RESEARCH PROPOSAL ON THE IMPACT OF HOUSING STATUS ON FIRST-YEAR COLLEGE STUDENT ACADEMIC ACHIEVEMENT

Yeran Tong

INTRODUCTION & OBJECTIVES

American higher education has been making efforts to increase the integration of undergraduate learning experiences, especially in combining the curricular and cocurricular worlds for students in recent history (Sriram, 2011). Among those efforts, Colleges and universities have started requiring traditional degree-seeking students to live on campus during their first year to provide a more immersive experience for freshmen. Dated back to 1993, several studies have investigated the effects of living on-campus on students' performance. Thompson et al. (1993) find that freshmen who live on campus are more likely to remain in school, make more progress in their program, and have better academic performance. Pascarella et al. (1993) also examined first-year students but explicitly focused on measures of critical thinking, reading comprehension, and mathematical skills. They find first-year students who lived on campus made more considerable gains in critical thinking skills, but no significant difference was found in terms of reading comprehension and mathematical skills compared to students who lived off-campus.

However, nowadays, with the rapid popularization of social media, streaming platforms, and wireless internet, it is much easier to interact with family and friends outside of the campus, which means students may not rely heavily on their living environment for social interaction than in the past (Gemmill & Peterson, 2006). And meanwhile, the internet nowadays is abundant in educational resources, which further mitigate the necessity of living on-campus. On the other hand, there is no guarantee that on-campus residents will attend student activities or athletic or cultural events (Gemmill & Peterson, 2006). With all being said, a re-examination is urgent to be conducted to check if the positive impact of living on-campus still exists in the age of information.

RESEARCH QUESTIONS

1. Will the housing status affect the academic performance of a first-year university student nowadays?
2. How does living on-campus and off-campus respectively influence academic performance? And will this effect persist throughout the first academic year?
3. Are there effects heterogeneity across the student's pre-enrollment status or non-STEM and STEM major?

THEORETICAL FRAMEWORK

In 1937, American Council on Education had propounded Student affairs theory and practice. The theory is grounded on the belief that opportunities for learning outside of the classroom are abundant (American Council on Education, 1937), and a great many of these experiences are shaped by the environments and programs provided to those living on-campus students. In debted to the theory, Pascarella and Terenzini (1991) hypothesized that residencing on campus was "the single most consistent within-college determinant of the impact of college." Since then, multiple related studies have been carried out and have indicated that students' daily and academic lives would be greatly influenced by the housing status they choose. Due to more immersive residential experiences provided by on-campus living, more pervasive communication and interactions are likely to take place among peers within the institution. Apart from that, students living on-campus tend to commit to campus learning associattion and have higher frequency engage with faculty through in-hall study groups and guest lectures, enriching their academic life beyond the classroom setting (Gemmill & Peterson, 2006). These students are timelier in their graduation and more often go on to graduate school and earn advanced degrees. Living on campus gives students an academic edge by earning higher grades, providing opportunities for learning communities and faculty access, and promoting their future growth by helping them stay connected to the college environment.

Campus housing effectively integrates learning and social development by providing students the opportunity to form an identity or a sense of community with the institution. Students who live on campus are found generally participate in more campus activities, take advantage of campus resources, and are more involved in leadership experiences (Gemmill & Peterson, 2006). The interaction that students have within the residence hall often frames their campus experience. Living in a community that offers all of the components of a small neighborhood assists students in gaining a better understanding of themselves and the diverse population that comprises their neighborhood. It also offers them informal and formal leadership opportunities on their floors or wings and within the entire residential and institutional community (Gemmill & Peterson, 2006).

DATA

Data used for this study is student-level records extracted from the Kaggle database. The dataset contains a total of 3159 completed individual first-year college students' records from an unknown university. The dataset covers students' demographic information (age when enrolled in university, gender, race, parents' education level), pre-enrollment academic records (SAT score or ACT score, high school name and high school GPA), first academic year academic records (major, minor, course taking status and final grades), housing status and socio-economic status (financial need of students, family contribution towards course fees, unmet financial need of the student). Student race information is coded into dummy variables in terms of White, Black, Hispanic, and Others. As for pre-enrollment college entrance exam scores, some students provided ACT scores, and some others submitted their SAT scores. To address that issue, all ACT scores were equated to SAT scores. A STEM major indicator column was later added to the dataset according to students' major information. However, it is common for a new freshman enter the university without deciding their major and the major information is marked as "Undeclared" in the major column. As result of this, there are three unique values in STEM major indicator column, which are 1 standing for STEM major, 0 standing for non-STEM mahor, and 2, standing for major undecide. The primary outcome of interest is how students' first semester GPA, second semester GPA, and cumulative GPA of whole academic year were related to the housing choices they made (on-campus living or off-campus living). And thus, housing status would be regraded as treatment.

