Final Report

Predicting the first four moments in particle turbulence

Introduction

This report is authored by **Matthew Cui, Phillip Harmadi, Glen Morgenstern, Joe Wang, Gaurav Sirdeshmukh,** and **Gautam Sirdeshmukh.**

Methodology

```
##                      St          Re  Fr R_moment_1 R_moment_2 R_moment_3
## St           1.00000000 -0.03169871 NaN  0.2147681  0.1479257  0.1647465
## Re          -0.03169871  1.00000000 NaN -0.7747206 -0.3932344 -0.3844289
## Fr                  NaN         NaN   1        NaN        NaN        NaN
## R_moment_1   0.21476813 -0.77472058 NaN  1.0000000  0.6298829  0.6217326
## R_moment_2   0.14792571 -0.39323445 NaN  0.6298829  1.0000000  0.9984335
## R_moment_3   0.16474648 -0.38442895 NaN  0.6217326  0.9984335  1.0000000
## R_moment_4   0.18004537 -0.37741773 NaN  0.6150484  0.9946671  0.9988414
##             R_moment_4
## St           0.1800454
## Re          -0.3774177
## Fr                 NaN
## R_moment_1   0.6150484
## R_moment_2   0.9946671
## R_moment_3   0.9988414
## R_moment_4   1.0000000
```

```
# We transform the variables using the sigmoid function so that this variable
# will be within a finite range.

train1 <- train %>%
  rename(M1 = R_moment_1, M2 = R_moment_2, M3 = R_moment_3, M4 = R_moment_4) %>%
  mutate(Fr_sigmoid = 1 / (1 + exp(-Fr)),
         Re_sigmoid = 1 / (1 + exp(-Re)),
         M1_sigmoid = 1 / (1 + exp(-M1)),
         M2_sigmoid = 1 / (1 + exp(-M2)),
         M3_sigmoid = 1 / (1 + exp(-M3)),
         M4_sigmoid = 1 / (1 + exp(-M4))) %>%
  mutate(Re_categorical = case_when(Re == 90 ~ "Low", Re == 224 ~ "Medium", Re == 398 ~ "High"),
         Fr_categorical = case_when(Fr == 0.052 ~ "Low", Fr == 0.3 ~ "Medium", Fr == Inf ~ "High"))

test1 <- test %>%
  mutate(Fr_sigmoid = 1 / (1 + exp(-Fr)),
         Re_sigmoid = 1 / (1 + exp(-Re))) %>%
  mutate(Re_categorical = case_when(Re == 90 ~ "Low", Re == 224 ~ "Medium", Re == 398 ~ "High"),
         Fr_categorical = case_when(Fr == 0.052 ~ "Low", Fr == 0.3 ~ "Medium", Fr == Inf ~ "High"))

train1
```

