Linear Models

Emily Schemanske & Nitanshi Mahajan

4/15/2023

Hugging Face BERTweet sentiment model

```
# read huggingface data
data_hf = read.csv('/Users/landise/Downloads/lm_huggingface.csv')

# remove date column
data_hf <- subset(data_hf, select = c(adidas_PercentChangeMA, djia_PercentChang
eMA, final_score_MA, pn_score_MA))
data_hf <- na.omit(data_hf)

head(data_hf)
```

```
##    adidas_PercentChangeMA djia_PercentChangeMA final_score_MA pn_score_MA
## 26             0.05230769            0.2015385      0.8596154   0.2671154
## 27             0.36538462            0.1503846      0.8584615   0.2725000
## 28             0.13038462            0.1846154      0.8580769   0.3425000
## 29             0.27192308            0.3538462      0.8557692   0.4082692
## 30             0.55192308            0.2976923      0.8553846   0.4055769
## 31             0.70923077            0.1519231      0.8557692   0.4101923
```

```
# scale pn_score_MA so all values are between 0 and 1
data_hf$pn_score_MA <- (data_hf$pn_score_MA - min(data_hf$pn_score_MA)) / (max
(data_hf$pn_score_MA) - min(data_hf$pn_score_MA))
head(data_hf, 1000)
```

```
##     adidas_PercentChangeMA djia_PercentChangeMA final_score_MA pn_score_MA
## 26            0.052307692          0.201538462      0.8596154 0.707079646
## 27            0.365384615          0.150384615      0.8584615 0.712035398
## 28            0.130384615          0.184615385      0.8580769 0.776460177
## 29            0.271923077          0.353846154      0.8557692 0.836991150
## 30            0.551923077          0.297692308      0.8553846 0.834513274
## 31            0.709230769          0.151923077      0.8557692 0.838761062
## 32            0.502307692          0.147307692      0.8523077 0.839469027
## 33            0.715000000          0.151153846      0.8515385 0.839469027
## 34            0.010000000          0.054230769      0.8517308 0.894690265
## 35            0.023846154         -0.008461538      0.8523077 0.892920354
## 36            0.318076923          0.147692308      0.8519231 0.923362832
## 37            0.463846154          0.056153846      0.8523077 0.913097345
## 39            0.511923077         -0.011923077      0.8509615 0.877699115
## 40            0.305000000          0.015384615      0.8521154 0.821415929
## 41            0.583076923          0.135000000      0.8501923 0.821061947
## 42            0.576923077         -0.010384615      0.8494231 0.876991150
## 44            0.811538462         -0.090000000      0.8432692 0.815929204
## 45            0.258076923         -0.358461538      0.8482692 0.755044248
## 47            0.075384615         -0.278846154      0.8515385 0.729557522
## 48            0.110000000         -0.372307692      0.8509615 0.725486726
## 49            0.161153846         -0.300769231      0.8513462 0.775752212
## 50            0.186923077         -0.274615385      0.8523077 0.749911504
## 51            0.494230769         -0.195384615      0.8521154 0.720884956
## 52            0.383846154         -0.116538462      0.8505769 0.668849558
## 53            0.642307692         -0.041153846      0.8496154 0.635929204
## 54            0.595384615         -0.201153846      0.8469231 0.569734513
## 55            0.580769231         -0.063846154      0.8523077 0.563362832
## 56            0.540384615         -0.022692308      0.8519231 0.566548673
## 57            0.650769231          0.022692308      0.8488462 0.503893805
## 58            0.575769231         -0.113461538      0.8503846 0.501769912
## 59            0.862307692         -0.168846154      0.8526923 0.445132743
## 60            0.762692308         -0.059230769      0.8536538 0.450442478
## 61            0.866923077         -0.040384615      0.8488462 0.445840708
## 62            0.705000000         -0.440000000      0.8525000 0.475398230
## 63            1.198076923         -0.304615385      0.8507692 0.505486726
## 64            1.045000000         -0.241923077      0.8486538 0.476637168
## 65            1.177307692         -0.234230769      0.8519231 0.476106195
## 66            0.843076923         -0.312307692      0.8523077 0.532035398
## 67            1.517307692          0.012307692      0.8548077 0.528672566
## 68            1.547692308          0.027692308      0.8540385 0.527964602
## 69            1.483846154          0.237307692      0.8563462 0.531150442
## 70            1.640000000          0.205000000      0.8607692 0.472212389
## 71            1.516153846          0.303461538      0.8584615 0.505309735
## 72            1.508076923          0.301153846      0.8586538 0.479646018
## 73            1.295000000          0.135000000      0.8573077 0.506548673
```

```
## 74            1.287307692          0.175384615       0.8596154 0.562477876
## 75            1.183461538          0.101923077       0.8603846 0.510088496
## 77            1.138846154          0.085769231       0.8596154 0.566725664
## 78            1.281538462          0.036538462       0.8611538 0.622654867
## 79            1.397692308          0.011538462       0.8621154 0.652035398
## 80            1.252307692          0.149615385       0.8640385 0.713982301
## 81            1.310384615          0.139230769       0.8605769 0.655221239
## 82            1.584615385          0.133461538       0.8632692 0.595044248
## 83            1.416923077          0.413846154       0.8671154 0.597522124
## 84            1.973846154          0.562307692       0.8655769 0.537168142
## 85            2.004615385          0.444230769       0.8644231 0.533628319
## 86            1.961923077          0.528461538       0.8617308 0.472743363
## 87            2.090769231          0.590384615       0.8628846 0.473805310
## 88            2.333076923          0.484615385       0.8596154 0.466725664
## 89            1.677307692          0.329230769       0.8596154 0.494867257
## 90            1.709230769          0.198076923       0.8605769 0.521415929
## 91            1.646923077          0.119230769       0.8588462 0.520707965
## 92            1.651923077          0.167307692       0.8580769 0.461769912
## 93            1.614230769          0.116538462       0.8559615 0.406725664
## 94            1.582692308          0.148846154       0.8601923 0.346548673
## 95            1.463846154          0.057307692       0.8586538 0.346194690
## 96            1.265000000          0.088846154       0.8550000 0.342300885
## 97            0.558846154          0.038461538       0.8567308 0.344247788
## 98            0.793461538          0.289230769       0.8575000 0.344955752
## 99            0.670384615          0.301153846       0.8563462 0.289203540
## 100           0.641923077          0.277307692       0.8532692 0.232920354
## 101           0.440769231          0.295384615       0.8528846 0.232920354
## 102           0.574615385          0.400384615       0.8525000 0.202300885
## 103           0.553846154          0.458076923       0.8530769 0.150442478
## 104           0.543846154          0.535384615       0.8519231 0.120707965
## 105           0.479615385          0.381153846       0.8519231 0.120707965
## 106           0.415000000          0.378076923       0.8526923 0.062477876
## 107           0.379615385          0.284615385       0.8546154 0.068141593
## 108           0.338076923          0.262307692       0.8548077 0.127610619
## 109          -0.304230769          0.346538462       0.8521154 0.124070796
## 110          -0.173846154          0.488461538       0.8540385 0.182654867
## 111           0.653461538          0.568076923       0.8528846 0.211858407
## 112           0.541153846          0.474615385       0.8559615 0.206194690
## 113           0.471538462          0.476923077       0.8571154 0.148495575
## 114           0.384230769          0.490000000       0.8582692 0.093097345
## 115           0.876923077          0.560000000       0.8559615 0.064247788
## 116           1.295000000          0.630000000       0.8567308 0.012389381
## 117           1.056923077          0.403461538       0.8544231 0.045309735
## 118           1.165000000          0.394230769       0.8548077 0.107079646
## 119           0.795769231          0.301153846       0.8555769 0.115929204
## 120           1.119230769          0.273461538       0.8519231 0.118407080
```

```
## 121         1.078846154         0.328461538       0.8542308 0.178938053
## 122         1.302307692         0.322307692       0.8565385 0.240353982
## 123         1.582692308         0.372692308       0.8532692 0.240176991
## 124         1.437307692         0.293846154       0.8513462 0.263716814
## 125         1.123846154         0.178846154       0.8530769 0.263716814
## 126         0.824230769         0.164615385       0.8553846 0.327433628
## 127         0.744615385         0.245384615       0.8542308 0.383008850
## 128         0.716538462         0.317692308       0.8559615 0.444424779
## 129         0.680384615         0.315000000       0.8586538 0.499292035
## 130         0.764615385         0.379615385       0.8605769 0.526902655
## 131         0.742307692         0.433846154       0.8605769 0.469911504
## 132         0.423846154         0.436923077       0.8590385 0.470442478
## 133         0.337692308         0.441923077       0.8601923 0.529203540
## 134         0.348461538         0.443076923       0.8580769 0.475752212
## 135         0.553846154         0.449615385       0.8582692 0.536991150
## 136         0.522692308         0.522307692       0.8588462 0.479646018
## 137         0.209615385         0.553076923       0.8584615 0.504247788
## 138         0.212692308         0.570000000       0.8561538 0.569026549
## 139         0.221153846         0.648076923       0.8569231 0.627787611
## 140         0.128846154         0.685384615       0.8575000 0.685663717
## 141         0.040000000         0.771538462       0.8603846 0.715044248
## 142        -0.071153846         0.566923077       0.8592308 0.773451327
## 143         0.037307692         0.592307692       0.8601923 0.802477876
## 144        -0.314230769         0.602307692       0.8605769 0.744070796
## 145         0.483461538         0.220769231       0.8613462 0.795398230
## 146         0.585384615         0.273076923       0.8619231 0.855221239
## 147         0.551538462         0.171153846       0.8571154 0.823539823
## 148         0.313846154         0.146153846       0.8544231 0.821769912
## 149         0.415000000         0.255384615       0.8555769 0.789557522
## 150         0.210000000         0.247692308       0.8582692 0.792212389
## 151         0.256153846         0.220384615       0.8580769 0.794867257
## 152         0.402307692         0.091923077       0.8582692 0.730619469
## 153         0.413461538         0.183076923       0.8588462 0.733097345
## 154         0.652307692         0.097692308       0.8530769 0.731327434
## 155         0.461538462         0.003461538       0.8515385 0.670088496
## 156         0.406153846        -0.039615385       0.8507692 0.669026549
## 157         0.195000000        -0.100000000       0.8511538 0.727433628
## 158         0.053846154        -0.123076923       0.8515385 0.734513274
## 159        -0.118846154        -0.143076923       0.8507692 0.734513274
## 160        -0.069615385         0.017307692       0.8526923 0.760176991
## 161         0.286923077         0.195000000       0.8548077 0.697876106
## 162         0.212307692         0.085769231       0.8538462 0.696814159
## 163         0.490769231         0.090000000       0.8534615 0.644778761
## 164         0.663461538         0.233461538       0.8557692 0.582831858
## 165         0.555769231         0.101153846       0.8553846 0.553097345
## 166         0.054615385         0.195769231       0.8546154 0.493982301
```

```
## 167          0.085000000         0.500384615        0.8546154 0.492920354
## 168          0.060769231         0.366153846        0.8517308 0.462477876
## 169         -0.100000000         0.391923077        0.8526923 0.457168142
## 170         -0.510000000         0.161923077        0.8488462 0.515575221
## 171         -0.351538462         0.161538462        0.8484615 0.515221239
## 172         -0.215000000         0.071153846        0.8500000 0.515575221
## 173          0.153846154         0.169615385        0.8540385 0.484955752
## 174          0.085769231         0.188846154        0.8567308 0.488849558
## 175          0.187307692         0.121538462        0.8561538 0.518584071
## 176          0.106153846        -0.055000000        0.8557692 0.488849558
## 177         -0.154230769        -0.117692308        0.8542308 0.481769912
## 178         -0.250000000        -0.315000000        0.8525000 0.482123894
## 179          0.079230769        -0.128076923        0.8542308 0.425840708
## 180          0.103076923        -0.124230769        0.8580769 0.366371681
## 181          0.221538462        -0.016153846        0.8561538 0.364601770
## 182          0.357692308        -0.071923077        0.8557692 0.361415929
## 183         -0.056538462         0.006153846        0.8561538 0.362831858
## 184         -0.155384615         0.070769231        0.8540385 0.382300885
## 185         -0.095384615        -0.143076923        0.8521154 0.382300885
## 186         -0.002692308        -0.037692308        0.8486538 0.350265487
## 187          0.321153846         0.037307692        0.8484615 0.381769912
## 188          0.543461538         0.196538462        0.8471154 0.439469027
## 189          0.468076923         0.202307692        0.8482692 0.438407080
## 190          0.666153846         0.218076923        0.8463462 0.470265487
## 191          0.819230769         0.086153846        0.8482692 0.502123894
## 192          1.023461538         0.230384615        0.8478846 0.564424779
## 193          1.121923077        -0.083846154        0.8471154 0.507787611
## 194          1.561923077         0.688461538        0.8461538 0.532920354
## 195          1.728076923         0.565384615        0.8498077 0.540000000
## 196          1.207307692         0.377307692        0.8532692 0.480884956
## 197          0.989230769        -0.045769231        0.8548077 0.483716814
## 198          0.796153846         0.071153846        0.8548077 0.481238938
## 199          0.905000000         0.213461538        0.8548077 0.541769912
## 200          0.984615385         0.224615385        0.8540385 0.540353982
## 201          0.821153846         0.043846154        0.8557692 0.573805310
## 202          0.739230769         0.073076923        0.8538462 0.633097345
## 203          0.531923077         0.133846154        0.8548077 0.667787611
## 204          0.252307692         0.262307692        0.8573077 0.671681416
## 205          0.206153846         0.333846154        0.8573077 0.726902655
## 206          0.397692308         0.482692308        0.8584615 0.783185841
## 207          0.182692308         0.296538462        0.8592308 0.848318584
## 208          0.321538462         0.248461538        0.8603846 0.853982301
## 209          0.263076923         0.291538462        0.8600000 0.851150442
## 210          0.218846154         0.243461538        0.8626923 0.880000000
## 211          0.299615385         0.253076923        0.8626923 0.880000000
## 212          0.132692308         0.255769231        0.8673077 0.942654867
```

