Case01-Report

Jingxuan Liu, Linda Tang, Justin Zhao, Mary Wang, Abbey List, and Jack Lichtenstein

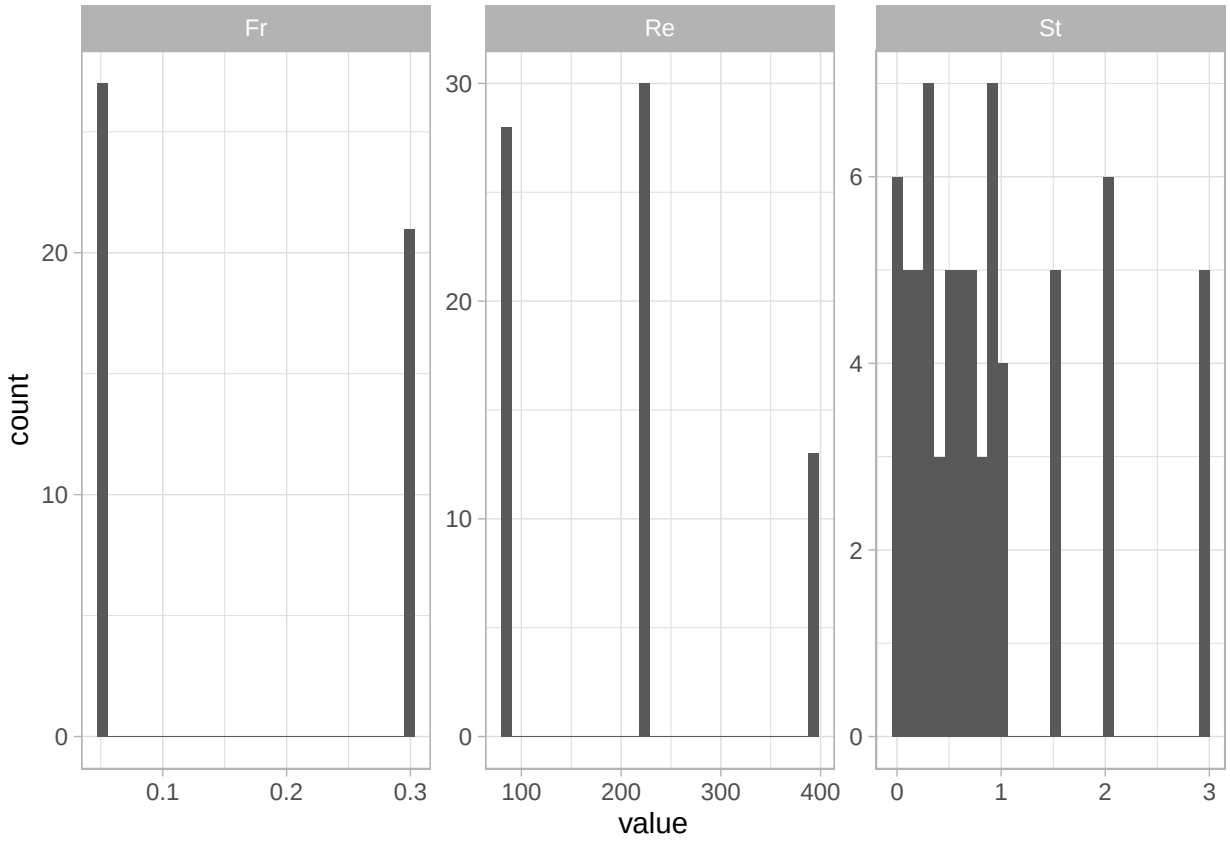
2021-10-10

Introduction

