STA 325 Case Study

Jeff Du, Sahil Tilak

Introduction

Understanding fluid motion and turbulence is one of the most challenging problems that physicists face. This case study will examine the relationship between three key properties of particles (Reynolds Number, Stokes Number, and Froude Number) and the probability distribution of particle cluster volumes. Through exploratory data analysis and nonlinear regression models, we are able to make predictions about the distribution of particle cluster volumes given parameter settings in terms of their four moments as well as better understand the way in which each parameter affects distribution.

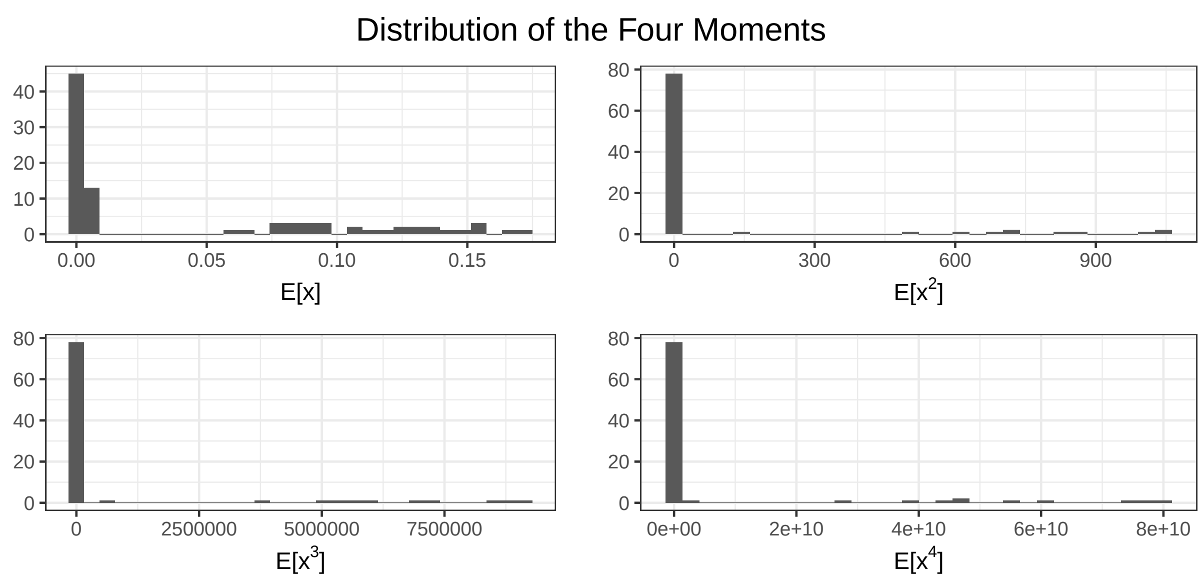
The dataset consists of 7 columns and 89 rows. Each row contains data from a simulation with different particle parameters. There are three predictor variables, or parameters:

Reynolds Number: a measure of the intensity of fluid turbulence of the particle. There are three values of Reynolds Number in the dataset: 90 (baseline value in regression), 224, and 338.

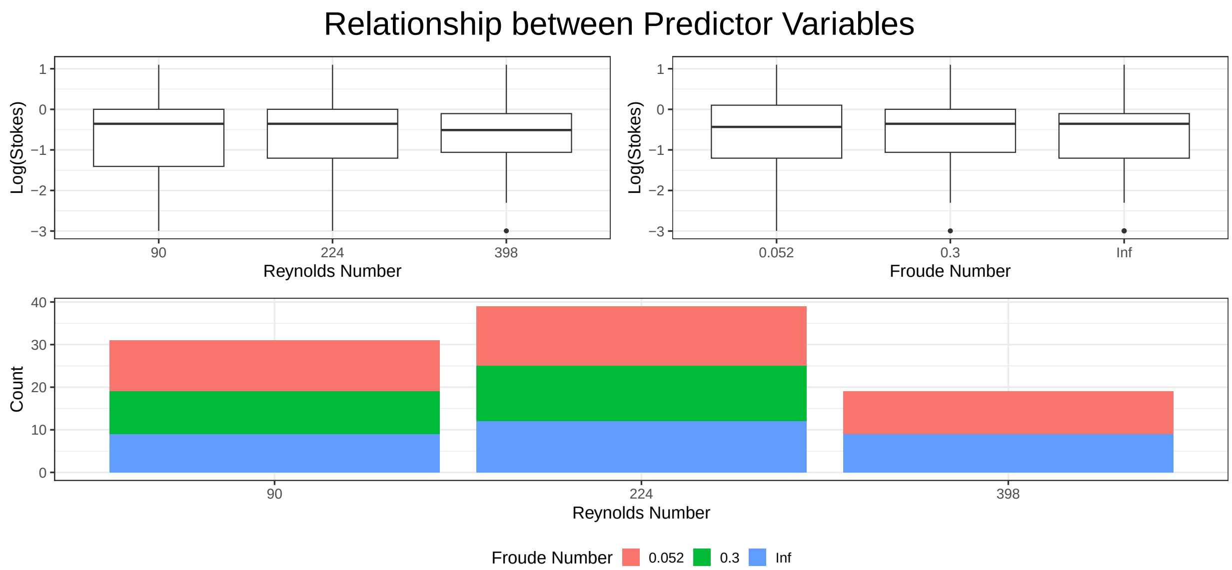
Stokes Number: a measure of the particle size and density. Values lie on the interval $[0, 3]$ in the dataset.

Froude Number: a measure of the gravitational acceleration of the particle. There are three values in the dataset: ∞, 0.3, and 0.052 (baseline value in regression).

There are four response variables, which are each of the four moments ($\mathbb{E}[x]$, $\mathbb{E}[x^2]$, $\mathbb{E}[x^3]$, $\mathbb{E}[x^4]$) of the probability distribution function of Voronoï volumes. We can examine the distributions through histograms.



We can see that all four moments are right skewed and not normally distributed. We also wanted to examine the relationship between the predictors to determine if there is value in including interaction effects. To do so, we create boxplots to illustrate the relationship between the continuous and categorical variables (i.e. the relationship between Stokes Number and Reynolds/Froude Number).



We notice potential relationships between all three combinations of predictor variables, with particular emphasis on the differences in range of log Stokes values based on different Reynolds and Froude numbers. We also notice that there are no particles with a Froude number of 0.3 and a Reynolds number of 398.

Methodology

To examine the relationship between a particle's fluid turbulence, gravitational acceleration, and density on the four moments, we fit four nonlinear regression models with each moment as the response. We considered a number of transformations on each moment, such as a Box-Cox transformation, but ultimately decided to log-transform each moment to have a clear interpretation of our subsequent regression coefficients. We also log-transformed each particle's Stokes number as this predictor was far from normally distributed as well.

Regarding our predictors, we first converted both the Reynolds number and the Froude number of each particle to categorical predictors because there only existed three unique values for each parameter in the dataset and because the numerical differences between such values of Reynolds and Froude numbers were not easily interpretable. We posited from our background research that fitting three interaction, effects for all three predictors would better model the relationship between such predictors. For example, it is well established in existing research that fluid particle acceleration (Fr) is innately related to the turbulence of a flow (Re), in line with the Kolmogorov microscales. Therefore, we included all three potential interactions in our model (Stokes number interactions are log-transformed).

We also considered adding higher order polynomials of the log of the Stokes number. We ran analysis of variance tests fitting models of varying degrees of log of the Stokes number and found that the quartic fit appeared to be reasonable for all moments. Therefore, our general model for each moment is as follows:

$$\begin{aligned} Y = \beta_0 &+ \beta_1 log(Stokes) + \beta_2 log(Stokes)^2 + \beta_3 log(Stokes)^3 + \beta_4 log(Stokes)^4 + \beta_5 Reynolds \\ &+ \beta_6 Froude + \beta_7 (log(Stokes) * Froude) + \beta_8 (log(Stokes) * Reynolds) + \beta_9 (Froude * Reynolds) \end{aligned}$$

Finally, we ran both stepwise forward and backward variable selections on each of our four models using AIC as our criteria. We personally care less about penalizing more complex models since we are given so few predictors to begin with. Forward and backward selection did not remove or add any variables to the second, third, and fourth moment models, but did remove the interaction term between the Reynolds number and the log of the Stokes number for the first moment model. However, we still decided to include this interaction because we believe it is scientifically grounded.

Results

Here we display the results for our four regression models.

