Turbulence Analysis

Tingnan Hu, Peter Liu, Islina Shan, Ken Ye, Nancy Zhang

2023-10-27

Introduction

Turbulence is one of the fascinating topics in the research in fluid dynamics. It is characterized by its chaotic motion, rapid fluctuations and lack of predictable patterns. Yet, there have been numerous attempts in scientific literature trying to model the behavior of turbulent flows, as turbulent flows are prevalent in our world and are the underlying forces that drive plenty of the physical processes, from wisps of smoking swirling up from the cigarette to mixing of chemicals in industrial processes. A better understanding and prediction of turbulent flow will help us gain a deeper insight into a wide range of applications, such as improved aerodynamics in airplane designs and better climatic modelling.

A subdomain in turbulent flow research deals with particle clustering in turbulent flow focusing on small particles" behavior in turbulent fluids. For our project, we are provided with a set of simulation results on small particle probability distribution. The outcome variable was originally a probability distribution for particle cluster volumes, but it was converted into its first four raw moments, $E[X]$ to $E[X^4]$, to facilitate analysis. The predictor set contains three variables:

- Reynolds number, `Re`, which provides information on the type of flow a fluid is experiencing. A low `Re` corresponds with laminar flow (smooth and orderly), while a high `Re` corresponds with turbulent flow.
- Gravitational acceleration, `Fr`, which measures the gravitational forces particles are experiencing.
- Stokes number, `St`, where larger value corresponds with larger particle size.

The main research objective of our project will be to build a viable statistical model to predict the response variable (first four raw moments of particle probability distribution) using the three predictors at hand and the provided traning set. Specifically, we are interested in the following:

- Does there exist a significant linear relationship between the predictors and the raw four moments?
- Is there any significant interaction effects between predictors on the response variables?
- Does a linear regression model suffice? Do we need a more complex model to better explain the relationship between the predictors and responses?
- Do the identified effects of the predictors vary for the four moments?

Ultimately, we aim for our model to capture adequate trends in our training data, so that for a new parameter setting of (`Re`, `Fr`, `St`), we can accurately predict its particle cluster volume distribution in terms of its four raw moments, as well as make inference on how each parameter affects the probability distribution for particle cluster volumes.

Methodology

First, we examine the predictor and response variables and perform adequate transformations. For predictor variables, we first noticed that `Fr` only takes on 0.052, 0.3, and Inf in both our training and testing data set, and directly using these values as they are is not viable as they contain infinity. Therefore, we create a new categorical variable called `gravity` using the following categorization:

Fr	Gravity
Fr < 0.1	low gravity
0.1 < Fr < 1	moderate gravity
Fr > 1	high gravity

We also noticed that the predictor variable `Re` only takes on 90, 224, and 398 in both our training and testing data set. We thus create a new categorical variable called `flow` using the following categorization:

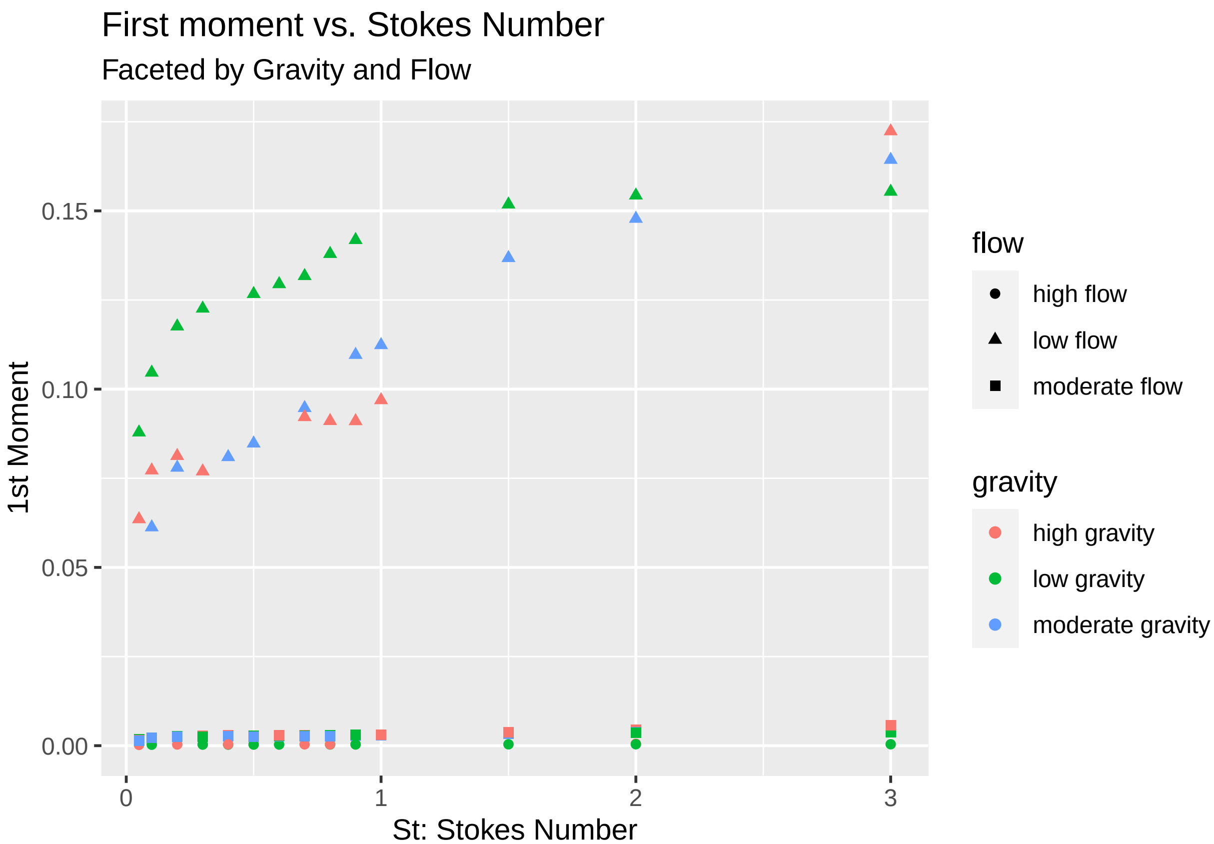
Re	Flow
Re < 100	low flow
100 < Re < 300	moderate flow
Re > 300	high flow

Simple Linear Regression

We explore the simple linear regression model first, which results in an adjusted R-squared value of 0.9251.

