CheapMass

Alain Duplan & Ariel Lutati

set up

```
## [1] "Year" "2019" "2018" "2017" "2016" "2015" "Nets"
```

linear model

note that all together, a proper linear could not be created using all the variables

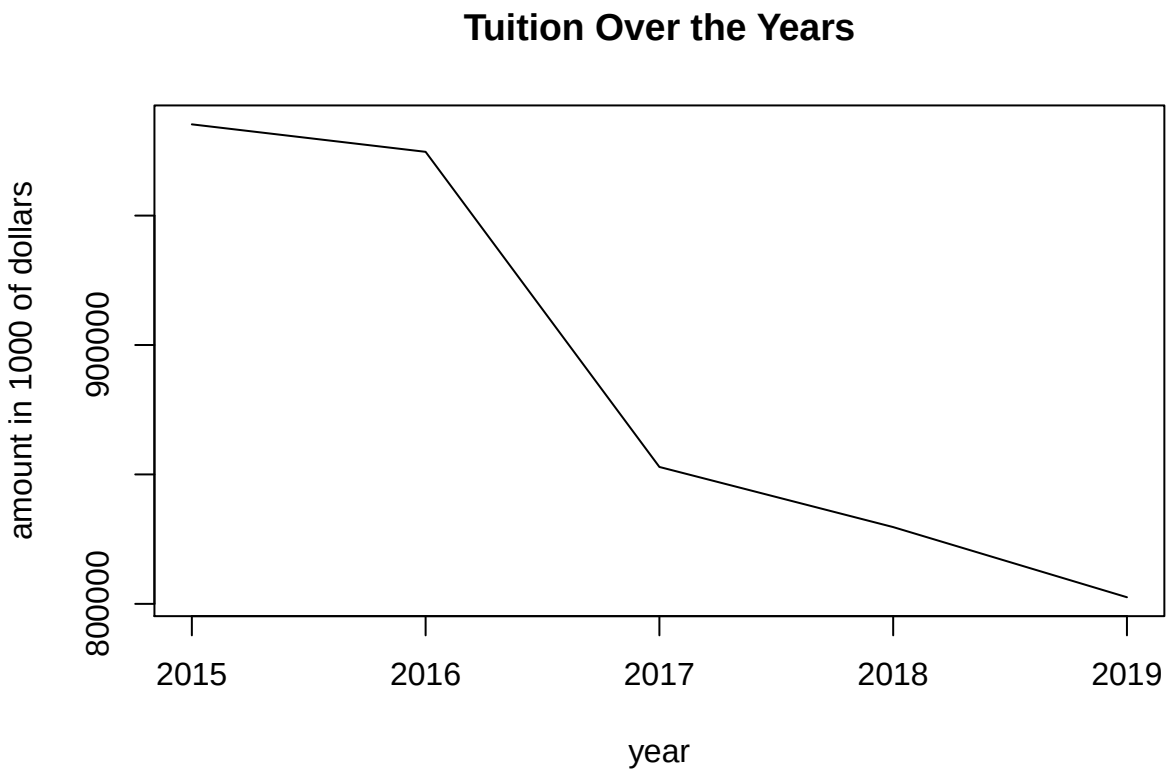
V3, V5, significant under 5% interval and V11 is significant under 1% interval. So we will use these three models of linear regresssion to try to predict tuittion and fees

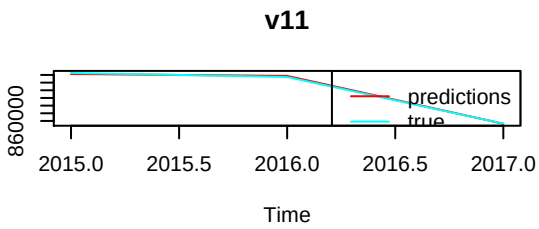
V11 - sumplimental disclosures of non cash activities V5 - net decrease in cash and noncash equivalents V3 - cash flows from capital and other financing activities

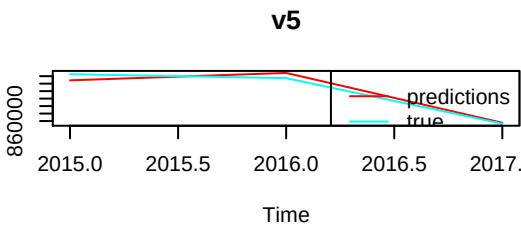
We also found that cash from non financing activities is important to umass since that is what they are using for refunds this semester

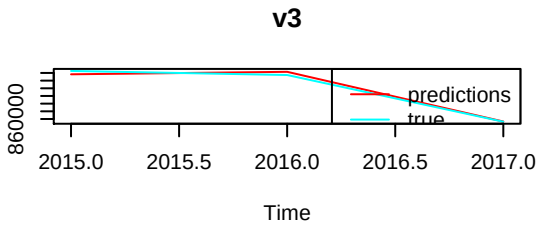
V7 and V2 are the only significant variables in this catergory so we will also use it V2 - grants and contracts v7 - student orginzation agency transactions

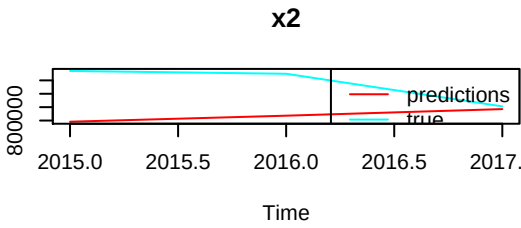
plot tuitiin and fees on its , if you want we can plot the other variables but that would be a lot of graphs

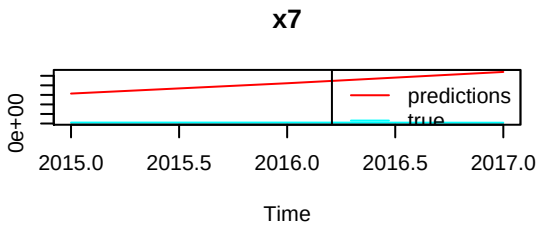

Tuition Over the Years
amount in 1000 of dollars
800000
900000
2015
2016
2017
2018
2019
year









