Box-cox transformation

Laura Chen

3/19/2021

```
# load data
library(readxl)
library(knitr)
library(broom)
library(gtsummary)
```

```
## #BlackLivesMatter
```

```
library(pastecs)
df = read_xlsx("~/desktop/FinProject/Analysis/IPO_data.xlsx")
df = subset(df, select=-c(search_avail, search_vol_avg,
                    original_high_filing_price,original_low_filing_price,
                    upward_adjustment, offer_price, closing_price))
df = df[!duplicated(df), ] # drop duplicates
```

```
# use min-max scaling to normalize data
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
df_numeric = df[sapply(df,is.numeric)]
preproc2 <- preProcess(df_numeric, method=c("range"))

norm2 <- predict(preproc2, df_numeric)
norm2 = norm2[!(norm2$underpricing == 0),] # get rid of zero value in underpricing
```

```
# OLS no transformation
model = lm(underpricing~., data=norm2)

# OLS with log(x+1) transformation on underpricing // use in paper
df_trans = norm2
df_trans$underpricing = log1p(df_trans$underpricing)
model_trans = lm(underpricing~., data=df_trans)
summary(model_trans)
```

```
##
## Call:
```

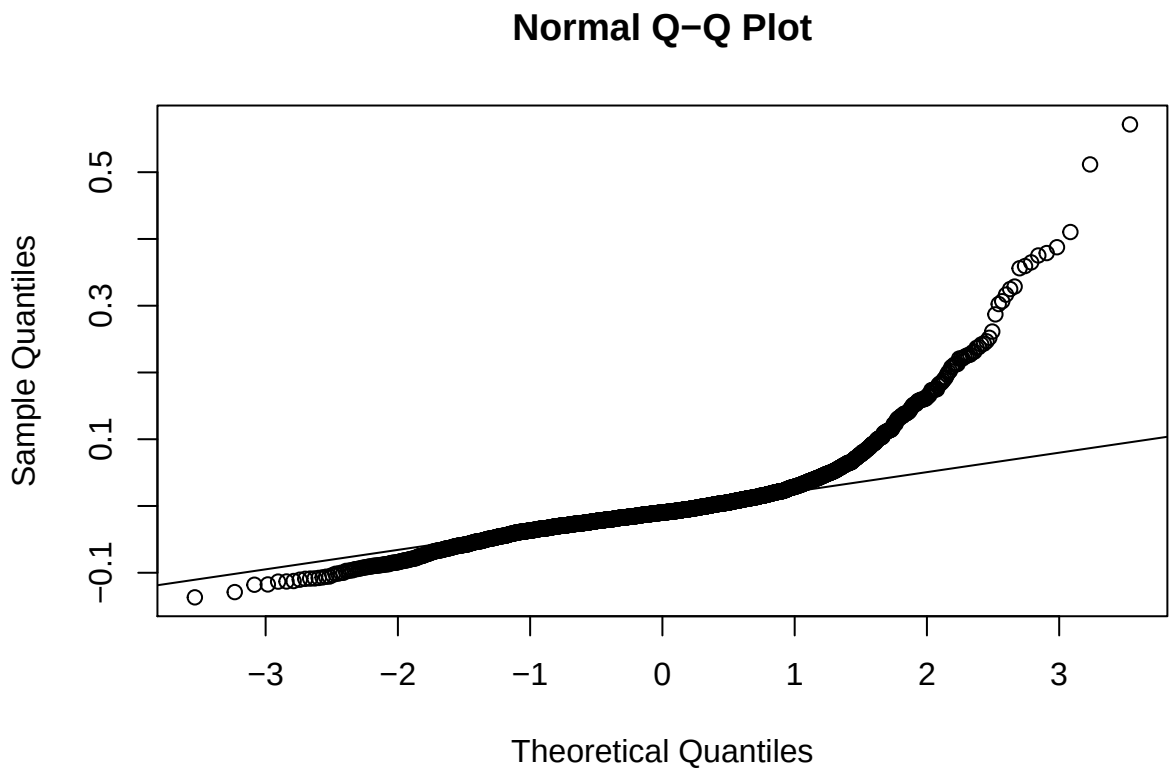
```
## lm(formula = underpricing ~ ., data = df_trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.13682 -0.02715 -0.00985  0.01214  0.57154
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                0.048588   0.011448   4.244 2.27e-05 ***
## proceeds_amt_mil           0.418478   0.111598   3.750 0.000181 ***
## primary_shares_offered    -0.166590   0.048022  -3.469 0.000531 ***
## secondary_shares_offered -0.438298   0.116291  -3.769 0.000168 ***
## venture_backed             0.016174   0.002661   6.078 1.41e-09 ***
## num_bookrunners            0.018107   0.030334   0.597 0.550614
## rank_no_leads             -0.037337   0.029032  -1.286 0.198549
## num_lead_colead_managers   0.003010   0.021613   0.139 0.889261
## c1                         0.005132   0.009341   0.549 0.582781
## c2                        -0.002004   0.008630  -0.232 0.816372
## c3                         0.033189   0.015366   2.160 0.030880 *
## c4                        -0.007388   0.011312  -0.653 0.513752
## word_length_sentiment     -0.065132   0.014521  -4.485 7.62e-06 ***
## negative                  -0.020151   0.010002  -2.015 0.044050 *
## positive                  -0.030624   0.008895  -3.443 0.000585 ***
## uncertainty               -0.022336   0.008381  -2.665 0.007752 **
## litigious                 -0.036727   0.017446  -2.105 0.035383 *
## strongmodal               -0.034777   0.021315  -1.632 0.102902
## weakmodal                  0.019575   0.012062   1.623 0.104738
## constraining               0.001686   0.009240   0.182 0.855212
## internet                   0.048350   0.003358  14.400  < 2e-16 ***
## nasdaq_returns             0.091782   0.014657   6.262 4.48e-10 ***
## vix_returns                0.045774   0.014849   3.083 0.002075 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05757 on 2431 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.2009, Adjusted R-squared:  0.1937
## F-statistic: 27.79 on 22 and 2431 DF,  p-value: < 2.2e-16
```

```
kable(tidy(summary(model_trans)))
```

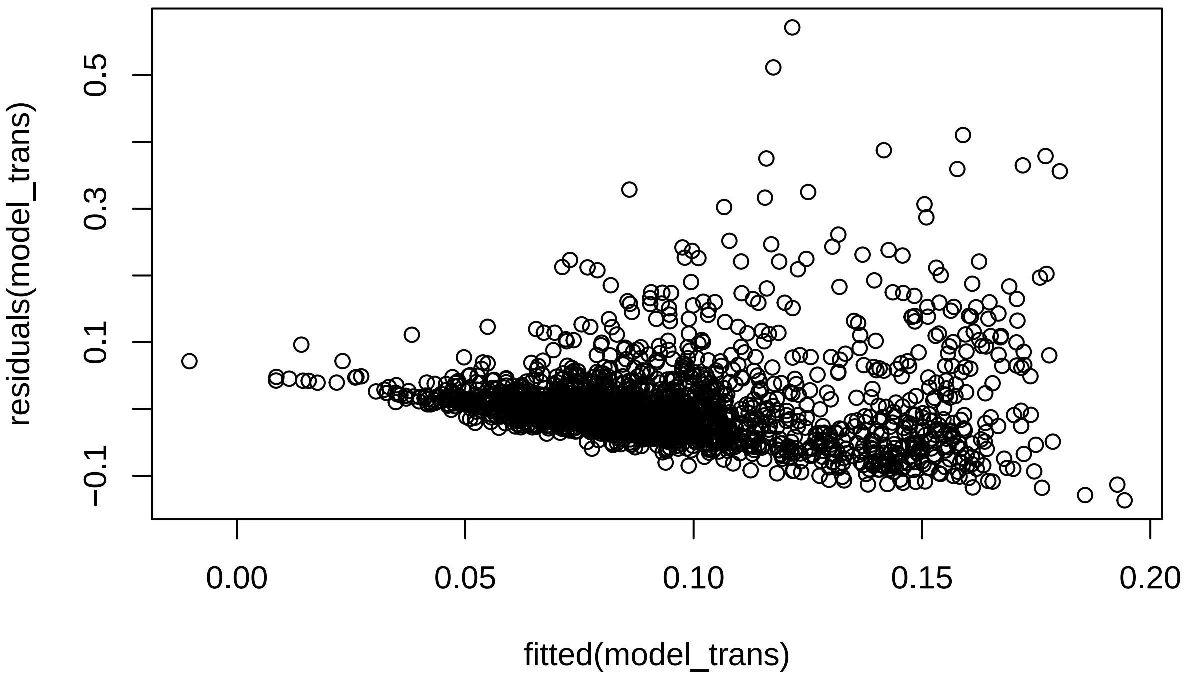
term	estimate	std.error	statistic	p.value
(Intercept)	0.0485881	0.0114478	4.2443331	0.0000227
proceeds_amt_mil	0.4184779	0.1115978	3.7498754	0.0001811
primary_shares_offered	-0.1665902	0.0480222	-3.4690272	0.0005314
secondary_shares_offered	-0.4382982	0.1162913	-3.7689672	0.0001678
venture_backed	0.0161736	0.0026610	6.0780817	0.0000000
num_bookrunners	0.0181073	0.0303343	0.5969239	0.5506138
rank_no_leads	-0.0373368	0.0290323	-1.2860464	0.1985494
num_lead_colead_managers	0.0030097	0.0216130	0.1392543	0.8892607
c1	0.0051320	0.0093410	0.5493999	0.5827815
c2	-0.0020043	0.0086303	-0.2322384	0.8163724
c3	0.0331888	0.0153661	2.1598777	0.0308796

term	estimate	std.error	statistic	p.value
c4	-0.0073877	0.0113116	-0.6531019	0.5137524
word_length_sentiment	-0.0651320	0.0145213	-4.4852863	0.0000076
negative	-0.0201508	0.0100021	-2.0146661	0.0440496
positive	-0.0306242	0.0088947	-3.4429679	0.0005851
uncertainty	-0.0223358	0.0083815	-2.6648908	0.0077522
litigious	-0.0367269	0.0174465	-2.1051191	0.0353833
strongmodal	-0.0347769	0.0213151	-1.6315592	0.1029019
weakmodal	0.0195755	0.0120619	1.6229107	0.1047382
constraining	0.0016862	0.0092398	0.1824919	0.8552120
internet	0.0483503	0.0033576	14.4004393	0.0000000
nasdaq_returns	0.0917817	0.0146567	6.2620987	0.0000000
vix_returns	0.0457736	0.0148489	3.0826287	0.0020748

```
#QQ plot
qqnorm(model_trans$residuals)
qqline(model_trans$residuals)
```




```
plot(residuals(model_trans)~fitted(model_trans))
```




```r
kable(summary(df_trans, digits=4))
```

