

Abstract Title Page

Title: Estimating the Impact of the PROMISE Scholarship Using Propensity Score Weighted Frontier Fuzzy Regression Discontinuity Design

Authors and Affiliations: Yetty Shobo, West Virginia Higher Education Policy Commission
Jen D. Wong, The Ohio State University
Angie Bell, West Virginia Higher Education Policy Commission

Background

Regression discontinuity (RD), an “as good as randomized,” research design is increasingly prominent in education research in recent years; the design gets eligible quasi-experimental designs as close as possible to experimental designs by using a stated threshold on a continuous baseline variable to assign individuals to a ‘treatment’. Fuzzy RD in which the threshold does not perfectly predict treatment receipt is a subset of this increasingly popular design. Lee and Lemieux (2010) identified only three education studies that used regression discontinuity design or its fuzzy subset between 1990 and 2000 compared to twenty-four using the design between 2000 and 2009. More studies have utilized the study design since 2009. However, two challenges hinder a wider adoption of RD designs: 1) its key requirement that individuals to the left and right of a stated threshold be exchangeable (Linden & Adams, 2012) and 2) the increasing use of multiple criteria for assigning ‘treatment’ in an environment of scarce resources. There is a need to explore ways to meet the key assumption of exchangeability and test the degree to which the requirement is met. Propensity scoring techniques offers a way to meet the assumption and calculate the degree to which the assumption is met (Linden & Adams, 2012). Reardon and Robinson (2010) also propose five ways of modeling multiple criteria threshold in RD design, one of which is frontier RD design. The combination of frontier RD design and propensity scoring techniques can be of great utility in the education sector. Specifically in examining the impact of merit-based grants, which are increasingly based on multiple criteria, combining frontier RD design with propensity scoring techniques can provide unbiased and efficient estimates of a grant’s impact; thereby, offering critical information for allocating resources in the context of today’s limited resources. This paper demonstrates the utility of combining frontier RD design and propensity scoring technique in estimating the effects of West Virginia’s Providing Real Opportunity for Maximizing In-state Student Excellence (PROMISE) grant.

In higher education, RD design is increasingly prominent in estimating the impact of merit-based aids on enrollment, school completion, grade point average (GPA), and other key outcomes. Merit-based aids are increasingly allocated based on individuals meeting multiple academic proficiency criteria. Since individuals are unlikely to be able to precisely sort around the proficiency thresholds for multiple criteria, it is not a far stretch to assume individuals to the immediate right and left of a threshold are similar (Lee & Lemieux, 2010). Consequently, award of merit-based aids based on multiple academic criteria is particularly similar to randomized experiments, where individuals who meet the multiple qualifying criteria receive the grant, and those just below the threshold do not receive the grant but are highly comparable and provide a control sample for estimating the effect of the grant.

A key premise in awarding merit-based grants is that the grants will positively impact the quantity and, possibly, the quality of schooling outcomes. In the context of college, the aim of merit-based grants is to lower the opportunity cost of schooling for academically qualified applicants, thereby increasing the likelihood of enrollment in college and, subsequently, on-time completion (Scott-Clayton, 2011). One would also hope that, by freeing up the time that would have been spent pursuing economic activities to pay for schooling, awardees would be more likely to enroll in college, enroll full-time, take a full load of courses, spend more time studying, and therefore have higher GPA. Merit-based aid programs with limits on years of awards will also likely accelerate student’s degree completion. In a period of limited financial resources and competing priorities, it is critical to examine whether these premises hold true for merit-based grants.

Few studies have examined whether receipt of merit-based grants is associated with higher likelihood of on-time college completion and high credit accumulation; even fewer studies have examined whether merit-based aid receipt is linked to college cumulative GPA. Further, current findings are varied, with some studies showing positive effects of financial aid on college completion (Dynarski, 2008; Scott-Clayton, 2011) and credit accumulation (Brock & Richburg-Hayes, 2006); whereas, few other studies show no effect (Angrist, Lang, and Oreopoulos, 2009). Cornwell, Lee, and Mustard (2005) also found that although Georgia State's HOPE scholarship reduced college dropout by 3 to 5 percentage points, it also reduced the likelihood of completing a full load of courses by six percentage points.