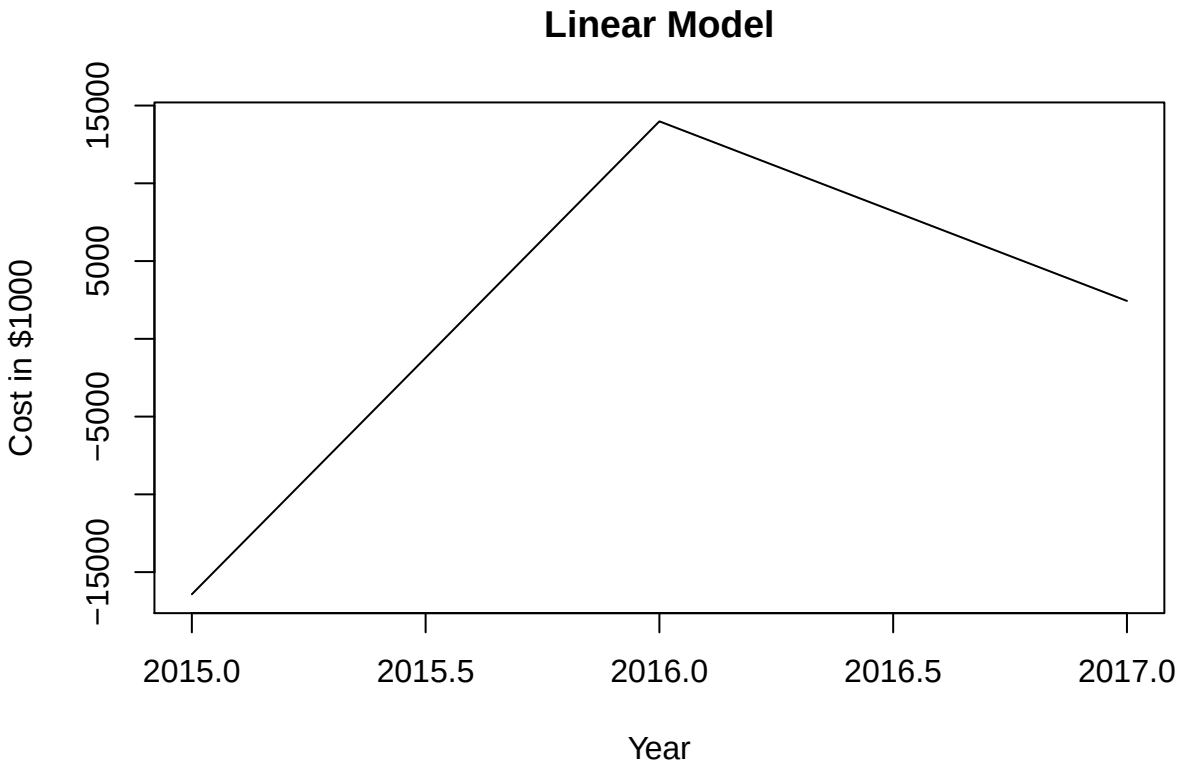


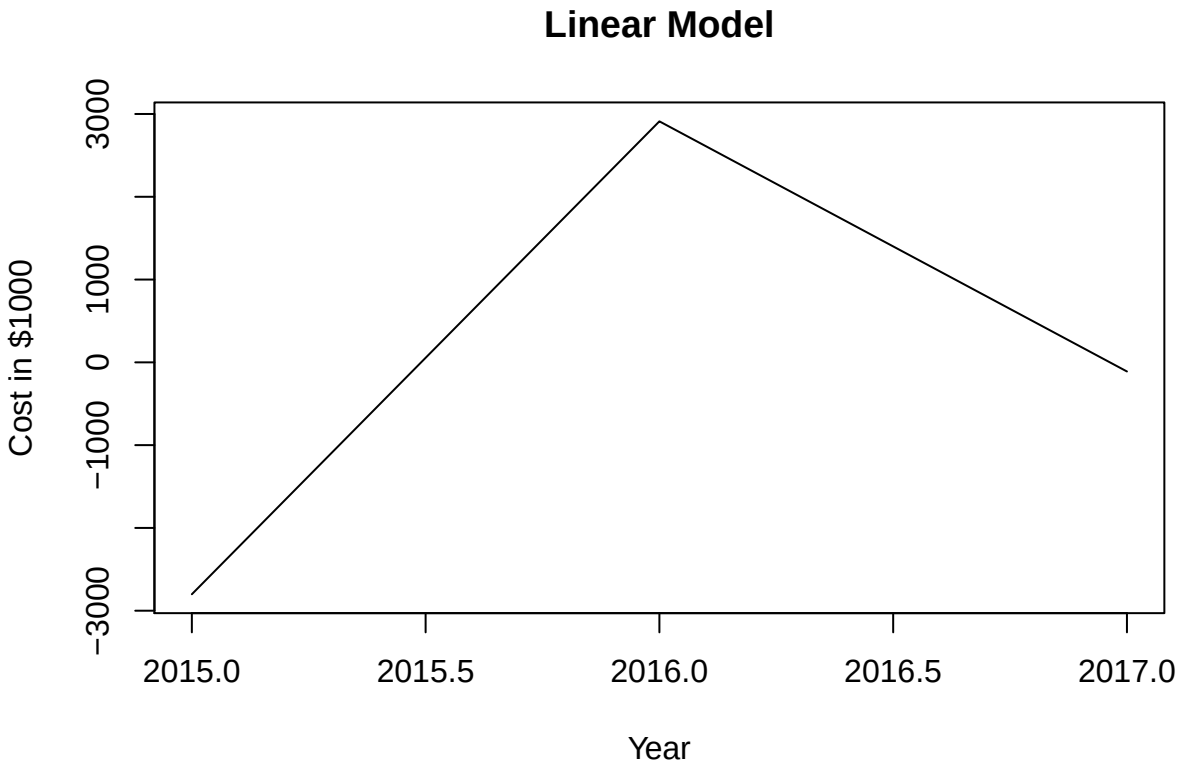
make our predictions
we have to add some labels lol but we can work on it after