Explore Interaction Effect

The plot below suggests a possible interaction effect between `St` and `flow` number for the 1st Moment, as the low flow (triangle shape) group reveals a positive slope, whereas the moderate flow (square shape) and high flow (circular shape) groups reveal close-to-zero slopes:



Applying this insight into our linear model for 1st moment, we add the interaction term `flow:St`, resulting a higher adjusted R-squared of 0.9676.

```
##
## Call:
## lm(formula = R_moment_1 ~ gravity + St + flow + flow:St, data = train)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.024358 -0.006729  0.003284  0.004195  0.023329
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -0.0038861  0.0036011  -1.079    0.284
## gravitylow gravity      0.0106824  0.0025101   4.256 5.56e-05 ***
## gravitymoderate gravity  0.0001185  0.0029338   0.040    0.968
## St                     -0.0016501  0.0031049  -0.531    0.597
## flowlow flow            0.0854744  0.0044191  19.342  < 2e-16 ***
## flowmoderate flow       0.0023091  0.0043881   0.526    0.600
## St:flowlow flow         0.0308060  0.0037499   8.215 2.83e-12 ***
## St:flowmoderate flow    0.0023811  0.0038207   0.623    0.535
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01005 on 81 degrees of freedom
## Multiple R-squared:  0.9702, Adjusted R-squared:  0.9676
## F-statistic: 376.3 on 7 and 81 DF,  p-value: < 2.2e-16
```

Explore Polynomial

We also tried using polynomial terms up to degree of 5 for stokes number, and judging from the p value for the associated F-statistics, only the first order term is necessary.

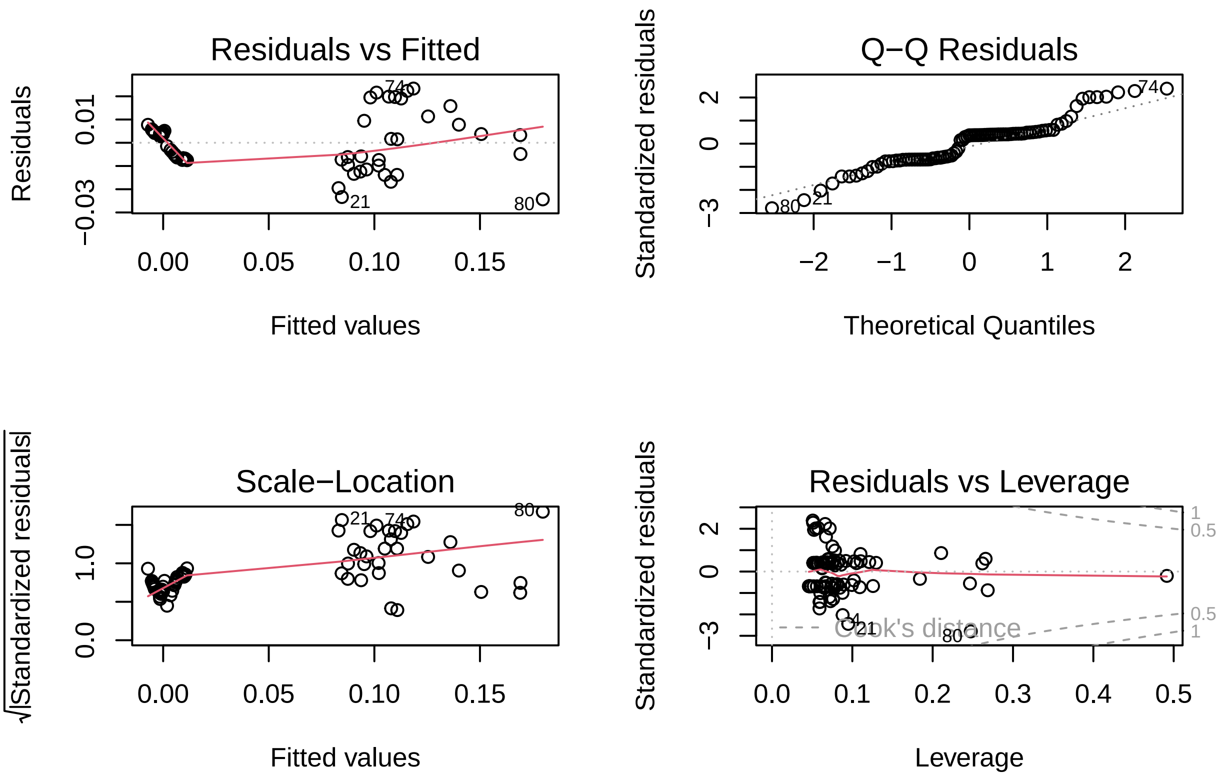
Ridge Regression

Splines

Results

Model 1: Linear Regression w/ Interaction Term

Considering a simple linear regression on the first moment: we have a 0.97 adjusted R squared value and significant F-statistics; however, the residual vs fitted values plot indicates a obvious non-linear trend, which suggests that the linearity assumption is violated.