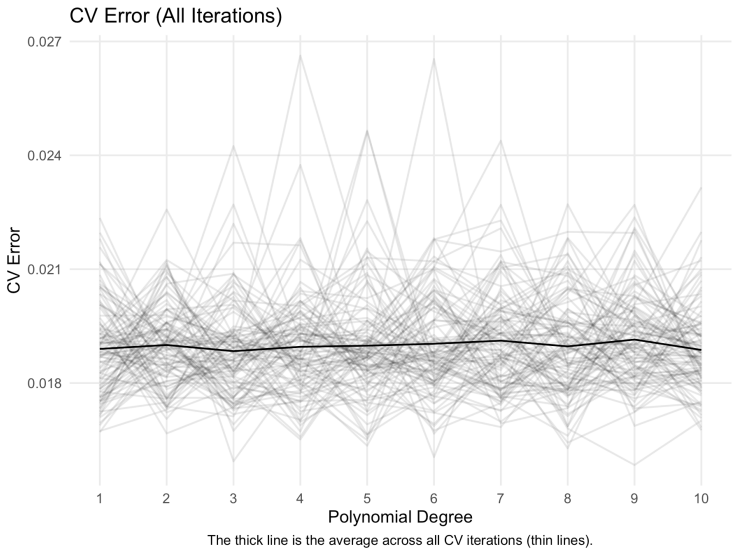
```
## # A tibble: 89 x 15
##       St    Re      Fr       M1      M2      M3       M4 Fr_sigmoid Re_sigmoid
##    <dbl> <dbl>   <dbl>    <dbl>   <dbl>   <dbl>    <dbl>      <dbl>      <dbl>
##  1  0.1    224   0.052 0.00216  0.130   14.4     1586.        0.513          1
```

```
##  2  3      224   0.052 0.00379  0.470   69.9    10404         0.513          1
##  3  0.7    224 Inf     0.00291  0.0435   0.822     15.6       1              1
##  4  0.05    90 Inf     0.0635   0.0907   0.467      3.27      1              1
##  5  0.7    398 Inf     0.000369 0.00622  0.126      2.57      1              1
##  6  2       90   0.3   0.148    2.01    36.2      672.        0.574          1
##  7  0.2     90 Inf     0.0813   0.324    3.04      33.0       1              1
##  8  3      224 Inf     0.00575  0.120    2.75      63.2       1              1
##  9  0.9    224 Inf     0.00302  0.0452   0.845     15.8       1              1
## 10  0.6    398   0.052 0.000314 0.00447  0.0821     1.51      0.513          1
## # ... with 79 more rows, and 6 more variables: M1_sigmoid <dbl>,
## #   M2_sigmoid <dbl>, M3_sigmoid <dbl>, M4_sigmoid <dbl>, Re_categorical <chr>,
## #   Fr_categorical <chr>
```

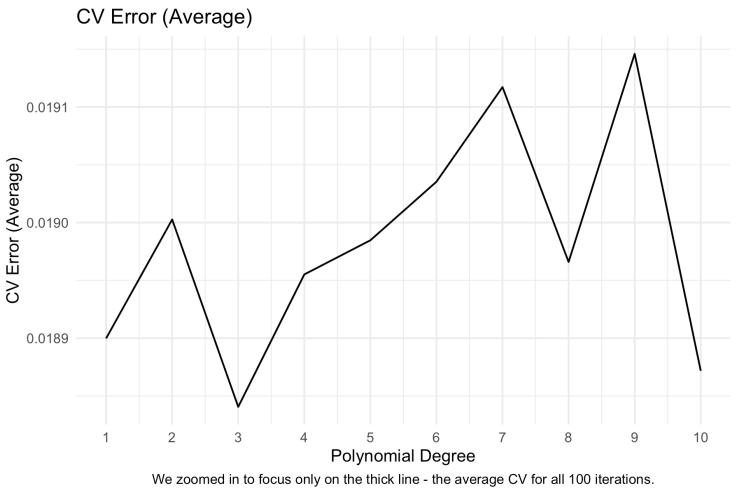
```
test1
```

```
## # A tibble: 23 x 7
##       St    Re      Fr Fr_sigmoid Re_sigmoid Re_categorical Fr_categorical
##    <dbl> <dbl>   <dbl>      <dbl>      <dbl> <chr>          <chr>
##  1  0.05   398   0.052      0.513          1 High           Low
##  2  0.2    398   0.052      0.513          1 High           Low
##  3  0.7    398   0.052      0.513          1 High           Low
##  4  1      398   0.052      0.513          1 High           Low
##  5  0.1    398 Inf          1              1 High           High
##  6  0.6    398 Inf          1              1 High           High
##  7  1      398 Inf          1              1 High           High
##  8  1.5    398 Inf          1              1 High           High
##  9  3      398 Inf          1              1 High           High
## 10  3      224   0.3        0.574          1 Medium         Medium
## # ... with 13 more rows
```