```
## 213          0.444230769          0.400000000       0.8655769 0.974513274
## 214          0.458076923          0.498076923       0.8673077 0.915398230
## 215          0.227692308          0.483076923       0.8684615 0.972035398
## 216         -0.266153846         -0.108461538       0.8680769 1.000000000
## 217         -1.460769231         -0.555769231       0.8673077 0.941592920
## 218         -0.769615385         -0.670384615       0.8673077 0.939115044
## 219         -1.133846154         -0.738846154       0.8667308 0.967787611
## 220         -0.658846154         -0.360000000       0.8705769 0.910088496
## 221         -0.376923077         -0.311153846       0.8661538 0.842654867
## 222          0.432307692          0.075769231       0.8665385 0.895044248
## 223         -0.040000000         -0.154230769       0.8653846 0.894336283
## 224         -0.014615385         -0.184615385       0.8630769 0.894690265
## 225         -0.452307692         -0.173076923       0.8625000 0.862654867
## 226         -0.411153846         -0.155000000       0.8611538 0.833805310
## 227         -0.163846154         -0.137307692       0.8603846 0.826017699
## 228          0.019230769         -0.046153846       0.8615385 0.770265487
## 229          0.263846154         -0.015384615       0.8621154 0.741946903
## 230          0.590000000          0.015000000       0.8588462 0.744247788
## 231          0.582692308          0.068076923       0.8596154 0.746725664
## 232          0.986153846          0.643076923       0.8578846 0.750619469
## 233          2.353461538          0.838846154       0.8582692 0.718761062
## 234          2.468076923          1.503076923       0.8567308 0.710619469
## 235          1.667307692          0.941923077       0.8550000 0.677168142
## 236          2.157307692          1.117692308       0.8551923 0.677345133
## 237          1.254615385          0.756153846       0.8561538 0.647787611
## 238          1.716923077          0.711538462       0.8530769 0.649203540
## 239          1.169230769          0.471153846       0.8546154 0.644601770
## 240          1.653076923          0.636923077       0.8553846 0.707256637
## 241          1.761153846          0.895769231       0.8550000 0.706548673
## 242          1.458461538          0.741153846       0.8548077 0.679292035
## 243          1.184615385          0.628076923       0.8536538 0.676460177
## 244          1.197692308          0.685000000       0.8548077 0.619115044
## 245          1.051153846          0.586923077       0.8553846 0.589734513
## 246          0.843076923          0.638846154       0.8555769 0.591504425
## 247          0.682307692          0.519230769       0.8567308 0.599646018
## 248          0.606538462          0.491538462       0.8555769 0.604601770
## 249          0.691923077          0.327692308       0.8559615 0.545132743
## 250          0.192692308          0.467692308       0.8573077 0.486194690
## 251          0.602307692          0.688076923       0.8565385 0.460000000
## 252          0.178461538          0.671538462       0.8559615 0.428672566
## 253         -0.016153846          0.791153846       0.8567308 0.430088496
## 254          0.378846154          0.720384615       0.8580769 0.453805310
## 255          0.024615385          0.739230769       0.8565385 0.452389381
## 256          0.262307692          0.968846154       0.8580769 0.450265487
## 257          0.230000000          0.807307692       0.8573077 0.418407080
## 258          0.543076923          0.615000000       0.8559615 0.361415929
```

```
## 259               0.584615385            0.551923077       0.8563462 0.325663717
## 260               0.365769231            0.543076923       0.8571154 0.269734513
## 261              -0.051923077            0.393461538       0.8588462 0.276283186
## 262               0.385384615            0.523076923       0.8576923 0.217876106
## 263               0.254230769            0.510384615       0.8571154 0.245663717
## 264               0.448076923            0.472692308       0.8582692 0.244955752
## 265               0.555000000            0.603461538       0.8578846 0.245132743
## 266               0.505000000            0.483846154       0.8565385 0.212566372
## 267               0.269230769            0.478846154       0.8542308 0.218584071
## 268               0.094615385            0.431923077       0.8546154 0.218230088
## 269               0.178846154            0.537307692       0.8528846 0.277876106
## 270               0.478846154            0.458076923       0.8513462 0.335221239
## 271               0.129615385            0.218846154       0.8509615 0.391504425
## 272              -0.010384615            0.233846154       0.8476923 0.453805310
## 273               0.286538462            0.205384615       0.8492308 0.510442478
## 274              -0.184615385            0.115769231       0.8494231 0.485132743
## 275              -0.105769231            0.131153846       0.8488462 0.545663717
## 276               0.116153846            0.190384615       0.8498077 0.571858407
## 277               0.289230769            0.225000000       0.8519231 0.574159292
## 278               0.125384615            0.176923077       0.8534615 0.632920354
## 279              -0.326153846            0.196538462       0.8534615 0.577699115
## 280              -0.513461538            0.163076923       0.8513462 0.606725664
## 281              -0.881153846            0.051153846       0.8521154 0.610619469
## 282              -0.916153846           -0.009230769       0.8513462 0.636283186
## 283              -0.852692308            0.176153846       0.8505769 0.664247788
## 284              -0.672307692            0.244230769       0.8517308 0.717699115
## 285              -0.798076923            0.175384615       0.8521154 0.751858407
## 286              -0.922307692            0.177692308       0.8525000 0.812035398
## 287              -0.856153846            0.157307692       0.8532692 0.808672566
## 288              -0.933076923            0.143461538       0.8536538 0.869380531
## 289              -0.942692308           -0.074230769       0.8532692 0.870088496
## 290              -0.998076923           -0.008846154       0.8548077 0.815929204
## 291              -1.066538462            0.001538462       0.8538462 0.757168142
## 292              -1.254615385           -0.070384615       0.8525000 0.727787611
## 293              -0.952692308           -0.100769231       0.8544231 0.663362832
## 294              -1.395000000           -0.140000000       0.8548077 0.687433628
## 295              -2.005000000           -0.180769231       0.8569231 0.683362832
## 296              -1.821153846           -0.174615385       0.8569231 0.624070796
## 297              -0.956538462            0.039615385       0.8575000 0.595929204
## 298              -1.139615385            0.027692308       0.8598077 0.534513274
## 299              -1.026153846            0.075384615       0.8573077 0.476106195
## 300              -1.078846154            0.017692308       0.8578846 0.504070796
## 301              -1.404615385           -0.073076923       0.8592308 0.442831858
## 302              -1.391923077           -0.186153846       0.8584615 0.417522124
## 303              -1.693846154           -0.296538462       0.8561538 0.414336283
## 304              -2.129615385           -0.360384615       0.8559615 0.387610619
```

```
## 305          -1.907692308         -0.418461538      0.8540385 0.386902655
## 306          -2.076153846         -0.477307692      0.8544231 0.332743363
## 307          -1.401923077         -0.161153846      0.8555769 0.357522124
## 308          -1.131538462         -0.162692308      0.8571154 0.330442478
## 309          -1.524615385         -0.613461538      0.8567308 0.331858407
## 310          -1.726923077         -0.468461538      0.8575000 0.333274336
## 311          -1.706153846         -0.559615385      0.8563462 0.300884956
## 312          -1.846153846         -0.518461538      0.8563462 0.297699115
## 313          -1.552692308         -0.177307692      0.8548077 0.326548673
## 314          -1.215000000         -0.071923077      0.8555769 0.266548673
## 315          -1.592692308         -0.005000000      0.8559615 0.207610619
## 316          -1.563461538         -0.165000000      0.8538462 0.202300885
## 317          -0.913846154         -0.230000000      0.8532692 0.206194690
## 318          -1.231538462         -0.051153846      0.8540385 0.206902655
## 319          -1.990384615         -0.421923077      0.8544231 0.207256637
## 320          -2.258846154         -0.587307692      0.8540385 0.155221239
## 321          -2.530769231         -0.695000000      0.8532692 0.097876106
## 322          -2.560000000         -0.608461538      0.8546154 0.125663717
## 323          -2.193076923         -0.533076923      0.8542308 0.101592920
## 324          -2.519615385         -0.274615385      0.8546154 0.098938053
## 325          -2.438846154          0.040384615      0.8561538 0.097345133
## 326          -1.728846154         -0.003846154      0.8551923 0.035929204
## 327          -0.862692308          0.238076923      0.8551923 0.033451327
## 328          -0.700769231          0.349615385      0.8532692 0.032743363
## 329          -1.475384615          0.223461538      0.8532692 0.031681416
## 330          -1.265000000          0.293076923      0.8534615 0.002831858
## 331          -1.038846154          0.413846154      0.8542308 0.000000000
## 332          -0.171153846          0.307692308      0.8551923 0.000000000
## 333          -0.233461538          0.211538462      0.8538462 0.030796460
## 334           0.070769231          0.167307692      0.8538462 0.084955752
## 335           0.319230769          0.166923077      0.8557692 0.027610619
## 336          -0.401538462          0.048076923      0.8553846 0.026902655
## 337           0.088076923          0.096538462      0.8555769 0.090796460
## 338           0.582692308          0.159230769      0.8559615 0.097522124
```

```
# univariate model for reference
base_mod_hf <- lm(adidas_PercentChangeMA ~ djia_PercentChangeMA, data=data_hf)
summary(base_mod_hf)
```

```
##
## Call:
## lm(formula = adidas_PercentChangeMA ~ djia_PercentChangeMA, data = data_hf)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.45728 -0.51758  0.01816  0.54109  1.73822
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -0.04695    0.05169  -0.908    0.364
## djia_PercentChangeMA  1.61906    0.13898  11.650   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8016 on 307 degrees of freedom
## Multiple R-squared:  0.3065, Adjusted R-squared:  0.3043
## F-statistic: 135.7 on 1 and 307 DF,  p-value: < 2.2e-16
```

```
# check model with all predictors
hf_full <- lm(adidas_PercentChangeMA ~ ., data=data_hf)
summary(hf_full)
```

```
##
## Call:
## lm(formula = adidas_PercentChangeMA ~ ., data = data_hf)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.13579 -0.49079  0.01898  0.49054  1.71442
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)            9.4990     9.1093   1.043 0.297876
## djia_PercentChangeMA   1.6702     0.1373  12.161  < 2e-16 ***
## final_score_MA       -11.6114    10.6787  -1.087 0.277744
## pn_score_MA            0.7361     0.1903   3.869 0.000134 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7852 on 305 degrees of freedom
## Multiple R-squared:  0.339,  Adjusted R-squared:  0.3325
## F-statistic: 52.14 on 3 and 305 DF,  p-value: < 2.2e-16
```

```
# delete insignificant predictor
hf_mod <- lm(adidas_PercentChangeMA ~ djia_PercentChangeMA + pn_score_MA, data=
data_hf)
summary(hf_mod)
```

```
##
## Call:
## lm(formula = adidas_PercentChangeMA ~ djia_PercentChangeMA +
##     pn_score_MA, data = data_hf)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.16677 -0.49365 -0.01292  0.50323  1.76221
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -0.4052     0.1090  -3.718 0.000239 ***
## djia_PercentChangeMA   1.6531     0.1365  12.112  < 2e-16 ***
## pn_score_MA            0.6818     0.1837   3.712 0.000244 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7855 on 306 degrees of freedom
## Multiple R-squared:  0.3364, Adjusted R-squared:  0.3321
## F-statistic: 77.57 on 2 and 306 DF,  p-value: < 2.2e-16
```