METHODOLOGY

The main objective of this study is to produce a bias-reduced estimation of the effect of on-campus living on first-year students' academic performance. Due to students making their own decision about their residency type (artificial randomization not involved), the given records suggest a slightly unbalanced observation of the on-campus and off-campus living students (41.7% of students choose to live on-campus, and 58.3% of students prefer living off-campus). And thus, there might be concerns raised owing to the possible systematic bias between the on-campus and off-campus groups. Typically, previous research has indicated that students' first-year academic success is highly positively correlated with their pre-enrollment grades (High School GPA, ACT and SAT scores (Radunzel and Noble, 2011). From this view, model estimation should control for these potential systematic differences between control and treated students before measuring the effect of housing status.

To approach the estimation problem, this study utilizes doubly robust propensity score weighting (PSW) techniques to account for selection assignment differences between treatment and comparison groups so that bias-reduced Average Treatment Effects on Treated (ATT: [E(r1 −r0|b(x),z = 1)] = E(r1|b(x),z = 1) −E(r0|b(x),z = 1)]) can be appropriately derived. The propensity score is developed from the ideal concept of the balancing score ***b*(x)**, and is denoted by:

$$e(x) = pr(z = 1|x)$$

The propensity score gives the probability of being exposed to treatment 1 given the observed covariates x, and the treatment condition is assumed to be independent of the set of observed covariates x. In this study, propensity score $e(x)$ is obtained by regressing students' demographic information, pre-enrollment academic records, first-year GPA, and STEM subject indicator onto the treatment variable 'housing status' using the ps function in the TWANG package. In the ATT estimation, the characteristics of controlled participants will be reweighted to mirror treated counterparts' characteristics. That means all treated participants would be multiplied by a weight of one, and controlled participants would be multiplied a weight of $\frac{e(x)}{1-e(x)}$. After adjustments, the controlled group resembles the treated group in terms of the observed covariates.

Beyond the regular propensity score weighting process, a multilevel approach is used in this study to further investigate the cluster effect brought about by students' pre-enrollment academic performance. This strategy accounts for the potential confounding nature of initial academic achievement by including high school grade fixed effects in the models predicting the likelihood of experiencing the treatment. By clustering students with similar college entrance exam scores, we can reduce bias in the estimate attributed to pre-enrollment academic level. For this part, college entrance exams scores were cut into quarters, and thus, four grade clusters were formed. Matching algorithm would only match treated and controlled within the cluster at this stage.

RESULTS AND DISCUSSION

Table 1 gives the summary statistics of the treated (living on campus) and controlled (living off campus) groups. From the table, it is evident that baseline equivalence between controlled and treated was not initially established. The first row indicates that roughly 60% of the students decided to live off-campus while 40% of students choose to live on campus. Beyond the difference in the group number, the demographic constitution and pre-enrollment academic records of the treated group differ significantly from the controlled group indicated by the p-value on the right side of Table 1. The male-female ratio approaches 0.44 for the treated group, while this ratio is much higher in the controlled group, 0.67. Apart from that, white people only make up 48% of the population in the treated group, whereas, in the controlled group, the white people population constitutes 70% of the group size. Most importantly, the difference in pre-enrollment academic records between the two groups is the major source of bias in effect estimation, as first-year GPA is theoretically highly correlated with high school grades. In general, the treated group has poorer average high school academic performance in terms of high school GPA and college entrance exam score.

At first, without adjusting for the systematic difference between the two groups, two naïve fixed-effect models were applied to investigate the effects of housing status. The first model is a simple fixed-effect ANOVA where first-term GPA, second-term GPA, and cumulative GPA were regressed on the treatment indicator only. The second model incorporates all the useable covariates into the regression process, and the results of the two models are given in table 2 and table 3, respectively. According to the tables, when baseline equivalence was not achieved, none of the models suggest that the housing status is a significant predictor in predicting the GPA. In the second naïve model, with all covariates incorporated in the regression, pre-enrollment academic performance turned out to be a crucial factor in predicting the first year of academic success. Both high school GPA and school entrance exam scores are statistically significant at $p < 0.01$ level, which is consistent with the assumption.