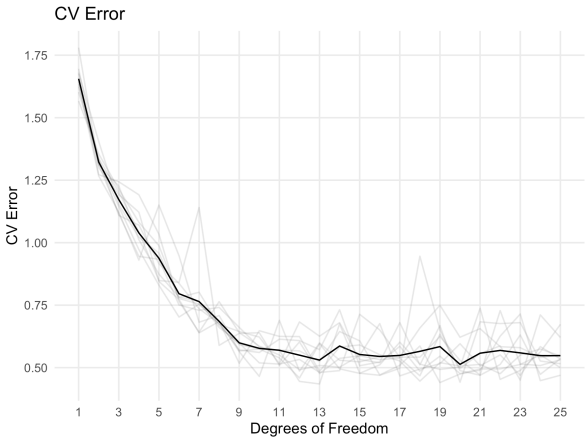
Quick Overview

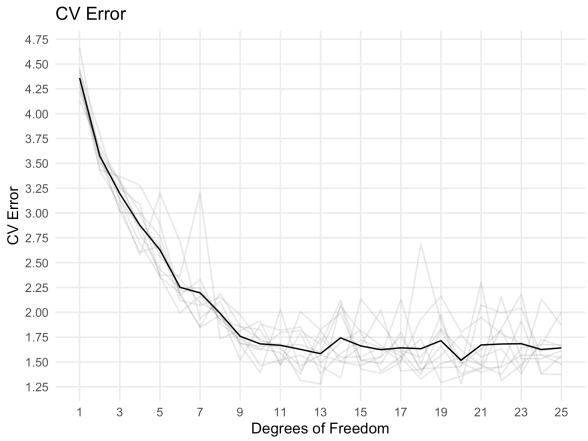


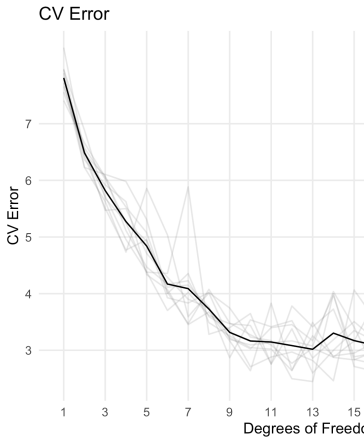
The thick line is the average across all CV iterations (thin lines).



We zoomed in to focus only on the thick line - the average CV for all 100 iterations.







