2022

Analyzing College On-Campus Residency Agglomeration Benefits and Their Impact on Student Outcomes

Emily Pedace
Scripps College

Analyzing College On-Campus Residency Agglomeration Benefits and Their Impact on Student Outcomes

By
EMILY PEDACE

Submitted to Scripps College in Partial Fulfillment of the Degree of Bachelor of Arts

PROFESSOR NICHOLAS KACHER
PROFESSOR NAYANA BOSE

November 27th, 2021

Table of Contents

Abstract

Economies of agglomeration have been shown to increase productivity in part due to technological/ knowledge spillovers. I look at public colleges and universities in California and Texas to examine if students living on-campus can generate enough agglomeration to boost productivity in the students. I use a unique data set with information collected from Common Data Sets, self-reported by the schools. When using an OLS regression with the averages of these variables, higher averages of on-campus residence is correlated to higher average graduation and retention rates for Texas and California. When using a fixed effects panel regression, I find changes in on-campus residency to be statistically insignificant for changes in six-year graduation rates, and statistically significant for changes in retention rates for both states. This determined that increases in on-campus residence positively affects retention rate.

Introduction

Economic agglomeration refers to the clustering of economic activity and the benefits that (in localization economies) are driven by pooled labor markets, specialized inputs and services, and technological/knowledge spillovers. There exists significant literature that finds evidence of agglomeration increasing regional productivity in cities. If this phenomenon can be generated at colleges and universities—like a smaller scale city—policy makers, college boards, and students can utilize this to generate increasing returns to students. This paper will explore if colleges with higher percentages of their student body living on campus or in university-owned housing experience agglomeration benefits of knowledge spillovers that impact student outcomes.

This research question provides opportunity to alter how students make their final college decisions: if knowledge spillover benefits are seen, the percentage of a school's student body living on campus can be indicative of the style of learning offered on a campus. For individuals with different learning habits, a community similar to cities that has high interpersonal relationships may be more beneficial than other college specific variables (such as student to faculty ratios). Additionally, findings could indicate a need for policy makers and school boards to invest in more college-owned housing. If causation is found between students living in school housing and superior student outcomes, then, over time this offers the possibility for higher rankings and acknowledgement of prestige. If the results are significant, potential economic gains could also arise from taking advantage of on-campus clustering boosting student productivity—if schools can produce better student outcomes by increasing on-campus housing this can have an impact of higher productivity in the labor markets their students join after school. Therefore, not only does this provide the opportunity to benefit students, but economic growth as well.

Using a uniquely created data set, this paper finds statistically significant results that on campus housing has positive effects on retention rates, but a statistically insignificant effect on a college's six-year graduation rate. This research question was analyzed by compiling information on public colleges and universities (17 in California and 10 in Texas) over multiple years into one data set including: on campus residence; six-year graduation rate; retention rate; total number of degree-seeking undergraduates; student to faculty ratio; percent of STEM degrees awarded; and percent of Business/Marketing degrees awarded. I begin with OLS regressions to first determine levels of correlation. Then, panel regressions with fixed effects were used with six-year graduation rate and retention rates serving as the measures for student outcomes. Looking at how changes in these independent variables affects changes in student outcomes with this model considers differences between individual colleges.

This paper uses theoretical agglomeration theories with a focus on the knowledge spillover approach, which relates firms' competitiveness and success to localized patterns of tacit (in person) knowledge creation, knowledge sharing, innovation, and learning. College and universities offer the perfect environment to study agglomeration, since the focus is tacit education. This approach, while driven on firms, will be applied to the clustering of students since they generate similar localized patterns centering around information.

Student housing is an important aspect for an individual when choosing which college or university to attend. However, even if there are found to be significant benefits for on-campus housing at schools, there are issues of space and funding. Many policy makers and school boards have grappled with issues of funding for more housing with increases in student enrollment. Recently, University of California Berkeley (not included in my data set for lack of reported years) has had significant issues with a lack of on-campus housing supply. Not only is U.C. Berkeley narrowly able to house twenty-two percent of their undergraduates, but housing in the surrounding area is extremely costly—especially for college students. Despite this, attempts by the college to add more on-campus housing has received resistance from organizations who want to enforce California's environmental laws (The New York Times, 2022).

If high rates of on-campus residency do improve the productivity and student outcomes for a school, this can provide even more support for policies that promote an increase in on-campus housing construction. Also, even if a school is well established and prestigious, such as the University of California Berkeley, inability to provide all their students with the experience of living at the school with knowledge spillovers that may contribute to an individual's success might rightfully sway a student's decision to attend.