Table 1: Model 1

	Estimate	Std. Error	Pr(>|t|)
Intercept	-2.05	0.0199	7.05e-81
Log(St)	1.5	0.141	1.84e-16
Log(St)^2	0.196	0.0698	0.0064
Log(St)^3	0.471	0.0702	3.71e-09
Log(St)^4	-0.153	0.0706	0.0334
Re 224	-3.81	0.0288	9.75e-89
Re 398	-6.01	0.0311	1.22e-100
Fr 0.3	-0.237	0.031	5.94e-11
Fr Inf	-0.272	0.0327	3.64e-12
Fr 0.3 * Log(St)	0.102	0.0184	4.7e-07
Fr Inf * Log(St)	0.0647	0.016	0.000123
Re 224 * Log(St)	0.0136	0.0155	0.382
Re 398 * Log(St)	-0.0159	0.0203	0.436
Re 224 * Fr 0.3	0.261	0.0401	8.57e-09
Re 244 * Fr Inf	0.381	0.0409	4.68e-14
Re 398 * Fr Inf	0.488	0.0446	4.81e-17

Table 3: Model 3

	Estimate	Std. Error	Pr(>|t|)
Intercept	14.6	0.252	9.39e-63
Log(St)	19.3	1.79	1.07e-16
Log(St)^2	-8.55	0.886	1.14e-14
Log(St)^3	6.72	0.891	1.01e-10
Log(St)^4	-3.61	0.895	0.000133
Re 224	-11.5	0.365	3.31e-44
Re 398	-18.1	0.395	1.76e-55
Fr 0.3	-13	0.393	8.41e-46
Fr Inf	-13	0.415	4.19e-44
Fr 0.3 * Log(St)	0.14	0.234	0.55
Fr Inf * Log(St)	-0.622	0.202	0.00298
Re 224 * Log(St)	-0.312	0.196	0.116
Re 398 * Log(St)	-1.08	0.257	7.65e-05
Re 224 * Fr 0.3	8.97	0.508	4.9e-28
Re 244 * Fr Inf	8.74	0.519	7.14e-27
Re 398 * Fr Inf	13.2	0.566	1.21e-35

Table 2: Model 2

	Estimate	Std. Error	Pr(>|t|)
Intercept	6.01	0.146	3.15e-52
Log(St)	12.1	1.04	2.4e-18
Log(St)^2	-5.01	0.514	7.86e-15
Log(St)^3	4.12	0.517	1.58e-11
Log(St)^4	-2.16	0.52	8.74e-05
Re 224	-7.61	0.212	4.05e-48
Re 398	-12	0.229	1.28e-59
Fr 0.3	-6.74	0.228	2.24e-42
Fr Inf	-6.76	0.241	7.56e-41
Fr 0.3 * Log(St)	0.108	0.136	0.428
Fr Inf * Log(St)	-0.327	0.118	0.00694
Re 224 * Log(St)	-0.136	0.114	0.236
Re 398 * Log(St)	-0.627	0.149	7.48e-05
Re 224 * Fr 0.3	4.75	0.295	1.01e-25
Re 244 * Fr Inf	4.68	0.301	7.22e-25
Re 398 * Fr Inf	6.99	0.329	5.15e-33

Table 4: Model 4

	Estimate	Std. Error	Pr(>|t|)
Intercept	23.2	0.348	3.92e-67
Log(St)	25.6	2.48	6.41e-16
Log(St)^2	-11.6	1.22	2.03e-14
Log(St)^3	9.01	1.23	2.59e-10
Log(St)^4	-4.87	1.24	0.000187
Re 224	-15.4	0.504	2.92e-43
Re 398	-24.1	0.545	1.97e-54
Fr 0.3	-19.2	0.543	8.88e-48
Fr Inf	-19.2	0.573	4.96e-46
Fr 0.3 * Log(St)	0.153	0.323	0.637
Fr Inf * Log(St)	-0.901	0.28	0.0019
Re 224 * Log(St)	-0.468	0.271	0.0888
Re 398 * Log(St)	-1.46	0.355	0.000103
Re 224 * Fr 0.3	13.1	0.702	1.85e-29
Re 244 * Fr Inf	12.7	0.717	3.12e-28
Re 398 * Fr Inf	19.4	0.782	2.6e-37

We performed model diagnostics for each model and determined that there were no substantial concerns on homoscedasticity and normality of residuals. We assume independence and linearity based on study design. Model diagnostic plots are included in the appendix. The adjusted R-squared values for each model were .9991, .9814, .9765, and .9757 respectively, indicating that essentially all of the variation in the log of the four moments can be explained by the combination of the predictor variables that we chose. Using an alpha of .05 as a threshold for statistical significance, we find that every individual predictor is statistically significant in each of the four models, though both the sign and magnitudes of the coefficients vary by moment. We also notice that most of the interaction terms are also statistically significant, with the two clear exceptions being the interaction between a Froude number of 0.3 and log(Stokes) and the interaction between a Reynolds number of 224 and log(Stokes).

Let's first interpret the individual regression coefficients, starting with Stokes number. Recall that a higher Stokes number represents a larger and denser particle. We modeled a non-linear relationship between the log of the Stokes number and the moments, and so an interpretation of, say, a .1 unit increase in the Stokes number holding the Reynolds and Froude numbers constant, is not so clear. However, we do notice that for the second, third, and fourth moment model, the degree one and three coefficients are positive, while the degree two and four coefficients are negative.

Next, we can examine the Reynolds number. Recall that high Reynolds numbers indicate that a particle's flow is turbulent and unpredictable. The coefficients for the standalone indicator values of Reynolds are negative, but the magnitude of the Reynolds indicator = 398 is larger than the Reynolds indicator = 224. Holding the size/density constant, a particle with Froude = 0.052 (baseline value) that is more turbulent (larger Froude values) is expected to have smaller moments than a particle that is less turbulent.

Finally, let's now examine the Froude number. Recall that high Froude numbers indicate that gravity can be neglected. The coefficients for the indicator variables of Froude are negative and essentially have the same magnitude for the two Froude values for all models, indicating that holding the size/density constant, a particle with Reynolds = 90 (baseline value) that experiences less influence from gravity (larger Froude values) is expected to have smaller moments than a particle that experiences influence from gravity.

Let's now examine the interaction between a particle's size and its gravitational influence. We see that the interaction coefficient between log(Stokes) and Froude = 0.3 is positive for all four models, but that the coefficient between log(Stokes) and a Froude value of infinity is negative for the second, third, and fourth models. For particles that have such a high Froude value that they are in free fall, our model predicts that as they become larger and denser, holding turbulence constant, their second, third, and fourth moments are expected to become smaller. This is in contrast to particles that experience more gravitational influence (Froude = 0.3), where the second, third, and fourth moments of the distribution are expected to become larger. However, this interaction was not significant.

We can also examine the interaction between a particle's size and its turbulence. We see that the interaction coefficient between log(Stokes) and Reynolds = 224 is only positive for the first moment model, but that the coefficient between log(Stokes) and Reynolds = 398 is negative for all the models (recall that the baseline value of Reynolds = 90). For particles that have such a high Reynolds value that they are extremely turbulent and unpredictable, our model predicts that as they become larger and denser, holding gravitational influence constant, their second, third, and fourth moments are expected to become smaller. This effect is less pronounced for less turbulent particles (Reynolds = 228); however, this interaction was not significant.