NLTK VADER sentiment model

```
# load nltk scores
d = read.csv('/Users/landise/Downloads/lm.csv')

# remove date column
data <- subset(d, select = c(adidas_PercentChangeMA, djia_PercentChangeMA, nltk
_positive_MA, nltk_negative_MA, nltk_neutral_MA, nltk_compound_MA))
head(data)
```

```
##   adidas_PercentChangeMA djia_PercentChangeMA nltk_positive_MA nltk_negative
_MA
## 1             0.05230769            0.2015385        0.1204407       0.03930
243
## 2             0.36538462            0.1503846        0.1163253       0.03980
243
## 3             0.13038462            0.1846154        0.1172099       0.03837
935
## 4             0.27192308            0.3538462        0.1185561       0.03811
012
## 5             0.55192308            0.2976923        0.1170561       0.03838
704
## 6             0.70923077            0.1519231        0.1162869       0.03854
089
##   nltk_neutral_MA nltk_compound_MA
## 1       0.8402207        0.7658166
## 2       0.8438361        0.7643704
## 3       0.8443745        0.8117012
## 4       0.8432592        0.8127974
## 5       0.8444899        0.8097512
## 6       0.8451438        0.8099897
```

```
# same base model as before
base_mod <- lm(adidas_PercentChangeMA ~ djia_PercentChangeMA, data=data)
summary(base_mod)
```

```
##
## Call:
## lm(formula = adidas_PercentChangeMA ~ djia_PercentChangeMA, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.45728 -0.51758  0.01816  0.54109  1.73822
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -0.04695    0.05169  -0.908    0.364
## djia_PercentChangeMA  1.61906    0.13898  11.650   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8016 on 307 degrees of freedom
## Multiple R-squared:  0.3065, Adjusted R-squared:  0.3043
## F-statistic: 135.7 on 1 and 307 DF,  p-value: < 2.2e-16
```

```
# check full model
nltk_full = lm(adidas_PercentChangeMA ~ ., data=data)
summary(nltk_full)
```

```
##
## Call:
## lm(formula = adidas_PercentChangeMA ~ ., data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.15391 -0.53721  0.02203  0.40926  1.71116
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1413.8086  1182.1883   1.196    0.233
## djia_PercentChangeMA     1.8305     0.1376  13.299  < 2e-16 ***
## nltk_positive_MA     -1385.2105  1184.7342  -1.169    0.243
## nltk_negative_MA     -1441.3410  1183.3975  -1.218    0.224
## nltk_neutral_MA      -1412.5541  1182.0382  -1.195    0.233
## nltk_compound_MA        -4.1132     0.6926  -5.939 7.85e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7148 on 303 degrees of freedom
## Multiple R-squared:  0.4558, Adjusted R-squared:  0.4468
## F-statistic: 50.76 on 5 and 303 DF,  p-value: < 2.2e-16
```

```
# delete most insignificant pred
nltk_less1 = lm(adidas_PercentChangeMA ~ djia_PercentChangeMA + nltk_negative_M
A + nltk_neutral_MA + nltk_compound_MA, data=data)
summary(nltk_less1)
```