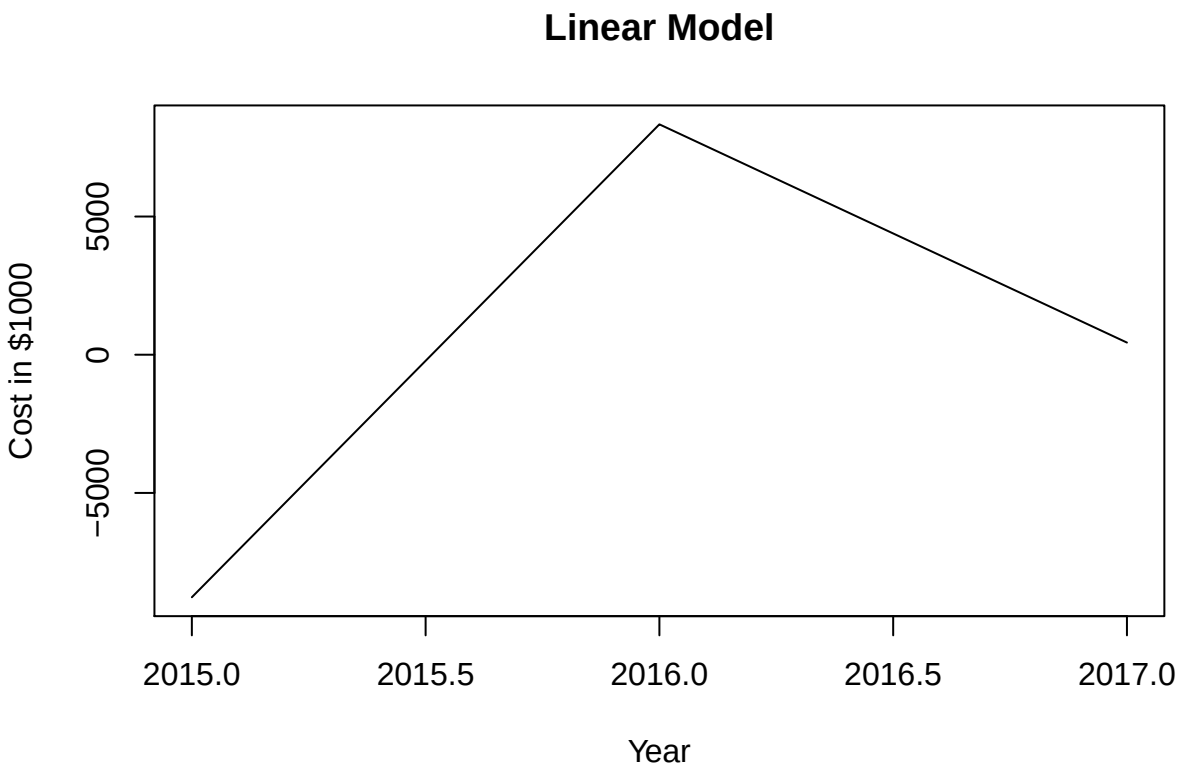
v11 is by far the closest predictors, follows by v5, v3 is meh and everything else goes to shit from there

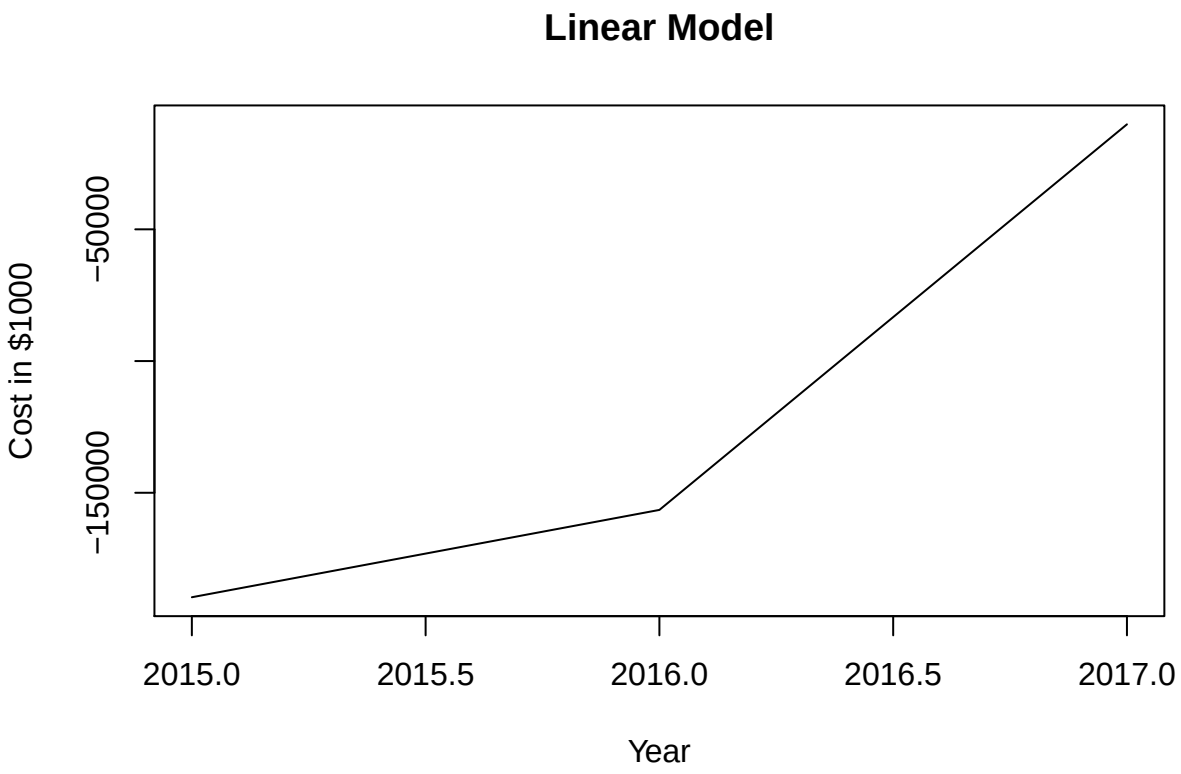
we will now plot the residual time series see how much we are off by