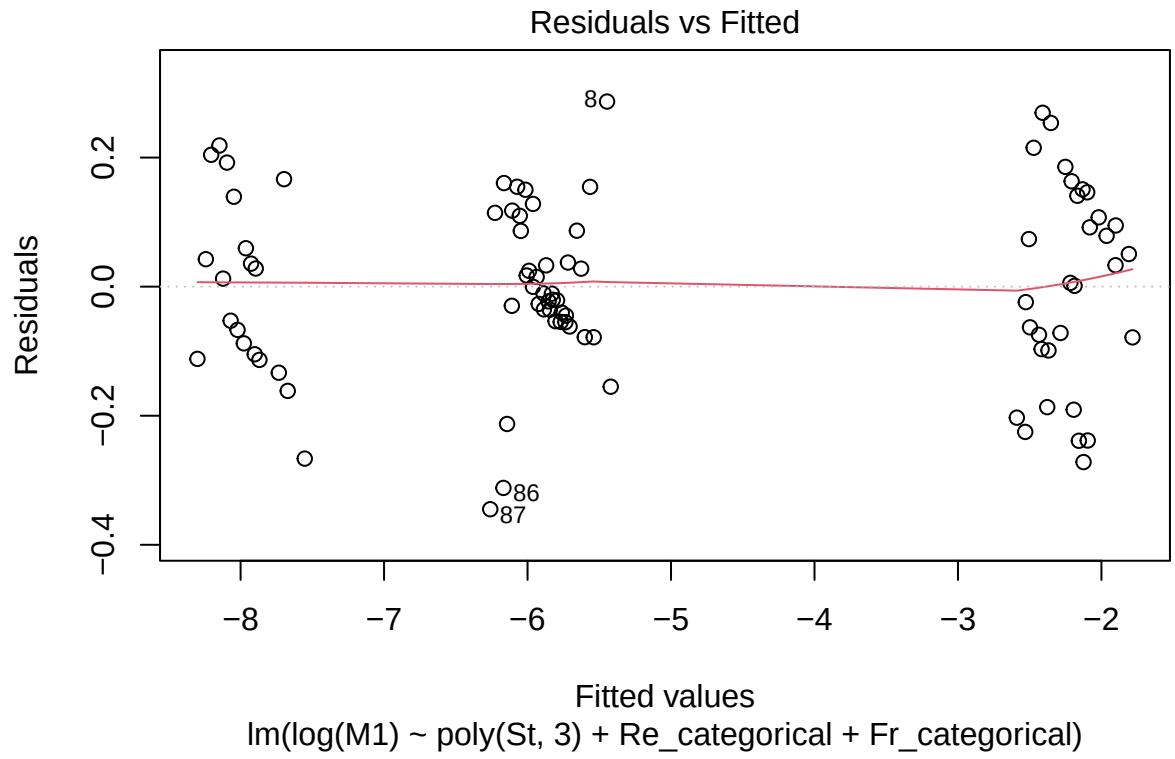
Results

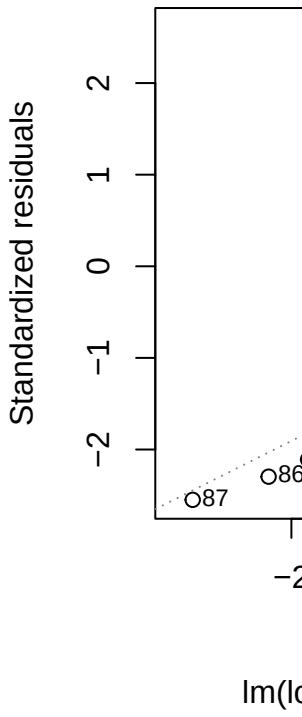
Final Models

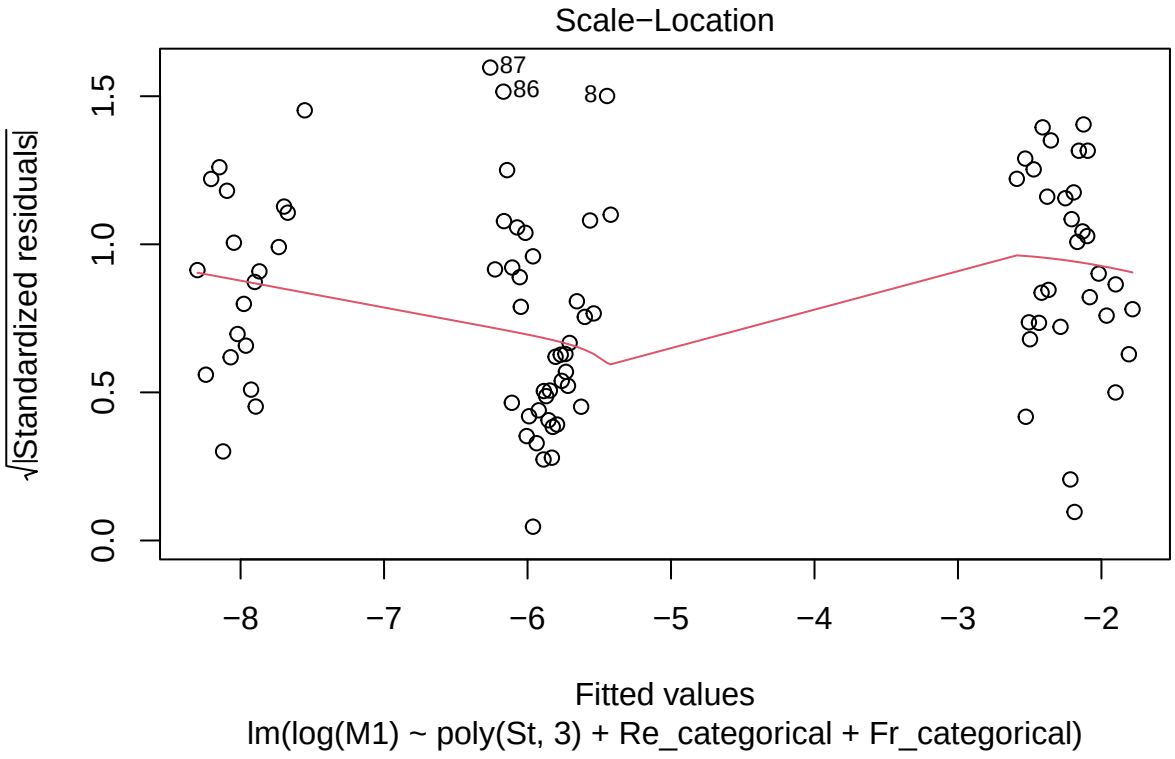
```
M1 <- lm(log(M1) ~ poly(St, 3) + Re_categorical + Fr_categorical, data = train1)
summary(M1)
```