A careful examination of merit-based scholarship is particularly critical in West Virginia, a state with the lowest percent of adults 25 and older who have a Bachelor's degree (US Census, 2006). Started in 2002, the West Virginia PROMISE grant aimed to "improve high school and post secondary academic achievement through scholarship incentives" and "promote access to higher education by reducing cost to students." Originally, the grant provided full tuition for attending public colleges (equivalent amount was provided for students in in-state not-for-profit private colleges) but it now provides a maximum of \$4750 towards tuition. Academic eligibility to receive and continue receiving the four-year maximum grant is increasingly stringent. Initially, eligibility was based on having at least 3.0 overall GPA, 3.0 GPA in core courses, and an ACT composite score of 21. Now to qualify for PROMISE, students have to meet the previous GPA requirements and have at least ACT composite score of 22 and subject scores of at least 20 (alternatively, scores of at least 490 in SAT verbal, 480 in SAT math, and 1020 in total). To date, there has been just one evaluation of the PROMISE program. Scott-Clayton (2011) found higher four- and five-year Bachelor's degree completion for PROMISE recipients as compared to non-recipients. The research also found PROMISE recipients were more likely to have completed 120 credits in four years and to have GPA higher than 3.0. However, the GPAs of PROMISE recipients were not significantly different from those of non-recipients.

Scott-Clayton's (2011) study makes an important contribution to the research gap on the impact of financial aid in West Virginia but does little to ensure that the key RD requirement was met. By just including covariates as control variables, Scott-Clayton's regression models also reduced sampling variability (Lee & Lemieux, 2010) and may lead to biased results (Linden & Adams, 2012). Her lack of evidence to reject the null hypothesis regarding the covariate balance at the threshold was merely a coincidence though it allowed the study to meet RD's key requirement. In situations where certain covariates are not monotonically associated with the threshold variable by nature, using RD design will not be an option (Linden & Adams, 2012); this should not be so. As such, it is critical to discover ways to ensure that the key assumption in RD design is met. Propensity score-based balancing techniques offer an attractive way to meet RD's key requirement. The propensity score or probability of being treated conditional on observed covariates, controls for baseline differences between the groups to the left and right of the threshold in a RD design, resulting in balance. Consequently, individuals with the same propensity score on both sides of the threshold will be balanced on all baseline covariates. Although Imbens and Lemieux (2008) argued that propensity score-based techniques are incompatible with RD because there is no overlap in the assignment variable, RD's assumption of exchangeability in the immediate area around the threshold makes it plausible to assume that the assignment variable is unassociated with the model around the threshold, thus, making an assumption of overlap plausible (Linden & Adams, 2012). Another drawback of Scott-Clayton's (2011) work is that the study did not conduct separate analysis for students in 2-year versus 4-

year institutions, thereby, questioning the meaning of her success indicators such as four-year college completion. The requirements for eligibility for the PROMISE scholarship have also become more complicated since Scott-Clayton's (2011) study. The multiple eligibility criteria now used to qualify for PROMISE today requires innovative RD designs such as the frontier RD proposed by Reardon and Robinson (2010). The present study addresses these concerns by using propensity scoring technique in a frontier RD design involving cohorts of West Virginia in-state freshmen students who enrolled in four-year public institutions in 2007/08 and 2008/09 academic years.

Objective

The present study utilizes propensity scoring technique in a frontier RD design to estimate the effects of West Virginia's (WV) PROMISE on the quantity (4- and 5-year graduate rates, sum of credits earned) and quality (cumulative GPA) of long-term college indicators.