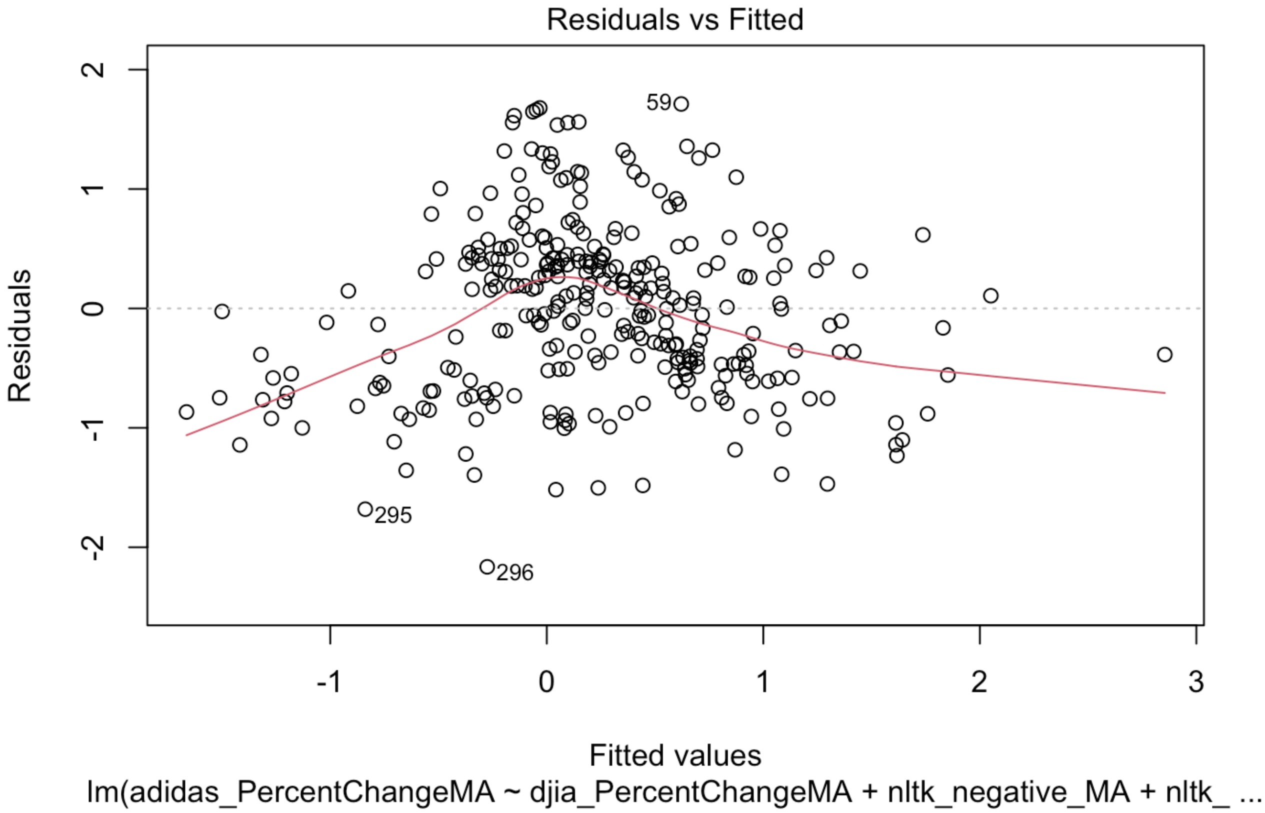
```
##
## Call:
## lm(formula = adidas_PercentChangeMA ~ djia_PercentChangeMA +
##     nltk_negative_MA + nltk_neutral_MA + nltk_compound_MA, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.16372 -0.51496  0.01823  0.41608  1.71222
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           31.5859     4.7350   6.671 1.20e-10 ***
## djia_PercentChangeMA   1.7929     0.1339  13.390  < 2e-16 ***
## nltk_negative_MA     -57.7529    10.9951  -5.253 2.82e-07 ***
## nltk_neutral_MA      -30.5085     5.0806  -6.005 5.45e-09 ***
## nltk_compound_MA      -4.4275     0.6386  -6.933 2.48e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7153 on 304 degrees of freedom
## Multiple R-squared:  0.4533, Adjusted R-squared:  0.4462
## F-statistic: 63.03 on 4 and 304 DF,  p-value: < 2.2e-16
```

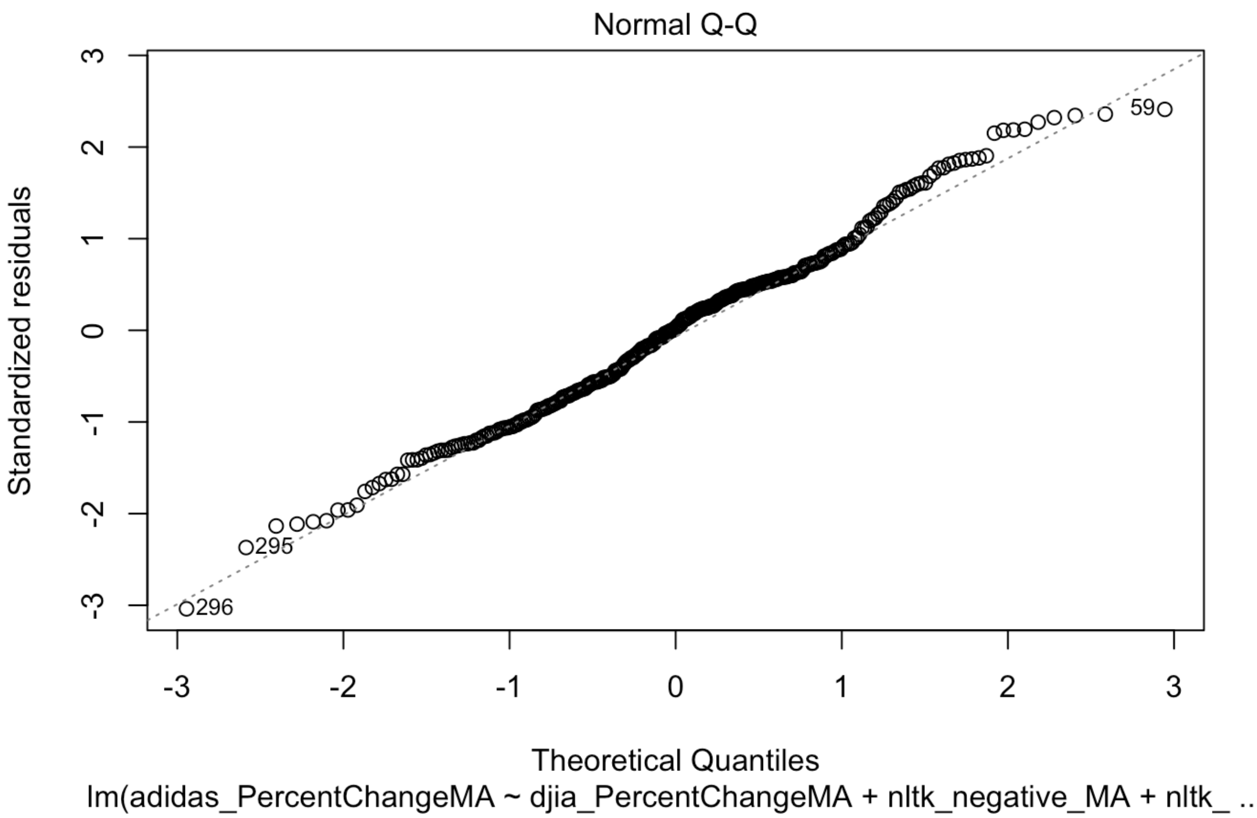
All predictors are significant.