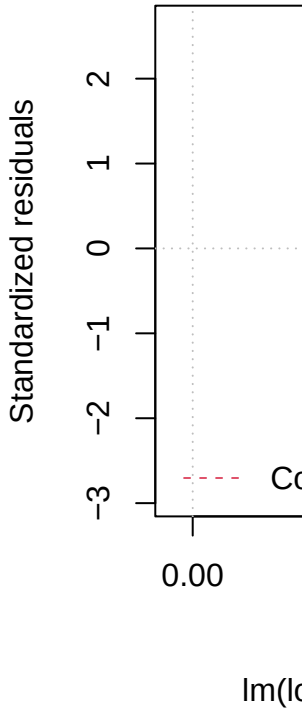
```
##
## Call:
## lm(formula = log(M1) ~ poly(St, 3) + Re_categorical + Fr_categorical,
##     data = train1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3449 -0.0783 -0.0003  0.1073  0.2868
##
## Coefficients:
##                     Estimate Std. Error  t value Pr(>|t|)
## (Intercept)         -7.98153    0.03784 -210.911  < 2e-16 ***
## poly(St, 3)1         1.81526    0.14370   12.633  < 2e-16 ***
## poly(St, 3)2        -0.63184    0.14418   -4.382 3.49e-05 ***
## poly(St, 3)3         0.21344    0.14347    1.488   0.1407
## Re_categoricalLow    5.76962    0.04355  132.489  < 2e-16 ***
## Re_categoricalMedium  2.13265    0.04191   50.881  < 2e-16 ***
## Fr_categoricalLow    0.02556    0.03552    0.720   0.4738
## Fr_categoricalMedium -0.09182    0.04155   -2.210   0.0299 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1433 on 81 degrees of freedom
## Multiple R-squared:  0.9962, Adjusted R-squared:  0.9959
## F-statistic:  3054 on 7 and 81 DF,  p-value: < 2.2e-16
```

```
plot(M1)
```










```
M2 <- gam(log(M2) ~ s(St, df = 9) + Re_categorical + Fr_categorical +
                    Re_categorical * Fr_categorical, data = train1)
summary(M2)
```

```
##
```