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :0.01367
1st Qu.:0.0023	1st Qu.:0.1270	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0606	1st Qu.:0.0000	1st Qu.:0.1540	1st Qu.:0.0000	1st Qu.:0.0506	1st Qu.:0.1170	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.1774	1st Qu.:0.0000	1st Qu.:0.0577	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.3905	1st Qu.:5.2174	1st Qu.:0.057[illegible]
Median :0.0043	Median :0.0195	Median :7.0000	Median :0.0000	Median :0.0000	Median :0.0000	Median :0.0909	Median :1.1295	Median :5.2403	Median :3.0000	Median :0.1170	Median :0.1373	Median :0.0000	Median :0.0000	Median :0.2651	Median :1.0000	Median :0.0869	Median :6.1088	Median :8.1370	Median :9.0000	Median :0.4496	Median :6.2741	Median :1.07182
Mean :0.0100	Mean :0.0313	Mean :3.0038	Mean :9.5465	Mean :0.0540	Mean :0.0767	Mean :0.0990	Mean :5.1448	Mean :8.2611	Mean :1.0600	Mean :5.1320	Mean :0.1586	Mean :0.0761	Mean :9.1014	Mean :4.2768	Mean :8.0400	Mean :0.0694	Mean :5.1185	Mean :5.1491	Mean :1.1611	Mean :1.4461	Mean :1.2815	Mean :5.09168
3rd Qu.:0.0079	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:1.1480	3rd Qu.:[illegible]	3rd Qu.:1.0000	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:0.1044	3rd Qu.:0.1932	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:[illegible]	3rd Qu.:0.096[illegible]
Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :1.0000	Max. :0.69315
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA's :4	NA

```r
# boxcox transformation
# (source: https://www.statology.org/box-cox-transformation-in-r/)
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

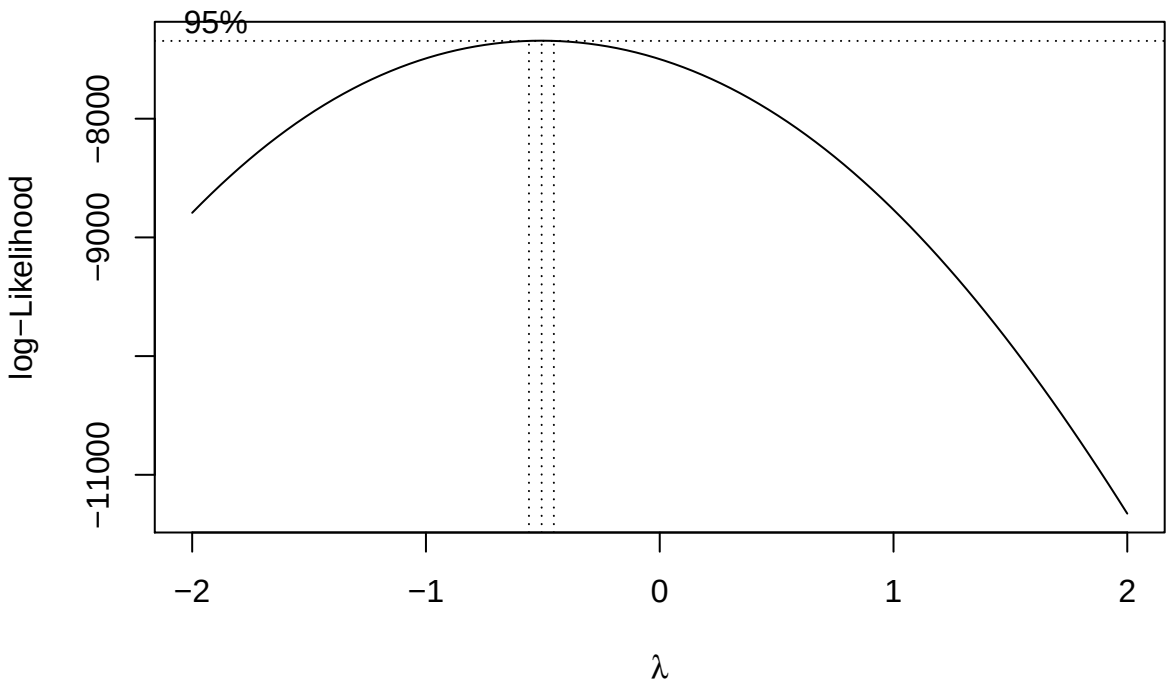
```
## The following object is masked from 'package:gtsummary':
##
##     select
```

```
library(mgcv)
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-34. For overview type 'help("mgcv-package")'.
```

```
bc = boxcox(underpricing~., data=df_trans)
```

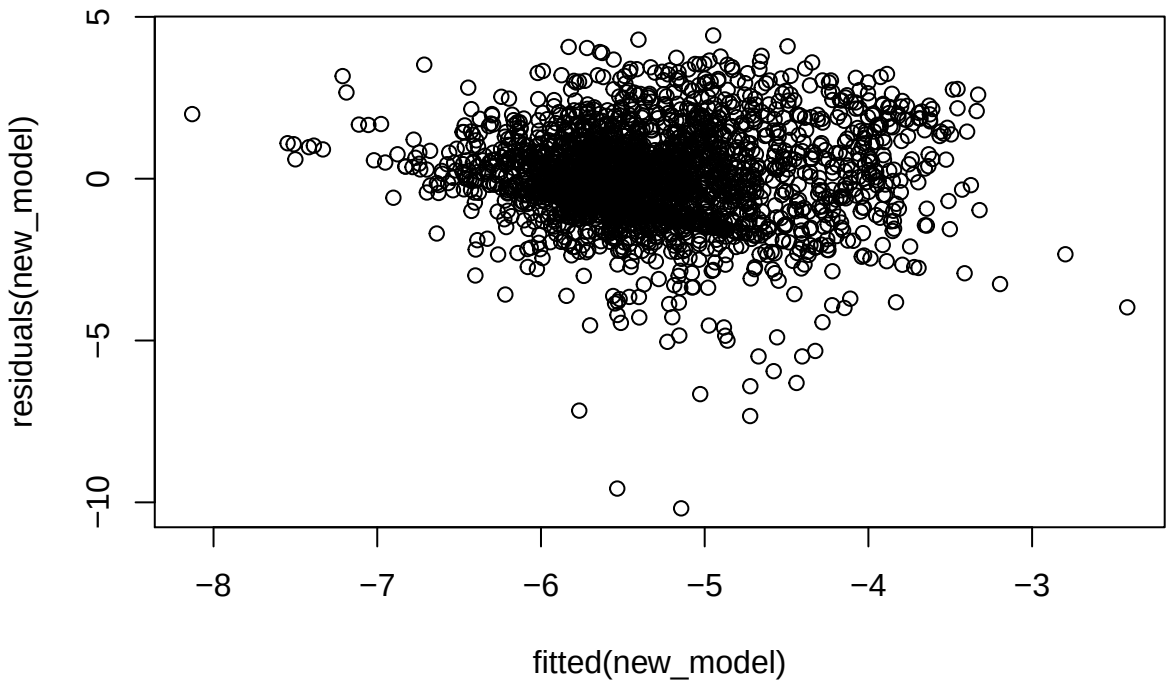



```
lambda <- bc$x[which.max(bc$y)]
new_model <- lm(((underpricing^lambda-1)/lambda) ~., data=df_trans)
summary(new_model)
```

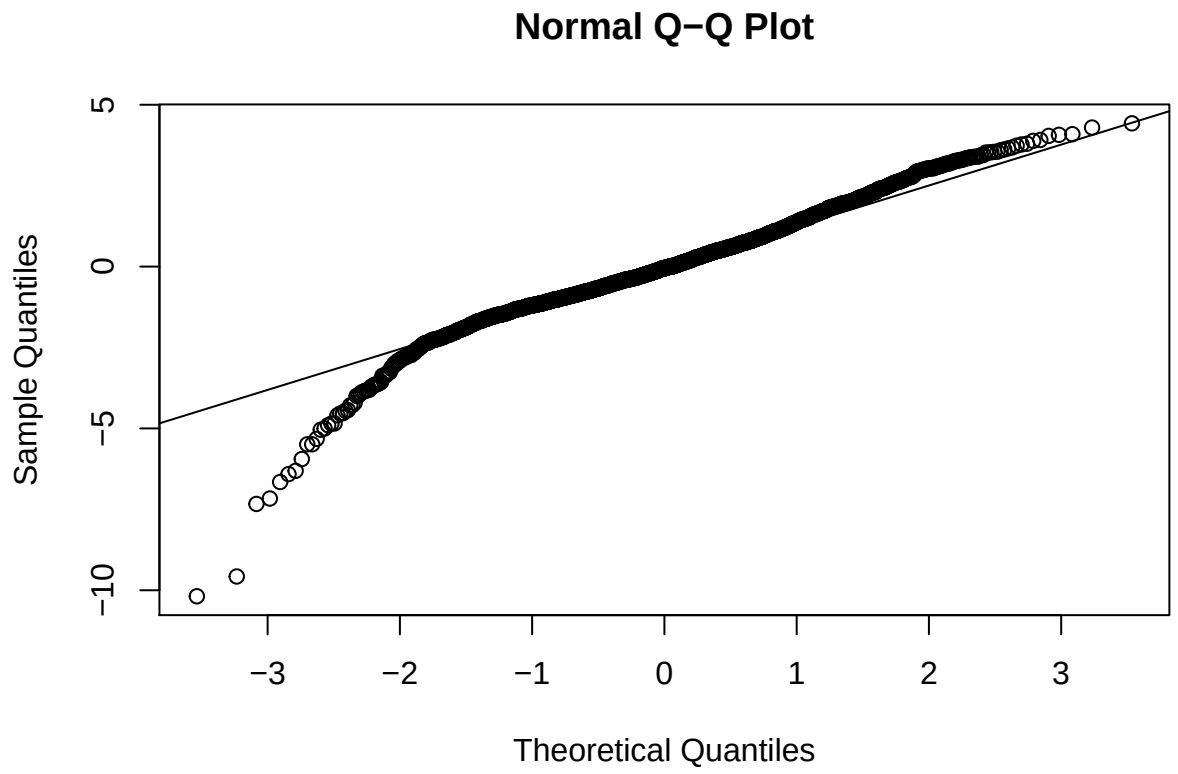
```
##
## Call:
## lm(formula = ((underpricing^lambda - 1)/lambda) ~ ., data = df_trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.1854  -0.8727  -0.0495   0.8313   4.4257
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -6.31795    0.28836 -21.910  < 2e-16 ***
## proceeds_amt_mil           12.38860    2.81102   4.407 1.09e-05 ***
## primary_shares_offered     -5.01380    1.20962  -4.145 3.52e-05 ***
## secondary_shares_offered -12.90267    2.92925  -4.405 1.10e-05 ***
## venture_backed             0.42878    0.06703   6.397 1.89e-10 ***
## num_bookrunners            1.16177    0.76409   1.520 0.128523
## rank_no_leads             -1.32949    0.73129  -1.818 0.069187 .
```