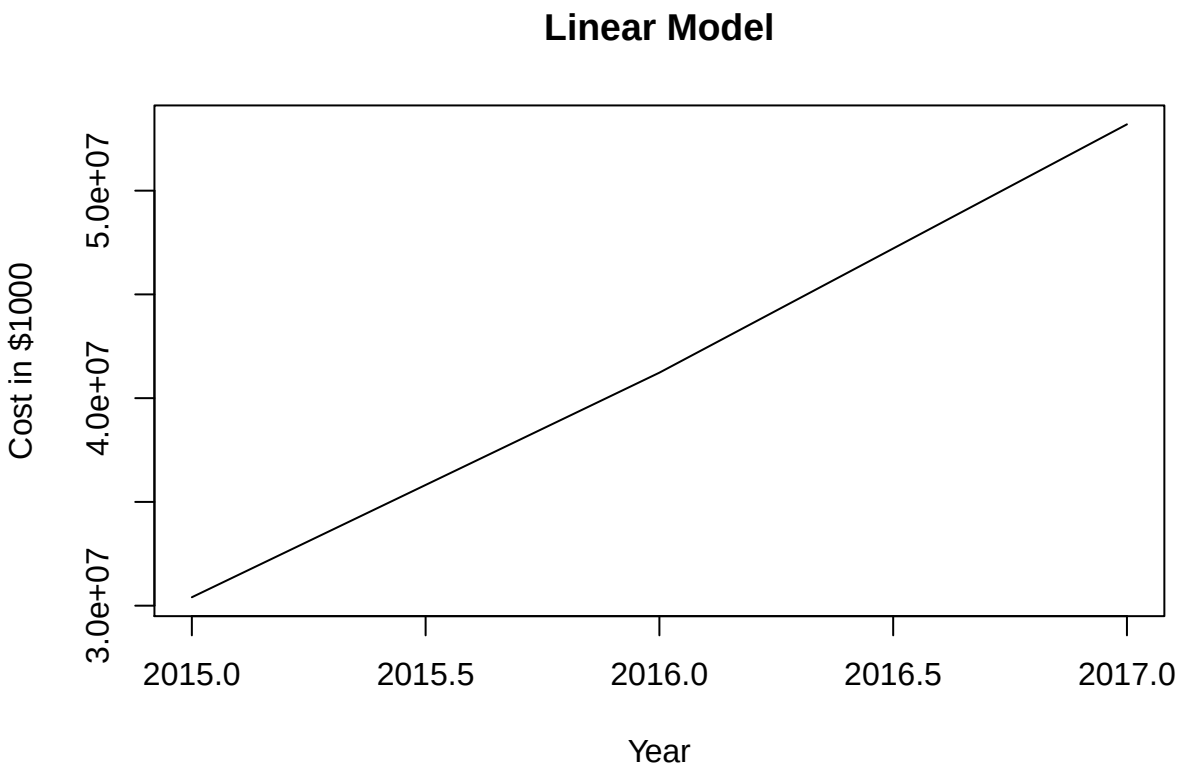
```
## Warning in par(mfro = c(3, 2)): "mfro" is not a graphical parameter
```


Linear Model
Cost in $1000
15000
5000
−5000
−15000
2015.0
2015.5
2016.0
2016.5
2017.0
Year


Linear Model
Cost in $1000
−3000
−1000
0
1000
3000
2015.0
2015.5
2016.0
2016.5
2017.0
Year


Linear Model
Cost in $1000
−5000
0
5000
2015.0
2015.5
2016.0
2016.5
2017.0
Year

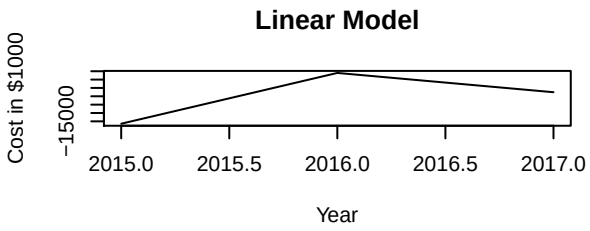

Linear Model
Cost in $1000
−50000
−150000
2015.0
2015.5
2016.0
2016.5
2017.0
Year

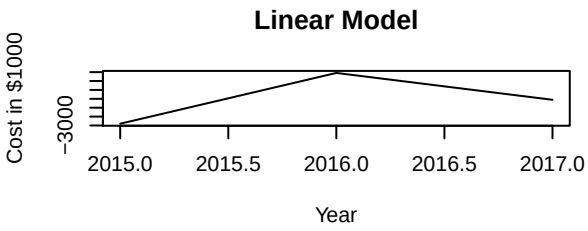


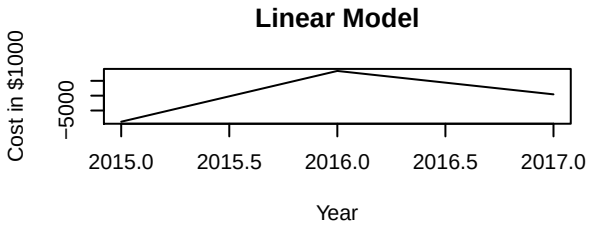
we have to add some labels lol but we can work on it after

