COHORT DEFAULT RATE

HOW EDUCATION INSTITUTION FACTORS AND DEMOGRAPHICS EFFECT THE COHORT DEFAULT RATE?

Akanksha Chavan, Praveen Samuel J, Rishabh Srivastava, Uday Reddy Vangala

ISM 6137 Statistical Data Mining | UNIVERSITY OF SOUTH FLORIDA.

Table of Contents

Executive Summary

When it comes to debts, Americans nearly owe about $1.73 trillion on their student loans, which is much higher than credit card debts. Student loan debts is the second-highest consumer debt category, of which make up of about 65% of today's college graduates. The impact of having such high rates of student loans equates to disturbingly high default numbers, which means that the education that students received did not give them the benefit of repaying their student loans in a timely fashion. In this research paper, our aim is to analyze the relevant Educational Institution factors and Demographic factors that impact the Cohort Default Rate (CDR) and formulate possible recommendations for target stakeholders to have viable solutions to minimize the problem.

The CDR data used in this research project is sourced from the Integrated Postsecondary Education Data System (IPEDS) and the National Student Loan Data System (NSLDS) databases. The IPEDS is a web-based data collection center since 2000, which incidentally is a large-scale survey that collects institutional-level data from postsecondary institutions in the United States and associated U.S. jurisdictions. The NSLDS is a database that collects data from many verified agencies like FFELP, ED Services, DCS, DLS, and schools for information on Federal Perkins Loan Program, student enrollment and federal aid overpayments. The NSLDS is part of the US Department of Education's central record for student aid and other student aid programs. In this paper, we are analyzing 22 variables that capture Education Institution factors and Demographics that we think will affect the CDR. Our data includes 7 years of data from 6694 universities from all US states, and universities from 6 major accreditation agencies. Each year of the data has around 2% of institutions of a CDR greater than 30%. Our data consists of 2089 public, 2005 private non-profit, and 2900 private for-profit universities.

To analyze the trend in CDR's, we have segregated and handled null variables from the above data which considered to be having impact on CDR. We have conducted row wise analysis of data i.e., Institutions. In this we have only considered Institutions from top accreditation agencies and which offer prominent degrees. Feature Engineering on courses awarded by the Institutes, exclusive columns and inculcate Carnegie classification of degrees. The models we focused on are Multi-Level Linear Regression for predicting the CDR.

Problem Definition & Significance

The target audience that would generally find our research interesting and helpful would be the financial leaders of universities and colleges (Commisso, 2017). Financial leaders are generally taxed with challenges in complying with standards from the US Department of Education Cohort Default Rate requirements. With the problem that our stakeholders (financial leaders) face, this proves to show a good opportunity and room for improvement. Analyzing the cohort default rate (CDR), along with the lack of effective student loan default management plans at institutions, stems from more than 194,000 people from US proprietary institutions defaulting on their federally granted student loans per year (Commisso, 2017). According to the US Department of Education, in 2011, more than 1 in 10 federal student loan borrowers entered default within 3 years of repayment, and these numbers represent the highest peak on record since the mid-1990s.

Interestingly, 95% of the borrowers with outstanding debt on their education still owe money to repay their student loans; and out of these numbers, only 52% of the students who have taken a student loan did not feel it was worth paying back. And out of this cohort, only 6.7% of eligible student borrower apply for loan forgiveness. The new "3-year" CDR (or cohort default rate) in our dataset is a more accurate representation of the defaults, thereby increasing universities and colleges' accountability for their

students' financial well-being. The new 'three-year CDR' in our dataset is a more accurate and meaningful measure of defaults, increasing colleges' accountability for their students' financial well-being. This will also have effect on schools as the default rates increases above certain levels it has a risk of losing access to federal student aid.

The importance of analyzing and devising a recommended solution for decreasing CDR for US universities can help lower student debt and increase student and college accountability to pay back student loan debt. To discover possible solutions, we need to analyze the causes and interventions for helping financial leaders to tackle the problem and implement a proper financial management system and programs in place for students to easily understand and reduce their student loan debt within a certain period after their graduation.

Prior Literature

There are studies that have identified that the cohort default rates have been used as a prominent indicator of whether if institutions are giving students meaningful education, in terms of preparing students to attain jobs that can help pay off their student debt loans (Charles et al, 2016). Having a CDR above 40% puts the institution at risk of losing their title (Title IV) and federal aid funding, and thereby closing the institution, because of the lack of federal aid. Having a CDR above or equal to 30% could also mean that the institution becomes ineligible from giving federal funding to students, such as Pell Grants and other federally guaranteed student loans. This signifies that colleges and universities are highly dependent on federal financial funding, and the threat of losing access to this source of revenue will not only impact educational opportunities for students (who lose this support), but it could also significantly impact the financial wellbeing of the institution (Charles et al, 2016).

Since federal aid policies for post-secondary education has changed, findings have showed that institutions need to ensure accurate enrollment reporting, ranging from institutional characteristics, student characteristics and background (i.e. race or ethnicity), socioeconomic contexts of students (i.e. parental income), college experiences, and more (Gross et al, 2010). According to Shen & Ziderman, loans are actively being targeted to certain social groups, that have a less guaranteed rate of paying back loans sooner, because of lack of education success factors and job securement (2009). Studies have also compared the types of borrowers (graduate vs undergraduate) from certain institutions that have certain socioeconomic and parental backgrounds. Federal loan limits are also dependent on how well a student does academically and the students' dependency on their parents for loan guarantee (such as filing for FAFSA) (Looney & Yannelis, 2019).

Another key point to highlight is that there are studies and research on emerging issues of non-profit, for-profit, accreditation, and minority-serving institutions. While CDR problem is not restricted to one portion of the post-secondary education, the for-profit sector has received more attention in terms of federal aid. According to the US Government Accountability Office (GAO), it has recently discovered that the average for-profit institution receives 66% of its total revenue from FSA programs (Hillman, 2013). However, accreditation of institutions plays a key role here since accreditation agencies are involved in serving as gatekeepers for granting Title IV funds to universities to earn revenue. There have been increased efforts change policies to boost graduation rates and investigate policy efforts to reduce CDR across a diverse range of students; however, the new changes are not immediate, and will proceed slowly but surely make an impact in the long run.

Data Source and Preparation

Data related to the Educational Institution factors was collected from Integrated Postsecondary Education Data System (IPEDS), which is a large-scale survey that collects institution-level data from postsecondary institutions in the United States and other U.S. jurisdictions. IPEDS defines a postsecondary institution as an organization that is open to the public and has the provision of postsecondary education or training beyond the high school level as one of its primary missions. Data related to the loans, students entering into Default period and their track of repayments was considered from The National Student Loan Data System (NSLDS®) database which is the U.S. Department of Education's central record for student aid. It contains data from schools, guaranty agencies, the William D. Ford Federal Direct Loan (Direct Loan) program, and other U.S. Department of Education programs. The NSLDS database provides a centralized, integrated view of federal student aid loans and grants that are tracked through their entire lifecycle from aid approval through disbursement and repayment

Data collected has 20 files of year wise data. Each datafile has information of Institutions on the columns and Institution names on the rows. We have 6694 Institutions in the rows and 2392 different attributes among the columns. Here are some more details of the original dataset:

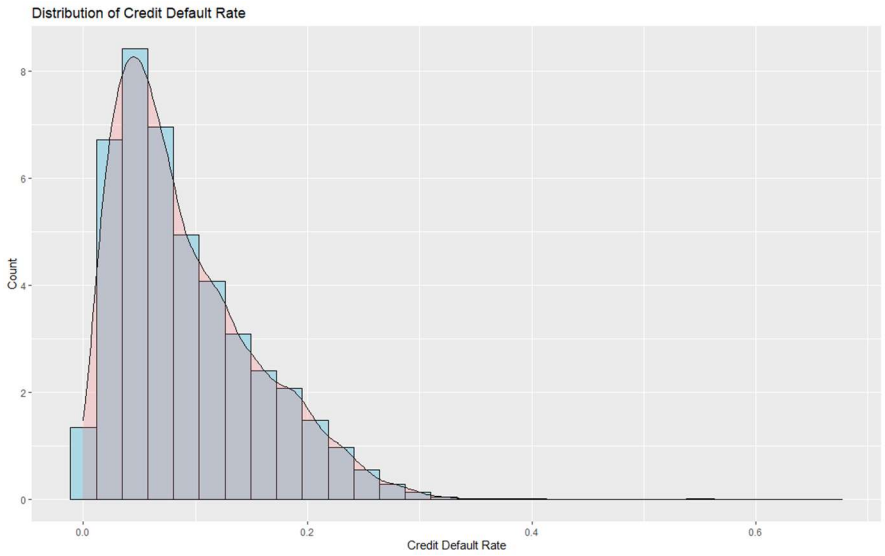
- 2089 Public, 2005 Private Non-Profit and 2600 Private For-Profit Educational Institutions.
- 65 Educational Institutions has CDR greater than 30%
- Educational Institutions from 41 accreditation agencies and 10 Regions.