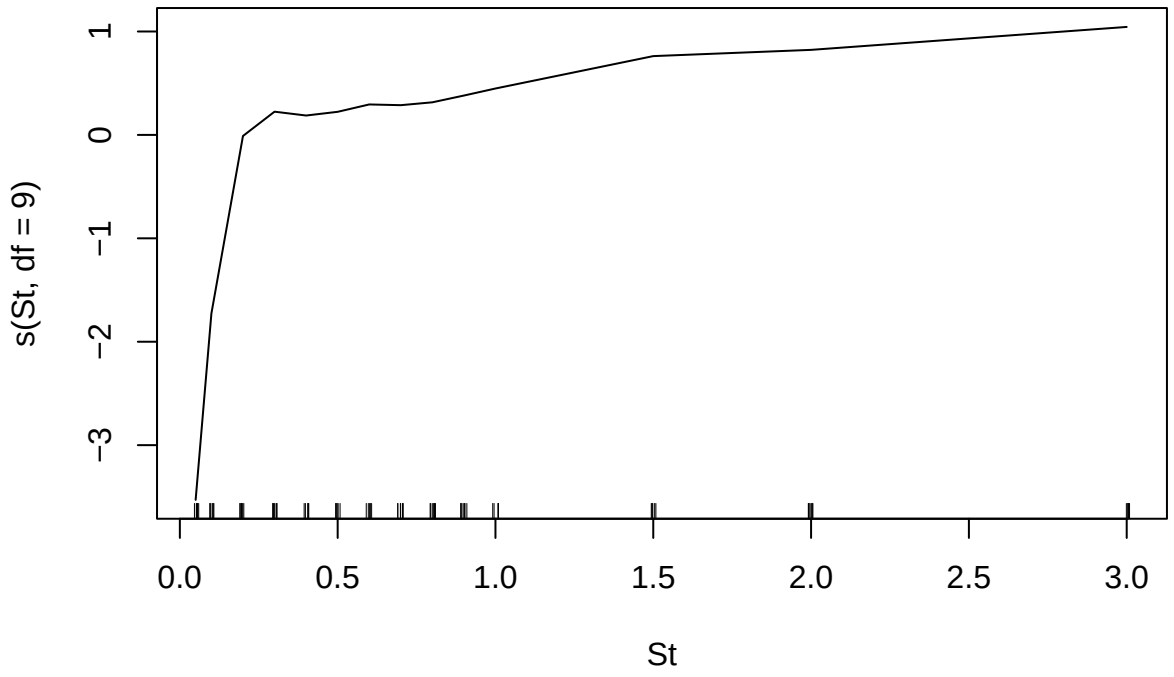
```
## Call: gam(formula = log(M2) ~ s(St, df = 9) + Re_categorical + Fr_categorical +
##     Re_categorical * Fr_categorical, data = train1)
## Deviance Residuals:
##       Min        1Q    Median        3Q       Max
## -3.100803 -0.185164  0.006113  0.231614  1.735705
##
## (Dispersion Parameter for gaussian family taken to be 0.5064)
##
##     Null Deviance: 1212.424 on 88 degrees of freedom
## Residual Deviance: 36.461 on 72.0001 degrees of freedom
## AIC: 209.1479
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##                               Df Sum Sq Mean Sq F value    Pr(>F)
## s(St, df = 9)                  1  49.47   49.47  97.699 4.819e-15 ***
## Re_categorical                 2 722.42  361.21 713.289 < 2.2e-16 ***
## Fr_categorical                 2 205.08  102.54 202.487 < 2.2e-16 ***
## Re_categorical:Fr_categorical  3 151.24   50.41  99.550 < 2.2e-16 ***
## Residuals                     72  36.46    0.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                               Npar Df Npar F     Pr(F)
## (Intercept)
## s(St, df = 9)                       8 20.692 6.661e-16 ***
## Re_categorical
## Fr_categorical
## Re_categorical:Fr_categorical
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(M2)
```

```
## Warning in preplot.Gam(x, terms = terms): No terms saved for "a:b" style
## interaction terms
```

```
## Warning in gplot.default(x = c("Medium", "Medium", "Medium", "Low", "High", :
## The "x" component of "partial for Re_categorical" has class "character"; no
## gplot() methods available
```

```
## Warning in gplot.default(x = c("Low", "Low", "High", "High", "High", "Medium", :
## The "x" component of "partial for Fr_categorical" has class "character"; no
## gplot() methods available
```




```
M3 <- gam(log(M3) ~ s(St, df = 9) + Re_categorical + Fr_categorical +
                    Re_categorical * Fr_categorical, data = train1)
summary(M3)
```

```
##
## Call: gam(formula = log(M3) ~ s(St, df = 9) + Re_categorical + Fr_categorical +
##     Re_categorical * Fr_categorical, data = train1)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -5.97775 -0.35007  0.03625  0.34668  3.06602
##
## (Dispersion Parameter for gaussian family taken to be 1.4732)
##
##     Null Deviance: 2837.669 on 88 degrees of freedom
## Residual Deviance: 106.0723 on 72.0001 degrees of freedom
## AIC: 304.189
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##                               Df  Sum Sq Mean Sq F value    Pr(>F)
## s(St, df = 9)                  1   94.67   94.67  64.263 1.429e-11 ***
## Re_categorical                 2 1162.83  581.42 394.656 < 2.2e-16 ***
## Fr_categorical                 2  798.47  399.24 270.994 < 2.2e-16 ***
## Re_categorical:Fr_categorical  3  542.39  180.80 122.722 < 2.2e-16 ***
## Residuals                     72  106.07    1.47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                               Npar Df Npar F     Pr(F)
## (Intercept)
```