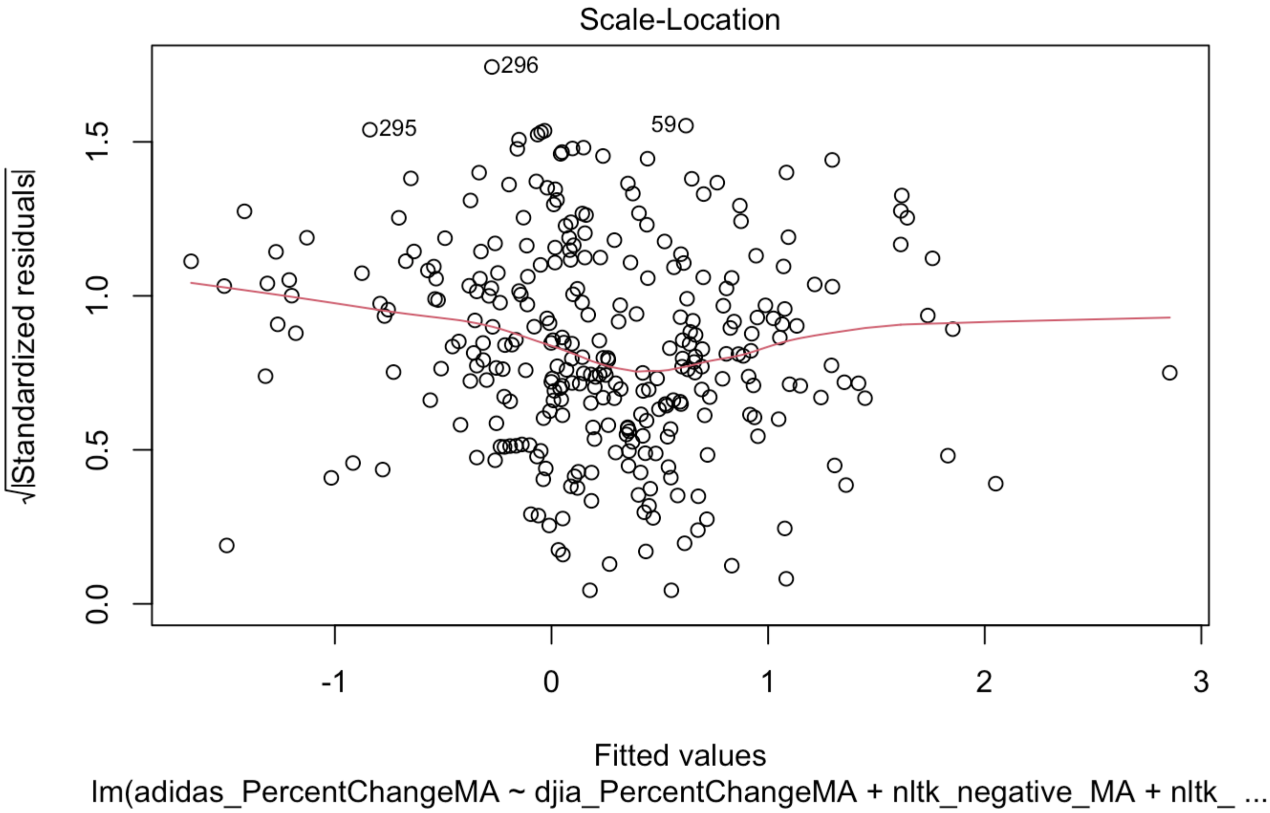
Diagnostics

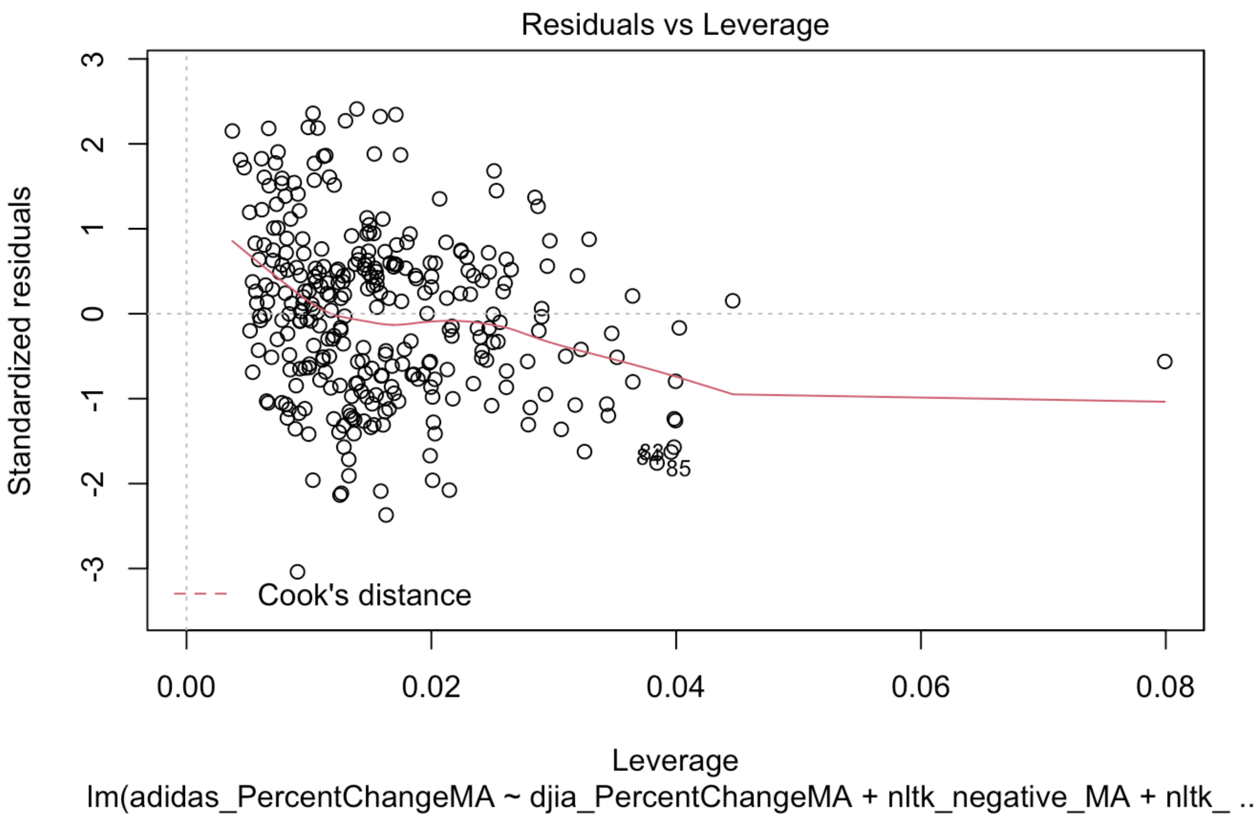
Plots

```
plot(nltk_less1)
```

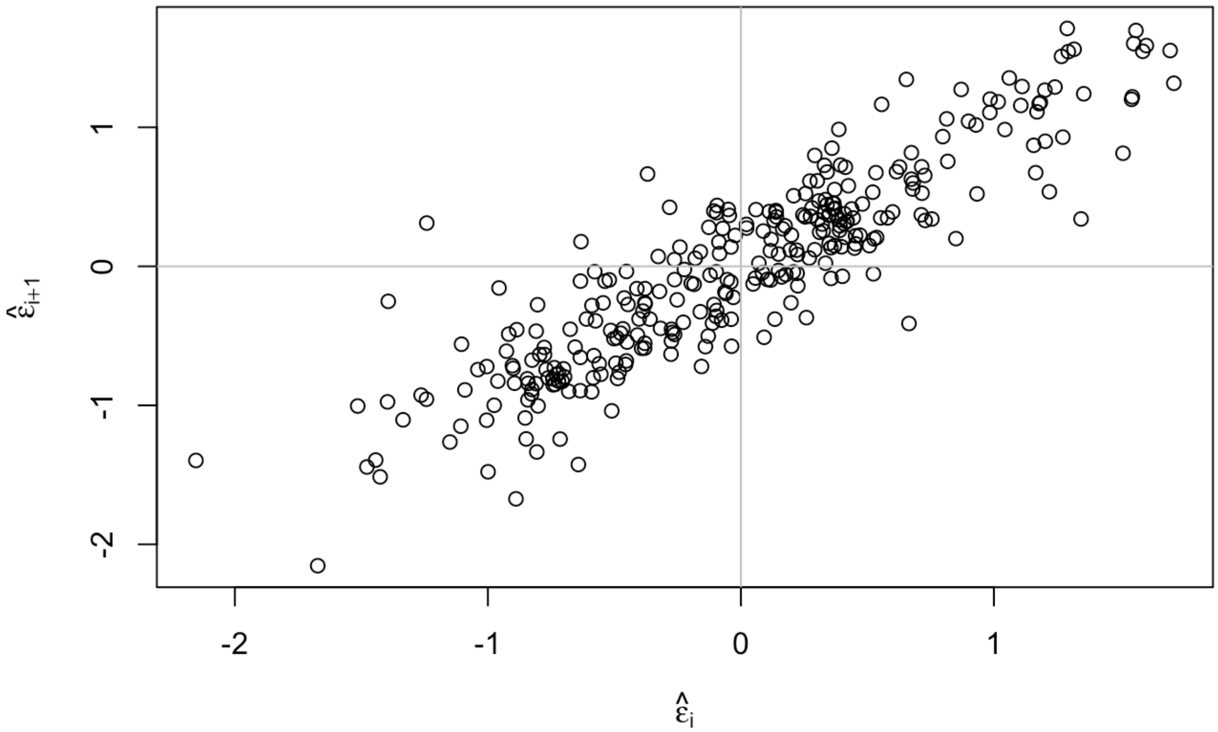

Residuals vs Fitted
Residuals
Fitted values
lm(adidas_PercentChangeMA ~ djia_PercentChangeMA + nltk_negative_MA + nltk_ ...
59
295
296
2
1
0
-1
-2
-1
0
1
2
3


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(adidas_PercentChangeMA ~ djia_PercentChangeMA + nltk_negative_MA + nltk_ ...
59
295
296
-3
-2
-1
0
1
2
3


Scale-Location
296
295
59
√|Standardized residuals|
0.0
0.5
1.0
1.5
-1
0
1
2
3
Fitted values
lm(adidas_PercentChangeMA ~ djia_PercentChangeMA + nltk_negative_MA + nltk_ ...




```
# lag plot to check for autocorrelation
n <- length(residuals(nltk_full))
plot(tail(residuals(nltk_full), n-1) ~ head(residuals(nltk_full), n-1), xlab=ex
pression(hat(epsilon)[i]), ylab=expression(hat(epsilon)[i+1]))
abline(h=0, v=0, col=grey(0.75))
```



There is strong evidence of autocorrelation - not uncommon in time series.

Transform the predictors

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.1.2
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.1.2
```

```
durbinWatsonTest(nltk_less1)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.8975952     0.2006623       0
##  Alternative hypothesis: rho != 0
```

p-value indiates strong evidence of autocorrelation.

```
# another package for DW test
library(orcutt)
```

```
## Loading required package: lmtest
```

```
## Warning: package 'lmtest' was built under R version 4.1.2
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 4.1.2
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
dwtest(nltk_less1)
```

```
##
##  Durbin-Watson test
##
## data:  nltk_less1
## DW = 0.20066, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

Source: https://rpubs.com/apricitea/handling-autocorrelation (https://rpubs.com/apricitea/handling-autocorrelation)

NLTK final model transformation

```
# get best RHO for transformation
mod_transformed <- cochrane.orcutt(nltk_less1, convergence=5, max.iter=1000)
mod_transformed
```

```
## Cochrane-orcutt estimation for first order autocorrelation
##
## Call:
## lm(formula = adidas_PercentChangeMA ~ djia_PercentChangeMA +
##     nltk_negative_MA + nltk_neutral_MA + nltk_compound_MA, data = data)
##
##  number of interaction: 4
##  rho 0.936129
##
## Durbin-Watson statistic
## (original):    0.20066 , p-value: 1.451e-58
## (transformed): 2.27455 , p-value: 9.914e-01
##
##  coefficients:
##          (Intercept) djia_PercentChangeMA     nltk_negative_MA
##            10.135928             0.926039             4.095486
##      nltk_neutral_MA     nltk_compound_MA
##           -11.207285            -1.045186
```

```
# transform predictors and target
rho <- mod_transformed$rho
y.trans <- data$adidas_PercentChangeMA[-1]-data$adidas_PercentChangeMA[-309]*rh
o
x.trans <- (data$djia_PercentChangeMA)[-1]-(data$djia_PercentChangeMA)[-309]*rh
o
x1.trans <- (data$nltk_negative_MA)[-1]-(data$nltk_negative_MA)[-309]*rho
x2.trans <- (data$nltk_neutral_MA)[-1]-(data$nltk_neutral_MA)[-309]*rho
x3.trans <- (data$nltk_compound_MA)[-1]-(data$nltk_compound_MA)[-309]*rho

# model with transformed preds
mod_transformed <- lm(y.trans ~ x.trans + x1.trans + x2.trans + x3.trans)
summary(mod_transformed)
```

