Academic Probation, Attrition, Future Performance and Gender in Canadian Post-Secondary Education

Evan Moir

15/04/2020

Abstract

This is an R Markdown document.

Introduction

Students in many post-secondary education institutions are placed on academic probation when they fail to keep up a certain level of academic performance.

Conceptual Background

Academic probation is a status given to students at post-secondary institutions who are not maintaining a certain level of academic performance. It is widely used in North American educational institutions. The arguments in favour of the use of academic probation are that it communicates to students the seriousness of their current academic standing, and gives them a status that may allow them access to additional help. However, it's also possible that the official nature of the probationary status adds extra pressure to an alread-struggling student. So the question of the effects of academic probation are of interest to institutions as they seek to promote continued enrolment, imporved academic performance, and later student success.

The possible psychological effects, gaming effects, and change in available resources on either side of the probation cutoff threshold make ths a potentially fruitful

Attrition and its complementary concept retention are used by institutions to understand which students are failing to progress beyond first year. It's defined as the rate at which first year students do not return for second yer. By looking for trends in the student populations with higher attrition rates, institutions may be able to lower those rates with broad or targeted initiative meant to ensure that the conditions are in place for students to progres beyond first year. Completing first year increasing the odds that a student will eventually graduate, which is good for the institution (higher graduate numbers, likely higher student success rate, and higher revenues) and presumably good for the student as well. Attrition can happen for many reasons, only some of which are within the control of the institution.

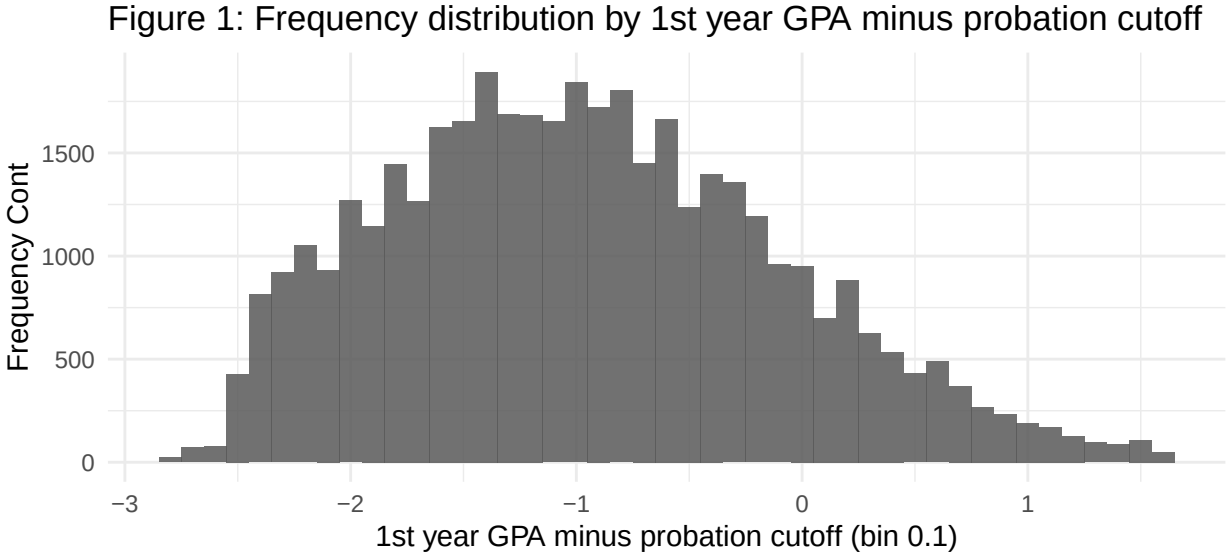
The Dataset

The chosen dataset of 44,362 rows is administrative data from a Canadian University. The data is at the student level and was collected between 1996 and 2005. The data includes variables describing:

- registration status
- 1st year GPA minus academic probation cutoff GPA

- 2nd year GPA minus academic probation cutoff GPA
- academic standing (i.e. probation status)
- gender
- age
- first language
- high school average

As shown in figure 1, the original dataset was heavily weighted in favour of students below the cutoff. It's not clear if this represented all students in the intended sample, or if the dataset was intentionally constructed in this way.