```
## num_lead_colead_managers   0.14876    0.54441   0.273 0.784690
## c1                         0.25206    0.23529   1.071 0.284160
## c2                         0.01836    0.21739   0.084 0.932706
## c3                         0.89640    0.38705   2.316 0.020643 *
## c4                        -0.17290    0.28493  -0.607 0.544020
## word_length_sentiment     -1.94297    0.36577  -5.312 1.18e-07 ***
## negative                  -0.52665    0.25194  -2.090 0.036688 *
## positive                  -0.70712    0.22405  -3.156 0.001618 **
## uncertainty               -0.72106    0.21112  -3.415 0.000647 ***
## litigious                 -0.84681    0.43946  -1.927 0.054104 .
## strongmodal               -0.67698    0.53690  -1.261 0.207468
## weakmodal                  0.37601    0.30383   1.238 0.215997
## constraining               0.08705    0.23274   0.374 0.708415
## internet                   0.86612    0.08457  10.241  < 2e-16 ***
## nasdaq_returns             2.34517    0.36919   6.352 2.52e-10 ***
## vix_returns                1.05866    0.37403   2.830 0.004686 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.45 on 2431 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.1675, Adjusted R-squared:   0.16
## F-statistic: 22.23 on 22 and 2431 DF,  p-value: < 2.2e-16
```

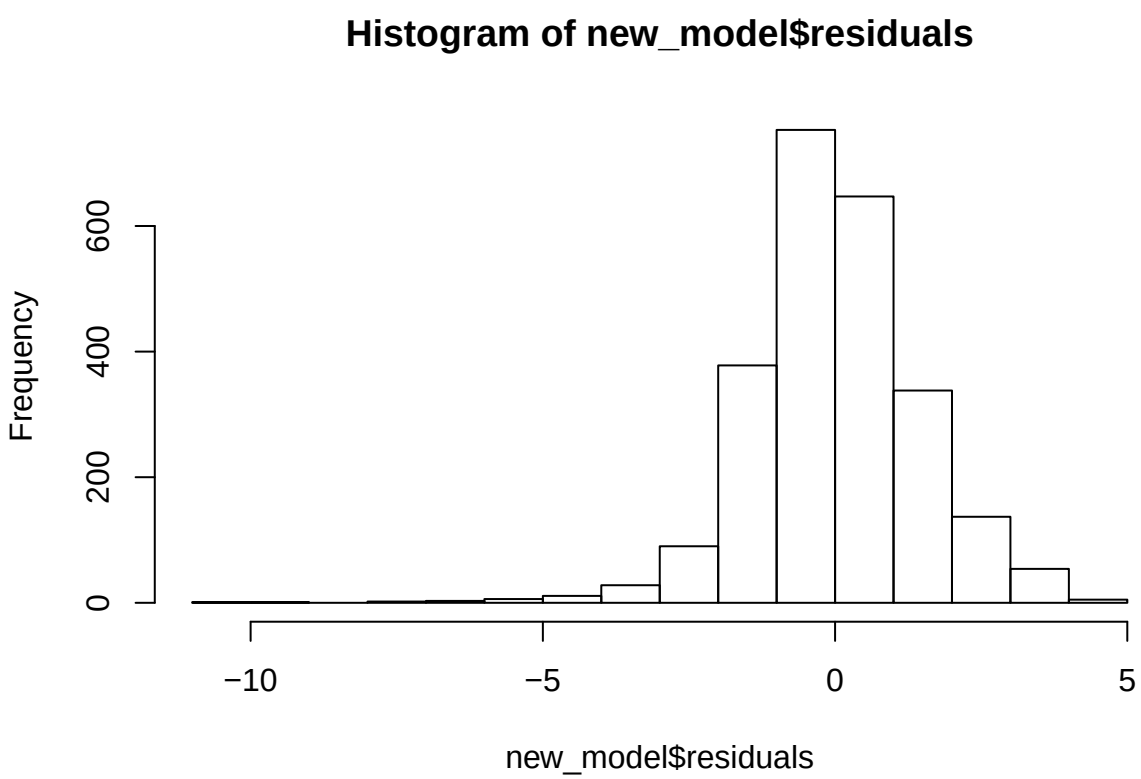
```r
plot(residuals(new_model)~fitted(new_model))
```




```r
#Q-Q plot for Box-Cox transformed model
qqnorm(new_model$residuals)
qqline(new_model$residuals)
```




```
hist(new_model$residuals)
```

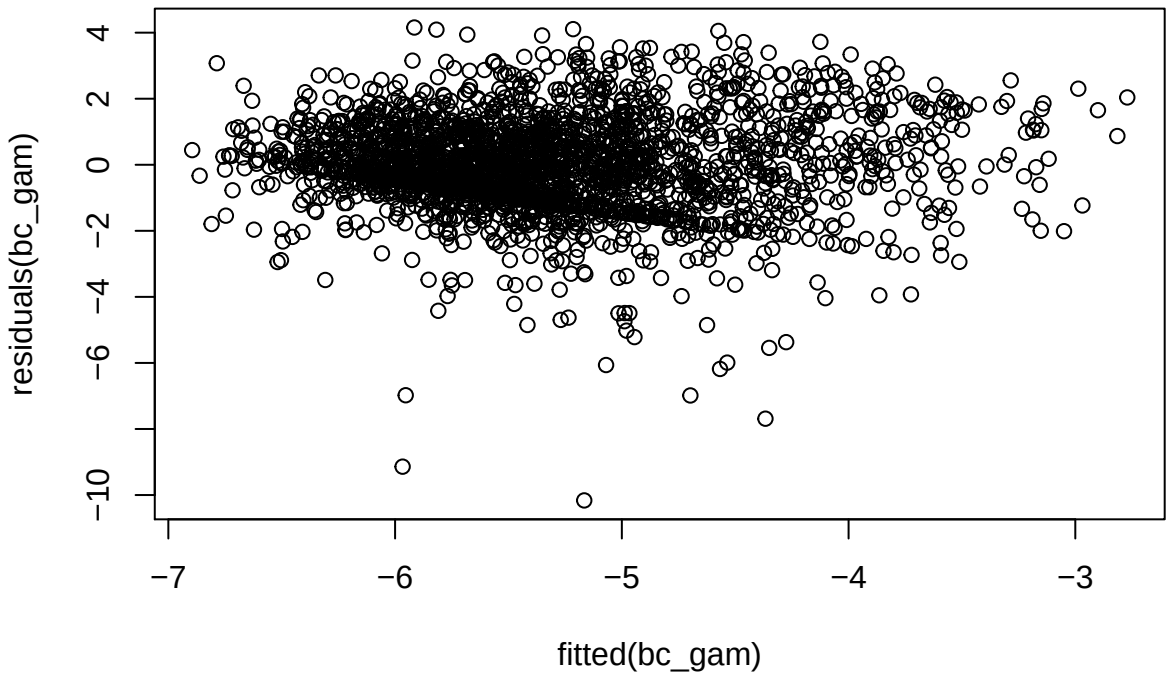


```r
# gam model using bocox transformation and log-transformed underpricing
# FINAL MODEL
bc_gam = gam(formula=((underpricing^lambda-1)/lambda)~venture_backed+
               s(num_bookrunners)+s(rank_no_leads)+
                 s(num_lead_colead_managers)+s(c1)+s(c2)+s(c3)+s(c4)+
                 s(word_length_sentiment)+s(negative)+s(positive)+
                 s(uncertainty)+s(litigious)+s(strongmodal)+s(weakmodal)+
                 s(constraining)+internet+s(nasdaq_returns)+s(vix_returns),
             data=df_trans)
summary(bc_gam)
```

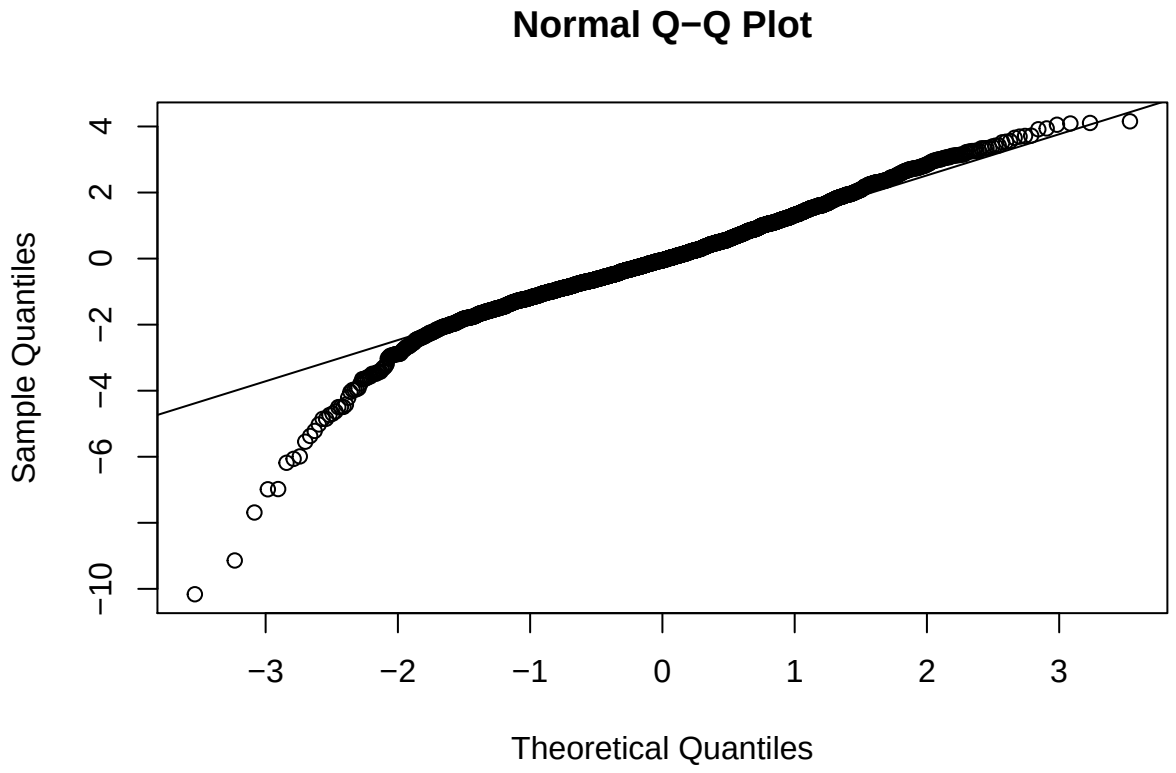
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## ((underpricing^lambda - 1)/lambda) ~ venture_backed + s(num_bookrunners) +
##     s(rank_no_leads) + s(num_lead_colead_managers) + s(c1) +
##     s(c2) + s(c3) + s(c4) + s(word_length_sentiment) + s(negative) +
##     s(positive) + s(uncertainty) + s(litigious) + s(strongmodal) +
##     s(weakmodal) + s(constraining) + internet + s(nasdaq_returns) +
##     s(vix_returns)
##
## Parametric coefficients:
##                Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    -5.60791    0.04347 -128.999  < 2e-16 ***
## venture_backed  0.36129    0.06838    5.283 1.38e-07 ***
## internet        0.76864    0.08374    9.179  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                               edf Ref.df      F  p-value
## s(num_bookrunners)          1.000  1.000  3.261  0.07107 .
## s(rank_no_leads)            3.544  4.333  3.987  0.00290 **
## s(num_lead_colead_managers) 4.711  5.647  2.976  0.00806 **
## s(c1)                       2.331  2.963  2.190  0.07025 .
## s(c2)                       2.075  2.654  1.722  0.18498
## s(c3)                       1.974  2.519  3.139  0.03684 *
## s(c4)                       1.000  1.000  0.196  0.65794
## s(word_length_sentiment)    2.492  3.166  8.123 1.81e-05 ***
## s(negative)                 1.921  2.408  2.851  0.05593 .
## s(positive)                 4.904  5.948  2.640  0.01781 *
## s(uncertainty)              1.000  1.000  9.150  0.00251 **
## s(litigious)                1.000  1.000  1.370  0.24193
## s(strongmodal)              1.000  1.000  2.386  0.12259
## s(weakmodal)                1.814  2.320  1.170  0.34236
## s(constraining)             1.000  1.000  0.000  0.99039
## s(nasdaq_returns)           5.670  6.893 14.543  < 2e-16 ***
## s(vix_returns)              3.216  4.096  6.024 7.28e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## R-sq.(adj) =  0.198   Deviance explained = 21.2%
## GCV = 2.0447  Scale est. = 2.0083    n = 2454
```

