UMass Amherst Enrollment on Tuition

Alain Duplan *University Of Massachuessetts, Amherst*

There has been a recent trend of the University of Massachusetts in regard to enrolling students. In order to fit the surplus students, the university made a portion of common rooms and z-rooms into triples and quads, and extended hours of dining services. Given that more students are admitted each year, it was brought to light that the current size of the enrolled student body was affecting the price of tuition and dormitories. To hypothesize, **The enrollment number of the University of Massachusetts, Amherst can be used to determine the cost of being a residential student.** Using the public data provided by the institution, a study was conducted to analyze the relationship between the cost of attending UMass and the enrollment statistics. To account for the increasing population and value of higher education, the rate of change was used to predict how much the cost of attending increases upon each given year. Using step linear regression and random foresting, 2 models were able to be created to predict the rate of change within an average 3% margin of error. Both were able to show that enrollment was the primary predictor with high statistical significance. This would lead to suggest that the rate of which cost of attendance increases by is determined by the rate of which student enrollment increases by.

Introduction

Amherst is the flagship campus of the University of Massachusetts. Being so it is one of the most populated colleges in the state. With more and more students wanting to pursue degrees, UMass, like most schools, is admitting more and more students each year. Initially this doesn't seem to be an issue as an extended student body would lead to greater student engagement and more income directly to the school to fund projects such as renovations. However, when the once cozy z-rooms are now becoming "economy triples," and the common rooms on every floor are now turning into quads and triples, we begin to question whether not the school is prepared to hold so many students.

When asked, a majority of people would say that UMass makes most of its income from tuition fees as it is a common misconception that the univerisity has direct control on the price of tuition. However, this is not necessarily true as tuition is set by the Department of Education and as UMass is a public institution, most students are sensitive with how much they pay for their tuition and fees. With national college debt on the rise, the growing cost of education worries the next generation on how they will afford it. Many pointed towards UMass's habit of over enrolling students to be the cause of the increasing cost of tuition.

Taking into consideration that many students are now struggling to find rooms for next semester, and that overall student debt is at all time high, it is hard not to question whether there is a correlation between enrollment and tuition and fees. While controlling as many variables as possible we aim to find if enrollment is a determinant of the cost of attending UMass.

Data

Finding the data was relatively simple. Since UMass is a public institution, they are required to publicly display that costs as well the enrollment statistics. The University Analytics and Institutional Research department provides a back catalogue of numerous datasets regarding UMass Amherst. The two main reports we will use are the Admission/Enrollment and the Undergraduate Student Charges reports.

The Admissions and Enrollment reports come as individual reports for each semester since the spring 1984 semester. Though they come in various formats, all included the total enrolled and newly admitted students by college. They also included the total number of applicants and accepted new students per first and second choice major. We will be using the sum of each college for every fall semester since 1984 to build our data frame.

The Undergraduate Student Charges report is a single PDF file, which is updated every year. Beginning from the 1987 school year, it lists the in-state and out of state tuition and fees cost and along with the costs of room and board. After pulling this data from the file, we concatenate it to our data frame using the corresponding year.