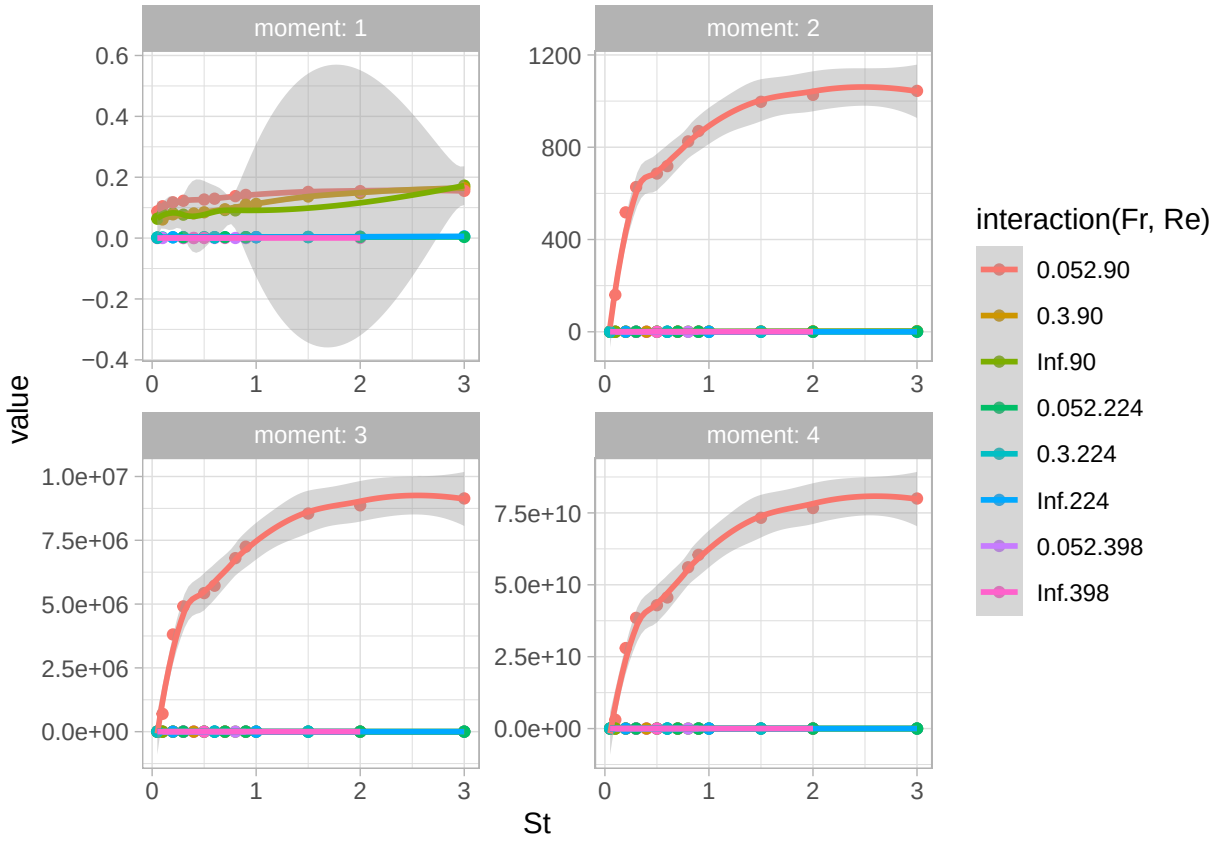
Methods

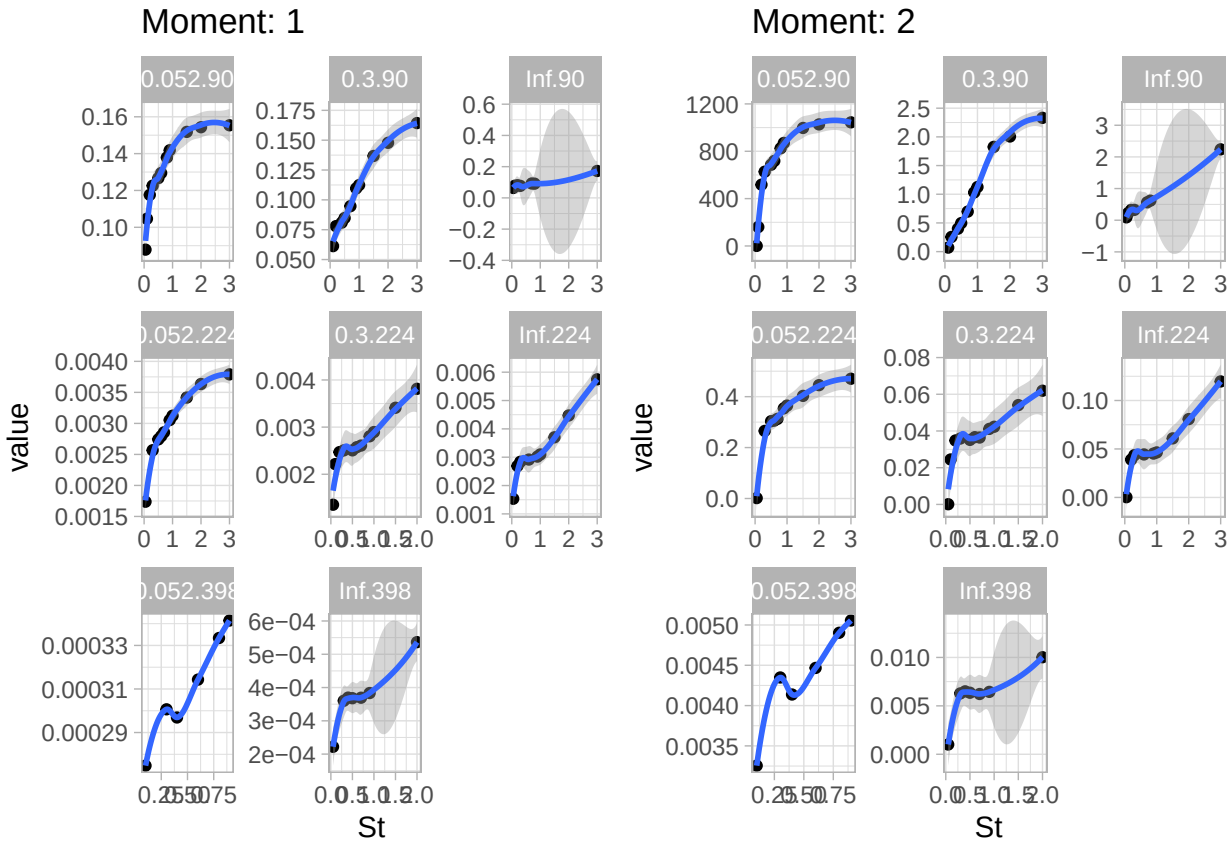
Before we begin any analysis, let's further split our training data into a smaller train and test set. We also create cross-validation folds with $K = 5$. We do this in an effort to reduce the likelihood of overfitting to the full training data.

