A guide to robust statistical methods in neuroscience: Figure 9

Rand R. Wilcox & Guillaume A. Rousselet

2023-03-02

Contents

Dependencies

```
library(ggplot2)
library(cowplot)
library(tibble)
library(akima)
library(DetMCD)
```

```
## Loading required package: robustbase
```

```
## Loading required package: pcaPP
```

```
source("./code/Rallfun-v40.txt")
source("./code/theme_gar.txt")
```

Fractional Anisotropy and Reading Ability

The data were supplied by Suzanne Houston.

The labels of the last six variables correspond to six tracts: Left Arcuate Fasciculus, Right Arcuate Fasciculus, Left Inferior Longitudinal Fasciculus, Right Inferior Longitudinal Fascisculus, Left Corticospinal, Right Corticospinal.

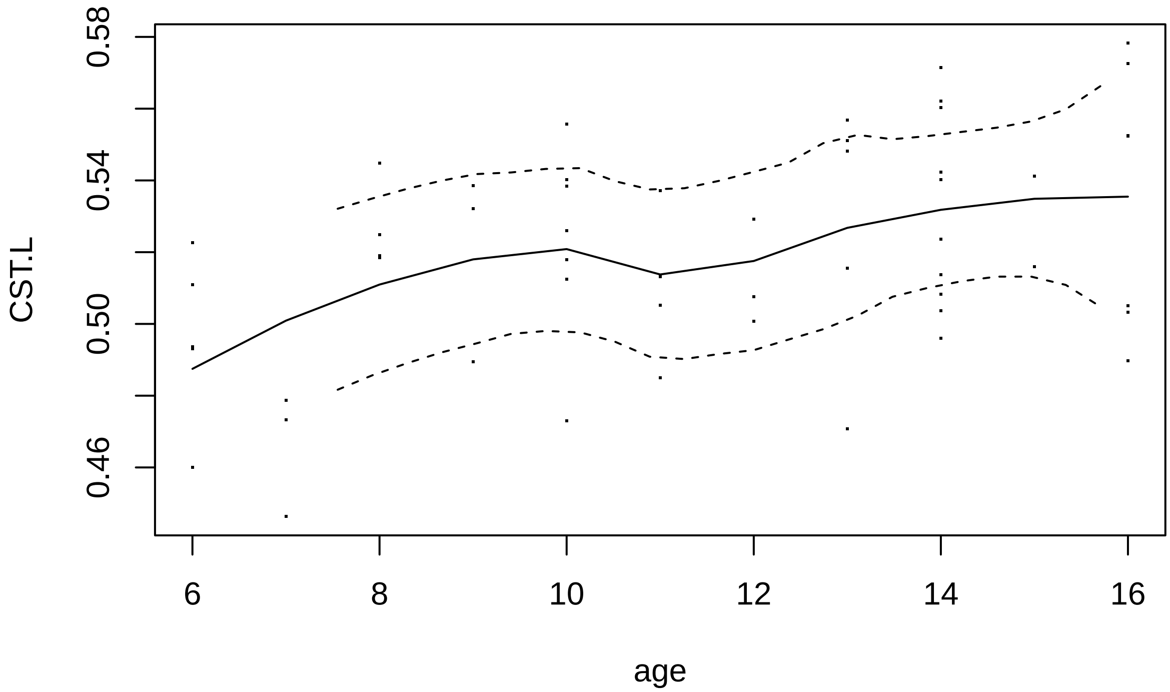
Get data

```
SH <- read.csv('./data/modeldata.csv',header=T,na.strings=999)
age <- SH[,3] # age
CST.L <- SH[,53] # CST.L = left corticospinal measures
GORT.FL <- SH[,26] # GORT.FL = GORT fluency measure
m <- elimna(cbind(age,CST.L,GORT.FL))
age <- m[,1]
CST.L <- m[,2]
GORT.FL <- m[,3]

# parameters for plots
axis_size <- 14
axis_title_size <- 16
```

Panel A: CST / age

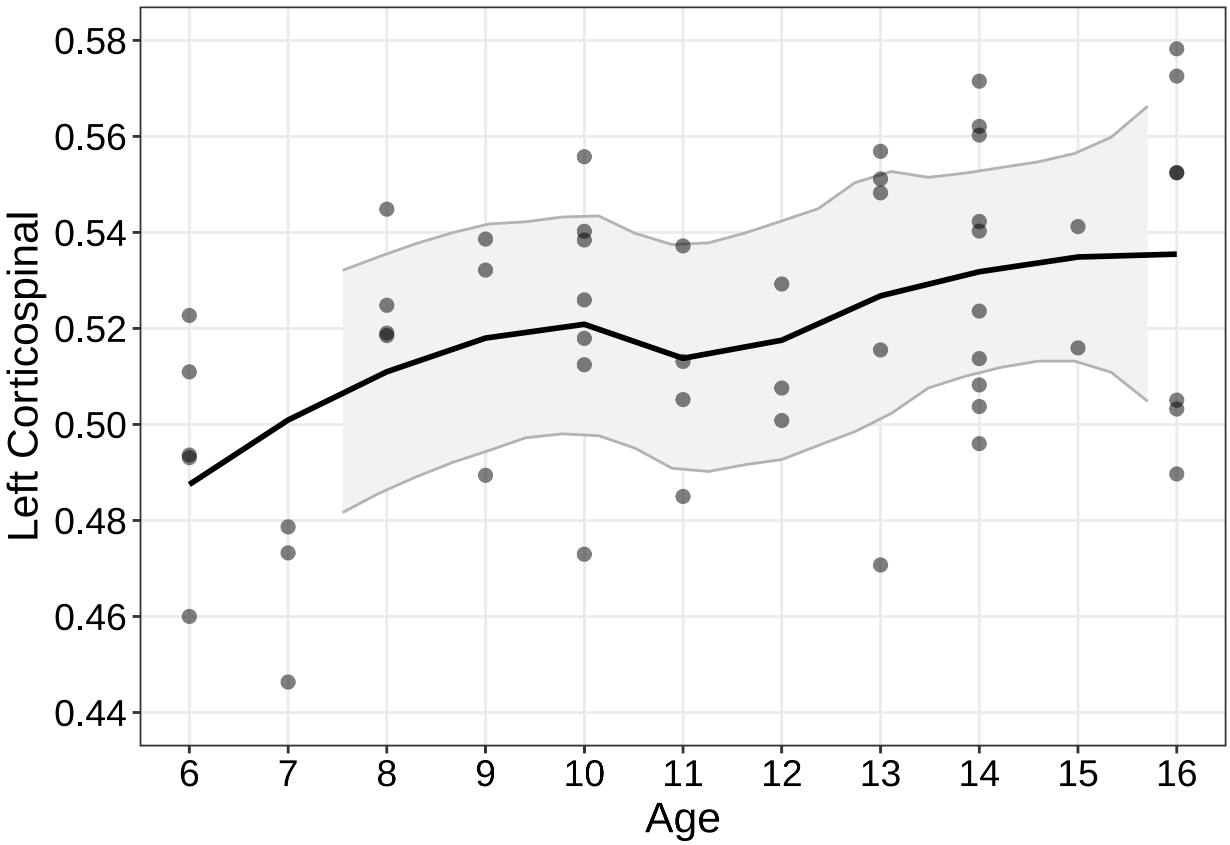
LOESS regression with confidence band

```
xord <- order(age)
age <- age[xord]
CST.L <- CST.L[xord]
GORT.FL <- GORT.FL[xord]

out <- lplotCI(age, CST.L, xlab='age', ylab='CST.L', CIV = TRUE, pts = age, plotit = FALSE)
out2 <- lplotCI(age, CST.L, xlab='age', ylab='CST.L', CIV = TRUE)
```



