Business Quantitative Techniques Assessment

Pradyumna Agrawal and Louise Higson

Replication of Lindo, Jason M., Nicholas J. Sanders, and Philip Oreopoulos. 2010. "Ability, Gender, and Performance Standards: Evidence from Academic Probation."

American Economic Journal: Applied Economics, 2 (2): 95-117.DOI: 10.1257/app.2.2.95

Overview

Universities, to ensure all students meet a minimum academic standard, implement a minimum Grade Point Average (GPA) that, if not met, results in the student being placed on academic probation. The student must then increase their subsequent year and average GPA to meet this minimum, otherwise risk academic suspension.
The above paper uses data of students academic standing over the period of their studies to assess the effect of placing students who fail to meet a minimum academic standard on probation.

The data includes a number of student characteristics (e.g. age, sex, high school grades, first language, birthplace) which are assessed to determine if any particular characteristic shows an increase in probability of a decision to leave or continue studying.

The regression discontinuity model is used to perform the analysis as it can assess if there is a significant difference in the outcomes between the treatment group (those placed on academic suspension) and the control group (those not placed on academic suspension).

The paper selected suggested that when students are placed on academic probation it can result in students not returning to university in the following year, and for those students who do return, an increase in their performance. When the observable characteristics are assessed individually the results are consistent in that it also shows the probability of an increase in performance.

The descriptive analyses replicated from the study are:
- Statistical data of the observational data

The inferential analyses replicated from the study are:
- discontinuity of the frequency of students on probation in the first year. This analysis can identify if there is any grouping of data suggesting a non random sorting across the threshold.
- assessment of the student characteristics to check for discontinuity between the treatment and control group. A discontinuity between these two groups suggests a particular characteristic may influence a students grade.
- how being placed on academic probation impacts the subsequent years GPA (increase, decrease or no effect).

A number of figures associated with each of these analysis will also be replicated.
The original report does not contain an assessment of linear regression assumption therefore these have been included with the assessment of discontinuity assumptions.

The free roam section of this report looks at some of the characteristics not assessed as part of the original report to determine if if there is a difference in response to probation based on these characteristics, and if these characteristics further support or contradict the study findings.

Most data cleaning had been performed by the original authors (removing students with missing data and those who did not have the academic standard assessed at the end of the first year). It was noted that there are discrepancies in the data range reported in the original study and the data available for review and replication. The report states academic standing data was collected in the years 1995 to 2004 however review of the reported first year confirmed the range of data supplied is 1996 to 2003.

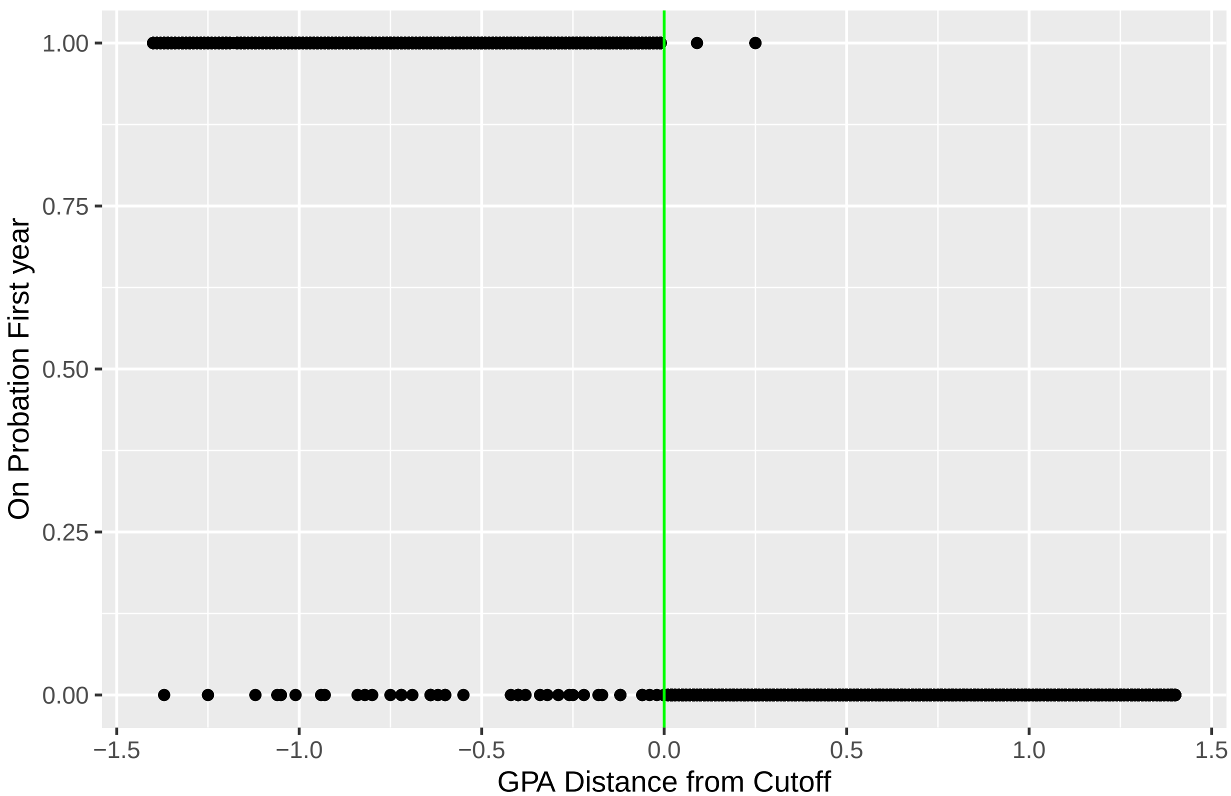
Table 1: First Year Range

	data_main (N = 44,362)
min	1996
max	2003

This has resulted in the inability replicate the results exactly as per the original report.

It was also noted that the original data had several points where students who did not meet the minimum GPA were not recorded as being placed on academic probation.

Figure 1: Distribution of Data Around the Cutoff



This data has been left in the analysis as we cannot confirm the accuracy of this data and therefore do not have sufficient justification to remove it.

Descriptive Replication

The data provided is based on GPA results from three different university campuses, one of which has a different cutoff GPA from the others. To allow for the use of all data based on the same cutoff point the original data includes an output variable of the students GPA distance from the specific campuses GPA cutoff. Using this value the cutoff point is set as 0. The data provided spans a bandwidth of 1.4 GPA points above and below the probation cutoff (0).

Table 2 is the descriptive statistics for the data and replicates Table 1 in the original study[1]. Results are similar to the original report descriptive statistics with male students representing 38% of the sample and 87% of students born in North America.

Table 2: Descriptive Summary

	data1 (N = 29,601)
High School Grade Percentile	40.13 ± 25.39
Credits Attempted in First Year	4.49 ± 0.52
Age at entry	18.70 ± 0.74
male	0.38 ± 0.48
English is First Language	0.72 ± 0.45
Born in North America	0.87 ± 0.33
At Campus 1	0.53 ± 0.50
At Campus 2	0.20 ± 0.40
At Campus 3	0.28 ± 0.45
Distance from Cutoff in 1st Year	0.48 ± 0.65
On Probation After 1st Year	0.23 ± 0.42
Ever on Academic Probation	0.28 ± 0.45
Left University after 1st Evaluation	0.05 ± 0.22
Distance from Cutoff Next Evaluation	0.71 ± 0.84
Ever Suspended	0.11 ± 0.32
Graduated by Year 4	0.37 ± 0.48
Graduated by Year 5	0.62 ± 0.49
Graduated by Year 6	0.71 ± 0.45

Table 3: Descriptive Summary

	data1 (N = 29,601)
High School Grade Percentile	40.13 ± 25.39
Credits Attempted in First Year	4.49 ± 0.52
Age at entry	18.70 ± 0.74
male	0.38 ± 0.48
English is First Language	0.72 ± 0.45
Born in North America	0.87 ± 0.33
At Campus 1	0.53 ± 0.50
At Campus 2	0.20 ± 0.40
At Campus 3	0.28 ± 0.45
Distance from Cutoff in 1st Year	0.48 ± 0.65
On Probation After 1st Year	0.23 ± 0.42
Ever on Academic Probation	0.28 ± 0.45
Left University after 1st Evaluation	0.05 ± 0.22
Distance from Cutoff Next Evaluation	0.71 ± 0.84
Ever Suspended	0.11 ± 0.32
Graduated by Year 4	0.37 ± 0.48
Graduated by Year 5	0.62 ± 0.49
Graduated by Year 6	0.71 ± 0.45

Lindo et al. (2008) Table 1 from original study[1]:

Table 1
Summary Statistics

	Mean	Standard Deviation
Characteristics		
High School Grade Percentile	40.12	25.36
Credits Attempted in First Year	4.49	0.52
Age at Entry	18.71	0.73
Male	0.38	0.48
English is First Language	0.73	0.45
Born in North America	0.88	0.33
At Campus 1	0.53	0.50
At Campus 2	0.20	0.40
At Campus 3	0.27	0.45
Outcomes		
Distance from Cutoff in 1st Year	0.48	0.65
On Probation After 1st Year	0.23	0.42
Ever on Academic Probation	0.28	0.45
Left University after 1st Evaluation	0.05	0.22
Distance from Cutoff Next Evaluation	0.71	0.84
Ever Suspended	0.12	0.32
Graduated by Year 4	0.37	0.48
Graduated by Year 5	0.62	0.49
Graduated by Year 6	0.71	0.46

Note: The sample consists of the 32,697 students within 1.4 grade points of the cutoff in their first year.