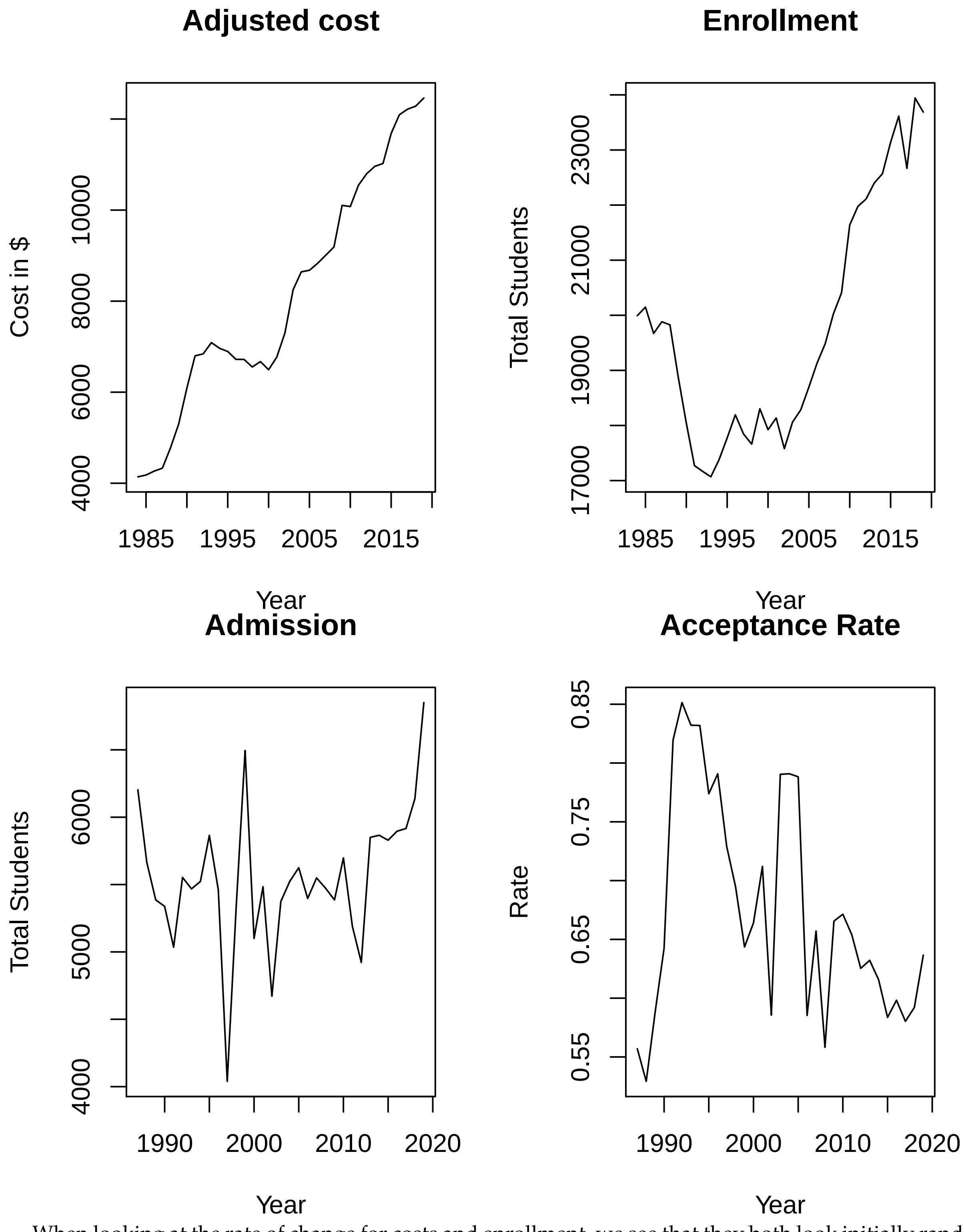
YEAR	IN-STATE	OUT-STATE	ROOM & BOARD	ENROLLMENT	APPLICATIONS	ACCEPTANCE
1984	1657	4481	2398	19993	NA	NA
1985	1947	4971	2244	20149	NA	NA
1986	1996	5020	2401	19670	NA	NA
1987	2048	5072	2584	19882	27339	15229
1988	2518	5950	2804	19826	28330	14993
1989	2980	7371	3064	18889	23287	13700

Once we have our data frame, we begin to adjust it to better fit our data. First we have to adjust all the costs for inflation. We deflate all of them to match the value of a dollar in 1984. Deflation was provided by the U.S. Bureau of Labor Statistics We then total three cost columns to find the average price to attend UMass. Since the reports were not really clear on both in-state vs out-of-state student ratios nor the amount of students living off campus, this became a road block. Using other sources provided inconsistent numbers, we simply used a 4 to 1 ratio between in-state vs out-of- state students respectively. This same ratio was used on students living on campus vs off-campus. This specific ratio was chosen as it seemed to be the average from all sources as they seem to range from as low as 60% all the way up to 90%. It is interesting to note that there has been a recent study that UMass has increased their acceptance rate for out-of-state students as they can make a larger profit due to higher tuition costs. These were summed to find the adjusted cost for a typical UMass student. Next we find the acceptance rate and admission rate by finding the ratio of the students who actually got accepted into UMass from those who applied, and those who admitted into UMass from those who were accepted. Since both enrollment and costs increase over time, we have to remove this factor from our data so we do not model natural growth. To do so, we added in the rate of change of enrollment and cost in comparison to the previous year.

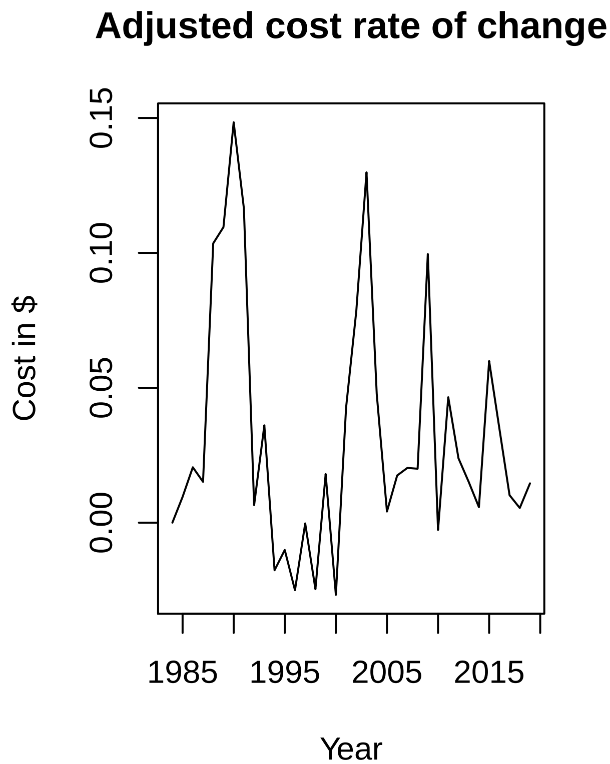
ADMISISSION	IN-STATE-ADJ	OUT-STATE-ADJ	ROOM & BOARD ADJ	ACCRATE	ADMRATE
NA	1657.000	4481.000	2398.000	NA	NA
NA	1872.115	4779.808	2157.692	NA	NA
NA	1883.019	4735.849	2265.094	NA	NA
6204	1878.899	4653.211	2370.642	0.5570431	0.4073807
5665	2208.772	5219.298	2459.649	0.5292270	0.3778430
5386	2504.202	6194.118	2574.790	0.5883111	0.3931387