ggplot2 version

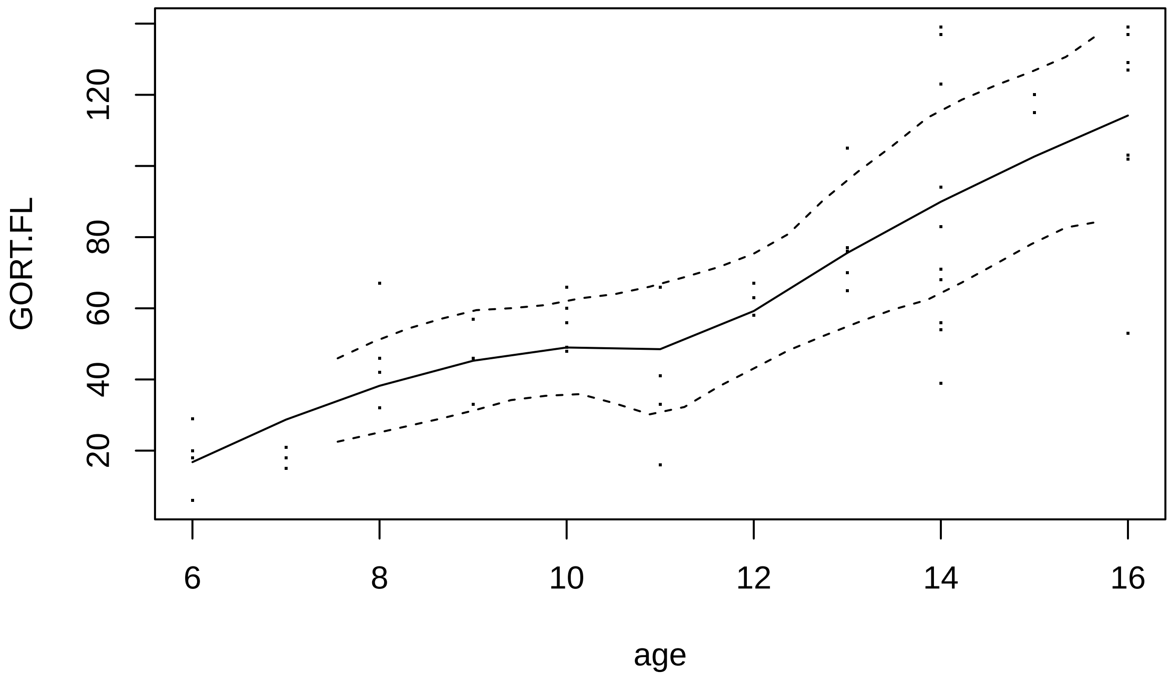
```
# make data frames
yhat <- out$Conf.Intervals[,2]
df <- tibble(age, CST.L, yhat)
x <- out2$Conf.Intervals[,1]
y <- out2$Conf.Intervals[,2]
lower <- out2$Conf.Intervals[,3]
upper <- out2$Conf.Intervals[,4]
dfhat <- tibble(x, y, lower, upper)

# make ggplot version
pA <- ggplot(dfhat, aes(x, y)) + theme_gar +
  geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey95") +
  geom_line(aes(y = lower), colour = "grey70") +
  geom_line(aes(y = upper), colour = "grey70") +
  geom_line(data = df, aes(age, yhat), colour = "black", size = 1) +
  geom_point(data = df, aes(age, CST.L), size = 2, alpha = 0.5) +
  scale_y_continuous(breaks = seq(0.44,0.58,0.02),
    limits = c(0.44, 0.58)) +
  scale_x_continuous(breaks = seq(6,16,1),
    limits = c(6, 16)) +
  theme(axis.text = element_text(size = axis_size),
    axis.title = element_text(size = axis_title_size),
    panel.grid.minor = element_blank()) +
  xlab("Age") +
  ylab("Left Corticospinal")
pA
```



Panel B: GORT / age

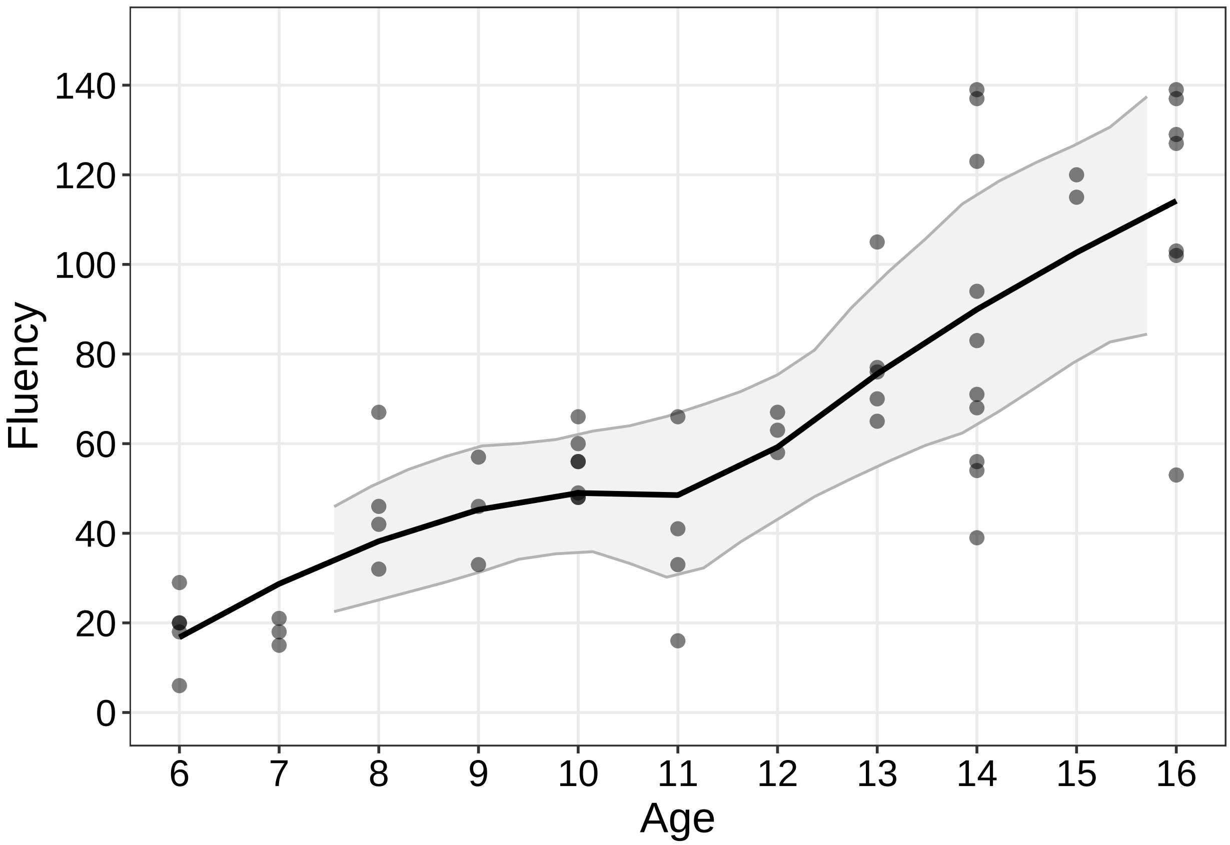
LOESS regression with confidence band

```
out <- lplotCI(age, GORT.FL, CIV = TRUE, pts = age, plotit = FALSE)
out2 <- lplotCI(age, GORT.FL, xlab='age', ylab='GORT.FL', CIV = TRUE, plotit = TRUE)
```



ggplot2 version

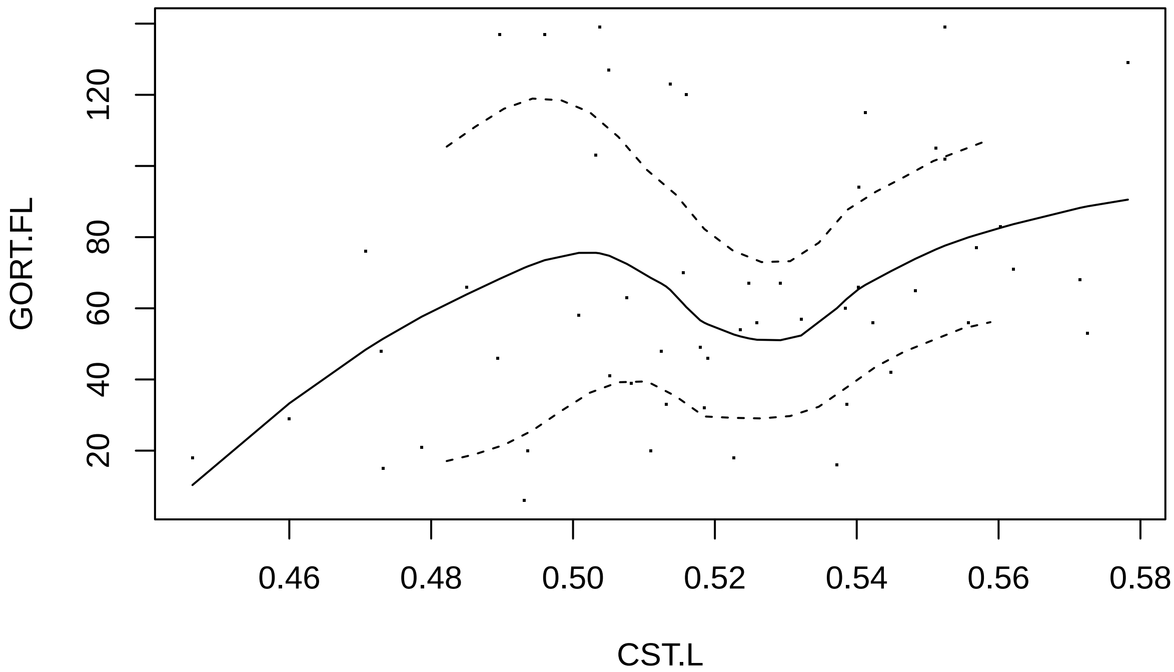
```
# make data frame
yhat <- out$Conf.Intervals[,2]
df <- tibble(age, GORT.FL, yhat)
x <- out2$Conf.Intervals[,1]
y <- out2$Conf.Intervals[,2]
lower <- out2$Conf.Intervals[,3]
upper <- out2$Conf.Intervals[,4]
dfhat <- tibble(x, y, lower, upper)

# make ggplot version
pB <- ggplot(dfhat, aes(x, y)) + theme_gar +
  geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey95") +
  geom_line(aes(y = lower), colour = "grey70") +
  geom_line(aes(y = upper), colour = "grey70") +
  geom_line(data = df, aes(age, yhat), colour = "black", size = 1) +
  geom_point(data = df, aes(age, GORT.FL), size = 2, alpha = 0.5) +
  scale_y_continuous(breaks = seq(0,140,20),
    limits = c(0, 150)) +
  scale_x_continuous(breaks = seq(6,16,1),
    limits = c(6, 16)) +
  theme(axis.text = element_text(size = axis_size),
    axis.title = element_text(size = axis_title_size),
    panel.grid.minor = element_blank()) +
  xlab("Age") +
  ylab("Fluency")
pB
```