Research Question & Method

The general research question posed in this report is: how do the outcome of students placed on academic probation differ from those who are not? The outcomes of interest can be described through the following questions:

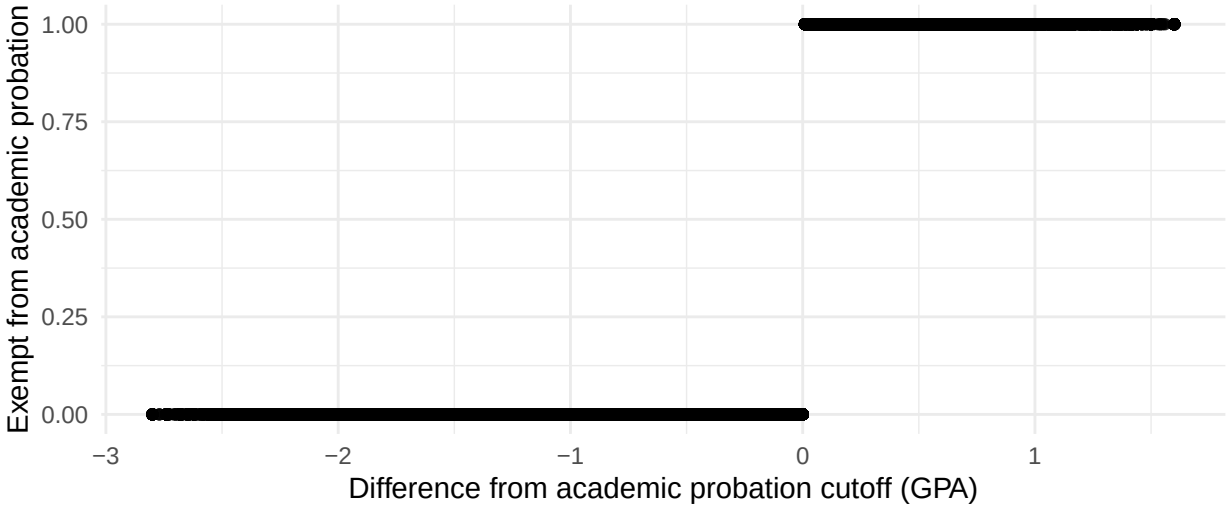
- Which students return for second year? (Retention / attrition)
- Do returning students see a change in their academic performance in second year?

Regression discontinuity (RD) will be used to explore how students on either side of the threshold respond in the areas of interest above. Both questions will be viewed through the lens of gender, i.e. do men and women differ on each side of the discontinuity threshold.

Establishing Existence of Discontinuity

If the intention is to use RD for analysis, it should be shown that there's an appropriate threshold that can be identified as a discontinuity. In this case, the discontinuity is the academic standing status (i.e. on academic probation, or not). Figure X shows that the academic standing is as expected on wither side of the discontinuit.

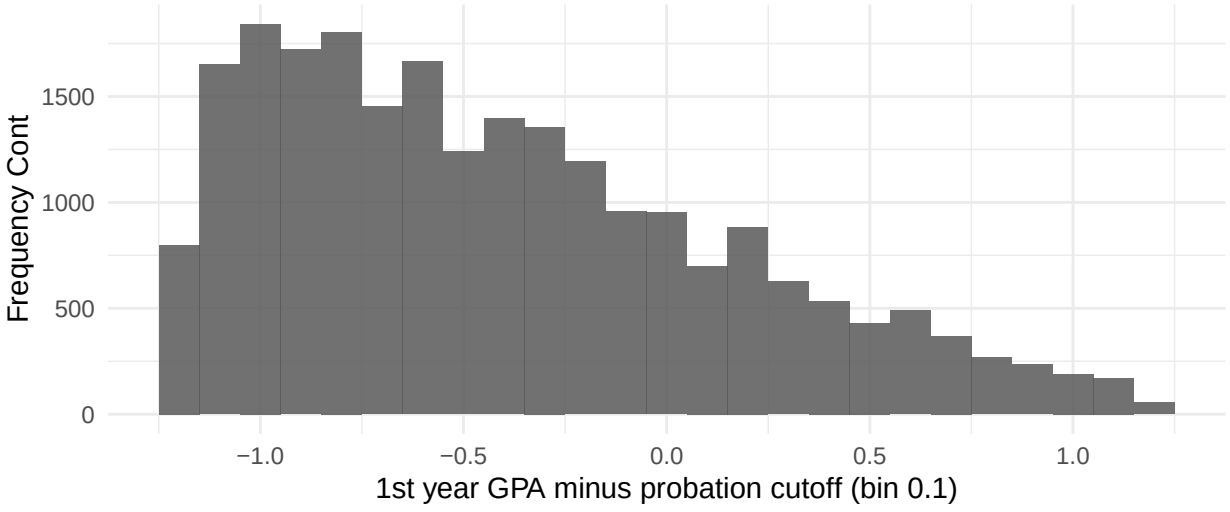
Figure 2: Probation status by 1st year GPA minus probation cutoff