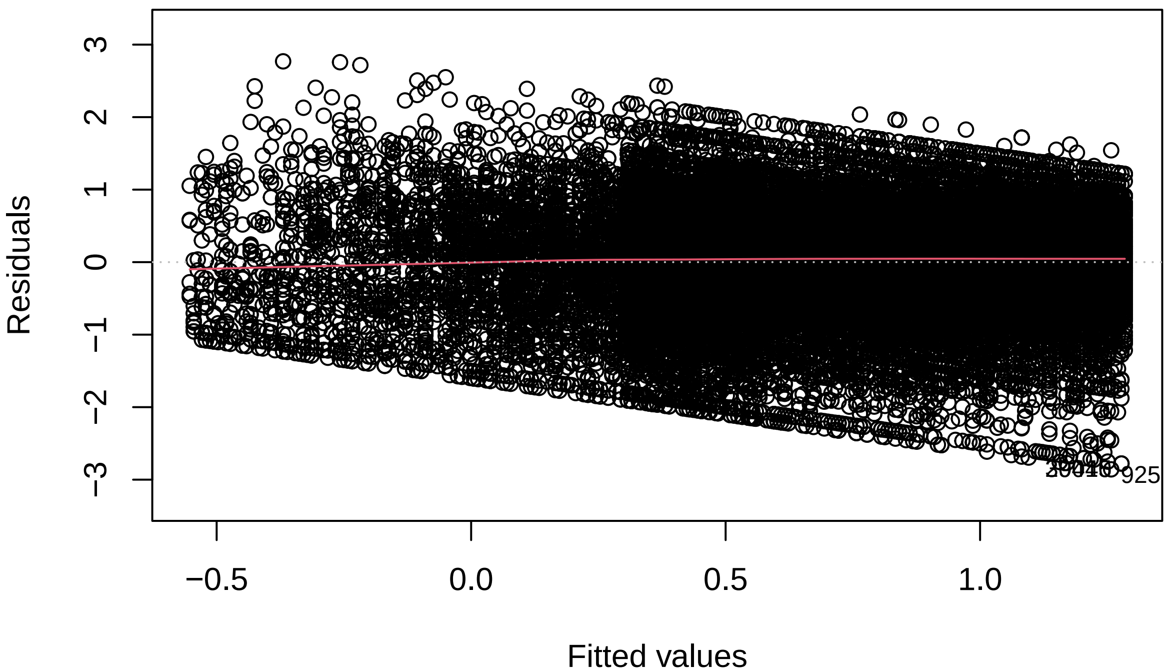
Inferential Replication

Testing for Regression Analysis Assumptions (Extention)

Lindo et al. (2008) does not include any specifc tests for assumptions of regression analysis. In order to use a Linear Regression model the following assumptions have to be met:
1. There is a linear relationship between the X and Y.
2. The variance of the residuals are constant for all values of X.
3. The error (residuals) of one observation is independent of the error of another.
4. For statistical inference, the residuals are normally distributed.

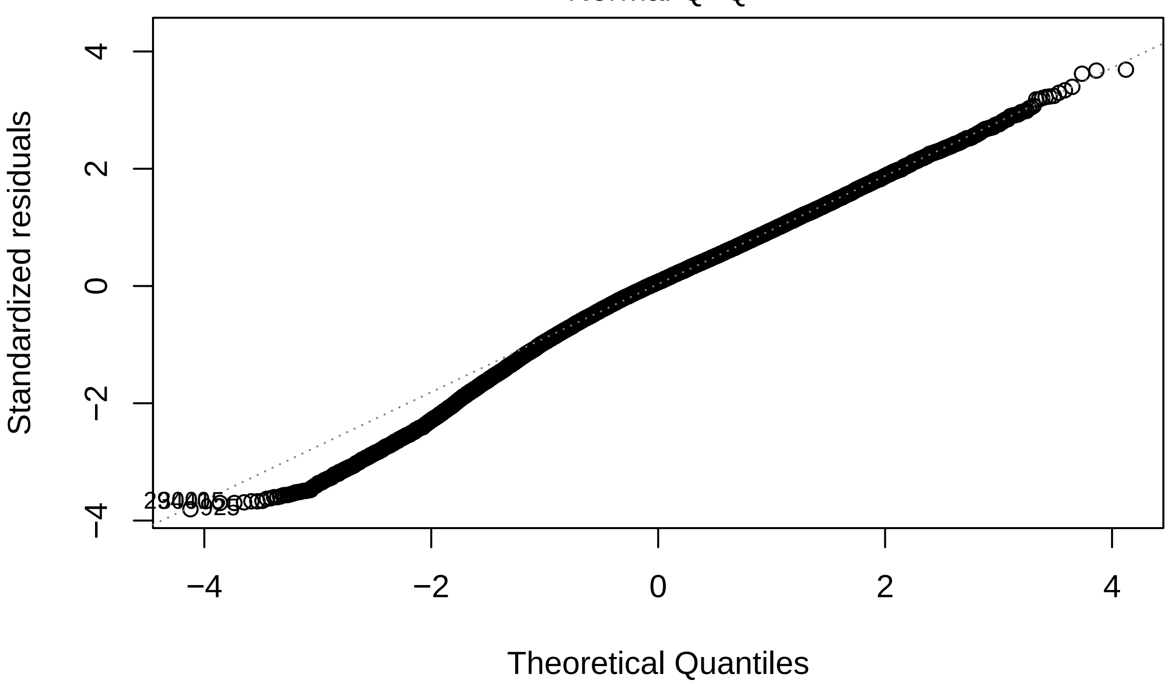
As the assessment we are replicating uses a linear regression model to assess the impact of being put on probation to a students next GPA result, this data was assessed to confirm that it meets the above assumptions.

