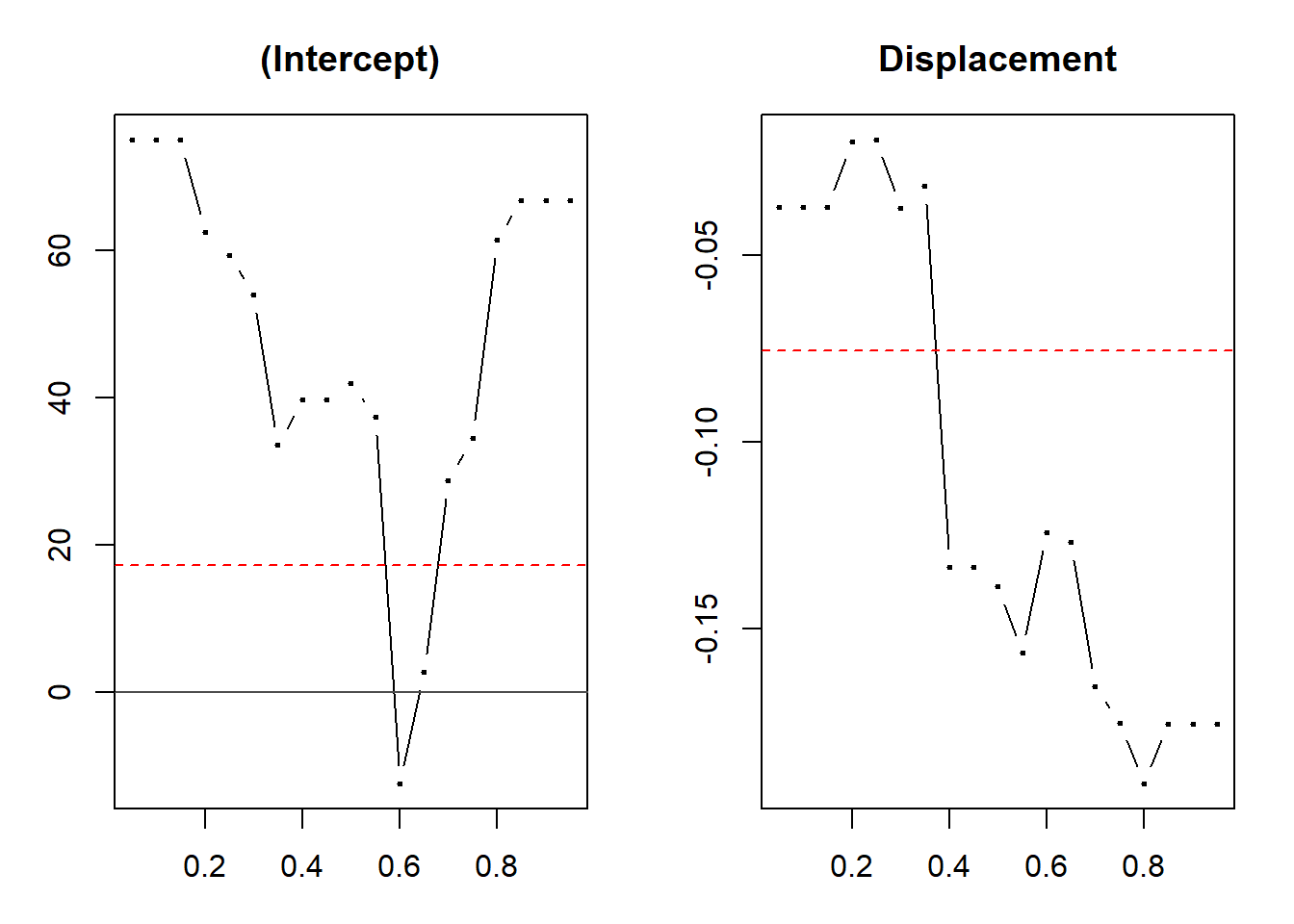
## Warning in rq.fit.br(x, y, tau = tau, ci = TRUE, ...): Solution may be

## nonunique

## Warning in rq.fit.br(x, y, tau = tau, ci = TRUE, ...): Solution may be

## nonunique

plot(fit1,mfrow=c(1,2)) *## plot did not work for sfm, so just used fit1*

****

fit2 <- rq(Mpg~.,tau=0.5, data=gasdata)

summary(fit2, se="boot")