Improvement Initiative / Intervention / Program / Practice

In 2002, WV PROMISE program offered recent high school graduates full tuition scholarship to in-state two- or four-year public or private not-for-profit degree granting institutions if they obtained a 3.0 high school GPA and 3.0 GPA in core courses, scored 21 in the ACT composite (or 1000 in the SAT). Subsequent enrollment is contingent on a minimum of 15 credits enrollment per semester, 2.75 cumulative GPA in first year of receiving PROMISE, and 3.0 cumulative college GPA afterwards. A student who fails to meet the enrollment requirement in one semester is no longer eligible for the program. The requirement for initial enrollment has evolved over the years and now, in addition to the GPA requirements, ACT composite score of at least 22, and ACT subject scores of at least 20 (490 for SAT verbal, 480 SAT math, and 1020 total) are required. The multiple criteria used for qualifying for PROMISE requires innovative research designs. The present study investigates a combination of such methods by employing a frontier RD design. Further, the use of propensity scores to create covariate balance on both sides of the threshold in the created frontier offers an easily interpretable effect (Reardon & Robinson, 2010)

Setting

The study examines the effects of receiving West Virginia's PROMISE scholarship for in-state students attending four-year public institutions.

Participants

Over 85 percent of PROMISE recipients attend public four-year institutions; around 1.6 percent attend public two-year and slightly over 10 percent attend private four-year institution (West Virginia Higher Education Policy Commission, 2009). This study focuses on four-year public institutions where majority of the recipients attend. The study examines student outcomes for 11,294 in-state freshmen students in West Virginia who enrolled in a public four-year institution in fall of 2007/08 and 2008/09 school years; 5407 or 47.9 percent of the full sample met the ACT/SAT test score criteria for PROMISE and form the frontier or the final sample used in this study.

As shown in Table 1, more than 50 percent of the full sample are female and White. Majority of the sample took the ACT rather than the SAT college entrance examination. Not

surprisingly, the mean GPA is higher for PROMISE recipients than the non-recipients. Majority of the sample received some type of grant in their first year of college. PROMISE recipients tended to have higher cumulative GPA at the end of their sixth year in college and were more likely to have graduate college by the end of their sixth year.

Insert Table 1 about here

Research Design

The present study uses fuzzy and frontier RD designs because PROMISE receipt is not perfect above the threshold and is based on multiple criteria. It uses a fuzzy RD design because and, as shown in Table 1, PROMISE uptake rate among seemingly eligible students is 91.5%; that is, not all students who appear eligible for the grant accept it. Imperfect PROMISE uptake could be due to students enrolling in out-of-state or other non-eligible institutions. Another possibility explanation could be due to their core GPA, another criterion to be PROMISE eligible, which is not included in the available data, is less than 3.0. Further, college applicants report their GPA to colleges in the fall semester before the year they intend to matriculate but their GPA may change enough in their last year to qualify or disqualify them for PROMISE. This lack of perfect receipt above the threshold requires a fuzzy RD which involves a two-stage regression estimation.

The multiple criteria required for PROMISE also requires first creating a frontier. In this study, the frontier is created by selecting all students who met the testing criteria. Based on Reardon and Robinson (2010), this study selects all in-state freshmen students in four-year public institutions in 2007/08 and 2008/09 academic years who obtained at least scores of 22 in ACT composite and 20 in each ACT subject (minimum of 490 for SAT verbal, 480 SAT math, and 1020 total). Having at least a 3.00 GPA then is used as the threshold for PROMISE receipt or non receipt; students who met the test eligibility and who have a high school GPA of at least 3.00 qualify to receive the PROMISE, whereas those who have less than 3.00 high school GPA do not qualify. The effect estimated in the analysis is therefore the local average effect of requiring a minimum of a 3.0 GPA for PROMISE among students meeting the testing criteria.