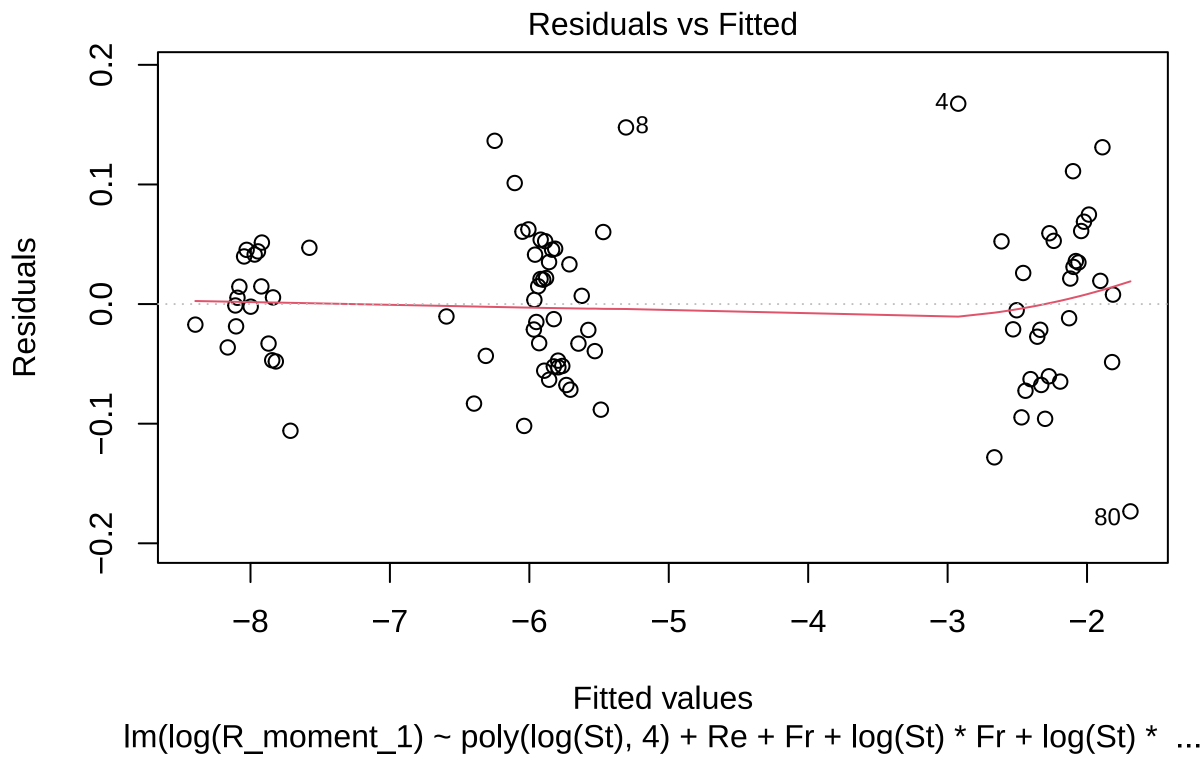
Finally, we can examine the interaction effects between a particle's turbulence and gravitational influence. We notice that all three interactions involving levels of the predictors are positive, but that the magnitude of the interaction term with high turbulence (Reynolds = 398) and no gravitational influence (Froude is infinity) is larger than the interaction term with medium turbulence (Reynolds = 224) and no gravitational influence. This suggests that holding the size/density of the particle constant, particles with high turbulence and no gravitational influence are expected to have larger moments than particles with less turbulence.

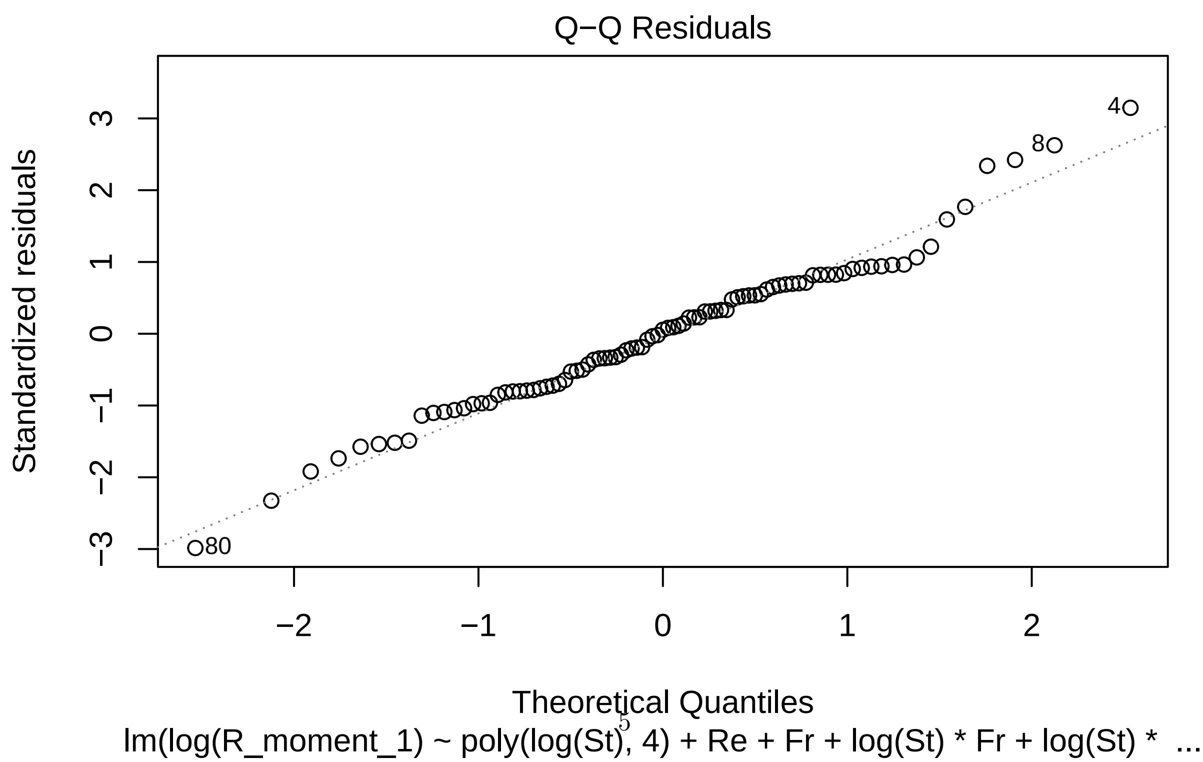
Conclusion

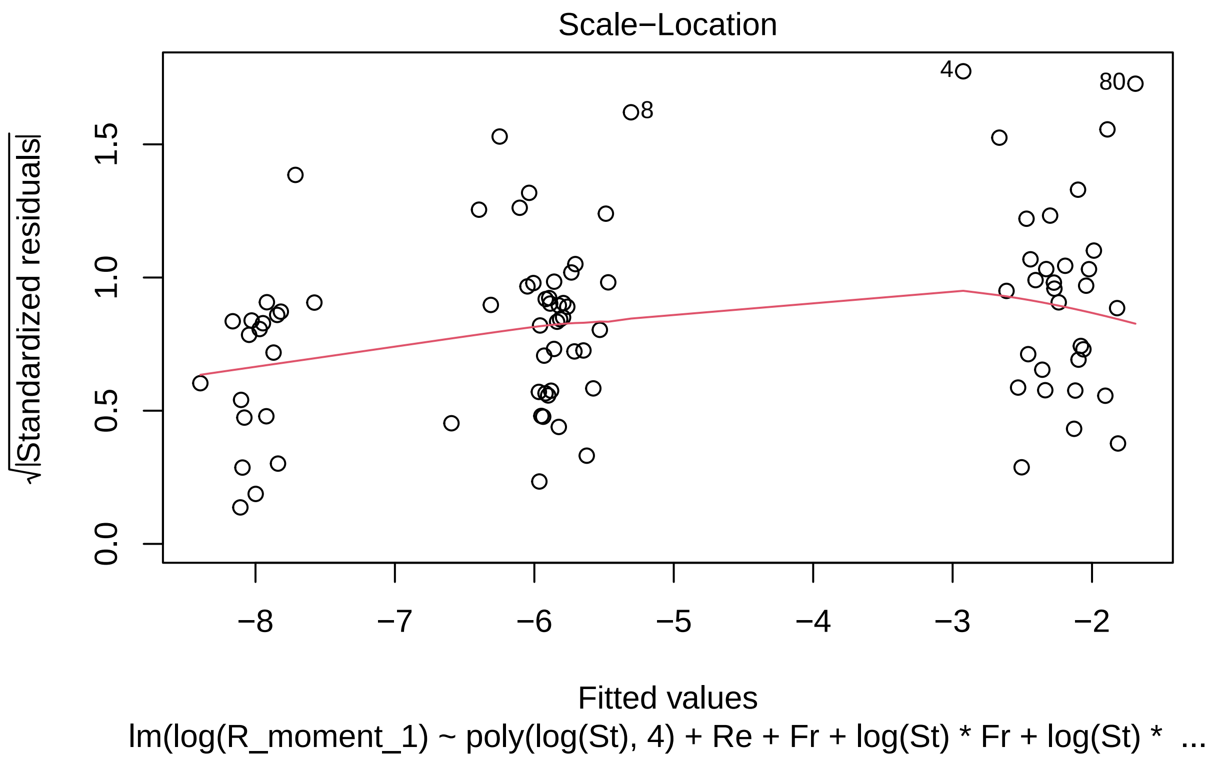
Through our analysis, we were able to examine how specific characteristics of a particle, such as its size/density, turbulence, and gravitational acceleration, effects the distribution of cluster volumes in terms of its four moments. We were able to examine the individual effects of each parameter as well as the interactions between each parameter, culminating in four separate models that were able to explain essentially all of the variation in each moment. We hope that our models can be utilized to not only predict the moments of a particle's cluster distribution but also serve as statistical support for scientifically grounded claims regarding particle turbulence, though we would need to obtain more data points to have a stronger conviction in our results.