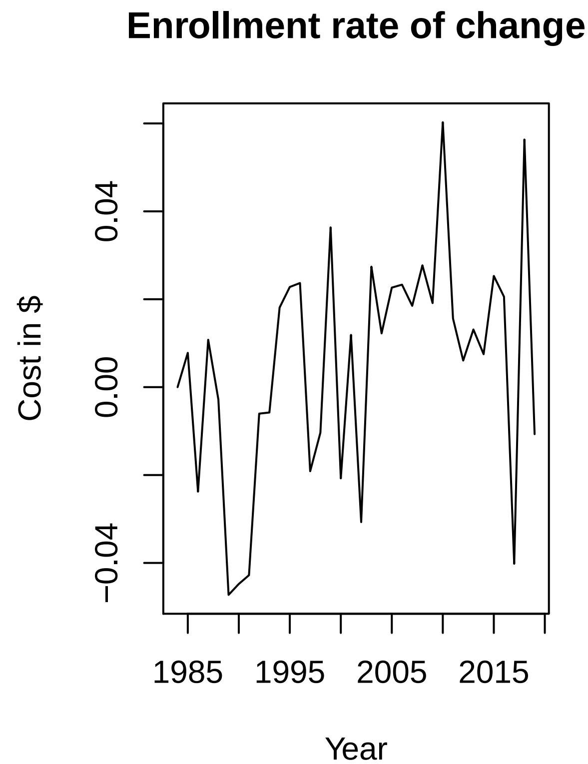
ADMRATE	COST	ENROLLMENTRATE	COSTRATE
NA	4140.200	0.0000000	0.0000000
NA	4179.808	0.0078027	0.0095666
NA	4265.660	-0.0237729	0.0205399
0.4073807	4330.275	0.0107778	0.0151477
0.3778430	4778.596	-0.0028166	0.1035318
0.3931387	5302.017	-0.0472612	0.1095343

Upon initial inspection, we see that regardless of inflation, UMass's costs seem to gradually increase as the years go by, showing proof that the value of UMass's education is growing. The enrollment looks to also gradually increase but not at a linear rate. However it is noticeable that both acceptance and admission rates are gradually decreasing.



When looking at the rate of change for costs and enrollment, we see that they both look initially random; however, we also see obvious peaks corresponding to different economic turning points, being 1990, 2003, and 2008, but no trends seem to be apparent.





All the graphs however share a common trait. They all show a huge dip in performance between the years 1992-1997. This could not be ignored and after doing some research it was discovered that during this time there was a lot of racial tension happening at UMass. These issues all seemed to stem from the 1992 ALANA agreement and the administrative response to it during 1994. It resulted in a sit in at Goddell in 1997 due to the increasing unrest coming from minority students on campus. Without getting into too much detail about the situation, the important thing to note is that it affected every graph negatively and couldn't be explained through natural or economic trends. To account, a dummy variable was made to only include those years 1994-1998 and the variable will be used when formulating the model.

Modeling

To begin we will reserve the last 6 years of the dataframe as our testing set, and everything else as our training set.