Dataset contains 2392 features, but in those we have nearly 2000 features either all Null's or "Privacy Suppressed". Nearly 250 features we have more than 75% Null values. We filtered put those variables and end up with nearly 142 meaningful variables. Among those we have 41 columns which consists of percentage of degree's awarded in 41 different streams. We did feature engineering on those variables and derived top 3 degree's awarded by an Institution based on percentages. We have few exclusive variables like NUM_PRIV and NUM_PUB which represents the number of students in an Institute, we combined such variables and made it one. Few columns are filled with codes, we replaced those with actual strings e.g., Regions are coded with numerical values, we changed to actual Region names. In the final dataset we ended up with 35 columns.

Dataset has 6694 institutions in the rows, from 41 accreditation agencies and 10 regions. We have considered institutions from major accreditation agencies and offering prominent major's/degrees and removed the Institutions which offer vocational degrees by accreditation filters and Institution wise keyword filters. We have factorized, releveled and recoded categorical variables before modelling.

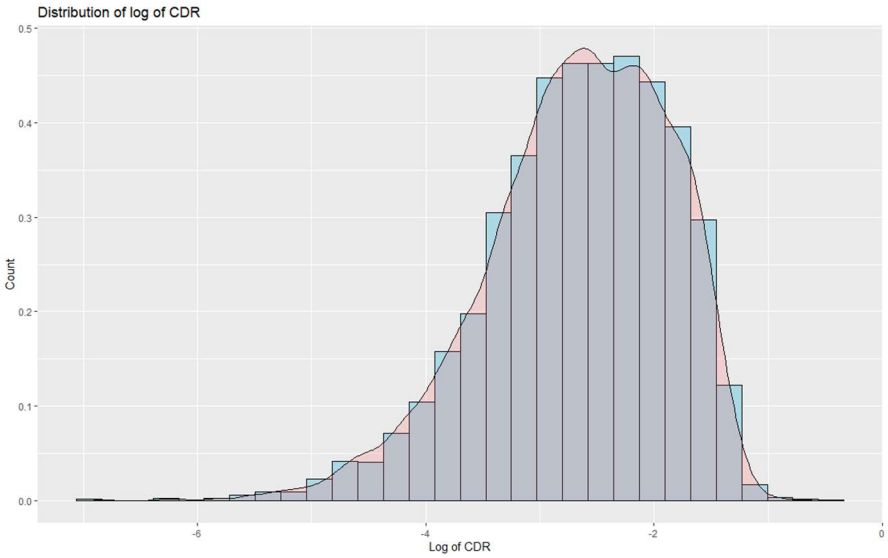
Exploratory Data Analysis & Visualizations

Distribution of Target variable - CDR: Our assumption here is that all the colleges would try to minimize their CDR and most of the colleges should lie in a certain range and data should be normally distributed in the ideal world.



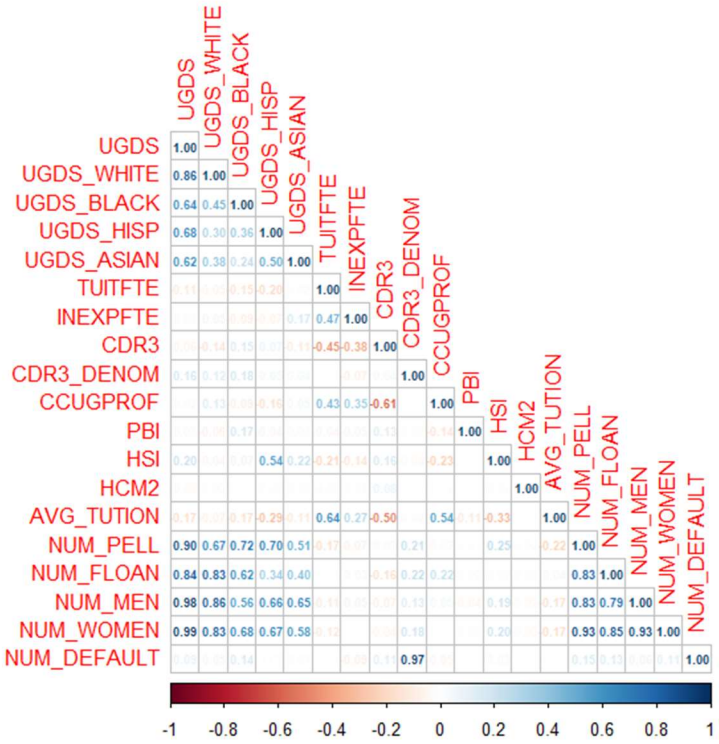
As you could see here there is range between 0.02 to 0.1 where most of the colleges fall and they are trying to keep their CDR as low as possible which is what we were expecting. But you would notice that data doesn't seems to be normally distributed and is right skewed. Also, we can see that CDR value ranges from 0.0 to ~0.7

Distribution of log of CDR: Taking log of CDR may end up making the data normally distributed.



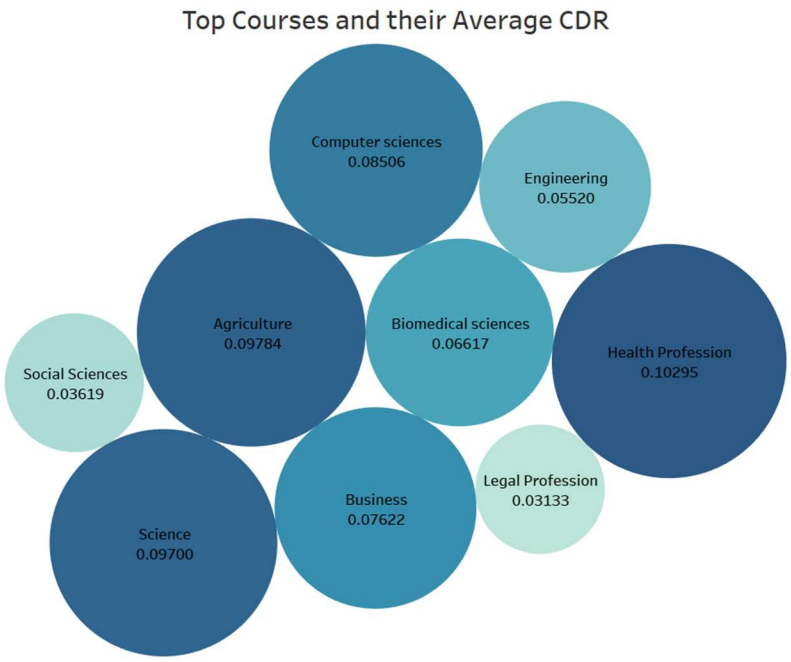
This graph shows the distribution of log CDR across the data. As you can see after taking log graph seems to follow a normal distribution. So, when applying model techniques we would be using log(CDR) instead of CDR.

Checking correlation among the continuous variables: This matrix would help us understand the predictors which have high correlation among themselves as well as with the target variable. Predictors that have higher correlation could not be used together and would penalize the model. Any correlation between -0.7 to 0.7 should not have higher impact and could be used together.