Over the course of this thesis, I will review literature that provides the theoretical groundwork on agglomeration. This section will also cover additional knowledge spillover research that supports universities being an important source of this spillover, as well as literature that provides the basis for some of the control variables in the coming model. The next section will contain the specifics on data collection, the use and importance of the panel regression model, another additional OLS model, and results. The last section will conclude the paper with the importance of both the statistically significant and insignificant results, as well as provide insight for possible future research and policies.

Literature Review

A multitude of economic theories focus on explaining economic agglomeration, or the clustering of economic activities in specific geographic locations.

The theoretical framework places significant importance regarding firm competitiveness on recurring knowledge creation and sharing, innovation, and learning (Bekele and Jackson, 2006). Tacit knowledge spillovers are the exchanging of ideas that require face-to-face information sharing with common cultural, linguistic, and societal norms. They discuss how these tacit, in person, knowledge spillovers are particularly important which leads to the emphasis of location and proximity. A critical distinction referenced by the authors are localization and urbanization economies: urbanization economies attribute advantages gained by clustering from all firms, overall size, and diversity; localization economies refer to firms in a specific sector receiving benefits from other same-sector firm clusters.

The concept of geographical proximity being an important factor relates to the work of Audretsch and Feldman (1996). They control for the geographical concentration of production and look at three sources of economic knowledge—industry research and development, skilled labor, and university research—using OLS and 3SLS methods of data analysis. They find that the tendency for innovation activity to cluster is higher in industries where knowledge plays an important role. These findings provide justification for this paper's topic centering around colleges and universities because knowledge externalities are of extreme value in those communities. Since their work shows activity clusters near these sources of knowledge, I will look directly at sources of knowledge to examine if the external benefits produced from these sources can be found internally.

In contrast to Audretsch and Feldman (1996) measuring clustering by geographical sections of land (such as states or cities), the work by Ciccone and Hall (1993) focuses on providing evidence for density as the major driver for productivity. Though the authors do find some outliers, density has a positive impact on employment and gross output. While this work uses location (city bounds) as a measure to look for knowledge spillover—opposed to others using population density to measure this—the findings are conclusive with other literature since cities contain uniquely high levels of human population density. With this as a basis, the empirical analysis of this paper will focus on the density of students at their respective schools, as opposed to the size of the plot of land a school has.

With respect to both population density and land coverage, universities are significantly smaller than cities. Liu (2015) and Kantor and Whalley (2009) look at spillover effects from a university on its surrounding economy. Liu (2015) uses the exogenous variable of the 1862 Land-Grant universities with synthetic control counties to provide the outcome trajectory a country would have had without the land-grant university. Since the Land-Grant focused on providing education to the agricultural and industrial society, they look at manufacturing effects. They found that after 80 years, long-run manufacturing output per worker had greatly increased, generating a robust finding that these results are an impact of direct spillovers from universities and agglomeration economies.

Kantor and Whalley (2009) also use non-educational labor markets in their empirical analysis along with research universities. Since knowledge sharing is a primary function of a school's economy, Liu (2015) and Kanton and Whalley (2009) use colleges and universities when examining knowledge spillovers. A differentiating factor in the Kantor and Whalley (2009) study is they use exogenous price changes (stock market shocks and initial market values of endowment values) along with endogenous institutional expenditures and initial endowment market values to explore possible agglomeration benefits found in average labor income by county. While in agreeance with Liu and statistically significant, the positive effects university knowledge spillovers have on local economies are small.

In addition, Kantor and Whalley (2009) do find positive effects calculated solely from general city size as well. One possible reason for these findings being less drastic than in Liu (2015) is the difference in having knowledge spillovers being the only variable examined for causality— whereas Liu (2015) has results that incorporate population increase effects, which can also generate agglomeration benefits (Ciccone and Hall, 1993). Since information spillover effects have a negative relationship with distance to density—and studies use universities as the center of the measure—it can be expected that our results will find higher positive effects within the university than those seen by Kantor and Whalley (2009) on the cities surrounding universities. It is important to note that in both literature, labor markets are used; since college and university economies likely have very different characteristics, this can be reasoning for this paper diverging in empirical results.

The type of human capital being shared can impact the quality of productivity derived from knowledge spillovers, as certain sectors of education may benefit more than others from tacit knowledge spillovers. Liu (2017), Hansen, Owan, and Pan (2006), and Patton (2015) explore how certain college majors and dynamics of sharing information play a role in the positive causation between knowledge spillovers and agglomeration economic benefits. Liu (2017) examines full-time workers with their corresponding college majors and Metropolitan Statistical Areas (serving as labor market bounds) along with urban wage premiums to find productivity differences. Since Liu (2015) found direct spillover effects from universities on cities in part due to population density, this study is in agreeance with others that localization effects promote productivity far more than urbanization effects.