Panel C: GORT / CST

LOESS regression with confidence band

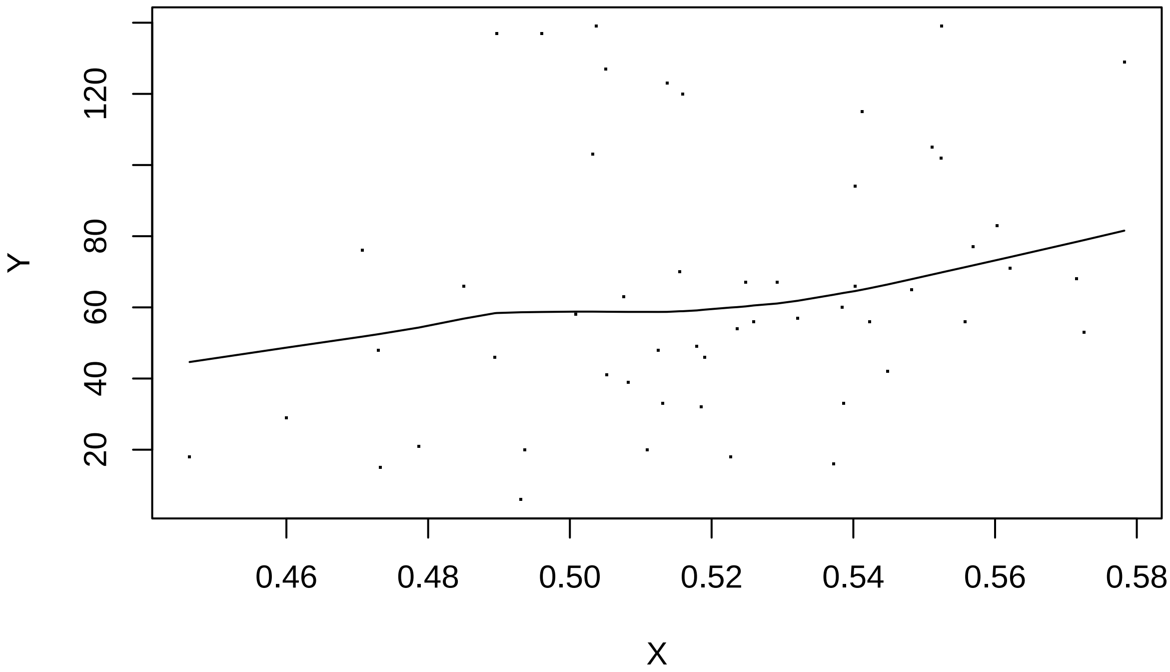
```
xord <- order(CST.L)
CST.L <- CST.L[xord]
GORT.FL <- GORT.FL[xord]

out <- lplotCI(CST.L, GORT.FL, CIV = TRUE, pts = CST.L, plotit = FALSE)
out2 <- lplotCI(CST.L, GORT.FL, xlab='CST.L', ylab='GORT.FL', CIV = TRUE, plotit = TRUE)
```



Running interval smoother suggests a straighter line

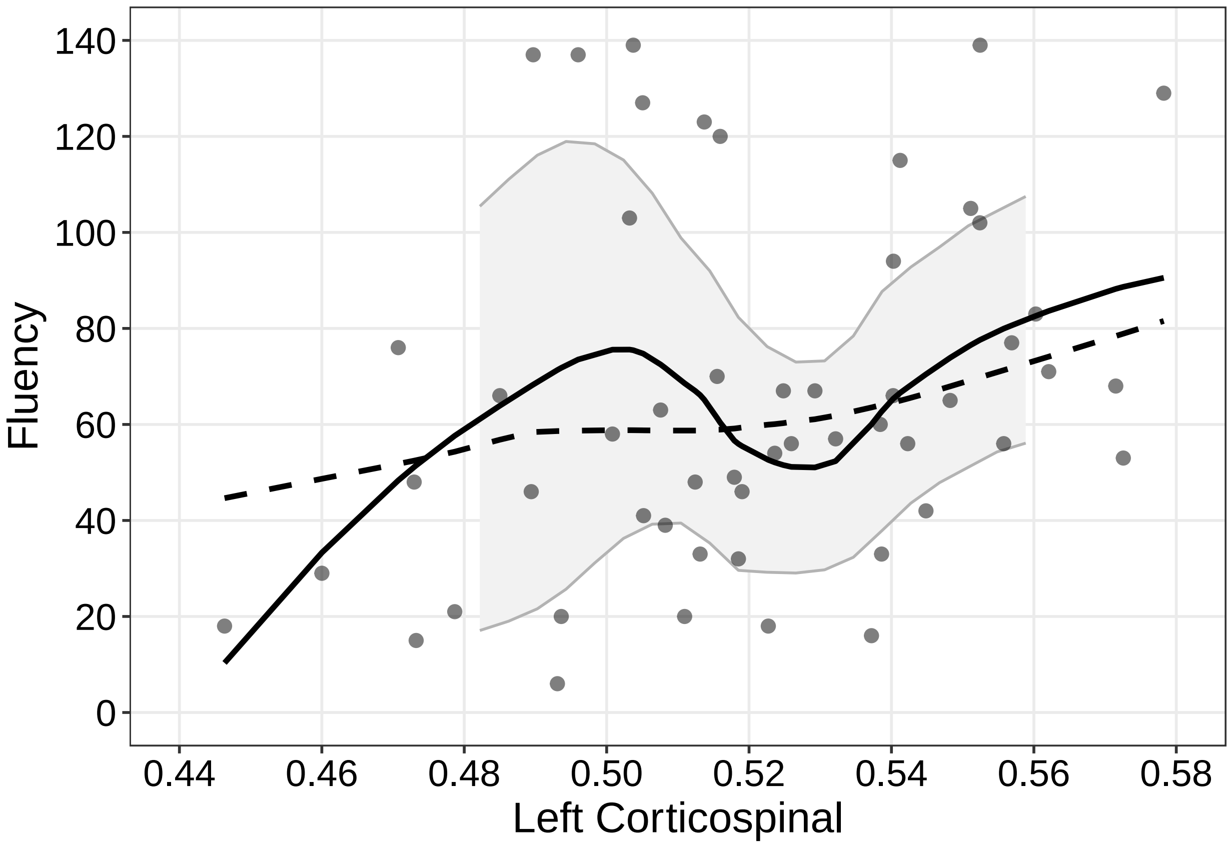
```
yhat.ris <- rplot(CST.L,GORT.FL,xout=F,pyhat=TRUE)$yhat
```

```
## [1] "Suggest also looking at result using xout=TRUE"
## [1] "A new estimate of the strength of the association is used by default."
## [1] " To get the old estimate, set OLD=TRUE"
```