Model 2: Ridge Regression

Model 3: Ridge Regression

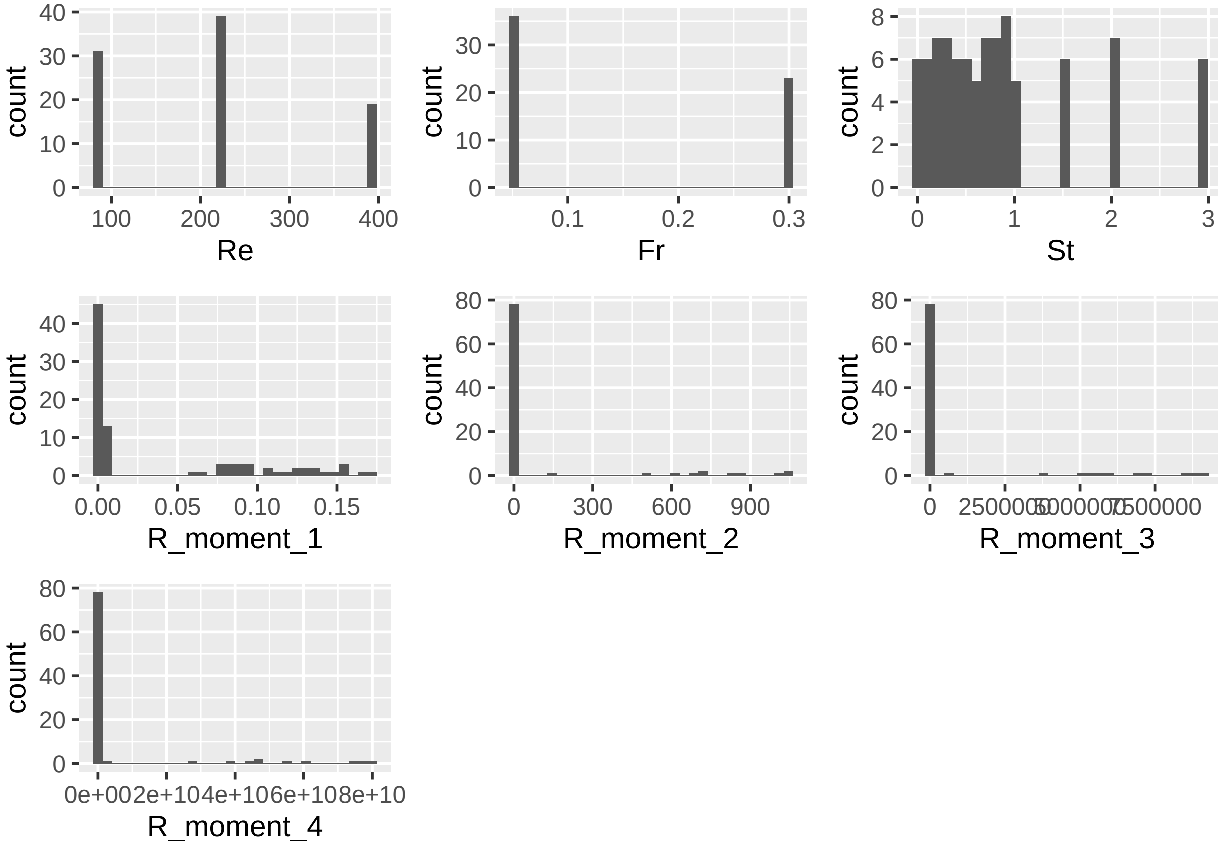
Final Model: ?