A main finding of their paper is workers with a bachelor's degree from a STEM field has large within-field agglomeration benefits—simultaneously finding other degrees to have little to no within-field agglomeration benefits. Human capital from STEM fields has such large impacts on their surrounding economies that strong heterogeneity in positive knowledge spillover effects from STEM majors to non-STEM majors is seen. Consistent with their within-field results for non-technical categories of human capital, across-field knowledge spillover benefits do not arise from the presence of non-STEM majors.

Patton (2015) investigates and attempts to use different empirical data to reproduce similar findings from papers on human capital relating to economic benefits. Some results they find strongly concur with Liu (2017), as their research concludes that individuals with STEM field backgrounds have the greatest to gain from other STEM field localization. Another strong finding in line with Liu (2017) is any field of study or work has the greatest benefits from STEM field interactions.

This implies that colleges with more students studying in these sectors will generate greater benefits to the surrounding college community when there are high percentages of this type of student body living on campus, since this tacit spillover is so valuable. Therefore, the percentage of students in a college or university that are majoring in science, technology, engineering, or mathematic fields will be controlled for in this paper's model, though the main variable is on-campus residency. To reiterate another justification for this paper finding slightly different results, these studies focusing on STEM majors and fields all focus on the labor market, which has different incentives and goals than the economies found in colleges and universities. Economies at schools focus more on generalized knowledge, whereas labor markets are more targeted and consistent in the information necessary to succeed. These studies also failed to address the quality and type of school where the majors and degrees were received. Furthermore, Patton (2015) finds urbanization effects to be stronger than localization effects. This contradicts Liu (2015) whose findings show the inverse.

Hansen, Owan, and Pan (2006) look at interaction factors that might affect the increases in productivity from information spillovers. They use a business management course at Washington University to assess the effects different characteristics (such as race, gender, or age) have on both group and individual performance in the classroom. Liu (2015) and Patton (2015) generally regard business and economic fields as more technical than other majors, and therefore estimate a positive spillover effect greater than the little-to-no effect seen in arts, history, and social sciences (excluding economics from social sciences). They find that gender diversity is useful in groups with well-educated students. While the reason for this is up for debate, I find this part of the study slightly flawed. Part of their study controls for self-selection using exogenous group assignments, but I believe this finding is partly due to the self-selection of students to attend co-ed schools. Students who learn more efficiently in their same gender dominated spaces may choose to attend a college that has gender demographics that meet this need. Similar self-selection may occur for individuals who benefit from diverse gender environments, so I think claims on gender diversity may contain self-selection bias.

In addition, Hansen, Owan, and Pan (2006) did not detect group performance variances due to racial diversity. I find the use of one school and one college course too small of a sample to accurately predict if this finding will hold for all schools—unfortunately, accessible data will not allow my empirical analysis to account for student body racial compositions. Hansen, Owan, and Pan (2006) utilize SAT scores to proxy for an individual's skill since some of their knowledge spillover results were found to be correlated with smarter students. SAT scores are becoming more and more obsolete since they have been criticized as perpetuating inequality such that many schools have stopped requiring them as part of the admissions process and thus have no SAT data to report on their student bodies. Therefore, my data analysis will instead look at changes that occur within a college or university, since variation in educational ability should be less within a school than across all schools in the study.

As seen, agglomeration is vital to the growing and successful economies in cities. Literature shows the positive knowledge spillover benefits colleges and universities generate for their surrounding cities. Along with significant lack of literature of cities generating spillover effects for colleges, I was also unable to find regional economic literature that analyzed agglomeration benefits within colleges. This paper contributes to the literature by analyzing whether agglomeration can be found at school levels, opposed to the literature that only looks at data on city-sized levels. In addition to giving new insight on the amount of density needed for an economy of agglomeration, I will examine if the variables other researchers found to boost in and out-field knowledge spillovers (such as STEM majors) have an impact at this smaller scale. If high percentages of students living on campus can serve as a smaller scale city and generate agglomeration benefits, this can be used to advance significant growth in students and colleges.