Residuals vs Fitted

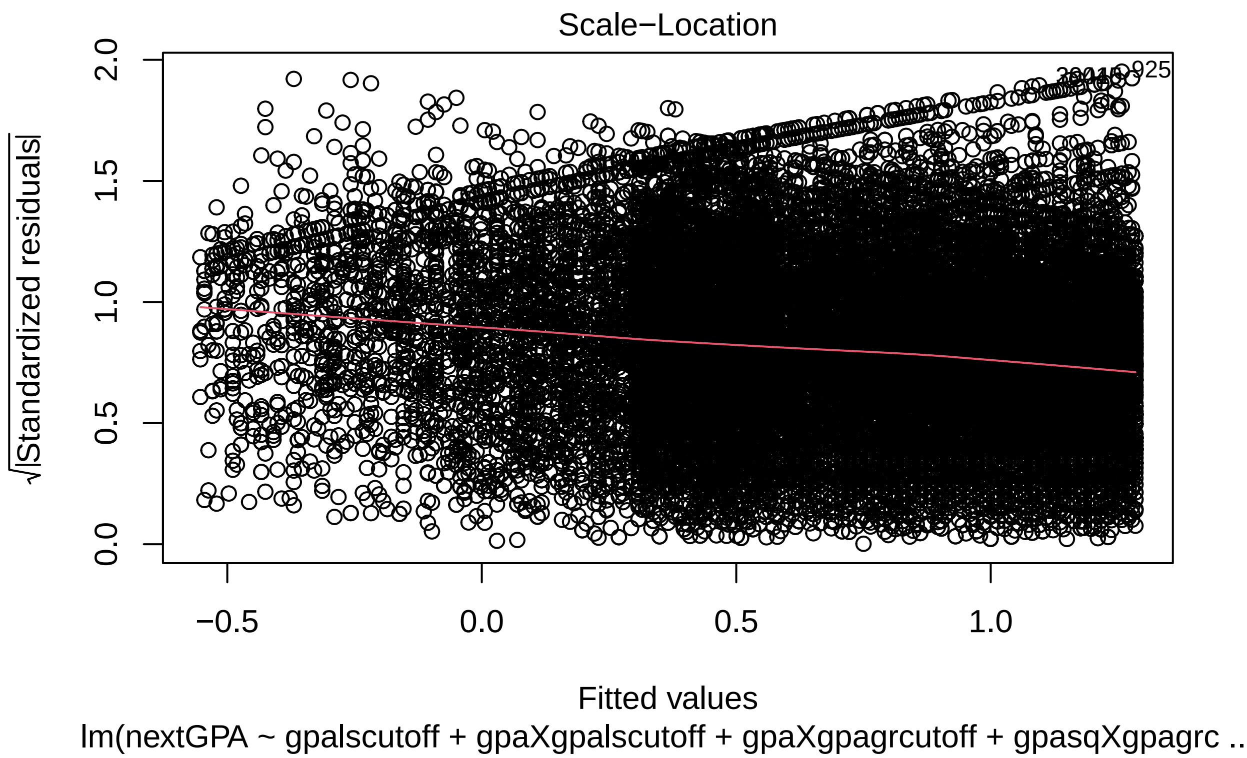


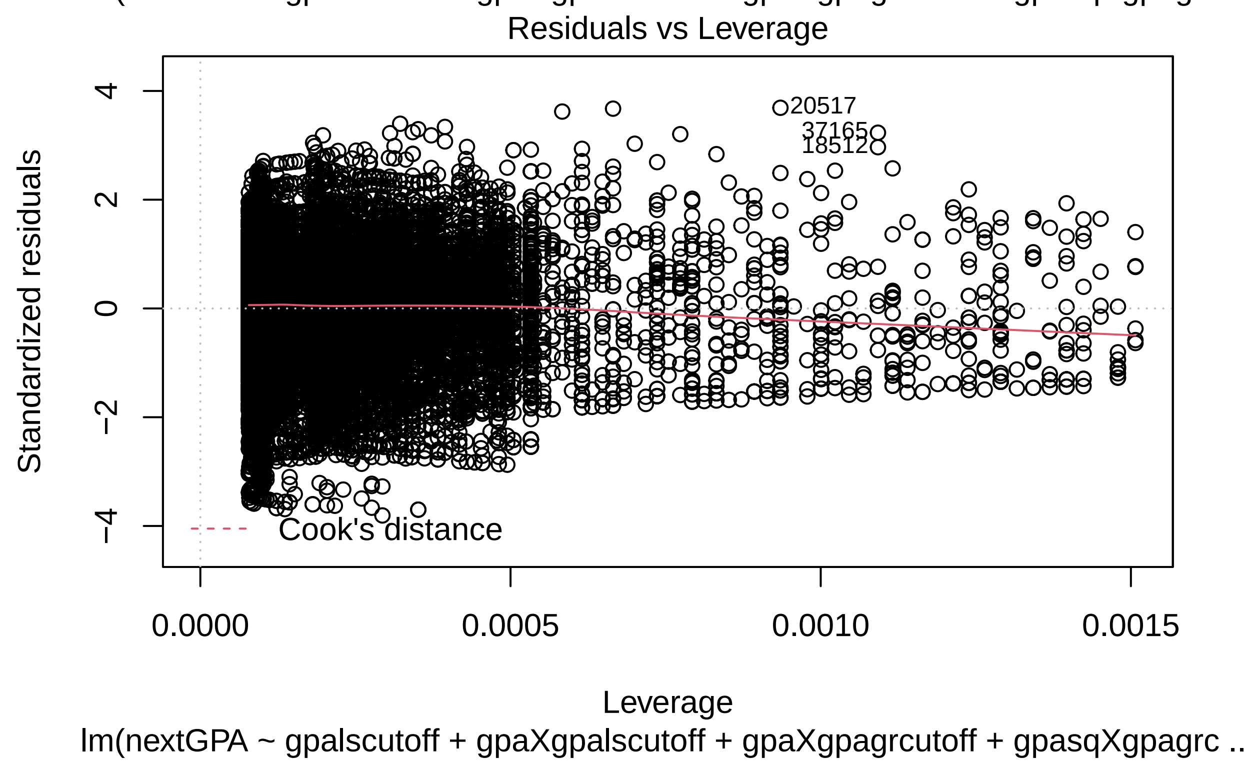
lm(nextGPA ~ gpalscutoff + gpaXgpalscutoff + gpaXgpagrcutoff + gpasqXgpagrc ...

Normal Q−Q



lm(nextGPA ~ gpalscutoff + gpaXgpalscutoff + gpaXgpagrcutoff + gpasqXgpagrc ...





The residuals vs fitted values plot shows the data is distributed above and below the line however is noted that there is a decreasing trend in the data where more points are below the line as the X variable increases. The scale location plot suggest the residuals of the data do not have constant variance as the width of the data spread appears to shift as X increases. The reason is that the actual GPA has a fixed range. sum of a fitted value with the corresponding residual is equal to the actual GPA. Now for a specific fitted value, the range of residual would be equal to the range of GPA minus the fitted value. It means that residuals have a

dependency on fitted values. Therefore, this validy check violates the assumptions of linear regression model. It is a case of heteroscedasticity. It is an issue for regression analysis as it annuls tests of significance which assume equal variance of residuals. GPA is a discrete variable with fixed number of values possible. For each value, standard errors have a different range. Hence, it can be concluded that errors are clustered at GPA. In the inference section, Clustered Covariance Matrix Estimation is done while performing tests of significance as suggested in (Peterson 2009) The the data below approximately -1 quantile on the Q-Q plot departs from the line therefore it cannot be assumed that the data is normally distributed.
There are no significant outliers observed in the Residuals vs Leverage plot.

The above test of assumptions suggests the data does not fit well to a normal distribution and there is unequal variance in the residuals. There may be a more suitable model to use for this data or further transformations to improve the adherence to the required assumptions, however, the data will be used as is to perform the study replication.

Testing for conditions of RDD

The Regression Discontinuity Design (RDD) model uses a continuous variable W (running variable) and assesses β_1 close to the cutoff point c to assess for an approximation of the treatment affect. Value of W (in this study; the distance from cutoff) below c (in this study; the cutoff is 0) are the treatment group and values above c did not receive the treatment[2].

$$Yi = \hat{\beta}_0 + \hat{\beta}_1 Xi + \hat{\beta}_2 Wi + u_i$$

$$Xi = \left\{ \begin{array}{ll} 1, & Wi \geq c \\ 0, & Wi < c \end{array} \right\}$$

RDD measures the change in the probability at c (the discontinuity) and is used to estimate the effect of the treatment based on the size of the discontinuity at c.
RDD requires that outputs close to the cutoff are continuous and do not group just above or below the cutoff, and, in the absence of treatment, the student would get the same GPA result. If grouping or a jump occurs at the cutoff, it suggests that the data is biased. In the current study this may occur if the students, knowing the GPA cutoff, manipulated their results to increase their GPA to just above the cutoff.
The first year GPA results were chosen by the original report authors to reduce the likelihood of bias, as students are less likely to be aware of the GPA assessment and cutoff values in their first year of university. As we do not have a sample of students who were not put on academic probation when below the cutoff to compare with, the frequency of the GPA results is used to assess if there is an observed jump at the cutoff. Figure 2 shows the regression of the frequency of the distance from cutoff values. The data is continuous through the cutoff point therefore there is no evidence of bias around the cutoff.

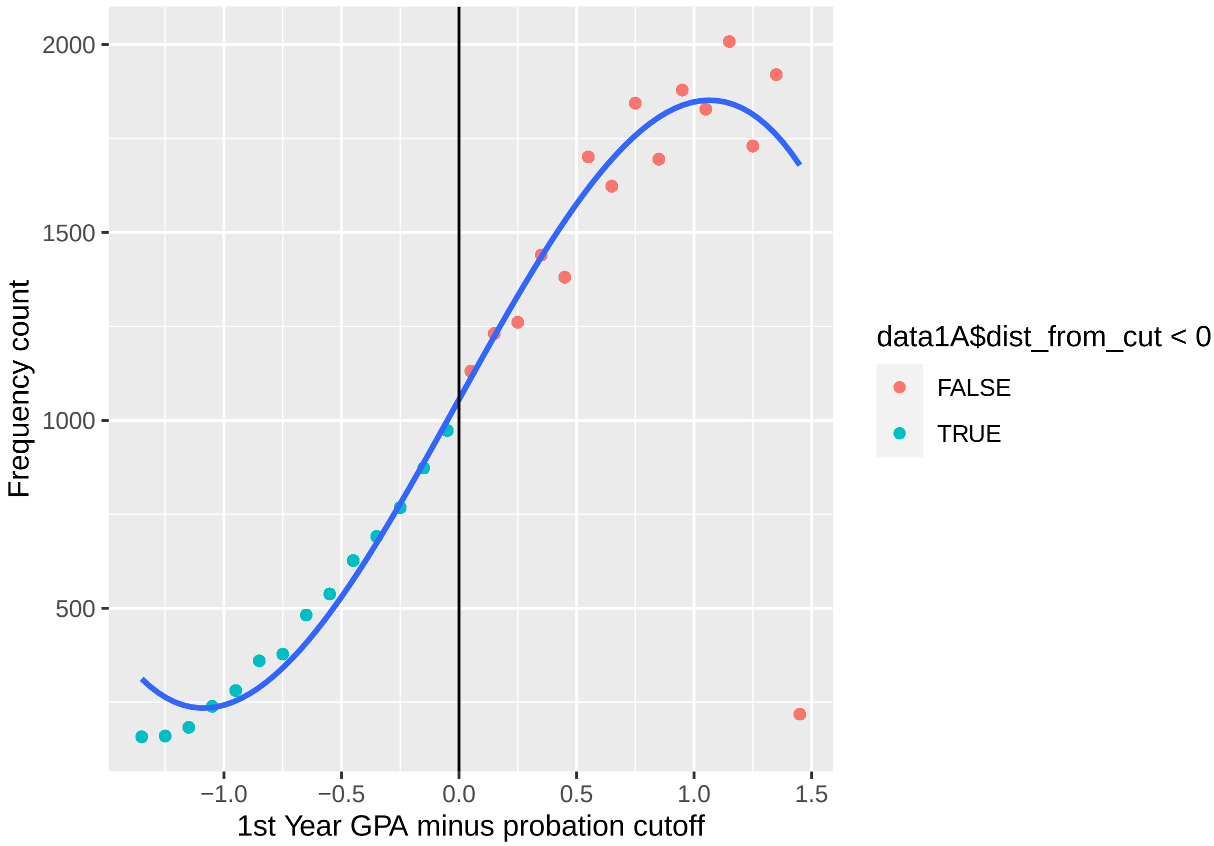
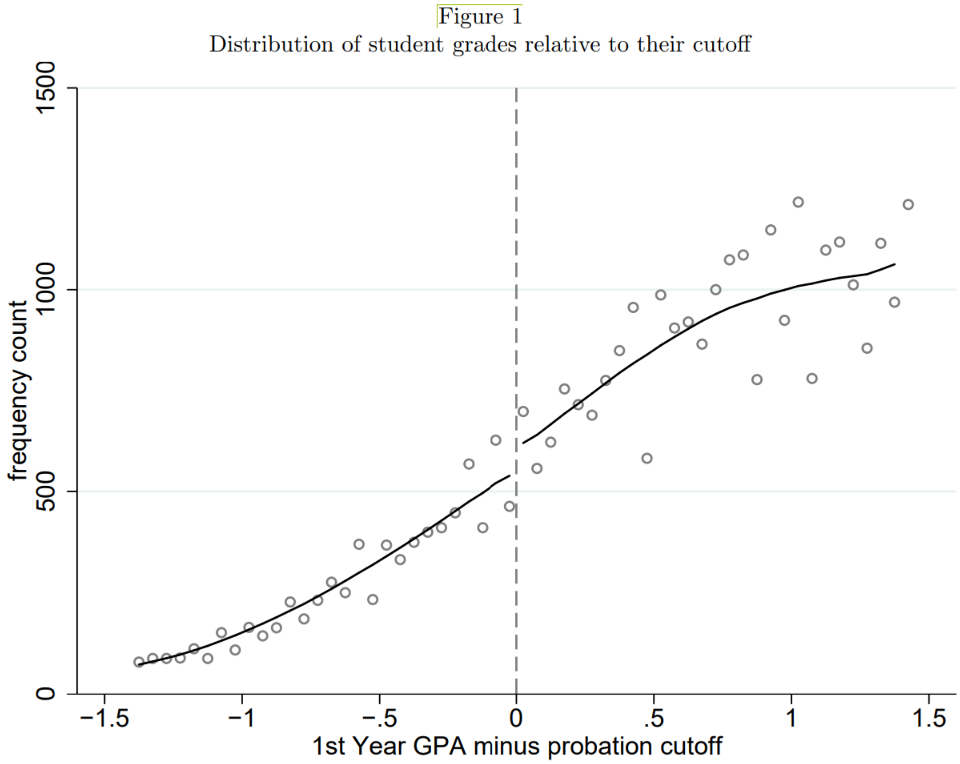

2000
1500
1000
500
Frequency count
−1.0
−0.5
0.0
0.5
1.0
1.5
1st Year GPA minus probation cutoff
data1A$dist_from_cut < 0
FALSE
TRUE

Figure 2 replicates Figure 1 from (Lindo et al. 2008)[1].