Conclusion

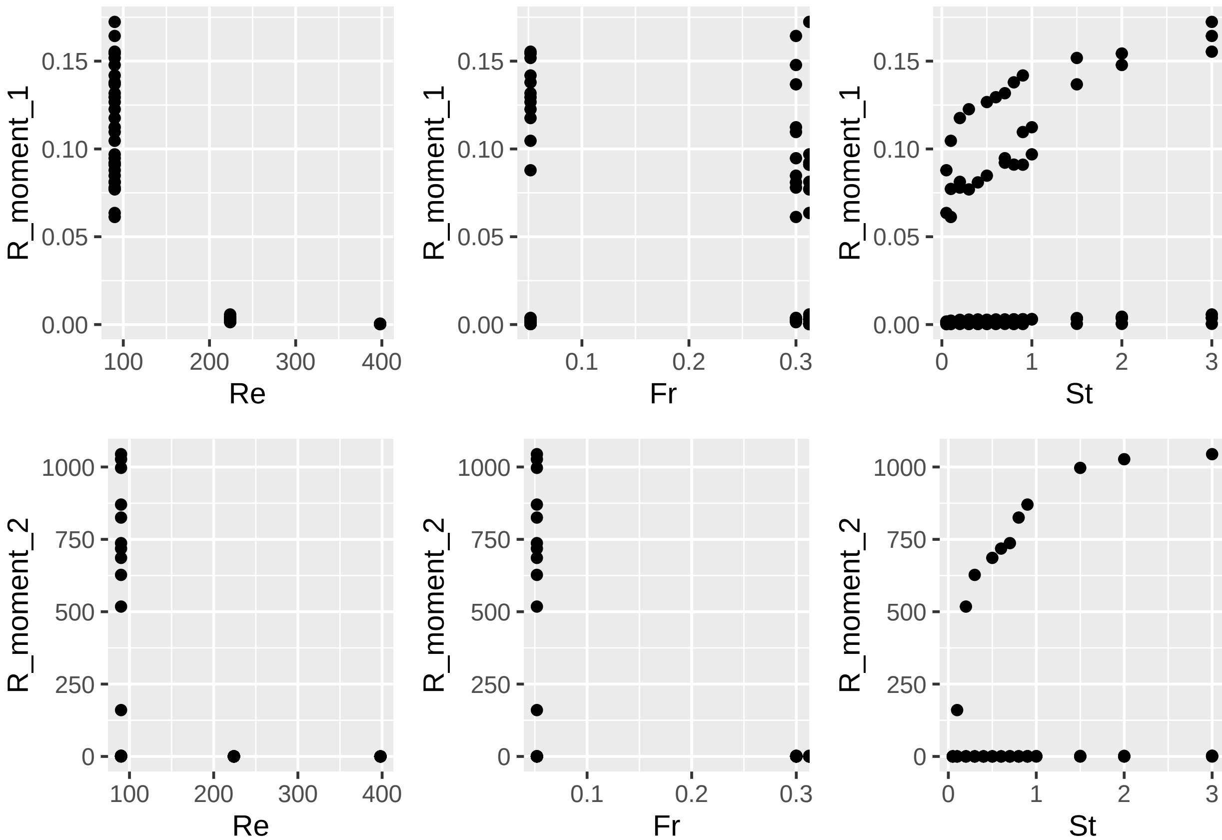
Appendix

EDA

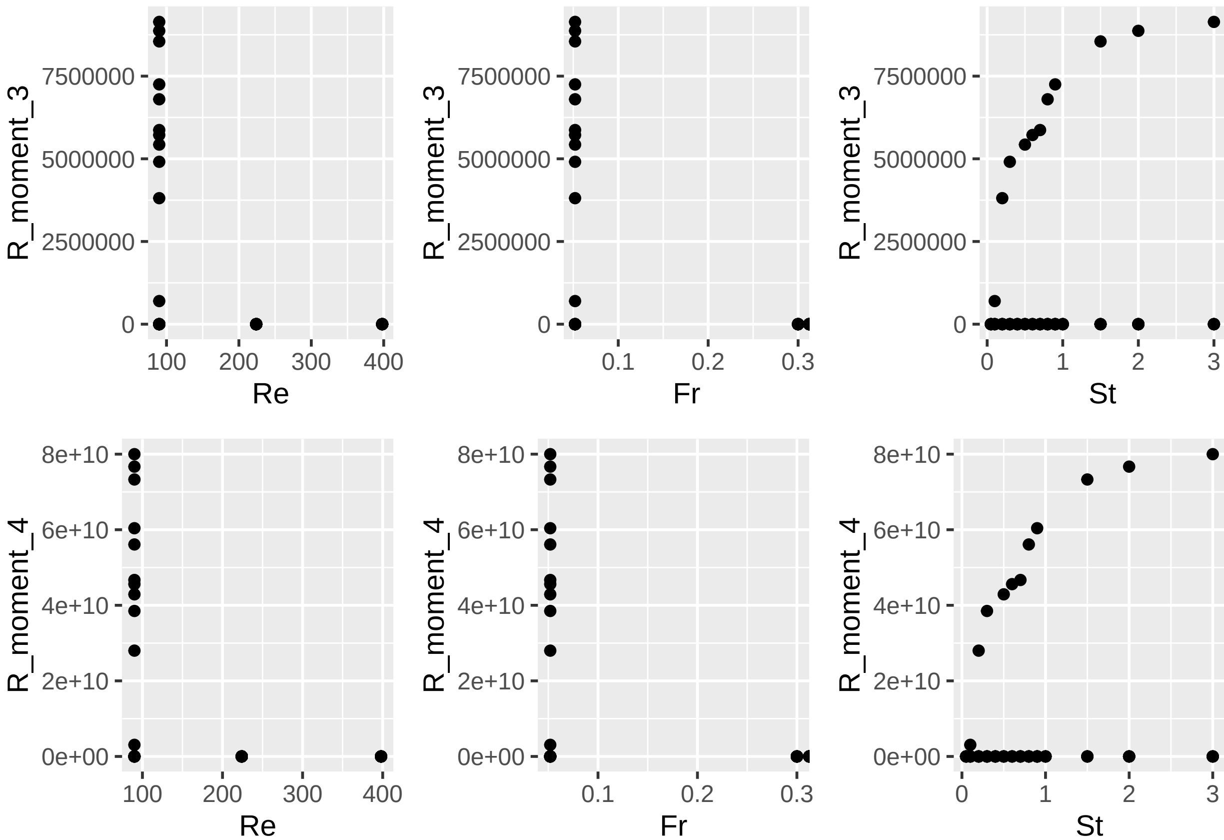
```
##        St               Re              Fr          R_moment_1
##  Min.   :0.0500   Min.   : 90.0   Min.   :0.052   Min.   :0.000222
##  1st Qu.:0.3000   1st Qu.: 90.0   1st Qu.:0.052   1st Qu.:0.002157
##  Median :0.7000   Median :224.0   Median :0.300   Median :0.002958
##  Mean   :0.8596   Mean   :214.5   Mean   :  Inf   Mean   :0.040394
##  3rd Qu.:1.0000   3rd Qu.:224.0   3rd Qu.:  Inf   3rd Qu.:0.087868
##  Max.   :3.0000   Max.   :398.0   Max.   :  Inf   Max.   :0.172340
##    R_moment_2         R_moment_3        R_moment_4          gravity
##  Min.   :   0.0001   Min.   :      0   Min.   :0.000e+00   Length:89
##  1st Qu.:   0.0245   1st Qu.:      0   1st Qu.:3.000e+00   Class :character
##  Median :   0.0808   Median :      1   Median :2.100e+01   Mode  :character
##  Mean   :  92.4902   Mean   : 753370   Mean   :6.194e+09
##  3rd Qu.:   0.5345   3rd Qu.:     40   3rd Qu.:5.345e+03
##  Max.   :1044.3000   Max.   :9140000   Max.   :8.000e+10
##      flow
##  Length:89
##  Class :character
##  Mode  :character
##
##
##
```




```
## TableGrob (3 x 3) "arrange": 7 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (1-1,3-3) arrange gtable[layout]
## 4 4 (2-2,1-1) arrange gtable[layout]
## 5 5 (2-2,2-2) arrange gtable[layout]
## 6 6 (2-2,3-3) arrange gtable[layout]
## 7 7 (3-3,1-1) arrange gtable[layout]
```




```
## TableGrob (2 x 3) "arrange": 6 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (1-1,3-3) arrange gtable[layout]
## 4 4 (2-2,1-1) arrange gtable[layout]
## 5 5 (2-2,2-2) arrange gtable[layout]
## 6 6 (2-2,3-3) arrange gtable[layout]
```




```
## TableGrob (2 x 3) "arrange": 6 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (1-1,3-3) arrange gtable[layout]
## 4 4 (2-2,1-1) arrange gtable[layout]
## 5 5 (2-2,2-2) arrange gtable[layout]
## 6 6 (2-2,3-3) arrange gtable[layout]
```

Simple Linear Regression

First Moment

Using 5-fold cross-validation to estimate the test set error

```
## Linear Regression
##
## 89 samples
##  3 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold)
## Summary of sample sizes: 73, 72, 69, 72, 70
## Resampling results:
##
```

```
##   RMSE        Rsquared   MAE
##   0.01060141  0.9700254  0.008485167
##
## Tuning parameter 'intercept' was held constant at a value of TRUE
```

Trying using polynomial terms up to degree of 5 for stokes number:

```
## Analysis of Variance Table
##
## Model 1: response ~ St + gravity + flow
## Model 2: response ~ poly(St, 2) + gravity + flow
## Model 3: response ~ poly(St, 3) + gravity + flow
## Model 4: response ~ poly(St, 4) + gravity + flow
## Model 5: response ~ poly(St, 5) + gravity + flow
##   Res.Df      RSS Df  Sum of Sq      F Pr(>F)
## 1     83 0.019399
## 2     82 0.019352  1 4.7187e-05 0.1959 0.6593
## 3     81 0.019180  1 1.7206e-04 0.7142 0.4006
## 4     80 0.019134  1 4.5704e-05 0.1897 0.6643
## 5     79 0.019031  1 1.0305e-04 0.4278 0.5150
```

Judging from the p value for the associated F-statistics, only the first order term is necessary.