Data

Collection Methods and Descriptive Statistics:

The data for the empirical analysis of this paper is collected from college and university self-reported Common Data Sets. This is part of the Common Data Set Initiative, which is a collaborative effort from those who provide data on higher education to provide transparency and accurate information—all the schools are required to use the same data template to report their results. These results are available for free online to the public. This initiative was created in 1997 and collects detailed information under these main categories: General Information, Enrollment, First-Time, First-Year Admission, Transfer Admission, Academic Offerings and Policies, Student Life, Annual Expenses, Financial Aid, Instructional Faculty and Class Size, and Degrees Conferred. However, many colleges and universities have extreme variations in the years the started participating in the Common Data Set Initiative.

I used specific information from these Common Data Sets to compile a unique, conglomerated data set that contains observations on 17 California and 10 Texas public universities and colleges over multiple years. The average number of years collected for California public schools is 13.06 with a minimum number of years collected for a school being 7 (starting at 2013), and a maximum being 21 years (starting at 1999). For Texas public schools the average number of years collected is 16.6 with a minimum number of years collected being 10 (starting at 2010), and a maximum being 20 (starting at 2000). I did not include 2020 college and university data to avoid my data being affected by the COVID-19 shock on all schools. All but one school had data up to 2019.

Table One: California Colleges and Respective Years

College/University	First Year of Data	Last Year of Data
California State University San Bernardino*	2005	2019
San Diego State University	2009	2019
California State Polytechnic University Pomona	2002	2018
California State University Los Angeles	2008	2019
University of California Los Angeles	2002	2019
University of California San Diego	2005	2019
University of California Riverside	2010	2019
University of California Merced	2005	2019
San Jose State University	2000	2019
California Poly State University San Luis Obispo	1999	2019
California State University San Marcos	2000	2019
California State University Monterey Bay	2007	2019
California State University Long Beach	2012	2019
University of California Santa Cruz	2013	2019

University of California Irvine	2013	2019
California State University Fresno	2012	2019
California State University Fullerton	2013	2019

*California State University San Bernardino had missing data for the years 2008 and 2009. This table contains the 17 California public colleges and universities in which data was collected from for this paper's analysis. It details the name of the college along with their first year of available data, last year of available data, and missing years of data. If a college was missing aspects of data for a certain year but did not miss all the data for said year, it will not be listed in the missing year.

Table Two: Texas Colleges and Respective Years

College/University	**First Year of Data**	**Last Year of Data**
University of Texas at Austin	2000	2019
University of Texas at Dallas	2010	2019
University of Houston	2001	2019
University of North Texas	2002	2019
University of Texas at Tyler	2002	2019
Sam Houston State University	2006	2019
University of Texas at San Antonio	2002	2019

Tarleton State University	2004	2019
Texas Tech University	2005	2019
Texas A&M University	2002	2019

This table contains the 10 Texas public colleges and universities in which data was collected from for this paper's analysis. It details the name of the college along with their first year of available data, last year of available data, and missing years of data. If a college was missing aspects of data for a certain year but did not miss all the data for said year, it will not be listed in the missing year.

In my model, the dependent variables are:

- Six-year graduation rates: This uses the cohort year that is six years prior to the data release year. For example, for year 2019, it is the six-year graduation rate of the 2013 cohort. This is the total number of students graduating within six years divided by the final 2013 cohort after adjusting for allowable exclusions[1].
- Retention rate: This is the percentage of first years who continue next year at the school. For example, for year 2019, this is the percentage of full-time bachelor's degree-seeking undergraduates who were first-years in 2018 and were still currently enrolled by Fall 2019.

I chose to use two dependent variables to ensure that my model could capture any possible benefit the student may have. The six-year graduation rate encompasses possible productivity boosts that assist students in the completion of their learning. Retention rates can be

[1] Allowable exclusions include death, permanent disability, service in the armed forces, and foreign aid service of the federal government or official church missions.

representative of an agglomeration economy providing enough knowledge spillover that a student is able to continue at a school. I believe the combination of the two adequately cover the main ways a student could be successful in college. In addition, the other variables that could possibly proxy for student outcomes (such as a student's grade point average) did not have accessible data. Both dependent variables are reported as percentages.

My model consists of five independent variables:

- On-Campus Residence: This variable is given as a percentage and is the number of degree-seeking undergraduates that live in college owned, operated, or affiliated housing, divided by the total number of degree-seeking undergraduates.
- Student to Faculty Ratio: This is always reported as number of students to one faculty member. This ratio is in terms of full-time plus one-third part-time with respect to students and instructional faculty—this does not include teaching assistants as faculty.
- Total Undergraduates: This is the number of enrolled students at a college or university that are actively seeking a degree for that year.
- STEM Degrees Awarded: This is reported as a percentage and was measured by degrees per major, not headcount. Students who double majored and received one degree are represented twice in the report. This reports the degrees conferred between July 1st of the prior year and June 30th of the reported year. I calculated the total percentage, and for the purposes of this paper a STEM degree consists of the following majors: are Natural Resources/ Environmental Science, Computer and Information Sciences, Engineering, Engineering Technologies, Biological/ Life

Sciences, Mathematics and Statistics, Physical Sciences, and Health Professions and Related Sciences.