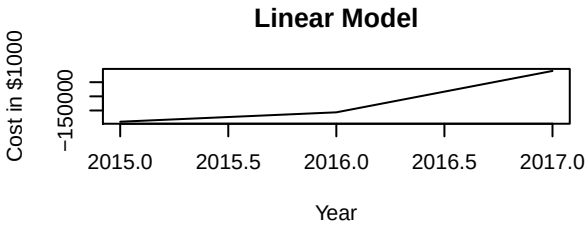
v11 is by far the closest predictors, follows by v5, v3 is meh and everything else goes to shit from there

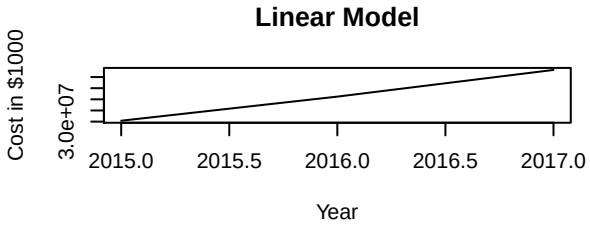
we will now plot the residual time series see how much we are off by









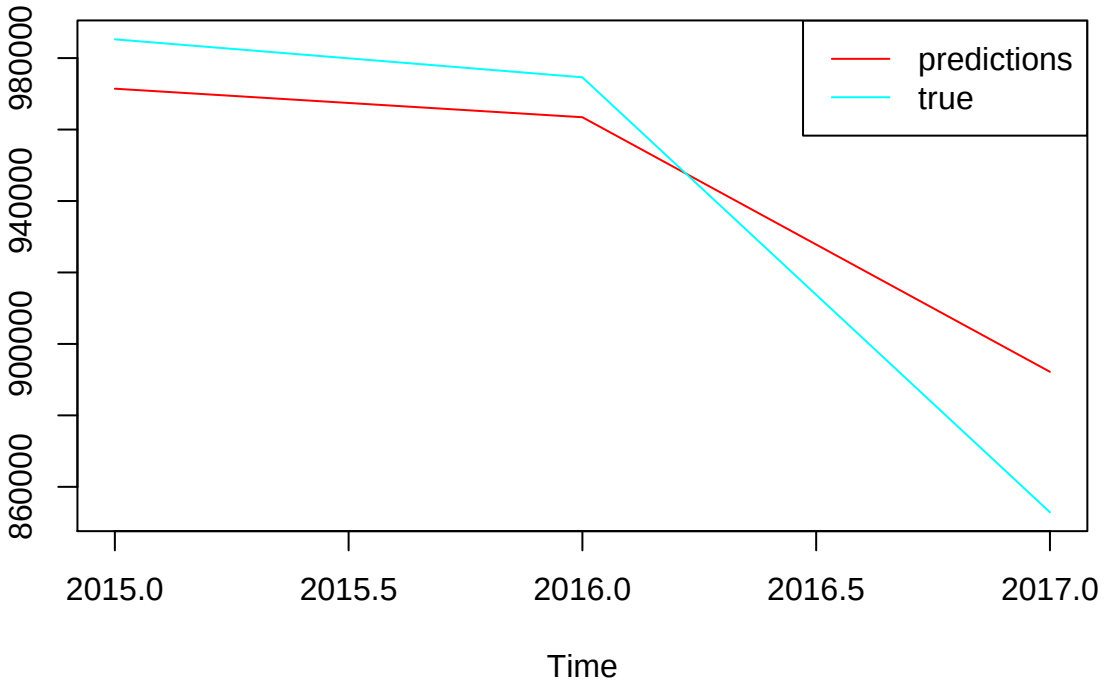


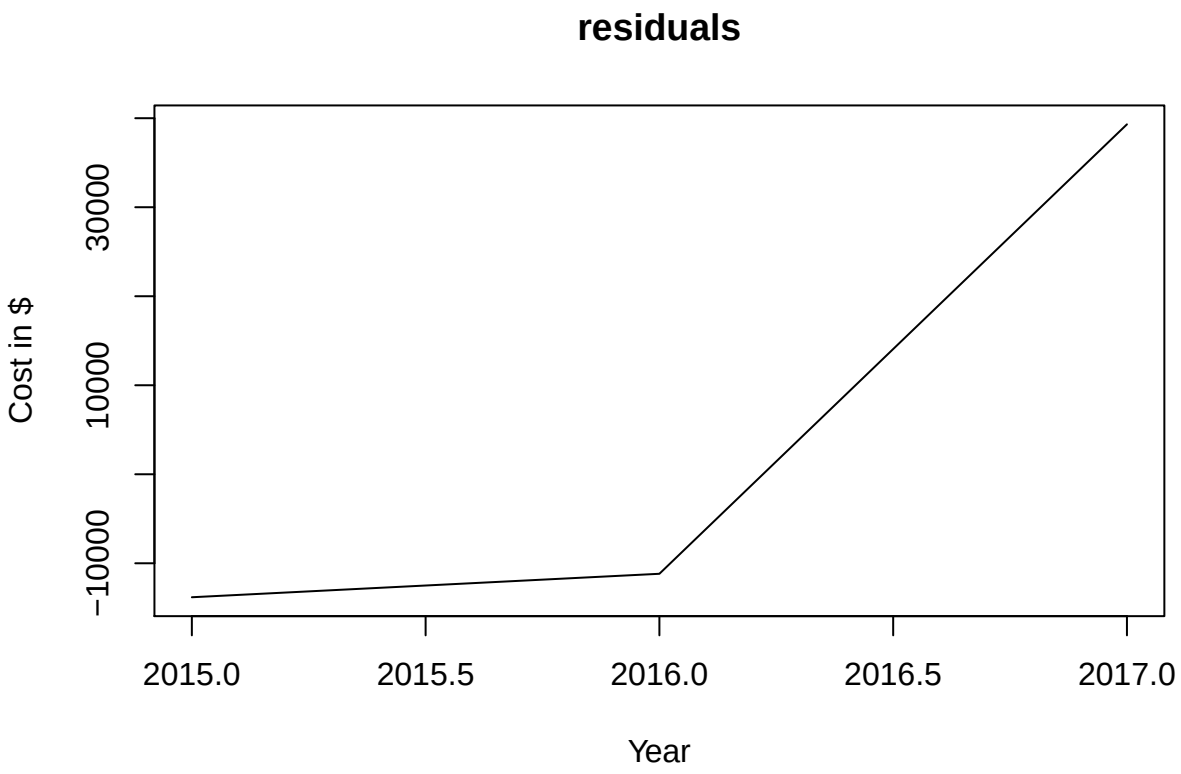
all centers at 0 but last 2, the second seems to have the smallest range which means its the best, so lets use the first three as our test