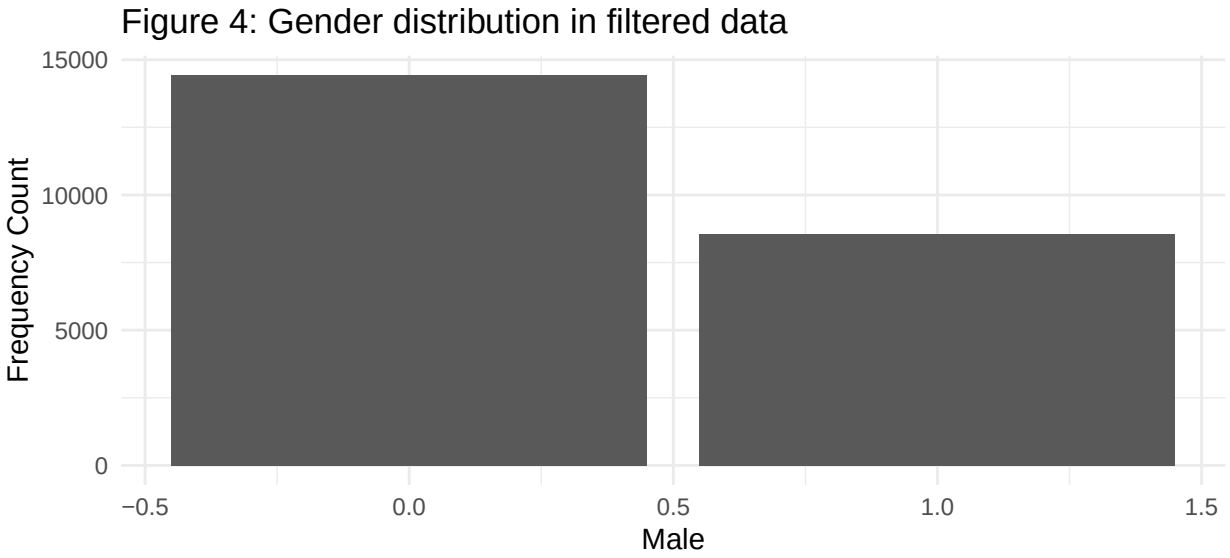
Data Preprocessing

The data was filtered to a band around the GPA cutoff of 1.2 points in both directios. This effectively removed failing students, and very high performing students. The resulting dataset contained 25,389 observations. The resulting frequency distribution is much more symetric around the cutoff.

Figure 3: Frequency distribution by 1st year GPA minus probation cutoff



Since gender is of interest, it's important to show that there are significant numbers of both genders in the filtered dataset:



To prepare the data for plotting and analysis, it was binned along by the "GPA minus probation cutoff" variable in increments of 0.1. The average "GPA minus probation cutoff" in each bin was then used as the basis for plotting (on the X axis in the plots to follow). the approrpiate value of interest (eg. attrition rate) was then calculated on the data in each bin for plotting on the Y axis.

Analysis & Discussion

Attrition

Figure 6 shows attrition rate for groups on both sides of the probation cutoff threshold, in an attempt to answer the question of whether being put on academic probation has an effect on attrition rate. A LOESS regression is used on the negative GPA minus cutoff values because the variation of the point coordinates made linear regression unsuitable. This result communicates two things clearly: students with lower GPAs show higher attrition rates, and there is minimal difference in attrition between the two groups around the threshold.