The observable characteristics were also assessed to determine if these variables are also continuous through the threshold. If any of the observations show significant discontinuity it is possible that a student with that characteristic is able to manipulate their grade to avoid being placed on academic probation.

Table 4: Observable Characteristics Estimated Discontinuity

	GPA below cutoff (1)	P-Value (1)	Std.err (1)	Intercept (0)	P-Value (0)	Std.err (0)
hsgrade_pct	0.0939305	0.9293879	1.0599761	30.6334367	0	0.7721364
totcredits_year1	0.0221858	0.7328393	0.0649933	4.3878668	0	0.0493337
age_at_entry	-0.0070179	0.8008362	0.0278194	18.7202783	0	0.0197531
male	-0.0183192	0.2922178	0.0173923	0.3771503	0	0.0112820
english	-0.0289750	0.1649021	0.0208632	0.7231576	0	0.0158699
bpl_north_america	0.0011262	0.9259052	0.0121095	0.8648038	0	0.0084997
loc_campus1	0.0013944	0.9639391	0.0308406	0.4459348	0	0.0231752
loc_campus2	0.0007837	0.9713773	0.0218417	0.2127378	0	0.0167962
loc_campus3	-0.0021781	0.9566283	0.0400484	0.3413274	0	0.0309865

Table 3 above replicates Table 2 from Lindo et al. (2008) [1].

Table 2
Estimated Discontinuities in Observable Characteristics

Bandwidth	1.40	1.00	0.60
Observations with non-zero weight	32,697	23,861	13990
	(1)	(2)	(3)
Dependent variable: High school grade (percentile ranking)			
1st Year GPA < probationary cutoff	1.290	1.045	1.920
	(1.213)	(1.375)	(1.651)
Constant (control mean)	30.639***	30.857***	30.788***
	(0.738)	(0.845)	(1.060)
Dependent variable: Credits attempted in 1st year			
1st Year GPA < probationary cutoff	0.045	0.058	0.079
	(0.079)	(0.088)	(0.110)
Constant (control mean)	4.382***	4.373***	4.366***
	(0.051)	(0.057)	(0.071)
Dependent variable: Age at entry			
1st Year GPA < probationary cutoff	0.017	0.032	0.019
	(0.035)	(0.040)	(0.051)
Constant (control mean)	18.734***	18.731***	18.725***
	(0.020)	(0.025)	(0.035)
Dependent variable: Male			
1st Year GPA < probationary cutoff	0.005	0.013	-0.015
	(0.023)	(0.027)	(0.033)
Constant (control mean)	0.377***	0.375***	0.392***
	(0.011)	(0.013)	(0.013)
Dependent variable: Born in North America			
1st Year GPA < probationary cutoff	0.007	0.013	0.004
	(0.012)	(0.014)	(0.018)
Constant (control mean)	0.871***	0.871***	0.872***
	(0.008)	(0.009)	(0.011)
Dependent variable: English is 1st Language			
1st Year GPA < probationary cutoff	-0.035	-0.055**	-0.049
	(0.022)	(0.026)	(0.034)
Constant (control mean)	0.736***	0.739***	0.734***
	(0.014)	(0.017)	(0.021)
Dependent variable: Attending Campus 1			
1st Year GPA < probationary cutoff	0.012	0.019	0.007
	(0.035)	(0.038)	(0.043)
Constant (control mean)	0.451***	0.445***	0.452***
	(0.023)	(0.025)	(0.026)
Dependent variable:Attending Campus 2			
1st Year GPA < probationary cutoff	-0.011	-0.001	-0.016
	(0.026)	(0.029)	(0.036)
Constant (control mean)	0.215***	0.217***	0.219***
	(0.017)	(0.019)	(0.022)
Dependent variable: Attending Campus 3			
1st Year GPA < probationary cutoff	-0.002	-0.018	0.008
	(0.046)	(0.052)	(0.062)
Constant (control mean)	0.334***	0.338***	0.329***
	(0.031)	(0.035)	(0.037)

Notes: Same as Table 3

* significant at 10%; ** significant at 5%; *** significant at 1%

All results are not significant at the 10% level (all p-values are greater than 0.1) therefore there is no evidence of discontinuity at the cutoff for any of the observable characteristics.
Assessment of the assumption that data is continuous through the threshold (and hence, not biased) is met.

Regression Discontinuity Design Analysis

Impact of Academic Probation on the next GPA result The inferential data replicated is the impact of being placed on academic probation on the students subsequent GPA (of students who remain at university).

Results were replicated using the polynomial linear regression equation stated in foot note 11 of the Lindo et al original report[1]:

[11]Our regression equation is given by:

$$\begin{aligned} Y_{ic} = \alpha &+ \delta_{le} 1(GPA_{ic}^{year\,1} < GPACUT_c) + \beta(GPANORM_{ic}^{year\,1}) \\ &+ \gamma_1(GPANORM_{ic}^{year\,1}) \times 1(GPA_{ic}^{year\,1} < GPACUT_c) \\ &+ \gamma_2(GPANORM_{ic}^{year\,1})^2 \times 1(GPA_{ic}^{year\,1} < GPACUT_c) \\ &+ u_{ic} \end{aligned}$$

where δ_{le} is the estimated impact of being placed on academic probation after the first year. Our figures suggest this functional form is a good fit. In prior versions of this paper, we found similar results using higher order polynomials.

It is noted that in the original report the assessment of impact on the subsequent GPA may be biased due to the loss of students (not returning to university in the second year) after the first year. The absence of second year results for these student can result in an increase in the probability of an improved GPA, as a number of students who would have a GPA below the cutoff in the second year if they had remained at university, are not longer included in the sample. Therefore students who have left school after the first year are removed from the dataset to reduce this bias.

Table 5: Next GPA Assessment

	GPA below cutoff (1)	P-Value (1)	Std.err (1)	Intercept (0)	P-Value (0)	Std.err (0)
All Data	0.2568016	0.0000000	0.0241821	0.3079885	0	0.0175804
HSgrade above Median	0.2252212	0.0004718	0.0643960	0.4260004	0	0.0347265
HSgrade below Median	0.2631084	0.0000000	0.0264614	0.2747851	0	0.0205670
Male	0.2064144	0.0000000	0.0377955	0.2709296	0	0.0268917
Female	0.2823507	0.0000000	0.0319481	0.3301747	0	0.0224475
Eng 1st language	0.2586673	0.0000000	0.0301759	0.3066086	0	0.0194409
Eng not 1st language	0.2538903	0.0000001	0.0470223	0.3119306	0	0.0341098

Results of the above table are compared to the first column of the first section of Table 6 (1.4 bandwidth, Control variable = no, No adjustment for students not returning), and Table 7 in the original report. Although results are not replicated exactly it still shows the probability of an improvement of the students subsequent GPA, calculated at 0.26 grade points (vs 0.19 grade points in the original report). All results are significant at the 1% level (p-values are less than 0.01).

Lindo et al. (2008) Table 6 and 7 [1]

Table 6
Estimated Discontinuities in Subsequent GPA

Bandwidth	1.40	1.40	1.00	1.00	0.60	0.60
Control variables	no	yes	no	yes	no	yes
	(1)	(2)	(3)	(4)	(5)	(6)
No adjustment for students not returning						
1st Year GPA < probationary cutoff	0.190***	0.185***	0.217***	0.210***	0.191***	0.181***
	(0.027)	(0.026)	(0.032)	(0.031)	(0.039)	(0.038)
Constant (control mean)	0.301***	-	0.298***	-	0.308***	-
	(0.016)	-	(0.021)	-	(0.025)	-
Observations with non-zero weight	29,745	29,745	21,628	21,259	12,626	12,626
Assuming students who left after being placed on AP would have performed the same in year 2 as in year 1						
1st Year GPA < probationary cutoff	0.128***	0.125***	0.149***	0.142***	0.130***	0.123***
	(0.024)	(0.022)	(0.029)	(0.027)	(0.035)	(0.033)
Constant (control mean)	0.301***	-	0.298***	-	0.308***	-
	(0.016)	-	(0.021)	-	(0.025)	-
Observations with non-zero weight	30,985	30,985	22,639	22,258	13,289	13,289
Assuming students on AP who left would have GPA change in the 10th percentile of similar students						
1st Year GPA < probationary cutoff	0.062**	0.058**	0.082***	0.074***	0.066*	0.058*
	(0.025)	(0.023)	(0.029)	(0.027)	(0.035)	(0.032)
Constant (control mean)	0.301***	-	0.298***	-	0.308***	-
	(0.016)	-	(0.021)	-	(0.025)	-
Observations with non-zero weight	30,985	30,985	22,639	22,258	13,289	13,289

Notes: Estimated standard errors, clustered on GPA, are displayed in parentheses. Estimates are based on regressions including a second order polynomial in adjusted first year GPA flexible on each side of the cutoff using rectangular kernel weights and bandwidths noted in the table header. Constants are included in all specifications, though they but are not shown for models with covariates as they do not represent the control mean in such cases. In the third panel, students that left after being placed on probation have their second year GPA imputed as their first year GPA minus 0.52—only 10% of students within 0.6 of the cutoff who return after being placed on probation do worse.