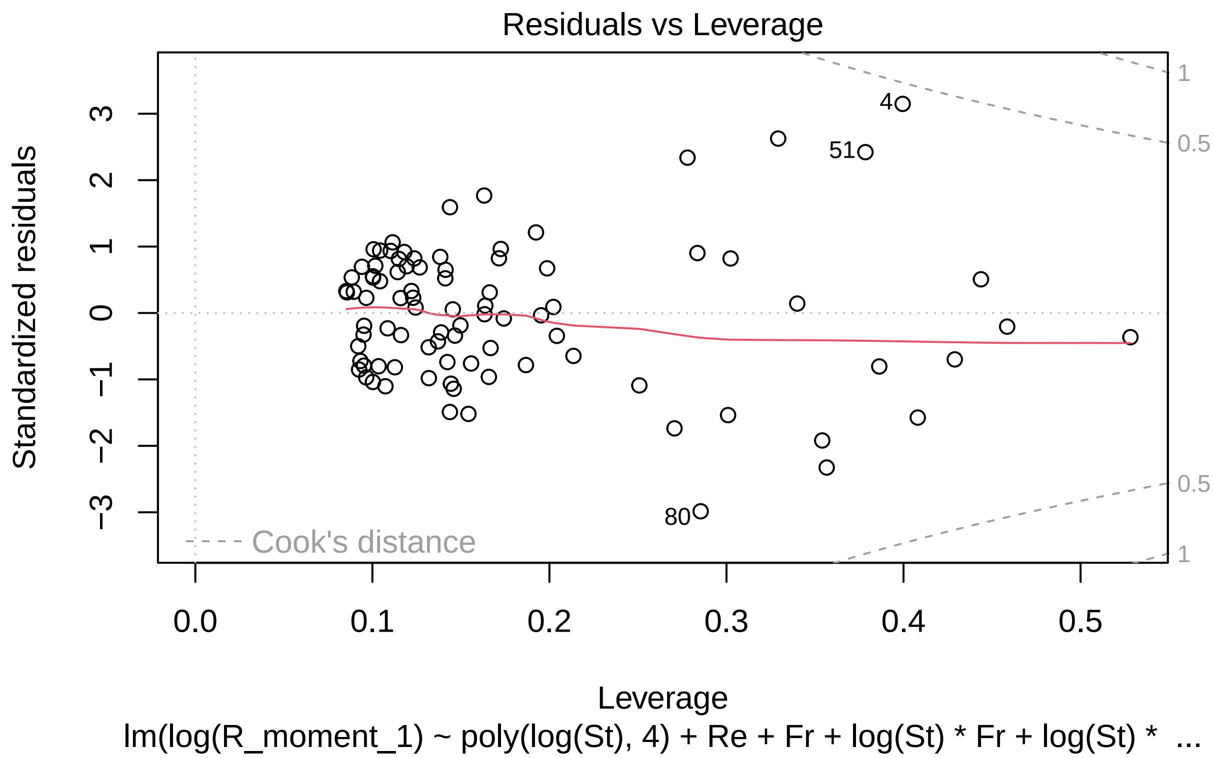
Appendix

First Moment Model Diagnostics



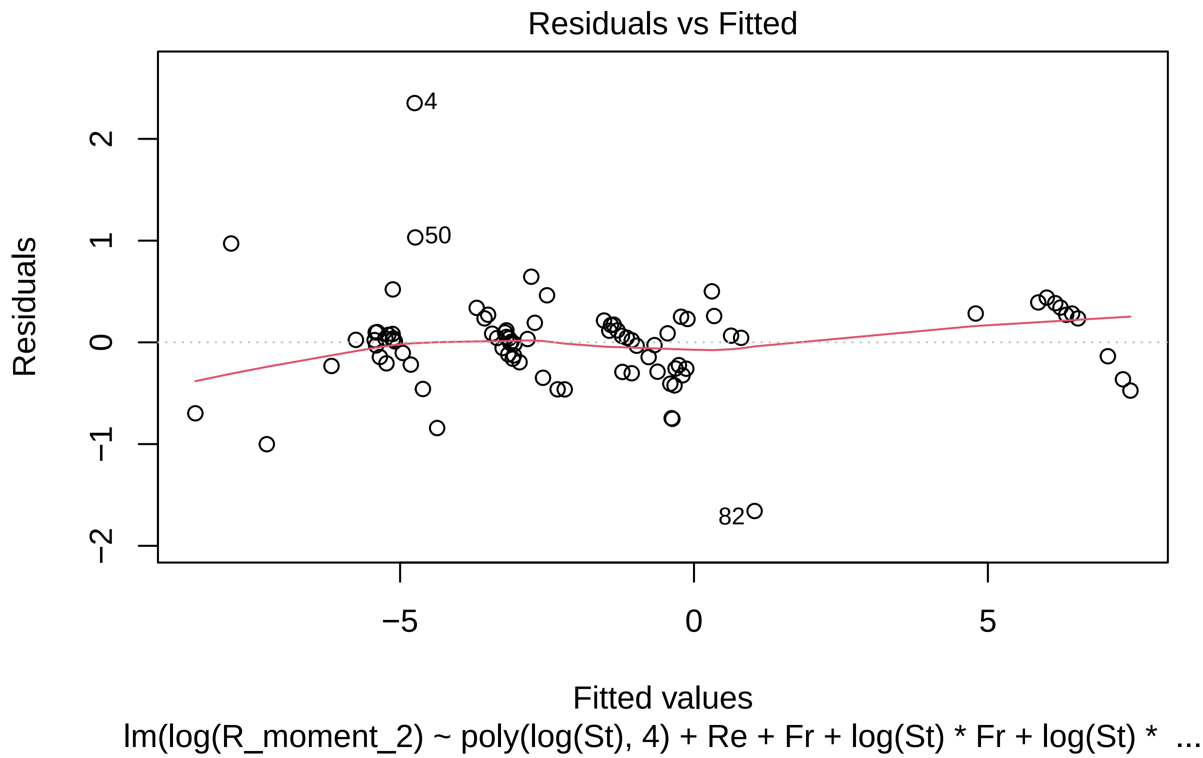


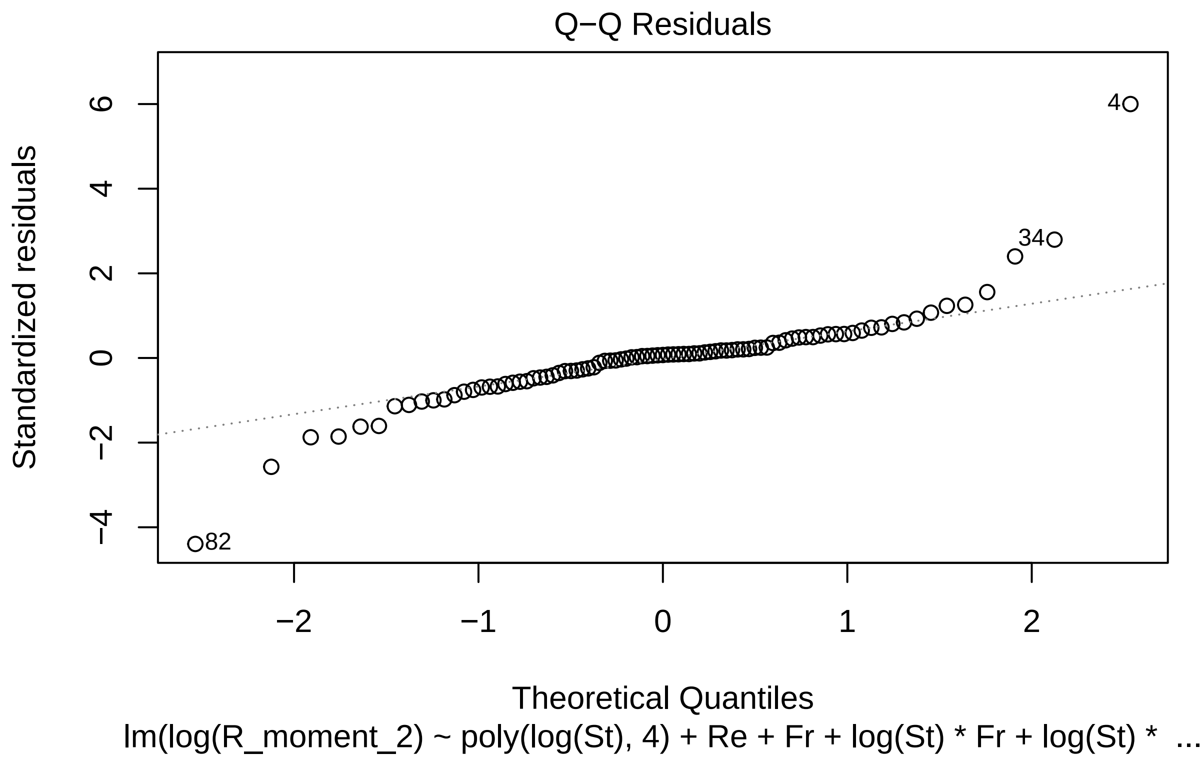

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
4
80
8
−8
−7
−6
−5
−4
−3
−2
Fitted values
lm(log(R_moment_1) ~ poly(log(St), 4) + Re + Fr + log(St) * Fr + log(St) * ...