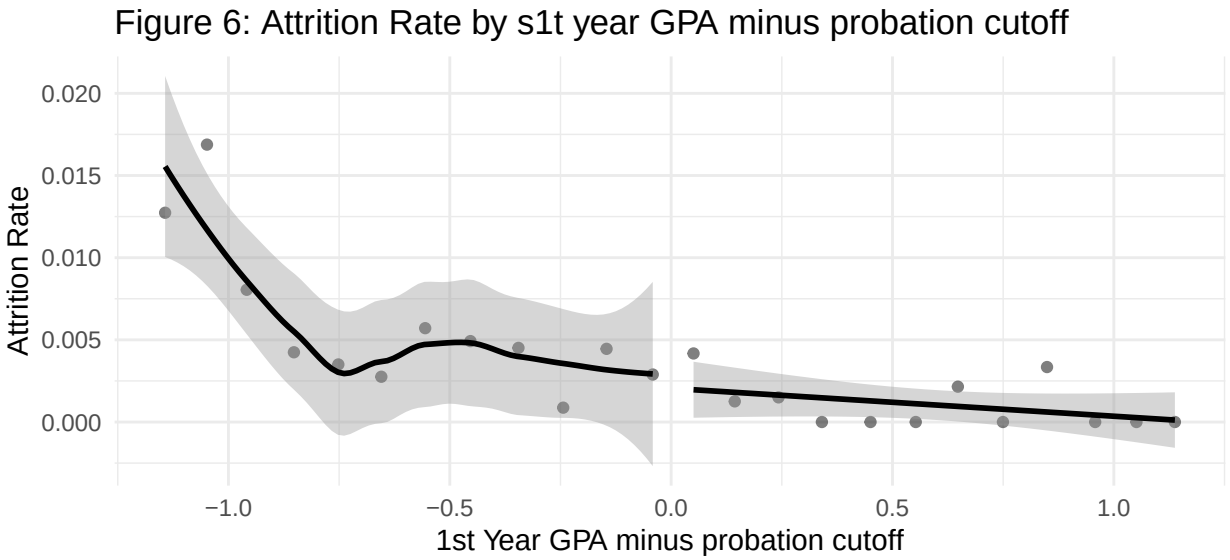
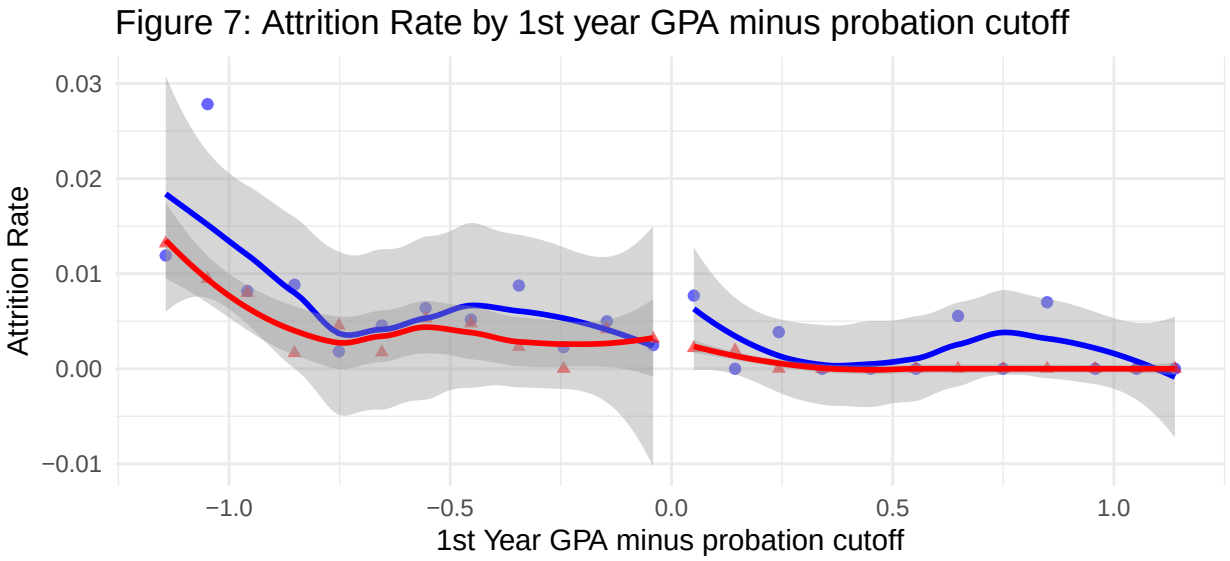
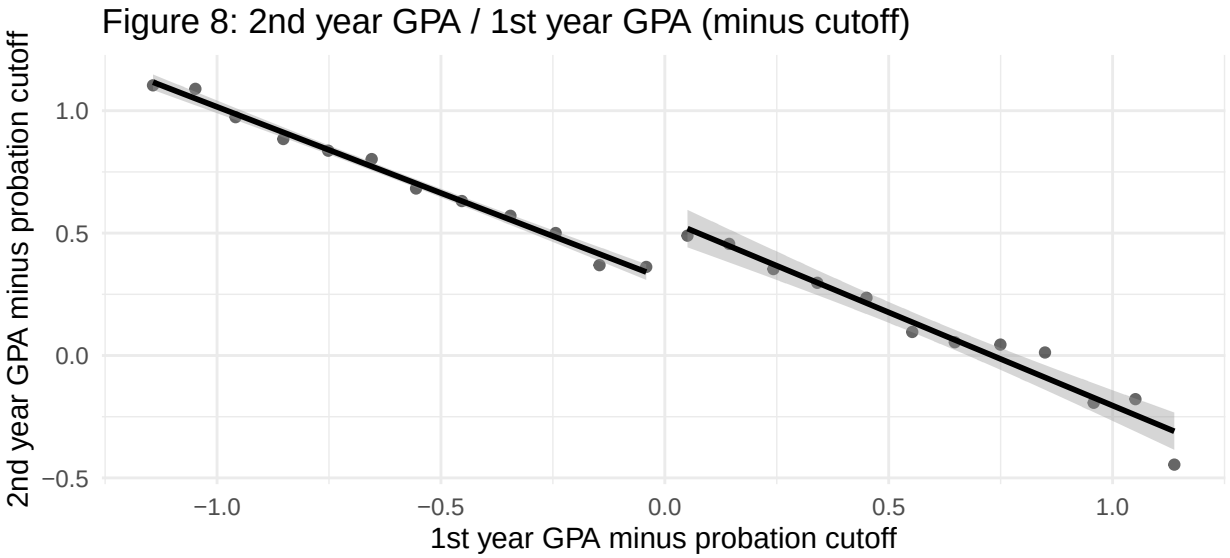