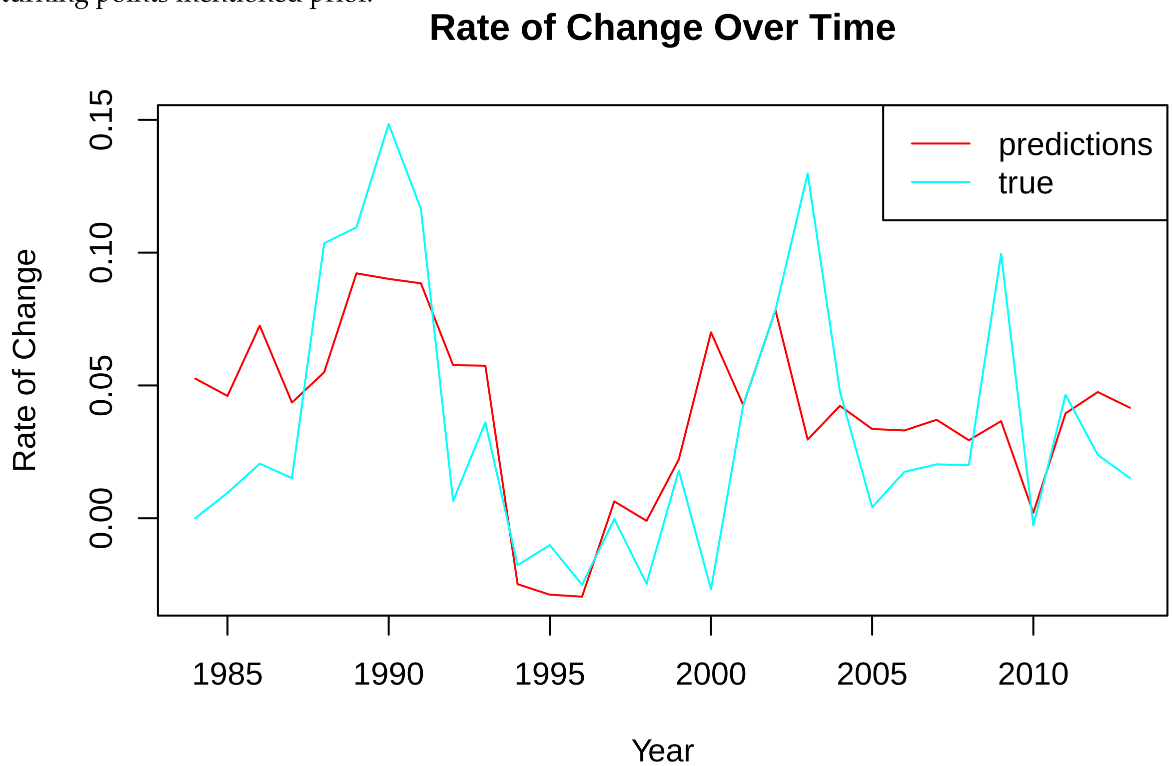
The first model we will use is a linear regression model. We will try to predict the rate of which cost changes by using admission rate, acceptance rate, enrollment rate of change, and our dummy turmoil variable as our determinants. After the first run we found that together none of the variables are significant. To fix this we run a step function which iteratively removes the least statistically significant variable until the remaining are significant. The conclusion of this function tells us that turmoil and enrollment rate are most significant. After running a linear regression on those two variables we find that they both are significant with p-values under .01. This would give us the equation:

$$COSTRATE = .052595 - (.062254)TURMOIL - (.838094)ENROLLMENTRATE$$

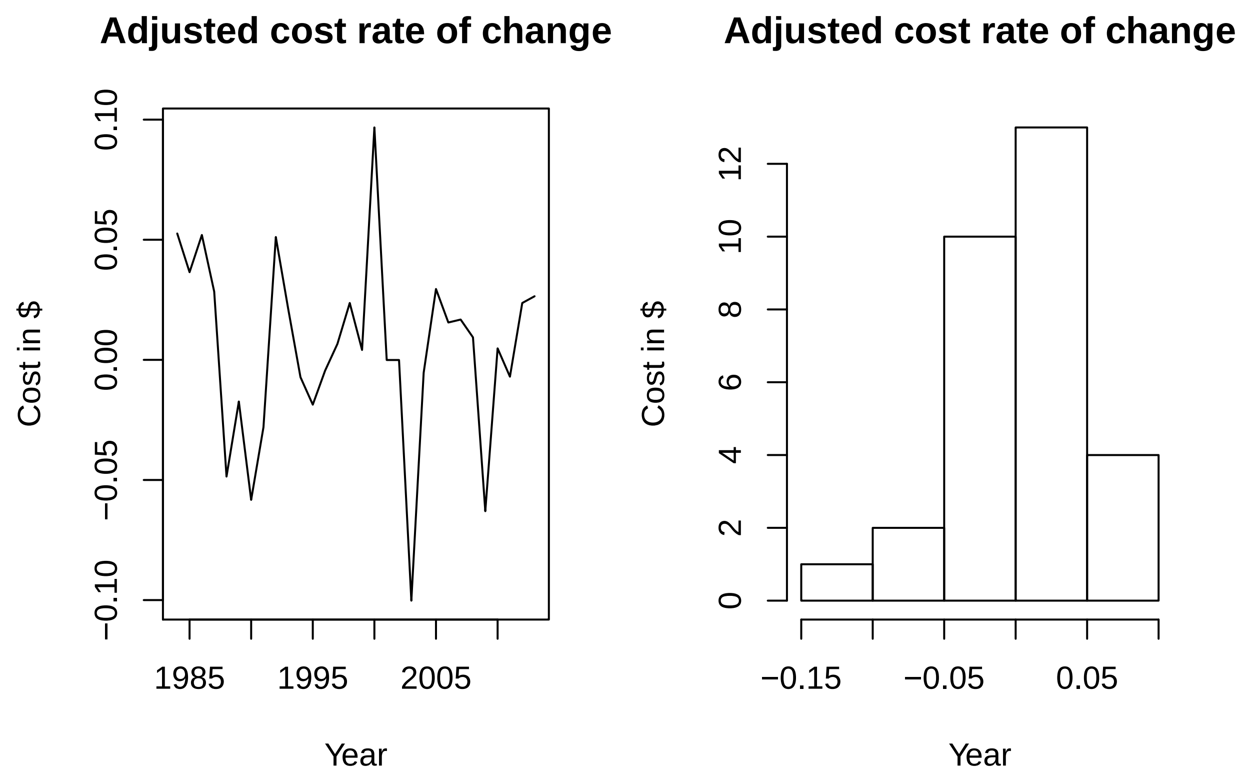
This would mean that on average, the turmoil of 1995 decreased the rate of change of cost by 6% and doubling enrollment(rate of change of 1 would mean that enrollment has doubled in size), would cause the rate of which cost increases to decrease by a huge 83%!