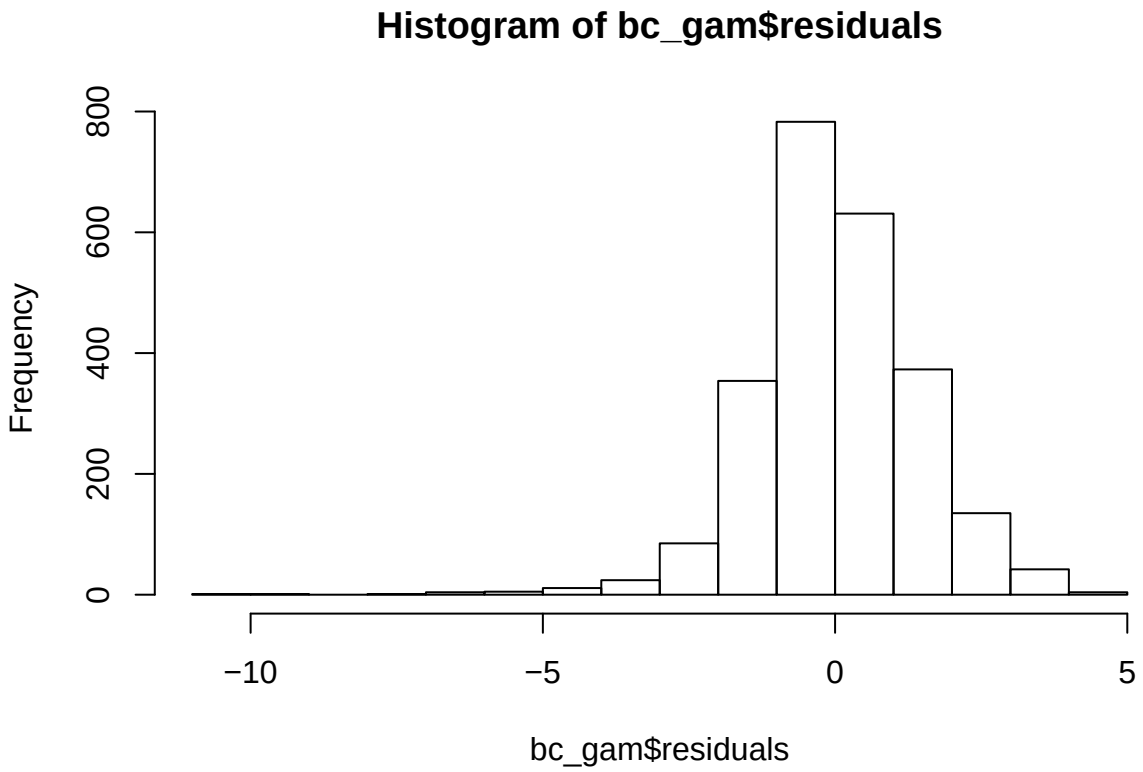
```
plot(residuals(bc_gam)~fitted(bc_gam))
```



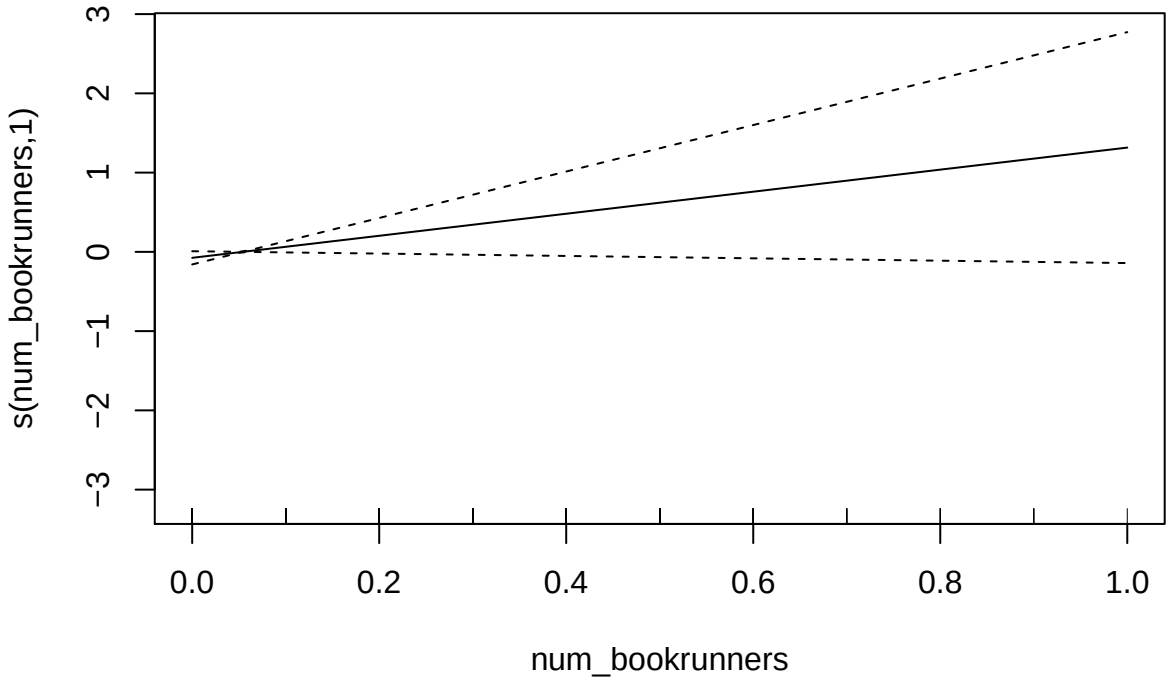

```
qqnorm(residuals(bc_gam))
qqline(residuals(bc_gam))
```