Data Collection and Analysis

Administrative data was obtained from West Virginia Higher Education Policy Commission (WVHEPC). WVHEPC is the state agency that administers and awards PROMISE, and regulates higher education institutions in general. This study used data for the cohorts who were freshmen in the first two years in which the most recent PROMISE award eligibility requirement changes were made. Using the 2007/08 and 2008/09 data also provide time needed to have information on four- and five-year graduation. Unfortunately, six-year graduation data were not available for the 2008/09 cohort at the time of the analysis so that outcome was not examined. Credit taken and GPA for spring of the sixth year was, however, available for the 2008/09 cohort. Similar, data were generated for the 2007/08 cohort.

To obtain the propensity scores, we regressed the probability of PROMISE receipt on dummy variables indicating Caucasian, African American or Black races, Hispanic ethnicity, gender, and a continuous variable of age in a logistic regression. We saved the predicted propensity scores and then computed the inverse probability of treatment weights (IPTW) as $1/\text{propensity score}$ for PROMISE recipients and $1/(1-\text{propensity score})$ for non-recipients.

Following Imbens and Lemieux (2008), we predicted PROMISE receipt using test discontinuity in equation 1. Using linear regressions, we then estimated the effect of predicted receipt on the outcomes of interest in the second stage in equation 2 for students who were test-eligible. Equation 1 is the RD estimate of the effect of crossing the GPA threshold of 3 whereas the fuzzy RD in Equation 2 estimates the effect of PROMISE receipt. The IPTWs were used as weights in the Fuzzy RD analyses. A bandwidth of 0.5 on each side of the GPA threshold was used in the analysis. For test sensitivity to bandwidth and show robustness, analyses with two different bandwidths were also conducted. For comparison, and to highlight the efficiency of using propensity scoring techniques, we also ran unweighted regressions reflecting equations 1 and 2.

$$(1) P_i = \psi + \alpha(\text{above}_i) + \alpha(\text{GPA}_{\text{dist}_i} * \text{above}_i) + \mu(\text{GPA}_{\text{dist}_i} * \text{below}_i) + \epsilon_i$$

$$(2) y_i = \Omega + \beta(\hat{p}_i) + \phi(\text{GPA}_{\text{dist}_i} * \text{above}_i) + \Theta(\text{GPA}_{\text{dist}_i} * \text{below}_i) + \lambda \text{Covariates}_i + \epsilon_i$$

Where P_i indicates PROMISE receipt, \hat{p}_i is the predicted promise receipt, above_i indicates that a student is above the GPA threshold and below_i indicates a student is below the threshold.

$\text{GPA}_{\text{dist}_i}$ is the distance between a student's GPA and the threshold GPA of 3.0. covariates are dummy variables for White, Black, Hispanic, and Pell grant receipt, and a continuous variable of age. β estimates the difference in outcomes at the threshold.

Findings

Figure 1 presents the distribution of high school GPA; the top left quadrant presents data for the full sample whereas the top right quadrant shows the distribution for those in the frontier. There is no evidence of precise sorting around the threshold. The lower half of Figure 1 also shows the plot of mean IPTW derived from the covariates by high school GPA. Similar graphs for each covariate (not included) suggested balance between the group to the right and left of the threshold. Table 2 also shows that weighting removed systematic differences in the baseline characteristics of PROMISE eligibles and ineligibles. Figure 2 plots two of the outcomes of interests examined in this study by high school GPA. For the dummy variable indicating that students earned at least 30 credits in their first year of college, there is evidence of discontinuity or treatment effect at the threshold for both the full and frontier samples in the top part of Figure 2. However, there is no evidence of treatment effect in the charts of earning a bachelor's degree in four years shown in the lower half of Figure 2.

Insert Table 2 and Figures 1 and 2 about here.