- Business/Marketing Degrees Awarded: This was also reported as a percentage, using majors not persons as the counting measure. This means that if one-person double majors, it reflects as two separate degrees awarded. This is the percent of degrees conferred between July 1st of the prior year and June 30th of the reported year.

On-campus residence is the variable that I predict, through previously mentioned economic theory, will be statistically and economically significant in explaining variance in our student outcome dependent variables. Previously mentioned literature by Liu (2015), Kantor and Whalley (2009), and Ciccone and Hall (1993) that find positive knowledge spillover effects from density and population provide the basis for controlling for overall degree-seeking undergraduates. So, while I propose that the economy of agglomeration and subsequent student outcomes generated at colleges and universities is more dependent on the percentage of student body living on campus (the main independent variable), overall population is still likely to be statistically significant.

I control for STEM degrees since Liu (2017) and Patton (2015) find STEM fields create, both for themselves and others, the highest amount of knowledge spillover effects. In the discussed literature, economics and business majors were also estimated to have positive spillover effects given that they are more technical than other majors. For this reason, business degrees are controlled for in my model. Since economic degrees are reported in the Common Data set under social science degrees conferred, they could not be included due to other social science degrees being found to have no spillover effects. Some of the Common Data Sets for certain years contained information on some variables while others were left unaccounted for.

This is seen in Table 3 as the different variables have different numbers of observations.

Table Three: California Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
On-Campus Residence	198	24.516	18.018	.1	99
Graduation Rate	214	62.488	17.152	27	91.45
Retention Rate	218	86.301	6.872	64	97.1
Number of Undergraduates	222	19331.527	8479.977	130	34921
Student Faculty Ratio	215	22.008	3.683	12	30
STEM Degrees Awarded	210	29.829	13.229	4.8	66
Business Degrees Awarded	210	17.473	9.025	0	42.4

This table shows the number of observations, mean, standard deviation, minimum, and maximum for both the independent and dependent variables for the schools in California.

Table Four: Texas Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
On-Campus Residence	160	19.29294	6.243453	3	35
Graduation Rate	161	54.0825	16.69984	25	85.6
Retention Rate	157	78.36503	11.19898	56	95.7
Number of Undergraduates	164	24628.41	12057.17	2973	53202
Student Faculty Ratio	160	20.58919	2.648388	14	30
STEM Degrees Awarded	164	24.965	10.42927	7.64	51.7
Business Degrees Awarded	164	21.74494	6.51235	0	38.78

This table shows the number of observations, mean, standard deviation, minimum, and maximum for both the independent and dependent variables for the schools in Texas.

Texas schools see a lower mean on-campus residence than schools in California. The minimum on-campus residency for the two states are similar, both nearing zero. However, the maximum reported on-campus residence for a California school is 99% which is remarkably higher than the maximum of 35% for Texas schools. It is important to note that neither of these states have high averages—this may affect the regression results, as I predict impacts from high rates of this variable and my data is limited to means far under 50%. My dependent variables have a difference of less than 10% when comparing the means across states. This would coincide with my prediction of correlation between my dependent variables and on-campus residency, since the on-campus residency across states also had mean values with less than a 10% difference.

Model and Results:

In order to look for correlation between my independent variables (on-campus residence, student to faculty ratio, number of undergraduates, STEM degrees awarded, and Business degrees awarded) and my dependent variables (six-year graduation rate and retention rate), I use the following OLS models:

OLS Models:

$$\widehat{Six - Year\ Graduation\ Rate} = \beta_0 + \beta_1 On - Campus\ Residence + \beta_2 Student\ to\ Faculty\ Ratio + \beta_3 Number\ of\ Undergraduates + \beta_4 STEM\ Degrees\ Awarded + \beta_5 Business\ Degrees\ Awarded$$

$$\widehat{Retention\ Rate} = \beta_0 + \beta_1 On - Campus\ Residence + \beta_2 Student\ to\ Faculty\ Ratio + \beta_3 Number\ of\ Undergraduates + \beta_4 STEM\ Degrees\ Awarded + \beta_5 Business\ Degrees\ Awarded$$

Since I have panel data, I generated new variables that represent the mean for each independent and dependent variable, aggregating the data at the university level. These OLS regressions examine if average higher rates of on-campus residency have higher average graduation and retention rates.