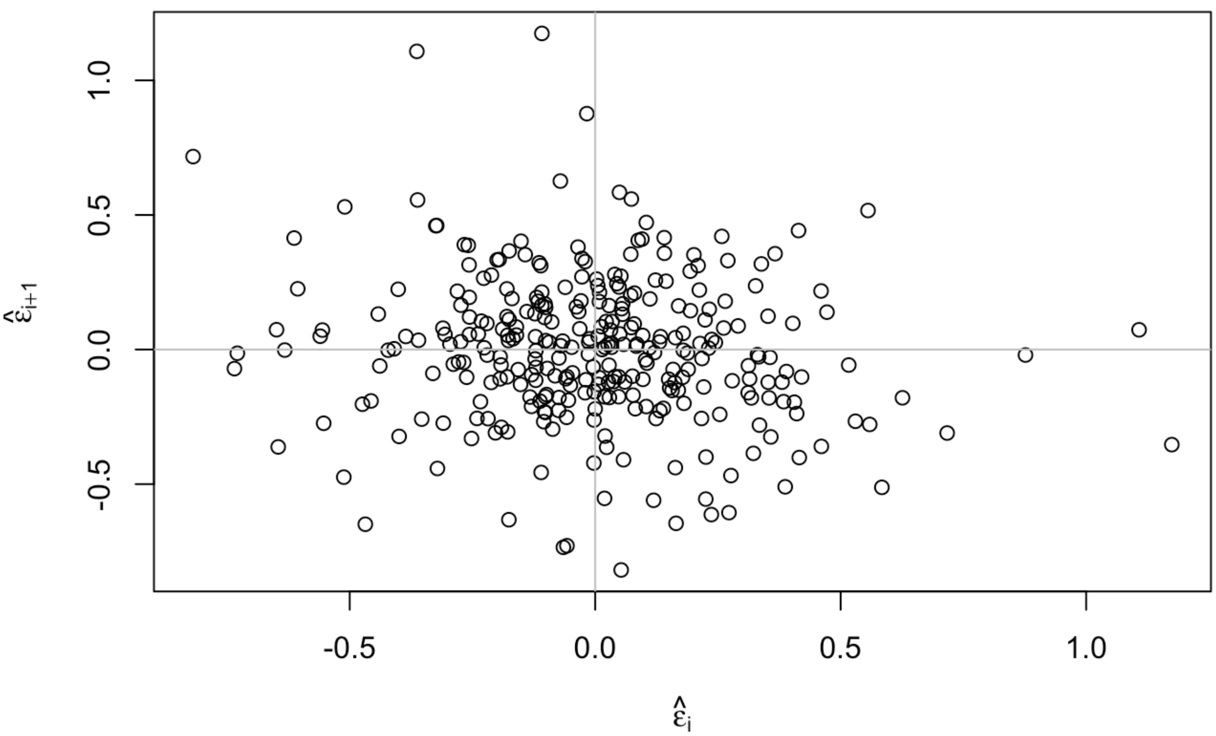
```
##
## Call:
## lm(formula = y.trans ~ x.trans + x1.trans + x2.trans + x3.trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81899 -0.17101  0.00396  0.16234  1.17445
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6474     0.5988   1.081    0.280
## x.trans       0.9260     0.1061   8.728   <2e-16 ***
## x1.trans      4.0955    19.7523   0.207    0.836
## x2.trans    -11.2073    10.0563  -1.114    0.266
## x3.trans     -1.0452     0.9860  -1.060    0.290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2789 on 303 degrees of freedom
## Multiple R-squared:  0.2094, Adjusted R-squared:  0.199
## F-statistic: 20.06 on 4 and 303 DF,  p-value: 1.135e-14
```

```
dwtest(mod_transformed)
```

```
##
##  Durbin-Watson test
##
## data:  mod_transformed
## DW = 2.2746, p-value = 0.9914
## alternative hypothesis: true autocorrelation is greater than 0
```

p-value near 1 indicates we have fixed the issue of autocorrelation

```
n <- length(residuals(mod_transformed))
plot(tail(residuals(mod_transformed), n-1) ~ head(residuals(mod_transformed), n
-1), xlab=expression(hat(epsilon)[i]), ylab=expression(hat(epsilon)[i+1]))
abline(h=0, v=0, col=grey(0.75))
```



Plot shows no more evidence of autocorrelation.

Select final vars

```
summary(mod_transformed)
```

```
##
## Call:
## lm(formula = y.trans ~ x.trans + x1.trans + x2.trans + x3.trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81899 -0.17101  0.00396  0.16234  1.17445
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.6474     0.5988   1.081    0.280
## x.trans       0.9260     0.1061   8.728   <2e-16 ***
## x1.trans      4.0955    19.7523   0.207    0.836
## x2.trans    -11.2073    10.0563  -1.114    0.266
## x3.trans     -1.0452     0.9860  -1.060    0.290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2789 on 303 degrees of freedom
## Multiple R-squared:  0.2094, Adjusted R-squared:  0.199
## F-statistic: 20.06 on 4 and 303 DF,  p-value: 1.135e-14
```

```
# y.trans <- data$adidas_PercentChangeMA[-1]-data$adidas_PercentChangeMA[-309]*
rho
# x.trans <- (data$djia_PercentChangeMA)[-1]-(data$djia_PercentChangeMA)[-309]*
rho
# x1.trans <- (data$nltk_negative_MA)[-1]-(data$nltk_negative_MA)[-309]*rho
# x2.trans <- (data$nltk_neutral_MA)[-1]-(data$nltk_neutral_MA)[-309]*rho
# x3.trans <- (data$nltk_compound_MA)[-1]-(data$nltk_compound_MA)[-309]*rho

mod_transformed <- lm(y.trans ~ x.trans + x2.trans + x3.trans)
summary(mod_transformed)
```

```
##
## Call:
## lm(formula = y.trans ~ x.trans + x2.trans + x3.trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81750 -0.17159  0.00408  0.16131  1.17448
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.7316     0.4395   1.665   0.0970 .
## x.trans       0.9256     0.1059   8.739   <2e-16 ***
## x2.trans    -12.4144     8.1867  -1.516   0.1305
## x3.trans     -1.1947     0.6713  -1.780   0.0761 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2785 on 304 degrees of freedom
## Multiple R-squared:  0.2093, Adjusted R-squared:  0.2015
## F-statistic: 26.82 on 3 and 304 DF,  p-value: 2.035e-15
```

```
mod_transformed <- lm(y.trans ~ x.trans + x3.trans)
summary(mod_transformed)
```

```
##
## Call:
## lm(formula = y.trans ~ x.trans + x3.trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.79132 -0.15269 -0.00324  0.17338  1.18297
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.06741    0.03636   1.854   0.0647 .
## x.trans      0.91428    0.10587   8.636 3.31e-16 ***
## x3.trans    -1.24891    0.67178  -1.859   0.0640 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2791 on 305 degrees of freedom
## Multiple R-squared:  0.2033, Adjusted R-squared:  0.1981
## F-statistic: 38.92 on 2 and 305 DF,  p-value: 8.842e-16
```

Try full model transformation

```
full_transformed <- cochrane.orcutt(nltk_full, convergence=5, max.iter=1000)
full_transformed
```

```
## Cochrane-orcutt estimation for first order autocorrelation
##
## Call:
## lm(formula = adidas_PercentChangeMA ~ ., data = data)
##
##  number of interaction: 4
##  rho 0.937493
##
## Durbin-Watson statistic
## (original):    0.20810 , p-value: 1.767e-58
## (transformed): 2.29765 , p-value: 9.952e-01
##
##  coefficients:
##          (Intercept) djia_PercentChangeMA     nltk_positive_MA
##         -1464.511470             0.925967          1477.609626
##     nltk_negative_MA      nltk_neutral_MA     nltk_compound_MA
##          1481.532644          1462.961715            -1.124369
```

```
rho <- full_transformed$rho
y.trans <- data$adidas_PercentChangeMA[-1]-data$adidas_PercentChangeMA[-309]*rh
o
x.trans <- (data$djia_PercentChangeMA)[-1]-(data$djia_PercentChangeMA)[-309]*rh
o
x1.trans <- (data$nltk_negative_MA)[-1]-(data$nltk_negative_MA)[-309]*rho
x2.trans <- (data$nltk_neutral_MA)[-1]-(data$nltk_neutral_MA)[-309]*rho
x3.trans <- (data$nltk_compound_MA)[-1]-(data$nltk_compound_MA)[-309]*rho
x4.trans <- (data$nltk_positive_MA)[-1]-(data$nltk_positive_MA)[-309]*rho

full_transformed <- lm(y.trans ~ x.trans + x1.trans + x2.trans + x3.trans + x4.
trans)
summary(full_transformed)
```

```
##
## Call:
## lm(formula = y.trans ~ x.trans + x1.trans + x2.trans + x3.trans +
##     x4.trans)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8089 -0.1657  0.0085  0.1555  1.1146
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -91.5424    56.0868  -1.632    0.104
## x.trans        0.9260     0.1058   8.754   <2e-16 ***
## x1.trans    1481.5326   899.0916   1.648    0.100
## x2.trans    1462.9617   897.0064   1.631    0.104
## x3.trans      -1.1244     0.9854  -1.141    0.255
## x4.trans    1477.6096   899.0695   1.643    0.101
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2781 on 302 degrees of freedom
## Multiple R-squared:  0.2162, Adjusted R-squared:  0.2032
## F-statistic: 16.66 on 5 and 302 DF,  p-value: 1.557e-14
```

```
full_transformed <- lm(y.trans ~ x.trans + x1.trans + x2.trans + x4.trans)
summary(full_transformed)
```

```
##
## Call:
## lm(formula = y.trans ~ x.trans + x1.trans + x2.trans + x4.trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82112 -0.15737  0.00793  0.16000  1.12659
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -88.3953    56.0468  -1.577    0.116
## x.trans        0.9277     0.1058   8.767   <2e-16 ***
## x1.trans    1442.0998   898.8745   1.604    0.110
## x2.trans    1411.7380   896.3283   1.575    0.116
## x4.trans    1421.7429   898.1823   1.583    0.114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2783 on 303 degrees of freedom
## Multiple R-squared:  0.2128, Adjusted R-squared:  0.2024
## F-statistic: 20.48 on 4 and 303 DF,  p-value: 5.974e-15
```