Finally, Table 3 presents the average PROMISE effect for those who met the GPA requirements and the testing criteria, compared to those who just met the testing requirements, for different outcomes. Conditional on having met the testing requirements, receiving PROMISE resulted in higher likelihood of earning a Bachelor's degree in four or five years, having at least a cumulative GPA of 3.0, and having 30 or more credits at the end of year one and 120 or more credits at the end of year four. Further, PROMISE receipt resulted in higher credit earned at the end of years one, four, and spring semester of year six. Recipients who met the GPA and testing requirements had nearly 20 credits more at the end of the spring semester in year six. They also had cumulative GPA that were 0.33 and 0.34 points higher at the end of year four and the spring semester of year six, respectively. All the effects were significant at either .01 or .001 levels.

Table 3 also shows that the results are quite robust to alternative specifications. The tests of robustness using a narrower and a wider bandwidth confirm our findings. All the effects were

significant in the two models with alternate bandwidths although there were slight fluctuations in the magnitude of the effects. Finally, we tested whether the results were sensitive to our choice of functional form. A quadratic equation like Equation 2 was specified with two quadratic terms added for GPAdist on both sides of the threshold. Other than having higher standard errors, the differences in the estimates were subtle.

Finally, we compared the baseline model to the unweighted model. Not only were the estimates in this model barely significant and in unexpected directions when significant, the standard errors were of several magnitude to those in the baseline model. Consequently, the confidence intervals were wider, offering less precise estimates.

Insert Table 3 about here

Conclusions

This study demonstrates the significant potential that the frontier RD and the propensity score weighting techniques hold for education research. Estimates obtained from combining the two are quite robust to alternative bandwidth and alternate functional form. More importantly, they are more precise and significant than those obtained when using covariates instead of the score.

Although the frontier RD design reduces power in analysis, this impact is not consequential in this paper. We can confidently conclude that WV PROMISE has significant impact on several key long-term outcomes for students who met both the GPA and testing requirements compared to those who met just the testing requirements. The estimates also have strong internal validity as the ability of individuals who cannot precisely control their scores to reach a threshold is random. The results from this paper will likely facilitate the popularity of frontier RD designs in education research and the broader use of propensity scoring techniques in such research.

Limitations

The findings in this study are based on certain assumptions. For instance, this study's use of a propensity score-based technique assumes an overlap in the characteristics of students in the immediate area on either side of the threshold, which is not far-fetched as it mirrors RD's assumption of exchangeability. This study also assumes that all important confounders for the propensity score model are observed and included based on expert opinions. Further, it assumes that the reason why seemingly PROMISE-eligible students do not receive the scholarship is likely because they were no longer eligible at the time of college enrollment. That is, their GPA may have dropped in their last semester of high school or their core GPA may be lower than the required 3 points. This study also assumes that such reasons do not matter for the purpose of estimating the average treatment effect. In cases in which PROMISE-eligible students chose to enroll in an ineligible institution, it is likely because the institution is offering merit-based aid that has more value than PROMISE. If such is the case, then our average treatment effect estimates are valid but likely conservative. While it is not possible to validate these assumptions, their degree of plausibility lends credence to this study.

Appendices

Appendix A.

References

- Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1, 136-163.
- Brock, T., & Richburg-Hayes, L. (2006). *Paying for persistence: Early results of a Louisiana scholarship program for low-income parents at tending community college*. New York: MDRC.
- Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student responses to merit scholarship retention rules. *The Journal of Human Resources*, 40, 895-917.
- Dynarski, S. (2008). Building the stock of college-educated labor. *Journal of Human Resources*, 43, 576-610.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142, 615-635.
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48, 281-355.
- Linden, A. and Adams, J. L. (2012). Combining the regression discontinuity design and propensity score-based weighting to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, 18, 317-325.
- Reardon, S. F. and Robinson, J. P. (2010). Regression discontinuity designs with multiple rating score variables. *Journal of Research on Educational Effectiveness*, 5, 83-104.
- Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, 46, 614-646.
- West Virginia Higher Education Policy Commission. (2009). *Final Report of the PROMISE scholarship ad-hoc advisory committee*. <http://www.wvhepc.com/wp-content/uploads/2014/01/PROMISEfinalreport.pdf>