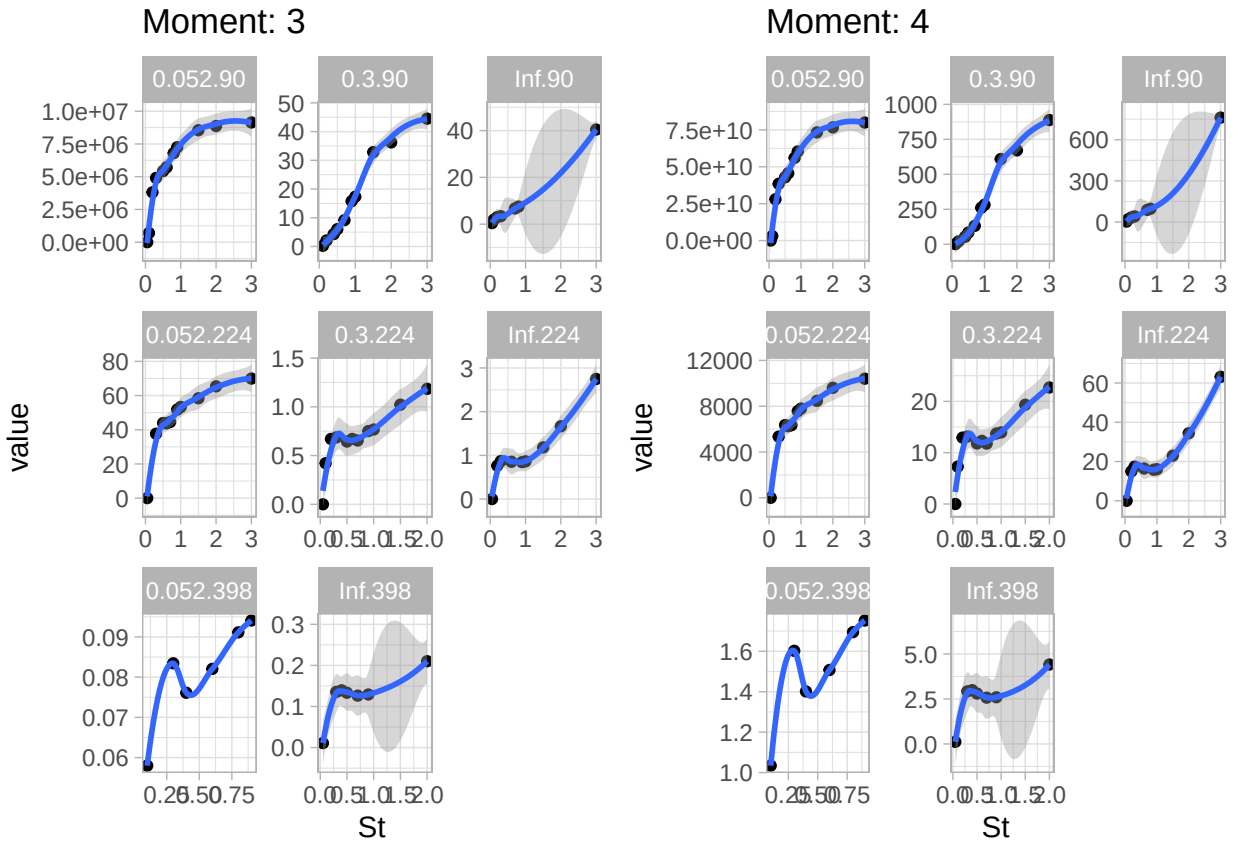
Below we visualize a histogram of the three numerical predictors in these data, `Fr`, `Re`, and `St`. We notice that while the predictors `Fr` and `Re` seem continuous, they are more like factor levels than numeric. As such, we convert both variables to factors.



Below, we plot the relationship between the three predictor variables and each of the four response variables. Note, the curves fit to the points are via local polynomial regression.


moment: 1
moment: 2
moment: 3
moment: 4
value
St
interaction(Fr, Re)
0.052.90
0.3.90
Inf.90
0.052.224
0.3.224
Inf.224
0.052.398
Inf.398