Table 4 shows the summary statistics of two groups after propensity score weighting. As briefly stated above, individuals in the controlled group (off-campus living group) were reweighted according to their propensity score so that the propensity score distribution of the controlled group would resemble the treated group. From Table 3, the significant differences between all covariates have been eliminated by the reweighting process. Compared ct.mn column with the first tr.mn column, p-values are far beyond the rejection level for all covariates. Based on this information, any potential difference in the effect of housing status is assumed not to be driven by participants' baseline disparities. Hence, a reduced bias estimation was able to be made. Table 5 showed an ANOVA result when baseline disparities were eliminated. Worth noticing is that the housing status becomes a significant predictor in predicting academic achievement and is likely to improve GPA by 0.07 on average if students live on-campus. Once all covariates were used in the regression model, the findings provided more evidence to support the original assumptions. In Table 6, with the exception of the STEM subjects indicator, all other covariates turned out to be statistically significant in this model. As anticipated, pre-enrollment academic performance is still positively correlated with the cumulative GPA. Housing status, in this model, tends to contribute more to the improvement of cumulative GPA than the previous model shows, with living on-campus enhancing the overall GPA by 0.135 on average. At last, table 7 gives the regression result after a within grade-cluster matching, a multilevel approach. However, the statistical analysis does not indicate any sign that the coefficient is significant, which suggests that the effects of living on campus on academic performance might be homogeneous across the span of college entrance exam scores.

SIGNIFICANCE OF THE STUDY

This study performed a re-examination of the effectiveness of living on campus on first year undergraduate students' academic performance using latest student level administrative records. This report provides bias reduced evidence that living on-campus is in favor of the academic success for freshmen students by using matching and weighting techniques that manipulate the pre-enrollment academic records for controlled and treated group. Furthermore, the research result responses to the question that whether popularization of the internet products and social media would potentially undermine the mechanisms of how living on-campus benefits students. The study result also produces some important implications in school policy. Schools can appeal to new students to live on campus for their first year or mandate first year students live on-campus to help them make the transition from high school to college.

The study also has highlighted limitations that other similar studies should void. The student records are from single universities, so the sample's representativeness is a concern. Plus, from the student records, college entrance exam scores of those students who appeared in the dataset remain below-average to the average level. It is likely that living on campus has a more substantial influence on those students but less effective on a group of students above the average high school graduates academic level. More future research has to be done to investigate the ceiling effects of living on campus.

Appendix

Table 1: Summary Statistics

	Off Campus	On Campus	p
N	1843	1316	
Average Student Age	18.01	17.96	0.01
% of Male	41	31	0.00
% of Female	59	69	0.00
% of White	70	48	0.00
% of Non-White	30	52	0.00
Average HS GPA	3.26	3.18	0.00
Average exam score	1007.92	992.78	0.00

Table 2: Naive Model 1

	Dependent variable:		
	culGPA	gpa1	gpa2
	(1)	(2)	(3)
HOUSING_STS	−0.027	−0.036	−0.004
	(0.025)	(0.029)	(0.031)
Constant	2.924***	2.922***	2.856***
	(0.016)	(0.018)	(0.020)
Observations	3,159	3,159	3,159
R^2	0.0004	0.001	0.00001
Adjusted R^2	0.0001	0.0002	−0.0003
Residual Std. Error (df = 3157)	0.692	0.793	0.866
F Statistic (df = 1; 3157)	1.173	1.581	0.016
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 3: Naive Model 2