ggplot2 version

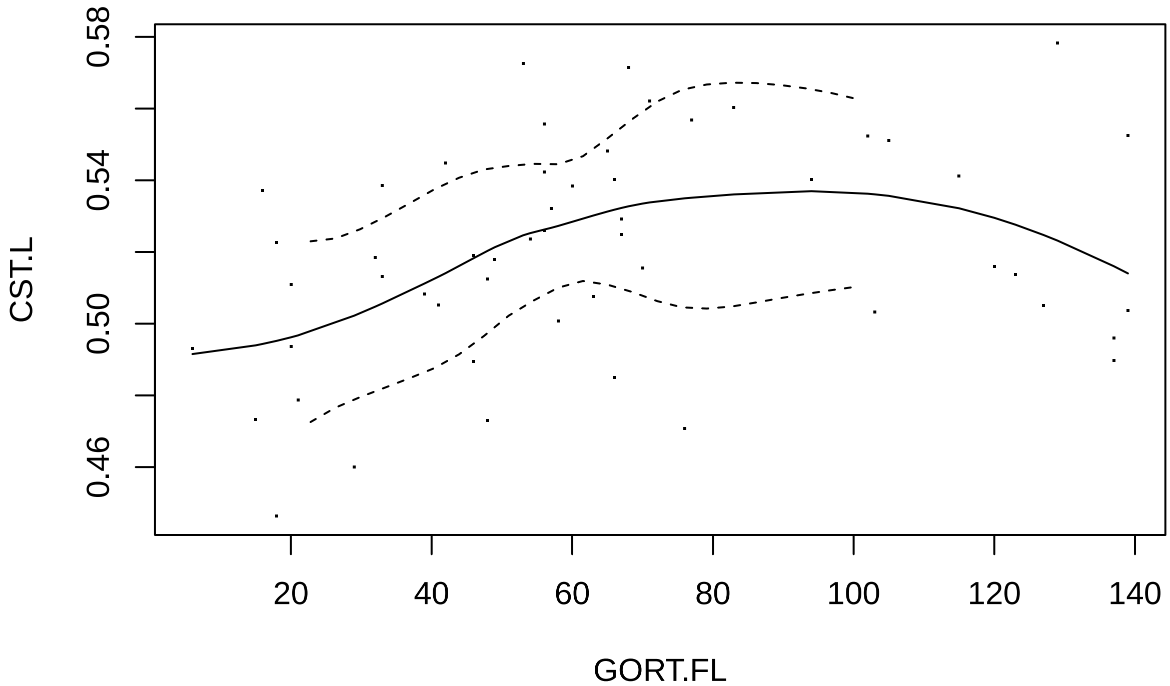
```
# make data frame
yhat <- out$Conf.Intervals[,2]
df <- tibble(CST.L, GORT.FL, yhat, yhat.ris)
x <- out2$Conf.Intervals[,1]
y <- out2$Conf.Intervals[,2]
lower <- out2$Conf.Intervals[,3]
upper <- out2$Conf.Intervals[,4]
dfhat <- tibble(x, y, lower, upper)

# make ggplot version
pC <- ggplot(dfhat, aes(x, y)) + theme_gar +
  geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey95") +
  geom_line(aes(y = lower), colour = "grey70") +
  geom_line(aes(y = upper), colour = "grey70") +
  geom_line(data = df, aes(CST.L, yhat), colour = "black", size = 1) +
  geom_line(data = df, aes(CST.L, yhat.ris), colour = "black", linetype = 2, size = 1) +
  geom_point(data = df, aes(CST.L, GORT.FL), size = 2, alpha = 0.5) +
  scale_y_continuous(breaks = seq(0,140,20),
    limits = c(0, 140)) +
  scale_x_continuous(breaks = seq(0.44,0.58,0.02),
    limits = c(0.44, 0.58)) +
  theme(axis.text = element_text(size = axis_size),
    axis.title = element_text(size = axis_title_size),
    panel.grid.minor = element_blank()) +
  xlab("Left Corticospinal") +
  ylab("Fluency")
pC
```



Panel D: CST / GORT

LOESS regression with confidence band

```
xord <- order(GORT.FL)
CST.L <- CST.L[xord]
GORT.FL <- GORT.FL[xord]

out <- lplotCI(GORT.FL, CST.L, CIV = TRUE, pts = GORT.FL, plotit = FALSE)
out2 <- lplotCI(GORT.FL, CST.L, xlab='GORT.FL', ylab='CST.L', CIV = TRUE, plotit = TRUE)
```

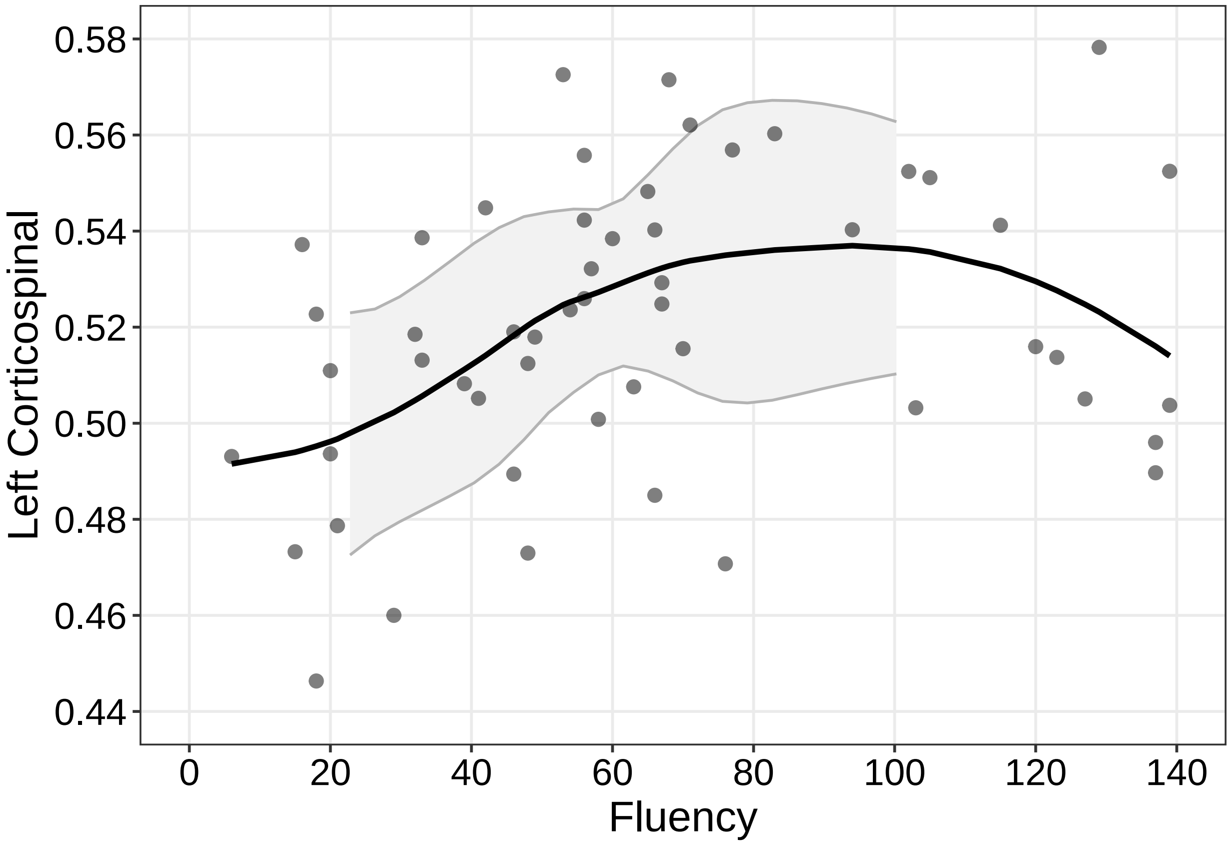



```
# outxout <- lplotCI(GORT.FL, CST.L, xout = TRUE, xlab='GORT.FL', ylab='CST.L', CIV = TRUE, plotit = TR
```

ggplot2 version

```
# make data frames
yhat <- out$Conf.Intervals[,2]
df <- tibble(GORT.FL, CST.L, yhat) # individual points
x <- out2$Conf.Intervals[,1]
y <- out2$Conf.Intervals[,2]
lower <- out2$Conf.Intervals[,3]
upper <- out2$Conf.Intervals[,4]
dfhat <- tibble(x, y, lower, upper) # fit
# x <- outxout$Conf.Intervals[,1]
# y <- outxout$Conf.Intervals[,2]
# dfhatxout <- tibble(x, y) # fit without IV outliers

# make ggplot version
pD <- ggplot(dfhat, aes(x, y)) + theme_gar +
  geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey95") + # confidence band
  geom_line(aes(y = lower), colour = "grey70") +
  geom_line(aes(y = upper), colour = "grey70") +
  geom_line(data = df, aes(GORT.FL, yhat), colour = "black", size = 1) + # fit
  geom_point(data = df, aes(GORT.FL, CST.L), size = 2, alpha = 0.5) + # individual points
  # geom_line(data = dfhatxout, aes(x, y),
  #           colour = "black", size = 0.5) + # fit without IV outliers
  scale_x_continuous(breaks = seq(0,140,20),
    limits = c(0, 140)) +
  scale_y_continuous(breaks = seq(0.44,0.58,0.02),
    limits = c(0.44, 0.58)) +
  theme(axis.text = element_text(size = axis_size),
```

```
    axis.title = element_text(size = axis_title_size),
    panel.grid.minor = element_blank()) +
  xlab("Fluency") +
  ylab("Left Corticospinal")
pD
```



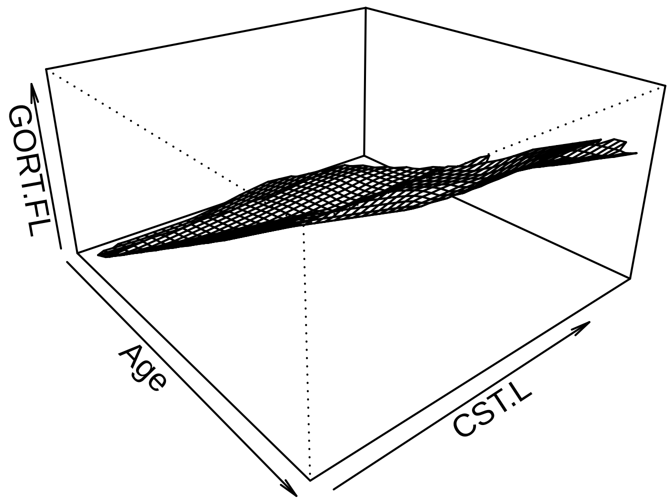
Combine panels into one figure

```
cowplot::plot_grid(pA, pB, pC, pD,
  labels=c("A", "B", "C", "D"),
  ncol = 2,
  nrow = 2,
  rel_widths = c(1, 1, 1, 1),
  label_size = 20,
  hjust = -0.5,
  scale=.95,
  align = "h")

# save figure
ggsave(filename='./figures/figure9.pdf',width=10,height=10)
```

Analyses presented in the text

First, look at the importance of age versus CST.L (Left Corticospinal) when predicting GORT.FL (The GORT Fluency reading score). Step 1: does the smoother suggest that the regression surface is approximately a plane?

```
lplot(SH[,c(3,53)],SH[,26],xlab='Age',ylab='CST.L',zlab='GORT.FL',xout=T)
```




```
## $Strength.Assoc
## [1] 0.917282
##
## $Explanatory.power
## [1] 0.8414063
##
## $yhat.values
## NULL
##
## $n
## [1] 53
##
## $n.keep
## [1] 53
```

Note: `xout=T` means that leverage points are removed.