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0525952	0.0085977	6.117337	0.0000026
ENROLLMENTRATE	-0.8380942	0.2984705	-2.807963	0.0097471
TURMOIL	-0.0622536	0.0197196	-3.156935	0.0042619

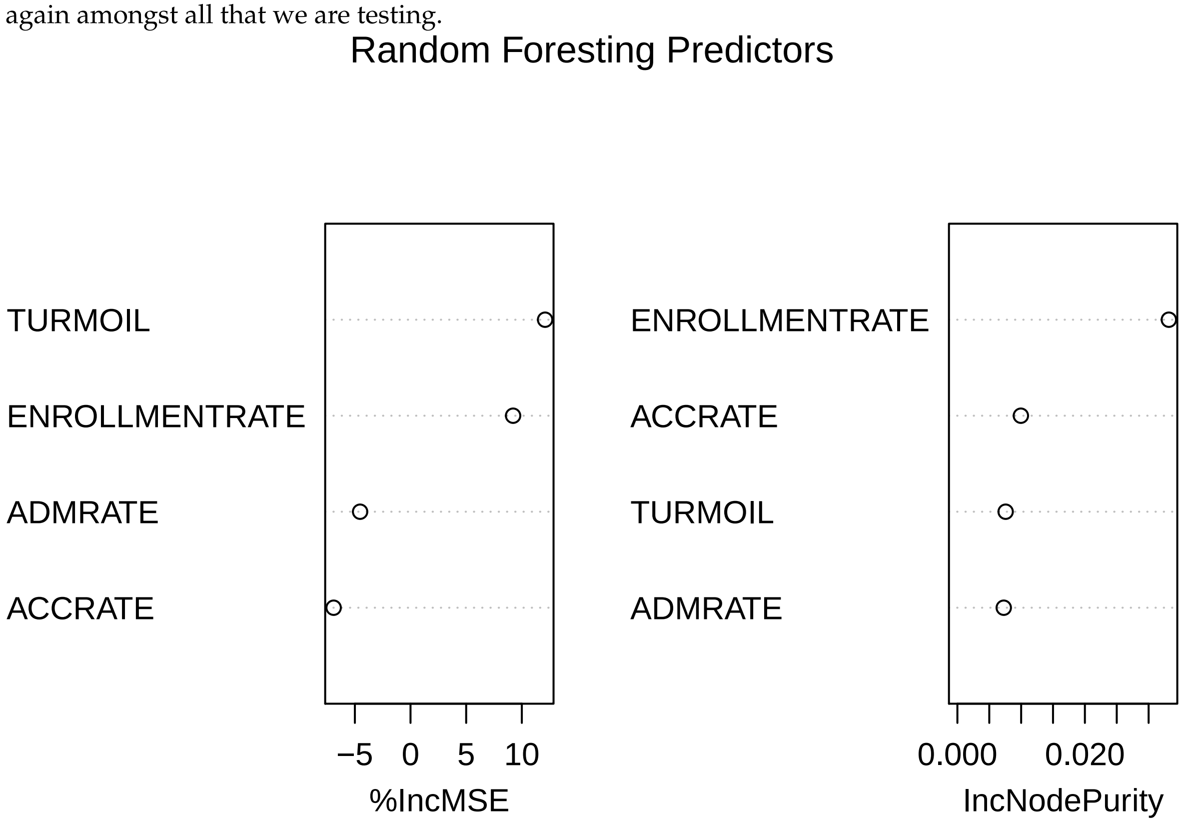
Once we are able to run this model on our training set, we see that it is a fairly decent predictor of the time series and does a great job of smoothing the data. However it seems to underestimate the economic turning points mentioned prior.