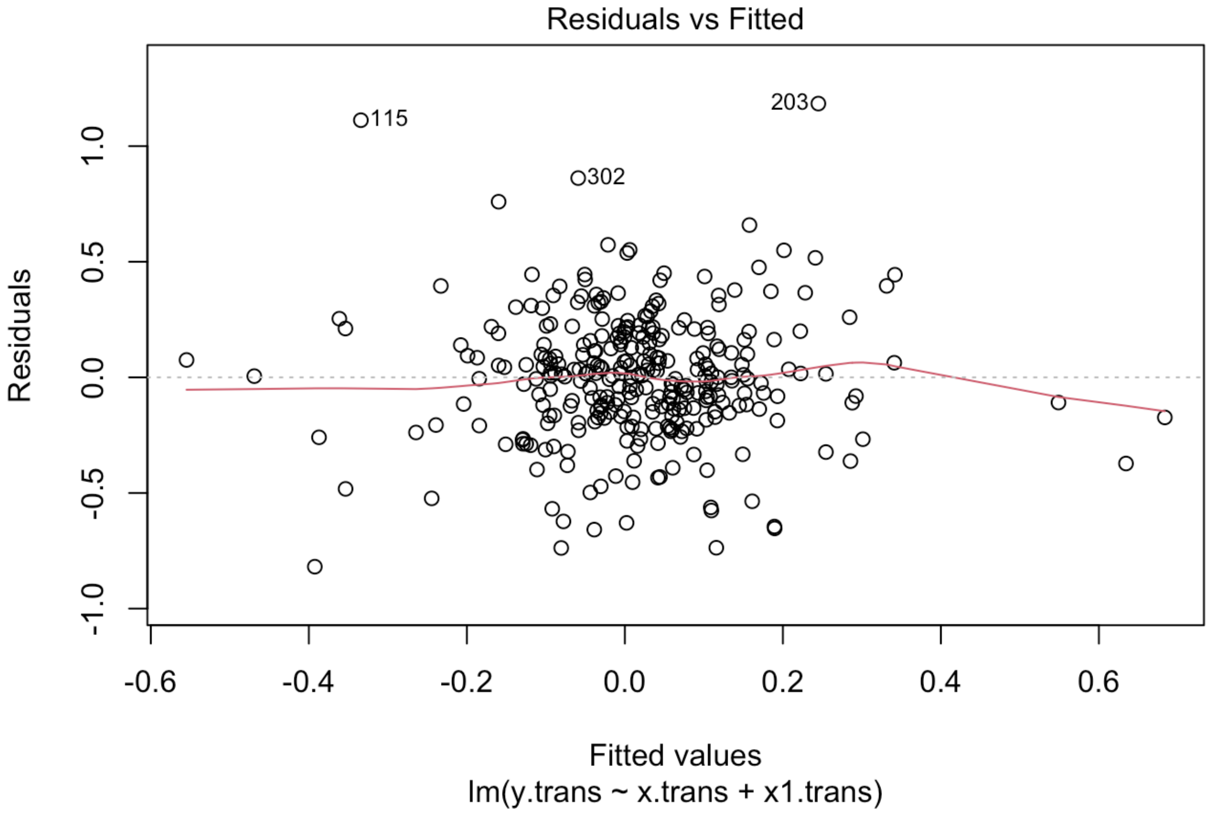
```
full_transformed <- lm(y.trans ~ x.trans + x1.trans + x4.trans)
summary(full_transformed)
```

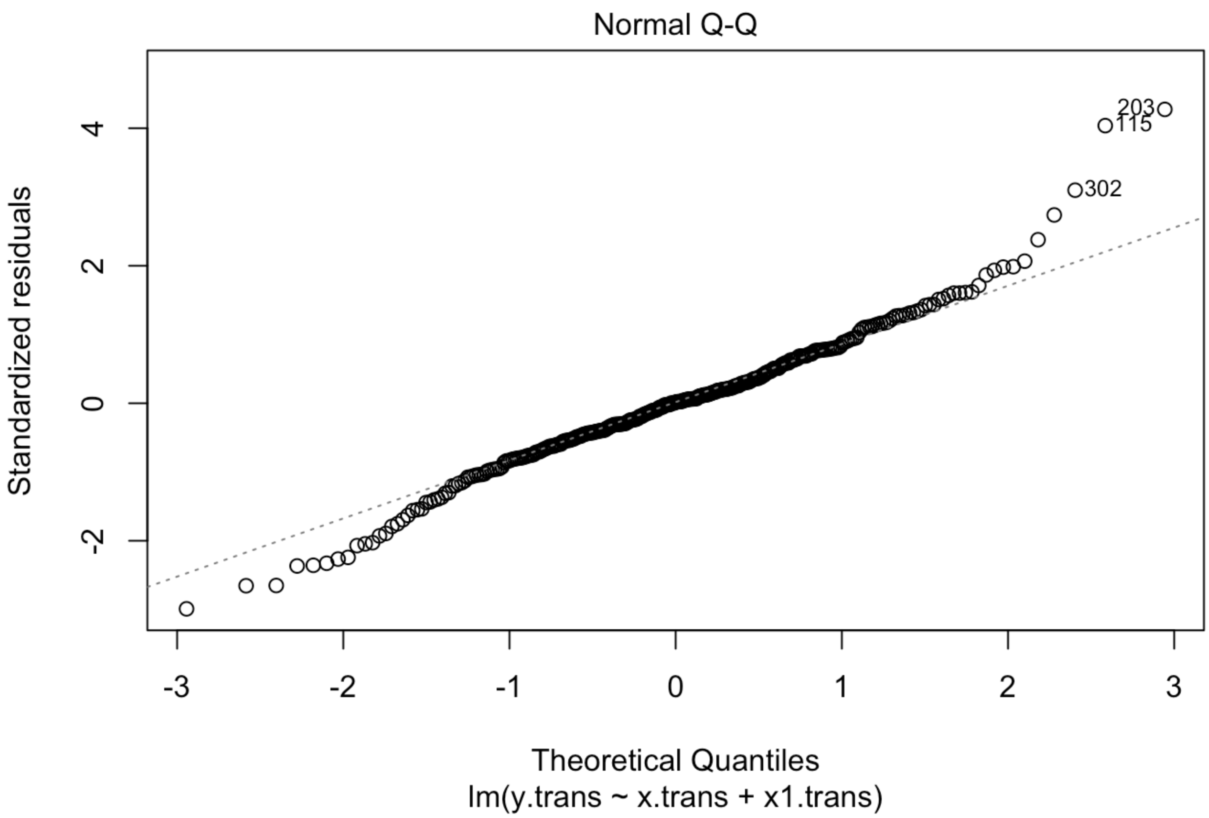
```
##
## Call:
## lm(formula = y.trans ~ x.trans + x1.trans + x4.trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.83038 -0.16090  0.00429  0.16325  1.18299
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.12031    0.08338  -1.443   0.1501
## x.trans      0.92713    0.10607   8.741   <2e-16 ***
## x1.trans    26.48435   12.34193   2.146   0.0327 *
## x4.trans     7.16004    9.27983   0.772   0.4410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.279 on 304 degrees of freedom
## Multiple R-squared:  0.2064, Adjusted R-squared:  0.1986
## F-statistic: 26.35 on 3 and 304 DF,  p-value: 3.532e-15
```