Moments 2-4

```
##
## Call:
## lm(formula = log(R_moment_2) ~ gravity + St + flow, data = train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0075 -1.2112 -0.1009  1.1631  3.0215
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -7.3531     0.5141 -14.302  < 2e-16 ***
## gravitylow gravity        2.7650     0.4473   6.182 2.27e-08 ***
## gravitymoderate gravity  -0.6773     0.5217  -1.298  0.19784
## St                        0.7167     0.2455   2.920  0.00451 **
## flowlow flow              7.7930     0.5472  14.242  < 2e-16 ***
## flowmoderate flow         3.1608     0.5280   5.987 5.26e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.806 on 83 degrees of freedom
## Multiple R-squared:  0.7768, Adjusted R-squared:  0.7633
## F-statistic: 57.76 on 5 and 83 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = log(R_moment_3) ~ gravity + St + flow, data = train)
##
```

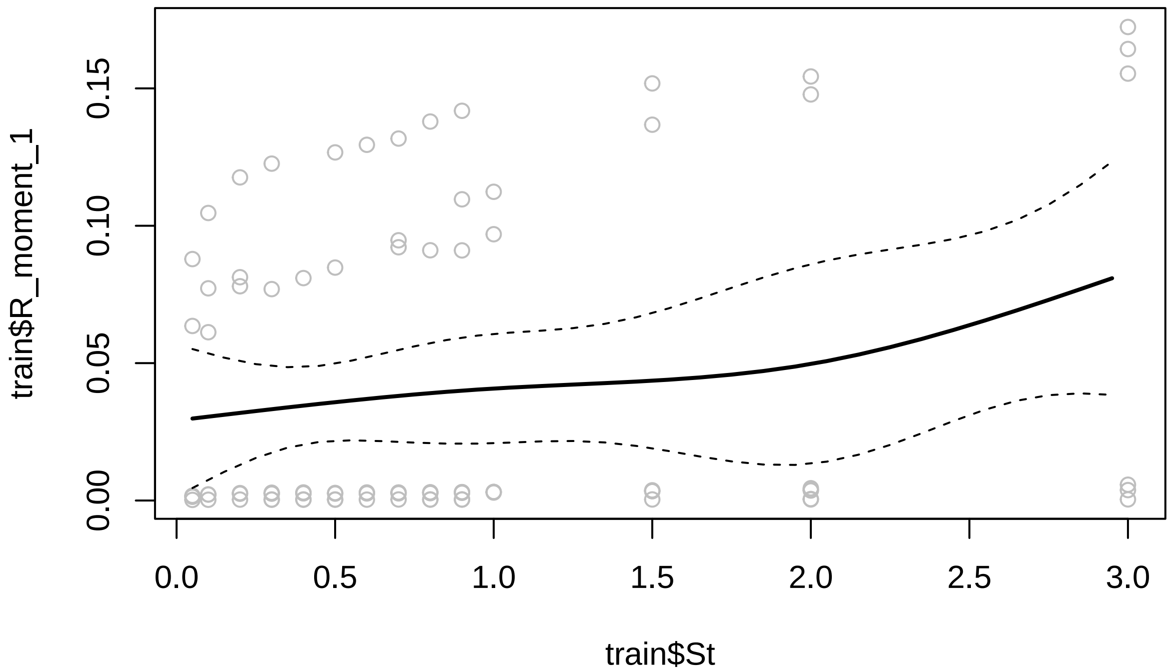
```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.7282 -2.3839 -0.4306  2.1123  5.4634
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -6.0426     0.9144  -6.609 3.49e-09 ***
## gravitylow gravity       5.4848     0.7955   6.895 9.77e-10 ***
## gravitymoderate gravity  -1.3207     0.9279  -1.423   0.1584
## St                       0.9452     0.4366   2.165   0.0332 *
## flowlow flow            10.2176     0.9731  10.500  < 2e-16 ***
## flowmoderate flow        4.3380     0.9390   4.620 1.39e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.211 on 83 degrees of freedom
## Multiple R-squared:  0.6983, Adjusted R-squared:  0.6802
## F-statistic: 38.43 on 5 and 83 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = log(R_moment_4) ~ gravity + St + flow, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.1076  -3.5768  -0.7964   3.0052   7.8067
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -4.6815     1.2975  -3.608 0.000526 ***
## gravitylow gravity       8.1791     1.1288   7.246 2.02e-10 ***
## gravitymoderate gravity  -1.9646     1.3166  -1.492 0.139457
## St                       1.1304     0.6195   1.825 0.071636 .
## flowlow flow            12.7305     1.3808   9.219 2.38e-14 ***
## flowmoderate flow        5.5438     1.3324   4.161 7.70e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.557 on 83 degrees of freedom
## Multiple R-squared:  0.6716, Adjusted R-squared:  0.6518
## F-statistic: 33.95 on 5 and 83 DF,  p-value: < 2.2e-16
```