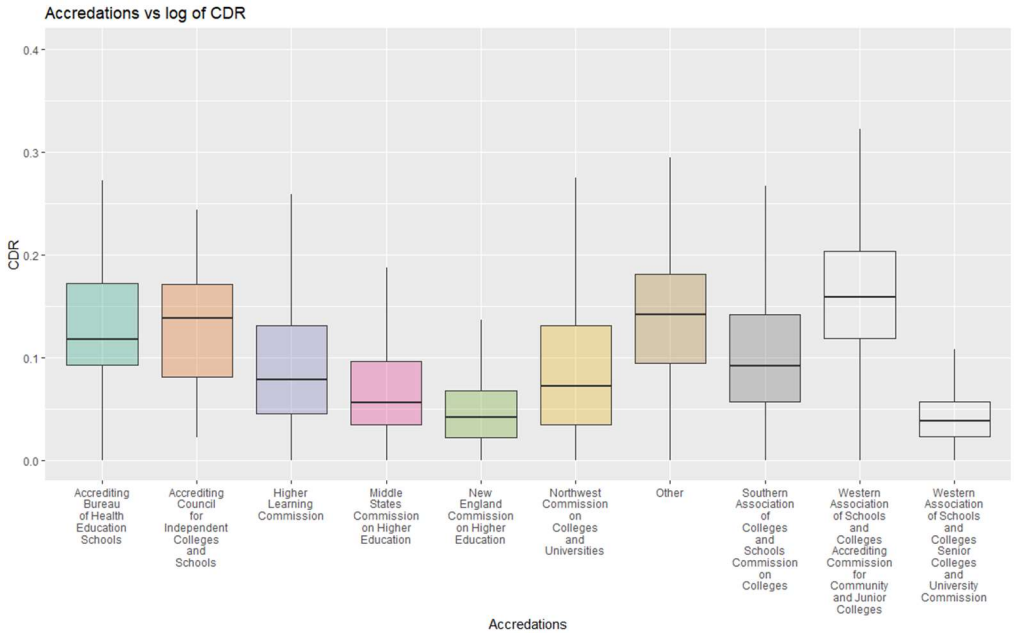
In the above graph you can see that UGDS is having higher correlation with UGDS_WHITE, UGDS_BLACK, UGDS_HISP, UGDS_ASIAN which is expected are they come from the same category of UGDS. Also, NUM_PELL, NUM_FLOAN, NUM_MEN, NUM_WOMEN have higher correlation among themselves as well as with all the categories of UGDS.

Top Course and their average CDR: There might be some of the courses which may have higher and lower CDR rates, and this could be a good predictor in understanding the effect of higher or lower CDR rates in certain courses.



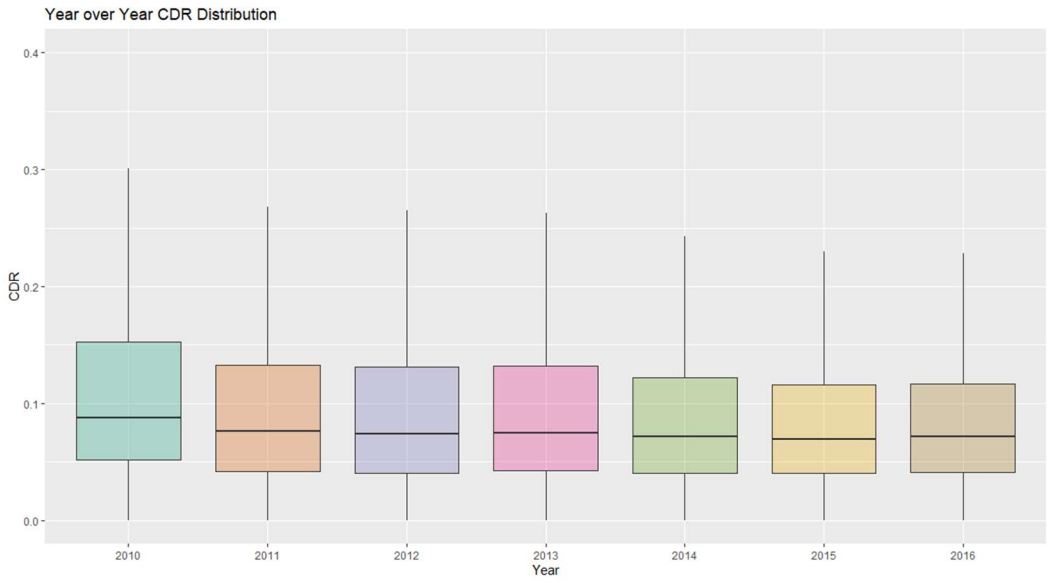
Above graph shows bigger bubbles for the courses that have higher CDR and vice versa. Among all the different course these are some of the most popular courses in all the colleges. Here you could see that Health Profession has the highest CDR rate among all other and Legal Profession is has the lowest.

Accreditations vs CDR: Out of many different accreditations we have filters couple of them which are not relevant to our analysis such as Biblical associations, Theological associations, etc. It would be good to understand if CDR varies much based on these top associations that we have considered in our dataset.



Based on above graph we can conclude that all the associations have very comparable CDR ranges which varies from 0.0 to 0.3. Also, we could notice that New England commission and Western Association of Schools and Colleges Senior Colleges and University Commission gas very tight range between 0.0 to 0.15 which is very good in comparison to other associations.

Year over Year CD Distribution: CDR should technically decrease over the years due to new policies and restrictions colleges implements from time to time, they also keep on finding the way by which values could be reduced. Our assumption is average value should decreases as we progress.



Looking at above graph we could see that CDR value ranges does change over the year, but the effect seems to be very insignificant. In year 2010 CDR values ranges from 0.0 to 0.3 which in year 2016 it ranges from 0.0 to 0.24 with an average value of 0.07 which is lower to that in year 2010.

Predictor Table

Predictor	Datatype	Effect	Rationale
NUMBRANCH	Categorical	+/-	We would like to analyze how the number of branches/campuses for an institute impact the CDR for an institute.
PREDDEG	Categorical	?	This variable shows the predominant kind of degree awarded by the institute.
CONTROL	Categorical	?	We would like to analyze how CDR varies with in Private and Public category Institutes.
REGION	Categorical	?	This variable includes the demographic factor of an Institute. We would like to analyze how CDR varies among the Regions.
UGDS_ASIAN UGDS_BLACK UGDS_HISP UGDS_WHITE	Numerical	+/-	Number of undergrad students with respect to Ethnicity. Ethnicity shouldn't affect the CDR rate. But as we know about America's history of systemic racism and red lining we can observe some bias.

TUITFTE	Numerical	+	Average Net tuition paid by full-time equivalent student, after deducting grands from full tuition fee.
INEXPFTE	Numerical	+	Average amount spend by the Institute on per full-time student.
CCUGPROF_C	Categorical	?	This variable shows the prominent full-time or part-time students in the Institute. We would like to analyze how CDR varies among these two groups.
PBI	Categorical	?	This indicates an Institute is a Predominantly Black Institute.
HIS	Categorical	?	This Indicates an Institute is a Hispanic-serving institution
top_course_1	Categorical	+/-	Major degree awarded by Institute with respect to the stream.
NUM_PELL	Numerical	-	This Indicates the number of Pell Grant awarded students in the Institute. As the Grant number increases CDR decreases.
NUM_FLOAN	Numerical	+	This Indicates the Number of students under the Federal Loan. As this increases the CDR might increase.
ENDOWBEGIN	Numerical	-	This Indicates the Endowments at the beginning of the fiscal year for an Institute. As the Endowments increases the CDR might decrease.
YEAR	Numerical	+/-	We want to analyze the trend of CDR over the years.
ACCREDCODE	Categorical	+/-	This indicates the Accreditation agency for an Institute. We would like to analyze how CDR varies with respect to Accreditation agencies.
UGDS	Numerical	0	Colleges with higher number of students may decrease the CDR as the denominator in CDR increases , this vale is highly correlated with ethnic student counts above.
NUM_MEN NUM_WOMEN	Numerical	+/-	Number of undergrad students with respect to gender. It shouldn't affect the CDR rate. But we want to check if there is any bias.
CDR	Numerical	Predictor	This is the predictor variable we are trying to inspect.
NUM_DEFAULT	Numerical	Predictor	This is the predictor variable, Number of students who defaulted within that institute in that particular year.

Models

With a handful of models, we looked at the CDR ratio, the primary model is a linear regression model using the Log of CDR as the predictor variable and the key effects described in the variable tables as predictors. We then looked at the data on a multi-level basis, with Institutions as level 1 and Accreditation Agencies as level 2. As a result, we have multi-level data with a few interaction variables. The interactions include the impact of public/private universities offering just 2/4 year courses on CDR rates, as well as the impact of a low/high number of branches within public/private schools on CDR ratios.