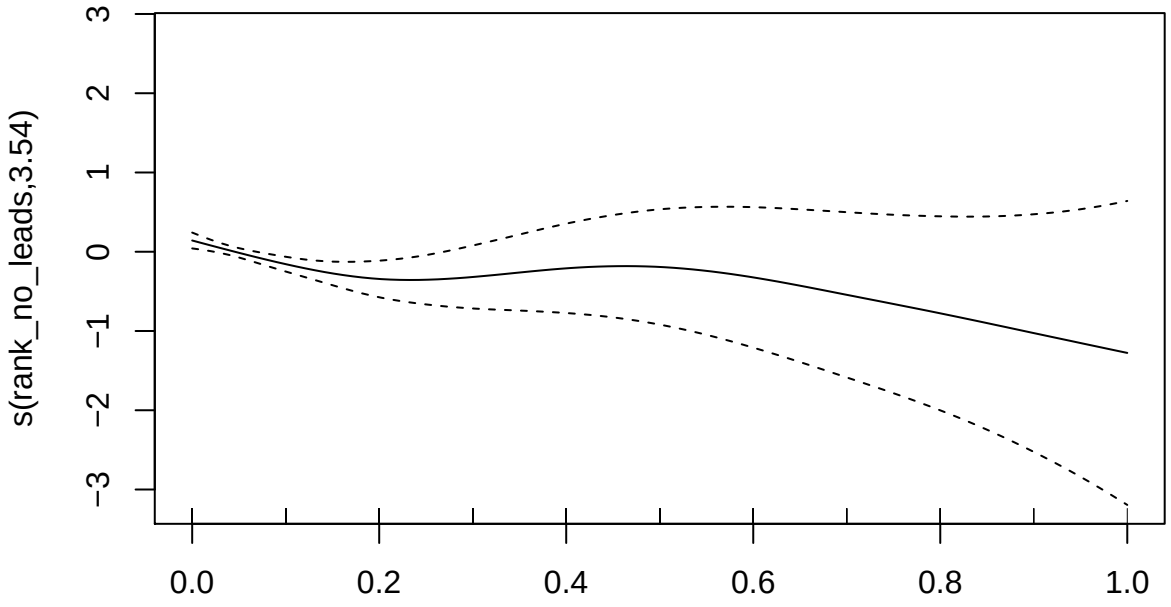


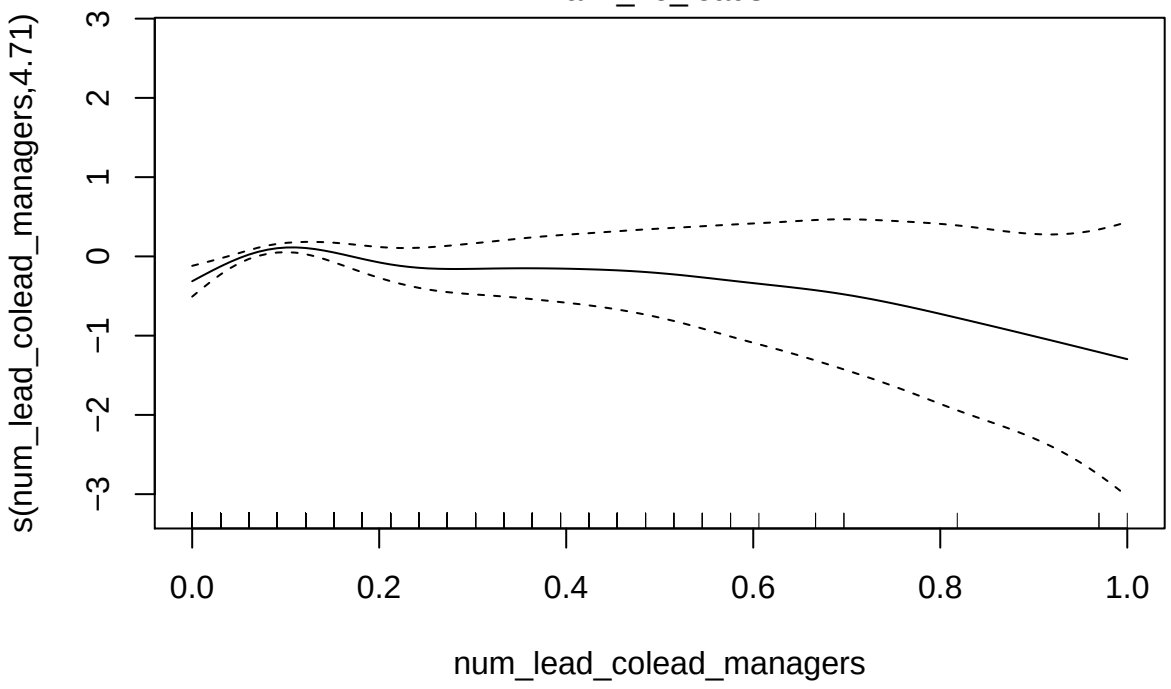
```
hist(bc_gam$residuals)
```

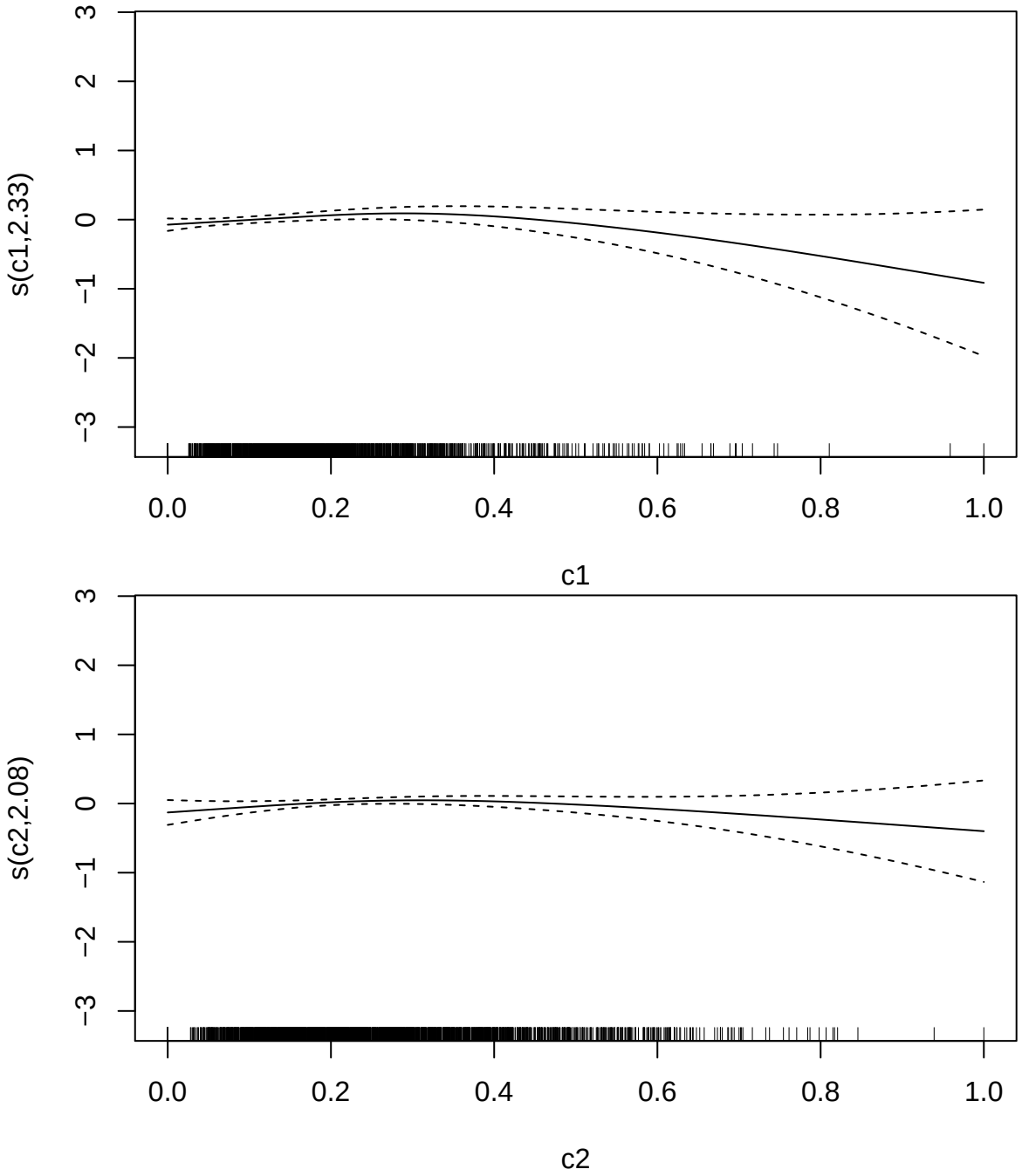


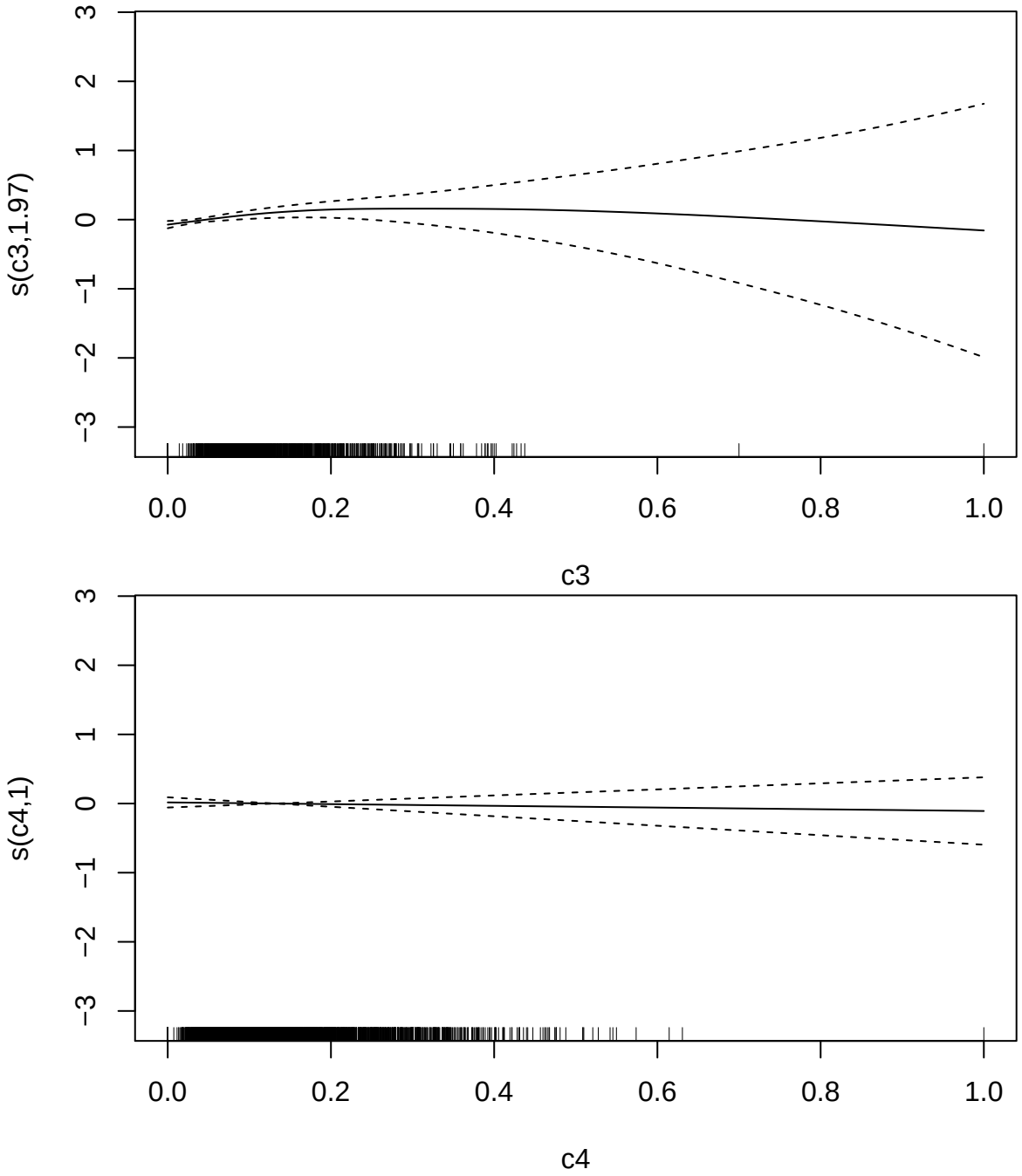

```
plot(bc_gam)
```

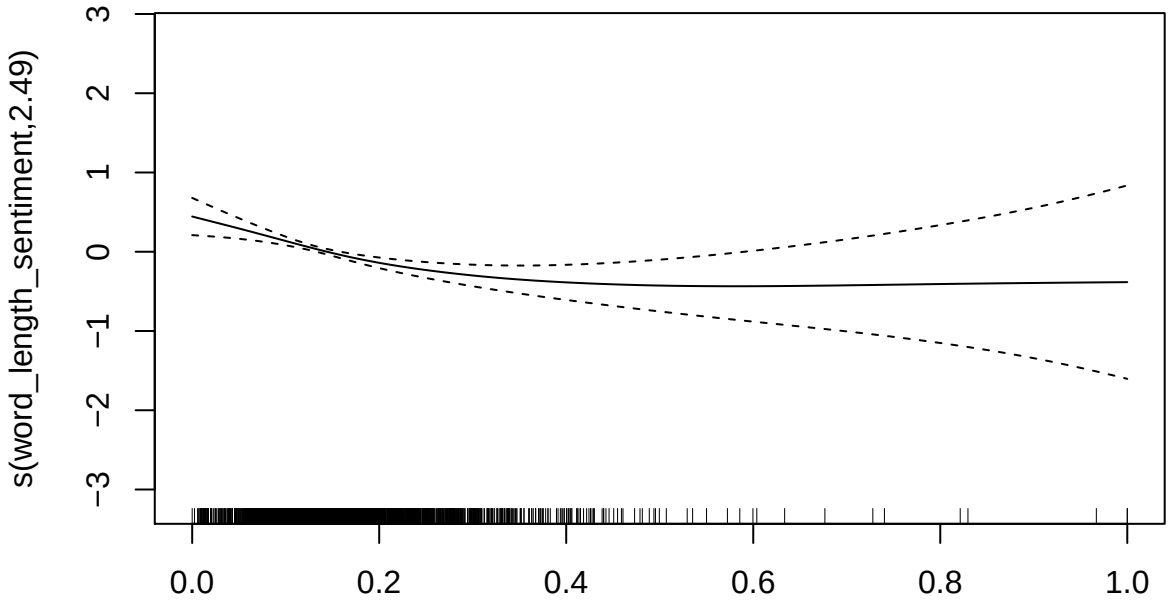


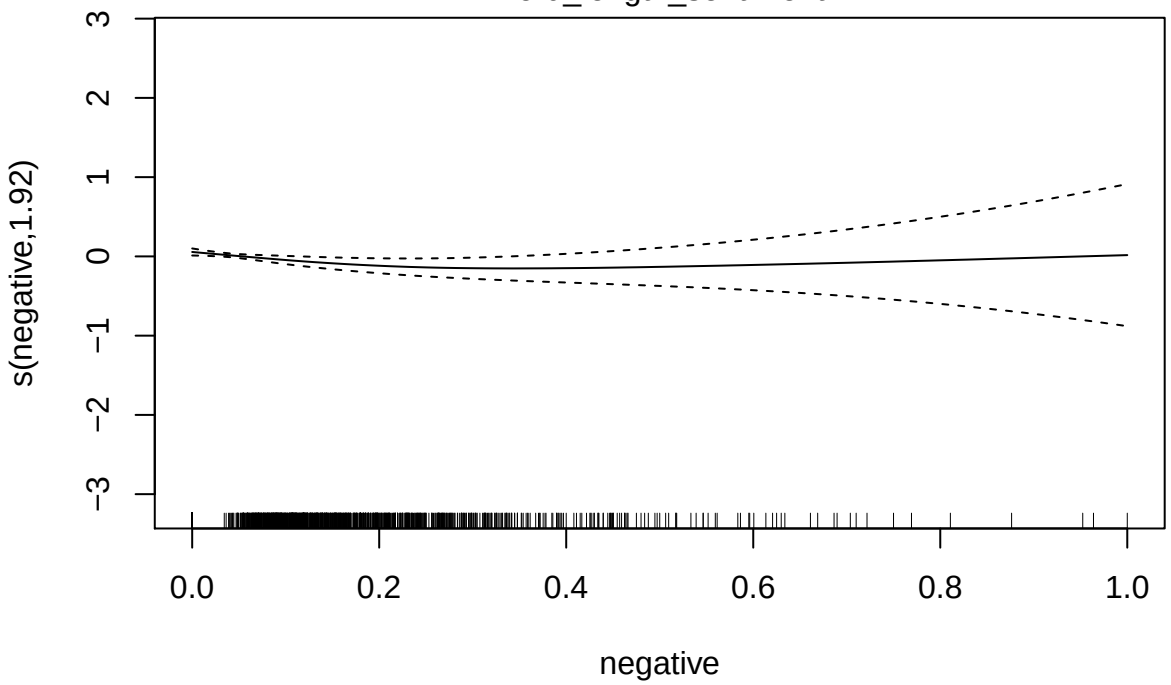

s(rank_no_leads,3.54)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
rank_no_leads