Appendix B. Tables

Table 1

Descriptive Statistics for West Virginia In-state Four-year College Enrollees

	All Sample	Promise Eligible	Promise Ineligible
Percent Female	54.7	53.2	55.9
Percent White, Non-Hispanic	92.2	94.8	90
Mean High School GPA	2.64	3.71	3.04
Took ACT	94.9	95.1	91.33
Took SAT	18.6	32.6	3.89
Received Pell Grant in First Year	38.1	27	37.5
Received Promise in First Year	44.9	91.6	1.9
Received any type of grant	91.6	97.4	84.9
Cumulative GPA by final year of data	2.65	3.11	2.29
% Graduated college by 6th year*	46.2	64.9	31.3
Average Award Received in first year	\$9,294	\$13,051	\$5,828
Average credits earned in final year	99.9	122	82.38
Sample Size	11,294	5,252	5,038

*6th year graduation data is not yet available for 2008/09 cohort

Table 2

Weighted and Unweighted Summary Statistics of Variables Used to Create IPTW

	All Sample	Promise Eligible		Promise Ineligible	
		Unweighted	Weighted	Unweighted	Weighted
Percent Female	54.7	53.2	55.0	55.7	54.7
Percent White, Non-Hispanic	92.2	94.8	92.2	90.2	92.2
Percent Black, Non-Hispanic	3.8	1.2	3.8	5.7	3.8
Percent Hispanic	1.0	0.8	1.0	1.1	1.0
Mean Age	18.4	18.3	18.4	18.4	18.4

Table 3

RD Estimates of the Effect of the WV PROMISE Scholarship, Using Estimated Eligibility as Instrument for Receipt (First Stage=0.92)

	Robustness Check				Using Covariates
	Using Inverse Probability of Treatment Weighing (Propensity Scoring)			Local Quadratic	
	Baseline Model	Alternate Bandwidths			
	HSGPA: 2.5- 3.5	HSGPA: 2.7 - 3.3	HSGPA: 2.3- 3.8		
Earned Bachelor's Degree in 4 Years	1.91(0.15)***	2.30(0.19)***	1.78(0.11)***	2.17(0.16)***	0.08(1.40)
Earned Bachelor's Degree in 5 Years	1.64(0.11)**	1.457(0.13)**	1.57(0.09)***	1.69(0.11)***	.40(1.08)
3+ GPA in Year 4	2.03(0.11)***	2.02(0.14)***	1.57(0.05)***	2.24(0.12)***	.27(1.09)
Has 30+ credits in year 1	3.18(0.11)***	3.75(0.14)***	2.80(0.09)***	3.30(0.12)***	1.96(1.10)
Has 120+ credits in year 4	2.31(0.12)***	2.34(0.15)***	2.57(0.10)***	2.51(0.13)***	.46(1.18)
Credits Earned, End of Year 1	3.12(0.43)***	3.07(0.55)***	2.75(0.29)***	3.30(0.44)***	-1.48(4.23)
Credits Earned, End of Year 4	17.18(2.18)***	15.35(2.71)***	16.88(1.62)***	18.33(2.22)***	-22.59(21.61)
Credits Earned, End of Spring Year 6	19.59(2.63)***	17.02(3.25)***	19.32(1.97)***	20.88(2.68)***	-21.97(26.09)
Cumulative GPA, End of Year 4	.33(0.04)***	.33(0.06)***	.26(.03)***	.37(0.04)***	-.44(.22)*
Cumulative GPA, Spring of Year 6	.34(0.04)***	.30(0.06)***	.27(.03)***	.38(0.04)***	-1.15(.43)*
Sample size	1490	792	3010	1490	1490

Note: * $p < .05$, two-tailed. ** $p < .01$, two-tailed. *** $p < .001$, two-tailed.