CDR Model – Linear Regression.

```
cdr_model0 = lmer(log(CDR3) ~ NUMBRANCH_c + PREDDEG_c + CONTROL + REGION +
UGDS_ASIAN + UGDS_BLACK + UGDS_HISP + UGDS_WHITE + log(TUITFTE) +
log(INEXPFTE) + CCUGPROF_c + PBI + HIS + top_course_1 + NUM_PELL +
NUM_FLOAN + log(ENDOWBEGIN) + YEAR +  CCUGPROF_c*CONTROL +
NUMBRANCH_c*CONTROL, REML=FALSE)
```

CDR Model – Multilevel model with Levels as Accreditation Agencies.

```
cdr_model1 = lmer(log(CDR3) ~ NUMBRANCH_c + PREDDEG_c + CONTROL + REGION +
UGDS_ASIAN + UGDS_BLACK + UGDS_HISP + UGDS_WHITE + log(TUITFTE) +
log(INEXPFTE) + CCUGPROF_c + PBI + HIS + top_course_1 + NUM_PELL +
NUM_FLOAN + log(ENDOWBEGIN) + YEAR +  CCUGPROF_c*CONTROL +
NUMBRANCH_c*CONTROL + (1|ACCREDCODE), REML=FALSE)
```

CDR Model – Multilevel model with Levels as Accreditation Agencies and Year.

```
cdr_model2 = lmer(log(CDR3) ~ NUMBRANCH_c + PREDDEG_c + CONTROL + REGION +
UGDS_ASIAN + UGDS_BLACK + UGDS_HISP + UGDS_WHITE + log(TUITFTE) +
log(INEXPFTE) + CCUGPROF_c + PBI + HIS + top_course_1 + NUM_PELL +
NUM_FLOAN + log(ENDOWBEGIN) +  CCUGPROF_c*CONTROL + NUMBRANCH_c*CONTROL +
(1|ACCREDCODE) + (1|YEAR), REML=FALSE)
```

(Please Refer Appendix for fixed and variable effects of this Model)

Marginal Effects:

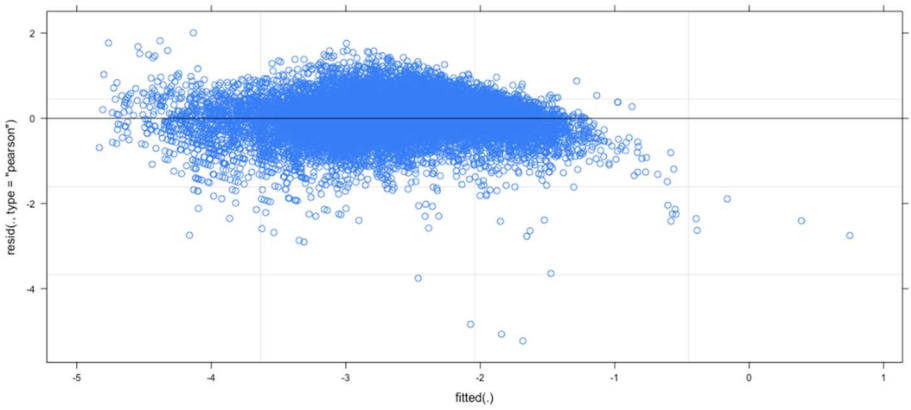
- Predominantly Graduate Degree offering Institutions have 100% less CDR when compared to Predominantly Under-Graduate Degree offering Institutions.
- Private for-profit and Private non-profit Institutions have 110%, 95% more CDR when compared to public Institutions.
- Institutions under Southwest (AZ, NM, OK, TX) region has higher CDR's approximately 30% more than Institutions in Far West (AK, CA, HI, NV, OR, WA), Plains (IA, KS, MN, MO, NE, ND, SD) which have low CDR's.
- Percentage of population of students under different Ethnicities have very very low or no impact on the CDR's.
- Institutions which are recognized as Predominantly Black Institutions have 2% less and Hispanic Serving Institutions have 5% more CDR's.
- If an Institute increases it's average spending on a student by 100%, the CDR decreases by 27%.

Quality Checks

LMER uses the MLE estimation technique (like GLM), which does not require normality or homoscedasticity assumption.

However, we still need to consider multicollinearity and independence. Estimates shown below (note: last column GVIF^(1/(2*Df)) is equivalent to VIF in OLS models) shows no multicollinearity. Durbin-

Watson test of independence is not defined for LMER models. We definitely had some correlated errors in the data, but hopefully, our multi-level specification takes out most of those errors.



Multi-colliniearity: We can see that there is littile multicolliniearity for CONTROL.

	GVIF	Df	GVIF^(1/(2*Df))
NUMBRANCH_c	77.162872	3	2.063327
PREDDEG_c	4.847610	3	1.300932
CONTROL	3711.678875	2	7.805354
REGION	3.330746	8	1.078099
UGDS_ASIAN	1.608014	1	1.268075
UGDS_BLACK	2.801822	1	1.673864
UGDS_HISP	3.911453	1	1.977739
UGDS_WHITE	4.526445	1	2.127544
log(TUITFTE)	4.895415	1	2.212559
log(INEXPFTE)	2.083180	1	1.443323
CCUGPROF_c	67.897586	4	1.694267
PBI	1.141650	1	1.068480
HSI	1.645872	1	1.282916
top_course_1	4.182393	32	1.022609
NUM_PELL	14.364129	1	3.790004
NUM_FLOAN	11.828336	1	3.439235
log(ENDOWBEGIN)	1.648986	1	1.284129
CONTROL:CCUGPROF_c	33291.233570	8	1.917104
NUMBRANCH_c:CONTROL	118.057896	6	1.488246

Recommendations

- Institutions should try increase their expenditure per full-time students, as this can marginally decrease the CDR rate.
- Institutions should try to concentrate on awarding more number of degrees(by increasing number of seats) in English Literature, Legal Profession, Engineering, sciences which can yeild lower CDR's. Colleges can provide additional support to students in majors with higher risk of future default and nonrepayment, or steer students into majors that lead to a higher likelihood of repaying loans.
- Found that compared to Institutes that award Certificates or Associates, Institutes that award Bachelor's and Graduate degree's have lower CDR's across all years.

References

1. Charles, K. D., Sheaff, S., Woods, J., & Downey, L. (2016). Decreasing Your Student Loan Cohort Default Rate: Leading a College-Wide Change Initiative at Mohave Community College. Community College Journal of Research and Practice, 40(7), 597–606. https://doi.org/10.1080/10668926.2015.1125814

2. Commisso, L. (2017). Identification of Best Practices to Assist Financial Leaders of Proprietary Institutions to Comply with Cohort Default Rate Requirements: A Phenomenological Study. ProQuest Dissertations Publishing.

3. Flint, Thomas A. (1994) "The Federal Student Loan Default Cohort: A Case Study," Journal of Student Financial Aid: Vol. 24 : Iss. 1 , Article 2. Available at : https://ir.library.louisville.edu/jsfa/vol24/iss1/2

4. Gross, Jacob P.K.; Cekic, Osman; Hossler, Don; and Hillman, Nick (2010) "What Matters in Student Loan Default: A Review of the Research Literature," Journal of Student Financial Aid: Vol. 39 : Iss. 1 , Article 2. Available at: https://ir.library.louisville.edu/jsfa/vol39/iss1/2

5. Hillman, N. W. (2013). Cohort default rates. Educational Policy, 29(4), 559–582. https://doi.org/10.1177/0895904813510772

6. Kelchen, R., & Li, A. Y. (2017). Institutional accountability: A comparison of the predictors of student loan repayment and default rates. The ANNALS of the American Academy of Political and Social Science, 671(1), 202–223. https://doi.org/10.1177/0002716217701681

7. Knapp, L. G., & Seaks, T. G. (1992). An analysis of the probability of default on federally GURANTEED student loans. The Review of Economics and Statistics, 74(3), 404. https://doi.org/10.2307/2109484

8. Looney, A., & Yannelis, C. (2019). How useful are default rates? borrowers with large balances and student loan repayment. Economics of Education Review, 71, 135–145. https://doi.org/10.1016/j.econedurev.2018.10.004

9. Lundgren, J. M. (2013). The effect of changing unemployment rates on student loan cohort default rates. ProQuest Dissertations Publishing. https://www.proquest.com/docview/1355174764?pq-origsite=primo&accountid=14745

10. Shen, H., & Ziderman, A. (2009). Student loans repayment and recovery: International comparisons. Higher Education, 57(3), 315–333. https://doi.org/10.1007/s10734-008-9146-0