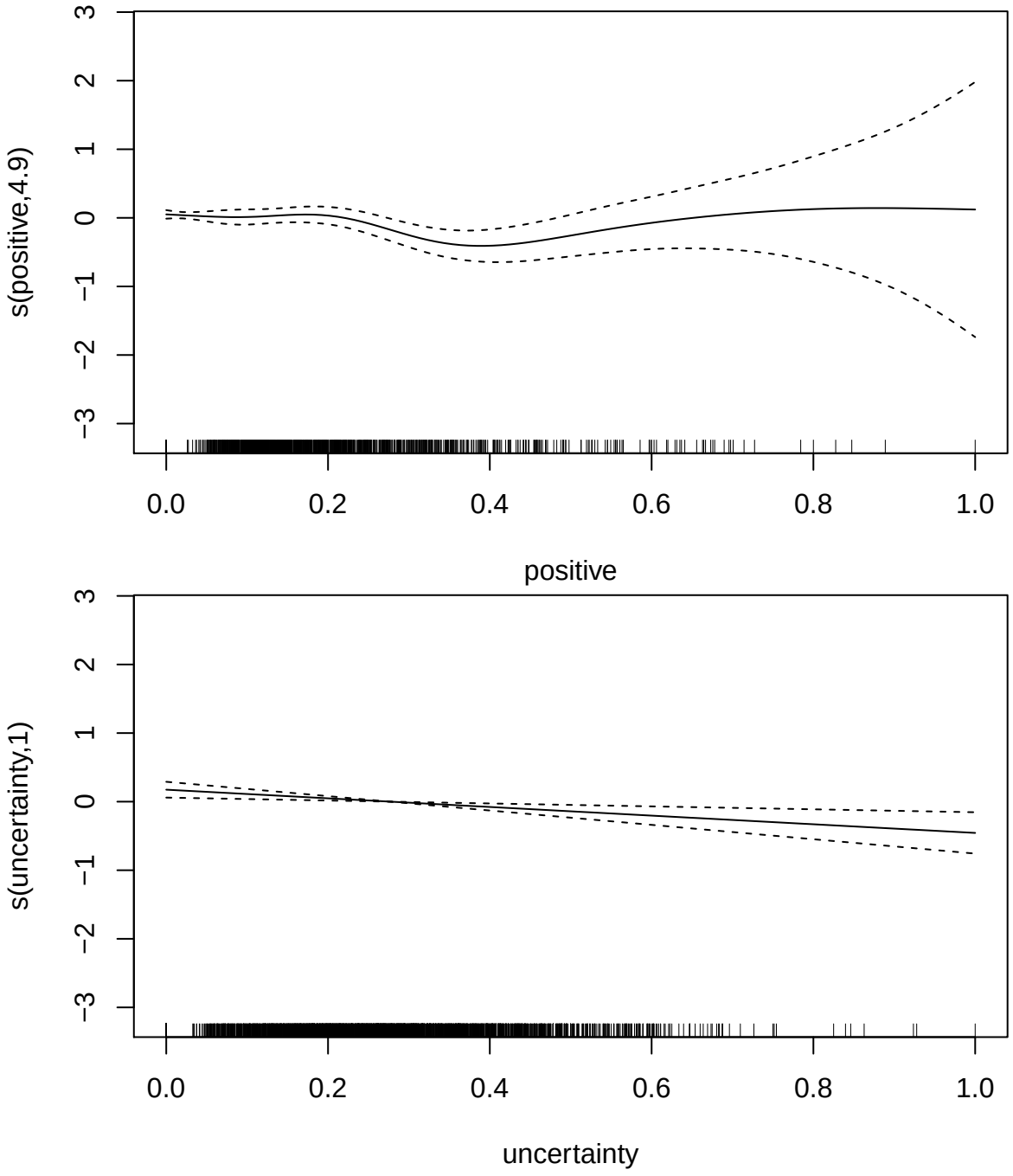

s(num_lead_colead_managers,4.71)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
num_lead_colead_managers

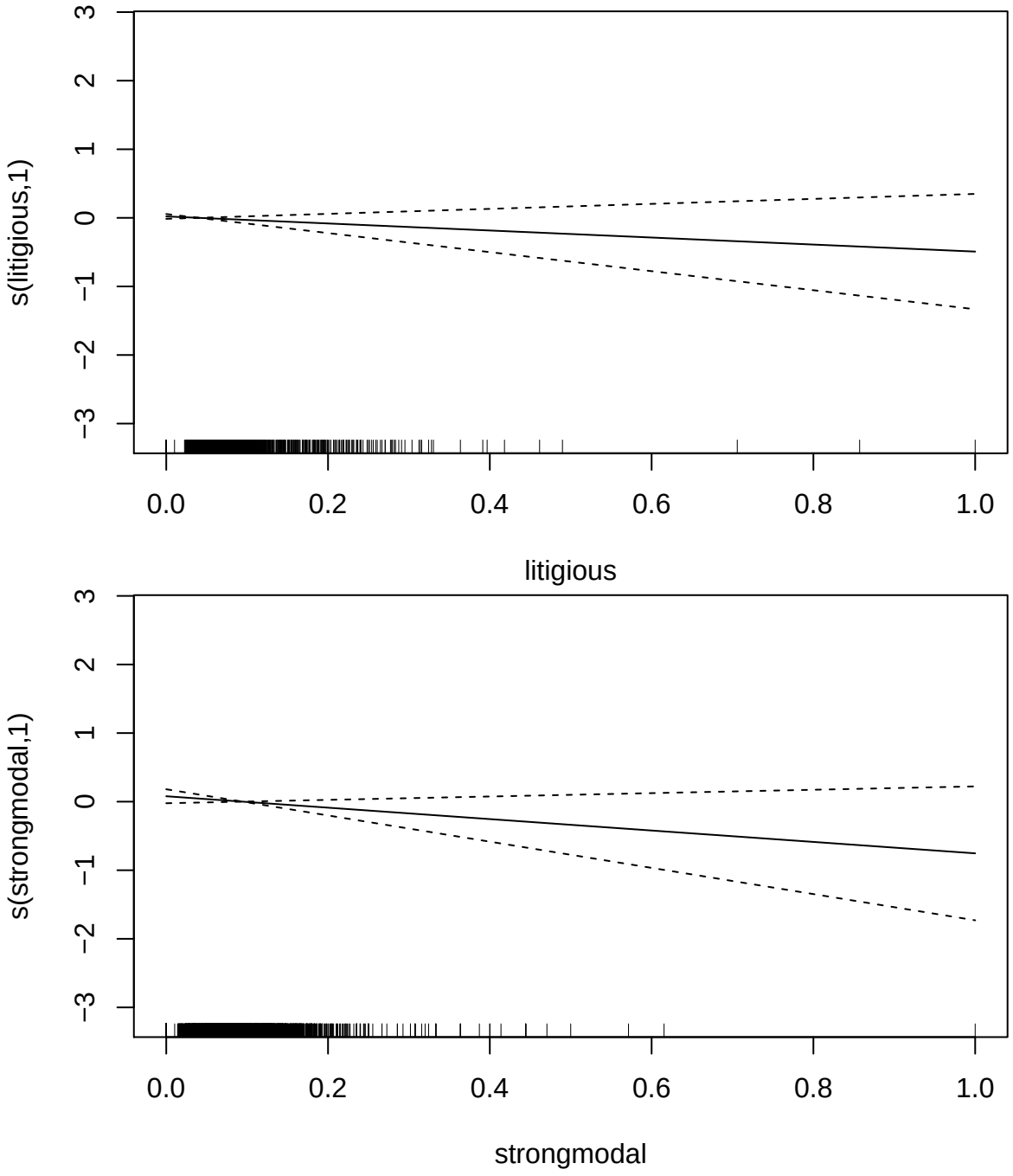

s(c1,2.33)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
c1
s(c2,2.08)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
c2

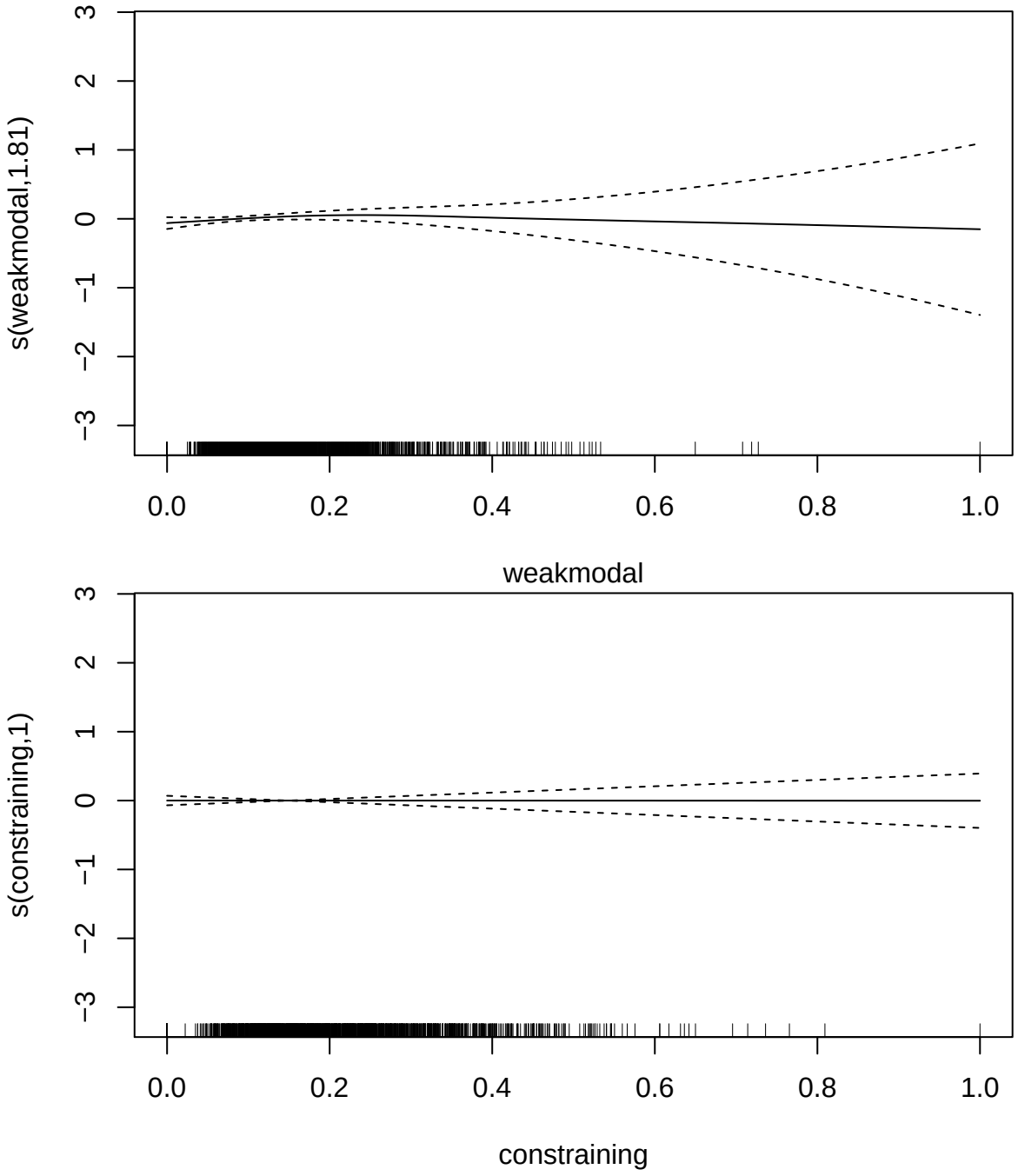

s(c3,1.97)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
c3
s(c4,1)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
c4

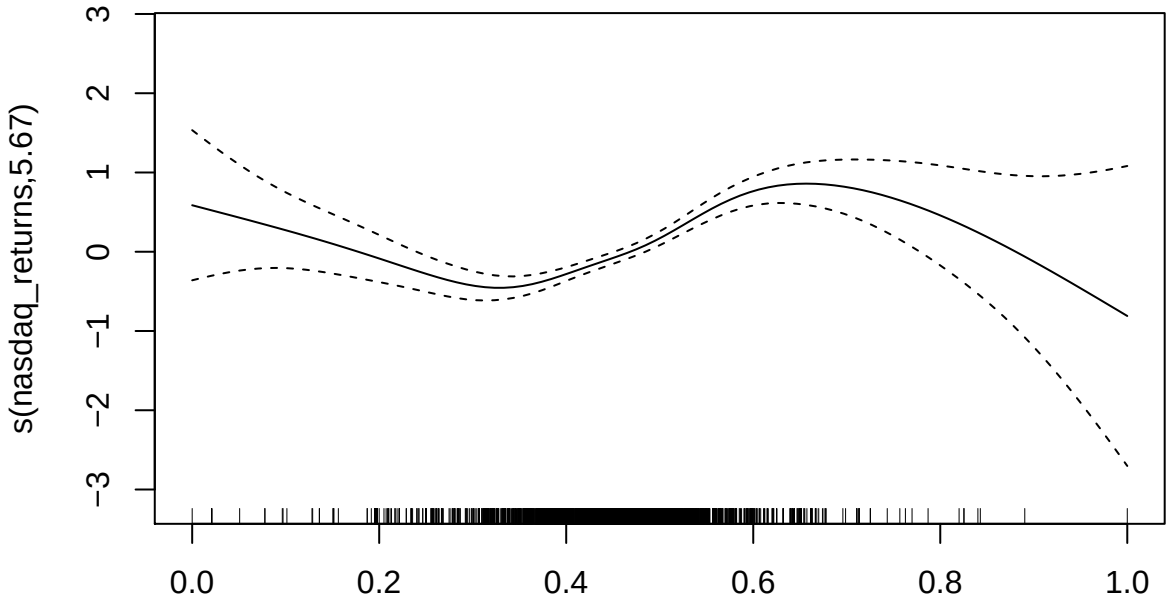

s(word_length_sentiment,2.49)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
word_length_sentiment