Compare importance of age versus CST.L when predicting GORT.FL:

First compare Pearson correlations.

```
TWOpov(SH[,c(3,53)],SH[,26])
```

```
## $est.rho1
## [1] 0.7807353
##
## $est.rho2
## [1] 0.2769435
##
## $dif
## [1] 0.5037918
##
```

```
## $ci
## [1] 0.2768364 0.7884781
```

So this suggests that age is more important. Now consider what happens when both IVs are used in the model.

```
#regIVcom(SH[,c(3,53)],SH[,26],xout=T)
regIVcom(cbind(SH$age,SH$CST_L_Mean),SH$GORT_FL_R)
```

```
## $n
## [1] 53
##
## $n.keep
## [1] 53
##
## $est.1
## [1] 505.1673
##
## $est.2
## [1] 0.1796199
##
## $e.pow1
## [1] 0.7518191
##
## $e.pow2
## [1] 0.0002673207
##
## $strength.assoc.1
## [1] 0.867075
##
## $strength.assoc.2
## [1] 0.01634995
##
## $ratio
## [1] 2812.423
##
## $strength.ratio
## [1] 53.03229
##
## $p.value
## [1] 0
```

So again, based on the Theil–Sen estimator, the results suggest that age is more important. But in addition, the association between CST.L and GORT.FL is now much weaker compared to the situation where CST.L is ignored. Using instead least squares regression gives a similar result:

```
#regIVcom(SH[,c(3,53)],SH[,26],xout=T)
regIVcom(cbind(SH$age,SH$CST_L_Mean),SH$GORT_FL_R,regfun=ols)
```

```
## $n
## [1] 53
##
## $n.keep
## [1] 53
##
## $est.1
```

```
## [1] 572.3714
##
## $est.2
## [1] 5.898085
##
## $e.pow1
## [1] 0.851836
##
## $e.pow2
## [1] 0.00877787
##
## $strength.assoc.1
## [1] 0.9229496
##
## $strength.assoc.2
## [1] 0.09369029
##
## $ratio
## [1] 97.04359
##
## $strength.ratio
## [1] 9.85107
##
## $p.value
## [1] 0
```

Next, compare importance of age versus GORT.RA.R (raw rate) when predicting WJ.AT.R (raw word attack scores). First establish evidence for an association for each IV taken separately.

```
olshc4(SH[,c(3)],SH[,40],xout=T)
```

```
## $n
## [1] 80
##
## $n.keep
## [1] 80
##
## $ci
##             Coef. Estimates  ci.lower ci.upper      p-value Std.Error
## (Intercept)     0  2.602822 -2.199303 7.404948 2.838829e-01 2.4121018
## Slope           1  1.686390  1.307391 2.065388 2.031708e-13 0.1903705
##
## $cov
##            [,1]        [,2]
## [1,]  5.8182353 -0.44433195
## [2,] -0.4443319  0.03624092
##
## $test.stat
## [1] 8.858462
##
## $R.squared
## [1] 0.4678271
```

```
olshc4(SH[,c(22)],SH[,40],xout=T)
```

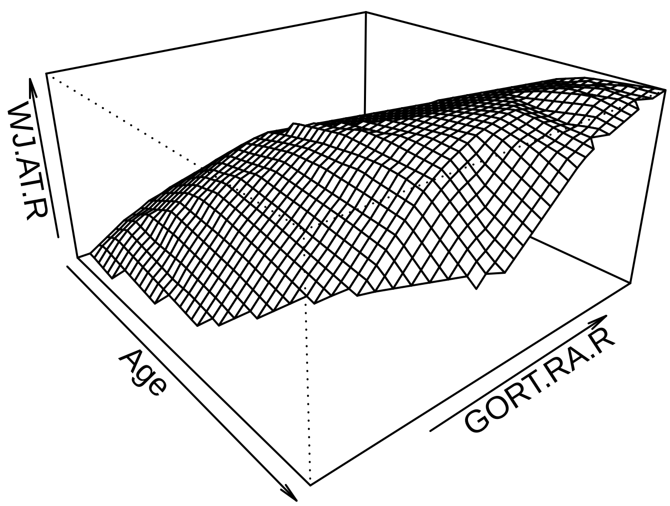
```
## $n
```

```
## [1] 73
##
## $n.keep
## [1] 73
##
## $ci
##             Coef.  Estimates  ci.lower   ci.upper      p-value  Std.Error
## (Intercept)     0 11.0939329 8.6769015 13.5109642 1.230127e-13 1.21218656
## Slope           1  0.3121753 0.2620571  0.3622936 0.000000e+00 0.02513525
##
## $cov
##              [,1]          [,2]
## [1,]  1.46939626 -0.0278012442
## [2,] -0.02780124  0.0006317809
##
## $test.stat
## [1] 12.41982
##
## $R.squared
## [1] 0.6225051
```

Next, perform a partial check on the assumption that the regression surface is a plane via the plot returned by the R function lplot.

```
lplot(SH[,c(3,22)],SH[,40],xlab='Age',ylab='GORT.RA.R',zlab='WJ.AT.R')
```

```
## [1] "Suggest also looking at result using xout=TRUE"
```



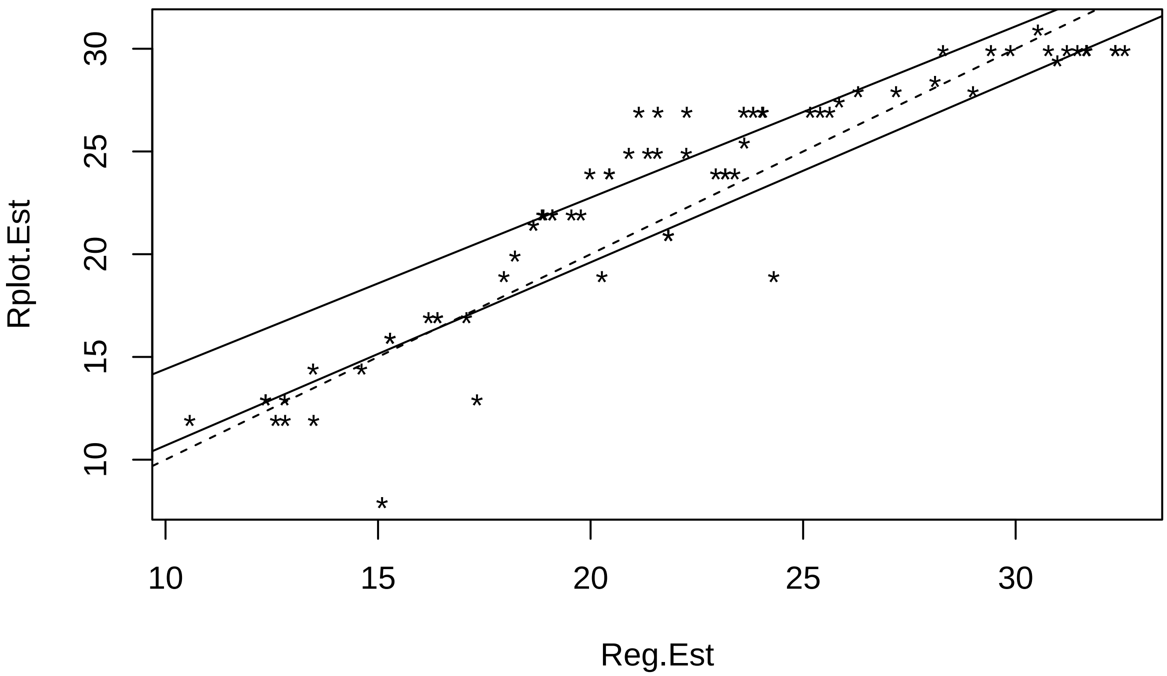

```
## $Strength.Assoc
## [1] 0.8387222
##
## $Explanatory.power
## [1] 0.703455
##
## $yhat.values
## NULL
##
## $n
## [1] 73
```

```
##
## $n.keep
## [1] 73
```

As an additional check, compare estimates of WJ.AT.R based on smooth versus a the usual linear regression line. If the usual linear model is reasonably accurate, a plot of the predicted values based on each method should be centered around a line with slope 1 and intercept zero.

```
# pmodchk(SH[,c(3,22)],SH[,40],op=0) # DV=WORD ATTACK
reg.vs.rplot(SH[,c(3,22)],SH[,40]) # DV=WORD ATTACK
```

```
## [1] "This function was updated July 2022"

## Loading required package: SparseM

##
## Attaching package: 'SparseM'

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



Now consider whether age (col. 3) is more important.

Theil-Sen