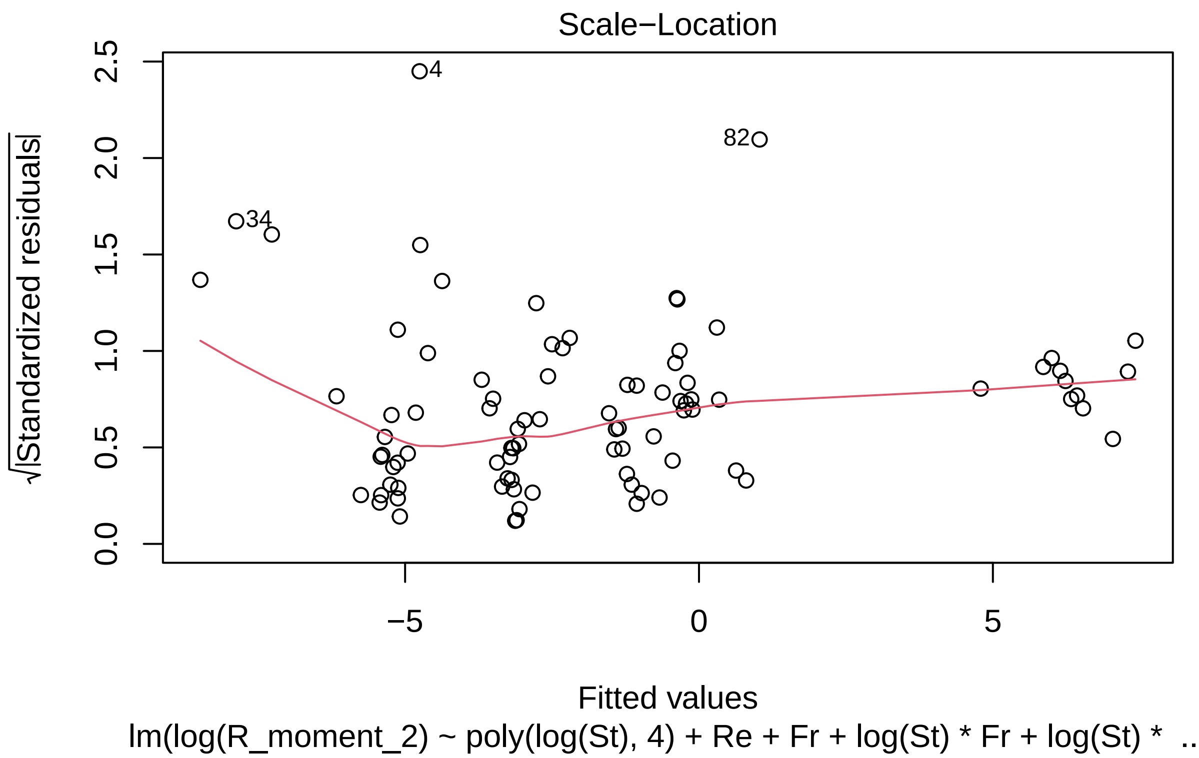


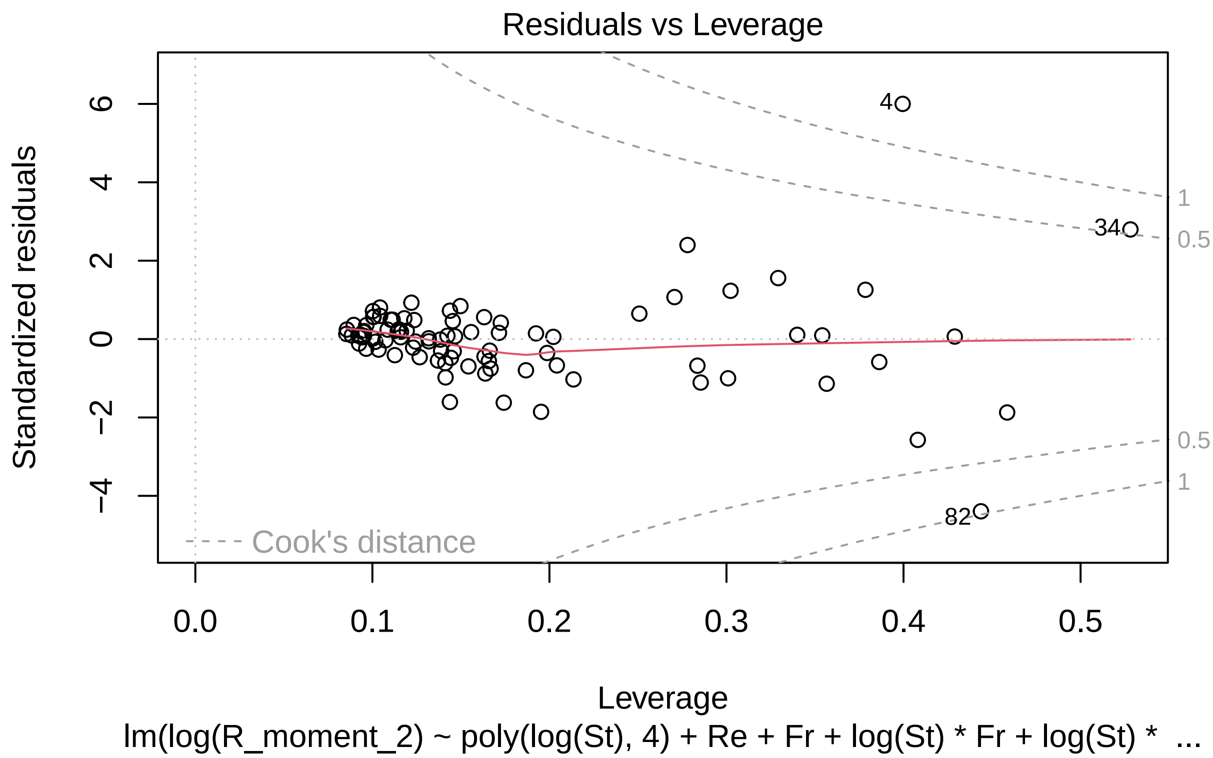
Residuals vs Leverage
Standardized residuals
Leverage
lm(log(R_moment_1) ~ poly(log(St), 4) + Re + Fr + log(St) * Fr + log(St) * ...
Cook's distance
0.0
0.1
0.2
0.3
0.4
0.5
−3
−2
−1
0
1
2
3
1
0.5
0.5
1
4
51
80