Data Sources

The Integrated Postsecondary Education Data System

https://nces.ed.gov/ipeds/

The National Student Loan Data System (NSLDS®)

https://nsldsfap.ed.gov/nslds_FAP/

Appendix

Data Pre-Processing

```
#We used python for data pre-processing.gt

import pandas as pd
import numpy as np
data = pd.read_csv("path/MERGED2019_20_PP.csv")
columns_df = pd.read_excel("University_records_filtered.xlsx")  #Filtered the unwanted columns in one Excel file.
data_f_cols = columns_df.columns
data_14_15_f = data[data_f_cols]  #taking only those filtered columns on all datasets.
data_14_15_f.drop(['CCUGPROF', 'PBI', 'HSI', 'ACCREDCODE', 'ACCREDAGENCY',
                                        'HCM2', 'CURROPER', 'DOLPROVIDER'], inplace = True, axis = 1)
data_14_15_f = pd.merge(data_14_15_f, columns_df[['OPEID', 'CCUGPROF', 'PBI', 'HSI', 'ACCREDCODE',
'ACCREDAGENCY',  'HCM2']], on = 'OPEID', how = 'right')
# dropping unwanted columns
data_14_15_f = data_14_15_f.drop(['NUM41_PUB', 'NUM42_PUB', 'NUM43_PUB',
                        'NUM44_PUB', 'NUM45_PUB', 'NUM41_PRIV',
                        'NUM42_PRIV', 'NUM43_PRIV', 'NUM44_PRIV', 'NUM45_PRIV', 'ICLEVEL'], axis = 1)
df_sub=data_14_15_f.iloc[:,7:45]  #Fearute Engineering for top_course columns
columns=["Agriculture", "Natural Resources", "Architecture", "Culture studies", "Communication", "Communication2"
, "Computer sciences", "Personal services", "Education", "Engineering", "Engineering2", "Foreign literature", "Family
Sciences", "Legal Profession", "English Literature", "Liberal Arts", "Library Science", "Biomedical sciences",
"Mathematics", "Military", "Multi disciplinary studies", "Fitness Studies", "Philosophy", "Theology", "Physical
sciences", "Science", "Psychology", "Homeland security", "Public administration", "Social Sciences", "Construction
Trade", "Repair technologies", "Precision production", "Transportation", "Visual arts", "Health Profession", "Business",
"History"]
df_sub.columns=columns
df_sub.fillna(0) # filling null values with 0
df_sub.replace(np.nan,0)
df_sub['Engineering']=df_sub['Engineering']+df_sub['Engineering2'] #adding exclusive columns
df_sub['Communication']=df_sub['Communication']+df_sub['Communication2']
df_sub.drop(['Engineering2'],inplace=True,axis=1) #dropping the extra column after adding
df_sub.drop(['Communication2'],inplace=True,axis=1)
#adding the top_courses columns after after feature engineering.
df = df_sub.set_index(df_sub.index)
df_top = pd.DataFrame(df_sub.columns.values[np.argsort(-df_sub.values, axis=1)[:, :3]], index=df.index,
            columns = ['top_course_1','top_course_2','top_course_3']).reset_index()
df_top['OPEID']=data_14_15_f['OPEID']
data_14_15_ful=pd.merge(data_14_15_f,df_top, on='OPEID', how="inner")
#dropping the degree awarding percentage columns
data_14_15_ful = data_14_15_ful.drop(['PCIP01', 'PCIP03', 'PCIP04', 'PCIP05', 'PCIP09', 'PCIP10',
     'PCIP11', 'PCIP12', 'PCIP13', 'PCIP14', 'PCIP15', 'PCIP16', 'PCIP19',
     'PCIP22', 'PCIP23', 'PCIP24', 'PCIP25', 'PCIP26', 'PCIP27', 'PCIP29',
     'PCIP30', 'PCIP31', 'PCIP38', 'PCIP39', 'PCIP40', 'PCIP41', 'PCIP42',
     'PCIP43', 'PCIP44', 'PCIP45', 'PCIP46', 'PCIP47', 'PCIP48', 'PCIP49',
```

```
      'PCIP50', 'PCIP51', 'PCIP52', 'PCIP54', 'index',], axis = 1)
data_14_15_final = data_14_15_ful[data_14_15_ful['INSTNM'].notna()]
data_14_15_final[['UGDS_WHITE', 'UGDS_BLACK', 'UGDS_HISP', 'UGDS_ASIAN',
'UGDS']]=data_14_15_final[['UGDS_WHITE','UGDS_BLACK','UGDS_HISP','UGDS_ASIAN','UGDS']].fillna(0)
data_14_15_final['UGDS_WHITE'] = data_14_15_final['UGDS_WHITE']*data_14_15_final['UGDS']
data_14_15_final['UGDS_BLACK'] = data_14_15_final['UGDS_BLACK']*data_14_15_final['UGDS']
data_14_15_final['UGDS_HISP'] = data_14_15_final['UGDS_HISP']*data_14_15_final['UGDS']
data_14_15_final['UGDS_ASIAN'] = data_14_15_final['UGDS_ASIAN']*data_14_15_final['UGDS']
data_14_15_final[['UGDS_WHITE', 'UGDS_BLACK', 'UGDS_HISP', 'UGDS_ASIAN', 'UGDS']].round(0)
data_14_15_final = data_14_15_final[data_14_15_final['CDR3'].notna()]
list_w=["Barber", "Nursing", "Beauty", "Cosmetology", "Hair","Community",
      "Online",'Salon',"Spa","Profession",'Culinary','Art', 'Career', 'Center', 'Baptist', 'Memorial']
def remove_words(list_w,df):
   for i in list_w:
      df = df[~df['INSTNM'].str.contains(i)]
   return df
temp2=remove_words(list_w,temp)
temp2=temp2[~temp2["ACCREDCODE"].isin(['TRACS','COE','ACCSC','ACCET','DETC','NACCAS','MSACSS','NAST','ABHE','
NASM','AARTS','NYBRE','NLNAC'])]
temp2[['NPT4_PUB','NPT4_PRIV','PCTPELL','PCTFLOAN','UGDS_MEN','UGDS_WOMEN']]=temp2[['NPT4_PUB','NPT4_P
RIV','PCTPELL','PCTFLOAN','UGDS_MEN','UGDS_WOMEN']].fillna(0)
temp2['AVG_TUTION'] = temp2['NPT4_PUB'] + temp2['NPT4_PRIV']
temp2['NUM_PELL']=temp2['PCTPELL']*temp2['UGDS']
temp2['NUM_FLOAN']=temp2['PCTPELL']*temp2['UGDS']
temp2['NUM_PELL']=temp2['NUM_PELL'].round(0)
temp2['NUM_FLOAN']=temp2['NUM_FLOAN'].round(0)
temp2['NUM_MEN']=temp2['UGDS_MEN']*temp2['UGDS']
temp2['NUM_WOMEN']=temp2['UGDS_WOMEN']*temp2['UGDS']
temp2['NUM_MEN']=temp2['NUM_MEN'].round(0)
temp2['NUM_WOMEN']=temp2['NUM_WOMEN'].round(0)
temp2['NUM_DEFAULT']=temp2['CDR3']*temp2['CDR3_DENOM']
temp2['NUM_DEFAULT']=temp2['NUM_DEFAULT'].round(0)
temp2['YEAR']="2016"
temp2 = temp2[temp2['UGDS']>500]
temp2_final = temp2.drop(['NPT4_PUB', 'NPT4_PRIV', 'NUM4_PUB', 'NUM4_PRIV', 'PCTPELL', 'PCTFLOAN',
     'D_PCTPELL_PCTFLOAN', 'UGDS_MEN', 'UGDS_WOMEN'], axis = 1)
temp3 = temp2_final.drop_duplicates(subset=['OPEID'])
temp3.to_csv("path/cdr19_20_final2.csv")

#Concatinating the pre-processed data files into one single file.
data_13_14_f=pd.read_csv("path/cdr13_14_final2.csv")
data_14_15_f=pd.read_csv("path/cdr14_15_final2.csv")
data_15_16_f=pd.read_csv("path/cdr15_16_final2.csv")
data_16_17_f=pd.read_csv("path/cdr16_17_final2.csv")
data_17_18_f=pd.read_csv("path/cdr17_18_final2.csv")
data_18_19_f=pd.read_csv("path/cdr18_19_final2.csv")
data_19_20_f=pd.read_csv("path/cdr19_20_final2.csv")
```

```
cdr_full = pd.concat([data_13_14_f,data_14_15_f,
data_15_16_f,data_16_17_f,data_17_18_f,data_18_19_f,data_19_20_f], ignore_index=True)
cdr_full['REGION'] = cdr_full['REGION'].map({1 : "New England",
                                    2 : "Mid East",
                                    3 : "Great Lakes",
                                    4:"Plains",
                                    5:"Southeast",
                                    6:"Southwest",
                                    7:"Rocky Mountains",
                                    8:"Far West",
                                    9:"Outlying Areas"})
cdr_full['CONTROL'] = cdr_full['CONTROL'].map({1 : "PUBLIC", 2 : "PRIVATE NON PROFIT", 3 : "PRIVATE FOR PROFIT"})
cdr_full.to_csv("path/final_cdr_sdm.csv")
```