Figures

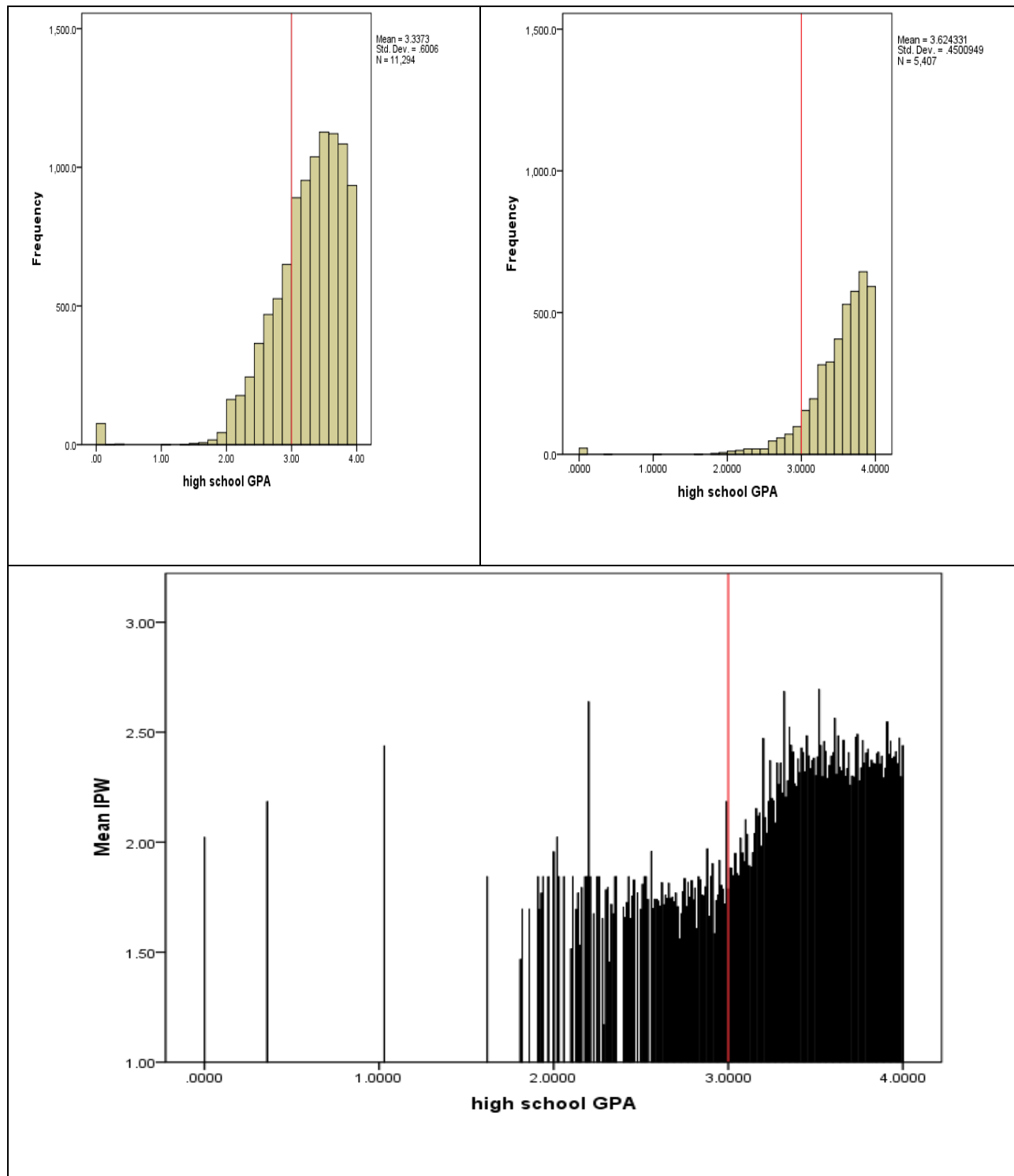


Figure 1. Distribution of high school GPA and inverse probability treatment weights