Figure 7 splits this into male and female plots, with LOESS regressions used on both sides for both genders. The results indicate that, in addition to the general trend of lower GPA leading to higher attrition rate, males have a higher attrition rate across the range of GPAs. There appears to be some discontinuity effect at the threshold with the male data, with the best fit line actually being lower on the negative side of the threshold - but the size of the error region makes it difficult to sy this is conclusive.



Performance Change

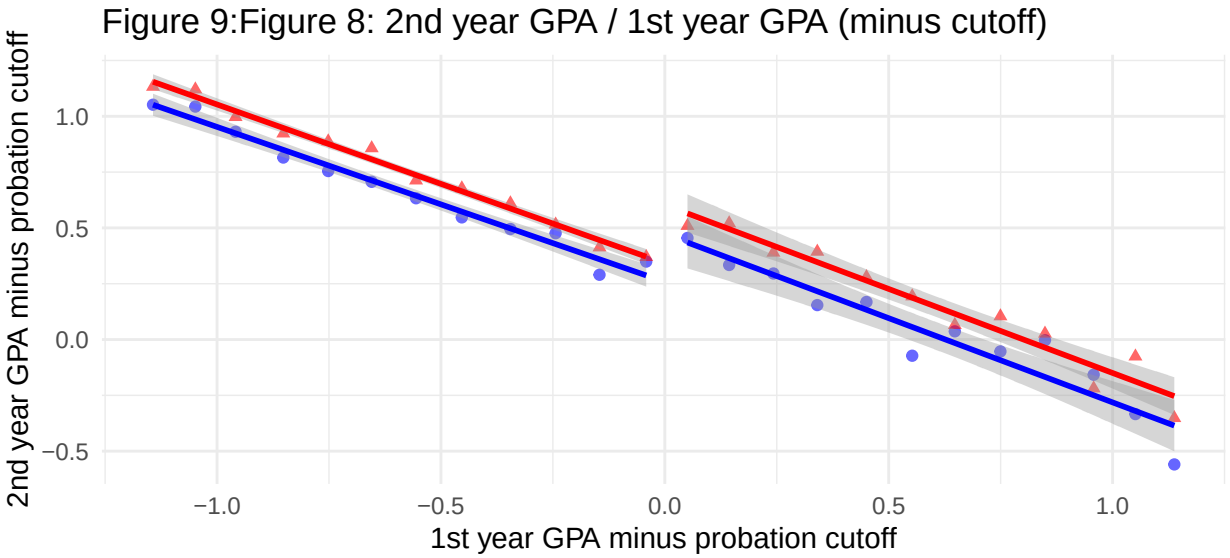
Figure 8 shows 1st year GPA minus probation cutoff values by 1st year GPA minus probation cutoff values for the filtered dataset in an attempt to answer the question of whether being put on academic probation in first year leads to improved performance in seond year. Linear regression is used to model both sides of the threshold, as the reationship appears quite linear.