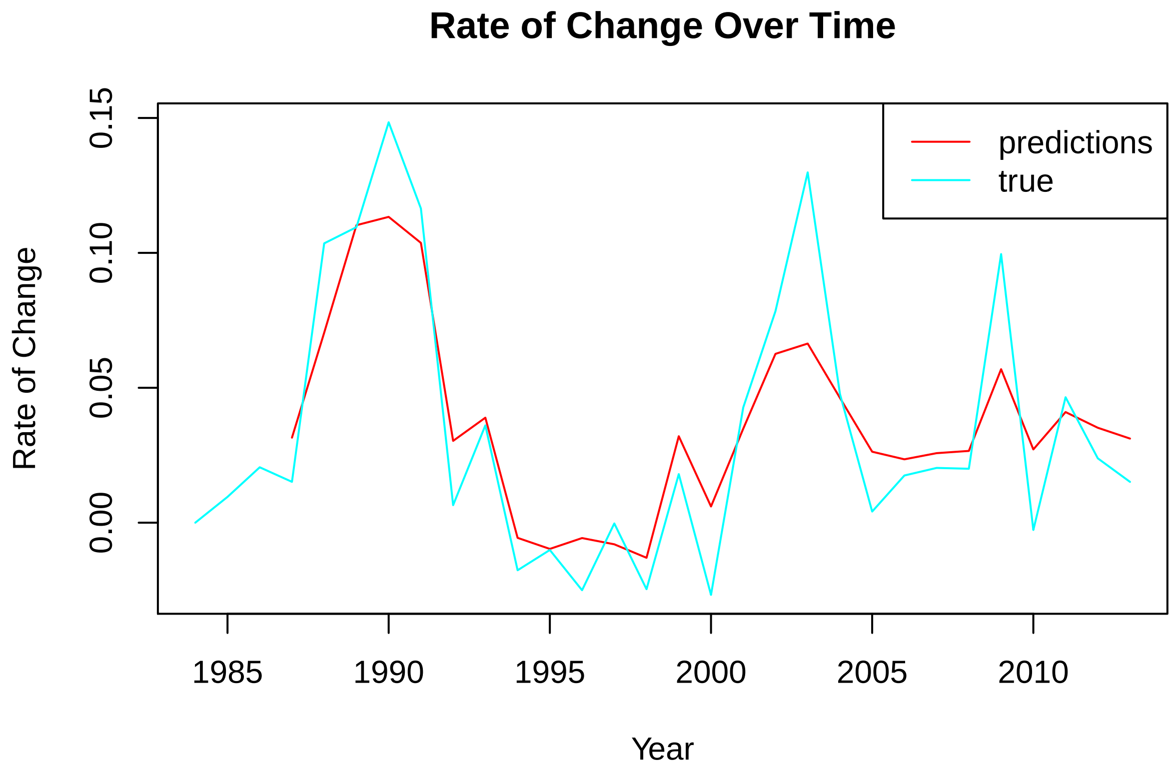
After we run the model, we plot the residual. We find that the residual plots show a slightly skewed distribution to the left with no trends, while being able to remain with a 10% range. We also find that the root mean square error is only 3.86% and the sum of the residuals is 14%, so it slightly over predicts. So we can conclude this a fairly accurate predictor that overestimates.



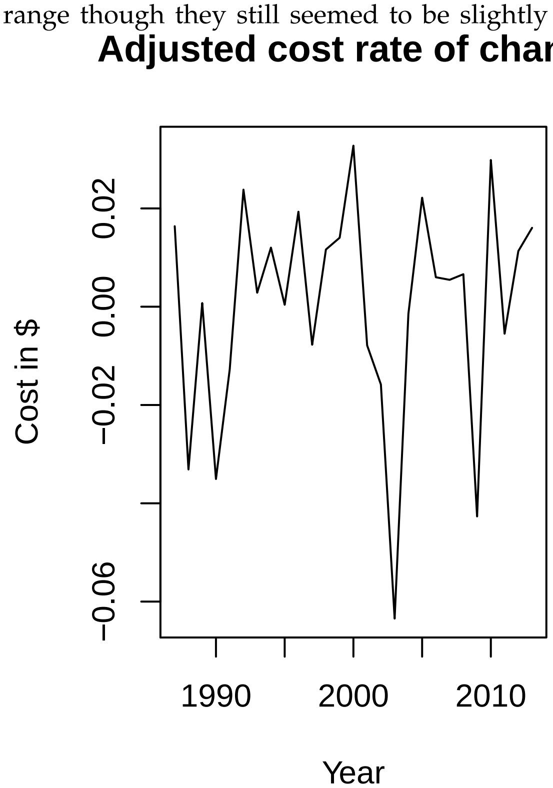
After we run a random foresting model to try to get a better model. Setting the seed to 1234, we see that enrollment rate and turmoil comes up the strongest predictors by means squared error and node purity, again amongst all that we are testing.