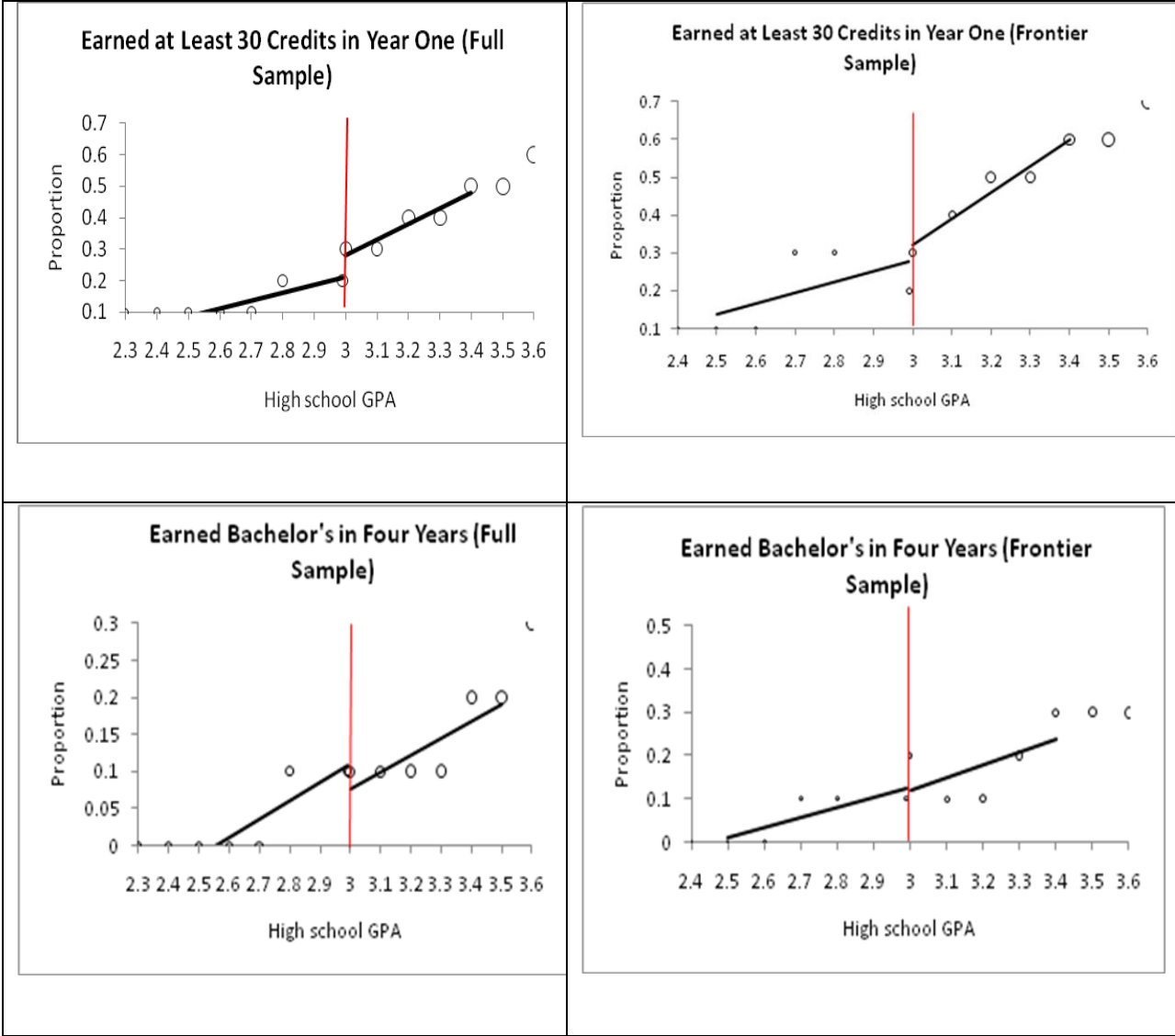


Figure 2. Selected outcomes by high school GPA for full and frontier samples