* significant at 10%; ** significant at 5%; *** significant at 1%

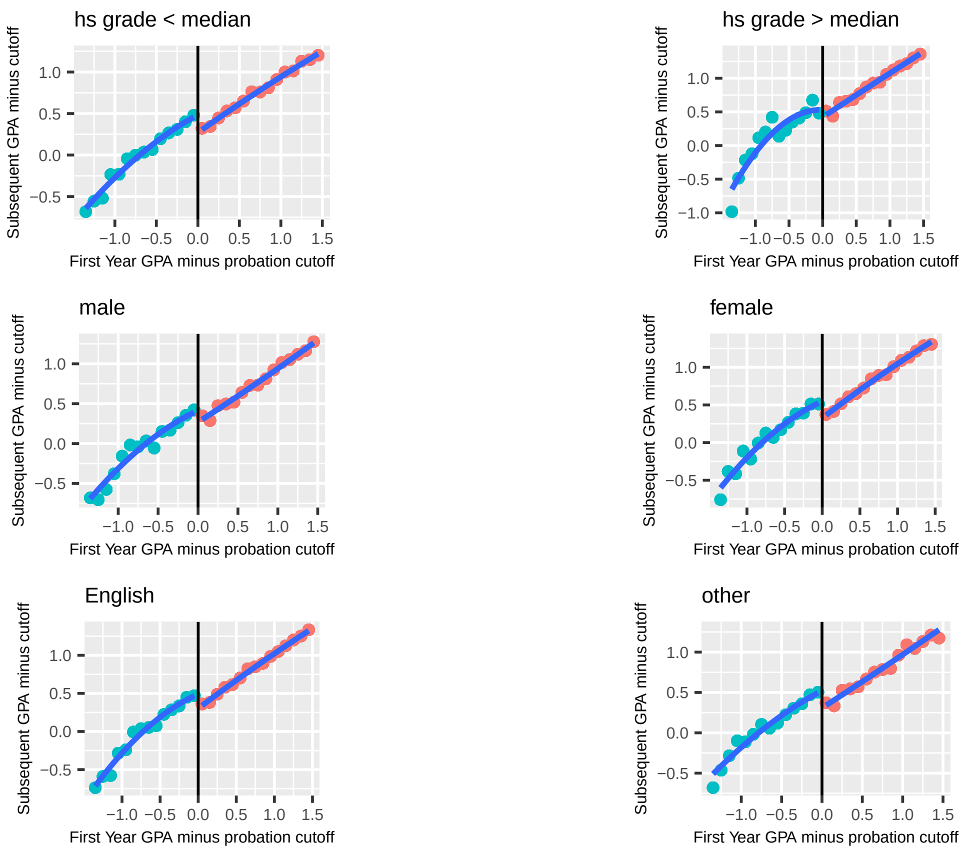
Table 7
Estimated Discontinuities in Subsequent GPA for Subgroups

Subgroup	HS Grades < Median	HS Grades >= Median	Men	Women	Native English	Non-native English
	(1)	(2)	(3)	(4)	(5)	(6)
1st Year GPA < probationary cutoff	0.205***	0.127*	0.134***	0.222***	0.177***	0.216***
	(0.030)	(0.076)	(0.046)	(0.035)	(0.036)	(0.052)
Constant (control mean)	0.271***	0.412***	0.280***	0.313***	0.310***	0.323***
	(0.019)	(0.035)	(0.025)	(0.023)	(0.032)	(0.037)
Observations with non-zero weight	18,866	10,524	11,111	18,634	21,443	8,302

Notes: Same as Table 5

* significant at 10%; ** significant at 5%; *** significant at 1%

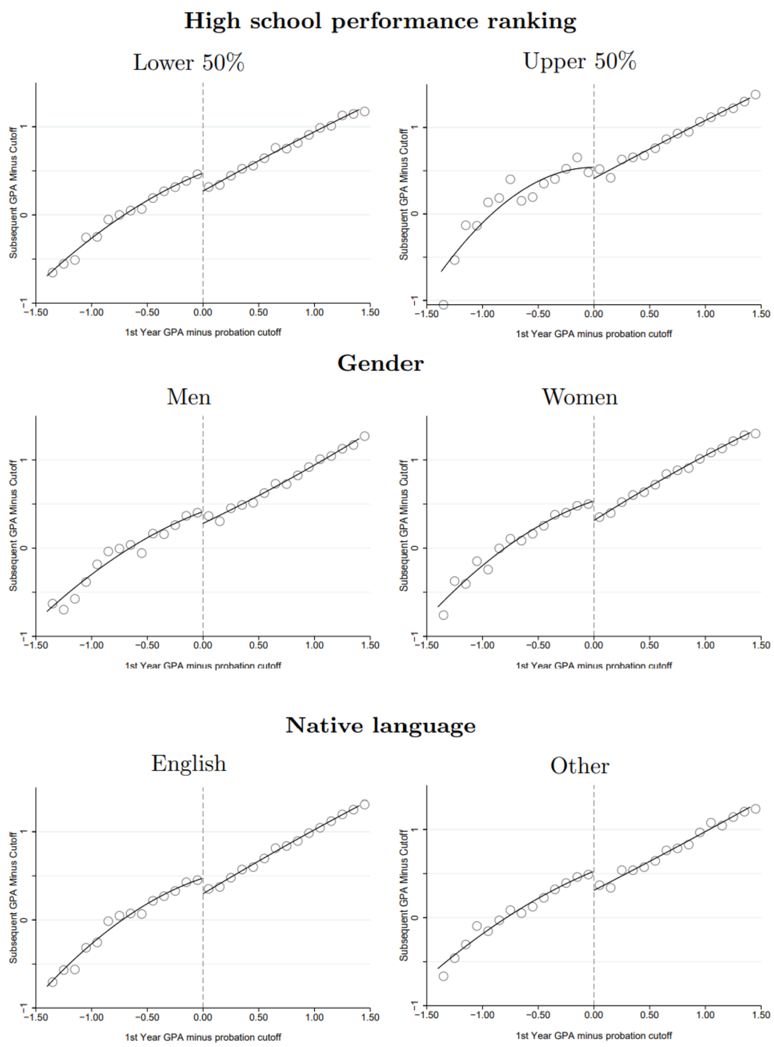
The subsequent GPAs for the different subgroups also showed and increase in the subsequent GPA as per the original report. Comparison of the groups shows the probability of improving the students subsequent GPA is greater in females than males, greater in students with high schools grades below the median than those above, and similar for those with English as their first language or not. It is noted that there is a difference in the native language results between the original report and this replication as the original report shows a greater difference in the increase in GPA between these two groups:
Original report difference in increase in GPA = 0.039 points
Replicated difference in increase in GPA = 0.005 points



Visual representation of the data in the following plots confirms the assessments made and align with the plots in teh original report.

Lindo et al. (2008) Figure 8 [1]