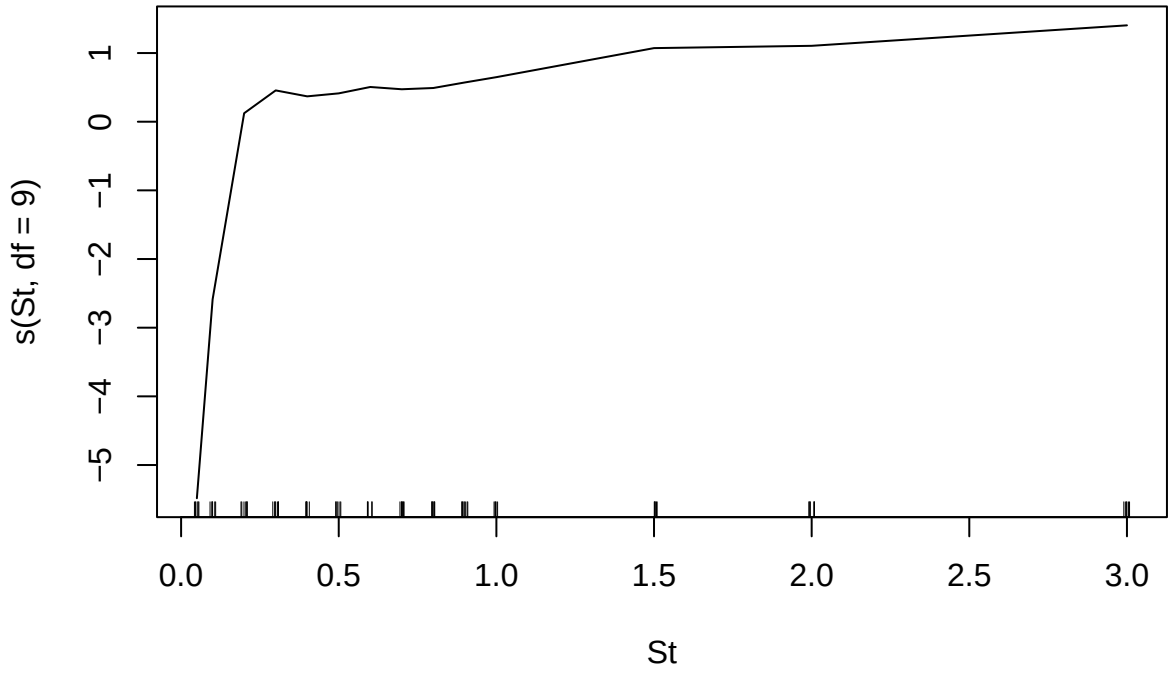
```
## s(St, df = 9)                       8 17.921 2.087e-14 ***
## Re_categorical
## Fr_categorical
## Re_categorical:Fr_categorical
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(M3)
```

```
## Warning in preplot.Gam(x, terms = terms): No terms saved for "a:b" style
## interaction terms
```

```
## Warning in gplot.default(x = c("Medium", "Medium", "Medium", "Low", "High", :
## The "x" component of "partial for Re_categorical" has class "character"; no
## gplot() methods available
```

```
## Warning in gplot.default(x = c("Low", "Low", "High", "High", "High", "Medium", :
## The "x" component of "partial for Fr_categorical" has class "character"; no
## gplot() methods available
```




```
M4 <- gam(log(M4) ~ s(St, df = 9) + Re_categorical + Fr_categorical +
                    Re_categorical * Fr_categorical, data = train1)
summary(M4)
```

```
##
## Call: gam(formula = log(M4) ~ s(St, df = 9) + Re_categorical + Fr_categorical +
##     Re_categorical * Fr_categorical, data = train1)
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

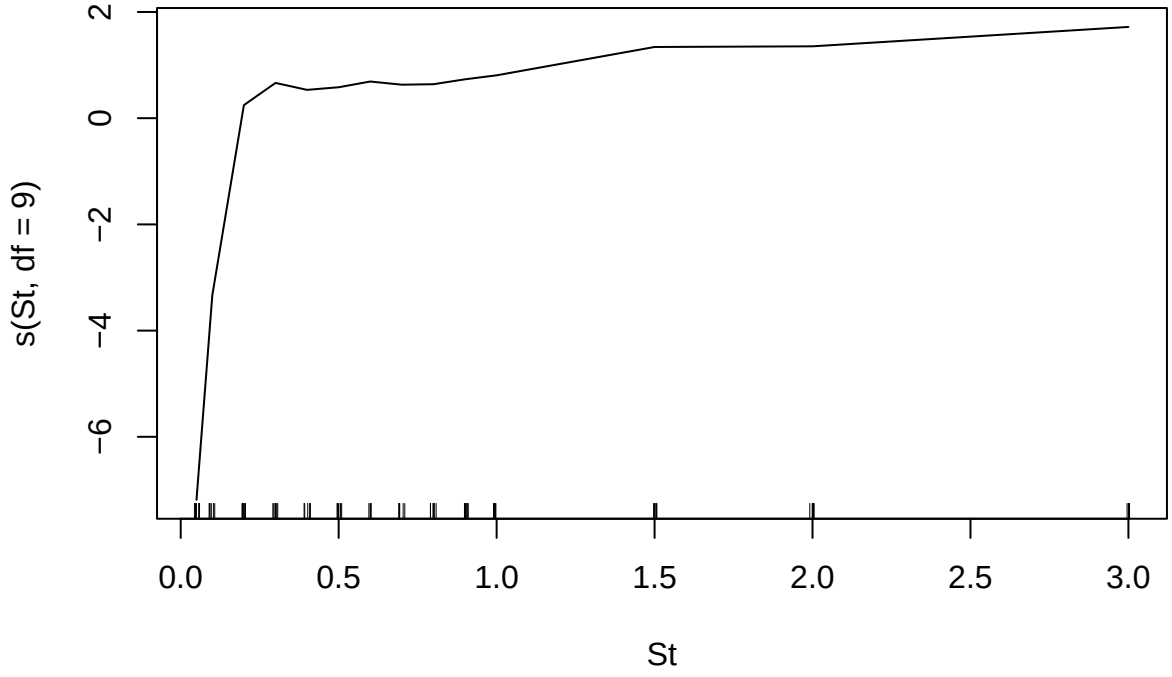
```
## -8.61481 -0.44595  0.04067  0.50573  4.15489
##
## (Dispersion Parameter for gaussian family taken to be 2.7642)
##
##     Null Deviance: 5248.685 on 88 degrees of freedom
## Residual Deviance: 199.0259 on 72.0001 degrees of