random forest

random forest on all of cash flows nets, var imp plot doesnt want to work so we will address that in the plot

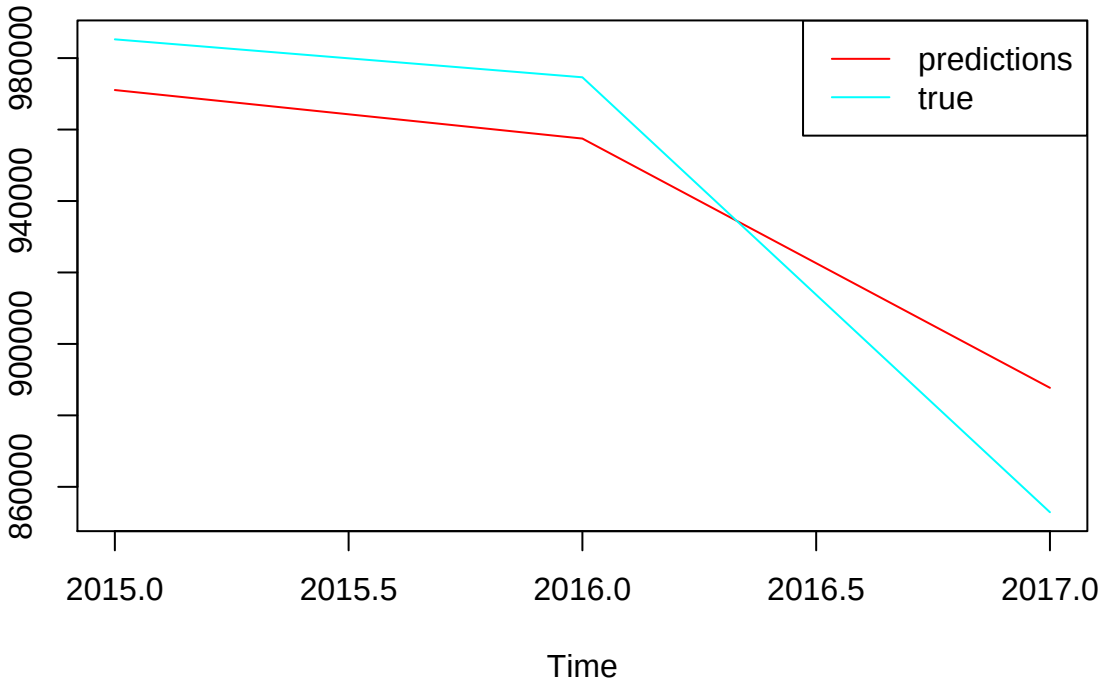
```
## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?
```


predictions
true
980000
940000
900000
860000
2015.0
2015.5
2016.0
2016.5
2017.0
Time