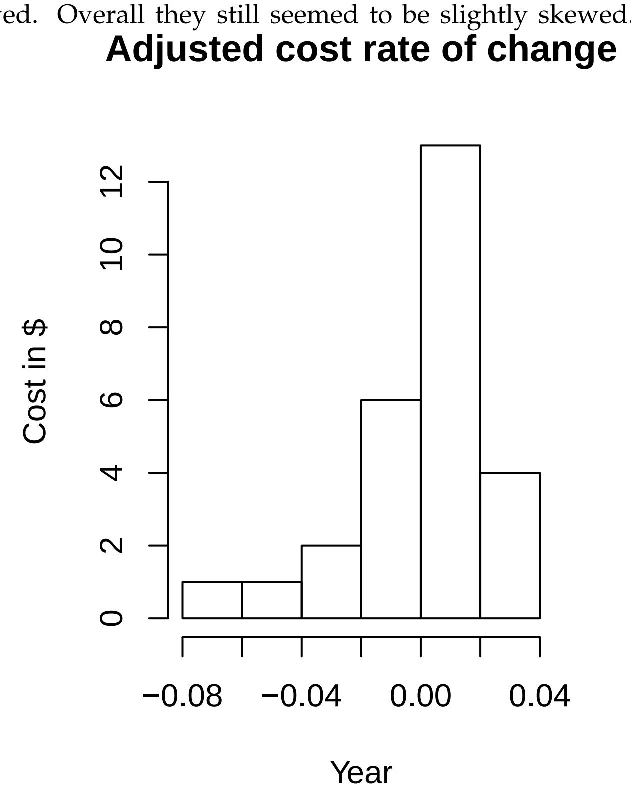


Running the random forest model on the training set we see it is a much smoother time series then the linear model and does a better job at predicting the peaks, though it is not perfect at doing it.



After checking the residuals, we find a root mean squared error to be 2.22% and the sum of residuals to be 2.5% which is much lower than the prior model. The residuals stayed within about a 5% range though they still seemed to be slightly skewed. Overall they still seemed to be slightly skewed.

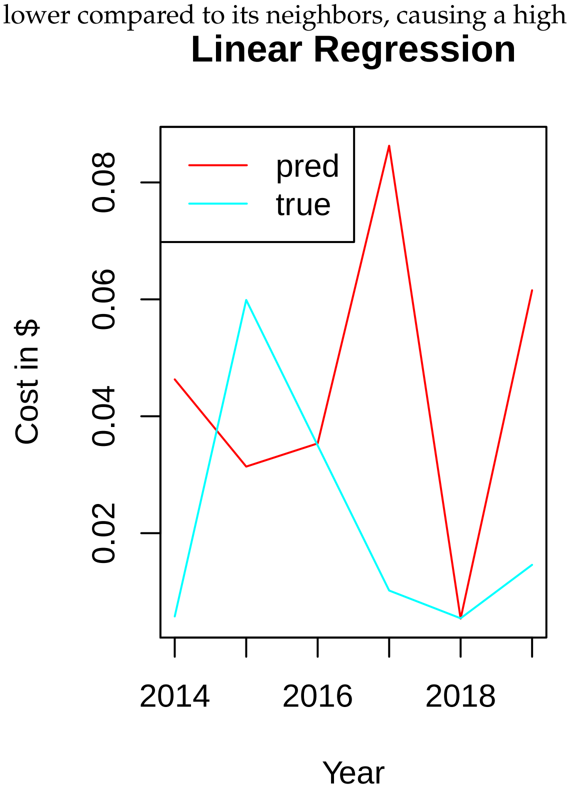


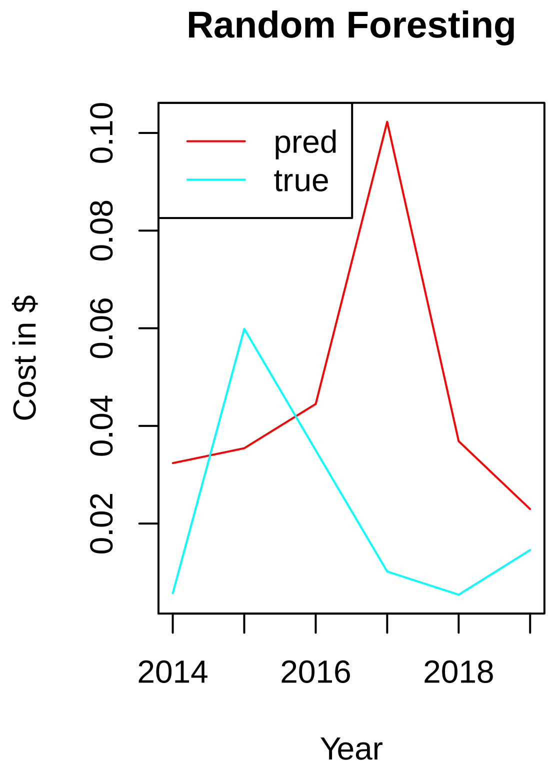


Together, it is believed that both models are accurate enough that we can use on our testing set.