Data Visualizations

```
library(readxl)
library(ggplot2)
library(dplyr)
library(viridis)
library(stringr)
library(PerformanceAnalytics)

df = read.csv("final_cdr_sdm.csv")
View(df)
attach(df)
str(df)
colSums(is.na(df))

df$GRADS <- NULL
df$ENDOWBEGIN <- NULL
df$ENDOWEND <- NULL
df$ADM_RATE <- NULL

# dropping rows with missing values
df = df[complete.cases(df),]

df$NUMBRANCH = as.factor(df$NUMBRANCH)
table(df$PREDDEG)
df$PREDDEG = replace(df$PREDDEG, df$PREDDEG == '1','Predominantly certificate-degree granting')
df$PREDDEG = replace(df$PREDDEG, df$PREDDEG == '2',"Predominantly associate's-degree granting")
df$PREDDEG = replace(df$PREDDEG, df$PREDDEG == '3',"Predominantly bachelor's-degree granting")
df$PREDDEG = replace(df$PREDDEG, df$PREDDEG == '4',"Predominantly graduate-degree granting")
PREDDEGPREDDEG
df$PREDDEG = as.factor(df$PREDDEG)
df$CONTROL = as.factor(df$CONTROL)
```

```
df$REGION = as.factor(df$REGION)
df$ACCREDAGENCY = replace(df$ACCREDAGENCY, df$ACCREDAGENCY == '','Other')
df$ACCREDAGENCY = as.factor(df$ACCREDAGENCY)
df$top_course_1 = as.factor(df$top_course_1)
df$YEAR = as.factor(df$YEAR)
df$YDegree = as.factor(df$YDegree)
# Basic box plot
p1 <- ggplot(df, aes(x=(df$CDR3))) +
 geom_histogram(aes(y=..density..),
                colour = 'black',
                fill = 'lightblue') +
 geom_density(alpha=.2, fill="#FF6666")
p1 + labs(x = 'Credit Default Rate'
        ,y= 'Count'
        ,title = 'Distribution of Credit Default Rate')

p2 <- ggplot(df, aes(x=log(df$CDR3))) +
 geom_histogram(aes(y=..density..),
                colour = 'black',
                fill = 'lightblue') +
 geom_density(alpha=.2, fill="#FF6666") #+
p2 + labs(x = 'Log of CDR'
        ,y= 'Count'
        ,title = 'Distribution of log of CDR')

#Checking correlation
df_corr = cbind(df[9:20],df[23],df[27:32])
chart.Correlation(df_corr)

library(corrplot)
M<-cor(df_corr)
corrplot(M, method="number", type = "lower",number.cex=0.55)

p4 <- ggplot(df, aes(x=ACCREDAGENCY, y=CDR3, fill=ACCREDAGENCY)) +
 geom_boxplot(alpha=0.3, outlier.shape = NA) +
 scale_y_continuous(limits = c(0.0,0.4))+
 theme(legend.position="none") +
 scale_fill_brewer(palette="Dark2") +
 scale_x_discrete(labels = function(ACCREDAGENCY) str_wrap(ACCREDAGENCY, width = 10))
p4 + labs(x = 'Accredations'
        ,y= 'CDR'
        ,title = 'Accredations vs log of CDR')


P5 <- ggplot(df, aes(x=YEAR, y=CDR3, fill=YEAR)) +
 geom_boxplot(alpha=0.3, outlier.shape = NA) +
 scale_y_continuous(limits = c(0.0,0.4))+
```

```
  theme(legend.position="none") +
  scale_fill_brewer(palette="Dark2") +
  scale_x_discrete(labels = function(YEAR) str_wrap(YEAR, width = 10))
p5 + labs(x = 'Year'
        ,y= 'CDR'
        ,title = 'Year over Year CDR Distribution')
```