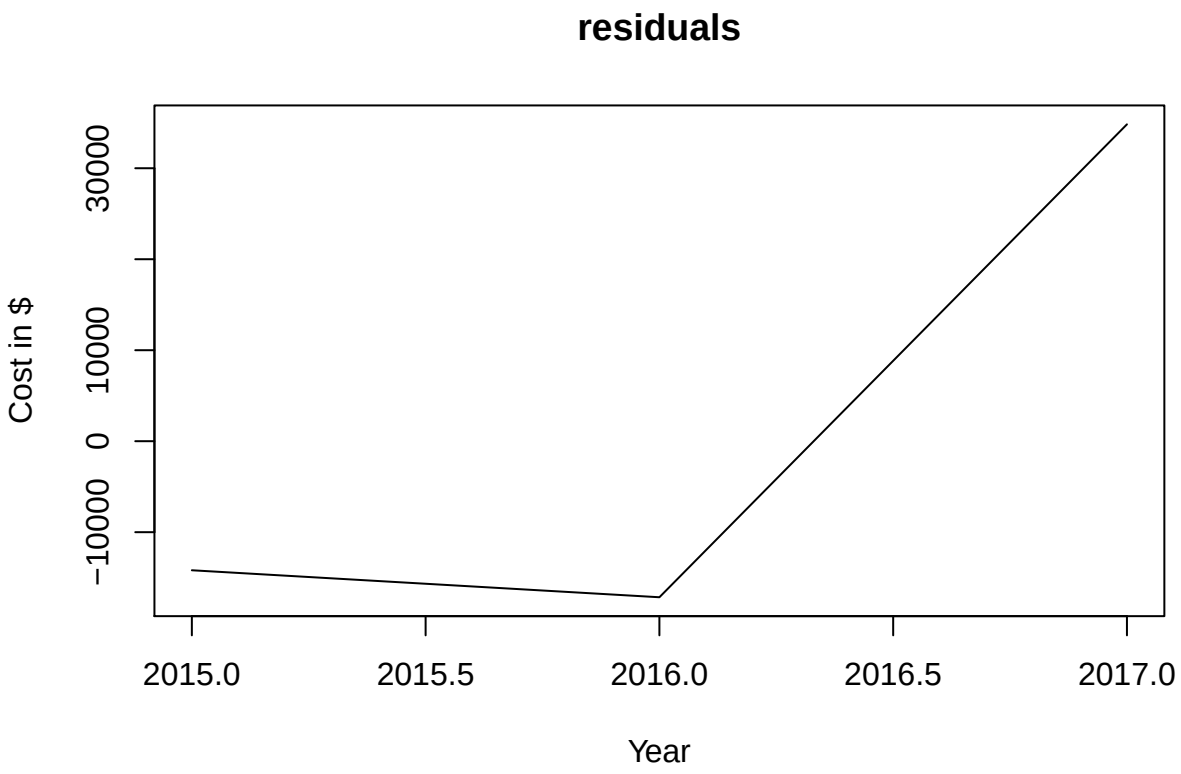


alot of varaiance so lets test on x

```
## Warning in randomForest.default(m, y, ...): The response has five or fewer
## unique values. Are you sure you want to do regression?
```


predictions
true
860000
900000
940000
980000
2015.0
2015.5
2016.0
2016.5
2017.0
Time

```
## [1] 3483.772
```

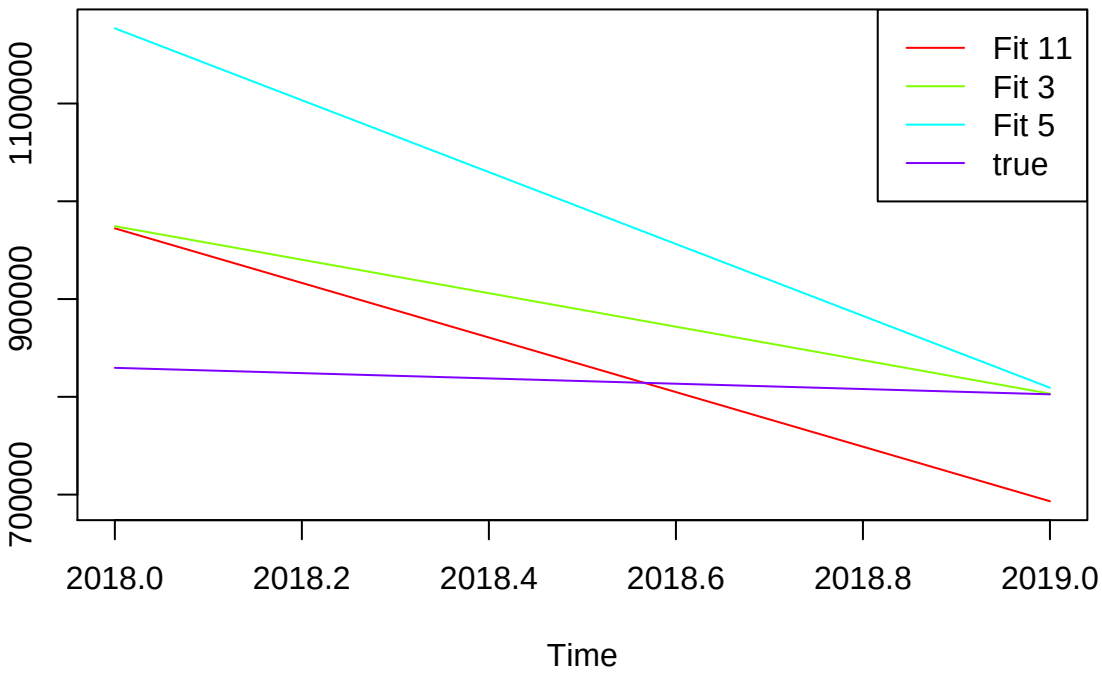


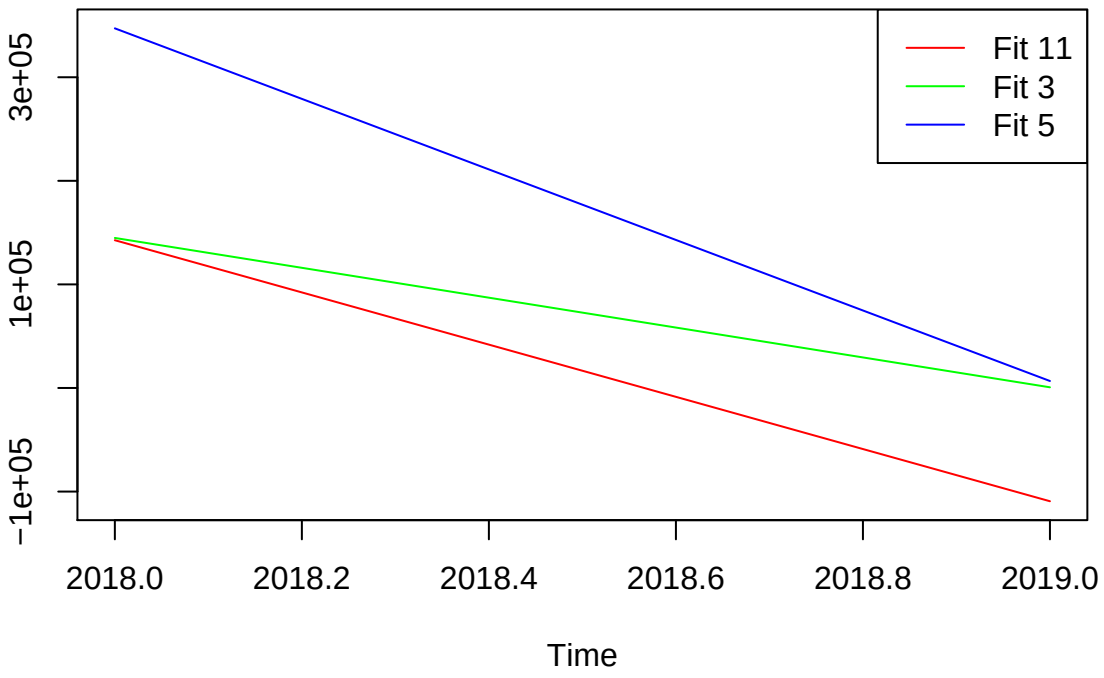
Running our random tree forest on our test: Having soe trouble working out the graphs for this