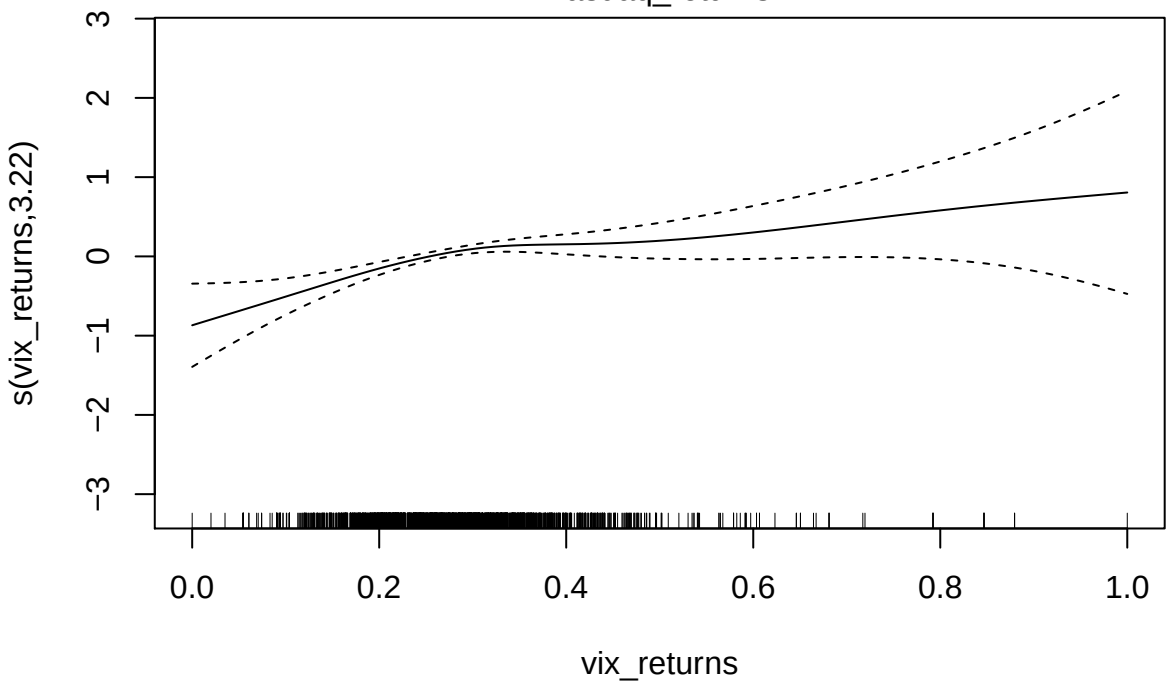

s(negative,1.92)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
negative


s(positive,4.9)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
positive
s(uncertainty,1)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
uncertainty


s(litigious,1)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
litigious
s(strongmodal,1)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
strongmodal

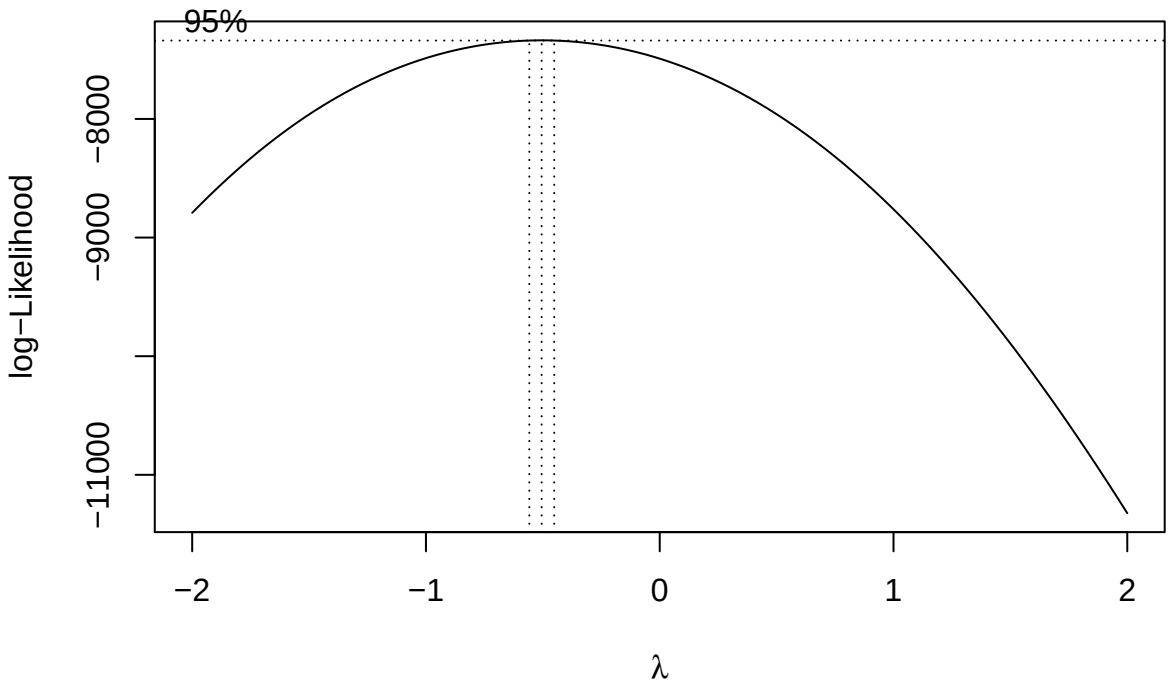

s(weakmodal,1.81)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
weakmodal
s(constraining,1)
3
2
1
0
−1
−2
−3
0.0
0.2
0.4
0.6
0.8
1.0
constraining






```
new = df_trans
new$proceeds_amt_mil = log1p(new$proceeds_amt_mil)
new$primary_shares_offered = log1p(new$primary_shares_offered)
new$secondary_shares_offered = log1p(new$secondary_shares_offered)
new$c3 = log1p(new$c3)
new$c4 = log1p(new$c4)
new$uncertainty = log1p(new$uncertainty)
new$litigious = log1p(new$litigious)
new$strongmodal = log1p(new$strongmodal)
new$weakmodal = log1p(new$weakmodal)
new$constraining = log1p(new$constraining)
```

```r
# boxcox transformation on log transformed response variable and log transformed
# predictor variables
bc_new = boxcox(underpricing~., data=new)
```

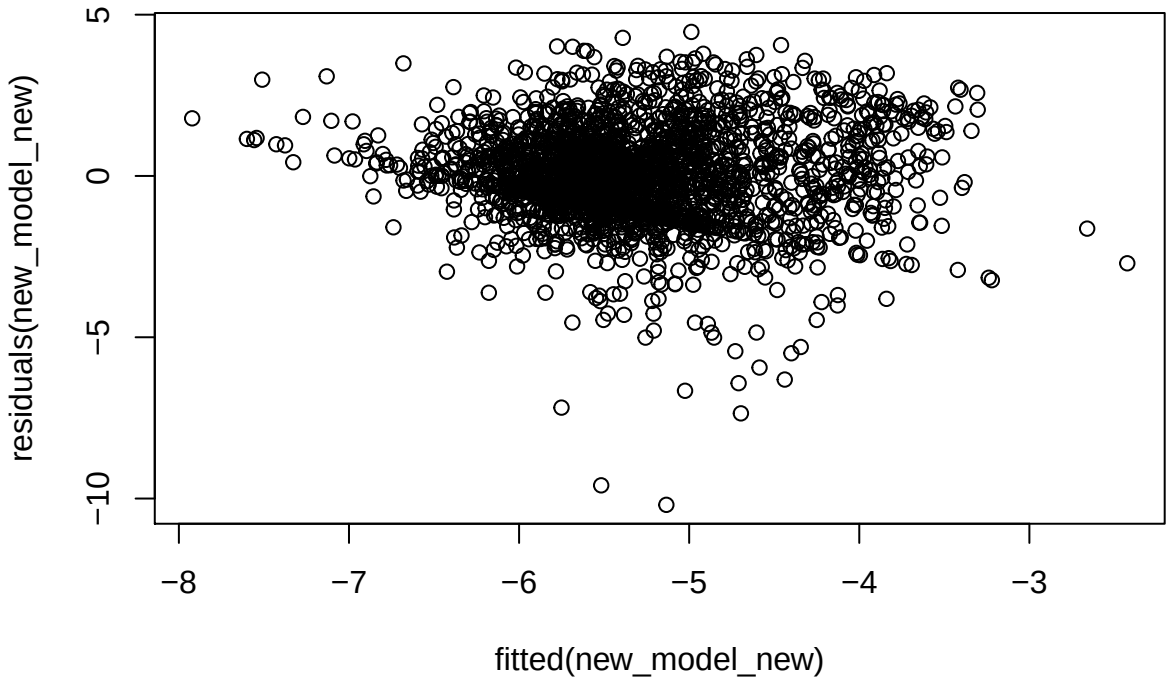



```r
lambda_new <- bc_new$x[which.max(bc_new$y)]
new_model_new <- lm(((underpricing^lambda_new-1)/lambda_new) ~., data=new)
summary(new_model_new)
```