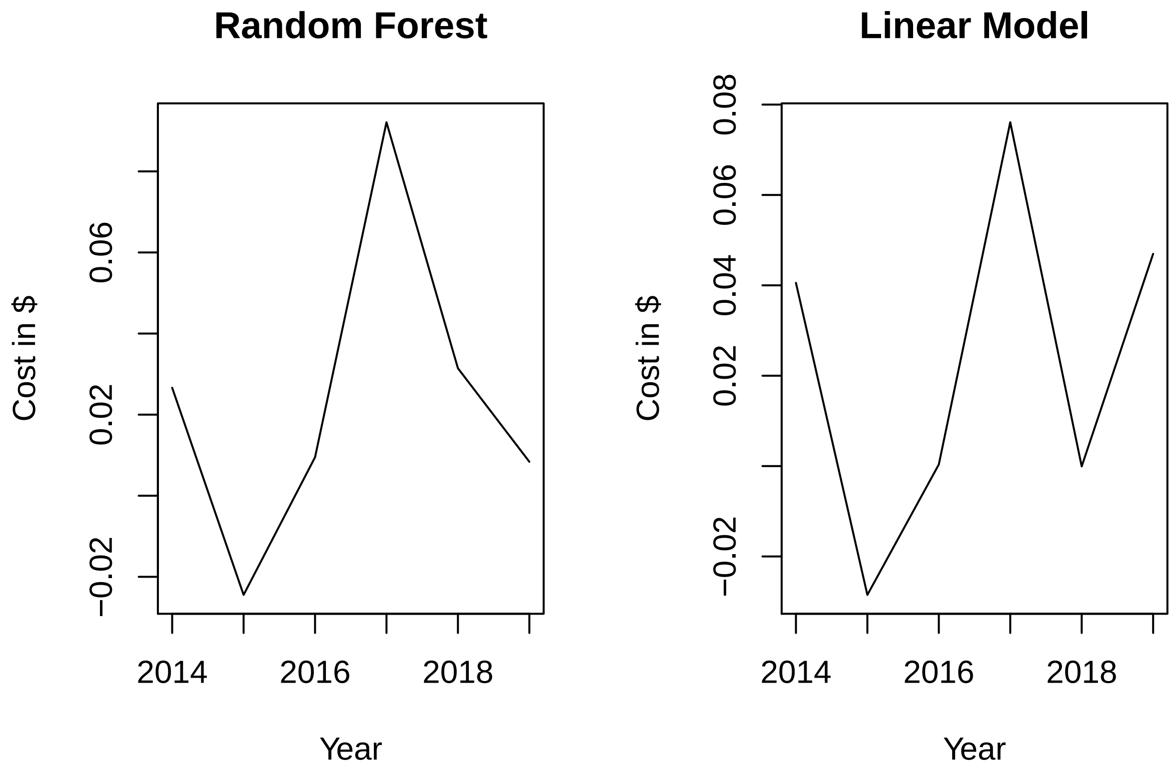
Results

Now that we have our models, we will run them both on our test data set. Both models stayed relatively close to the true prediction but predicted a much higher than expected rate for 2017. This could be due to the fact the rate of enrollment for that year doesn't seem to follow the trend other years did as it is much lower compared to its neighbors, causing a high error.





Upon testing the residuals, both models managed to stay within 8% of the true values, and this value is only high due to the error of 2017. Both of the time series show no trend in errors and the given small test sample makes the histogram obsolete. The linear model has a root mean squared error of 4.17% and a total error of 13.54%. The random forest model has a root mean squared error of 5.86% and a total error of 14.66%. This huge error in both models is due to the year 2017 so if it was removed from the set the totals would only sum up to 6%.



In this regard, it is safe to say both models were able to accurately predict the rate of change of cost within a roughly 5% interval.

Conclusion

We were able to draw a clear correlation between the rate of which enrollment increases and the rate of which costs increases. While controlling for not only inflation, but the effect of time, we were still able to predict the increase of cost to attend UMass within a roughly 5% interval. Though this is not a direct causation between the two, we can conclude that expected enrollment is at least considered when deciding the costs for next semester. The models weren't able to accurately account for business and economic cycles which can affect the price of which tuition is set. We are also only looking at purely the populous of UMass, not other factors such as staffing, programs, and resources provided by the school. With all the new buildings and renovations, there is a direct affect on room and board cost as it increases the value of life off campus. We must also keep in mind that UMass doesn't directly choose their tuition as that is the decision of the Department of Education.

Taking all of these into consideration, it remains undeniable that the relationship between the two holds true. There is in fact a correlation present; therefore, we are unable to reject our hypothesis. There is a significant relationship between the amount of students enrolled at the University of Massachusetts, Amherst and the cost of attending while staying on campus.

References

"Admissions And Enrollment Reports." UMass Amherst

Couloute, Lucius. "The 1997 Goodell Sit-in: A Movement Tired of Broken Promises." Racial Umass, 2015

Thys, Fred. Study: Out-Of-State Students With Lower Test Scores Admitted To UMass Amherst More. WBUR, 29 May 2018

"U.S. Bureau of Labor Statistics." U.S. Bureau of Labor Statistics, U.S. Bureau of Labor Statistics, 28 Apr. 2020

"Undergraduate Student Charges." UMass Amherst