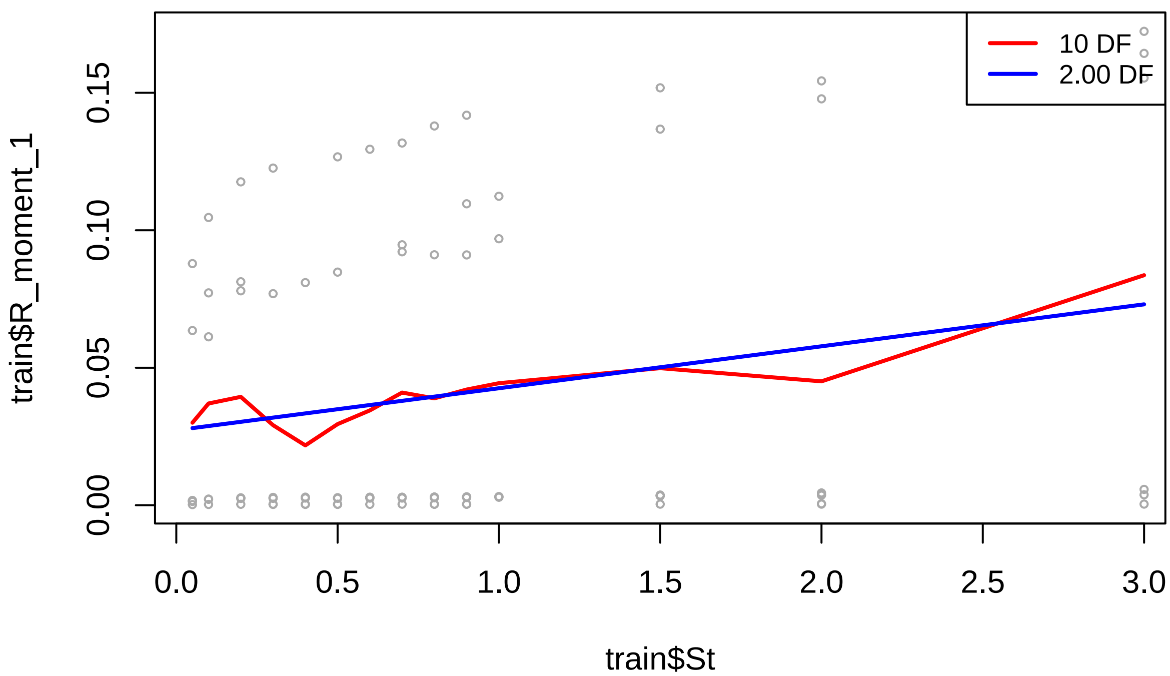
Considering a simple linear regression on the first moment: we have a 0.97 adjusted R squared value and significant F-statistics; however, the residual vs fitted values plot indicates a obvious non-linear trend, which suggests that the linearity assumption is violated.