The results appear to show a trend to wards higher 2nd year GPAs for lower first year GPAs, which is somewhat unintuitive (one might expect higher 1st year GPAs to correlate with higher 2nd year GPAs in general), but possible. However, there is also a very clear threshold effect - students just below the threshold

have a lower GPA than those just above the threshold. The data is highly linear here, so there is very little error and no overlap in the error regions immediately adjacent to the threshold.

Figure 9 shows the same data, but split by gender.



This shows a similar trend, though the errors are larger here so it's less conclusive. It also shows that women have a higher GPA in general, which is entirely possible.

Discussion

It became clear while performing this analysis that the question of whether academic probation affects attrition rates would be better answered by having a control group that's not put on probation. While the results above do show an discontinuous increase in attrition rate on the negative side of the threshold for the sample as a whole, the error ranges are large enough that this is inconclusive. Likewise for the gendered attrition rates - it's not possible to conclude that there are significant threshold effects.

The results found for performance change were somewhat unitutive, but for the sample as a whole there was a clear threshold effect where the students immediately on the negative side of threshold had a lower GPA than thoseimmediately on its positive side.

What could cause such an effect? Possibly the psycholigical effects alluded to in the introduction, e.g. the status of "probation" can affect a student's confidence or motivation. If additional availability of resources (e.g. academic counseling, tutoring) or increased motivation were the dominant effect, one would expect higher GPAs on the negative side of the threshold.

More study is needed to understand the causes behind the observed effect.

Ethical Considerations

There are a few motivations behind the desire on the part of post-secondary institutions to better understand the effect of academic probation. They can be broken int otwo categories: serving the interests of the students, serving the interests of the institution. (These interests undoubtedly overlap, but not entirely.) These institutions are large, complex organizations whose different members have differing motivations. It is likely simultaneously true that some individuals are primarily concerned with the interests of the students, while others are more concerned with the interests of the institution.

This partial divergence of interests is mentioned because it may be the case that academic probation, and analyses of it, may be used in ways that do not work in the interest of every individual student (or even populations of students). Though this analysis was concerned with a Canadian insitution, this is of partial concern in private education systems in other countries, where the profit motive further distorts institutions' incentives.

Regarding the use of gender, so long as the data is being anonymized, the most pressing potential ethical issues lie in the response to attrition and continued poor performance. If an analysis such as this leads to institutions prioritizing the needs of one gender over in other in supporting students on academic probation, they must be very certain that the analysis used was done in a rigorous way to ensure the reallocation of resources is fully justified.

References

- Ability, Gender, and Performance Standards: Evidence from Academic Probation. Jason M. Lindo, Nicholas J. Sanders, Philip Oreopoulos.
- Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
- Sebastian Calonico, Matias D. Cattaneo, Max H. Farrell and Rocio Titiunik (2020). rdrobust: Robust Data-Driven Statistical Inference in Regression-Discontinuity Designs. R package version 0.99.6. https://CRAN.R-project.org/package=rdrobust

Appendix

For full code, see Github repository:

INF2178 Problem Set 5