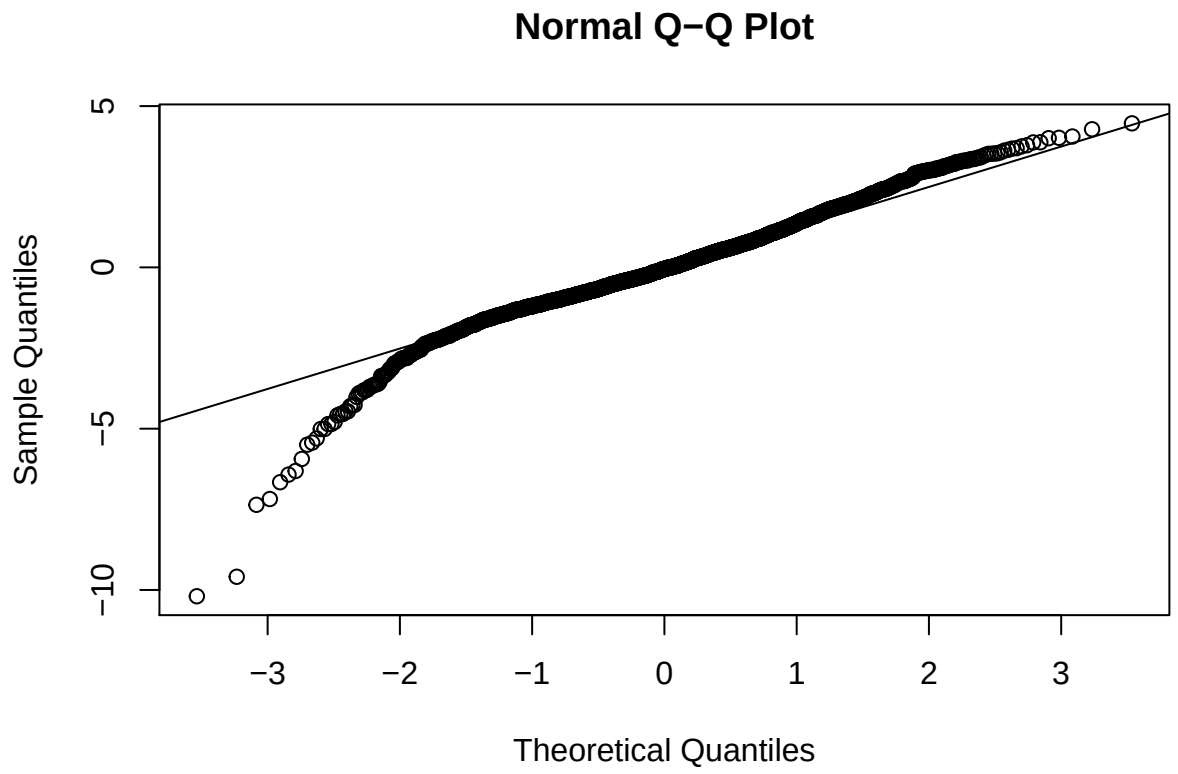
```
##
## Call:
## lm(formula = ((underpricing^lambda_new - 1)/lambda_new) ~ .,
##     data = new)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.1947  -0.8561  -0.0487   0.8334   4.4649
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -6.27845    0.29043 -21.618  < 2e-16 ***
## proceeds_amt_mil           19.21928    3.54210   5.426 6.33e-08 ***
## primary_shares_offered     -7.88424    1.57686  -5.000 6.14e-07 ***
## secondary_shares_offered -18.69884    3.51181  -5.325 1.10e-07 ***
## venture_backed              0.42466    0.06719   6.321 3.09e-10 ***
## num_bookrunners             1.16123    0.76188   1.524 0.127595
## rank_no_leads              -1.23696    0.73031  -1.694 0.090442 .
## num_lead_colead_managers    0.07370    0.55149   0.134 0.893702
## c1                          0.22321    0.23539   0.948 0.343094
## c2                          0.05612    0.21689   0.259 0.795843
## c3                          1.08615    0.43863   2.476 0.013344 *
```

```
## c4                        -0.14922    0.33532  -0.445 0.656362
## word_length_sentiment     -1.91088    0.36621  -5.218 1.96e-07 ***
## negative                  -0.49312    0.25163  -1.960 0.050145 .
## positive                  -0.72208    0.22380  -3.226 0.001270 **
## uncertainty               -0.93695    0.27341  -3.427 0.000621 ***
## litigious                 -1.03185    0.50154  -2.057 0.039758 *
## strongmodal               -0.73165    0.62675  -1.167 0.243176
## weakmodal                  0.51485    0.35726   1.441 0.149687
## constraining               0.10153    0.27880   0.364 0.715758
## internet                   0.86401    0.08444  10.232  < 2e-16 ***
## nasdaq_returns             2.33519    0.36833   6.340 2.73e-10 ***
## vix_returns                1.05447    0.37303   2.827 0.004741 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.447 on 2431 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.1714, Adjusted R-squared:  0.1639
## F-statistic: 22.85 on 22 and 2431 DF,  p-value: < 2.2e-16
```

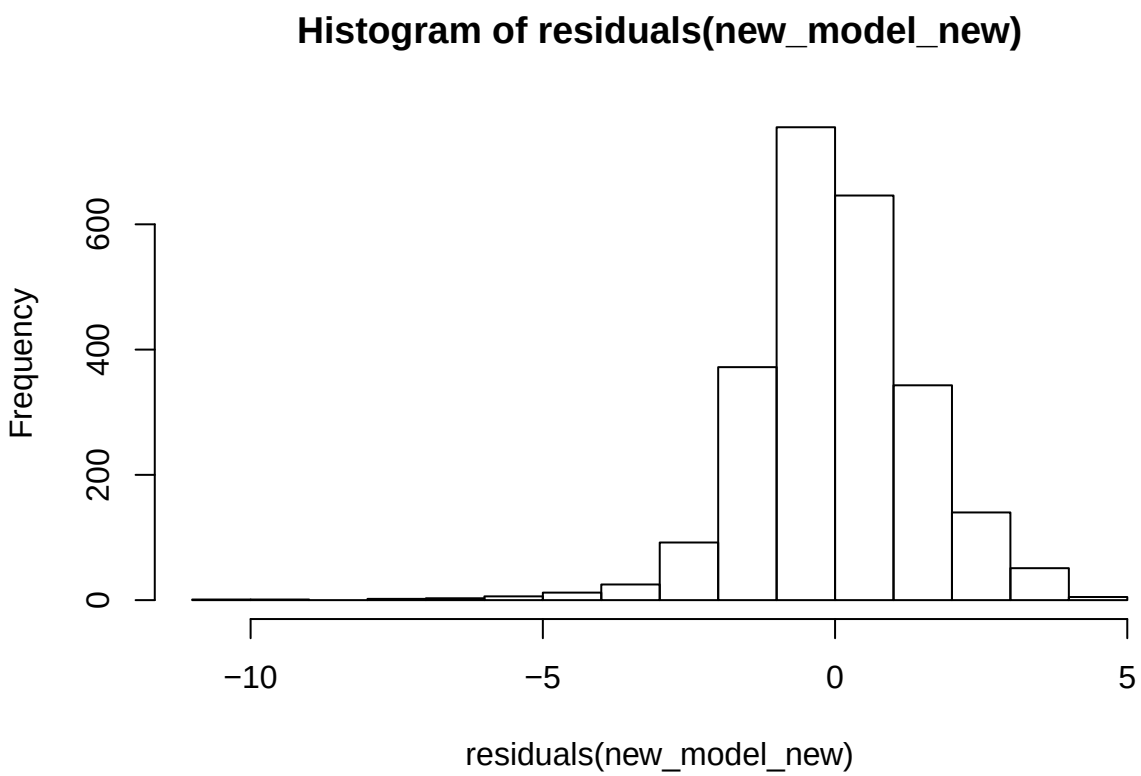
```
plot(residuals(new_model_new)~fitted(new_model_new))
```




```
qqnorm(residuals(new_model_new))
qqline(residuals(new_model_new))
```



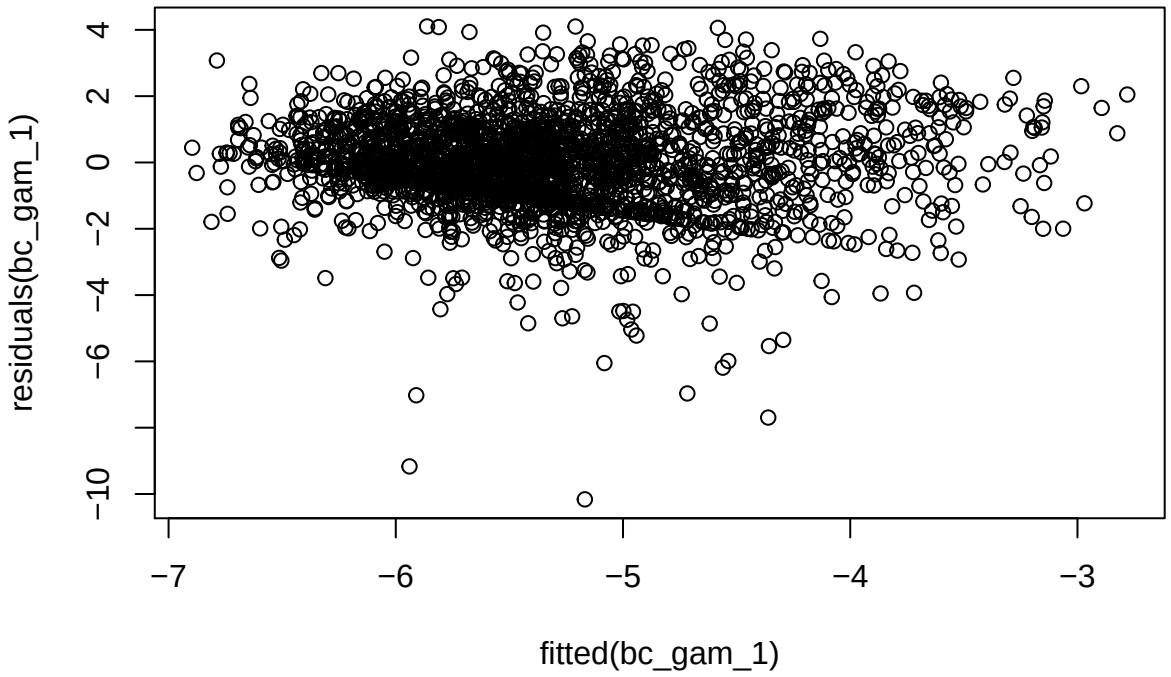

```
hist(residuals(new_model_new))
```



```r
bc_gam_1 = gam(formula=((underpricing^lambda_new-1)/lambda_new)~venture_backed+
               s(num_bookrunners)+s(rank_no_leads)+
                 s(num_lead_colead_managers)+s(c1)+s(c2)+s(c3)+s(c4)+
                 s(word_length_sentiment)+s(negative)+s(positive)+
                 s(uncertainty)+s(litigious)+s(strongmodal)+s(weakmodal)+
                 s(constraining)+internet+s(nasdaq_returns)+s(vix_returns),
             data=new)
summary(bc_gam_1)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## ((underpricing^lambda_new - 1)/lambda_new) ~ venture_backed +
##     s(num_bookrunners) + s(rank_no_leads) + s(num_lead_colead_managers) +
##     s(c1) + s(c2) + s(c3) + s(c4) + s(word_length_sentiment) +
##     s(negative) + s(positive) + s(uncertainty) + s(litigious) +
##     s(strongmodal) + s(weakmodal) + s(constraining) + internet +
##     s(nasdaq_returns) + s(vix_returns)
##
## Parametric coefficients:
##                Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    -5.60666    0.04353 -128.797  < 2e-16 ***
## venture_backed  0.35884    0.06851    5.238 1.77e-07 ***
## internet        0.76795    0.08378    9.166  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                               edf Ref.df      F  p-value
## s(num_bookrunners)          1.000  1.000  3.273  0.07055 .
## s(rank_no_leads)            3.548  4.337  4.027  0.00269 **
## s(num_lead_colead_managers) 4.702  5.637  2.961  0.00836 **
## s(c1)                       2.345  2.980  2.211  0.07154 .
## s(c2)                       2.093  2.678  1.767  0.17661
## s(c3)                       1.856  2.366  3.384  0.02805 *
## s(c4)                       1.000  1.000  0.144  0.70467
## s(word_length_sentiment)    2.285  2.909  8.512 2.39e-05 ***
## s(negative)                 1.923  2.411  2.859  0.05576 .
## s(positive)                 4.943  5.991  2.636  0.01612 *
## s(uncertainty)              1.000  1.000  8.951  0.00280 **
## s(litigious)                1.000  1.000  1.613  0.20413
## s(strongmodal)              1.000  1.000  2.387  0.12251
## s(weakmodal)                1.781  2.259  1.222  0.28869
## s(constraining)             1.001  1.001  0.004  0.95285
## s(nasdaq_returns)           5.632  6.855 14.652  < 2e-16 ***
## s(vix_returns)              3.184  4.057  6.069 7.00e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.198   Deviance explained = 21.2%
## GCV = 2.0441  Scale est. = 2.0081    n = 2454
```

```
plot(residuals(bc_gam_1)~fitted(bc_gam_1))
```




```
qqnorm(residuals(bc_gam_1))
qqline(residuals(bc_gam_1))
```