Table Five: California and Texas OLS Regressions for Graduation and Retention Rates

	California Graduation Rates	California Retention Rates	Texas Graduation Rates	Texas Retention Rates
On-Campus Residence	.403*** (.035)	.107*** (.018)	1.539*** (.078)	.81*** (.068)
Number of Undergraduates	.001*** (0)	.001*** (0)	.001*** (0)	.001*** (0)
Student to Faculty Ratio	-2.061*** (.215)	-.524*** (.111)	.233 (.342)	-.269 (.3)
STEM Degrees Awarded	.056 (.036)	.101*** (.018)	.672*** (.056)	.246*** (.049)
Business Degrees Awarded	-.195*** (.069)	-.008 (.036)	-.592*** (.109)	.247** (.095)
R-sq	.913	.847	.904	.837
N	222	222	166	166

**** p<.01, ** p<.05, * p<.1*

This table shows the coefficients and standard errors in parenthesis for the four OLS regressions.

The results of this regression on California schools find on-campus residence to be statistically significant, with a one-unit increase in on-campus residence correlating to a .4 increase in a school's graduation rate. This is consistent with my hypothesis that higher rates of students living in college or university owned housing will produce agglomeration benefits to the school's population that boost student outcomes. Consistent with what the literature would predict, the number of undergraduates, Business degrees awarded, and student to faculty ratio were all statistically significant. STEM degrees awarded were statistically insignificant—

contrary to the literature that anticipated this variable would be more significant than Business degrees since they are more technical than the latter. While the p-value was .12, it would still be economically insignificant since it had a very small coefficient.

In agreeance with the California data, when running an OLS regression for the averages of our variables from the Texas data, on-campus residence was extremely statistically significant with a coefficient of 1.54. This shows an even greater increase in a college's average graduation rate due to on-campus residency than found with the California data. The only variable not statistically significant is the student to faculty ratio. Opposite to my expectations and as found in my previous regressions, Business degrees awarded have a negative coefficient.

When running the OLS regression for retention rates, there were many similarities in the results for California and Texas: on-campus residence, number of undergraduates, and STEM degrees awarded were found to be statistically significant. An important difference between the two is the coefficient for on-campus residence for Texas was nearly eight times larger than the reported coefficient for California. This would suggest that while both states on average have higher retention rates when there are higher on-campus residency rates, the relation is more important depending on state circumstances.

It is important to note that this OLS model cannot account for time invariant differences—this means that if the behavior of some of the variables is not affected by time, this is not controlled for. If a college or university has constant high rates of both on-campus residency and six-year graduation or retention rates due to an external factor specific to the school (this could include anything from higher endowments, more selective acceptance processes, etc.) some of this correlation between average six-year graduate rates and on-campus residency has the potential to be due to omitted variable bias. Despite this, since literature shows multiple colleges and universities produce agglomeration benefits to their surrounding economies, I still find the correlation between the two variables in the OLS results economically significant. To account for the possible biases of this model, I use panel regressions with fixed effects to examine if causality can be attributed to on-campus residency rates and student outcomes in conjunction to the found correlation.

Panel Fixed Effects Models:

$$Six - Year\ \widehat{Graduation}\ Rates = \beta_{0+}\beta_{it1}On - Campus\ Residence + \beta_{it2}Student\ to\ Faculty\ Ratio + \beta_{it3}Number\ of\ Undergraduates + \beta_{it4}STEM\ Degrees\ Awarded + \beta_{it5}Business\ Degrees\ Awarded$$

$$\widehat{Retention}\ Rates = \beta_{0+}\beta_{it1}On - Campus\ Residence + \beta_{it2}Student\ to\ Faculty\ Ratio + \beta_{it3}Number\ of\ Undergraduates + \beta_{it4}STEM\ Degrees\ Awarded + \beta_{it5}Business\ Degrees\ Awarded$$

In order to control for other factors that can affect student outcomes and create bias (such as different levels of endowment, different college resources, or different skill levels of accepted students), I use a panel regression model with college as my fixed effect. This looks at changes in six-year graduation and retention rates (student outcomes) for each college over several years as a function of changes in percentage living in college owned, operated, or affiliated housing; total degree-seeking undergraduates; student to faculty ratio; percent of STEM degrees awarded; and percent of Business degrees awarded. This regression will look at if changes within on-campus residence affects changes within our dependent variables. Fixed effects are essential to control for differences in location opportunities at an even more precise level than state, and issues that might arise with different colleges having different levels of resources or endowment. Two different panel regressions are employed (with no time lags) to examine six-year graduate rates and retention rates separately.