Figure 8
Stratified results for GPA in next enrolled term



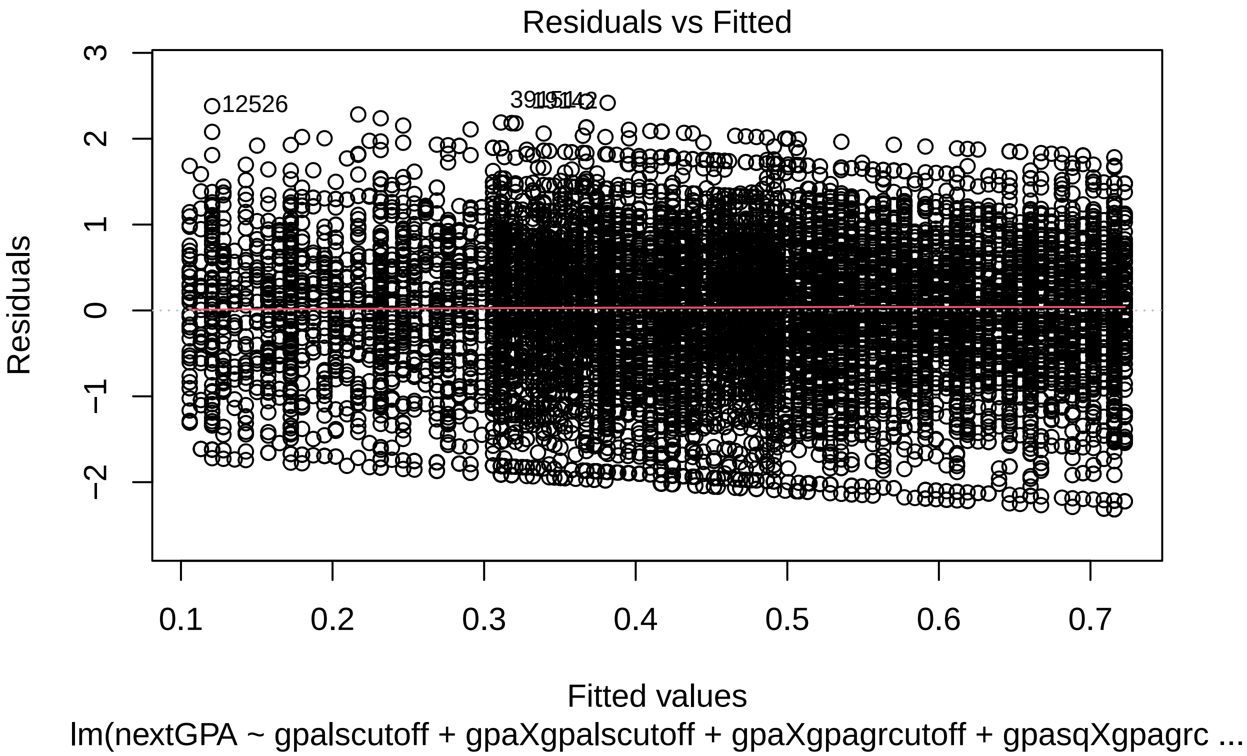
Critical Assessment

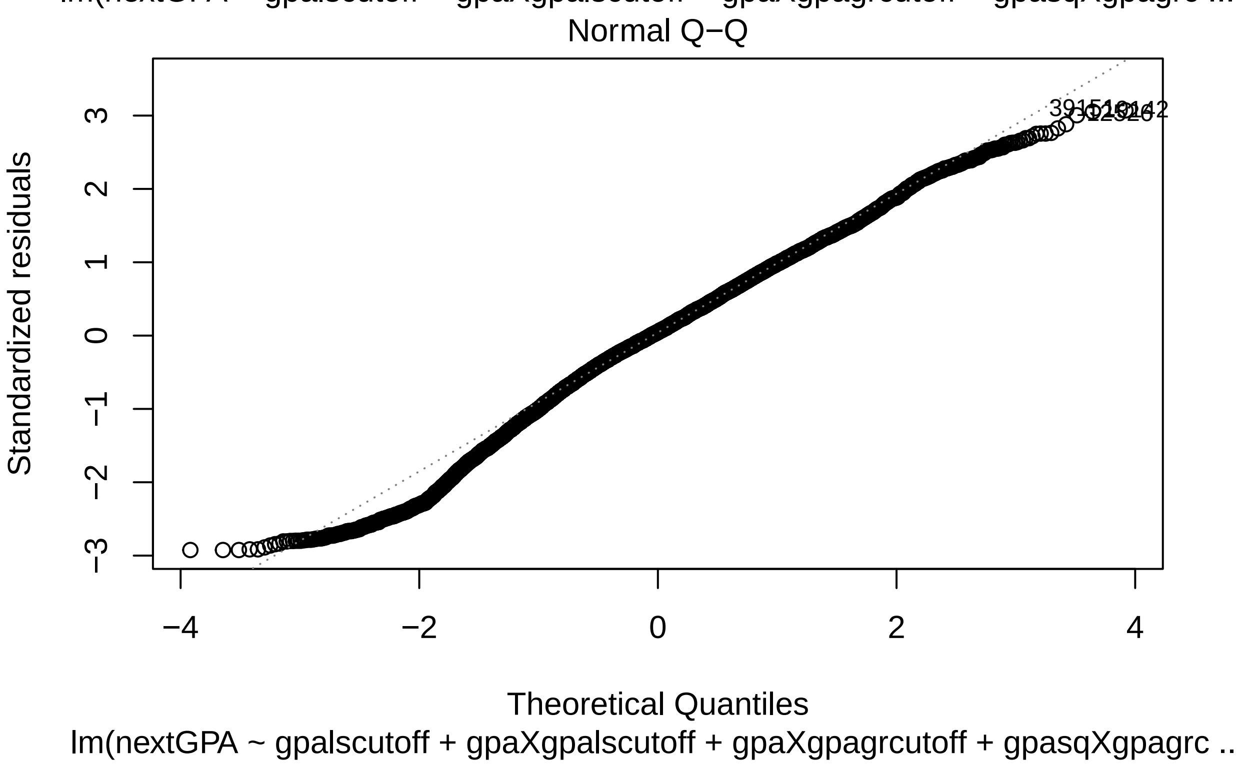
In replicating this report a number of issues and inconsistencies were encountered: - Data provided for replication is not the same as the data used in the report. The number of input variable using bandwidth of ± 1.4 in the report is 32,697 whereas the number of input variable in the provided data when the bandwidth is ± 1.4 is 29,601. This was a results of the data provided having missing result for the years 1995 and 2004. The replication 'Replication of Lindo, Sanders & Oreopoulos (2010), Student Project'[3] confirmed the dataset used in out assessment is correct as the validation check of observable characteristics performed in this report, when the data is limited to a bandwidth of ± 0.6, matches table 1 and the Extension table of the replicated report exactly.

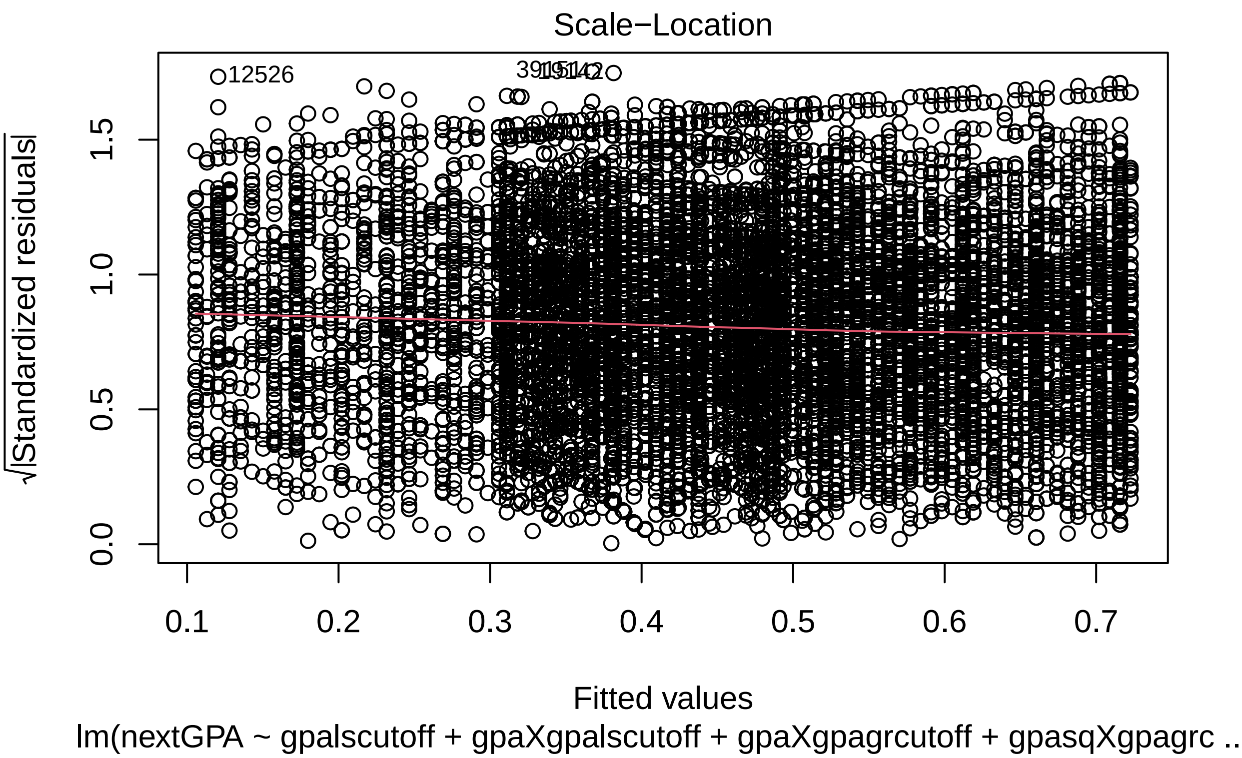
- There is no confirmed assessment of linear regression assumptions in the original report, only RDD assumptions have been addressed. Our assessment of assumptions of linear regression suggests the data is not normally distributed and there is inconsistent variance in residuals with a pattern observed in both the Residuals vs Fitted plot and the Scale-Location plot. Therefore a linear regression may not be the best model for analysing this data.

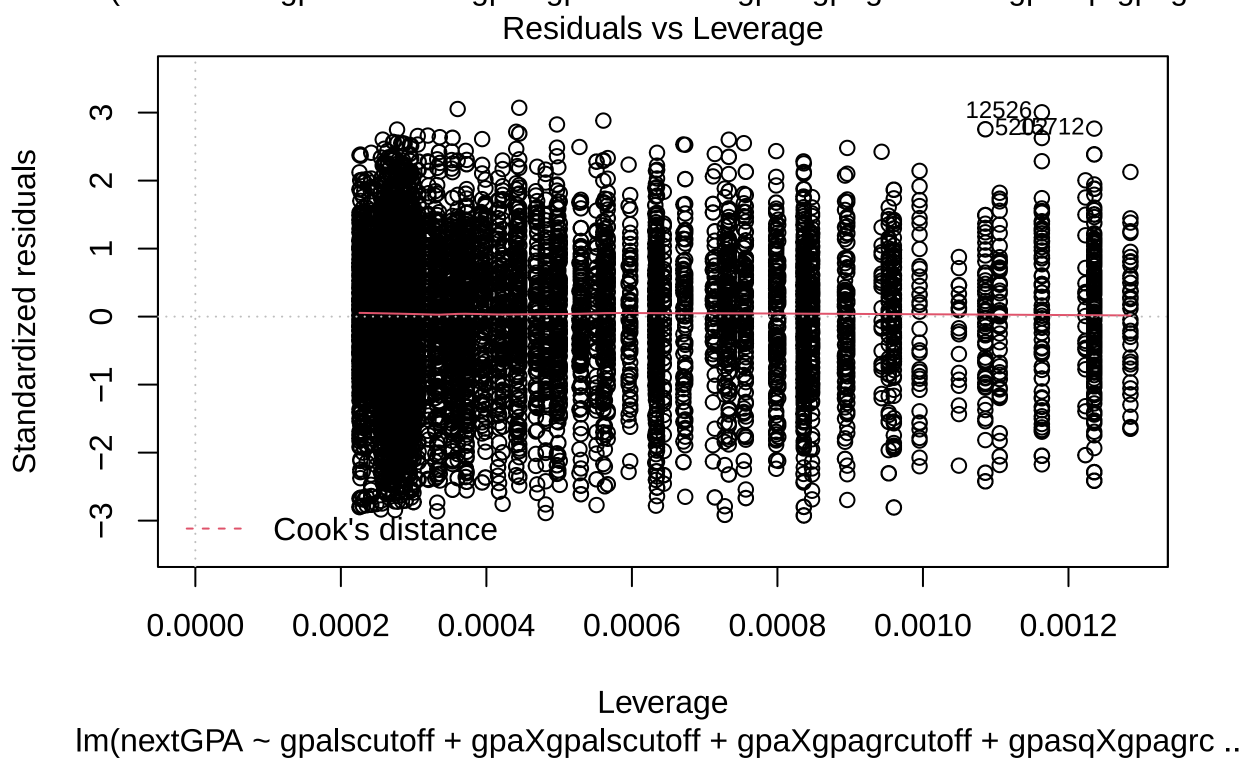
Because of these inconsistencies in the original report the data could not be replicated exactly.