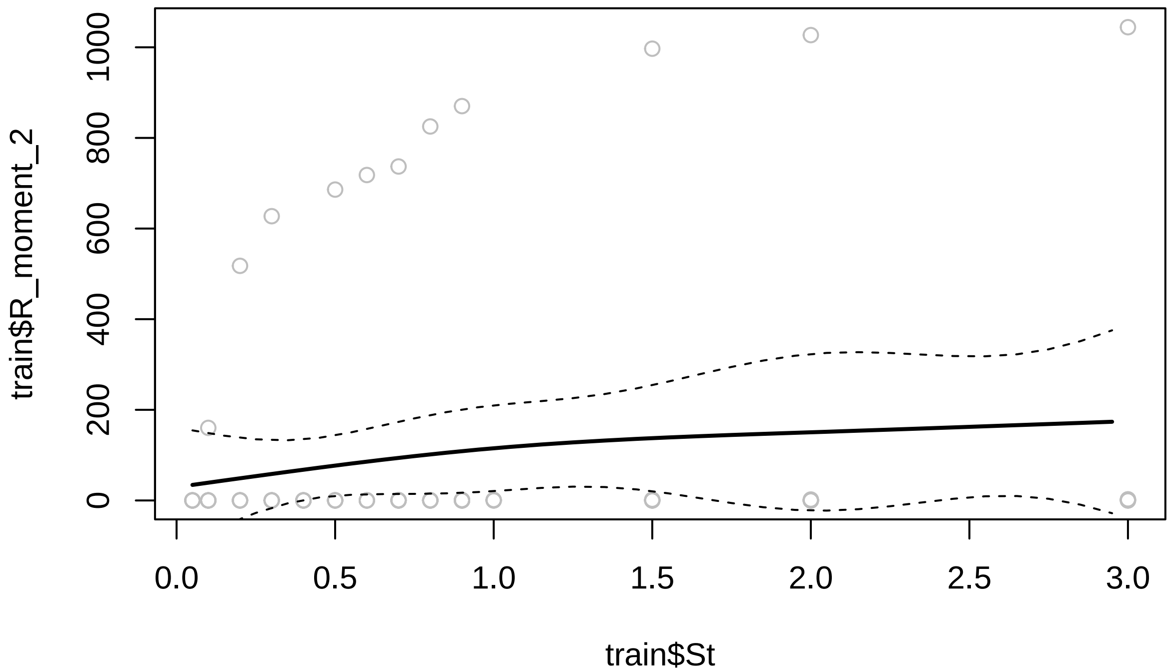
Ridge Regression

Splines

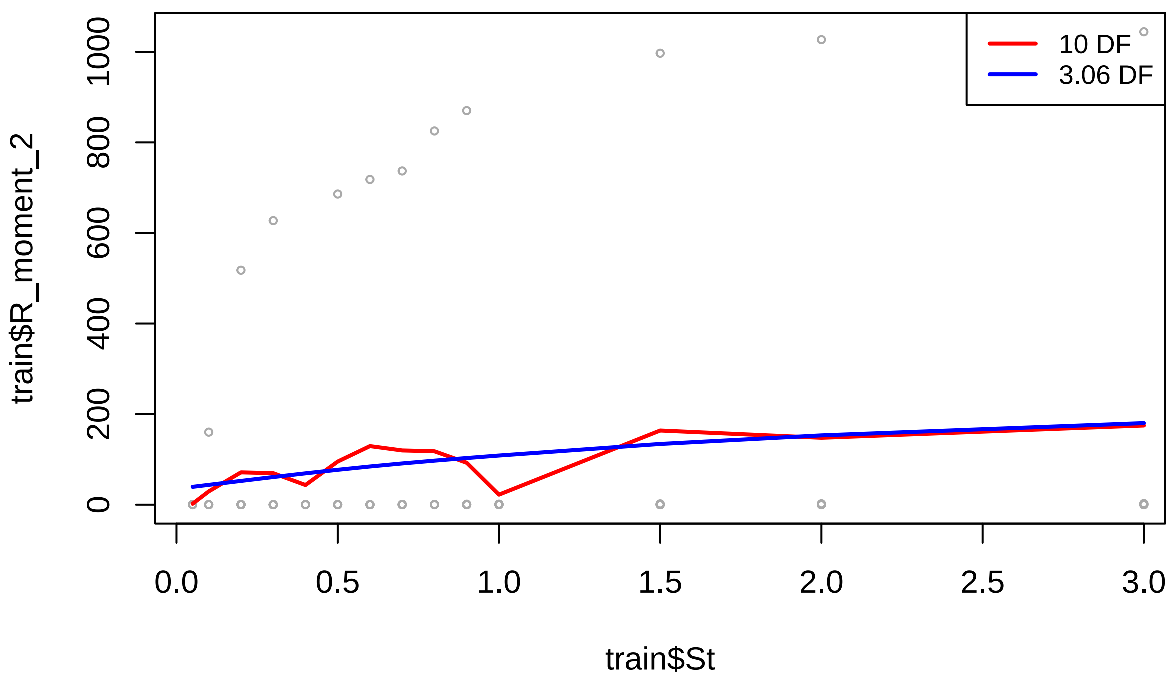


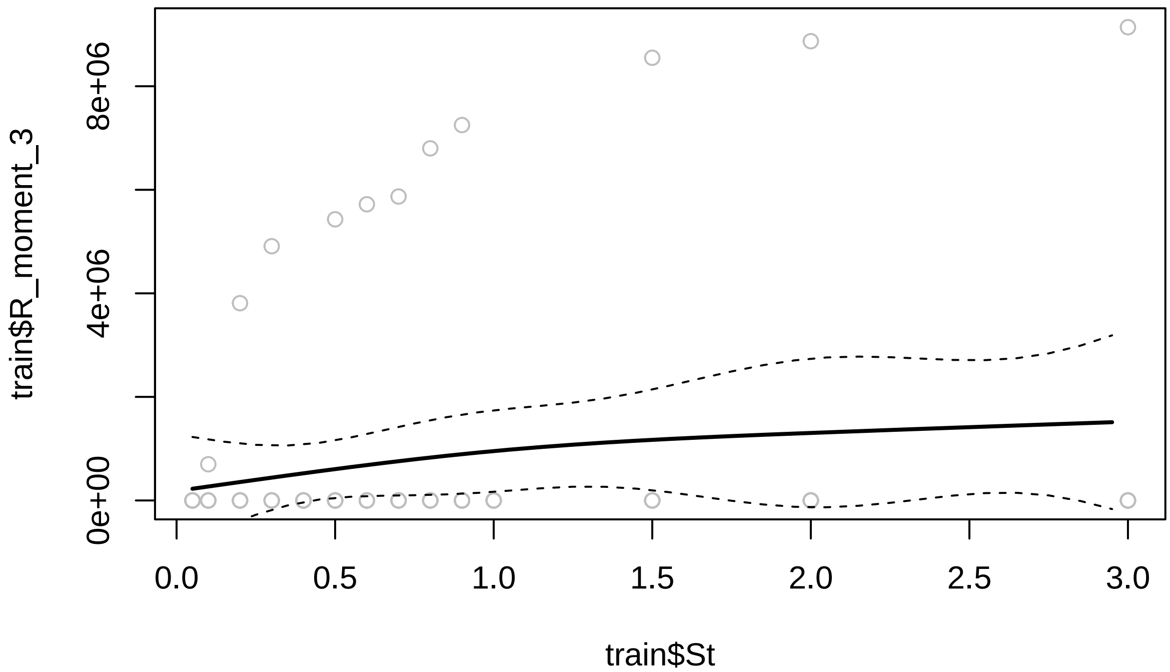
Smoothing Spline

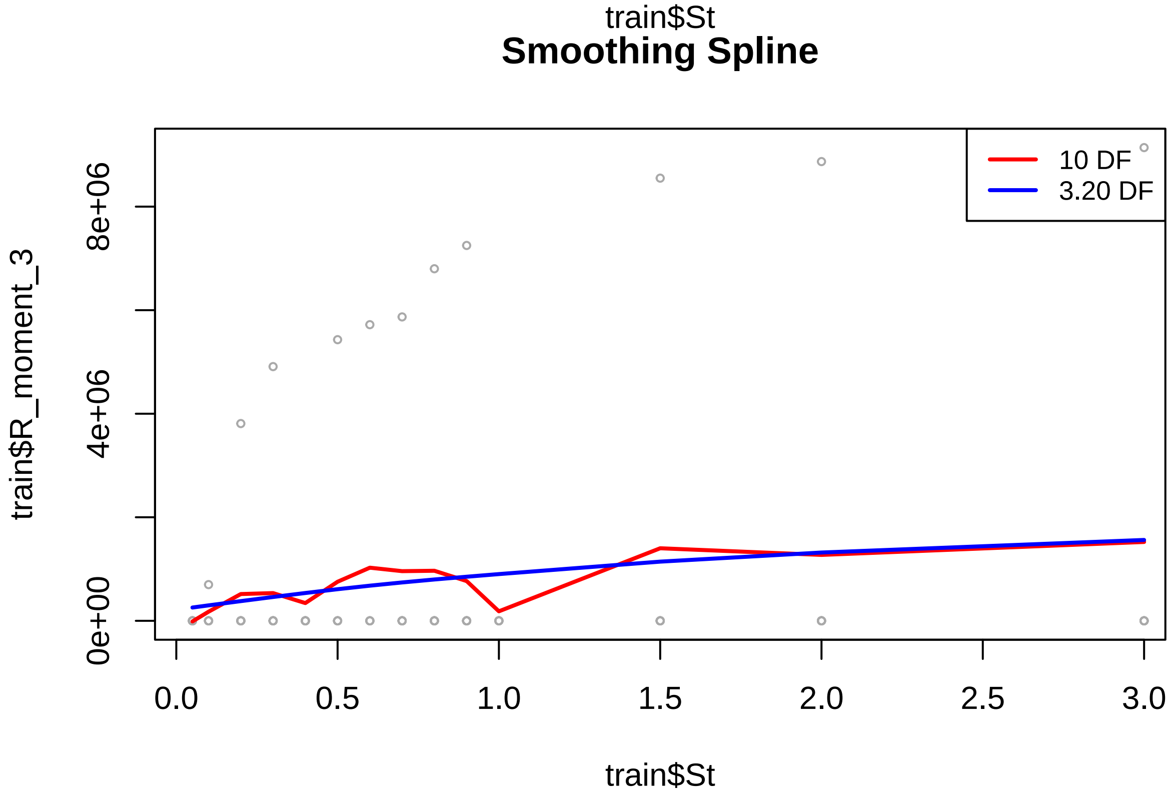


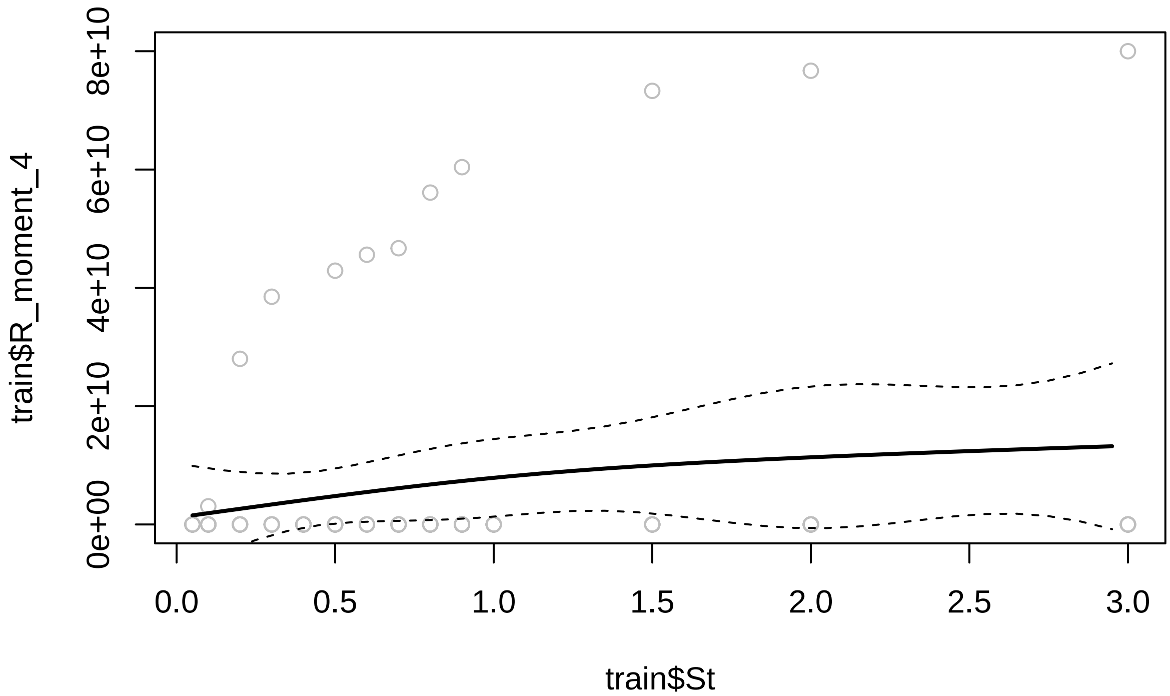


Smoothing Spline

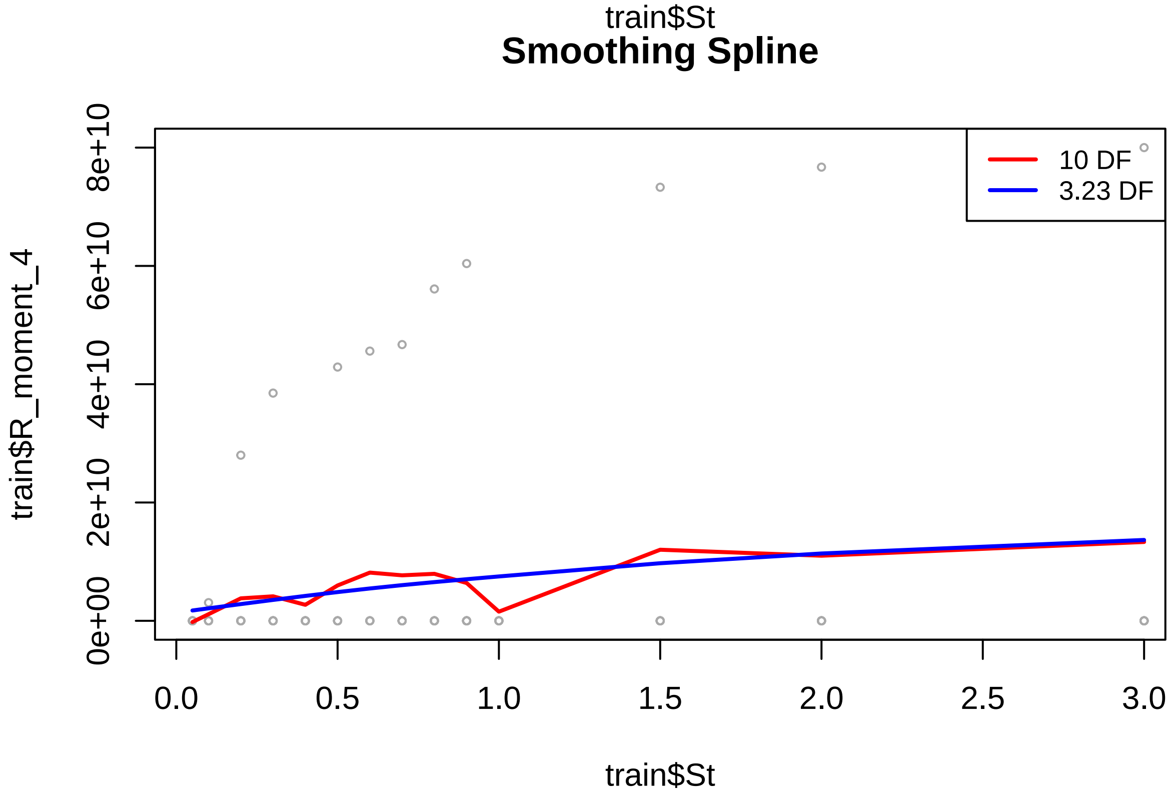



train$R_moment_3
0e+00
4e+06
8e+06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St


Smoothing Spline
10 DF
3.20 DF
train$R_moment_3
0e+00
4e+06
8e+06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St



train$R_moment_4
0e+00
2e+10
4e+10
6e+10
8e+10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St



Smoothing Spline
10 DF
3.23 DF
train$R_moment_4
0e+00
2e+10
4e+10
6e+10
8e+10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
train$St