Second Moment Model Diagnostics

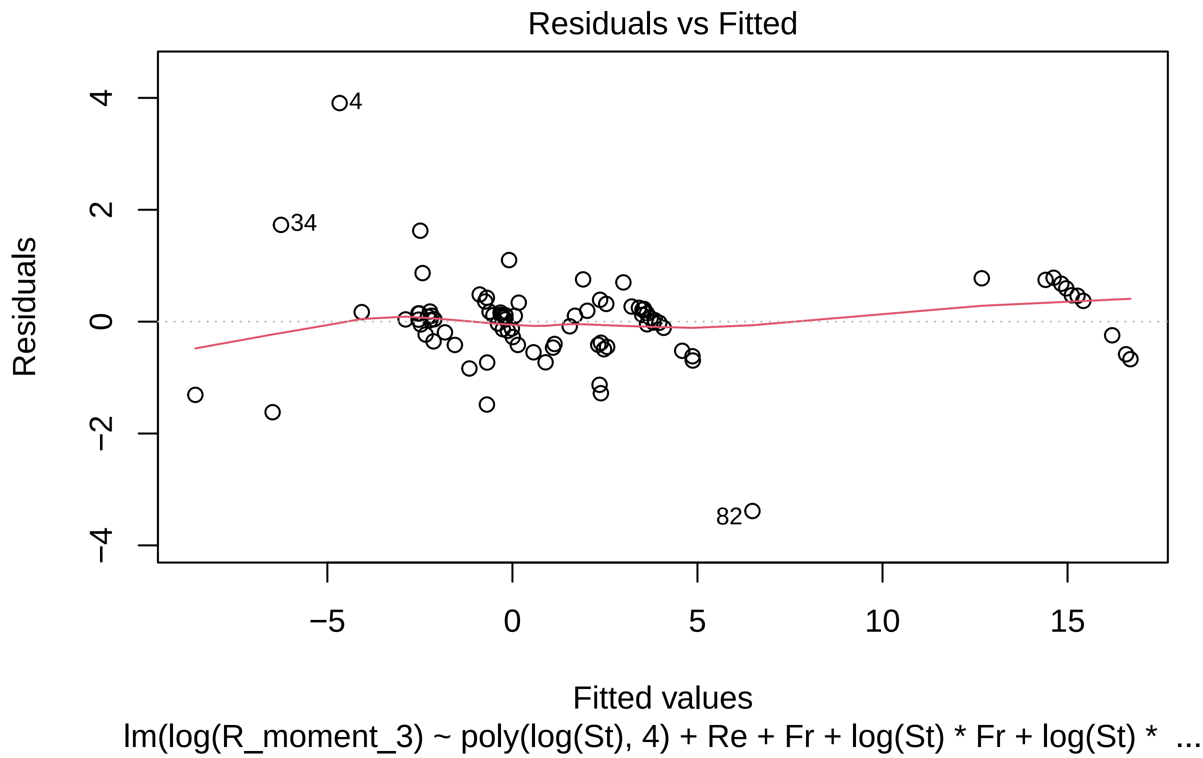


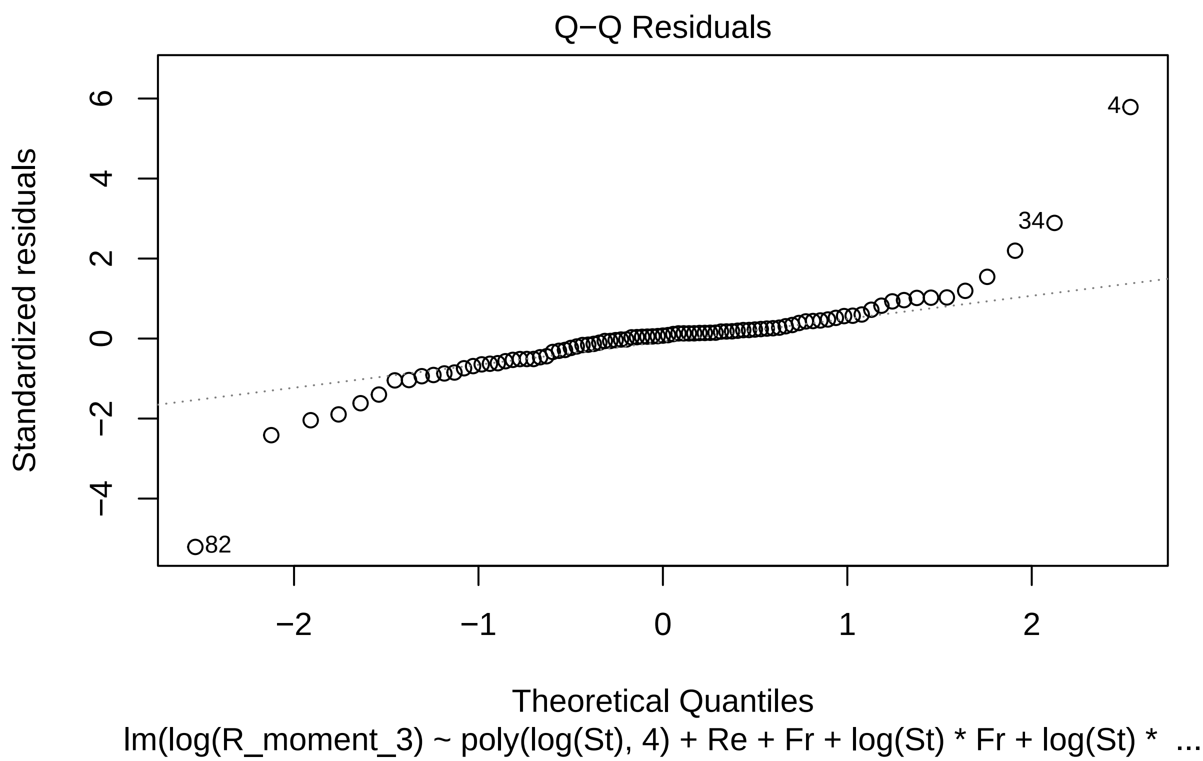


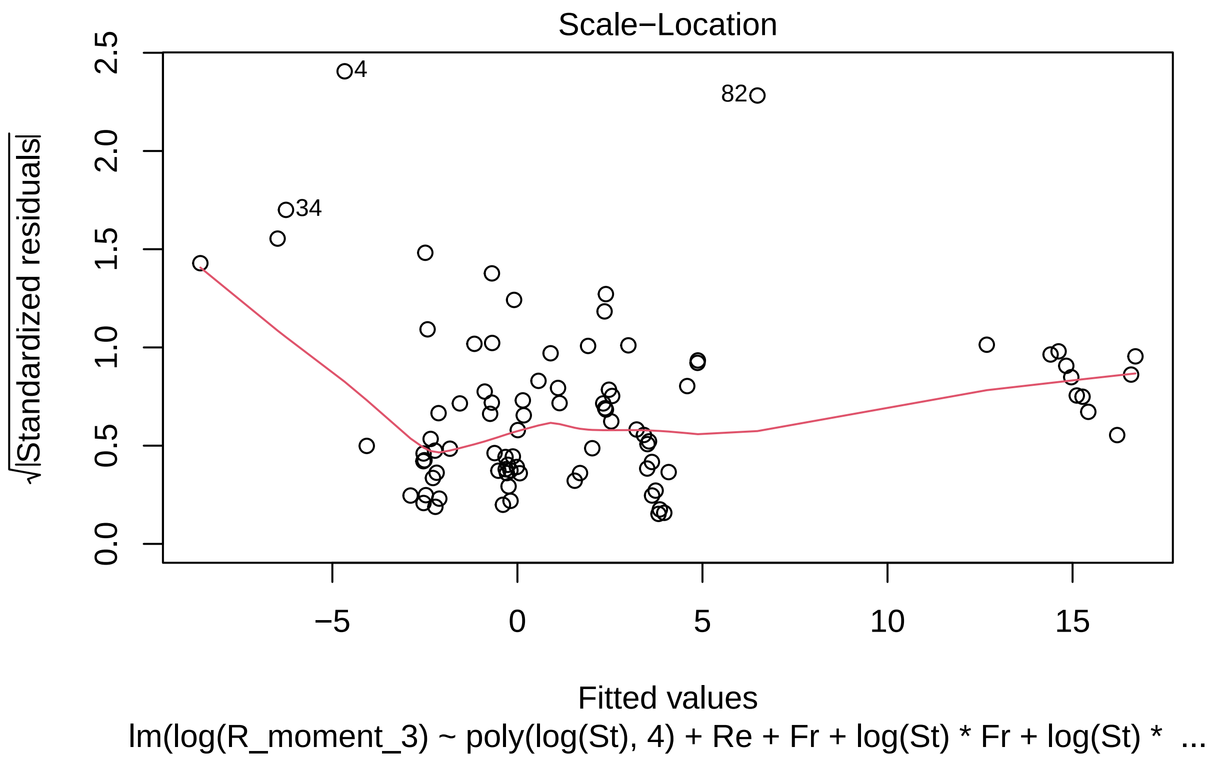

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
−5
0
5
4
82
34
Fitted values
lm(log(R_moment_2) ~ poly(log(St), 4) + Re + Fr + log(St) * Fr + log(St) * ...