Table Six: California and Texas Fixed Effects Panel Regressions for Graduation and Retention Rates

	California Graduation Rates	California Retention Rates	Texas Graduation Rates	Texas Retention Rates
On-Campus	-.021	.045**	.065	.154**
Residence	(.035)	(.021)	(.087)	(.068)
Number of	.002***	0***	0***	0**
Undergraduates	(0)	(0)	(0)	(0)
Student to	.121	.337***	.486***	.045
Faculty Ratio	(.173)	(.103)	(.161)	(.131)
STEM Degrees	-.138	.001	.29***	.157***
Awarded	(.112)	(.047)	(.072)	(.057)
Business	-.511***	-.274***	-.153	-.053
Degrees	(.157)	(.071)	(.102)	(.08)
Awarded				
R-sq	.53	.468	.525	.599
N	177	180	154	152

****p<.01, **p<.05, *p<.1*

This table shows the coefficients and standard errors in parenthesis for the four fixed effects panel regressions.

For both California and Texas, counter to my predictions, on-campus residence is statistically insignificant on graduation rates. However, also for both states, looking at graduation rates the number of undergraduates at a college or university is statistically significant. This is anticipated by the literature which shows high density populations generate agglomeration benefits. Although it is shown as significant from the corresponding p-values, the coefficients are small enough that the number of undergraduates at a school is unexpectedly not economically significant.

Differences amongst the states begin to appear when analyzing the other independent variables in the graduation rate regressions. For the California schools, the only other statistically significant variable is Business degrees awarded. Similar to the three of the four OLS regressions, the coefficient for this variable counter literature by showing a negative relationship. The only OLS regression that shows statistical significance and a positive relationship for Business degrees awarded is when using Texas data and the dependent variable of retention rates. The fixed effects panel regression for Texas graduation rates portrays the opposite, with student to faculty ratio and STEM degrees awarded being statistically significant.

While only statistically significant for the California retention rate regression and Texas graduation rate regression, the student to faculty ratio having a positive coefficient is an interesting result. As the student to faculty ratio increases, the number of students on average in a classroom is increasing. Generally, smaller classrooms are thought of as creating a better learning environment for students since it allows more personalized learning. In fact, some prestigious schools have strict limits on the number of students granted access into a course— U.S. News and World Report even uses student to faculty ratios as an aspect of their school ranking system. So, while this positive relationship is unforeseen, it does on a miniscule scale follow patterns of agglomeration that say benefits arise from more density.

The findings that are conclusive to my hypothesis of high rates of on-campus residency providing agglomeration benefits to the school's community of students is seen with the student outcome of retention rates. California and Texas both had statistically significant on-campus residence with this dependent variable (retention rates). The coefficient for Texas is higher than that of California, showing a greater importance depending on the state. As with graduation rates, when using retention rates the number of undergraduates is statistically significant but with a coefficient too small to attribute economic importance.

When looking at results from the retention rate regression for California, student to faculty ratio and Business degrees awarded are also statistically significant. For Texas, the only other statistically significant variable is STEM degrees. It is important to note with these findings that there are limited consistencies across these regressions. Regardless of the student outcome being measured, only Business degrees awarded are consistently statistically significant for California; for Texas it is STEM degrees awarded. For across state consistencies, the only statistically significant commonality is the positive relationship with on-campus residency and retention rates.

Conclusion

Interpreting the Results:

When using the OLS regression to look at whether average high average rates of on-campus residence have average high graduation rates, it is statistically significant that a one unit increase in on-campus residence has a .40 (California) and 1.54 (Texas) increase in graduation rates. Since it would only take an estimate of a 2.5 increase in on-campus residence to increase graduation rates by one unit in California (which would be one percent) this is economically significant. The OLS regression on average high rates of on-campus residence and retention rates finds a one-unit increase in the percentage on-campus residence has a .11(California) and .81 (Texas) increase in retention rates.

However, the use of panel regressions shows that the on-campus residence percentage is only statistically significant for retention rates, with a one unit increase in on-campus residence producing a .05 (California) and .15 (Texas) increase in retention rate percentage. On-campus residence was statistically insignificant in the panel regression for graduation rates in both states. These results would tentatively indicate that high percentages of students living in college owned, operated, or affiliated housing do not cause agglomeration benefits that positively impact a school's graduation rates. It would also suggest that the knowledge spillover effect that is produced from agglomeration in these localization school economies has a very small positive impact on retention rates. It would require approximately a 23 unit increase in on-campus percentage to produce a full one unit increase in retention percentage—indicating a lack of economic significance.