	Dependent variable:		
	culGPA	gpa1	gpa2
	(1)	(2)	(3)
STDNT_AGE	0.039**	0.055**	0.016
	(0.019)	(0.023)	(0.025)
STDNT_GENDER	−0.138***	−0.157***	−0.133***
	(0.022)	(0.026)	(0.029)
STDNT_TEST_ENTRANCE_COMB	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)
HIGH_SCHL_GPA	0.723***	0.742***	0.794***
	(0.023)	(0.028)	(0.031)
isSTEM	−0.025*	−0.042**	−0.010
	(0.015)	(0.017)	(0.019)
HOUSING_STS	0.026	0.016	0.055*
	(0.021)	(0.025)	(0.028)
Constant	−0.732**	−1.144***	−0.544
	(0.362)	(0.429)	(0.478)
Observations	3,159	3,159	3,159
R^2	0.296	0.246	0.215
Adjusted R^2	0.295	0.245	0.214
Residual Std. Error (df = 3152)	0.581	0.690	0.768
F Statistic (df = 6; 3152)	221.006***	171.672***	144.000***
Note:	*p<0.1; **p<0.05; ***p<0.01		

	tx.mn	ct.mn	p	ct.mn	p
STDNT_AGE	17.96	17.96	0.7820	18.01	0.01
STDNT_GENDER	0.31	0.31	0.7460	0.41	0.00
STDNT_TEST_ENTRANCE_COMB	992.78	995.44	0.6020	1007.92	0.00
HIGH_SCHL_GPA	3.18	3.18	0.7370	3.26	0.00
isSTEM	0.53	0.53	0.9350	0.66	0.00
white	0.48	0.50	0.3610	0.70	0.00
black	0.44	0.42	0.4110	0.17	0.00
hisp	0.06	0.05	0.7060	0.08	0.01
other	0.02	0.03	0.7940	0.04	0.00

Table 4: Balance of the treatment and comparison groups

Table 5: Robust Model

	Dependent variable:
	culGPA
HOUSING_STS	0.070^{***}
	(0.025)
Constant	2.827^{***}
	(0.018)
Observations	3,159
R^2	0.003
Adjusted R^2	0.002
Residual Std. Error	0.620 (df = 3157)
F Statistic	8.155^{***} (df = 1; 3157)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: Doubly Robust Model

	Dependent variable:
	culGPA
HOUSING_STS	0.135***
	(0.032)
STDNT_AGE	0.072**
	(0.036)
STDNT_GENDER	−0.136***
	(0.034)
STDNT_TEST_ENTRANCE_COMB	0.001***
	(0.0001)
HIGH_SCHL_GPA	0.701***
	(0.040)
isSTEM	−0.021
	(0.025)
Constant	−1.483**
	(0.649)
Observations	3,159
Log Likelihood	−3,648.853
Akaike Inf. Crit.	7,311.707
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Multiple Approach Model

	Dependent variable:
	culGPA
HOUSING_STS	0.065
	(0.043)
Constant	2.832***
	(0.124)
Observations	3,159
Log Likelihood	−3,593.633
Akaike Inf. Crit.	7,191.266
Note:	*p<0.1; **p<0.05; ***p<0.01

References

American Council on Education. (1937). The student personnel point of view. Washington, DC: Author

De Araujo, P., & Murray, J. (2010). Channels for Improved Performance from Living on Campus. American Journal of Business Education, 3(12), 57-64.

Gemmill, E. L., & Peterson, M. (2006). Technology use among college students: Implications for student affairs professionals. NASPA Journal, 43(2), 280–300. doi:10.2202/0027 6014.1640

Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

López Turley, R. N., & Wodtke, G. (2010). College residence and academic performance: Who benefits from living on campus?. Urban Education, 45(4), 506-532.

Pascarella, E. T., & Terenzini, P. T. (1991). How college affects students: Findings and insights from twenty years of research. San Francisco, CA: Jossey-Bass.

PAUL R. ROSENBAUM, DONALD B. RUBIN, The central role of the propensity score in observational studies for causal effects, Biometrika, Volume 70, Issue 1, April 1983, Pages 41–55, https://doi.org/10.1093/biomet/70.1.41

Polly A. Graham, Sarah Socorro Hurtado & Robert M. Gonyea (2018). The Benefits of Living on Campus: Do Residence Halls Provide Distinctive Environments of Engagement?, Journal of Student Affairs Research and Practice, 55:3, 255-269, DOI: 10.1080/19496591.2018.1474752

Radunzel, J., & Noble, J. (2012). Predicting Long-Term College Success through Degree Completion Using ACT [R] Composite Score, ACT Benchmarks, and High School Grade Point Average. ACT Research Report Series, 2012 (5). ACT, Inc

Sriram, R., Shushok Jr, F., Perkins, J., & Scales, T. L. (2011). Students as Teachers: What Faculty Learn by Living on Campus. Journal of College and University Student Housing, 38(1), 40-55.