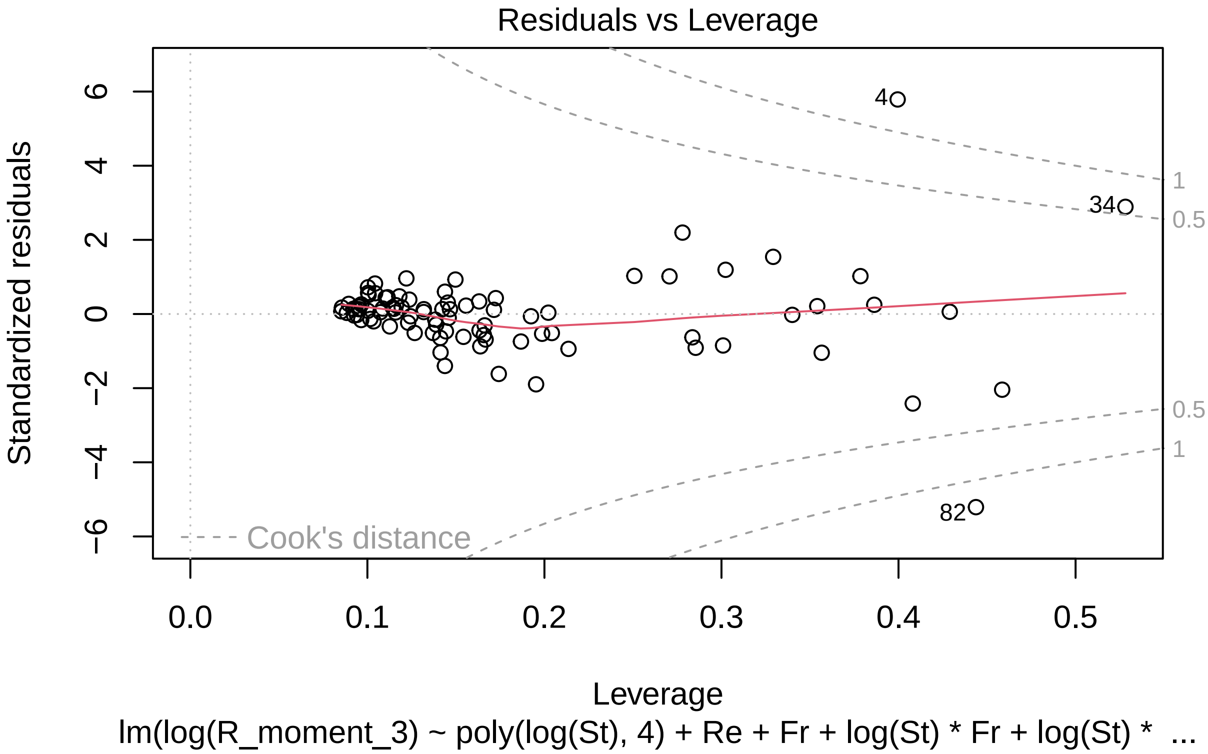

Residuals vs Leverage
Standardized residuals
Leverage
lm(log(R_moment_2) ~ poly(log(St), 4) + Re + Fr + log(St) * Fr + log(St) * ...
Cook's distance
4
34
82
0.5
1
0.0
0.1
0.2
0.3
0.4
0.5
6
4
2
0
−2
−4