Data Modelling

```
rm(list=ls())
library(rio)
library(moments)
library(car)
library(stargazer)
library(lme4)
library(lmtest)

cdr_data=import('final_cdr_sdm.csv')

cdr_data$NUMBRANCH_c <- ifelse(cdr_data$NUMBRANCH <= 5,"Less than 5",
ifelse(cdr_data$NUMBRANCH > 5 & cdr_data$NUMBRANCH <= 10,"Between 5 and 10",
ifelse(cdr_data$NUMBRANCH > 10 & cdr_data$NUMBRANCH <= 15,"Between 10 and 15",
ifelse(cdr_data$NUMBRANCH > 15,"More than 15",0))))

cdr_data$PREDDEG_c <- ifelse(cdr_data$PREDDEG == 1,"Predominantly certificate-degree granting",
ifelse(cdr_data$PREDDEG == 2,"Predominantly associate's-degree granting",
ifelse(cdr_data$PREDDEG == 3,"Predominantly bachelor's-degree granting",
ifelse(cdr_data$PREDDEG == 4,"Predominantly graduate-degree granting",0))))

cdr_data$CCUGPROF_c <- ifelse(cdr_data$CCUGPROF == 1 |cdr_data$CCUGPROF == 2,"Two year - part time",
ifelse(cdr_data$CCUGPROF == 3 |cdr_data$CCUGPROF == 4,"Two year - full time",
ifelse(cdr_data$CCUGPROF == 5 ,"Four year - part time",
ifelse(cdr_data$CCUGPROF > 5 ,"Four year - full time","Not applicable"))))
cdr_data$PREDDEG_c=as.factor(cdr_data$PREDDEG_c)
cdr_data$PREDDEG_c = relevel(cdr_data$PREDDEG_c, "Predominantly certificate-degree granting")
cdr_data$CONTROL=as.factor(cdr_data$CONTROL)
cdr_data$CONTROL = relevel(cdr_data$CONTROL, "PUBLIC")
cdr_data$REGION=as.factor(cdr_data$REGION)
##cdr_data$CCUGPROF=as.factor(cdr_data$CCUGPROF)
cdr_data$PBI=as.factor(cdr_data$PBI)
cdr_data$HSI=as.factor(cdr_data$HSI)
##cdr_data$HCM2=as.factor(cdr_data$HCM2)
cdr_data$top_course_1=as.factor(cdr_data$top_course_1)
cdr_data$CCUGPROF_c=as.factor(cdr_data$CCUGPROF_c)
cdr_data$CCUGPROF_c = relevel(cdr_data$CCUGPROF_c, "Not applicable")
cdr_data$YEAR=as.factor(cdr_data$YEAR)
```

```
cdr_data$ACCREDCODE=as.factor(cdr_data$ACCREDCODE)
cdr_data$NUMBRANCH_c=as.factor((cdr_data$NUMBRANCH_c))
cdr_data$NUMBRANCH_c = relevel(cdr_data$NUMBRANCH_c, "Less than 5")
colSums(is.na(cdr_data))
cdr_data$ENDOWBEGIN[is.na(cdr_data$ENDOWBEGIN)] <- mean(cdr_data$ENDOWBEGIN,
na.rm = TRUE)
str(cdr_data$ENDOWBEGIN)
str(cdr_data$INEXPFTE)
cdr_data = subset(cdr_data, cdr_data$CDR3 != 0)
cdr_data$INEXPFTE=cdr_data$INEXPFTE+1
cdr_data$ENDOWBEGIN=cdr_data$ENDOWBEGIN+1
hist(CDR3)
hist(log(CDR3))
hist(NUM_DEFAULT)
hist(log(NUM_DEFAULT))
boxplot(cdr_data$CDR3~cdr_data$YEAR, cdr_data,outline=TRUE,main = "CDR across years",
xlab = "Year", ylab = "CDR value",
names = c(levels(cdr_data$YEAR)))
attach(cdr_data)
cdr_model1=lm(log(CDR3) ~ NUMBRANCH_c+PREDDEG_c+CONTROL+REGION
+UGDS
+log(TUITFTE)+log(INEXPFTE)
+CCUGPROF_c+PBI+HSI+NUM_PELL+NUM_FLOAN+log(ENDOWBEGIN)+YEAR)
cdr_model2=lm(log(CDR3) ~ NUMBRANCH_c+PREDDEG_c+CONTROL+REGION
+UGDS_ASIAN+UGDS_BLACK+UGDS_HISP+UGDS_WHITE
+log(TUITFTE)+log(INEXPFTE)
+CCUGPROF_c+PBI+HSI+top_course_1
+NUM_PELL+NUM_FLOAN+log(ENDOWBEGIN)
+ CCUGPROF_c*CONTROL+NUMBRANCH_c*CONTROL
+ ACCREDCODE+YEAR)
cdr_model3=lmer(log(CDR3) ~ NUMBRANCH_c+PREDDEG_c+CONTROL+REGION
+UGDS_ASIAN+UGDS_BLACK+UGDS_HISP+UGDS_WHITE
+log(TUITFTE)+log(INEXPFTE)
+CCUGPROF_c+PBI+HSI+top_course_1
+NUM_PELL+NUM_FLOAN+log(ENDOWBEGIN)
+ CCUGPROF_c*CONTROL+NUMBRANCH_c*CONTROL
+ (1|ACCREDCODE)+(1|YEAR),data = cdr_data, REML=FALSE)
summary((cdr_model3))
stargazer(cdr_model3,type="text", single.row=TRUE)
ranef(cdr_model3)
vif(cdr_model3)
plot(cdr_model3)
dwtest(cdr_model3)
```

```
> stargazer(cdr_model3,type="text", single.row=TRUE)

================================================================================================
                                                                  Dependent variable:
                                                              ---------------------------
                                                                       log(CDR3)
------------------------------------------------------------------------------------------------
NUMBRANCH_cBetween 10 and 15                                        -0.050 (0.138)
NUMBRANCH_cBetween 5 and 10                                          0.011 (0.043)
NUMBRANCH_cMore than 15                                            -0.274*** (0.051)
PREDDEG_cPredominantly associate's-degree granting                 -0.059*** (0.020)
PREDDEG_cPredominantly bachelor's-degree granting                  -0.302*** (0.027)
PREDDEG_cPredominantly graduate-degree granting                    -1.548*** (0.294)
CONTROLPRIVATE FOR PROFIT                                          1.158*** (0.112)
CONTROLPRIVATE NON PROFIT                                          0.951*** (0.103)
REGIONGreat Lakes                                                  0.132*** (0.038)
REGIONMid East                                                     0.184*** (0.054)
REGIONNew England                                                   0.230** (0.093)
REGIONOutlying Areas                                               0.176*** (0.063)
REGIONPlains                                                        -0.001 (0.039)
REGIONRocky Mountains                                               0.076** (0.033)
REGIONSoutheast                                                    0.292*** (0.039)
REGIONSouthwest                                                    0.305*** (0.038)
UGDS_ASIAN                                                       -0.00003*** (0.00001)
UGDS_BLACK                                                       0.00005*** (0.00001)
UGDS_HISP                                                        -0.00004*** (0.00000)
UGDS_WHITE                                                       -0.00003*** (0.00000)
log(TUITFTE)                                                       -0.250*** (0.012)
log(INEXPFTE)                                                      -0.273*** (0.010)
CCUGPROF_cFour year - full time                                    0.649*** (0.070)
CCUGPROF_cFour year - part time                                    0.625*** (0.073)
CCUGPROF_cTwo year - full time                                     0.671*** (0.092)
CCUGPROF_cTwo year - part time                                     0.690*** (0.072)
PBI1                                                                -0.024 (0.032)
HSI1                                                               0.052*** (0.016)
top_course_1Architecture                                            0.525** (0.231)
top_course_1Biomedical sciences                                     0.158** (0.074)
top_course_1Business                                               0.169*** (0.065)
top_course_1Communication                                           0.131 (0.092)
top_course_1Computer sciences                                       0.063 (0.096)
top_course_1Construction Trade                                      0.064 (0.137)
top_course_1Culture studies                                         0.340 (0.498)
top_course_1Education                                               0.112 (0.071)
top_course_1Engineering                                            -0.005 (0.068)
top_course_1English Literature                                     -0.583*** (0.152)
top_course_1Family Sciences                                         0.194 (0.129)
top_course_1Fitness Studies                                        0.356*** (0.091)
top_course_1Foreign literature                                      0.245 (0.212)
top_course_1Health Profession                                      0.167** (0.065)
top_course_1Homeland security                                      0.382*** (0.074)
top_course_1Legal Profession                                       -0.212 (0.293)
top_course_1Liberal Arts                                            0.085 (0.066)
top_course_1Military                                                0.351 (0.293)
top_course_1Multi disciplinary studies                              0.011 (0.083)
top_course_1Natural Resources                                       0.008 (0.100)
top_course_1Personal services                                      0.445*** (0.097)
top_course_1Philosopy                                               0.010 (0.152)
top_course_1Physical sciences                                      -0.089 (0.163)
top_course_1Precision production                                    0.081 (0.158)
top_course_1Psychology                                              0.122 (0.080)
top_course_1Public administartion                                  0.393*** (0.100)
top_course_1Repair technologies                                     0.220** (0.086)
```

```
top_course_1Science                                                  -0.158 (0.356)
top_course_1Social Sciences                                         -0.203*** (0.068)
top_course_1Theology                                                 0.172** (0.079)
top_course_1Transportation                                           0.186 (0.122)
top_course_1Visual arts                                             0.288*** (0.073)
NUM_PELL                                                          0.00003*** (0.00001)
NUM_FLOAN                                                          0.00001* (0.00000)
log(ENDOWBEGIN)                                                     -0.055*** (0.002)
CONTROLPRIVATE FOR PROFIT:CCUGPROF_cFour year - full time            -1.014*** (0.111)
CONTROLPRIVATE NON PROFIT:CCUGPROF_cFour year - full time            -1.057*** (0.103)
CONTROLPRIVATE FOR PROFIT:CCUGPROF_cFour year - part time            -1.050*** (0.116)
CONTROLPRIVATE NON PROFIT:CCUGPROF_cFour year - part time            -1.082*** (0.107)
CONTROLPRIVATE FOR PROFIT:CCUGPROF_cTwo year - full time             -0.919*** (0.125)
CONTROLPRIVATE NON PROFIT:CCUGPROF_cTwo year - full time             -0.886*** (0.137)
CONTROLPRIVATE FOR PROFIT:CCUGPROF_cTwo year - part time              -0.276 (0.209)
CONTROLPRIVATE NON PROFIT:CCUGPROF_cTwo year - part time             -0.929*** (0.304)
NUMBRANCH_cBetween 10 and 15:CONTROLPRIVATE FOR PROFIT                -0.270* (0.147)
NUMBRANCH_cBetween 5 and 10:CONTROLPRIVATE FOR PROFIT                -0.176** (0.082)
NUMBRANCH_cMore than 15:CONTROLPRIVATE FOR PROFIT                    0.600*** (0.066)
NUMBRANCH_cBetween 10 and 15:CONTROLPRIVATE NON PROFIT               0.618*** (0.186)
NUMBRANCH_cBetween 5 and 10:CONTROLPRIVATE NON PROFIT                0.350*** (0.109)
NUMBRANCH_cMore than 15:CONTROLPRIVATE NON PROFIT                     0.429** (0.174)
Constant                                                            2.305*** (0.148)
-------------------------------------------------------------------------------------
Observations                                                              14,166
Log Likelihood                                                         -10,124.280
Akaike Inf. Crit.                                                      20,410.560
Bayesian Inf. Crit.                                                    21,022.800
=====================================================================================
Note:                                                    *p<0.1; **p<0.05; ***p<0.01

> ranef(cdr_model3)
$ACCREDCODE
        (Intercept)
        0.033118412
ABHES  -0.004915593
ACICS  -0.118330850
MSACHE -0.085336322
NCACHE  0.092296405
NECHE  -0.194717890
NWCCU   0.092426343
SACSCC  0.034242475
WASCJC  0.248488494
WASCSR -0.097271473

$YEAR
      (Intercept)
2010  0.130920777
2011 -0.020151433
2012 -0.031725743
2013  0.007355273
2014 -0.015867457
2015 -0.057243399
2016 -0.013288018

with conditional variances for "ACCREDCODE" "YEAR"
```