##

## Call: rq(formula = Mpg ~ ., tau = 0.5, data = gasdata)

##

## tau: [1] 0.5

##

## Coefficients:

## Value Std. Error t value Pr(>|t|)

## (Intercept) 41.98707 59.04470 0.71111 0.48613

## Displacement -0.13873 0.12072 -1.14921 0.26551

## Hpower -0.17596 0.22320 -0.78833 0.44076

## Torque 0.24692 0.18586 1.32854 0.20060

## Comp\_ratio -1.14223 5.79125 -0.19723 0.84586

## Rear\_axle\_ratio 9.03682 7.04463 1.28280 0.21584

## Carb\_barrels 1.14349 2.66037 0.42982 0.67242

## No.\_speeds -3.91968 7.58634 -0.51668 0.61167

## Length 0.17526 0.31678 0.55327 0.58688

## Width -0.54095 0.63552 -0.85119 0.40585

## Weight -0.00472 0.01111 -0.42444 0.67628

## Trans.\_type 1.99845 8.05889 0.24798 0.80696

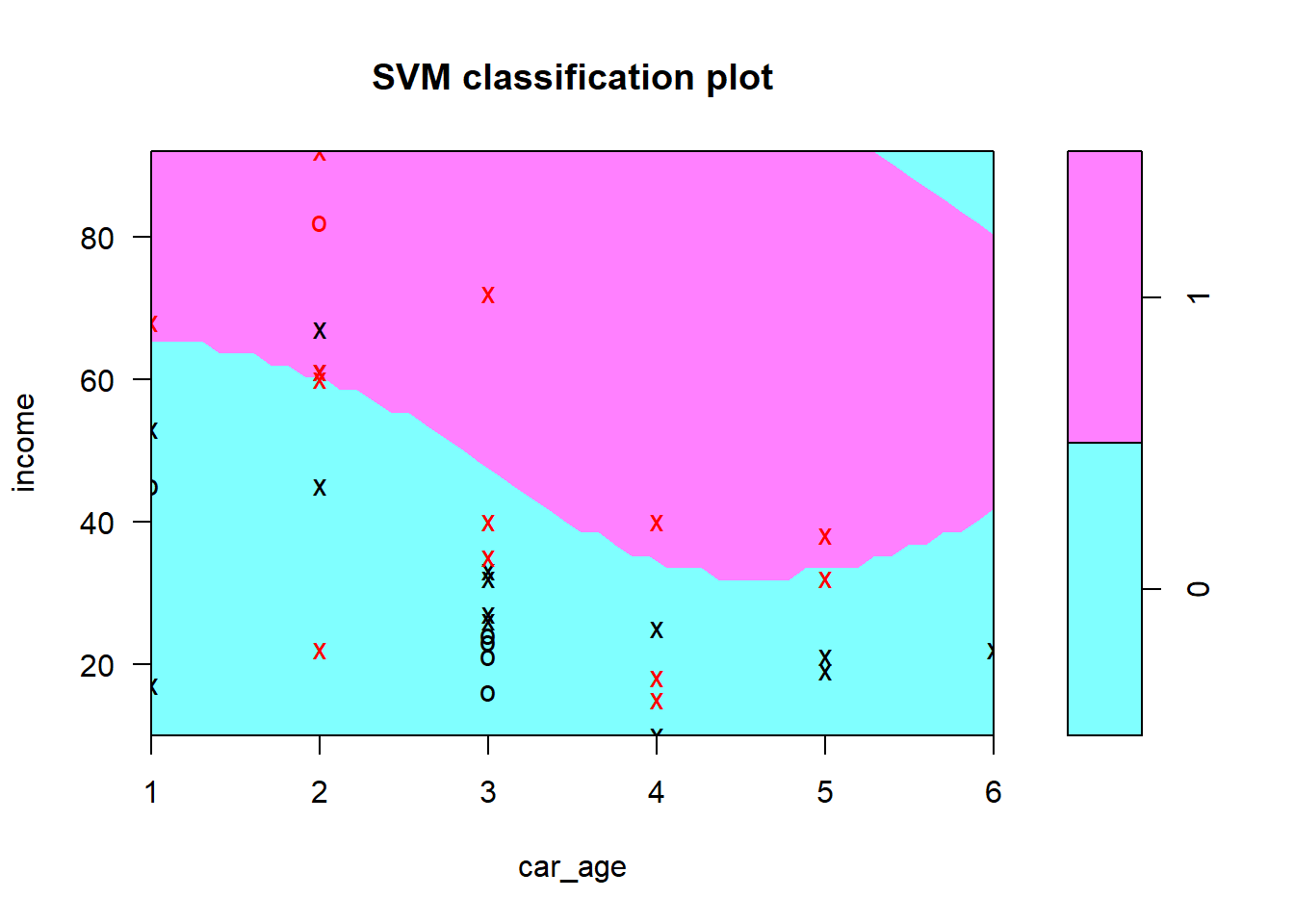
Problem 3

cardata <- read.csv('car.csv')

**library**(e1071)

svm = svm(factor(y)~ ., data = cardata)

plot(svm,cardata,income~car\_age )

****

newdata <- with(cardata, data.frame(income=50, car\_age = 5))

newdata$ypred<-predict(svm, newdata = newdata, type = "response")

print(paste("Response for carage=5 and income=50:",newdata$ypred))

## [1] "Response for carage=5 and income=50: 1"