testing / results

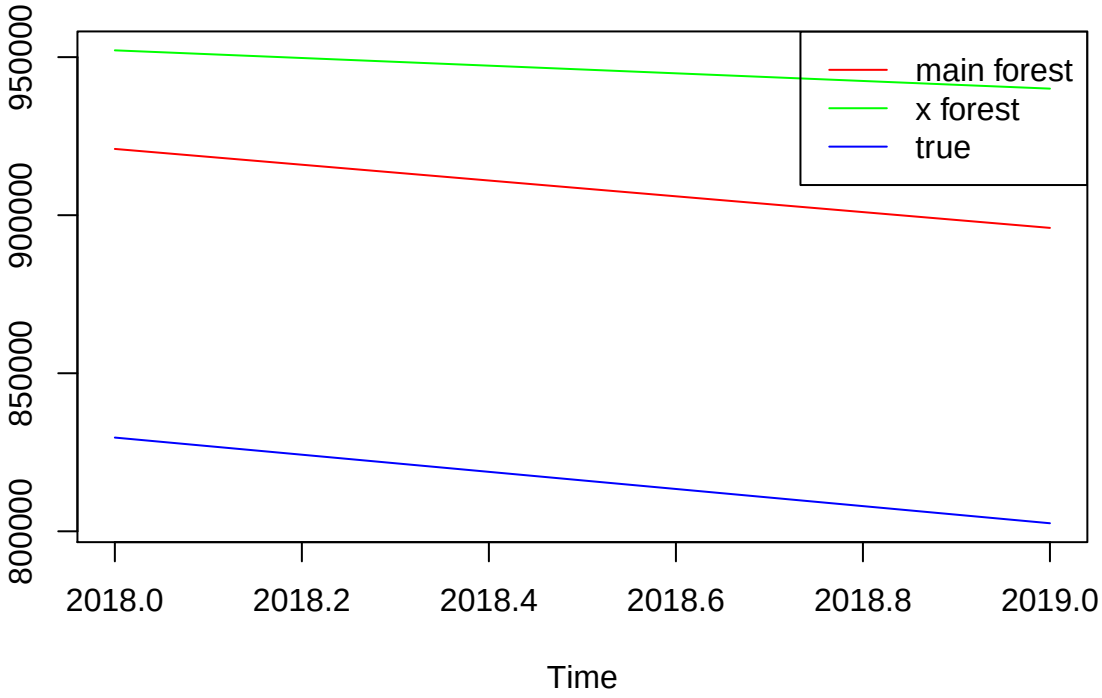
lets use the best fitted models on our training set

we will use fit 11, fit 3, fit 5, and both forests

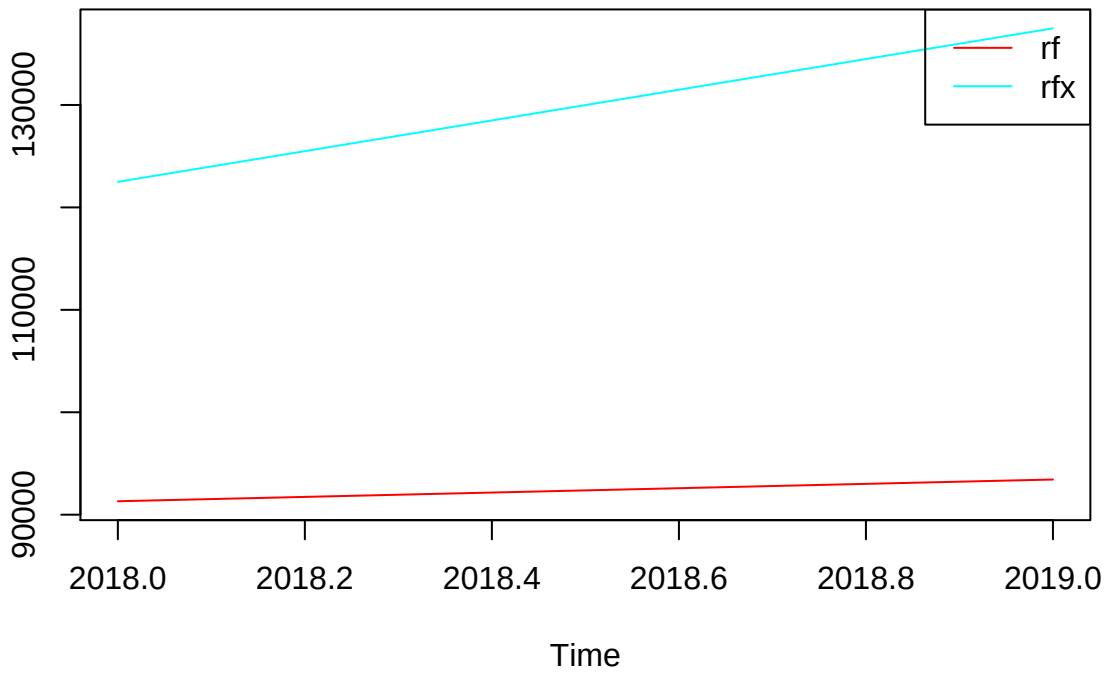

Fit 11
Fit 3
Fit 5
true
1100000
900000
700000
2018.0
2018.2
2018.4
2018.6
2018.8
2019.0
Time



check forest



a lot of error



conclusin

none of the models we got could really good predictors so we can conclude that there is no correlation with cashflow from other sources and cash flows from tutition and fees. make note that our data was weak given we only had 5 years to work with.