Third Moment Model Diagnostics

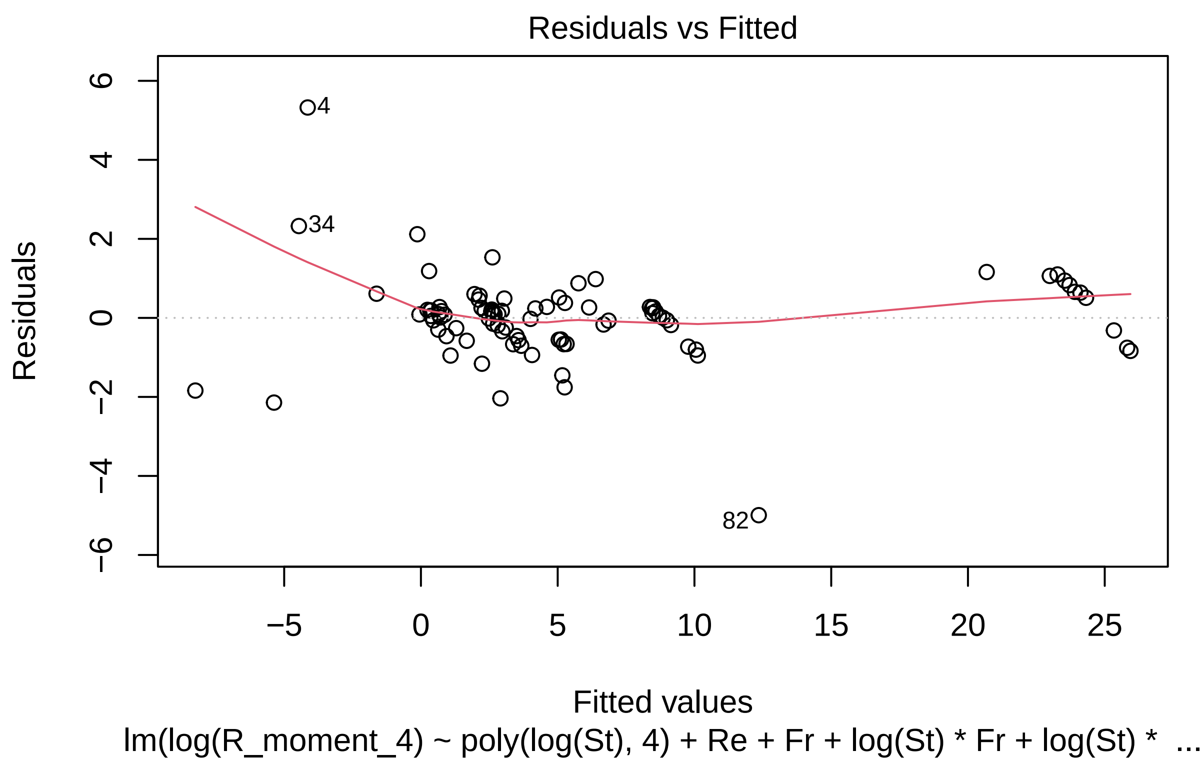


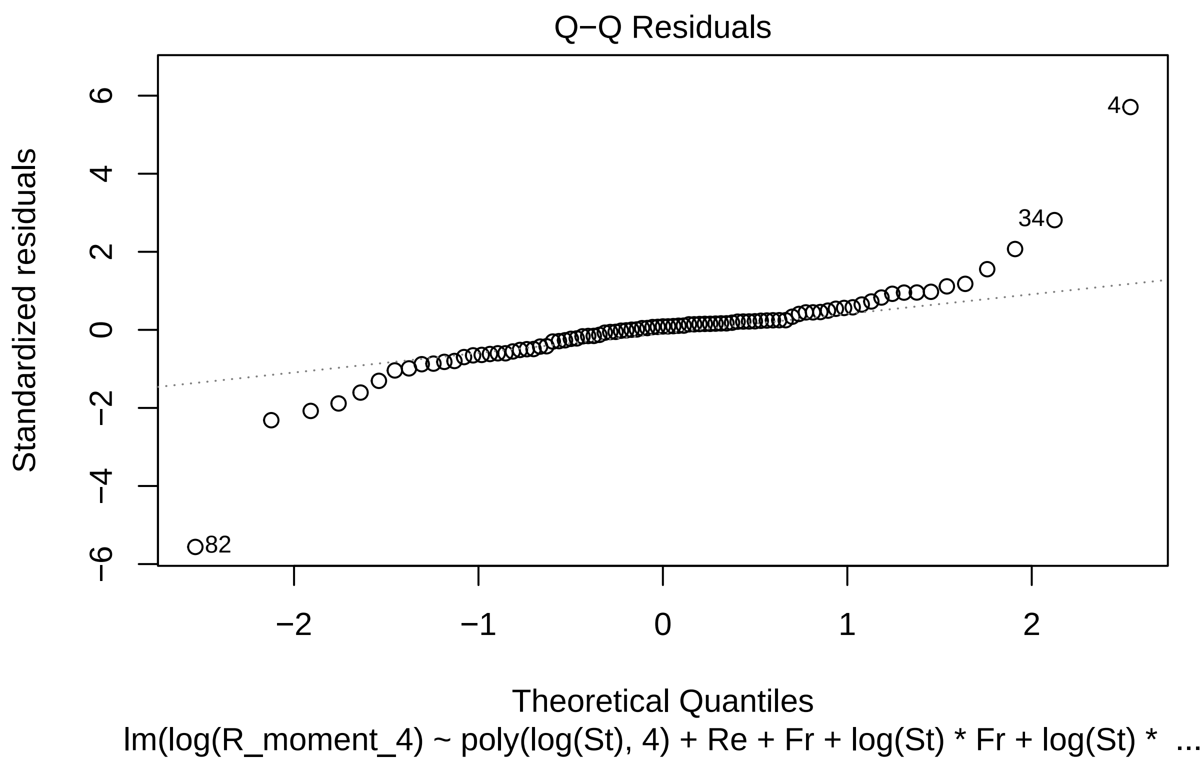


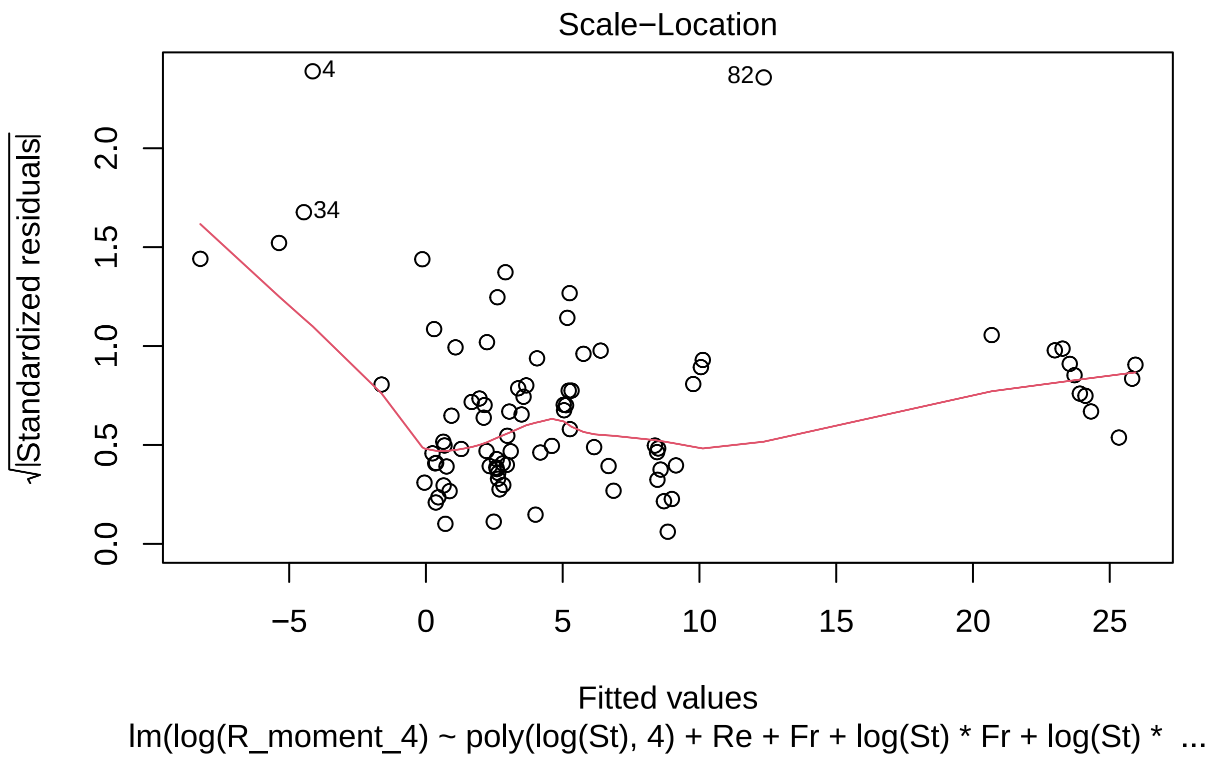

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
4
82
34
−5
0
5
10
15
Fitted values
lm(log(R_moment_3) ~ poly(log(St), 4) + Re + Fr + log(St) * Fr + log(St) * ...

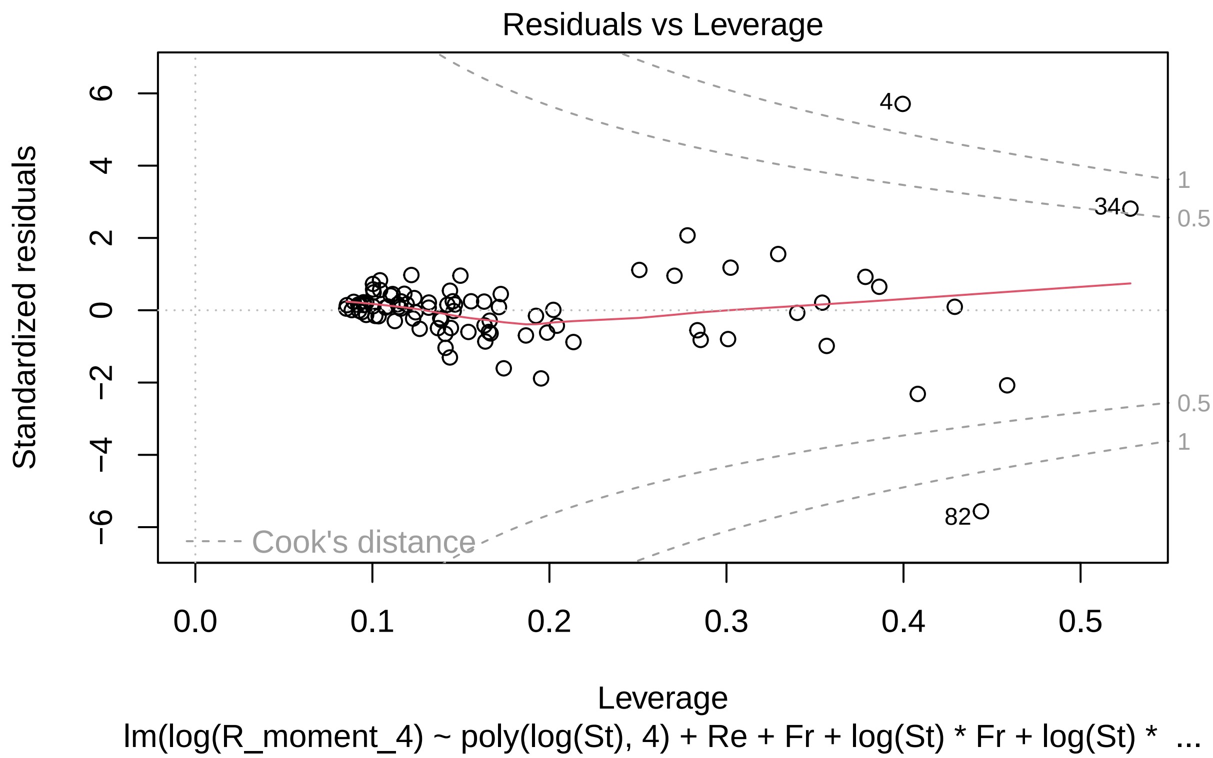

Residuals vs Leverage
Standardized residuals
Leverage
lm(log(R_moment_3) ~ poly(log(St), 4) + Re + Fr + log(St) * Fr + log(St) * ...
Cook's distance
4
34
82
1
0.5
0.5
1
0.0
0.1
0.2
0.3
0.4
0.5
6
4
2
0
−2
−4
−6

Fourth Moment Model Diagnostics






Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
4
82
34
−5
0
5
10
15
20
25
Fitted values
lm(log(R_moment_4) ~ poly(log(St), 4) + Re + Fr + log(St) * Fr + log(St) * ...


Residuals vs Leverage
Standardized residuals
6
4
2
0
−2
−4
−6
4
34
82
1
0.5
0.5
1
Cook's distance
0.0
0.1
0.2
0.3
0.4
0.5
Leverage
lm(log(R_moment_4) ~ poly(log(St), 4) + Re + Fr + log(St) * Fr + log(St) * ...