Implications and Future Research:

Being that average high rates of on-campus residency does appear to generate knowledge spillovers that positively impact average graduation rates but on-campus residency has slight positive impacts on retention rate, it might prove to be beneficial for colleges to invest in more quality on-campus housing.

It is important to note that the average on-campus residency percentage in this data set is approximately 24.5 (California) and 19.29 (Texas). The basis of this argument uses high percentages of residency. Since this is lacking from most of our observations, using colleges in California or other states that have higher rates for this explanatory variable could provide further insight on the causal effects of localization economies that result in universities and colleges. This would ensure the lack of statistical significance for a change in on-campus residence affecting a change in graduation rate is not due to a possible threshold limit. However, considering there are low rates of on-campus residency, and it was still statistically significant for some of the regressions, it presents motivation to look for agglomeration benefits at other categories of colleges and universities that have higher rates.

In 2021 to 2022, California Governor Gavin Newsom has proposed increasing funding by 39.6 billion dollars to California public universities and schools (Burke, 2022). A caveat of this funding is in part that schools must increase their graduation rates (along with reducing college costs, eliminating achievement gaps, and prepping students for the workforce). While further research is needed, using the results of my OLS regression, it would appear useful for schools to use part of this additional funding for housing. As discussed earlier with University of California Berkeley, lack of housing supply is a large issue for schools in California. Not only would using this funding for building more on-campus housing help to resolve that issue, but if there were indicators it would also boost average graduation rates it would help meet the program requirements to continue receiving this funding.

For Texas' public higher education funding, one of the more recent increases in financial support was in 2020 with the Governor's Emergency Education Relief Fund (McGee, 2022). This program set aside 307 million dollars, however, since this program was focused on COVID-19 relief the main goal was for schools to distribute this funding to the students. Not only have Texas institutions had difficulty adequately transferring this money to students due the strict regulations surrounding the program, but it focuses on students having the ability to continue to enroll in college. Based off the positive relationships found with increases in on-campus residency positively affecting changes in retention rates, this might indicate support for programs that give funding to Texas public colleges and universities for more housing instead of focusing on what is currently a challenging plan of distributing money directly to students.

References:

Audretsch, D.B. and Feldman, M.P. (1996) ‘R&D Spillovers and the Geography of Innovation and Production’, *The American Economic Review*, 86(3), pp.630-640.

Bekele, W.G. and Jackson, R. (2006) ‘Theoretical Perspectives on Industry Clusters’, *Regional Research Institute Publications and Working Papers*, 92, pp.1-26.

Burke, M., 2022. *Newsom Offers New Money IF California College Systems Meet Equity Goals.* Edsource. Available at: https://edsource.org/2022/how-gov-newsoms-budget (Accessed: 28 April 2022).

Ciccone, A., Hall, R. E., and National Bureau of Economic Research (1993) *Productivity and the Density of Economic Activity*. Cambridge, MA: National Bureau of Economic Research (NBER working paper series, no. 4313).

Hansen, Z. K. *et al.* (2006) *The Impact of Group Diversity on Performance and Knowledge Spillover--An Experiment in a College Classroom*. Cambridge, Mass.: National Bureau of Economic Research (NBER working paper series, no. 12251).

Kantor, S. and Whalley, A. (2009) *Do Universities Generate Agglomeration Spillovers? Evidence From Endowment Value Shocks*. Cambridge, Mass: National Bureau of Economic Research (NBER working paper series, no. w15299).

Liu, S. (2015) “Spillovers from Universities: Evidence from the Land-Grant Program,” *Journal of Urban Economics*, 87, pp.25–41.

Liu, S. (2017) “Agglomeration, Urban Wage Premiums, and College Majors,” *Journal of Regional Science*, 57(4), pp. 611–630.

McGee, K., 2022. *Texas Created a Program to Help Students Reenroll in College During the*

Pandemic. Schools Struggled to Find Students Who Qualified. Texastribune.org. Available at: https://www.texastribune.org/2022/03/24/texas-college-students-reskilling-grant/ (Accessed: 28 April 2022).

Patton, M. and University of Alabama. Department of Economics, Finance, and Legal Studies (2015) *Essays on Human Capital Heterogeneity and Agglomeration*. Dissertation. University of Alabama Libraries.

The New York Times (2022) *Berkeley vs Berkeley is a Fight Over the California Dream.* Available at: https://www.nytimes.com/2022/03/10/us/uc-berkeley-student-housing.html (Accessed: 26 April 2022).