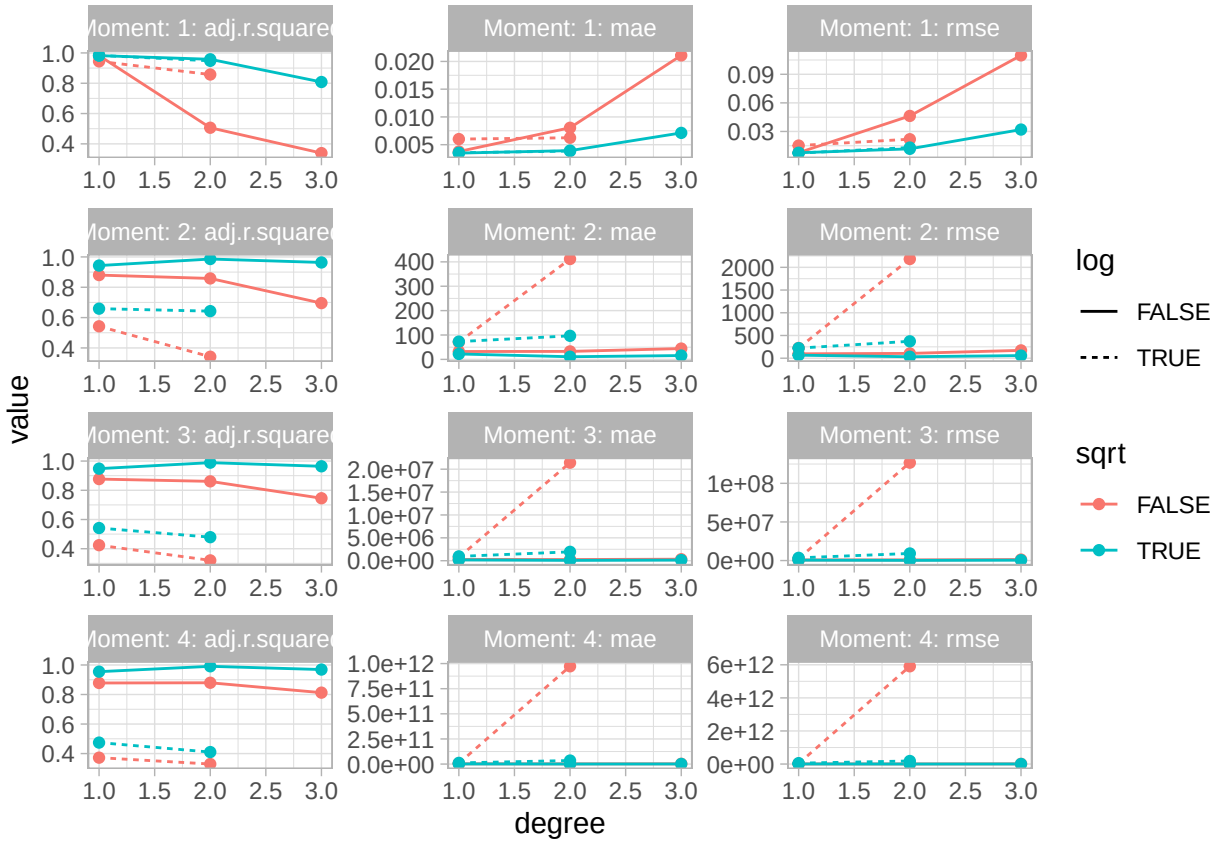

Moment: 1
Moment: 2
0.052.90
0.3.90
Inf.90
0.052.224
0.3.224
Inf.224
0.052.398
Inf.398
value
St



In general, these seem to be very strong relationships. We also notice that the interaction between `Fr` and `Re` seems to explain a lot of the variance in the response. That is, the relationship between the third variable `St` and the response depends a lot on the specific interaction between `Fr` and `Re`. We will very likely need to include this interaction in any model we build for these data.

Additionally, we notice that the relationship between `St` and the response may benefit from taking the square root of `St`.

Below, we perform cross validation of a number of candidate models for each moment. Specifically, for each moment, we train a model to predict the moment with the general formula `~ poly(St, degree)*interaction` where `interaction` is the factor interaction between `Fr` and `Re`. We vary the `degree` parameter from 1 to 3. Additionally, we may choose to take the square root of `St` or take the log of the response.



moment	degree	sqrt	log	name	value
1	1	TRUE	FALSE	mae	0.00348
1	1	TRUE	TRUE	rmse	0.00730
1	1	TRUE	TRUE	adj.r.squared	0.98384
2	2	TRUE	FALSE	rmse	32.42433
2	2	TRUE	FALSE	mae	10.84030
2	2	TRUE	FALSE	adj.r.squared	0.98613
3	2	TRUE	FALSE	rmse	240,845.80324
3	2	TRUE	FALSE	mae	76,364.57174
3	2	TRUE	FALSE	adj.r.squared	0.98910
4	2	TRUE	FALSE	rmse	1,837,866,060.54046
4	2	TRUE	FALSE	mae	558,091,625.13870
4	2	TRUE	FALSE	adj.r.squared	0.99101

Results

Discussion

References

Appendix