```
regIVcom(SH[,c(3,22)],SH[,40],xout=T) # Theil--Sen, DV=WORD ATTACK
```

```
## $n
## [1] 73
##
## $n.keep
## [1] 72
##
## $est.1
## [1] 0
##
```

```
## $est.2
## [1] 14.98479
##
## $e.pow1
## [1] 0
##
## $e.pow2
## [1] 0.4318465
##
## $strength.assoc.1
## [1] 0
##
## $strength.assoc.2
## [1] 0.6571503
##
## $ratio
## [1] 0
##
## $strength.ratio
## [1] 0
##
## $p.value
## [1] 0.00065
```

Least squares

```
regIVcom(SH[,c(3,22)],SH[,40],xout=T,regfun=ols) # Least Sq.
```

```
## $n
## [1] 73
##
## $n.keep
## [1] 72
##
## $est.1
## [1] 0.5347689
##
## $est.2
## [1] 10.49163
##
## $e.pow1
## [1] 0.0154115
##
## $e.pow2
## [1] 0.3023582
##
## $strength.assoc.1
## [1] 0.1241431
##
## $strength.assoc.2
## [1] 0.5498711
##
## $ratio
## [1] 0.05097101
##
```

```
## $strength.ratio
## [1] 0.2257676
##
## $p.value
## [1] 0.016
```

In contrast, when the independent variables are are considered separately, rather than including both in the model, and if Pearson's correlations are compared, fail to reject:

```
TWOpovPV(SH[,c(3,22)],SH[,40])
```

```
## $p.value
## [1] 0.043
##
## $est.rho1
## [1] 0.6839788
##
## $est.rho2
## [1] 0.7889899
##
## $ci
## [1] -0.2272516754 -0.0004480487
```

As another example, take age and CST.L as the independent variables and WJ.WA.R (Woodcock word attack score) as the dependent variable.

```
olshc4(SH[,53],SH[,26],xout=T)
```

```
## $n
## [1] 53
##
## $n.keep
## [1] 53
##
## $ci
##             Coef. Estimates   ci.lower  ci.upper    p-value Std.Error
## (Intercept)     0 -111.8054 -273.98219  50.37149 0.17237454   80.7821
## Slope           1  341.2089   33.95483 648.46291 0.03021582  153.0467
##
## $cov
##            [,1]      [,2]
## [1,]   6525.748 -12341.30
## [2,] -12341.301  23423.29
##
## $test.stat
## [1] 2.229443
##
## $R.squared
## [1] 0.07669772
```

```
olshc4(SH[,3],SH[,26],xout=T)
```

```
## $n
## [1] 74
##
## $n.keep
## [1] 74
```

```
##
## $ci
##              Coef.  Estimates  ci.lower  ci.upper      p-value Std.Error
## (Intercept)     0 -39.014456 -55.14523 -22.88368 7.705507e-06 8.0918349
## Slope           1   9.111319   7.49560  10.72704 0.000000e+00 0.8105081
##
## $cov
##            [,1]       [,2]
## [1,] 65.477791 -6.3605351
## [2,] -6.360535  0.6569235
##
## $test.stat
## [1] 11.24149
##
## $R.squared
## [1] 0.6095476
```

```
regIVcom(SH[,c(3,53)],SH[,26],xout=T)  # Theil--Sen
```

```
## $n
## [1] 53
##
## $n.keep
## [1] 53
##
## $est.1
## [1] 505.1673
##
## $est.2
## [1] 0.1796199
##
## $e.pow1
## [1] 0.7518191
##
## $e.pow2
## [1] 0.0002673207
##
## $strength.assoc.1
## [1] 0.867075
##
## $strength.assoc.2
## [1] 0.01634995
##
## $ratio
## [1] 2812.423
##
## $strength.ratio
## [1] 53.03229
##
## $p.value
## [1] 0
```

```
regIVcom(SH[,c(3,53)],SH[,26],xout=T,regfun=ols)  # Least squares
```

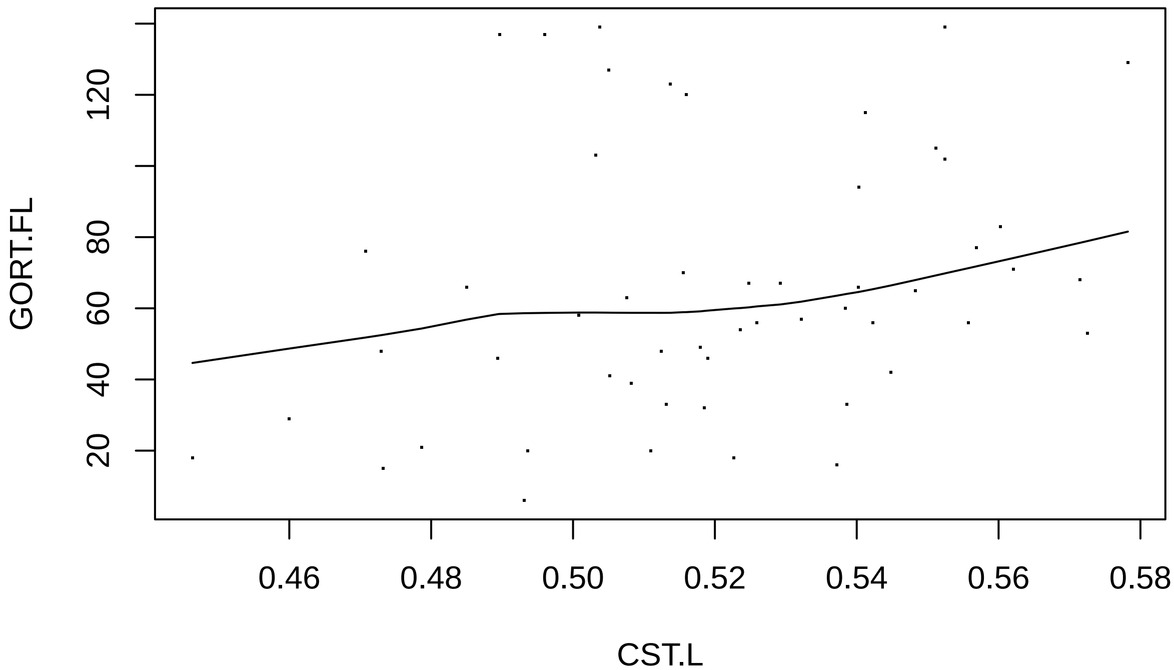
```
## $n
```

```
## [1] 53
##
## $n.keep
## [1] 53
##
## $est.1
## [1] 572.3714
##
## $est.2
## [1] 5.898085
##
## $e.pow1
## [1] 0.851836
##
## $e.pow2
## [1] 0.00877787
##
## $strength.assoc.1
## [1] 0.9229496
##
## $strength.assoc.2
## [1] 0.09369029
##
## $ratio
## [1] 97.04359
##
## $strength.ratio
## [1] 9.85107
##
## $p.value
## [1] 0
```

It is noted that the regression line for predicting CST.L with GORT.FL appears to be reasonably straight based on the running interval smoother:

```
rplot(SH[,53],SH[,26],ylab='GORT.FL',xlab='CST.L',xout=T)
```

```
## [1] "A new estimate of the strength of the association is used by default."
## [1] " To get the old estimate, set OLD=TRUE"
```




```
## $n
## [1] 53
##
## $n.keep
## [1] 53
##
## $Strength.Assoc
## [1] 0.3038042
##
## $Explanatory.Power
## [1] 0.09229699
##
## $xvals
## NULL
##
## $yhat
## NULL
```

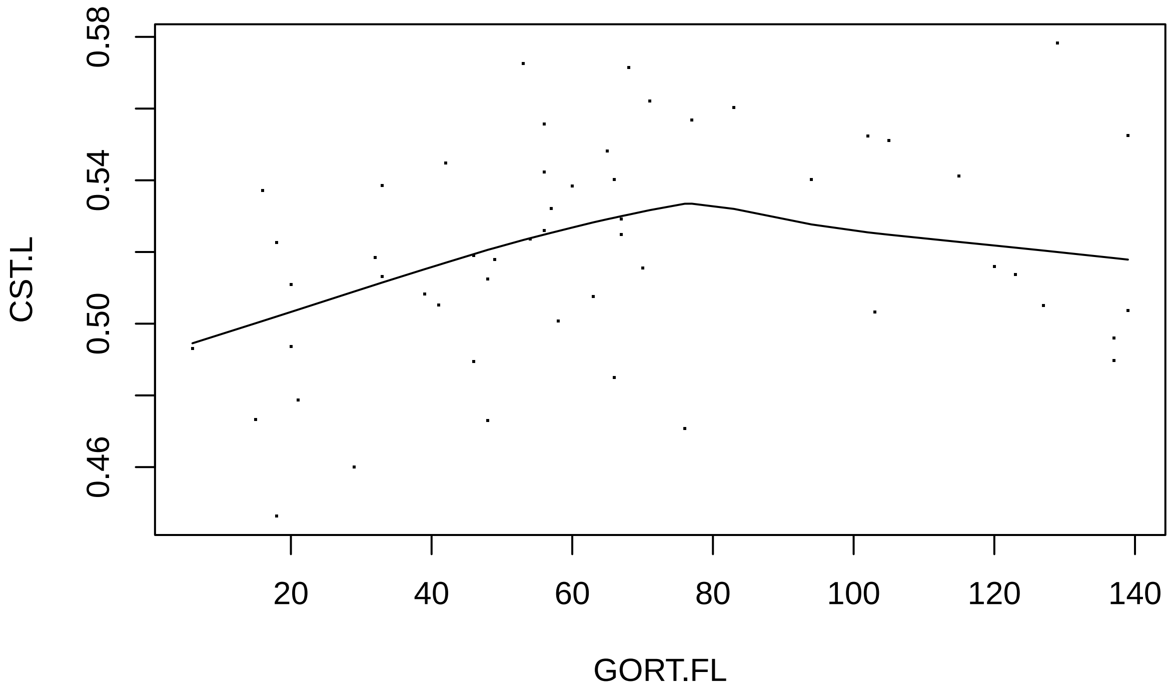
This straighter line is plotted in Figure 9C of the article.

Moreover, using both independent variables, the usual linear model appears to provide a reasonable approximation of the regression surface based on nonparametric estimator LOESS, which is applied by the R function lplot (leverage points removed).

But in general, do not assume that the usual linear model is satisfactory or that curvature can be addressed simply by adding a quadratic term. Here is an example where raw GORT fluency (GORT.L) scores are used to predict Left Corticospinal measures (CST.L). Now assuming a straight regression line seems dubious at best:

```
rplot(SH[,c(26)],SH[,53],xlab='GORT.FL',ylab='CST.L')
```

```
## [1] "Suggest also looking at result using xout=TRUE"
## [1] "A new estimate of the strength of the association is used by default."
## [1] " To get the old estimate, set OLD=TRUE"
```




```
## $n
## [1] 53
##
## $n.keep
## [1] 53
##
## $Strength.Assoc
## [1] 0.4138911
##
## $Explanatory.Power
## [1] 0.1713058
##
## $xvals
## NULL
##
## $yhat
## NULL
```

Is it possible to straighten the regression line by replacing X with X^a for some a? Half slope ratio is negative, which indicates that the answer is no:

```
hratio(SH[,26],SH[,53])
```

```
##            [,1]
## [1,] -1.232936
```

Now, suppose we split the data where CST.FL is less than or equal to 70. We can compare the slope when CST.FL is less than or equal to 70 to the slope when CST.FL is greater than 70:

```
flag=SH[,c(26)]<=70
reg2ci(SH[flag,c(26)],SH[flag,53],SH[!flag,c(26)],SH[!flag,53],plotit=F)
```

```
## $n
## [1] 36 17
```

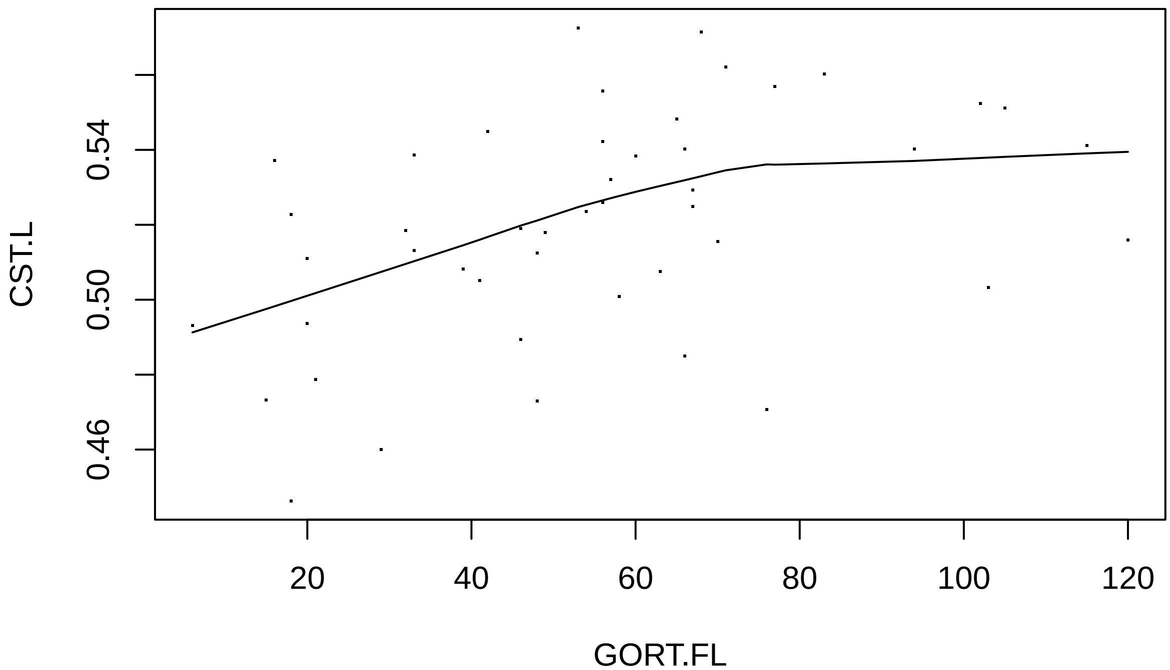
```
##
## $output
##      Parameter      ci.lower    ci.upper    p.value      Group 1       Group 2
## [1,]         0 -0.1751743170 0.017605084 0.07345576 0.4855441083  0.6054597822
## [2,]         1  0.0002508617 0.002178409 0.01335559 0.0007132645 -0.0006934263
```

So there is evidence that the nature of the association changes at or near CST.FL close to 70. In particular, it is possible that there is a positive association to about CST.FL=70, and then the strength of the association weakens considerably. The corresponding Pearson correlations are 0.48 and -0.26 and differ significantly at the 0.05 level; based on a robust (Winsorized) correlation, the p-value = 0.007.

However, the apparent curvature might be due in part to outliers among the independent variable. Here is a plot of the regression line when they are removed.

```
rplot(SH[,c(26)],SH[,53],xlab='GORT.FL',ylab='CST.L',xout=T)
```

```
## [1] "A new estimate of the strength of the association is used by default."
## [1] " To get the old estimate, set OLD=TRUE"
```




```
## $n
## [1] 53
##
## $n.keep
## [1] 46
##
## $Strength.Assoc
## [1] 0.44176
##
## $Explanatory.Power
## [1] 0.1951519
##
## $xvals
## NULL
##
## $yhat
```

```
## NULL
```

So again there is evidence of curvature with the bend now near GORT.FL=80.

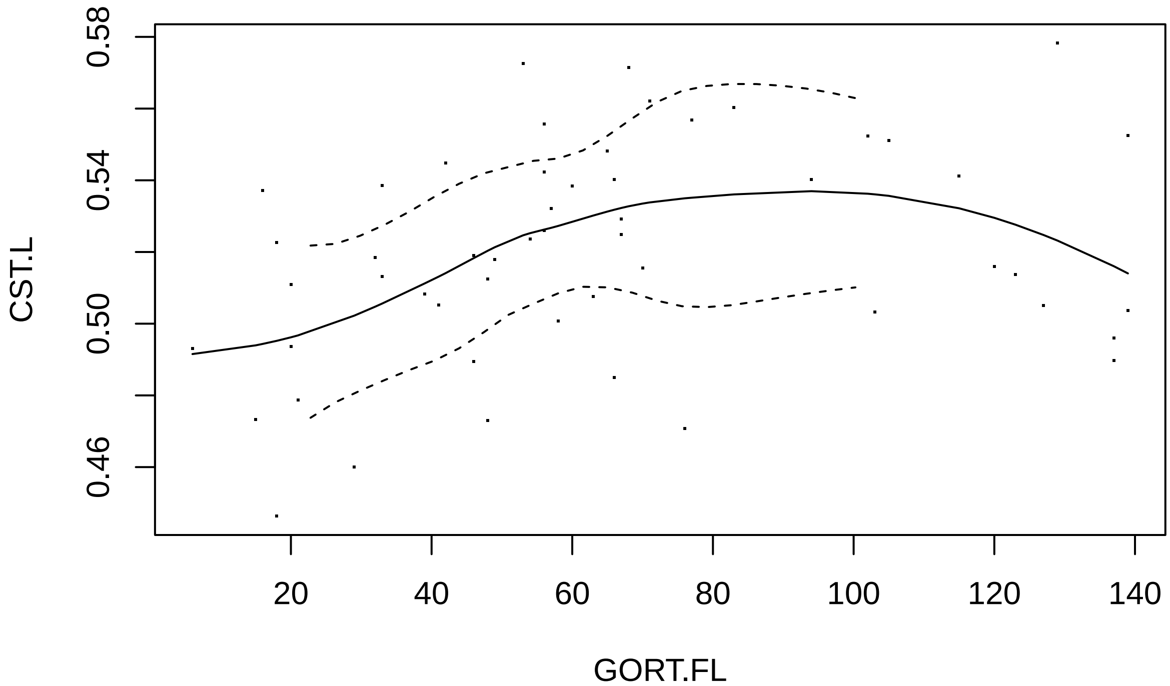
But when these two variables are interchanged, now the regression line appears to be reasonably straight as previously noted. One possible explanation is that there are regression outliers (outliers among the dependent variable), when predicting GORT.FL with CST.L:

```
reglev(SH[,53],SH[,26],plotit=F)
```

```
## $levpoints
## integer(0)
##
## $regout
## [1] 13 22 24 32 33 43 45 50
##
## $bad.lev.points
## integer(0)
##
## $keep
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
## [26] 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
## [51] 51 52 53
##
## $dis
## NULL
##
## $stanres
##  [1] -0.91139415  0.22023708  0.61426837  0.49702366  0.26886867 -0.61426837
##  [7] -0.42080795  0.85408108  1.31838348 -0.23997650 -1.03771440  0.60489318
## [13]  3.92573156 -0.00315712  0.42653410  1.06175756 -1.04979000 -0.82785629
## [19]  0.02647875  0.51229430 -1.26040289  4.17056714 -1.67235156  4.02525514
## [25] -0.12920003  1.49144314 -0.07657844  0.15675194  0.05290223 -0.45121090
## [31]  0.15352576  3.06812889  2.80465050 -0.21941094 -0.08636939  0.13187407
## [37]  0.06623673  1.50038601  0.94611949  0.81343457 -0.10304668 -1.03400043
## [43]  3.42430752  1.35311499  2.89897793  1.81854586  2.12110473  0.61426837
## [49] -0.51272331  2.52462696  2.21227660 -0.52964012  0.04563648
##
## $crit
## [1] 2.241403
```

It is noted that if instead lplotCI is used, and outliers are retained, there appears to be curvature for the higher GORT.FL values:

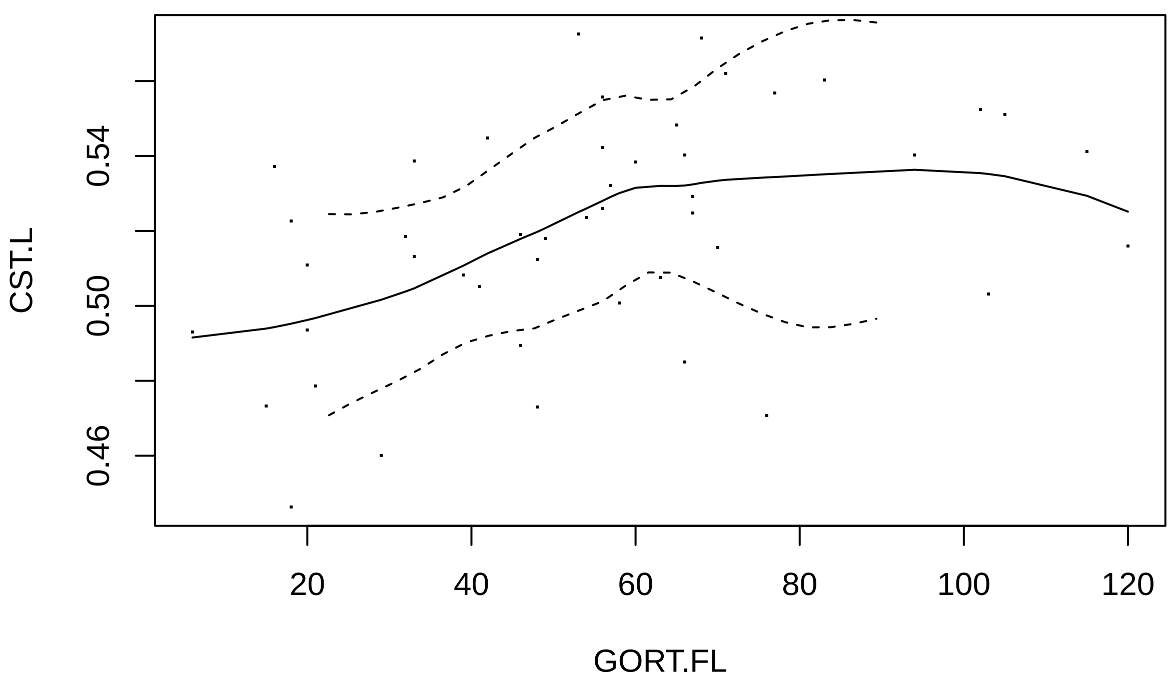
```
lplotCI(SH[,c(26)],SH[,53],xlab='GORT.FL',ylab='CST.L',xout=F)
```




```
## $p.crit
## [1] 0.002546526
##
## $Conf.Intervals
## NULL
```

Here is the result when outliers among the independent variable are removed.

```
lplotCI(SH[,c(26)],SH[,53],xlab='GORT.FL',ylab='CST.L',xout=T)
```




```
## $p.crit
## [1] 0.002660931
```

```
##
## $Conf.Intervals
## NULL
```

This provides another illustration of why it can be important to always check on the impact of removing outliers.

Compare the correlation between Age and WJ.WA.R (word attack) versus GORT.FL.R.

```r
TWOpov(SH[,c(3,26)],SH[,40])
```

```
## $est.rho1
## [1] 0.6839788
##
## $est.rho2
## [1] 0.7998406
##
## $dif
## [1] -0.1158618
##
## $ci
## [1] -0.240438710 -0.009646211
```

Now compare these two IVs when both are included in the model.

```r
regIVcom(SH[,c(3,26)],SH[,40])
```

```
## $n
## [1] 73
##
## $n.keep
## [1] 73
##
## $est.1
## [1] 0.05051815
##
## $est.2
## [1] 18.08512
##
## $e.pow1
## [1] 0.00145985
##
## $e.pow2
## [1] 0.5226155
##
## $strength.assoc.1
## [1] 0.03820799
##
## $strength.assoc.2
## [1] 0.7229215
##
## $ratio
## [1] 0.002793354
##
## $strength.ratio
## [1] 0.05285219
##
```

```
## $p.value
## [1] 9e-04
```

Using least squares, switching from a homoscedastic method to a heteroscedastic method might make a substantial difference: Here is an example dealing with age and GORT standardized rate score.

```r
ols(SH[,c(3)],SH[,23]) # Use a hom. method
```

```
## $n
## [1] 74
##
## $n.keep
## [1] 74
##
## $summary
##              Estimate Std. Error  t value     Pr(>|t|)     low.ci      up.ci
## (Intercept) 7.4923749  1.4579528 5.138969 2.276934e-06  4.5859991 10.3987507
## x           0.2070691  0.1198211 1.728152 8.824772e-02 -0.0317899  0.4459281
##
## $coef
## (Intercept)           x
##   7.4923749   0.2070691
##
## $F.test
##   value
## 2.98651
##
## $Ftest.p.value
##      value
## 0.08824772
##
## $F.test.degrees.of.freedom
## numdf dendf
##     1    72
##
## $R.squared
## [1] 0.0398273
##
## $residuals
##  [1] -2.3913423  0.2298650 -4.5984114 -4.3913423  2.2652105  0.6086577
##  [7]  2.4369341  3.1945195  4.8510723 -3.7701350 -0.9418586  4.4015886
## [13] -0.5630659  4.4015886 -0.1842732 -0.3913423 -0.1489277 -3.3913423
## [19] -1.9772041 -1.1489277 -1.1489277  1.6086577 -1.9418586 -2.3559968
## [25] -0.3559968  5.6086577 -2.1842732 -0.3913423 -1.5630659 -2.7347895
## [31] -3.3913423  1.8510723  1.4369341 -0.9772041  0.2652105  1.0227959
## [37]  3.1945195 -7.7701350  5.6086577  4.8510723 -1.9418586  0.2298650
## [43]  1.8510723  5.8157268  0.4369341  4.6440032 -5.7701350 -2.3913423
## [49]  0.4369341  3.6086577  3.1945195 -3.3913423 -3.3913423 -4.3913423
## [55] -0.5630659  2.2298650  4.8157268 -0.1842732 -0.9772041 -0.3913423
## [61] -6.3913423 -4.8054805 -4.3913423  2.1945195 -0.8054805  3.4015886
## [67]  2.1945195  2.4015886  2.2652105  0.2652105  0.1945195  1.8157268
## [73]  2.4369341 -3.7701350
```

```r
olshc4(SH[,c(3)],SH[,23]) # Use a het. method
```

```
## $n
```

```
## [1] 74
##
## $n.keep
## [1] 74
##
## $ci
##             Coef. Estimates     ci.lower  ci.upper      p-value Std.Error
## (Intercept)     0 7.4923749  5.056911344 9.9278385 4.196988e-08 1.2217246
## Slope           1 0.2070691 -0.008339671 0.4224779 5.929869e-02 0.1080575
##
## $cov
##            [,1]        [,2]
## [1,]  1.4926111 -0.12653223
## [2,] -0.1265322  0.01167643
##
## $test.stat
## [1] 1.916286
##
## $R.squared
## [1] 0.0398273
```

Here, the p-value drops when using a heteroscedastic method, but the reverse can happen because the homoscedastic method uses an incorrect estimate of the standard error when there is homoscedasticity.