> summary((cdr_model3))

Linear mixed model fit by maximum likelihood ['lmerMod']

Formula: log(CDR3) ~ NUMBRANCH_c + PREDDEG_c + CONTROL + REGION + UGDS_ASIAN +
UGDS_BLACK + UGDS_HISP + UGDS_WHITE + log(TUITFTE) + log(INEXPFTE) +
CCUGPROF_c + PBI + HSI + top_course_1 + NUM_PELL + NUM_FLOAN +

log(ENDOWBEGIN) + CCUGPROF_c * CONTROL + NUMBRANCH_c * CONTROL +
(1 | ACCREDCODE) + (1 | YEAR)
Data: cdr_data
AIC BIC logLik deviance df.resid
20410.6 21022.8 -10124.3 20248.6 14085
Scaled residuals:
Min 1Q Median 3Q Max
-10.5874 -0.5398 0.0657 0.6099 4.0612
Random effects:
Groups Name Variance Std.Dev.
ACCREDCODE (Intercept) 0.01614 0.12706
YEAR (Intercept) 0.00369 0.06074
Residual 0.24348 0.49344
Number of obs: 14166, groups: ACCREDCODE, 10; YEAR, 7
Fixed effects:
Estimate Std. Error t value
(Intercept) 2.305e+00 1.483e-01 15.543
NUMBRANCH_cBetween 10 and 15 -5.016e-02 1.383e-01 -0.363
NUMBRANCH_cBetween 5 and 10 1.128e-02 4.275e-02 0.264
NUMBRANCH_cMore than 15 -2.743e-01 5.061e-02 -5.420
PREDDEG_cPredominantly associate's-degree granting -5.857e-02 2.037e-02 -2.875
PREDDEG_cPredominantly bachelor's-degree granting -3.024e-01 2.678e-02 -11.290
PREDDEG_cPredominantly graduate-degree granting -1.548e+00 2.940e-01 -5.267
CONTROLPRIVATE FOR PROFIT 1.158e+00 1.124e-01 10.300
CONTROLPRIVATE NON PROFIT 9.509e-01 1.029e-01 9.239
REGIONGreat Lakes 1.319e-01 3.818e-02 3.454
REGIONMid East 1.842e-01 5.409e-02 3.406
REGIONNew England 2.303e-01 9.289e-02 2.480
REGIONOutlying Areas 1.761e-01 6.255e-02 2.816
REGIONPlains -1.251e-03 3.932e-02 -0.032
REGIONRocky Mountains 7.599e-02 3.328e-02 2.284
REGIONSoutheast 2.923e-01 3.907e-02 7.481
REGIONSouthwest 3.055e-01 3.829e-02 7.977
UGDS_ASIAN -2.957e-05 5.606e-06 -5.275
UGDS_BLACK 4.708e-05 5.737e-06 8.205
UGDS_HISP -4.301e-05 3.602e-06 -11.940
UGDS_WHITE -3.016e-05 2.109e-06 -14.298
log(TUITFTE) -2.499e-01 1.150e-02 -21.729
log(INEXPFTE) -2.725e-01 1.037e-02 -26.276
CCUGPROF_cFour year - full time 6.489e-01 6.991e-02 9.282
CCUGPROF_cFour year - part time 6.255e-01 7.275e-02 8.598
CCUGPROF_cTwo year - full time 6.710e-01 9.157e-02 7.328
CCUGPROF_cTwo year - part time 6.900e-01 7.220e-02 9.557
PBI1 -2.404e-02 3.182e-02 -0.755
HSI1 5.192e-02 1.567e-02 3.313
top_course_1Architecture 5.255e-01 2.309e-01 2.276
top_course_1Biomedical sciences 1.585e-01 7.373e-02 2.149
top_course_1Business 1.691e-01 6.541e-02 2.585
top_course_1Communication 1.310e-01 9.171e-02 1.428
top_course_1Computer sciences 6.299e-02 9.637e-02 0.654
top_course_1Construction Trade 6.371e-02 1.374e-01 0.464
top_course_1Culture studies 3.400e-01 4.984e-01 0.682

top_course_1Education 1.123e-01 7.093e-02 1.584
top_course_1Engineering -4.646e-03 6.834e-02 -0.068
top_course_1English Literature -5.827e-01 1.519e-01 -3.835
top_course_1Family Sciences 1.939e-01 1.289e-01 1.505
top_course_1Fitness Studies 3.561e-01 9.082e-02 3.921
top_course_1Foreign literature 2.451e-01 2.120e-01 1.156
top_course_1Health Profession 1.667e-01 6.518e-02 2.557
top_course_1Homeland security 3.816e-01 7.418e-02 5.144
top_course_1Legal Profession -2.119e-01 2.927e-01 -0.724
top_course_1Liberal Arts 8.468e-02 6.602e-02 1.283
top_course_1Military 3.511e-01 2.928e-01 1.199
top_course_1Multi disciplinary studies 1.093e-02 8.260e-02 0.132
top_course_1Natural Resources 7.576e-03 9.997e-02 0.076
top_course_1Personal services 4.454e-01 9.680e-02 4.601
top_course_1Philosopy 1.021e-02 1.517e-01 0.067
top_course_1Physical sciences -8.905e-02 1.627e-01 -0.547
top_course_1Precision production 8.083e-02 1.582e-01 0.511
top_course_1Psychology 1.221e-01 7.989e-02 1.528
top_course_1Public administartion 3.927e-01 1.004e-01 3.913
top_course_1Repair technologies 2.198e-01 8.552e-02 2.570
top_course_1Science -1.576e-01 3.564e-01 -0.442
top_course_1Social Sciences -2.033e-01 6.790e-02 -2.994
top_course_1Theology 1.715e-01 7.870e-02 2.180
top_course_1Transportation 1.860e-01 1.220e-01 1.525
top_course_1Visual arts 2.877e-01 7.325e-02 3.928
NUM_PELL 3.095e-05 5.290e-06 5.850
NUM_FLOAN 6.905e-06 4.167e-06 1.657
log(ENDOWBEGIN) -5.545e-02 2.480e-03 -22.362
CONTROLPRIVATE FOR PROFIT:CCUGPROF_cFour year - full time -1.014e+00 1.108e-01 -9.153
CONTROLPRIVATE NON PROFIT:CCUGPROF_cFour year - full time -1.057e+00 1.027e-01 -10.287
CONTROLPRIVATE FOR PROFIT:CCUGPROF_cFour year - part time -1.050e+00 1.159e-01 -9.060
CONTROLPRIVATE NON PROFIT:CCUGPROF_cFour year - part time -1.082e+00 1.070e-01 -10.108
CONTROLPRIVATE FOR PROFIT:CCUGPROF_cTwo year - full time -9.190e-01 1.252e-01 -7.341
CONTROLPRIVATE NON PROFIT:CCUGPROF_cTwo year - full time -8.859e-01 1.374e-01 -6.447
CONTROLPRIVATE FOR PROFIT:CCUGPROF_cTwo year - part time -2.760e-01 2.087e-01 -1.322
CONTROLPRIVATE NON PROFIT:CCUGPROF_cTwo year - part time -9.292e-01 3.037e-01 -3.060
NUMBRANCH_cBetween 10 and 15:CONTROLPRIVATE FOR PROFIT -2.699e-01 1.474e-01 -1.830
NUMBRANCH_cBetween 5 and 10:CONTROLPRIVATE FOR PROFIT -1.760e-01 8.227e-02 -2.140
NUMBRANCH_cMore than 15:CONTROLPRIVATE FOR PROFIT 6.000e-01 6.646e-02 9.028
NUMBRANCH_cBetween 10 and 15:CONTROLPRIVATE NON PROFIT 6.180e-01 1.859e-01 3.324
NUMBRANCH_cBetween 5 and 10:CONTROLPRIVATE NON PROFIT 3.503e-01 1.089e-01 3.215
NUMBRANCH_cMore than 15:CONTROLPRIVATE NON PROFIT 4.293e-01 1.742e-01 2.465
Correlation matrix not shown by default, as p = 78 > 12.
Use print(x, correlation=TRUE) or
vcov(x) if you need it
fit warnings:
Some predictor variables are on very different scales: consider rescaling