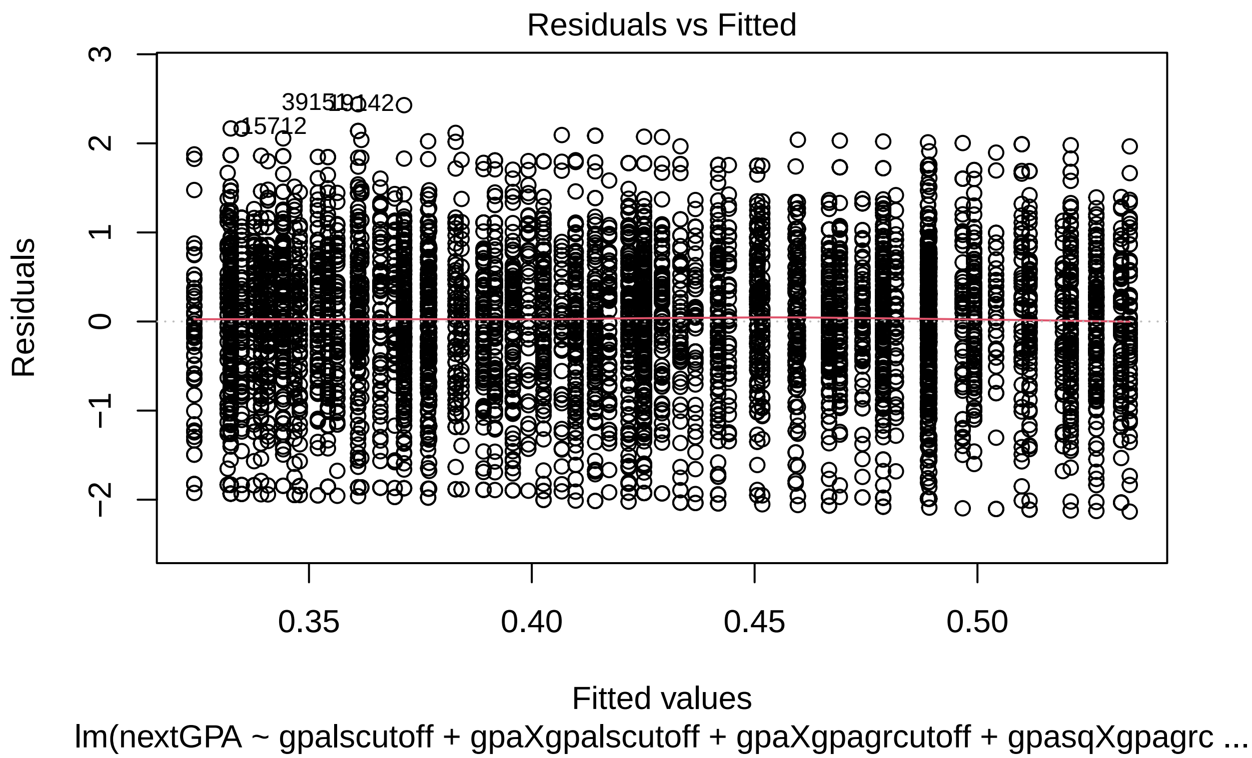
The bandwidth was reduced to ± 0.6 and ± 0.3 to test if this improved the fit to a linear regression model:

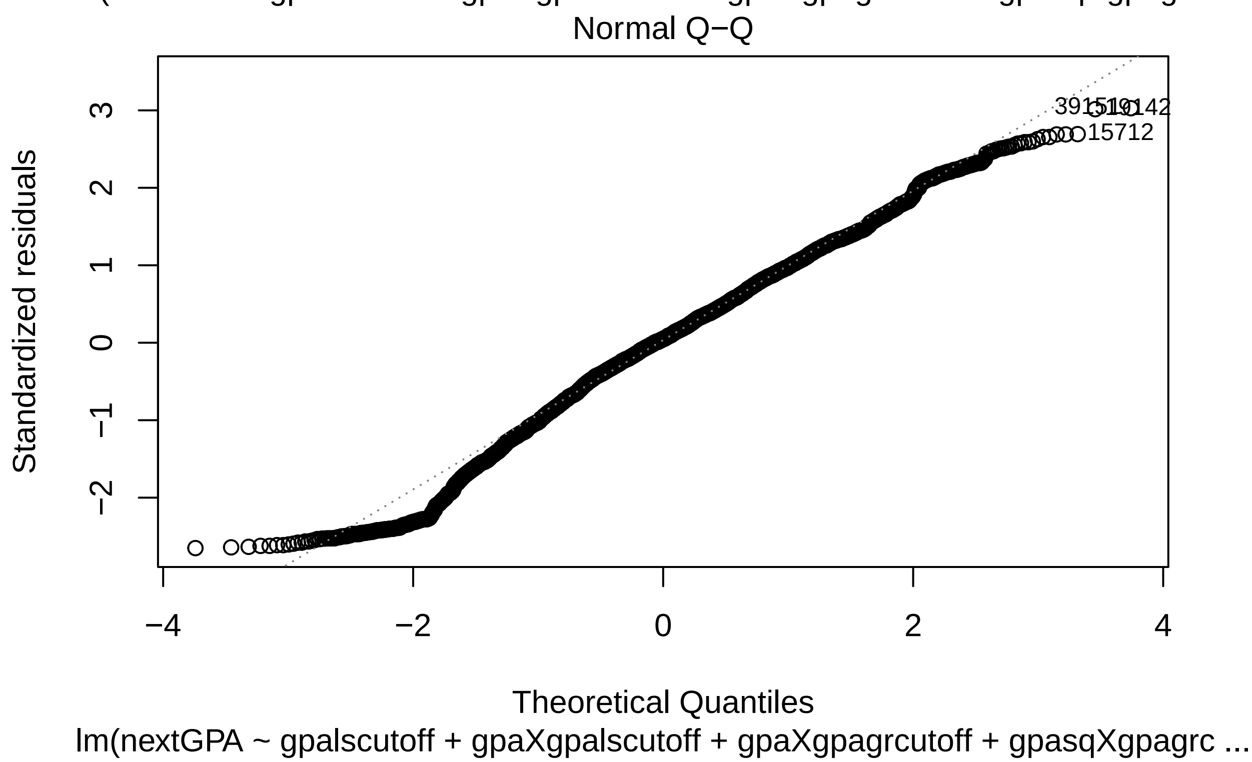


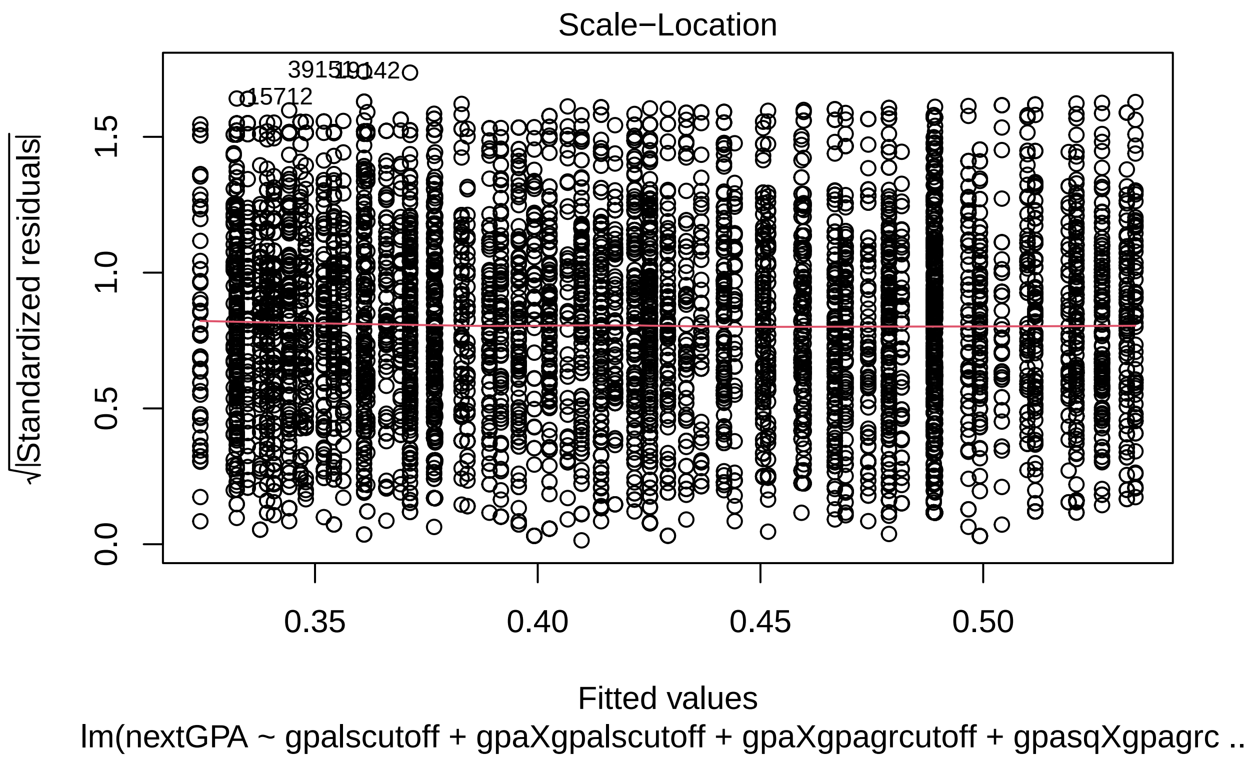


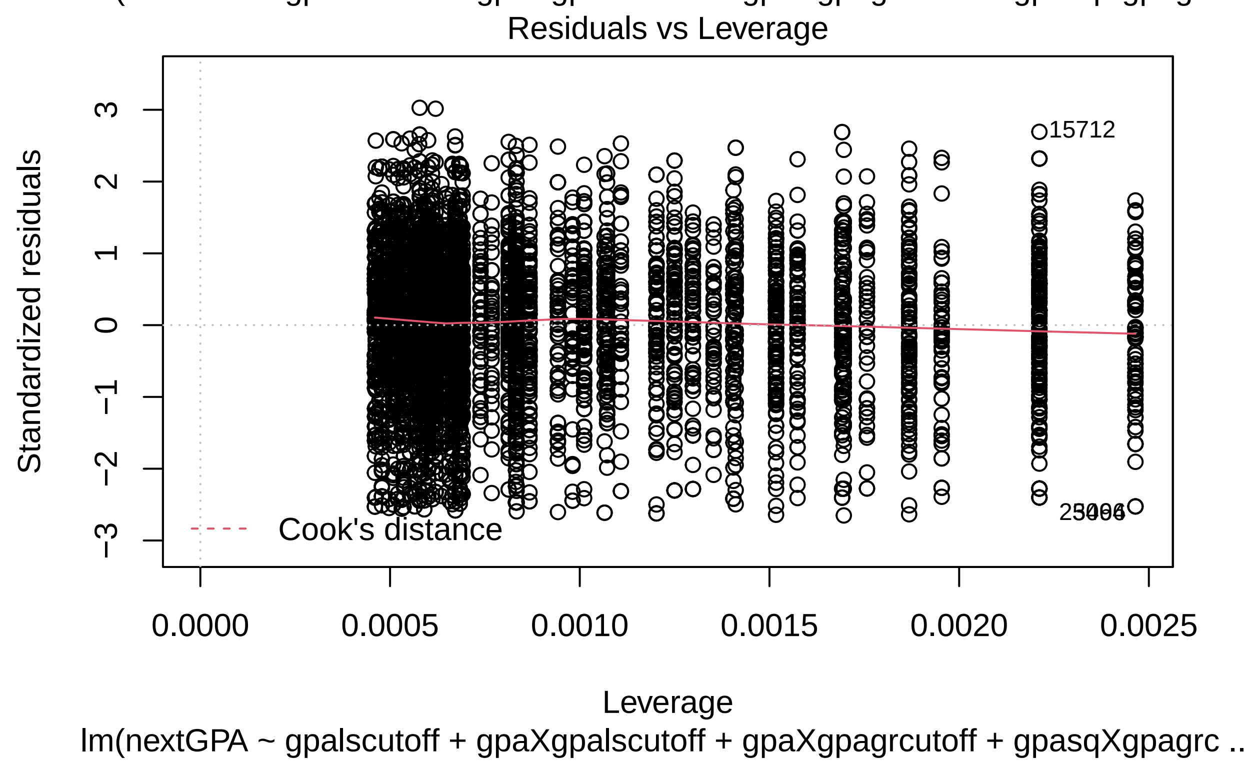

Scale−Location
√|Standardized residuals|
Fitted values
lm(nextGPA ~ gpalscutoff + gpaXgpalscutoff + gpaXgpagrcutoff + gpasqXgpagrc ...
12526


Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(nextGPA ~ gpalscutoff + gpaXgpalscutoff + gpaXgpagrcutoff + gpasqXgpagrc ...
12526


Residuals vs Fitted
Residuals
Fitted values
lm(nextGPA ~ gpalscutoff + gpaXgpalscutoff + gpaXgpagrcutoff + gpasqXgpagrc ...
391519
10142
15712
3
2
1
0
−1
−2
0.35
0.40
0.45
0.50


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(nextGPA ~ gpalscutoff + gpaXgpalscutoff + gpaXgpagrcutoff + gpasqXgpagrc ...
391519
10142
15712
3
2
1
0
−1
−2
−4
−2
0
2
4





The diagnostic plots are similar to earlier results. residuals variance does not improve the fit to a normal distribution. Further transformations of the output variable 'nextGPA' are required to improve the fit of the data.

The replication of the next GPA results were chosen as this data is continuous output data which fits the requirements of a linear model. Other outputs assessed in the original report are the effect of probation on the decision to leave university, the effect of ever being suspended, and the effect on graduation. All of these outputs are binary therefore the authors have transformed these outputs to probabilities to create continuous data suitable for use in a linear regression model. Further research performed has found that a binary model can be used for regression discontinuity which may be more suitable for these output types (section 3 of A Practical Introduction to Regression Discontinuity Designs: Volume II[4]).

The discussion section of the original report states that there is an unexpected response to academic probation in that it was expected that students with a lower ability (assessed using the observable characteristic of high school percentage) were expected to not show an increase in their subsequent GPA score or be more likely to drop out. However the students with a high school grade below the median showed a greater increase in their subsequent GPA than those above the median. In the original report this variation in the expected outcome has been proposed as occurring because the impact to the self confidence of the students with high schools grades above the median is greater than those students who are below the median. The example given in the study is that the higher ability student (based on high school grades) expects to do well in their first year. When placed on academic probation the students self confidence is reduced enough to prevent them performing well in subsequent years or ultimately drop out. Whereas the student who performed below the median in high school is expecting the experience to be difficult and is not impacted as much in future grades. The authors assume that the decision to return to school is a measure of self confidence and assesses the impact of probation on the measured student ability (high school grade) and self confidence (the decision to return to university). This is a limited view point with an alternative provided here:
The authors do not consider other observable characteristics in the above assessment. The replication of data in Table 4 also shows that both female students and students where English is a second language also show a greater increase in their GPA following academic probation when compared to the alternative subgroup (males and English as a first language). Considering both these other subgroups it could be proposed that students who performed below the median in high school, are female and have English as a second language have ‘more to lose’ based on societal pressures if they drop out or do not improve their subsequent GPA scores. This can be further assessed if additional data looking at socioeconomic status, name of the high school attended or address of childhood residence. This data could then be used to determine if there is also a correlation between increasing subsequent GPA scores.

Free Roam

Data for the observable attributes of age and birthplace were not assessed as part of the original report. The free roam section of the report below assessed the impact of academic probation in the first year on the subsequent GPA based on the age of the student (when starting university) and the student birthplace (limited to North America or other).

Table 6: Free Roam Next GPA Assessment

	GPA below cutoff (1)	P-Value (1)	Std.err (1)	Intercept (0)	P-Value (0)	Std.err (0)
All Data	0.2494352	0.0000000	0.0201727	0.3153549	0.0000000	0.0114563
Age entry = 17	0.3115107	0.0058122	0.1127412	0.2965530	0.0000003	0.0573077
Age entry = 18	0.2689409	0.0000000	0.0381668	0.3399820	0.0000000	0.0220345
Age entry = 19	0.2403013	0.0000000	0.0307291	0.3122061	0.0000000	0.0161216
Age entry = 20	0.1759932	0.0197977	0.0754600	0.2789598	0.0000000	0.0408608
Age entry = 21	0.3048772	0.0715958	0.1687802	0.1733606	0.1428656	0.1180915
Birth place Nth America	0.2557916	0.0000000	0.0229510	0.3232390	0.0000000	0.0113720
Birth place Other	0.2194292	0.0001749	0.0584060	0.2608772	0.0000000	0.0350878

All results show an increase in subsequent GPA results of 0.2 to 0.3 grade points at a minimum 10% significance. This is in agreement with the original paper that being put on probation increases the students probability of increasing their subsequent GPA (of those students that remain at university).The greatest increase is seen for those students who entered university at age 17 and 21 which are the upper and lower age limits of the data.

Meet The Authors

Author 1:

Name: Louise Higson
Student ID: 20631312
Email: louise.higson@postgrad.curtin.edu.au
Contributions to the project: Validity check plot, Inferential analysis and free roam analysis.

Author 2:

Name: Pradyumna Agrawal
Student ID: 20046165
Email: pradyumna.agrawal@postgrad.curtin.edu.au
Contributions to the project: Descriptive Analysis, Validity Check, and Plots to support inferential analysis.

References

Cattaneo, Matias D, Nicolas Idrobo, and Rocio Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Cambridge University Press.

Gehlen, Annica. 2020. "Replication of Jason m. Lindo, Nicholas j. Sanders & Philip Oreopoulos (2010)." *GitHub Repository Replication-Performance-Standards*. https://github.com/amageh/replication-performance-standards/blob/master/replication-notebook.ipynb; GitHub.

Hanck, Christoph, Martin Arnold, Alexander Gerber, and Martin Schmelzer. 2019. "Introduction to Econometrics with r." *University of Duisburg-Essen*.

Lindo, Jason M, Nicholas J Sanders, and Philip Oreopoulos. 2010. "Ability, Gender, and Performance Standards: Evidence from Academic Probation." *American Economic Journal: Applied Economics* 2 (2): 95–117.

Petersen, Mitchell A. 2009. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *The Review of Financial Studies* 22 (1): 435–80.

Shi, Hong. 2021. "Lack of Evidence?: Academic Probation Does Not Affect Student Dropout and Subsequent Performance." https://github.com/honn-ishinn/RDD_academic_probation; GitHub.