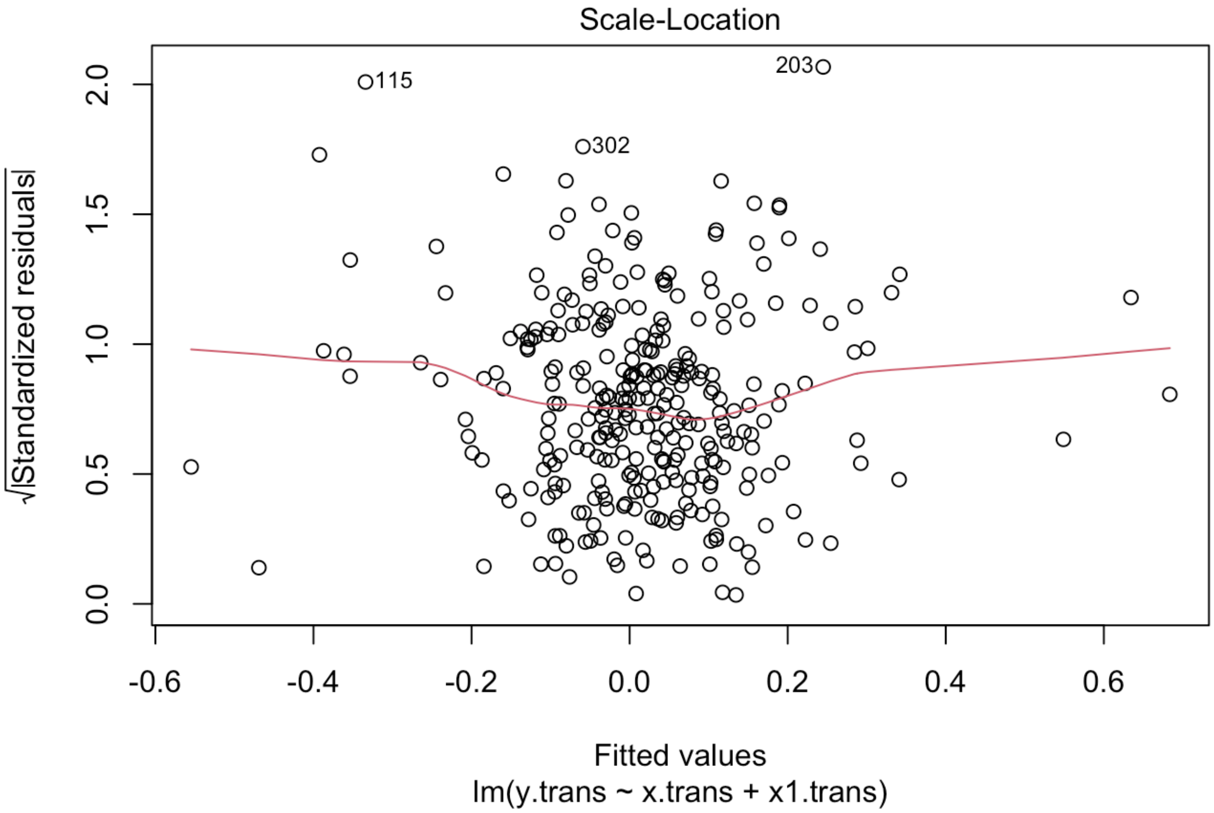
```
full_transformed <- lm(y.trans ~ x.trans + x1.trans)
summary(full_transformed)
```

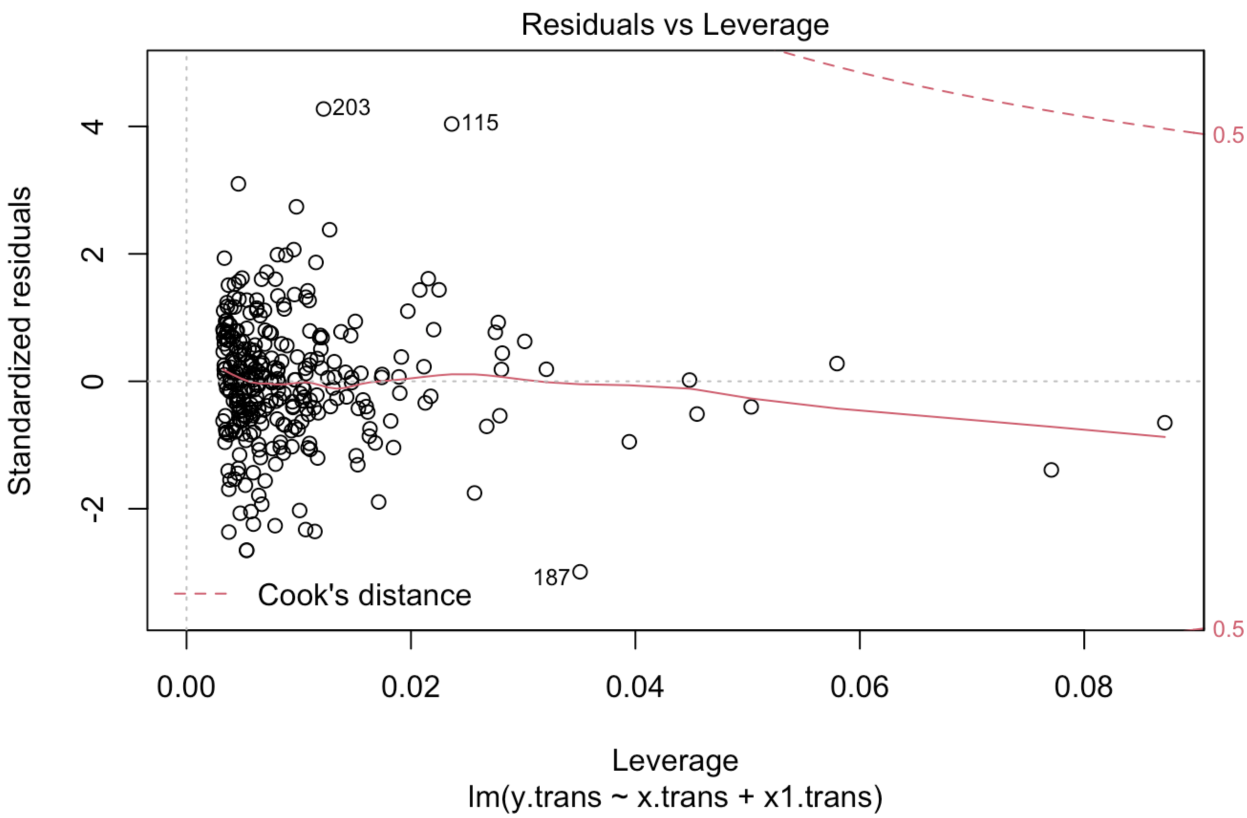
```
##
## Call:
## lm(formula = y.trans ~ x.trans + x1.trans)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81880 -0.15390  0.00414  0.16331  1.18406
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.06305    0.03797  -1.661   0.0978 .
## x.trans      0.92271    0.10584   8.718   <2e-16 ***
## x1.trans    24.15942   11.96050   2.020   0.0443 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2788 on 305 degrees of freedom
## Multiple R-squared:  0.2048, Adjusted R-squared:  0.1996
## F-statistic: 39.28 on 2 and 305 DF,  p-value: 6.61e-16
```

```
plot(full_transformed)
```

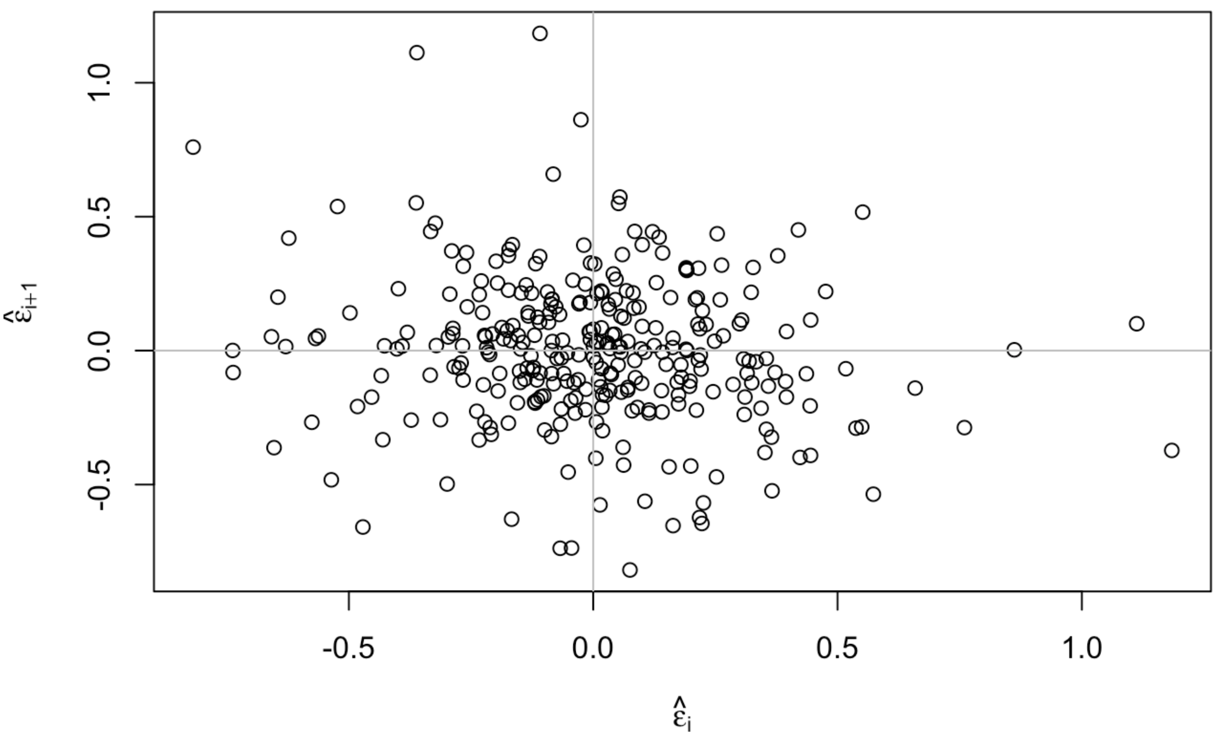



Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(y.trans ~ x.trans + x1.trans)
203
115
302
-3
-2
-1
0
1
2
3
4


Scale-Location
√|Standardized residuals|
Fitted values
lm(y.trans ~ x.trans + x1.trans)
115
203
302
2.0
1.5
1.0
0.5
0.0
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6




```
n <- length(residuals(full_transformed))
plot(tail(residuals(full_transformed), n-1) ~ head(residuals(full_transformed),
n-1), xlab=expression(hat(epsilon)[i]), ylab=expression(hat(epsilon)[i+1]))
abline(h=0, v=0, col=grey(0.75))
```




```
base_transformed <- lm(y.trans ~ x.trans)
summary(base_transformed)
```

```
##
## Call:
## lm(formula = y.trans ~ x.trans)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8035 -0.1684  0.0061  0.1603  1.2071
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.00657    0.01600   0.411    0.682
## x.trans      0.91243    0.10625   8.587  4.6e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2802 on 306 degrees of freedom
## Multiple R-squared:  0.1942, Adjusted R-squared:  0.1916
## F-statistic: 73.74 on 1 and 306 DF,  p-value: 4.596e-16
```

Both models show simliar results.