freedom
## AIC: 360.1979
##
## Number of Local Scoring Iterations: NA
##
## Anova for Parametric Effects
##                              Df  Sum Sq Mean Sq F value    Pr(>F)
## s(St, df = 9)                 1  146.72  146.72  53.076 3.281e-10 ***
## Re_categorical                2 1732.68  866.34 313.410 < 2.2e-16 ***
## Fr_categorical                2 1769.38  884.69 320.048 < 2.2e-16 ***
## Re_categorical:Fr_categorical  3 1162.78  387.59 140.217 < 2.2e-16 ***
## Residuals                     72  199.03    2.76
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Anova for Nonparametric Effects
##                              Npar Df Npar F     Pr(F)
## (Intercept)
## s(St, df = 9)                      8 16.705 1.025e-13 ***
## Re_categorical
## Fr_categorical
## Re_categorical:Fr_categorical
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(M4)
```

```
## Warning in preplot.Gam(x, terms = terms): No terms saved for "a:b" style
## interaction terms
```

```
## Warning in gplot.default(x = c("Medium", "Medium", "Medium", "Low", "High", :
## The "x" component of "partial for Re_categorical" has class "character"; no
## gplot() methods available
```

```
## Warning in gplot.default(x = c("Low", "Low", "High", "High", "High", "Medium", :
## The "x" component of "partial for Fr_categorical" has class "character"; no
## gplot() methods available
```



Conclusion